

Computational Study of Conformational Changes in Nuclear Receptors Upon Ligand Binding

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Azam Rashidian
aus Mamasani/Iran

Tübingen
2023

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der
Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:

07.11.2023

Dekan:

Prof. Dr. Thilo Stehle

1. Berichterstatter:

Prof. Dr. Antti Poso

2. Berichterstatter:

Prof. Dr. Stefan Laufer

3. Berichterstatter:

Prof. Dr. Paavo Honkakoski

Contents

Contents.....	i
I. Summary.....	iii
II. Zusammenfassung.....	v
III. List of publications.....	vii
IV. Author contributions.....	viii
V. List of Poster presentations.....	xv
VI. List of abbreviations.....	xvi
VII. Introduction.....	1
1. Nuclear Receptors.....	1
1.1 Structure and Activity.....	1
1.2 Nuclear receptor modulators.....	7
1.3 Retinoid X receptor.....	9
1.4 Pregnane X Receptor.....	10
1.5 Constitutive Androstane Receptor.....	13
1.6 Farnesoid X receptor.....	14
2. In silico Drug discovery and Molecular Modelling.....	16
2.1 3-D structure of the target and druggability.....	18
2.2 Homology Modelling.....	19
2.3 Protein and ligand preparation.....	22
2.4 Force field.....	24
2.4.1 Geometry optimization.....	26
2.5 Molecular docking and pose selection.....	27
2.5.1 Limitation of molecular docking.....	30
2.6 Molecular Dynamic Simulations and its challenges.....	31
2.7 Analysis of MD simulation trajectories.....	35
2.7.1 Principal Component Analysis.....	35
2.7.2 Markov State Modelling.....	36
2.7.3 Free energy (prime MM-GBSA).....	38
2.7.4 Interaction analysis.....	40
2.7.5 Angle and distance calculation.....	41
2.7.6 Ligand trajectory analysis.....	42
3. Experimental Methods to Analyse NR Activity.....	43
3.1 Time Resolved Fluorescence resonance energy transfer.....	43
3.2 Luciferase reporter gene transactivation assay.....	44
3.3 Coregulator-recruitment (mammalian two-hybrid, CARLA).....	44

4. Concluding Remarks and Outlook.....	45
5. References.....	48
6. List of figures.....	76
VIII. Result and Discussion	77
7. Publication I.....	77
8.Publication II.....	111
9. Publication III.....	154
10. Publication IV	181
11. Publication V [Submitted Manuscript].....	219
IX. Acknowledgements.....	258

I. Summary

The nuclear receptor superfamily (NR), as a major group of intracellular receptors, regulate broad aspects of cell functions including cell growth, differentiation, and metabolism in distinct organs. The activation of these receptors is regulated by endogenous or exogenous lipophilic compounds and regulatory proteins. They share a highly similar structure, particularly, in ligand binding domains (LBD) and DNA binding domains (DBD). Having DBD reveals their genome transcriptional role as they are known as transcriptional factors. These features all reveal NRs remarkable role in organism survival (within the scope of managing metabolic rates, energy stores, salt homeostasis, responding to exogenous toxins, and inflammation to regulate growth, reproduction, and development) and highlight them as promising targets for therapeutic development.

Meanwhile, advancements in computer-aided drug discovery including target identification and validation, high-throughput virtual screening, ADME and Toxicity prediction, lead optimization and molecular dynamic simulations have facilitated the time-consuming and expensive conventional methods of drug discovery.

Leveraging the computational advancements in the drug discovery process, our research on NRs has led to four publications, with two being authored as the first author. Additionally, a manuscript has been mentioned in the results and discussion section. In our projects, we employed microsecond-long all-atom molecular dynamics (MD) simulations with relevant systems to investigate the dynamic behaviour and conformational changes induced by ligand binding.

Our first publication focused on our newly identified dual kinase and Pregnane X Receptor (PXR) inhibitor, where we observed ligand-specific influence on conformations of different PXR-LBD regions, including $\alpha 6$ region, α AF-2, $\alpha 1$ - $\alpha 2'$, $\beta 1'$ - $\alpha 3$ and $\beta 1$ - $\beta 1'$ loop. The insight into conformational behaviour of PXR-LBD promotes PXR antagonism. This work was complemented by an additional co-authored virtual screening-based paper where we describe the identification of the C100 compound ¹. We employed *in silico screen* and experimental cellular reporter assay to identify small molecule kinase inhibitor from Tübingen kinase inhibitor collection (TÜKIC) compound library which act also as PXR inhibitor. Further biochemical binding and cellular protein interaction assays categorized the novel compounds as mixed competitive/noncompetitive, passive antagonist which disrupt PXR coregulatory

binding by direct binding to PXR. The results show the possibility of dual PXR, and kinase inhibitors which could be beneficial in cancer treatment.

In the third publication, we focused on the conformational behaviour of the Farnesoid X receptor (FXR), comparing known agonists to our newly discovered antagonists. We studied the antagonist-induced conformational changes in the FXR ligand-binding domain, also comparing monomer against FXR/RXR heterodimers, in the presence and absence of the coactivator peptides. This work provides new insights into the conformational behaviour of FXR.

In the fourth publication, we helped in the design of a series of novel compounds acting as CAR agonists and selective toward the PXR. The behaviour of these novel compounds and their induced CAR conformation was studied through MD simulation besides in vitro assays.

Finally, in our manuscript, we examined interactions of individual branched 4-nonylphenols (22NP, 33NP, and 353NP) and linear 4-nonylphenol (4-NP) with (wt)CAR1 and its variant CAR3 using molecular dynamics (MD) simulations, and cellular experiments. To the best of our knowledge, this is the first work of describing interactions of individual 4-nonylphenoles in detail with the human CAR receptor and its dominant variant CAR3. We hope this work contributes to the safer use of NPs.

II. Zusammenfassung

Als eine bedeutende Gruppe intrazellulärer Rezeptoren, reguliert die Superfamilie der nukleären Rezeptoren (NRs) eine Vielzahl von Zellfunktionen in verschiedenen Organen, einschließlich Zellwachstum, Differenzierung und Stoffwechsel. Die Aktivierung dieser Rezeptoren wird sowohl durch endogene oder exogene lipophile Verbindungen als auch durch regulatorische Proteine gesteuert. Sie besitzen äußerst ähnliche Strukturen, insbesondere hinsichtlich ihrer Ligandenbindenden Domänen (LBD) und DNA-bindenden Domänen (DBD). *Durch die DBD wird ihre genomische transkriptionelle Rolle offenbart, da sie als Transkriptionsfaktoren bekannt sind.* All diese Eigenschaften zeigen die bemerkenswerte Rolle der NRs beim Überleben von Organismen (im Rahmen der Regulierung von Stoffwechselraten, Energiespeichern, Salz-Homöostase, Reaktion auf exogene Toxine und Entzündungen zur Regulierung von Wachstum, Fortpflanzung und Entwicklung) und machen sie zu vielversprechenden Zielen für die therapeutische Entwicklung.

Mittlerweile haben Fortschritte in der computergestützten Wirkstoffforschung, vor allem im Hinblick auf Targetidentifizierung und -validierung, virtuellem Hochdurchsatz-Screening, ADME- und Toxizitätsprognose, Leitstrukturoptimierung und molekulardynamischen Simulationen, die zeitaufwändigen und kostspieligen konventionellen Methoden der Wirkstoffforschung vereinfacht.

Durch die Nutzung der oben genannten computergestützten Fortschritte in der Wirkstoffforschung fand die in dieser Dissertation behandelte Arbeit Eingang in vier Veröffentlichungen, von denen zwei in Erstautorschaft verfasst wurden. Des Weiteren wurde ein Manuskript im Abschnitt Ergebnisse und Diskussion erwähnt. In unseren Projekten verwendeten wir Molekulardynamik-Simulationen (MD) mit relevanten Systemen, um das dynamische Verhalten und die Konformationsänderungen durch die Bindung von Liganden zu untersuchen.

In der ersten Veröffentlichung konzentrierten wir uns auf unseren neu identifizierten dualen Kinase- und Pregnane-X-Rezeptor (PXR)-Inhibitor, bei dem wir einen ligandenspezifischen Einfluss auf die Konformationen der verschiedenen PXR-LBD-Regionen beobachteten, einschließlich der $\alpha 6$ -Region, $\alpha AF-2$, $\alpha 1-\alpha 2'$, $\beta 1'-\alpha 3$ und $\beta 1-\beta 1'$ -Schleife. Der Einblick in das Konformationsverhalten von PXR-LBD fördert den PXR-Antagonismus. Diese Arbeit wurde durch eine zusätzliche Veröffentlichung in Co-Autorschaft ergänzt, in der wir einen Virtual-screening-basierten Ansatz

beschreiben, in dem wir die Identifizierung der C100-Verbindung ¹ beschreiben. Wir verwendeten ein In-silico-Screening und experimentelle zelluläre Reporterassays, um einen niedermolekularen Kinaseinhibitor aus der Tübinger Kinaseinhibitor-Sammlung (TÜKIC) als PXR-Inhibitor zu identifizieren, der auch als PXR-Inhibitor wirkt. Weiterführende biochemische Bindungs- und zelluläre Proteininteraktionsassays kategorisierten die neuartigen Verbindungen als gemischt kompetitive/nichtkompetitive, passive Antagonisten, die die PXR-Co-Regulator-Bindung durch direkte Bindung an PXR stören. Die Ergebnisse zeigen die Möglichkeit von dualen PXR- und Kinaseinhibitoren, die in der Krebsbehandlung von Vorteil sein könnten.

In der dritten Veröffentlichung konzentrierten wir uns auf das Konformationsverhalten des Farnesoid-X-Rezeptors (FXR) und verglichen bekannte Agonisten mit unseren neu entdeckten Antagonisten. Wir untersuchten die antagonistisch induzierten Konformationsänderungen in der Ligandenbindungsdomäne von FXR und verglichen auch Monomere mit FXR/RXR (Retnoid X Rezeptors) -Heterodimeren, in Anwesenheit und Abwesenheit der Coaktivatorpeptide. Diese Arbeit liefert neue Erkenntnisse zum Konformationsverhalten von FXR.

In der vierten Veröffentlichung haben wir bei der Konzipierung einer Serie neuartiger Verbindungen mit CAR-Agonisten-Wirkung und Selektivität für PXR mitgewirkt. Das Verhalten dieser neuartigen Verbindungen und die induzierte Konformation des humanen CAR wurden mittels MD-Simulationen und In-vitro-Assays untersucht.

Schließlich haben wir in unserem Manuskript die Interaktionen einzelner verzweigter 4-Nonylphenole (22NP, 33NP und 353NP) und linearen 4-Nonylphenols (4-NP) mit (wt)CAR und seiner Variante CAR3 mithilfe von Molekulardynamiksimulationen und zellulären Experimenten untersucht. Nach unserem besten Wissen handelt es sich dabei um die erste Arbeit, die die Interaktionen einzelner 4-Nonylphenole detailliert mit dem humanen CAR-Rezeptor und seiner dominanten Variante CAR3 beschreibt. Wir hoffen, dass diese Arbeit zu einem sichereren Einsatz von NPs beiträgt.

III. List of publications

Publication I

Discrepancy in interaction and conformational dynamics of pregnane X receptor (PXR) bound to an agonist and a novel competitive antagonist.

Publication II

Target Hopping from Protein Kinases to PXR: Identification of Small-Molecule Protein Kinase Inhibitors as Selective Modulators of Pregnane X Receptor from TüKIC Library.

Publication III

When Two Become One: Conformational Changes in FXR/RXR Heterodimers Bound to Steroidal Antagonists.

Publication IV

Discovery of Receptor Agonists with the Imidazo[1,2-*a*]pyridine Structure Novel Human Constitutive Androstane.

Publication V [submitted manuscript]

Filling the blank space: Branched 4-nonylphenols isomers are responsible for 2 robust constitutive androstane receptor (CAR) activation by nonylphenol

IV. Author contributions

Publication I

Discrepancy in interaction and conformational dynamics of pregnane X receptor (PXR) bound to an agonist and a novel competitive antagonist

Azam Rashidian

Conceptualization and investigation
Methodology and formal analysis, Data curation,
Writing original draft preparation,
Review and editing, visualization.
Discussion of the results and interpretation

Dr Thales Kronenberger:

Conceptualisation and methodology
Review and editing
Discussion of the results and interpretation

Prof. Dr. Stefan A. Laufer

Study supervision
Proofreading and final approval of the manuscript

Dr. Tatu Pantsar:

Study supervision
Conceptualisation and methodology
Data curation and visualisation
Discussion of the results and interpretation
Corresponding author

Other co-authors: Experimental assays, experimental data visualization, writing, review and editing. All authors have read and agreed to the published version of the manuscript.

Publication II

Target Hopping from Protein Kinases to PXR: Identification of Small-Molecule Protein Kinase Inhibitors as Selective Modulators of Pregnane X Receptor from TükIC Library

Enni-Kaisa Mustonon

Conceptualization and investigation
Methodology, formal analysis, Data curation,
Writing original draft preparation,
Review and editing, visualization.
Discussion of the results and interpretation

Azam Rashidian

Investigation
Methodology
Visualization
Writing the manuscript, review and editing

Prof. Dr. Stefan A. Laufer

Study supervision
Resources, access to TükIC
Proofreading and final approval of the manuscript

Dr. Tatu Pantsar:

Conceptualisation and methodology
Data curation and visualisation
Discussion of the results and interpretation
Corresponding author

Prof. Dr. Oliver Burk/experimental

Conceptualisation and methodology
Validation, formal analysis, investigation

Resources, data curation

Writing original draft preparation,

Study supervision

Proofreading and final approval of the manuscript

Other co-authors: Experimental assays, experimental data visualization, writing, review, and editing. All authors have read and agreed to the published version of the manuscript.

Publication III

When Two Become One: Conformational Changes in FXR/RXR Heterodimers Bound to Steroidal Antagonists

Azam Rashidian

Investigation

Methodology and formal analysis,

Writing the manuscript, review and editing

Discussion and interpretation

Students' supervision and shared orientation

Alejandro Díaz-Holguín

Investigation

Methodology and formal analysis,

Visualization

Review and editing

Prof. Dr. Antti Poso:

Study supervision

Proofreading and final approval of the manuscript

Dr Thales Kronenberger:

Study supervision

Conceptualisation and methodology

Data curation and visualisation

Discussion of the results and interpretation

Corresponding author

Other co-authors: Writing, review, and editing. All authors have read and agreed to the published version of the manuscript.

Publication IV

Discovery of Receptor Agonists with the Imidazo[1,2-a]pyridine Structure Novel Human Constitutive Androstane

Ivana Mejdrová

Conceptualization and investigation
Methodology, formal analysis, Data curation,
Writing original draft preparation,
Review and editing, visualization.
Discussion of the results and interpretation

Azam Rashidian

Investigation
Methodology and formal analysis,
Data curation, visualization
Writing the manuscript, review and editing

Dr Thales Kronenberger:

Conceptualisation and methodology
Discussion of the results and interpretation
Review and editing

Prof. Dr Petr Pavek:

Study supervision
Resources
Proofreading and final approval of the manuscript

Radim Nenka:

Study supervision
Conceptualisation and methodology
Data curation and visualisation
Discussion of the results and interpretation
Corresponding author

Other co-authors: Experimental assays, experimental data visualization, writing, review, and editing. All authors have read and agreed to the published version of the manuscript.

Publication V [submitted manuscript]

Filling the blank space: Branched 4-nonylphenols isomers are responsible for robust constitutive androstane receptor (CAR) activation by nonylphenol

Azam Rashidian

Investigation

Methodology and formal analysis,

Writing the manuscript, review, and editing

Discussion and interpretation

Students' supervision and shared orientation

Students' supervision and shared orientation

Prof. Dr. Antti Poso:

Study supervision

Proofreading and final approval of the manuscript

Dr Thales Kronenberger:

Study supervision

Conceptualisation and methodology

Data curation and visualisation

Discussion of the results and interpretation

Prof. Dr Petr Pavek:

Study supervision

Proofreading and final approval of the manuscript

Resources

Other Co-authors: Experimental assays, experimental data visualization, writing, review, and editing. All authors have read and agreed to the published version of the manuscript.

V. List of Poster Presentations

DPHG (German Pharmaceutical Society) Annual meeting 2021, virtual event, 29-1 October.

Investigating the conformational changes of pregnane X receptor Ligand Binding Domain upon binding Endocrine Disrupting chemicals

Azam Rashidian, Thales Kronenberger, Tatu Pantsar, Antti Poso

23rd European Symposium on Quantitative Structure-Activity Relationship (23rd EuroQSAR), 2022, 26-30 September, Heidelberg, Germany.

Conformation is the key: Discovery of PXR agonist Vs competitive antagonist

Azam Rashidian, Enni-Kaisa Mustonen, Thales Kronenberger, Matthias Schwab, Oliver Burk, Stephan A Laufer, Tatu Pantsar

VI. List of abbreviations

3-D	Three-dimensional structure
ADME	Absorption, distribution, metabolism, and excretion
AF-1	Activation Function 1
AF-2	Activation Function 2
AI	Artificial intelligent
AR	Androgen receptor
BPA	Bisphenol A
CADD	Computer-aided drug design
CAR	Constitutive androstane receptor
CoA	Coactivator
CoR	Corepressor
DDI	drug–drug interactions
DDT	Dichlorodiphenyltrichloroethane
DES	Diethylstilbestrol
ED	Endocrine Disruptor
EDCs	Endocrine Disrupting Chemicals
EE2	17- α -ethinylestradiol
EGF	Epidermal growth factor receptor
ER α , β	Estrogen receptor α , β
FDA	Food and Drug Administration
FLIP	Fluorescence Loss in Photobleaching
FRAP	Fluorescence Recovery After Photobleaching
FXR	Farnesoid X receptor
GPCR	G-protein-coupled receptor
HTS	High throughput screening
HREs	Hormone response elements
HTVS	High throughput virtual screening
LBD	Ligand Binding Domain
LBDD	Ligand-based drug design
LBP	Ligand Binding Pocket
LXR α , β	Liver X receptor

MD	Molecular dynamics
MHC	Major histocompatibility complex
MM	Molecular Mechanics
MSMs	Markov State models
NCOR	Nuclear Receptor Coregulator
NMR	Nuclear magnetic resonance
NPT	Isothermal-Isobaric Ensemble
NR	Nuclear Receptor
NR-CoA	Nuclear Receptor Coactivator
NVE	Microcanonical ensemble
NVT	Canonical ensemble
PBC	Periodic boundary conditions
PCA	Principal component analysis
PCBs	Polychlorinated biphenyls
PES	Potential Energy Surface
PDB	Protein Data Bank
PPARs	peroxisome proliferator-activated receptors
PR	Progesterone receptor
PXR	Pregnane X receptor
QM	Quantum Mechanics
QSAR	Quantitative Structure-Activity Relationships
RAR	Retinoic acid receptor
REMD	Replica exchange molecular dynamics
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
RP	Ramachandran plot
RXR	Retinoid X receptor
SAR	Structure-activity relationship
SASA	Solvent-accessible surface area
SBDD	Structure-based drug discovery
SMRT	Silencing mediator of retinoic acid and thyroid hormone
SP	Standard precision docking
SPC	Simple Point Charge

SPC/E	Simple Point Charge/Extended
SRC1	Steroid receptor coactivator 1
SXR	Steroid and xenobiotic receptor
TBT	Tri-butyl-tin
TIP3P	Transferable Intermolecular Potential 3 Points
TIP4P	Transferable Intermolecular Potential 4 Points
TNC	Transnanochlor
TR	Thyroid hormone receptor
TR-FRT	Time-resolved fluorescence energy transfer
TüKIC	Tübingen Kinase Inhibitor Collection
VDR	Vitamin D receptor
XP	Extra precision docking

VII. Introduction

1. Nuclear Receptors

1.1 Structure and Activity

Nuclear receptors represent, besides other drug targets such as G protein-coupled receptors, ion channels, receptor tyrosine kinases and immunoglobulin-like receptors, a major receptor target class for drug development. Human Nuclear Receptors (NRs) are a superfamily of intra-cellular receptors consisting of 48 members. However, the number of functionally different NR proteins is by far larger, due to alternative splicing processes and post-translational modification, such as ubiquitination and phosphorylation². In 1974, the correlation between hormone action and alterations in the gene expression status was reported³. Later studies revealed the now-called classic model of the NR signalling pathway⁴. The first NRs were cloned and investigated in 1985, this represented the starting point of modern NR research⁵⁻⁷. Subsequently, additional NRs were identified⁸⁻¹⁰ (Fig. 1) and by now the family is composed of over 500 members spread among several metazoan species¹¹

Based on their mechanism of action and ability to bind to DNA, NRs are categorized into four groups¹²: types I, II, III and IV.

Type I, which belongs to subfamily 3¹³, are steroid hormone receptors. These receptors are homodimers existing in the cytoplasm and upon ligand binding, they are translocated to the nucleus and bind to specific sequences of DNA, known as hormone response elements (HREs), inverted repeat (Fig.1). Estrogen receptor (E α , E β), Androgen receptor (AR) Progesterone receptor (PR) and Glucocorticoid (GR) belong to this type.

Type II, belongs to subfamily 1¹⁴, which is found in the nucleus such as the Thyroid hormone receptor (TR), Retinoic acid receptor (RAR), Peroxisome proliferator-activated receptors (PPARs), Liver X receptor (LXR α , LXR β), Farnesoid X receptor (FXR), Vitamin D receptor (VDR), Pregnane X receptor (PXR) and Constitutive androstane receptor (CAR) heterodimerizing with Retinoic X receptor (RXR). Type II is usually complexed with corepressor proteins in the absence of ligand and binding to DNA direct repeat. Upon binding ligand, protein conformation is changed, leading to dissociation of corepressor protein and recruitment of coactivator. Subsequently,

this complex, in addition to other transcriptional machinery components, transcribes DNA.

Type III, subfamily 2¹⁵, are like type I as being homodimers but in contrast, they bind to DNA direct repeat. No ligand has been identified in this group.

Type IV nuclear receptors have the ability to bind to DNA in either a monomeric or dimeric form¹³. One representative member of this group is Steroidogenesis Factor-1 (SF-)¹¹. Like type III receptors, these receptors do not have any known natural ligands, the reason that they are referred to as orphan nuclear receptors. Both type III and type IV receptors are still not well understood in terms of their function and structure.

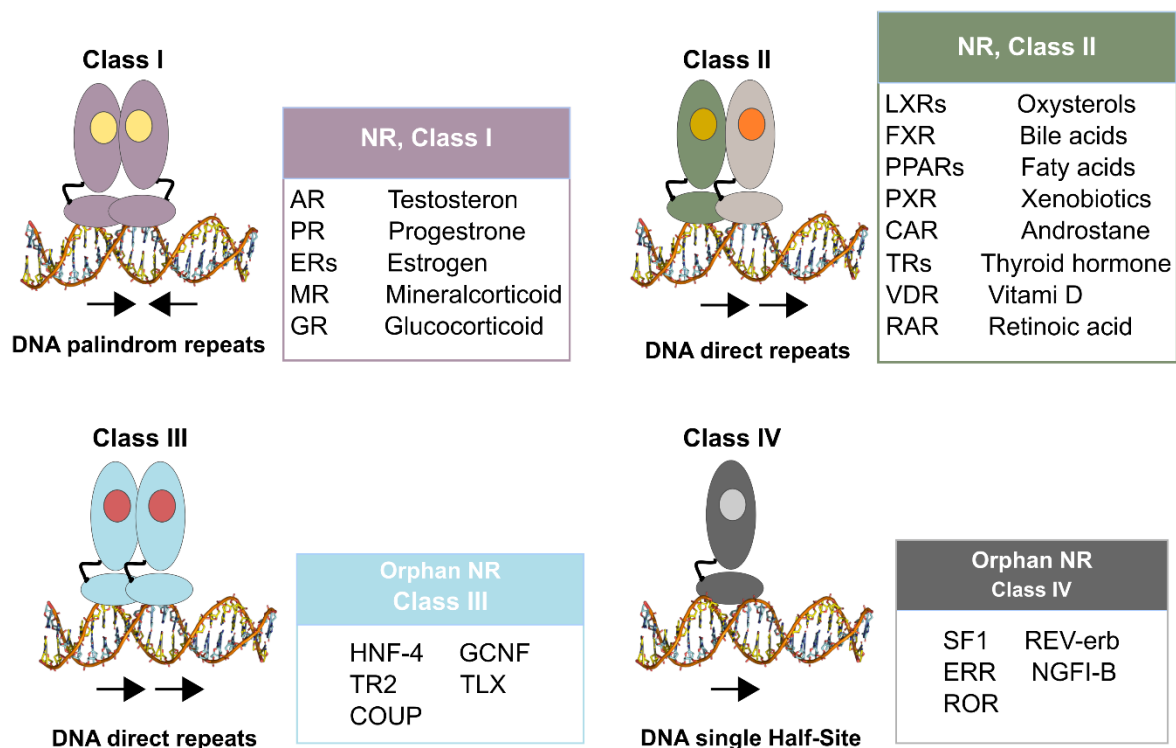


Fig. 1 Overview of the nuclear receptor superfamily. Four superfamilies of nuclear receptors are represented based on dimerization, DNA binding (direct or inverted repeat), and ligand specificity (required or not required); Class I: Steroid receptor (also known as hormone receptor); Class II: RXR heterodimers; Class III and Class IV: dimeric and monomeric orphan receptors, respectively, with unknown ligands.

All NRs have a similar structural organization and typically contain five structural domains as follows (Fig. 2):

1. N-terminal domain: which varies considerably among the receptors and is commonly unstructured; it contains a transactivation domain known as Activation Function 1 (AF-1) and is ligand-independent.
2. DNA-binding domain (DBD): which is highly conserved across various NR receptors, this region has four cysteines that coordinate to two zinc atoms which bind to DNA response elements (*e.g.*, DBD functions in a post-translational modification which happens at Thr38 in CAR ^{16,17}).
3. hinge: a highly flexible connecting region believed to regulate the cellular distribution of the NR.
4. the ligand binding domain (LBD): comprised by a very conserved bundle eleven α -helices, where the ligand binding pocket is located, However, the interior of the ligand-binding pockets exhibits significant variation, enabling nuclear receptors to bind a diverse array of endogenous and synthetic ligands ^{18,19}.
This binding capacity extends to include the activation function-2 (also referred to as the α AF-2 helix) and the three-stranded β -sheet, except for PXR. The eleven α -helices can be categorized into three distinct groups: H1/H3, H4/H5/H8/H9, and H7/H10/H11.
5. C-terminal domain: also varies considerably, in terms of sequence, among nuclear receptors ²⁰⁻²³.

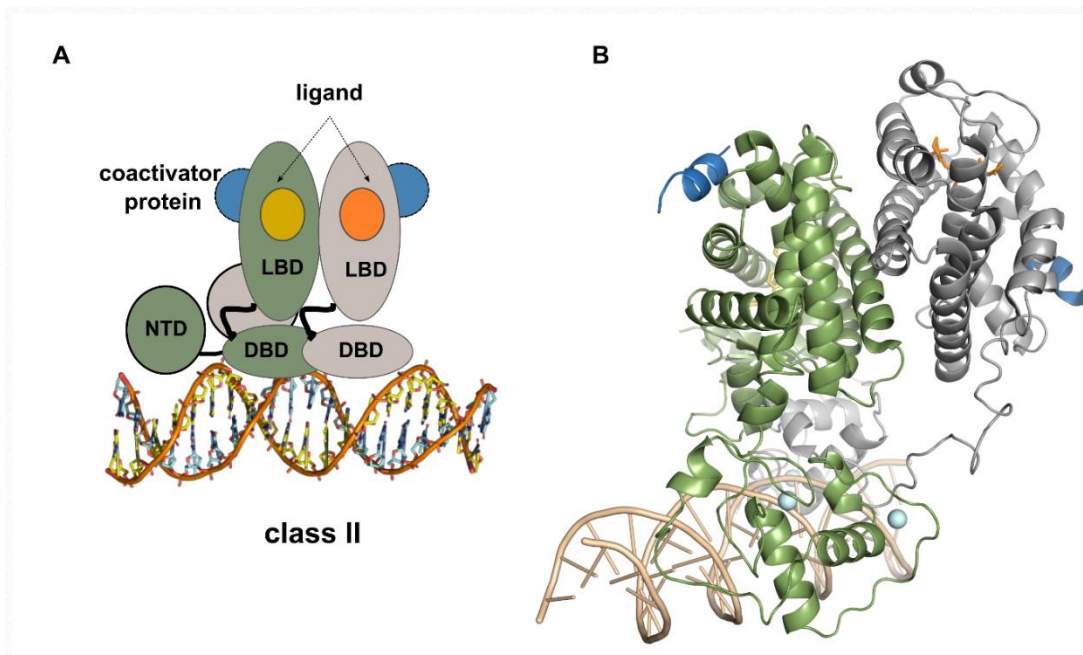


Fig. 1 Illustration of nuclear receptor structure (Class II). (A) Schematic view of the PPAR γ -RXR α complex. The N-terminal domain (shown as **NTD**) is ligand-independent; the DNA binding domain (shown as **DBD**) is conserved with two zinc fingers; the **Hinge** region is shown as a black loop connecting the DBD and LBD; the ligand binding domain (shown as **LBD**) is involved in dimerization, ligand binding, and coactivator binding (denoted in blue). (B) Cartoon representation of image A. Gold colour denotes DNA; cyan spheres depict zinc; the green and grey cartoons illustrate dimer structure of the PPAR γ -RXR α complexed with a coactivator in blue. PDB ID: 3DZY ²⁴.

Physiologically, NRs regulate the expression of genes involved in the development, metabolism, and even drug/xenobiotic responses. NRs are regulating genes involved in different physiological functions such as cell growth, differentiation, homeostasis, and metabolism and were conserved through evolution. They are transcription factors and commonly function by being activated by small lipophilic molecules (<1,000 Da), able to cross the membrane. Initially, they were solely identified as endocrine receptors, however, it was later discovered that NRs can also interact with xenobiotic compounds, such as Endocrine Disruptor Chemicals (EDCs or EDs, ²⁵⁻²⁷). EDCs can mimic the behaviour of endogenous ligands such as natural hormones and modify their metabolism and transport through NR-mediated signalling. This phenomenon causes a wide range of developmental, reproductive, or metabolic diseases ^{25,28,29}

Transcriptional activation is typically facilitated through the LBD. This complex domain consists of three distinct yet interconnected surfaces. These surfaces include:

1. Ligand-binding pocket: This pocket serves as a location for small molecules to bind.
2. Activation function 2 domain (AF-2): Composed of the helices 3/4/5/12 interface. AF-2 is responsible for ligand-dependent transactivation. It also functions as the surface for binding coregulators.
3. Dimerization surface: This surface enables interaction with other LBDs in partner molecules.

Known NR activator ligands show different interaction profiles with residues within this ligand-binding pocket. Starting from the ligand-free basal conditions, corepressors can interact via the short peptide motif LxxlxxxL (where L is leucine, I is isoleucine, and x can be any amino acid) on the AF-2 domain surface³⁰. After binding to a ligand, the LBD undergoes an allosteric conformational change that results in the movement and stabilization of H12. This conformational change leads to the release of corepressor binding (referred to as CoR, if present) and enables the recruitment of coactivator (known as CoA), which is commonly referred to as the "coregulator switching" model^{31,32}. Coactivators bind via LxxLL motifs (where L is leucine, and x can be any amino acid) to NRs³³.

As mentioned, human NRs are regulating the gene network involved in physiological phenomena such as cell growth, development of secondary characteristics, cell differentiation and metabolism regulation. The activated NRs bind to a conserved DNA region called response element (RE) downstream in the promoter of target genes³⁴. The canonical core motif has the consensus sequence 5'-AGGTCA-3'³⁵. The specificity and affinity of NR binding are dependent on the configuration and number of the core motif³⁶. Another factor to affect the NR-specificity is the linker region between the core motifs^{37,38}. Concomitantly with the DBD-DNA interaction, the NR complex has access to different nuclear co-regulatory proteins.

Transcriptional regulation of genes, whether it involves activation or repression, occurs through the interaction of nuclear receptors (NRs) with coactivators or corepressors, as well as other protein factors that interact with the promoter of the target gene. Corepressor proteins, such as the silencing mediator of retinoid and thyroid receptors (SMRT) or the nuclear receptor co-repressor (NCoR)³⁹⁻⁴² contribute to gene silencing by recruiting histone deacetylases, chromatin modifiers, and remodelling proteins. On the other hand, coactivator proteins, such as the steroid receptor coactivator (SRC)

family, can recruit histone acetyltransferases, histone methyltransferases, and histone kinases, resulting in chromatin unpacking, promoter opening, and activation of the target genes ^{31,39,43}. Numerous studies have highlighted the pivotal role of H12/ α AF-2, which is part of the AF-2 region, in controlling the activation and deactivation processes ⁴⁴.

In this sense, NR activator ligands can stabilize the active AF-2 conformation, forming a surface that binds co-regulatory proteins. Then, different CoA proteins can modulate the transcriptional activity. The CoA recruitment event together with the DNA interaction marks the beginning of the nuclear receptor activity and, depending on the coactivators/corepressors binding and on the cellular context, alternative transcriptional outcomes can take place ⁴⁵⁻⁴⁷. Alternatively, antagonists can act by destabilizing this relevant H12/ α AF-2 conformation and partial-agonists can partially trigger this molecular event. The nature of the ligand, occupancy in the binding pocket and interactions, determine the position of H12 and subsequently the CoA interaction ^{18,19}.

The corepressor and coactivator motifs form amphipathic α -helices, of which the hydrophobic residues interact with the AF-2 surfaces of the LBD ¹⁹. However, NR-LBD (in)-activation should probably not be seen as an “on/off” switch model. Rather, NR-LBD acts as a regulator fine-tuning the interaction between NR domains with the coregulators, which would allow a range of signalling outcomes¹⁹.

In this context, the biological role of NR is not determined by each protein individually but is rather a result of their interaction. The molecular determinants dictating specificity/selectivity in NR-CoA interactions remain understudied on a structural level. The pioneering work from Broekema *et al.*, (2014) ⁴⁸ suggests that amino acid sequences in both the NR-LBD and coregulator motif are relevant determinants in the NR-specific preferences for particular coregulator binding motifs. However, most of the NR crystal structures only offer a static vision of these individual components, lacking insights into the conformational changes induced by the different ligands and protein binding partners. **The essence of the problem is the difficulty of experimentally addressing conformational change in complex structures**; that is, how the effect of the ligand-binding propagates through the structure to affect other sites.

1.2 Nuclear receptor modulators

Many studies have shown the crosstalk of NRs that is followed by controlling the homeostasis of glucose, bile acids, lipids, hormones, and inflammation ⁴⁹. This ability stems from the flexibility and versatility of nuclear receptors, as their transcriptional activity can be regulated by ligands, partner proteins, coactivator, corepressor, and promoter genes. This mechanism underscores their role in a wide range of developmental, reproductive, or metabolic NR-related diseases. Given these characteristics, NRs have emerged as prominent therapeutic targets.

ER is the most targeted NR due to its druggable nature. It belongs to type I of nuclear receptor found in cytoplasm connected with heat shock proteins. Upon ligand binding it forms a homodimer and translocate to the nucleus. This receptor is found in two forms, ER α and ER β , both of which bind to the native ligand estradiol. Tamoxifen, approved in the 1970s ⁵⁰, and Raloxifene, approved in 1997 ⁵¹, are used for the treatment and prevention of ER α -dependent breast cancer in women as antagonists. Both compounds have been co-crystallized with the receptor, binding to the ligand binding pocket (LBP), They exert their effects by dislocating H12 from an active conformation to an inactive state. However, Tamoxifen can lead to endometrial cancer as an agonist because of the variability of coregulator proteins, whereas Raloxifene, also acting as an agonist, is used for osteoporosis treatment in women. Androgen Receptor (AR) is another NR found in the prostate and several other tissues with testosterone as native ligands. Several diseases, including prostate cancer, have been linked to this receptor. Although Enzalutamide is AR competitive antagonist approved drug in 2009, patient resistance happens after months of treatment ⁵².

As we mentioned earlier, EDCs are xenobiotic compounds interacting with NRs and by mimicking endogenous, causing a broad range of diseases. Epidemiological studies reported that exposure to xenoestrogens such as Diethylstilbesterol (DES) during fetal development and exposure to Dichlorodiphenyltrichloroethane (DDT) during puberty increase the risk of breast cancer ⁵³. Another study offers initial insight into the neural effects of human exposure to Bisphenol (BPA). The results propose that when expectant mothers are exposed to BPA during prenatal stages, it may cause modifications in the microstructure of white matter in preschool-aged children and these changes in white matter can mediate the connection between early-life exposure to BPA and the emergence of internalizing problems ⁵⁴. Similarly, a rising number of

cancerous testicular or malformations of the male genital tract also could be attributed to exposure to Endocrine disruptor chemicals ⁵⁵. A study has revealed the relationship between the level of plasma Polychlorinated biphenyls (PCBs), a persistent and lipophilic aromatic chemical, and reduced semen quality, particularly reduced sperm motility ⁵⁶.

Class II of NR have been also the target of therapeutics. Among them, PXR, CAR and FXR are of substantial interest. These receptors are introduced shortly in the next section followed by the employed *in silico* methods.

1.3 Retinoid X receptor

Retinoid X receptor (RXR) is a type II NRs and forms heterodimers with approximately one-third of the other NRs. In humans there are three isoforms of RXR: α , mainly found in the liver, kidney, and intestine; β , found in most human tissue; and γ , which mainly exists in the brain and muscles⁵⁷. Malfunctioning of these isoforms have been linked to various health issues^{23,58–60}.

RXR activation can be categorized into two groups: permissive and non-permissive heterodimer. In the permissive heterodimers⁶¹, activation can be induced by binding the agonist to either RXR or NR partner or both receptors. Examples of this group are RXR/FXR, RXR/LXR and RXR/PPAR. In the case of the non-permissive heterodimer, only agonist binding to the RXR partner triggers activation, but RXR can still bind to the agonist, releasing the corepressor and recruiting coactivator. Examples of this are RXR/TR and RXR/VDR. In this case, RXR can also bind to the agonist and lead to synergistic action in the presence of a heterodimer ligand^{61,62}.

The main small molecule compound that bound to RXR is 9-cis-Retinoic acid (9cRA). it belongs to a retinoid family and has a critical role in cell growth, development, differentiation, and apoptosis. Other RXR ligands are 9cRA-related compounds and indenoisoquinolines. Remarkably, the work from Bexarotene and Diarylamines²⁴ employed X-ray crystallography to describe the PPAR γ and RXR α structure. They explained how these two receptors interact with DNA, highlighting the influence of DNA in governing the interaction between the two receptor domains through specific rearrangements. They also revealed the cooperative nature of multiple domains of PPARs, which can modulate the properties of the PPAR γ -RXR α complex. In this complex, the LBD of PPAR γ tightly couples with RXR α domains. Accordingly, the conformational change induces a reposition of receptor domains responsible for DNA binding and optimizing their contact with DNA. They also examined the dynamic properties of structures using amide hydrogen/deuterium exchange mass spectrometry (H/D-Ex). The result revealed that the helices H10/H11, involved in LBD-LBD heterodimerization, adopt a slightly shifted/curved conformation when the protein is contacted with DNA, facilitating optimal contact of receptor with DNA²⁴. These findings established the concept of H10/H11 analysis in our *in silico* studies, as presented in our publications.

1.4 Pregnane X Receptor

Pregnane X receptor (PXR), also known as the steroid and xenobiotic sensing nuclear receptor (SXR) ⁶³, is encoded by the *NR1I2* (nuclear receptor subfamily1, group1, member 2) gene on chromosome 3 ^{64,65}. PXR heterodimerizes with RXR α , β and γ (NR2B1–3) ^{66–68} at H10/H11 region. Additionally, the PXR-LBD demonstrates a distinctive characteristic by homodimerizing at its $\beta 1'$ interface. Homodimerization occurs through the conserved W223 and Y225 residues in each monomer (Fig. 3) ⁶⁹. These amino acid residues involved in the interface exhibit high conservation among various species, including humans, rhesus monkeys, rabbits, mice, rats, pigs, and dogs. However, in dog's PXR, W223 is replaced by Q223 ⁶⁹. Noble *et al.* (2006) showed mutation of W223 and Y225 does not interfere with DNA, RXR, or ligand binding, rather it disrupts the homodimerization, reducing the recruitment of the coactivator SRC-1 and transcriptional activity ⁶⁹.

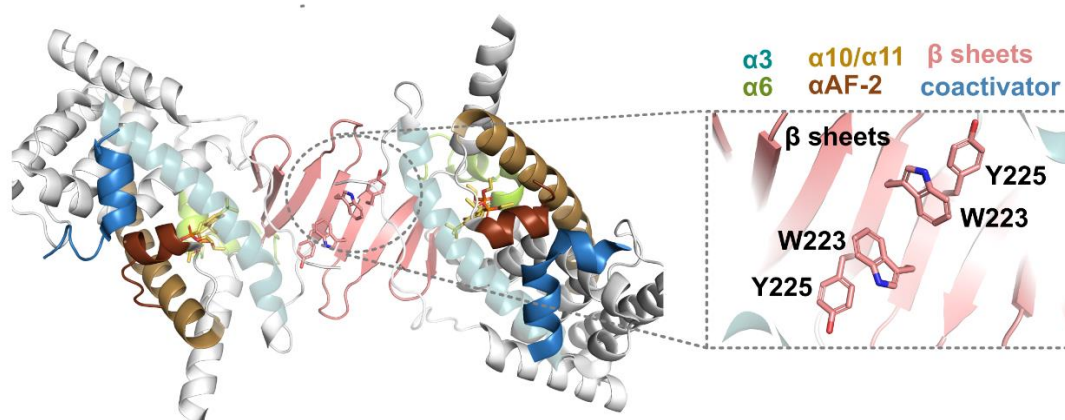


Fig. 2 Crystal structure of PXR homodimer. The amino acids W223 and Y225, located on $\beta 1'$, mediate the homodimerization. The interface (which are β -sheets) shown in pink. PDB ID: 1NRL ⁷⁰.

When an agonist ligand binds to the PXR-RXR α heterodimer in nucleus, it promotes coactivator binding and release of corepressor from AF-2 ⁷¹. Subsequently, this activated PXR complex induces expression of the target gene. PXR structure, like other NRs, has a large hydrophobic LBD. The primary PXR isoform is composed of 434 amino acids, featuring a notable hydrophobic triad consisting of F288, W299, and Y306. Unlike other nuclear receptors (NRs), the PXR-LBD lacks the typical stable H2' and H6 helices. Crystallographic structural data clearly illustrate the lack of stability in the H2' region, which appears disordered in all publicly accessible PXR structures.

These characteristics result in a more expansive and flexible LBD for PXR, distinguishing it from other NRs⁷²⁻⁷⁴. Consequently, the PXR-LBD can accommodate a diverse range of ligands. Presently, the Protein Data Bank (<https://www.ebi.ac.uk/pdbe/>) repository contains 52 crystal structures of human PXR in its active mode, complexed with the coactivator protein SRC1. Among these structures, 44 exist in a homodimeric assembly, while 8 are available in heterodimer form.

PXR is a ligand-dependent transcriptional factor involved in small molecule metabolism and regulation of diverse cellular processes including bile acid metabolism, glucose homeostasis, cell proliferation as well as inflammation. PXR mostly exist in the liver and intestine. It regulates the gene expression of enzymes and transporters that are responsible for the different pathways of endogenous and xenobiotic pharmacokinetics including absorption, distribution, metabolism, and excretion (ADME). Gene targets of PXR are cytochrome P450 genes (CYP2B, CYP2C, CYP3A) and efflux and uptake transporters of the ATP-binding cassette^{65,75-78}. Besides endogenous ligands, PXR is activated by a broad number of diverse small molecules, including drugs, environmental pollutants, and natural products. These various functions make PXR a potential therapeutic candidate. However, the activation of PXR can induce intestinal and hepatic first-pass metabolism and drug efflux transport⁷⁹ which in turn may lead to drug-drug interactions (DDI), adverse drug reactions or therapeutic failure of drugs⁸⁰⁻⁸². To exemplify, we can consider the report about the reduced effect of rifampicin on midazolam or contraceptive due to the increased expression of CYP3A4 when these medicines are co-administrated⁸³, or isavuconazonium which activates the expression of CYP2B6 through PXR-mediated induction and decrease the exposure of bupropion^{75,84}. These observations raised the interest in designing PXR antagonists along with the attempt to limit the activation of PXR in the presence of xenobiotics. As of 2002, several “azole” compounds have been identified as PXR inhibitors, such as ketoconazole, enilconazole, FLB-12 and SPA70. It is less known about the structural trigger of PXR-bound antagonists. For instance, SPA70 and SJB7 are close analogues⁸⁵ where SPA70 act as an antagonist of PXR but SJB7 is a PXR agonist which highlights the promiscuity of PXR-LBD. No co-crystal structures of PXR and antagonists are currently available to elucidate the details of the PXR-antagonist interactions.

Most known ligands bind to orthosteric PXR-LBD, however, allosteric sites can be an alternative region for the PXR modulator's accommodation ⁸⁶⁻⁸⁹. Allosteric sites are distant from orthosteric sites and accommodate structurally different ligands. So far, 202 allosteric modulators have been reported for nuclear receptors ⁸⁹. The proposed allosteric ligand binding sites are the AF-1 site, zinc fingers and response elements, LBP (synergistic), the AF-2 site, and the binding function 3 (BF-3 site) ^{89,90}. An example of an allosteric ligand binding pocket is reported by Delfosse *et al.* ⁹¹ where the simultaneous binding of 17- α -ethinylestradiol (EE2) and transnanchlor (TNC) enhanced the CYP3A4 induction much higher than the single binding of either of compounds when compared to potent agonist SR12813. Ketoconazole ⁹² is an example of a modulator binding to AF-2 and acts as a PXR inhibitor. To name, Fluconazole, Enilconazole, Pazopanib ⁸², Metformin ⁹³, Leflunomide ⁹³, FLB-12 ^{92,94}, Coumestrol ^{95,96}, Sulforaphane ⁹⁷, and Camptothecin ⁹⁸ are other AF-2 modulators. There has been a number of approaches from experimental methods to computational techniques such as pharmacophore, quantitative structural-activity relationship (QSAR), machine learning, and structure-based methods to investigate PXR conformation upon ligand binding ⁷⁵. Notably, due to the lack of crystal structures of PXR in complex with antagonist ligands, likely attributed to the complexity and high flexibility of the system, computational studies play a crucial role in unravelling the conformational dynamics of the PXR-antagonist complex. Taken together, the off-target effect of kinase inhibitors as cancer therapeutic agents pursued us to identify small-molecule kinase inhibitors of the Tübingen Kinase Inhibitor Collection (TüKIC) compound library that would act also as PXR antagonists. The discovery of drugs that can simultaneously inhibit both PXR and protein kinases could offer new possibilities in cancer treatment and help overcome drug resistance. To achieve this, we conducted a study utilizing structure-based virtual screening and molecular dynamics (MD) simulations to investigate the conformational changes of PXR-LBD, specifically when bound to an antagonist. Further details are discussed in publication I.

1.5 Constitutive Androstane Receptor

Constitutive Androstane Receptor (CAR), encoded by the *NR1I3* (nuclear receptor subfamily1, group1, member 3) gene belongs to class II NRs. CAR heterodimerizes by RXR α , β , γ (*NR2B1–3*) and like PXR, plays a critical role in regulating genes involved in exogenous and endogenous metabolism.

Alternative splicing generates multiple CAR isoforms in humans and other primates, but not in rodents^{99,100}. This mechanism, however, is not yet fully understood. Around 50% of transcripts encodes the wild-type CAR1 that displays high basal activity, CAR2 and CAR3 isoforms demonstrate ~10% and ~40%, respectively with low constitutive activity, likely due to their reduced interaction with RXR which results in weaker binding to DNA and coactivators^{101,102}.

The unique feature of CAR is its constitutive activation, distinguishing it from other nuclear receptors. Unlike other NRs, CAR does not require ligand binding for its transcriptional activity although ligand binding can modulate CAR activity as an inducer or inhibitor. CAR can bind to a vast number of chemical compounds¹⁰³. As a result, CAR regulates multiple genes involved in xenobiotic detoxification, which might overlap with or be distinctive from PXR target genes. CAR is predominantly expressed in the intestine and liver^{104,105} and primarily localized in the cytoplasm and forms a complex with heat shock proteins¹⁰⁶. The majority of CAR ligands act as direct activators, such as 6-(4Chlorophenyl)imidazo[2,1-b][1,3]thiazole-5-carbaldehyde-O-(3,4-dichlorobenzyl) oxime (CITCO)¹⁰⁷ in human and TCPOBOP¹⁰⁸ in mouse. On the other hand, synthetic compounds like phenobarbital¹⁰⁹ and acetaminophen¹¹⁰ (<https://pubmed.ncbi.nlm.nih.gov/12376703/>), as well as endogenous compound bilirubin, are examples of indirect CAR activators. It has been proposed that the Epidermal growth factor receptor (EGFR) signalling pathway is inhibited by phenobarbital, leading to the dephosphorylation of CAR at T38 within the cytoplasm, which enables its translocation to the nucleus^{16,17}. Additionally, flavonoids have been reported to function as both direct and indirect activators depending on cellular context¹¹¹.

In the nucleus, CAR plays a constitutive regulatory role in target genes, including CAR exclusive gene CYP2B6. To achieve this, CAR interacts with specific DNA motifs DR3, DR4, DR5, ER6, and ER8, which are located in the enhancer and promotor region of target genes¹¹². Moreover, certain CAR activators, such as phthalates, antivirals, and

artemisinin derivatives display some isoform selectivity ^{101,102,113–115}. Androstane metabolites, PK11195 ¹¹⁶, T0901317 ¹¹⁷ and S07662 ^{118,119} are examples of CAR inverse agonists (inhibitors). Inverse agonist can reduce the basal constitutive activity of CAR, acting as inhibitors. Another feature of CAR, as a member of NRs, is its permissive activity when complexed with RXR (section 1.3). A number of studies ^{118,120,121} have demonstrated the synergistic and additive effect of multiple xenobiotic compounds. Dauwe et al., (2023)¹²² conducted their *in vivo* study with several pesticides—recognized as ligands of CAR and Tri-butyl-tin (TBT) served the role of an RXR agonist. In mice subjects, the concurrent administration of dieldrin (pesticide) and TBT prompted a synergistic activation of CAR. Furthermore, combined effects were observed with propiconazole, bifenoxy, boscalid, and bupirimate. These findings highlight the need for further investigation into the structure and function of CAR, particularly considering its shared target genes and modulator compounds with PXR. With the limited availability of crystal structures depicting agonist-bound CAR, our objective was to employ *in silico* methods to model CAR isoforms and comprehensively study their structure and conformational changes when interacting with various chemical compounds. These compounds are ranging from CITCO analogies to endocrine disruptors (EDs).

Our results are presented in publication IV where we discovered several derivatives of 3-(1H-1,2,3-triazol-4-yl)imidazo [1,2-a]pyridine that directly activate human CAR in nanomolar concentrations and one manuscript where we conclude branched 4-nonylphenols isomers are responsible for enhancing constitutive androstane receptor. The methods were accompanied and supported by functional assays.

1.6 Farnesoid X receptor

Farnesoid X receptor (FXR) is a member of the NR superfamily (*NR1H4*), highly expressed in the liver and intestine, but to a lesser extent expressed in the kidney, adipose tissue, and adrenal. FXR regulates endogenous such as bile acid homeostasis, lipid and glucose metabolism, and inflammation ¹²³. Farnesol derivatives, which are metabolic intermediates of the mevalonate pathway ¹²⁴ chenodeoxycholic acid (CDCA) and cholic acid (CA) are endogenous ligands for FXR ¹²⁵. FXR α and FXR β are two known FXR genes ^{126,127}. The FXR α gene is conserved from fish to human ¹²⁸ and encodes four transcript isoforms, FXR α , FXR α 2, FXR α 3,

and FXR α 4 in humans and mice, whereas FXR β is a pseudogene in humans and primates^{123,129,130}. FXR, likewise other NRs family members, has a highly conserved domain. The binding of the ligand to FXR, which is in complex with the retinoid X receptor (RXR), induces conformational changes leading to the recruitment of either coactivators or corepressors, promoting or silencing the transcription of target genes, respectively. More details on FXR were explained in publication III.

2. In silico Drug discovery and Molecular Modelling

The process of introducing a new drug to market is time-consuming and expensive because each step has its own challenges. For the validation of a pathway as a drug target, the disease/pathology mechanism should be well-known, and the selected target should be druggable, meaning its activity can be effectively modulated by an exogenous compound. Next, the selected target should be validated by experiments such as gene knockdowns and antibody interaction¹³¹. This step is very crucial since any issue will propagate through the whole pipeline of discovery and development¹³². The purpose of lead identification is to identify or create a compound that interacts with a selected target. In this step, the mechanism of action of the drug, cell-based *in vitro* tests for initial safety¹³³ as well as the pharmacokinetic and pharmacodynamic property of identified lead are also conducted¹³⁴. Later, in lead optimization step, identified compound is modified to increase the effectiveness and safety, to prevent off-target effects, and to optimize the dosage^{135,136}. This step also includes safety testing in animal models and multiple pre-clinical *in vivo* animal models¹³⁷. To address potential side effect of the drug, it will pass through extensive testing in animal models. The accuracy of this step is crucial since the next stages of development are highly expensive¹³⁷. For this step, knock-out or genetically modified mice are used. After the Investigational New Drug (IND)¹³⁸ application, the drug candidate undergoes clinical tests during Phase I (healthy volunteers), II (preliminary data in patients) and III (only ~10% of the drug can reach this point¹³⁹) and FDA approval¹⁴⁰, and finally the drug can be introduced in the market. These processes require investment for efficient operation with a low rate of success¹³⁹. On the other hand, advances in computer-aided drug design (CADD) and protein crystallographic techniques have brought new approaches to drug discovery pipelines.

The CADD approach is a cost-effective method that can be utilized in the virtual screening of compound libraries against a target¹⁴¹. It can investigate the specificity of selected hits using molecular docking and molecular dynamics, predict ADMET properties of the selected hit, and optimize the lead compound for synthesis and testing. Imatinib, zanamivir and nelfinavir are examples of drugs designed or optimized with the aid of computer methodologies. The artificial intelligence (AI) and machine learning even further improved the CADD process^{142,143}.

Accordingly, different computational methods have been implemented for understanding the nuclear receptor functional activities. These computational approaches, generally, fall into two categories, namely ligand-based and structural-based drug design.

The ligand-based technique is used in the absence of the three-dimensional (3-D) structure of the target which relies on knowledge of the small molecules bound to the desired target. When the active ligand is known, QSAR and pharmacophore mapping are the commonly used approaches ¹⁴⁴, While in the presence of knowledge about both active and inactive ligands of the target of interest, a machine learning approach is employed. These methods will generate a predictive model for hit identification and lead optimization ^{145,146}.

In contrast, a **structure-based strategy** is utilized when the structure of the target is available which is mainly determined by experimental techniques such as X-ray crystallography, nuclear magnetic resonance (NMR) or cryo-electron microscopy. These techniques, however, are expensive and time-consuming. When the crystal structure of the drug target is not available, protein structure can be predicted by computational methods like homology modelling. However, in 2020, the emerge of AlphaFold, a new tool developed by DeepMind predicting the 3-D structure of the target from its sequence, even further accelerate protein structure prediction with high accuracy ¹⁴⁷.

I	II	III	IV	V
Target/ligand selection	Protein/ligand preparation	Binding site definition	Docking	Validation
3D structure of target	Add missing side chain/loop	Grid generation	Searching algorithm (generating conformational space)	Redocking of co-crystal structure (the result should be similar to the top-ranked pose)
3D structure of ligand	Add hydrogen	Binding site detection	Scoring function (ranking poses)	
	Assign bond order			
	Minimize the structure			MD simulation

Fig. 3 An example of typical workflow of structure-based drug discovery. The flowchart summarizing the methodology employed in this work (modified from ¹⁴⁸).

2.1 3-D structure of the target and druggability

The application of structure-based computational methods heavily relies on the availability of experimentally determined 3-D coordinates of macromolecule structures in a database. The knowledge of the 3-D structure of the target protein is a key requirement to apply computational methods to identify novel agonists and antagonists. The bio macromolecular 3-D structures are commonly stored in Protein Data Bank (PDB)¹⁴⁹, which is a public archive containing results from X-ray crystallography, NMR spectroscopy, and cryo electron microscopy (CryoEM). Most of the stored structures are determined by X-ray crystallography (89%) whereas the minority fraction belongs to NMR spectroscopy and electron microscopy (<https://www.rcsb.org/stats>)¹⁴⁹. In addition, the PDB oneDeep has emerged as a unified system for deposition, biocuration and validation of macromolecule structures¹⁵⁰.

Since the quality of the protein structure has a tremendous effect on the outcome in the modelling results, several parameters have been introduced to check the quality such as resolution (precision with which atom carbon is replaced in structure). As a rule of the thumb, a resolution value lower than 2 Å is more confident in placing atoms in structures. Nevertheless, the resolution does not provide information about the quality of the data. Therefore, the two other parameters are defined, namely: R-factor and R-free¹⁵¹. The R factor is a way to see how well a simulated diffraction pattern matches the one we observe in experiments. The R-factor is in the range of 0.6 for random fit and 0.2 for typical fit, whereas a perfect fit would have a value of zero. Therefore, the lower the R-factor, the more consistent model, and the better reflection of experimental data. This range could be lower for small molecules because of a better-ordered crystal. The R-factor is biased because it reflects the refinement model against the data used to train the model. Therefore, R-free was introduced as a less biased way where 10% of experimental data is removed to be used as later cross-validation and 90% is kept for refinement. Hence, the R-free value is calculated based on how accurately the model predicts the 10% of data that was not used in the refinement process. An ideal model has a similar R-factor and R-free about 0.26¹⁵². For selecting the crystal structure with good quality, one should also take care of the B-factor of the binding site to be close to the optimal value which is 12-20 Å²,

Deviations from this value indicate less reliable atomic coordinates within the binding site. B factor represents the relative motion exhibited by various regions of the protein. Regions characterized by well-ordered structures typically display low B-factor values, whereas flexible regions like loops or surfaces exhibit higher values ¹⁵³.

In drug design, it is important to understand the structure and function of the protein active site as well as finding any potential allosteric binding sites. In the absence of a native ligand, it is viable to utilize some tools, such as a SiteMap ^{154,155} from Schrödinger which explore putative binding sites based on size, functionality, and solvent exposure. Generated binding sites are ranked by SiteScore. Along with site definition, SiteMap also provides the possibility to optimize the lead compound with an enhanced receptor complementarity when evaluating the protein-ligand interactions (Schrödinger, LLC, New York, NY).

In the case of our NR work, we started with the generation of a protein structure curated database for the ligand binding domain (LDB) of nuclear receptors (NRs). All selected structures were manually curated based on their resolution (high-resolution structures, < 2.5Å were preferred whenever available). Structures with co-crystallized ligands (not soaking) for known relevant ligands (agonists, partial agonists, and antagonists) were also selected as controls for our studies and their ligands were evaluated in terms of electronic density to ensure their quality.

2.2 Homology Modelling

When an experimentally derived structure is unavailable for our protein of interest (query), a homology modelling strategy can be utilized to generate a 3-D structure. Steps include selecting sequence alignment tools (e.g., BlastP: <https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>)¹⁵⁶ template recognition (using the Protein Data Bank, PDB; <https://www.rcsb.org/>)¹⁴⁹, alignment correction, backbone and loops generation, and side chain modelling using rotamer library. For building a homology model, the degree of sequence identity/similarity between the query and the templates is deterministic where certain limits suggest the reliability of the generated model. There is good coverage with query-template identity higher than 80%. When this pairwise similarity ranges from 50-80% likely the proteins share the same fold ¹⁵⁷. 30-49% of identity model is with certain errors in the loop and side chain conformation, and less than 30% of sequences indicates a wrong model ¹⁵⁸. There are

several potential sources of error while generating a structure model resulting from low homology similarity between the target sequence and template, steric clashes of side chains, inaccuracy of bond length and angle, missing atom, and inaccuracy of underused rotamer library. Therefore, validation of the generated model is essential to assess the quality of the generated model. Among others, Root Mean Square Deviation (RMSD) is a measure of the average distance between the atoms of two superimposed proteins. If the RMSD value is 0.0-0.5, the model and template are essentially identical, up to 1.5 denotes a very good fit, 1.5-5.0 is a moderate fit and 5.0 to 12.0 is a completely unrelated model ^{159,160}. Another parameter for validation of model is stereochemical quality of protein which is doable through Ramachandran plot (RP). RP is distribution plot of protein torsion angles ϕ and ψ , analysing both the geometry of individual residues and the overall protein structure ¹⁶¹ (Fig. 5). The plot is divided into three distinct regions. The favoured region represents dihedral angles which exist in well-folded protein and is energetically the most favourable region. The extension of the favourable region is allowed area that are also acceptable but less common than those in the favoured region. In contrast, in the outlier region, the atoms come closer together than their van der Waals radii, resulting in steric clashes, making this region disallowed. Glycine residues are exceptions and due to the lack of a side chain, probable to be found in any region of the plot.

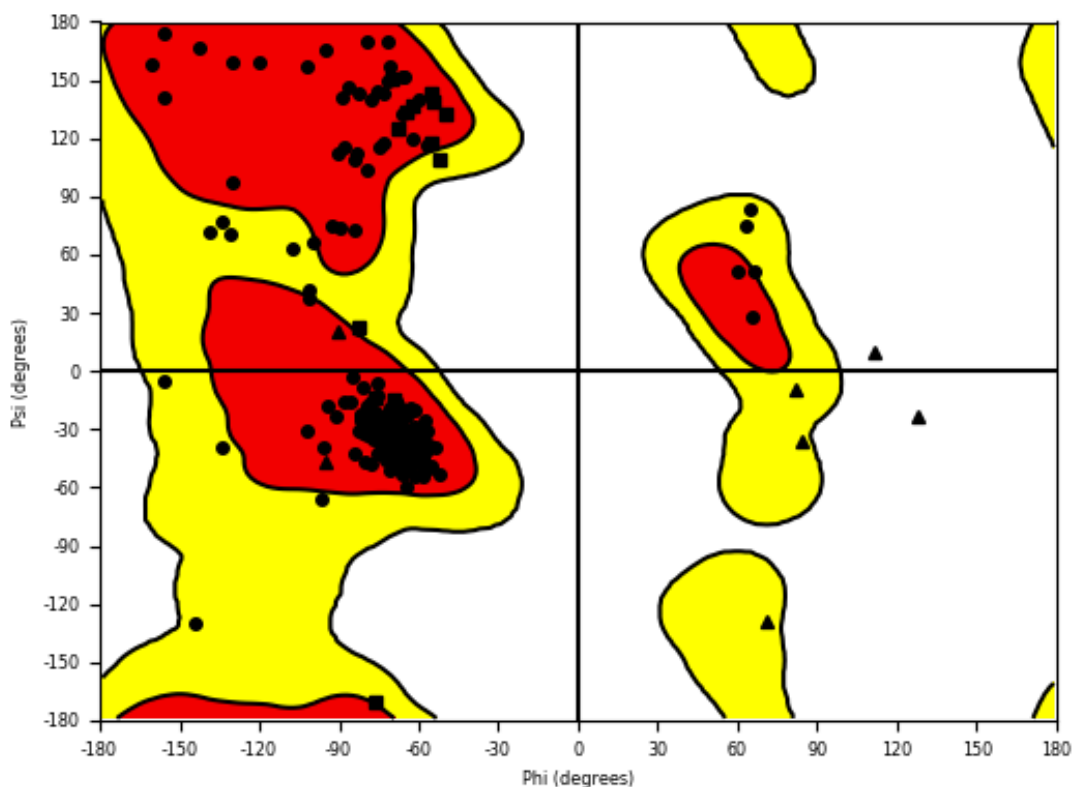


Fig. 4 Ramachandran plot of CAR3-LBD model. The plot describes dihedral angles for all CAR3 residues. The model was built using PDB ID: 1XVP¹⁶² as template. Glycine is marked with triangles, proline with squares, and other residues with circles. The plot shows that most residues fall within the allowed regions, except for Glycine.

Several software packages exist as a ready-to-use platform for homology modelling, such as EasyModeller 4.0 and MODELLER which have built-in functions for validation¹⁶³. Procheck¹⁶⁴, ProSA¹⁶⁵ and WhatCheck¹⁶⁶ are other examples of validation tools along with PDSUM and Molprobit¹⁶⁷ which are online tools for model quality validation. In summary, homology modelling is a time- and cost-efficient method and is easily applicable through freely available software. However, a few cons remain such as difficulties in loop modelling, the essential need for the experimental structure of homologous protein and disability to provide information about the protein folding, although AlphaFold¹⁶⁸ can offset this shortcoming. In conclusion, one can benefit from the homology modelling method in the absence of receptor crystal structure but should bear in mind that the generated models are not error-free and need validation by various methods.

2.3 Protein and ligand preparation

After retrieving the 3-D structure, either a crystal structure or a model, the structure should be prepared prior to docking. In the case of a crystal structure, highly flexible regions do not diffract with high enough resolution for assignment during crystallography, such as loops. Hence, these regions should be generated. In addition, because of the insufficient resolution, hydrogen atom coordinates cannot be determined in the crystal. Hence, the placement of missing hydrogen atoms is a relevant step in target preparation. Missing side chains, missing bonds, and bond orders should be added as well. Higher alternate occupancies should be selected, and, more importantly, optimizing the hydrogen bonding network including the prediction of protonation states of protein residues as well as identification of structural water should be considered.

In our **projects**, we utilised PROPKA in the Protein Preparation Wizard tool of Maestro¹⁶⁹ to select the most likely protonation states and tautomer for the Histidine residues. We followed the software suggestions and then optimized the generated H-bonding species accordingly. Nevertheless, for PXR project (publication I) we also tested whether different X-ray structures would yield distinct ionization states for the histidine residues. Fig. 6 demonstrates the application of PROPKA to these different crystal structures, each with different pH values (pH 7.0, and 8.0, which were used for crystallization). All histidine residues near the binding site exhibited the same ionization states, regardless of the pH values. Additionally, we checked the electron density of these residues, although their resolution did not allow us to accurately observe the hydrogens.

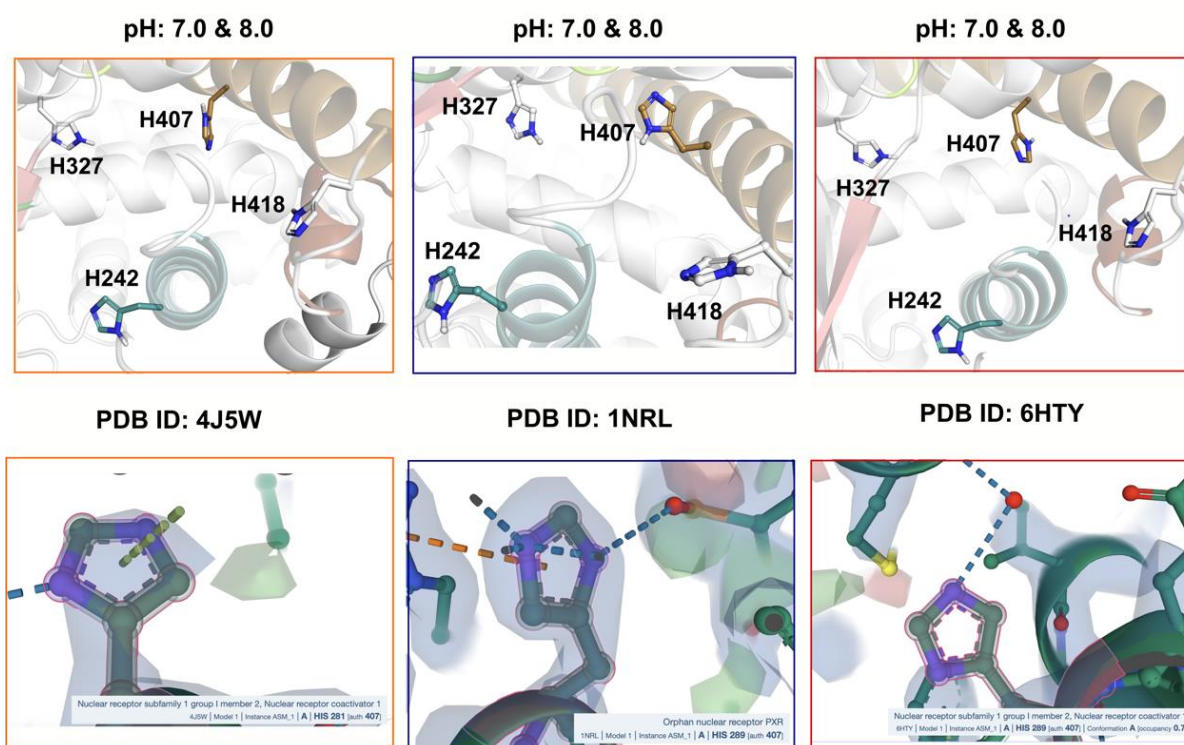


Fig. 5 Histidine 407 (H407) protonation state. Protonation obtained from different PXR-LBD crystal structures. The top panel depicts H407 after preparation and H-bond assignment using PROPKA algorithm in Schrodinger suite. Bottom panel displays the corresponded H407 with electron density before H-bond assignment.

Before docking, along with the target protein, the ligand structures should also be prepared, including the assignment of ionization states, partial charges, and the generation of their minimized 3-D conformations ¹⁶⁹.

In the case of our NR work, to obtain the starting configuration for EDs before docking, the ligands were prepared using LigPrep (Schrödinger, LLC, New York, NY) to assign the protonation state and partial charges utilizing the software Epik (at pH 7.0±2.0). For the PXR study, we screened the virtual TükIC library and implemented a SAR study on the resulting compounds followed by ligand preparation steps using the Schrödinger suite. The same preparation steps were implemented for other projects presented in this thesis work.

2.4 Force field

There are two main methods in molecular modelling, namely: molecular mechanics and quantum mechanics. Molecular mechanics (MM) is less computationally intensive compared to quantum mechanics wherein the electronic motion is disregarded and calculate the energy of a system based on the positions of the nuclei which makes energy calculations faster. It can be used to calculate properties like structure, docking, molecular dynamics, entropy, and energies and can handle thousands of atoms. However, MM does not consider electrons separately and treats atoms as balls connected by springs, which limits its ability to calculate electron-related properties. On the other hand, quantum mechanics (QM) can provide energy calculations for bond breaking and forming, thermochemistry, and electronic properties. It considers electrons, nuclei, and interaction energies but is computationally expensive and can only handle hundreds of atoms. However, the key region of large molecules like active site of enzyme can be studied with quantum mechanics. Therefore, a combination of molecular mechanics and quantum mechanics is commonly employed in drug design. In molecular mechanics, a force field is a function that describes the energy of a molecule and how bonded interactions (such as bond lengths, bond angles, and torsions), as well as non-bonded interactions (such as van der Waals forces, electrostatic forces, and hydrogen bonds), evolve. A force field consists of equations and parameters that define the geometry and behaviour of a specific molecule. The functional form is given by energy of equation, which is summation of bonded and non-bonded energy terms, and parameters are constants obtained from experimental data and ab-initio calculation¹⁷⁰. Non-bonded interactions are typically described by the Lennard Jones equation and Coulomb's potential¹⁷⁰.

Non-bonded energy consists of electrostatic components, which are represented by distance-dependent dielectric constants, Van der Waals forces (repulsion due to overlap between the electron densities of two or more closely packed atoms and attraction which happens due to London or dispersion forces) and hydrogen bonding interactions (H-bond, for short), which is a chemical bond that occurs between an electronegative atom, such as nitrogen, oxygen or fluorine, and a hydrogen atom that is covalently bonded to another electronegative atom in a different molecule and play an important role in the stability and properties of biomolecules.

Based on the functional form and implemented parameters (depending on the atom type the related parameter is changed) different types of force fields exist. They might be different in the number of terms in energy expression, the complexity of the term, and the method of obtaining parameters if experimental or *ab initio*. To name: MM2/MM3/MM4 ¹⁷¹ (for small molecule study), CHARMM ¹⁷², AMBER ¹⁷³ (the most common force-fields for protein study), OPLS ¹⁷⁴ (for liquid simulations), and GROMOS ¹⁷⁵ (for MD simulation of macromolecules) . Each force field has its equation which obtains its parameter values from different databases. Therefore, one should be cautious not to mix different force fields for the same set of calculations, otherwise, the result would be inconsistent ^{176,177}.

Force field is core of MD simulations where the accuracy of generated model is heavily depends on the quality of the underlying force field. Each force field has its own positive and negative points. GROMACS 54A7 ¹⁷⁸ and CHARMM 36m ¹⁷⁹ were developed to overcome the problem of sampling structured and disordered proteins, where AMBER ¹⁷³ was highly successful in the description of the alpha-helix formation ¹⁷⁷. The study conducted by Kamenik et al. (2020) provides information related to polarized and non-polarized force fields, suggesting that the additive force field is sufficient for studying ultra-fast-folding dynamics and partial unfolding and refolding processes within the provided computational time.¹⁷⁷.

2.4.1 Geometry optimization

The geometry of a molecule is described by its internal coordinates, which consist of bond lengths, bond angles, and torsion angles between its atoms. Given a set of atoms and vector “ r ”, defining the atoms’ position, it introduces the concept of energy as a function of the atom's position, $E(r)$. The goal is to optimize this energy by finding molecular geometries where the net inter-atomic force on each atom of the molecule is relatively close to zero ¹⁸⁰. Such positions on the potential energy surface (PES) correspond to stationary points (minima or equilibrium structure), whereas saddle points in PES is a transition state in theories of reaction mechanisms ¹⁸¹. MM methods search the PES to reach these minimum energy structures using various minimization algorithms. These algorithms typically categorized into two types: based on first-order derivatives (Gradients) such as steepest descent and Conjugate Gradients, and those involving second-order derivatives (Gradients and Hessians) like Newton-Raphson. The choice of the algorithm depends on multiple parameters including storage and computing requirements, availability of an analytical gradient and the size of the Hessian matrix ^{182–184}. It is worth noting that conformational search is a significant challenge, especially for large molecules. Therefore, MDs simulations can be utilized alternatively as it enhances conformational sampling by integrating Newton's laws of motion (further details are provided in section 2.6). Once the minimum conformation is attained, it is crucial to validate the prediction by cross-checking it against the real conformation using multiple experimental and theoretical methods. Examples of such methods include X-ray crystallography, Electron diffraction, Ultraviolet (UV) Spectroscopy, Infrared (IR) Spectroscopy, NMR Spectroscopy, and Photoelectron Spectroscopy, among others.

2.5 Molecular docking and pose selection

Since the 1980s molecular docking has been used as a core method in the SBDD pipelines¹⁸⁵. When a high-quality 3-D structure of the target is available molecular docking can be used to predict the position, orientation, and conformation of a small molecule compound (ligand) within a defined region/pocket of the target (receptor) using an automated computer algorithm.

Molecular docking is used for various purposes in drug discovery such as hit identification, lead optimization, structure-activity studies, and prediction of the potential binding mode of a known binder. The docking procedure consists of two steps: Sampling and scoring. In the sampling phase, the docking algorithm generates docking poses employing systematic, stochastic, Monto Carlo or genetic algorithms. In systematic search, all the degree of freedom of molecule is searched using an incremental algorithm whose examples are FlexX¹⁸⁶ and Glide¹⁸⁷⁻¹⁸⁹. The stochastic search method, also known as a random search, randomly moves from one region of conformational space to another unconnected region in a single step. It searches the conformational space either by changing the cartesian coordinate or the torsion angle of rotatable bonds. AutoDock¹⁹⁰ and GOLD¹⁹¹ are examples of the usage of genetic algorithms based on a computational model of Darwinian evolution.

After conformational sampling generation, the generated poses are ranked by a scoring function. The scoring functions are equations that approximate the binding affinity of the tested ligands towards the target. The classical scoring functions are empirical, knowledge-based, and molecular mechanistic-based (also known as Force-field based). The newer scoring functions are machine learning-based functions¹⁹². An example of a knowledge-based scoring function is GOLD¹⁹¹, while AutoDock¹⁹⁰ is mechanistic-based, and Glide is the empirical-based scoring function. Machine learning-based scoring functions use random forest (RF), support vector machine (SVM), and deep learning (DL)¹⁹². Each scoring function has its strengths and limitations.

Here, we briefly describe the workflow of the Glide docking algorithm from the Schrödinger suite, which we employed in our projects. Glide uses an exhaustive (incremental) algorithm to sample the various conformations of ligands and examines the complementarity of ligand-receptor using a grid-based method, representing the shape of the receptor. The initial screen aims to locate promising ligand poses.

Selected initial screening poses will be minimized using the OPLS force field ^{174,193–198}. Then, the energy-minimized poses are scored by the Glide score. Finally, the poses are ranked using the Emodel score, which is a combination of Glide score, non-bonded interaction energy, and ligand conformation energy. Glide has the feature to tune the van der Waals radii of protein or ligand atoms ¹⁸⁸. We conducted Glide docking^{187,188} in our projects with default settings. Docking was performed in a cubic grid (13 Å) defined by the co-crystallized ligand present in each selected structure, using the standard-precision (SP) ¹⁸⁷ and extra-precision (XP) ¹⁸⁹ level of accuracy. For each crystal structure, we conducted redocking experiments, where our aim was to evaluate the ability of the docking to reproduce the experimental ligand conformation.

Proteins are flexible molecules, displaying a wide range of motions from local side-chain rotations to global conformational changes, generating an ensemble of conformation that may accommodate a wide variety of ligands. While most of the docking algorithms treat the protein as rigid despite its flexible nature, some algorithms take the protein flexibility into account such as Induced Fit docking from Schrödinger ^{199,200}. However, such a method is computationally expensive. A more computationally efficient approach is ensemble docking method ¹⁴⁸. This docking programme dock flexible ligands against multiple conformation of the protein rather than only a single conformation. In ensemble docking, the ensemble of protein structures can be obtained from either different crystal structures or the different snapshots of an MD simulation trajectory. The structures are superimposed, and ligands are docking within the protein binding site. The ensemble-based screening takes protein flexibility into account, improving the predictive power of structure-based drug discovery. This can be applicable when a set of crystal structures is available for the target protein. However, only a limited number of proteins possess an extensive set of crystal structures of different and relevant conformations of the bound and unbound receptor. To validate the docking result, usually known active ligand(s) is docked before the docking of the intended molecule (redocking). In addition, the visual inspection of the generated pose is important when validating a docking pose. There is agreement on potential inaccuracies of the docking method. Hence, one of the most crucial factors in docking result assessment is visual inspection, although it has not been introduced a clear guideline for the visual assessment of docking results. In this regard, a survey conducted by Fischer *et al.*, (2021) ²⁰¹, collected 93 expert opinions from the academy

and industry sector, to identify agreeing principles for visual inspection along with its limitations. The result showed that the most frequent criteria in the visual assessment of docking pose are shape complementarity^{202,203}, hydrogen bonds and hydrophobic contacts and the least reliable criteria were scoring function. If there is a huge number of test ligands, in this case, generated poses can be filtered based on the docking score, and among those, the best score poses could be visually inspected.

In our studies, to evaluate the capability of the docking algorithm (Glide) in locating the ligands within the LBD, we redocked the cocrystal ligands. The redocking poses resulted in top-ranked docking poses with a heavy atom root-mean-square deviation (RMSD) below 1 Å in comparison with the original conformation.

The inspection was performed to evaluate our docked binding poses using the active ligand as the reference binding mode. Similarly, shape complementarity, hydrogen bonds and hydrophobic contacts were considered as decisive criteria when assessing the docked poses, and to a lesser degree, we took the docking score (glide score and Emodel score) into account. The docking poses of EDs (*e.g.*, nonylphenol) were diverse, binding in multiple conformations within the LBP and displaying little to no hydrogen bond interactions. We hypothesized that the EDs could have multiple binding modes, *i.e.* a single static model for the NR-ED interaction would not be sufficient to describe/predict its outcome²⁰⁴. We attributed this as a limitation of the docking itself and proceeded to evaluate the stability of the ligands within the LBP by using molecular dynamics simulations.

2.5.1 Limitation of molecular docking

Although molecular docking has several strengths, one of which is the method's ability to screen large compound databases at low cost and time, compared to experimental techniques, such as HTS, some limitations remain.

One of the docking limitations relates to the empirical nature of scoring functions.

The reliable prediction of ligand binding affinity is still challenging. However, the rising number of protein-ligand structures with measured binding affinities along with the increasing datasets containing active, decoy, and true inactive compounds, has facilitated the implementation of machine-learning techniques to obtain better-performing scoring functions.

Docking is also limited unless key interactions with the protein are already known either by previous crystal structures, reliable modelling on homologues and/or site-directed mutagenesis. Although molecular docking shows its efficiency in its prediction status, it needs to be subsequently validated by experimental methods²⁰⁵.

Another caveat is that while docking is a tool for virtual screening of the ligand and hit identification, it only demonstrates the binding affinity or binding energy of the complex rather than conformational change. To obtain further conformational space of the complex, we need to simulate the movement of the system through time. Further validation of a docking pose can be achieved by conducting an MD simulation of the selected docking pose,²⁰⁶ which also refines and improves the precision of docking.

In our reported work, selected binding poses (explained in section 2.3) were submitted to short MDs to examine ligand stability within LBP as final validation. The MDs protocol will be explained in the next section.

2.6 Molecular Dynamics Simulations and its challenges

NRs-LBD's high flexibility is critical for their function. Although the docking method can provide invaluable information for the initial study of the structure-function relationship, its limitations (section 2.5), such as their static model, provide an incomplete vision of biological systems. The biological function of an NR is a result of a conformational ensemble shifting from one state to another in dynamic equilibrium. The molecular Dynamic simulations (MDs) method can be utilized to overcome the shortcomings of the docking and study the ligand-induced conformational changes, as well as the relevant peptide recruitment surfaces. MD simulations use the initial model structure as the starting point to generate an ensemble of conformations that can reliably represent the protein-ligand behaviour in solution^{207,208}. MD simulations were first developed during the 1950s²⁰⁹. This approach follows the law of classical mechanics (Newton's second law, Eq. 1), the net force (F) for an atom (i) is equal to its mass multiplied by its acceleration (a). This corresponds to the empirical potential energy function $V(x)$.

$$\text{Eq. 1. } F = m_i a_i = -\frac{\partial V(x(t))}{\partial x_i(t)}$$

The generated atomic coordinates over time form the MD dynamics trajectory of a molecule. Since the conformational flexibility of a protein is extremely related to the surrounding environment, the MD contributing factors, apart from the biomolecules, such as solvent, temperature, and ions are extremely relevant. Different models can be used to represent water molecules in molecular dynamics simulations.

The most commonly used water models are: **TIP3P** (Transferable Intermolecular Potential 3 Points)²¹⁰ is a three-site model where each water molecule is represented by three interaction sites: two positive charges for hydrogen atoms and one negative charge for the oxygen atom. **TIP4P** (Transferable Intermolecular Potential 4 Points)²¹¹ is an extension of the TIP3P model. In this model, an additional interaction site is added to the oxygen atom instead of the electron's lone pairs. This model provides a more accurate description of water's properties compared to TIP3. Another used water model is **SPC** (Simple Point Charge)²¹² which is a three-site model that represents each water molecule as a point charge located at the oxygen atom and two points

charges representing the hydrogen atoms. This model is simpler than TIP3. **SPC/E** (Simple Point Charge/Extended) ²¹² is an extension of the SPC model where an additional site is added to the oxygen atom representing polarization effects ^{210,213–215}. This model improves the description of water's properties, especially at higher temperatures and densities. CHARMM includes its own water model, which is similar to the TIP3P model but includes Lennard-Jones interactions between hydrogens and oxygens to improve the description of the liquid phase ^{216–218}. The latest version is CHARMM 36m (sTIP3P) ²¹⁸, optimized based on the original TIP3P for proteins and lipids. Overall, the system being studied affects our choice of water model.

In terms of how the potential energy (V) is calculated, MD simulations can be divided into classical molecular mechanics and quantum mechanics-based MD simulations. As mentioned earlier, quantum mechanics-based MD simulation are resource intensive. Hence, currently the classical molecular mechanics-based simulation is the most common MD simulation method. There exist several MD simulations packages such as NAMD ²¹⁹ Gromacs ²²⁰ and Amber ¹⁷³. The steps involved in MD simulations are as follows:

1. Input preparation: in this step we need a prepared and minimized structure (section 2.3). Water is another element to be added around the molecule implicitly or explicitly. Adding ions (NA⁺, K⁺, CL⁻) neutralize the system. Hence molecule is in a neutral environment at pH 7.0. In addition, the system can be set up in salt as well. Next step for setting up the simulation system is applying the Periodic Boundary Conditions (PBC). PBC is a technique used in computer simulations, to mimic the behaviour of an infinite system through a finite simulation cell. This approach enables particles at one edge of the simulation cell to interact with particles on the opposite edge, mimicking an infinite repeating lattice. PBC allows for the generation of more representative and accurate statistical ensembles, such as the canonical ensemble (NVT²²¹) or the isothermal-isobaric ensemble (NPT²²¹). Temperature, pressure, and simulation time are other components of the system to be set up. In addition, we need to provide the simulation program with information including Topology, parameter, and force field (section 2.4).
2. Energy minimization/ equilibration: The purpose of this step is to relieve steric clashes existing in the system due to all participants elements (macromolecules, small molecule, water, and ions) to minimize the system from high energy contacts and define a starting point for equilibration step. It helps the molecule to be relax and reach

to a stable state. During minimization, molecule search for nearby local minima on the potential energy surface and to obtain favourable bond lengths and bond angles minimizing the force pulling or pushing atoms together, removing the possible steric clashes and adjusting the distribution of solvent molecules ²²². Then, in equilibration phase, the molecule escapes local minima with low energy barrier to find the lowest energy minimum (section 2.4.1). The aim of this step is to equilibrate the energy of the system, relax the structure and the solvent. It is important to select an appropriate ensemble class for equilibration to draw a meaningful conclusion from the MD simulations. The commonly used thermodynamic ensembles in MD simulations are **NVE** ²²¹ which describes an isolated system with a constant number of particle (N), Volume(V) and Energy (E). In the **NVT** ²²¹ ensemble, the system is in thermal contact with a heat reservoir allowing the exchange of energy at a constant Temperature (T). Particle (N) and Volume(V) are fixed. The **NPT**²²¹ ensemble, on the other hand, aims to equilibrate the system's internal pressure with the external pressure by permitting the exchange of energy between the system and the environment alongside allowing variation in volume.

3. Production: In this step, a series of structure at a specified times, velocities, and system coordinates are recorded as a trajectory.

Time is one of the main parameters to be considered when studying the dynamic of a system through MD simulations. The simulation times spans from femtoseconds (fs) to milliseconds (ms) depending on the size of the system and computational resource and the longer time scale provides more stable system (Fig. 7). However, bond vibration and side chain rotation happen below than nanosecond (local flexibility), larger motions occur at microsecond to millisecond. Depending on the region of interest and available computational power one can determine the timescale of MD simulations ²²³.

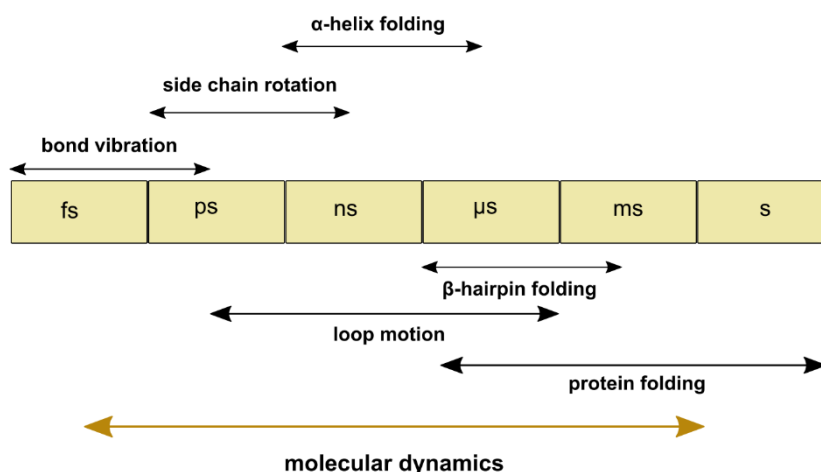


Fig. 6 Comparison of time scale for protein motion. Size of the protein determine the time scale for MD simulation. Bond vibration and side chain rotation happen in the femtosecond (fs) to nano- second (ns) timescale and protein folding happens in microsecond (μ s) to milliseconds (modified from Werner *et al.* 2012 ²²³).

Multiple replicas can increase the reliability of the observations. MD simulations are no exception to this rule. The minor differences in initial velocity, floating-point precision, or underlying hardware can lead each simulation replica in different paths of conformational landscape with different free energy landscapes. Knapp *et al.*, (2018) tested this concept based on 310 μ s of simulation time for a small system (10 amino acids and for a T-cell receptor/MHC system). They randomly chose a sample of each replica set and estimated the reproducibility and reliability that can be achieved by a given number of replicas at a given simulation time. They found that the single simulations are not reproducible whereas multiple shorter replicas are more reliable. They suggested five to 10 replicas as the rule of thumb ²²⁴.

Another crucial subject in simulation is sampling. It is important to assure if system under simulation visits all conformational states. These days, different enhanced sampling methodologies such as Replica exchange molecular dynamics (REMD) ^{225,226} and metadynamics ²²⁷ have been developed where multiple copies of same simulations are running at different temperatures, and replicas exchange their temperature at regular intervals. This exchange of temperatures between replicas allows the system to discover conformational space efficiently, preventing the system from being trapped in local energy minima, thus improving the efficiency of MD simulations ^{225,226}. Overall, MD simulations provide insights into the dynamics of

proteins and help us understand how they bind with small molecules or other proteins, which, in turn, aids in the development of selective inhibitors or activators.

To the best of our knowledge, the timescale of our simulations appears to be suitable for the study of nuclear receptors and is, in fact, the longest available.

2.7 Analysis of MD simulation trajectories

The most commonly used procedure to analyse a molecular dynamic simulation is an overview of the changes of root mean square deviation (RMSD) ²²⁸ and root mean square fluctuation (RMSF) ²²⁹, along the time and residues, respectively. Both approaches use a reference structure (starting conformation of simulation). For RMSD calculation, the deviation of each atom coordinates in newly generated conformation from the corresponding atom coordinate in the reference structure is calculated as a function of the simulation time. RMSF provides information about the fluctuation of residue or atoms throughout the simulation. These two parameters provide information about the validity of the simulation rather than offering deep insight into the simulation data.

Given the big number of atom coordinates, we have big data. To gain a better idea of the dynamic and direction of motion of protein data, it is good to implement a more specific algorithm such as principal component analysis (known as PCA).

2.7.1 Principal Component Analysis

PCA is an analysis method to capture the significant motion from the simulation trajectory. PC method can display the functionally important motions of proteins in the space spanned by a small number of their huge conformational modes, the reason that is called also essential dynamic. PC is a multivariate statistical technique which reduces the dimension of the data. The method is using a covariance matrix which is a linear transformation of data ²³⁰. Usually, the covariance matrix is generated by C α atom coordinate or C α torsion angles. The covariance matrix is decomposed to obtain its eigenvectors and eigenvalues. The eigenvectors represent the directions or principal components. The eigenvalue represents the amount of variance explained by each principal component. In MD analysis PCA is the projection of the MD data onto an eigenvector. Another statistical method to analyse data is time-lagged component analysis. This method is also the linear transformation to reduce the

dimension of the data where it also identifies the slow order parameters of the simulation.

In terms of our NR work, PCA was conducted for the backbone atoms using GROMACS tools (version specified in the specific papers) using the pre-implemented scripts `gmx_covar` and `gmx_anaeig`. For GROMACS analysis, the Desmond trajectories were aligned and transformed to xtc-format, keeping only backbone atoms. Figures describing the extreme motions can be generated and visualized using PyMOL-script `Modevectors`. Interestingly, one can decompose the principal component (PC) projection into an extreme motion, that describes the related protein movements. Results have been demonstrated in publication I, III, and IV.

2.7.2 Markov State Modelling

Due to the massive amount of generated simulation data, visual inspection of trajectories is an almost impossible task, besides not providing quantitative description of the system. Hence, statistical methods, such as Markov States Models (MSMs)^{231–233}, can be implemented to generate a quantitative model of dynamics of the system. It provides “state” and “rate” view of conformational space helping to capture all possible behaviour of macromolecule through a dynamic trajectory. The aim is to build a model which can predict kinetic, thermodynamic and structure.

There are two steps to build the model: 1) To define state, 2) To estimate rate between states. The method uses clustering algorithms and builds the matrix of counts, normalizes the count to construct transition probability matrix. This matrix implies the probability of jumping between states after every lag time and is the key dynamical model obtained from MSMs.

In MSMs method, the dynamics are approximated as a Markovian process between distinct microstates (Fig. 8). MSMs have been used to improve ensemble docking to optimize a specific conformation in a ligand, and to characterize ligand-binding processes as well as inactive-to-active transitions in signalling proteins^{234,235}.

In practice, the first step is to extract molecular features from the raw data obtained from MD simulation and transforming the features into a low dimensional subspace. Dimensionality reduction of the trajectory is performed by means of a dimensionality reduction method. Many of the studies used the time-lagged independent component analysis (tICA)²³³ to find a suitable low-dimensional representation of the system. The

low dimensional subsets would be discretised into state decomposition. Like the PCAs, TICA is a coordinate transformation. It yields an optimal slow subspace which decorrelated (at lag time τ) from the remaining coordinates ²³³. In TICA, different parameters can be used such as coordinates, distances, or torsion of selected atoms. One can apply those parameters to a small region of structure or the whole structure. For the rate step, a maximum likelihood or Bayesian MSM is estimated from the discrete trajectories. Model is validated based on Chapman-Kolmogorov ²³⁶. Next the stationary and kinetic properties of the MSM is analysing. Metastable microstate is built and applying transition path theory (TPT) to identify the pathways of conformational change. Expectation value will be calculated for experimental observation. PyEMMA software package ²³⁷ has made MSMs accessible to a wide range of academic researchers.

Against the advantages of MSM method, still several challenges remain. These challenges are related to the decomposition of conformational state. Another subject that needs to be take care of is related to the seeding procedure in MD simulations. While seeding provides a great variety of relevant conformations beneficial for MSM building, one should be aware that the simulations, starting from a seed that is kinetically far from the other starting points, may never overlap with the rest. This affects the determination of the transition rates between states ²³⁸.

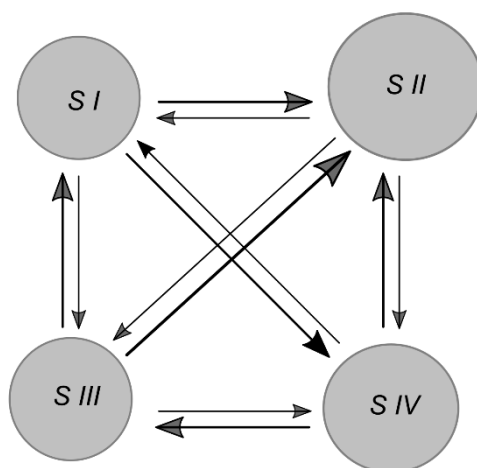


Fig. 7 Hypothetical scheme of Markov State Models. Metastable states are depicted as circles. 'S' stands for state. The size of each circle is in proportion to the probability of state occurrence, and state transitions are shown by arrows. The thickness of the arrows is related to the probability of the state transition.

We implemented this method, for example, to investigate the conformational behaviour of PXR upon antagonist binding to discriminate the agonistic and antagonist mode of PXR-LBD conformation in complex with a full agonist (SR12813) and competitive antagonist (compound 100) in publication I.

2.7.3 Free energy (prime MM-GBSA)

Application of computational method is a trade-off between accuracy and efficiency. As we discussed earlier, docking, which is the most common approach to estimate binding affinity of ligand, is an efficient approach but not highly accurate, particularly when it comes to discriminate between ligands differed with <6 KJ/mol in their ΔG ²³⁹. Poisson–Boltzmann or generalized Born and surface area continuum solvation, MM/PBSA and MM/GBSA, are well-known approaches combined with molecular mechanics energies to predict ligand-binding affinities by estimating the free energy of binding of the ligand to the biological macromolecules. The MM/PBSA was established in 1990 and developed through the years. The method can be employed in several tasks such as protein-protein interactions, conformer stability, rescoring, lead optimization, enhancing docking results, etc.^{240–249}

The method consists of several energy terms, namely: electrostatic term, solvation-free energy (polar solvation term, non-polar term). Binding free energy change of a system is calculated as follows (Eq. 2,²⁵⁰):

$$\text{Eq. 2. } \Delta G_{bind} = \Delta H - T\Delta S$$

where the ΔH is the enthalpy change in the system, T is the absolute temperature (in K), and ΔS is the entropy change in the system.

In computational study, the Binding free energy changes are calculated through a thermodynamic cycle as follows (Eq. 3²⁵⁰) :

$$\text{Eq. 3. } \Delta G_{bind,solv} = \Delta G_{bind,vacum} + \Delta G_{solv,complex} - (\Delta G_{solv,ligand} + \Delta G_{solv,protein})$$

Where the solvation free energy is divided into two terms: polar and non-polar (Eq. 4)

$$\text{Eq. 4. } \Delta G_{solv} = \Delta G_{pol} + \Delta G_{nonpol}$$

The binding free energy of a protein-ligand complex is calculated as follows (Eq. 5):
The $\langle \dots \rangle_i$, represents an average over i snapshots generated through MD simulations.

$$\text{Eq. 5.} \quad \Delta G_{bind} = \langle G_{complex}(i) - G_{protein}(i) - G_{ligand}(i) \rangle_i$$

Decomposition of each Binding-free energy (Eq. 6):

$$\text{Eq. 6.} \quad G = E_{int} + E_{el} + E_{vdw} + G_{pol} + G_{nonpol} - TS$$

Where E_{int} denotes the internal energy, E_{el} stands for electrostatic energy, and E_{vdw} for Van der Waals energies. The G_{pol} and G_{nonpol} are the polar and nonpolar solvation-free energies, respectively, T is the absolute temperature and S is the entropy.

MM/PBSA and MM/GBSA are based on implicit solvation models which the estimation is not as accurate as techniques with explicit solvent models²⁵⁰ but they are computationally more efficient. Another advantage of these methods over the explicit solvent method is the possibility to break down the resulted free energy into subcomponents and calculate their contribution to the total free energy separately. The electrostatic term, charges used for the protein and ligand are deterministic. It depends on the dielectric constant of protein.

The original entropy term is calculated by removing water and residues closer than 8 Å to the ligand and minimizing the rest. The entropy gives the largest uncertainty. Several approaches have been suggested to replace MM force field with QM either for ligand or for the complex. However, inconsistency between energy function used for simulation and energy calculations is problematic. MM/GBSA can determine the ligand efficiency, where the magnitude of order is divided by number of heavy atoms of ligand.

Overall, due to the development of MM/PBSA, MM/GBSA, users have more flexibility in using different constant dielectric, parameters for non-polar energy, radii for the MMPB or MMGB calculations, whether to incorporate entropy term, and whether to conduct MD or minimization. The study conducted by sun *et al* (2014) on tyrosine kinases' structure docked with AutoDock showed that the accuracy of the MM/GBSA method in binding affinity calculation is better than docking score²⁴². Hence, it is suggested to be a good approach for post-processing of docked structures. It has been

often recommended to use one snapshot of the minimized complex for the calculation of the MM/GBSA rather than ensemble of MD snapshots due to less computational cost, but in this case the dynamic of the system will be ignored which leads to the loss of statistical precision highlighting the importance of MD sampling ²⁵¹.

To gain insights into the binding affinity of EDs with NRs, we utilized MM/GBSA. This method is computationally efficient compared to rigorous alchemical perturbation methods like free energy perturbation. It is also more robust than molecular docking based solely on scoring functions ²⁴⁹. Another advantage of this approach is its ability to incorporate the influence of explicit solvent dynamics on ligand binding, thereby accurately estimating the entropic contribution. This feature is particularly valuable given the highly hydrophobic nature of our EDs. A similar strategy was employed to investigate the binding dynamics of per- and polyfluoroalkyl substances (PFASs) using short MD simulations ²⁵². The study revealed that the predicted absolute binding energies of PFASs were lower than the corresponding experimental values; however, a strong correlation was still observed.

2.7.4 Interaction analysis

Knowing about the interaction occurring within a molecule, both intra-molecular that stabilizes the protein conformation and intermolecular that happens between ligand and target, provides insights into the thermodynamics and kinetics of a system. MD simulations provide valuable insight into the dynamic system and displays the interactions as a function of time (frequency of interaction) using the trajectories of atoms in each time step. The commonly investigating interactions include hydrophobic interactions (including π - π interactions) and Hydrogen bonds, which hold significance in maintaining the stability of structural water within the ligand binding pocket and simultaneously influencing the stability of both the ligand and target. Additionally, other noteworthy interactions include water bridges, ionic interactions, and Van der Waals interactions. It's worth noting that diverse software applications employ distinct definitions when calculating these interactions. This information could be implemented on different occasions from deciphering the conformational alteration of a biomolecule to hit-lead optimization. We utilized such information to understand the reason for structural alteration of NR, reported in our publications.

2.7.5 Angle and distance calculation

Angle calculation provides insight into the orientation (Torsion angle) or bending of the molecule (bond angle). We utilised this calculation to understand our ligand properties using Schrödinger suite (Schrödinger LLC).

Distance calculation over an MD trajectory provides information into the interaction that occurs between atoms and in turn reflects the stability of the molecule (*e.g.*, the ligand in LBP), as well as structural changes.

In addition, to validate the PCA observations, we chose to analyse the distances between specific secondary structure elements. These distances were calculated using Maestro event analysis tool (Schrödinger, LLC) over the generated trajectories or using their centres of mass with the Maestro script `trj_asl_distance.py` (Schrödinger LLC). For most of the nuclear receptors, we monitored the distance between the H3-helix α AF-2/H12 residues, as well as the H12 folding, among other metrics. Together with the hydrophobic/hydrogen bond profile for each ligand along the simulations, these geometrical indicators can be used to cluster the compounds between potential NR-activators and non-activators. Further statistical studies will allow us to gain a picture of the characteristics that lead a particular ligand to bind to a given nuclear receptor.

In our approaches, described in the different projects, each system, composed of independent microsecond long simulations, were carefully examined through various techniques such as clustering, assessment of protein-ligand interaction frequencies, principal component analyses (PCA), and evaluation of geometric distances. The PCA analyses yielded relevant geometric distances by comparing the known binders with the test compounds. However, irrespective of the PCA's suggested regions, our assessment encompassed H12-H3 distances, H12 folding (SSE%), and H6 folding/distances.

2.7.6 Ligand trajectory analysis

When RMSD and RMSF provide general information about ligand stability and its conformational changes, there exist other parameters offering detailed insight into the ligand behaviour and its affinity towards the target. Solvent Accessible Surface Area (SASA), the radius of gyration and angles, among others, are the decisive criteria. SASA measures the surface area of a molecule which is accessible to solvent and in units of a square angstrom (\AA^2). The radius of gyration (R_g) measures the extendness of the molecule. Throughout the simulation, a ligand undergoes conformational strain to retain its protein-bound conformation. Therefore, gathering data on ligand torsions is beneficial to monitor the conformational evolution of its rotatable bonds throughout simulation time.

3. Experimental Methods to Analyse NR Activity

A variety of biochemical and biological assays is available, that help to explore and understand the structure, function, and genetic changes related to nuclear receptors and also elucidate of NRs ²⁵³. Examples are (Fluorescence Recovery After Photobleaching (FRAP) ²⁵⁴ and Fluorescence Loss in Photobleaching (FLIP) which provides insight into the dynamics and interaction of the molecules in various cell process ²⁵⁵, and chip-on-chip methodology to identify receptor-regulated genes ²⁵⁶. Study of NR-coregulator which is a major field of NR research ^{31,39,43,257,258}, methods for screening for binding partners and for quantifying specific receptor-target protein interactions ²⁵⁹, understanding the role of phosphorylation on receptor function ^{260,261}, advances in tissue-selective gene targeting and knock-out strategies for generating mouse models of receptor function in vivo ²⁶², and studying genetic alterations in hormone-dependent cancers ²⁶³ are some examples. In the following, several relevant biochemical and cell-based assays frequently used in NR research field are explained.

3.1 Time Resolved Fluorescence resonance energy transfer

Time-resolved fluorescence resonance energy transfer (TR-FRET) is a commonly used experimental technique to investigate the binding between nuclear receptor and ligand or protein-protein interaction (e.g., nuclear receptor-coregulator interaction). In this method, two fluorescent molecules are utilized: a donor fluorophore and an acceptor fluorophore. When the two molecule interacts, donor and acceptor fluorophores are brought together. When the doner is excited, it transfers its emission energy to the acceptor. This event leads to the emission of fluorescence at a specific wavelength. and the fluorescence emitted by the acceptor fluorophore is measured at a delayed time after the excitation pulse. This delay reduces the background noise and autofluorescence, resulting in a higher signal-to-noise ratio with improved sensitivity ²⁶³. FRET is transferring of energy from a donor fluorophore in an excited state to a nearby acceptor fluorophore. Output is proportional to amount of binding ²⁶⁴.

3.2 Luciferase reporter gene transactivation assay

Luciferase-based reporter assays measure the activity of the nuclear receptors on the target gene expression upon ligand binding, within a relevant cell system. The assay is very sensitive and reproducible. To determine if a protein (or Protein-ligand combination) can activate (or suppress) the transcription of a gene of interest, recombinant DNA technology is used to produce a construct in which the gene's promoter is placed adjacent to a luciferase reporter gene. The cultured cells are transfected with this construct, as well as a construct coding for the protein, in our examples the NRs. If the NR can activate transcription, the cell will translate and produce the luciferase reporter (often a firefly's luciferase). The amount of produced luciferase can be quantified using a luminometer. Protocols involve transient transfection of the receptor and a response element reporter gene construct ^{76,265}. Other protocols also have an additional luciferase (Renilla), which is constitutively and constantly being produced, to normalize by the number of transfected cells and act as a transfection control. Many cell lines that have been introduced as candidate recipients of these vectors, including CHO, HuH7, MCF-7, HEK293, HepG2 and Caco-2 cells. This method can identify NR activators and upon pre-treatment with known agonist and unknown ligands in concentration response, also identify antagonists ²⁶³.

3.3 Coregulator-recruitment (mammalian two-hybrid, CARLA)

An alternative transactivation assay system is the mammalian two-hybrid system which is a technique to detect protein-protein interaction in cells. The result is interpreted by expression or repression of reporter genes. Cells are treated with NR ligands. In this method the DBD of the yeast transcription factor GAL4 binds to specific NR response elements. When LBD of desired gene are fused to DBD, it promotes co-activator binding. The interaction between the NR and its co-activator is detected based on a reporter gene containing multiple copies of the GAL4 upstream activating system. A set of agonists and inverse agonist were identified to bind to the human CAR using similar system ²⁶⁶.

4. Concluding Remarks and Outlook

NRs are important intracellular receptors with the role of transcriptional factors that regulate several genes involved in physiological processes including development, differentiation, metabolism, and systemic homeostasis. These receptors are ligand-induced activated proteins and play their role in interaction with multiple cofactor proteins. Their first structure was determined in the 90s by X-ray diffraction and the number has been increasing considerably. Due to their essential role on the whole organism homeostasis and organ specificity, NRs are candidates for therapeutic targets for many human diseases. Therefore, the discovery and development of any modulator which regulates the activity of receptors is significant. To achieve this goal, it is important to have deep insight into the NR-ligand interactions, as well as cofactor proteins. Meanwhile, the computational method as an approach accelerates the process of drug discovery and development and advances in computer efficiency provides *in silico* time-efficient methods and technique for such studies. Therefore, we studied NRs by implementing *in silico* screening and structure-based drug design followed by *In vitro* experiment to validate our computational simulation. We exclusively employed long-timescale atomistic MD simulations. It is of great importance to take the protein flexibility into account when trying to understand the NR active/inactive conformation since one challenge in NR modelling is its plasticity.

MD simulation is recognized as a relevant and underused tool to elucidate NR conformational rearrangement upon ligand binding and also discover the ligand entry/exit pathway²⁶⁷. NRs are flexible molecules with highly conserved domains. This flexibility is highlighted in PXR, as exemplified by structural data and earlier studies in the presence of agonists^{268,269}. However, our PXR study displayed slightly lower flexibility in the presence of a partial antagonist (compound 100). Moreover, with the implementation of a computational approach, we could identify important residues interacting with modulators where mutation studies confirmed and expanded our knowledge of those interactions.

Our studies demonstrated the integration of computational technologies with experimental testing and validation in the field of drug discovery, leading to enhance efficiency in identifying therapeutic candidates. It is noteworthy that while computational approaches offer valuable insights, they do not guarantee absolute

certainty in their predictions. For example, the homology modelling process encompasses various potential sources of error, such as inadequate template selection, improper alignment of the target sequence with the template, inaccurate folding of loops, and suboptimal selection of side-chain conformations. It is evident that the primary sources for errors in this process are the inappropriate choice of template and misalignment, particularly in cases where the sequence identity is below 30%. Hence, it is crucial to complement computational methods with *in vitro* and *in vivo* assays for comprehensive validation and assessment of drug design ²⁷⁰.

Our computational modelling of NRs could enhance our understanding of receptor conformational rearrangement due to protein-ligand interaction (PXR and CAR study) and protein-protein interaction (FXR study). Meanwhile, our novel NR modulators could propose potential therapeutic for NR-related diseases.

However, we remember that nuclear receptors behave in a tissue-specific manner, having the ability to regulate different sets of genes based on the transcriptional complexes they form. This leads to the selection of cell-specific enhancers and results in distinct physiological consequences. Hence, the screening of modulators for nuclear receptors will equip future studies with essential tools for developing a detailed understanding of the regulation of nuclear receptors. Nuclear receptors are controlled not only by ligand and coregulator but also undergoes modifications such as epigenetic, ubiquitination, and protein maturation, as well as other unknown mechanisms that need to be considered during NRs study projects. Through decades, various studies have been conducted in the field of nuclear receptors, including the identification of new target genes, elucidation of mechanisms of gene regulation, and development of new drugs ^{75,271}. Nevertheless, the needs for designing ligands with higher affinity and selectivity as well as technologies that can help to understand the complex interactions remain.

Moreover, during CADD implementation, one should be aware of several potential pitfalls that might happen, otherwise, it generates false prediction which would provide false input data for the experimental pipeline, resulting in invalid prediction and consequently error propagation, which waste time and resources in the other way around. Overall, these pieces of work emphasize the utilization of computational approaches in Structure-Based Drug Design (SBDD), such as molecular docking and MD simulations, for the discovery of innovative bioactive substances and receptor

conformational changes in nuclear receptors considering the limitations of these methods.

5. References

1. Mustonen, E.-K. *et al.* Target Hopping from Protein Kinases to PXR: Identification of Small-Molecule Protein Kinase Inhibitors as Selective Modulators of Pregnane X Receptor from TüKIC Library. *Cells* **11**, 1299 (2022).
2. Zhou, J. & Cidlowski, J. A. The human glucocorticoid receptor: One gene, multiple proteins and diverse responses. *Steroids* **70**, 407–417 (2005).
3. Ashburner, M., Chihara, C., Meltzer, P. & Richards, G. Temporal control of puffing activity in polytene chromosomes. *Cold Spring Harb Symp Quant Biol* **38**, 655–662 (1974).
4. Yamamoto, K. R. Steroid receptor regulated transcription of specific genes and gene networks. *Annu Rev Genet* **19**, 209–252 (1985).
5. Hollenberg, S. M. *et al.* Primary structure and expression of a functional human glucocorticoid receptor cDNA. *Nature* **318**, 635–641 (1985).
6. Green, S. *et al.* Human oestrogen receptor cDNA: sequence, expression and homology to v-erb-A. *Nature* **320**, 134–139 (1986).
7. Miesfeld, R. *et al.* Genetic complementation of a glucocorticoid receptor deficiency by expression of cloned receptor cDNA. *Cell* **46**, 389–399 (1986).
8. Petkovich, M., Brand, N. J., Krust, A. & Chambon, P. A human retinoic acid receptor which belongs to the family of nuclear receptors. *Nature* **330**, 444–450 (1987).
9. Evans, R. M. The steroid and thyroid hormone receptor superfamily. *Science* **240**, 889–895 (1988).
10. Laudet, V. Evolution of the nuclear receptor superfamily: early diversification from an ancestral orphan receptor. *J Mol Endocrinol* **19**, 207–226 (1997).

11. Porter, B. A., Ortiz, M. A., Bratslavsky, G. & Kotula, L. Structure and Function of the Nuclear Receptor Superfamily and Current Targeted Therapies of Prostate Cancer. *Cancers (Basel)* **11**, 1852 (2019).
12. de Almeida, N. R. & Conda-Sheridan, M. A review of the molecular design and biological activities of RXR agonists. *Medicinal Research Reviews* **39**, 1372–1397 (2019).
13. Mangelsdorf, D. J. *et al.* The Nuclear Receptor Superfamily: The Second Decade. *Cell* **83**, 835–839 (1995).
14. Klinge, C. M., Bodenner, D. L., Desai, D., Niles, R. M. & Traish, A. M. Binding of type II nuclear receptors and estrogen receptor to full and half-site estrogen response elements in vitro. *Nucleic Acids Research* **25**, 1903–1912 (1997).
15. McKenna, N. J. & O'Malley, B. W. Combinatorial Control of Gene Expression by Nuclear Receptors and Coregulators. *Cell* **108**, 465–474 (2002).
16. Mutoh, S. *et al.* Dephosphorylation of threonine 38 is required for nuclear translocation and activation of human xenobiotic receptor CAR (NR1I3). *J Biol Chem* **284**, 34785–34792 (2009).
17. Mutoh, S. *et al.* Phenobarbital indirectly activates the constitutive active androstane receptor (CAR) by inhibition of epidermal growth factor receptor signaling. *Sci Signal* **6**, ra31 (2013).
18. Rastinejad, F., Huang, P., Chandra, V. & Khorasanizadeh, S. Understanding nuclear receptor form and function using structural biology. *J Mol Endocrinol* **51**, T1–T21 (2013).
19. Nagy, L. & Schwabe, J. W. R. Mechanism of the nuclear receptor molecular switch. *Trends in Biochemical Sciences* **29**, 317–324 (2004).

20. Robinson-Rechavi, M., Garcia, H. E. & Laudet, V. The nuclear receptor superfamily. *Journal of Cell Science* **116**, 585–586 (2003).
21. Bertrand, S. *et al.* Evolutionary Genomics of Nuclear Receptors: From Twenty-Five Ancestral Genes to Derived Endocrine Systems. *Molecular Biology and Evolution* **21**, 1923–1937 (2004).
22. Pawlak, M., Lefebvre, P. & Staels, B. General molecular biology and architecture of nuclear receptors. *Curr Top Med Chem* **12**, 486–504 (2012).
23. de Almeida, N. R. & Conda-Sheridan, M. A review of the molecular design and biological activities of RXR agonists. *Medicinal Research Reviews* **39**, 1372–1397 (2019).
24. Chandra, V. *et al.* Structure of the intact PPAR- γ -RXR- α nuclear receptor complex on DNA. *Nature* **456**, 350–356 (2008).
25. Küblbeck, J. *et al.* The EDCMET Project: Metabolic Effects of Endocrine Disruptors. *International Journal of Molecular Sciences* **21**, 3021 (2020).
26. Küblbeck, J., Niskanen, J. & Honkakoski, P. Metabolism-Disrupting Chemicals and the Constitutive Androstane Receptor CAR. *Cells* **9**, 2306 (2020).
27. Gore, A. C. *et al.* EDC-2: The Endocrine Society’s Second Scientific Statement on Endocrine-Disrupting Chemicals. *Endocrine Reviews* **36**, E1–E150 (2015).
28. Latini, G., Gallo, F. & Iughetti, L. Toxic environment and obesity pandemia: Is there a relationship? *Ital J Pediatr* **36**, 8 (2010).
29. Neel, B. A. & Sargis, R. M. The Paradox of Progress: Environmental Disruption of Metabolism and the Diabetes Epidemic. *Diabetes* **60**, 1838–1848 (2011).
30. Hu, X. & Lazar, M. A. The CoRNR motif controls the recruitment of corepressors by nuclear hormone receptors. *Nature* **402**, 93–96 (1999).

31. Glass, C. K. & Rosenfeld, M. G. The coregulator exchange in transcriptional functions of nuclear receptors. *Genes Dev.* **14**, 121–141 (2000).
32. Murakami, S., Nagari, A. & Kraus, W. L. Dynamic assembly and activation of estrogen receptor α enhancers through coregulator switching. *Genes Dev* **31**, 1535–1548 (2017).
33. Heery, D. M., Kalkhoven, E., Hoare, S. & Parker, M. G. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature* **387**, 733–736 (1997).
34. Chandler, V. L., Maler, B. A. & Yamamoto, K. R. DNA sequences bound specifically by glucocorticoid receptor in vitro render a heterologous promoter hormone responsive in vivo. *Cell* **33**, 489–499 (1983).
35. Beato, M., Herrlich, P. & Schütz, G. Steroid hormone receptors: Many Actors in search of a plot. *Cell* **83**, 851–857 (1995).
36. Juge-Aubry, C. *et al.* DNA Binding Properties of Peroxisome Proliferator-activated Receptor Subtypes on Various Natural Peroxisome Proliferator Response Elements. *Journal of Biological Chemistry* **272**, 25252–25259 (1997).
37. Näär, A. M. *et al.* The orientation and spacing of core DNA-binding motifs dictate selective transcriptional responses to three nuclear receptors. *Cell* **65**, 1267–1279 (1991).
38. Umesono, K., Murakami, K. K., Thompson, C. C. & Evans, R. M. Direct Repeats as Selective Response Elements for the Thyroid Hormone, Retinoic Acid, and Vitamin D3 Receptors. *Cell* **65**, 1255–1266 (1991).
39. Chen, J. D. & Li, H. Coactivation and Corepression in Transcriptional Regulation by Steroid/Nuclear Hormone Receptors. *CRE* **8**, (1998).

40. Dowell, P. *et al.* Identification of Nuclear Receptor Corepressor as a Peroxisome Proliferator-activated Receptor α Interacting Protein*. *Journal of Biological Chemistry* **274**, 15901–15907 (1999).
41. Ishizuka, T. & Lazar, M. A. The N-CoR/Histone Deacetylase 3 Complex Is Required for Repression by Thyroid Hormone Receptor. *Mol Cell Biol* **23**, 5122–5131 (2003).
42. Bailey, P. *et al.* The Nuclear Receptor Corepressor N-CoR Regulates Differentiation: N-CoR Directly Interacts with MyoD. *Molecular Endocrinology* **13**, 1155–1168 (1999).
43. Horwitz, K. B. *et al.* Nuclear receptor coactivators and corepressors. *Molecular Endocrinology* **10**, 1167–1177 (1996).
44. Ai, N., Krasowski, M. D., Welsh, W. J. & Ekins, S. Understanding nuclear receptors using computational methods. *Drug Discov Today* **14**, 486–494 (2009).
45. Weiss, R. E. *et al.* Mice deficient in the steroid receptor co-activator 1 (SRC-1) are resistant to thyroid hormone. *EMBO J* **18**, 1900–1904 (1999).
46. Tagami, T., Madison, L. D., Nagaya, T. & Jameson, J. L. Nuclear receptor corepressors activate rather than suppress basal transcription of genes that are negatively regulated by thyroid hormone. *Mol Cell Biol* **17**, 2642–2648 (1997).
47. Berghagen, H. *et al.* Corepressor SMRT functions as a coactivator for thyroid hormone receptor T3R α from a negative hormone response element. *J Biol Chem* **277**, 49517–49522 (2002).
48. Broekema, M. F. *et al.* Profiling of 3696 Nuclear Receptor–Coregulator Interactions: A Resource for Biological and Clinical Discovery. *Endocrinology* **159**, 2397–2407 (2018).
49. Pascussi, J.-M. *et al.* The Tangle of Nuclear Receptors that Controls Xenobiotic Metabolism and Transport: Crosstalk and Consequences. *Annual Review of Pharmacology and Toxicology* **48**, 1–32 (2008).

50. HU, R., HILAKIVI-CLARKE, L. & CLARKE, R. Molecular mechanisms of tamoxifen-associated endometrial cancer (Review). *Oncol Lett* **9**, 1495–1501 (2015).
51. Wooltorton, E. Osteoporosis treatment: raloxifene (Evista) and stroke mortality. *CMAJ* **175**, 147 (2006).
52. McCrea, E., Sissung, T. M., Price, D. K., Chau, C. H. & Figg, W. D. ANDROGEN RECEPTOR VARIATION AFFECTS PROSTATE CANCER PROGRESSION AND DRUG RESISTANCE. *Pharmacol Res* **114**, 152–162 (2016).
53. Ünüvar, T. & Büyükgebiz, A. Fetal and Neonatal Endocrine Disruptors. *J Clin Res Pediatr Endocrinol* **4**, 51–60 (2012).
54. Grohs, M. N. *et al.* Prenatal maternal and childhood bisphenol a exposure and brain structure and behavior of young children. *Environmental Health* **18**, 85 (2019).
55. Jeng, H. A. Exposure to Endocrine Disrupting Chemicals and Male Reproductive Health. *Front Public Health* **2**, 55 (2014).
56. Hauser, R., Williams, P., Altshul, L. & Calafat, A. M. Evidence of Interaction between Polychlorinated Biphenyls and Phthalates in Relation to Human Sperm Motility. *Environ Health Perspect* **113**, 425–430 (2005).
57. Mangelsdorf, D. J. *et al.* Characterization of three RXR genes that mediate the action of 9-cis retinoic acid. *Genes Dev.* **6**, 329–344 (1992).
58. Hamada, K. *et al.* H-2RIIBP, a member of the nuclear hormone receptor superfamily that binds to both the regulatory element of major histocompatibility class I genes and the estrogen response element. *Proc Natl Acad Sci U S A* **86**, 8289–8293 (1989).
59. Dollé, P., Fraulob, V., Kastner, P. & Chambon, P. Developmental expression of murine retinoid X receptor (RXR) genes. *Mechanisms of Development* **45**, 91–104 (1994).
60. Mangelsdorf, D. J. & Evans, R. M. The RXR heterodimers and orphan receptors. *Cell* **83**, 841–850 (1995).

61. Pérez, E., Bourguet, W., Gronemeyer, H. & de Lera, A. R. Modulation of RXR function through ligand design. *Biochim Biophys Acta* **1821**, 57–69 (2012).
62. Castillo, A. I. *et al.* A permissive retinoid X receptor/thyroid hormone receptor heterodimer allows stimulation of prolactin gene transcription by thyroid hormone and 9-cis-retinoic acid. *Mol Cell Biol* **24**, 502–513 (2004).
63. Blumberg, B. *et al.* SXR, a novel steroid and xenobioticsensing nuclear receptor. *Genes Dev* **12**, 3195–3205 (1998).
64. Bertilsson, G. *et al.* Identification of a human nuclear receptor defines a new signaling pathway for CYP3A induction. *Proc Natl Acad Sci U S A* **95**, 12208–12213 (1998).
65. Lehmann, J. M. *et al.* The human orphan nuclear receptor PXR is activated by compounds that regulate CYP3A4 gene expression and cause drug interactions. *J Clin Invest* **102**, 1016–1023 (1998).
66. Delfosse, V. *et al.* Mechanistic insights into the synergistic activation of the RXR-PXR heterodimer by endocrine disruptor mixtures. *Proc Natl Acad Sci U S A* **118**, e2020551118 (2021).
67. Willson, T. M. & Kliewer, S. A. PXR, CAR and drug metabolism. *Nat Rev Drug Discov* **1**, 259–266 (2002).
68. Oladimeji, P. O. & Chen, T. PXR: More Than Just a Master Xenobiotic Receptor. *Mol Pharmacol* **93**, 119–127 (2018).
69. Noble, S. M. *et al.* Human PXR Forms a Tryptophan Zipper-Mediated Homodimer. *Biochemistry* **45**, 8579–8589 (2006).
70. Watkins, R. E., Davis-Searles, P. R., Lambert, M. H. & Redinbo, M. R. Coactivator binding promotes the specific interaction between ligand and the pregnane X receptor. *J Mol Biol* **331**, 815–828 (2003).

71. Johnson, D. R., Li, C.-W., Chen, L.-Y., Ghosh, J. C. & Chen, J. D. Regulation and binding of pregnane X receptor by nuclear receptor corepressor silencing mediator of retinoid and thyroid hormone receptors (SMRT). *Mol Pharmacol* **69**, 99–108 (2006).
72. Watkins, R. E. *et al.* The human nuclear xenobiotic receptor PXR: structural determinants of directed promiscuity. *Science* **292**, 2329–2333 (2001).
73. Ekins, S. & Schuetz, E. The PXR crystal structure: the end of the beginning. *Trends Pharmacol Sci* **23**, 49–50 (2002).
74. Ekins, S. *et al.* Challenges Predicting Ligand-Receptor Interactions of Promiscuous Proteins: The Nuclear Receptor PXR. *PLoS Comput Biol* **5**, (2009).
75. Hall, A., Chanteux, H., Ménochet, K., Ledecq, M. & Schulze, M.-S. E. D. Designing Out PXR Activity on Drug Discovery Projects: A Review of Structure-Based Methods, Empirical and Computational Approaches. *J. Med. Chem.* **64**, 6413–6522 (2021).
76. Geick, A., Eichelbaum, M. & Burk, O. Nuclear receptor response elements mediate induction of intestinal MDR1 by rifampin. *J Biol Chem* **276**, 14581–14587 (2001).
77. Kast, H. R. *et al.* Regulation of multidrug resistance-associated protein 2 (ABCC2) by the nuclear receptors pregnane X receptor, farnesoid X-activated receptor, and constitutive androstane receptor. *J Biol Chem* **277**, 2908–2915 (2002).
78. Maeda, K. *et al.* Identification of the rate-determining process in the hepatic clearance of atorvastatin in a clinical cassette microdosing study. *Clin Pharmacol Ther* **90**, 575–581 (2011).
79. Chen, Y. *et al.* Nuclear receptors in the multidrug resistance through the regulation of drug-metabolizing enzymes and drug transporters. *Biochemical Pharmacology* **83**, 1112–1126 (2012).

80. Sinz, M. *et al.* Evaluation of 170 xenobiotics as transactivators of human pregnane X receptor (hPXR) and correlation to known CYP3A4 drug interactions. *Curr Drug Metab* **7**, 375–388 (2006).
81. Chen, T. Overcoming drug resistance by regulating nuclear receptors. *Adv Drug Deliv Rev* **62**, 1257–1264 (2010).
82. Burk, O. *et al.* Identification of approved drugs as potent inhibitors of pregnane X receptor activation with differential receptor interaction profiles. *Arch Toxicol* **92**, 1435–1451 (2018).
83. Wiesinger, H. *et al.* The Effects of Weak and Strong CYP3A Induction by Rifampicin on the Pharmacokinetics of Five Progestins and Ethinylestradiol Compared to Midazolam. *Clin Pharmacol Ther* **108**, 798–807 (2020).
84. Yamazaki, T. *et al.* Pharmacokinetic Effects of Isavuconazole Coadministration With the Cytochrome P450 Enzyme Substrates Bupropion, Repaglinide, Caffeine, Dextromethorphan, and Methadone in Healthy Subjects. *Clin Pharmacol Drug Dev* **6**, 54–65 (2017).
85. Lin, W. *et al.* SPA70 is a potent antagonist of human pregnane X receptor. *Nat Commun* **8**, 741 (2017).
86. De Smet, F., Christopoulos, A. & Carmeliet, P. Allosteric targeting of receptor tyrosine kinases. *Nat Biotechnol* **32**, 1113–1120 (2014).
87. Shan, Y. *et al.* How Does a Drug Molecule Find Its Target Binding Site? *J. Am. Chem. Soc.* **133**, 9181–9183 (2011).
88. Moore, T. W., Mayne, C. G. & Katzenellenbogen, J. A. Minireview: Not picking pockets: nuclear receptor alternate-site modulators (NRAMs). *Mol Endocrinol* **24**, 683–695 (2010).

89. Kamaraj, R., Drastik, M., Maixnerova, J. & Pavek, P. Allosteric Antagonism of the Pregnane X Receptor (PXR): Current-State-of-the-Art and Prediction of Novel Allosteric Sites. *Cells* **11**, 2974 (2022).
90. Changeux, J.-P. & Christopoulos, A. Allosteric Modulation as a Unifying Mechanism for Receptor Function and Regulation. *Cell* **166**, 1084–1102 (2016).
91. Delfosse, V. *et al.* Mechanistic insights into the synergistic activation of the RXR–PXR heterodimer by endocrine disruptor mixtures. *Proc Natl Acad Sci USA* **118**, e2020551118 (2021).
92. Venkatesh, M. *et al.* In Vivo and In Vitro Characterization of a First-in-Class Novel Azole Analog That Targets Pregnane X Receptor Activation. *Mol Pharmacol* **80**, 124–135 (2011).
93. Krausova, L. *et al.* Metformin suppresses pregnane X receptor (PXR)-regulated transactivation of CYP3A4 gene. *Biochemical Pharmacology* **82**, 1771–1780 (2011).
94. Das, B. C. *et al.* Synthesis of novel ketoconazole derivatives as inhibitors of the human Pregnane X Receptor (PXR; NR1I2; also termed SXR, PAR). *Bioorganic & Medicinal Chemistry Letters* **18**, 3974–3977 (2008).
95. Ekins, S. *et al.* Computational Discovery of Novel Low Micromolar Human Pregnane X Receptor Antagonists. *Mol Pharmacol* **74**, 662–672 (2008).
96. Wang, H. *et al.* The Phytoestrogen Coumestrol Is a Naturally Occurring Antagonist of the Human Pregnane X Receptor. *Molecular Endocrinology* **22**, 838–857 (2008).
97. Zhou, C. *et al.* The Dietary Isothiocyanate Sulforaphane Is an Antagonist of the Human Steroid and Xenobiotic Nuclear Receptor. *Mol Pharmacol* **71**, 220–229 (2007).
98. Chen, Y., Tang, Y., Robbins, G. T. & Nie, D. Camptothecin Attenuates Cytochrome P450 3A4 Induction by Blocking the Activation of Human Pregnane X Receptor. *J Pharmacol Exp Ther* **334**, 999–1008 (2010).

99. Lamba, J., Lamba, V. & Schuetz, E. Genetic variants of PXR (NR1I2) and CAR (NR1I3) and their implications in drug metabolism and pharmacogenetics. *Curr Drug Metab* **6**, 369–383 (2005).
100. Mbatchi, L. C., Brouillet, J.-P. & Evrard, A. Genetic variations of the xenoreceptors NR1I2 and NR1I3 and their effect on drug disposition and response variability. *Pharmacogenomics* **19**, 61–77 (2018).
101. Auerbach, S. S., Stoner, M. A., Su, S. & Omiecinski, C. J. Retinoid X receptor-alpha-dependent transactivation by a naturally occurring structural variant of human constitutive androstane receptor (NR1I3). *Mol Pharmacol* **68**, 1239–1253 (2005).
102. Auerbach, S. S., DeKeyser, J. G., Stoner, M. A. & Omiecinski, C. J. CAR2 Displays Unique Ligand Binding and RXR α Heterodimerization Characteristics. *Drug Metab Dispos* **35**, 428–439 (2007).
103. Yang, H. & Wang, H. Signaling control of the constitutive androstane receptor (CAR). *Protein Cell* **5**, 113–123 (2014).
104. Baes, M. *et al.* A new orphan member of the nuclear hormone receptor superfamily that interacts with a subset of retinoic acid response elements. *Mol Cell Biol* **14**, 1544–1552 (1994).
105. Wei, P., Zhang, J., Dowhan, D. H., Han, Y. & Moore, D. D. Specific and overlapping functions of the nuclear hormone receptors CAR and PXR in xenobiotic response. *Pharmacogenomics J* **2**, 117–126 (2002).
106. Kobayashi, K., Sueyoshi, T., Inoue, K., Moore, R. & Negishi, M. Cytoplasmic accumulation of the nuclear receptor CAR by a tetratricopeptide repeat protein in HepG2 cells. *Mol Pharmacol* **64**, 1069–1075 (2003).

107. Maglich, J. M. *et al.* Identification of a Novel Human Constitutive Androstane Receptor (CAR) Agonist and Its Use in the Identification of CAR Target Genes*. *Journal of Biological Chemistry* **278**, 17277–17283 (2003).
108. Tzamelis, I., Pissios, P., Schuetz, E. G. & Moore, D. D. The xenobiotic compound 1,4-bis[2-(3,5-dichloropyridyloxy)]benzene is an agonist ligand for the nuclear receptor CAR. *Mol Cell Biol* **20**, 2951–2958 (2000).
109. Honkakoski, P., Zelko, I., Sueyoshi, T. & Negishi, M. The Nuclear Orphan Receptor CAR-Retinoid X Receptor Heterodimer Activates the Phenobarbital-Responsive Enhancer Module of the CYP2B Gene. *Mol Cell Biol* **18**, 5652–5658 (1998).
110. Zhang, J., Huang, W., Chua, S. S., Wei, P. & Moore, D. D. Modulation of acetaminophen-induced hepatotoxicity by the xenobiotic receptor CAR. *Science* **298**, 422–424 (2002).
111. Chai, S. C., Cherian, M. T., Wang, Y.-M. & Chen, T. Small-molecule modulators of PXR and CAR. *Biochim Biophys Acta* **1859**, 1141–1154 (2016).
112. Honkakoski, P., Sueyoshi, T. & Negishi, M. Drug-activated nuclear receptors CAR and PXR. *Ann Med* **35**, 172–182 (2003).
113. DeKeyser, J. G., Laurenzana, E. M., Peterson, E. C., Chen, T. & Omiecinski, C. J. Selective Phthalate Activation of Naturally Occurring Human Constitutive Androstane Receptor Splice Variants and the Pregnane X Receptor. *Toxicological Sciences* **120**, 381–391 (2011).
114. Burk, O. *et al.* Differential effects of clinically used derivatives and metabolites of artemisinin in the activation of constitutive androstane receptor isoforms: Antimalarial drugs and CAR. *British Journal of Pharmacology* **167**, 666–681 (2012).
115. Honkakoski, P. Searching for Constitutive Androstane Receptor Modulators. *Drug Metab Dispos* **50**, 1002–1009 (2022).

116. Lynch, C. *et al.* Identification of Modulators That Activate the Constitutive Androstane Receptor From the Tox21 10K Compound Library. *Toxicol Sci* **167**, 282–292 (2019).
117. Kanno, Y., Tanuma, N., Takahashi, A. & Inouye, Y. TO901317, a potent LXR agonist, is an inverse agonist of CAR. *J Toxicol Sci* **38**, 309–315 (2013).
118. Küblbeck, J. *et al.* New in Vitro Tools to Study Human Constitutive Androstane Receptor (CAR) Biology: Discovery and Comparison of Human CAR Inverse Agonists. *Mol. Pharmaceutics* **8**, 2424–2433 (2011).
119. Jyrkkäinen, J. *et al.* Molecular Dynamics Simulations for Human CAR Inverse Agonists. *J. Chem. Inf. Model.* **52**, 457–464 (2012).
120. Keminer, O. *et al.* Identification of novel agonists by high-throughput screening and molecular modelling of human constitutive androstane receptor isoform 3. *Arch Toxicol* **93**, 2247–2264 (2019).
121. Lynch, C. *et al.* Identification of Novel Activators of Constitutive Androstane Receptor from FDA-Approved Drugs by Integrated Computational and Biological Approaches. *Pharm Res* **30**, 489–501 (2013).
122. Dauwe, Y. *et al.* Steatosis and Metabolic Disorders Associated with Synergistic Activation of the CAR/RXR Heterodimer by Pesticides. *Cells* **12**, (2023).
123. Tian, S., Chen, S., Pan, C. & Li, Y. FXR: structures, biology, and drug development for NASH and fibrosis diseases. *Acta Pharmacol Sin* **43**, 1120–1132 (2022).
124. Forman, B. M. *et al.* Identification of a nuclear receptor that is activated by farnesol metabolites. *Cell* **81**, 687–693 (1995).
125. Han, C. Y. Update on FXR Biology: Promising Therapeutic Target? *Int J Mol Sci* **19**, 2069 (2018).
126. Teodoro, J. S., Rolo, A. P. & Palmeira, C. M. Hepatic FXR: key regulator of whole-body energy metabolism. *Trends Endocrinol Metab* **22**, 458–466 (2011).

127. Lee, F. Y., Lee, H., Hubbert, M. L., Edwards, P. A. & Zhang, Y. FXR, a multipurpose nuclear receptor. *Trends in Biochemical Sciences* **31**, 572–580 (2006).
128. Maglich, J. M. *et al.* The first completed genome sequence from a teleost fish (*Fugu rubripes*) adds significant diversity to the nuclear receptor superfamily. *Nucleic Acids Res* **31**, 4051–4058 (2003).
129. Otte, K. *et al.* Identification of Farnesoid X Receptor β as a Novel Mammalian Nuclear Receptor Sensing Lanosterol. *Mol Cell Biol* **23**, 864–872 (2003).
130. Modica, S., Gadaleta, R. M. & Moschetta, A. Deciphering the nuclear bile acid receptor FXR paradigm. *Nucl Recept Signal* **8**, e005 (2010).
131. Bordeaux, J. *et al.* Antibody validation. *Biotechniques* **48**, 197–209 (2010).
132. Bunnage, M. E. Getting pharmaceutical R&D back on target. *Nat Chem Biol* **7**, 335–339 (2011).
133. Astashkina, A., Mann, B. & Grainger, D. W. A critical evaluation of in vitro cell culture models for high-throughput drug screening and toxicity. *Pharmacology & Therapeutics* **134**, 82–106 (2012).
134. Tuntland, T. *et al.* Implementation of pharmacokinetic and pharmacodynamic strategies in early research phases of drug discovery and development at Novartis Institute of Biomedical Research. *Front Pharmacol* **5**, 174 (2014).
135. Hughes, J., Rees, S., Kalindjian, S. & Philpott, K. Principles of early drug discovery. *Br J Pharmacol* **162**, 1239–1249 (2011).
136. Moore, K. & Rees, S. Cell-Based Versus Isolated Target Screening: How Lucky Do You Feel? *SLAS Discovery* **6**, 69–74 (2001).
137. Everitt, J. I. The Future of Preclinical Animal Models in Pharmaceutical Discovery and Development: A Need to Bring In Cerebro to the In Vivo Discussions. *Toxicol Pathol* **43**, 70–77 (2015).

138. Investigational New Drug (IND) Application. *FDA* <https://www.fda.gov/drugs/types-applications/investigational-new-drug-ind-application> (2022).
139. Sun, D., Gao, W., Hu, H. & Zhou, S. Why 90% of clinical drug development fails and how to improve it? *Acta Pharm Sin B* **12**, 3049–3062 (2022).
140. Drug Approvals and Databases. *FDA* <https://www.fda.gov/drugs/development-approval-process-drugs/drug-approvals-and-databases> (2022).
141. Zhao, L., Ciallella, H. L., Aleksunes, L. M. & Zhu, H. Advancing computer-aided drug discovery (CADD) by big data and data-driven machine learning modeling. *Drug Discov Today* **25**, 1624–1638 (2020).
142. Gupta, R. *et al.* Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Mol Divers* **25**, 1315–1360 (2021).
143. Dara, S., Dhamecherla, S., Jadav, S. S., Babu, C. M. & Ahsan, M. J. Machine Learning in Drug Discovery: A Review. *Artif Intell Rev* **55**, 1947–1999 (2022).
144. Acharya, C., Coop, A., Polli, J. E. & MacKerell, A. D. Recent Advances in Ligand-Based Drug Design: Relevance and Utility of the Conformationally Sampled Pharmacophore Approach. *Curr Comput Aided Drug Des* **7**, 10–22 (2011).
145. Aparoy, P., Kumar Reddy, K. & Reddanna, P. Structure and Ligand Based Drug Design Strategies in the Development of Novel 5-LOX Inhibitors. *Curr Med Chem* **19**, 3763–3778 (2012).
146. Kronenberger, T., de Oliveira Fernandes, P., Drumond Franco, I., Poso, A. & Gonçalves Maltarollo, V. Ligand- and Structure-Based Approaches of Escherichia coli FabI Inhibition by Triclosan Derivatives: From Chemical Similarity to Protein Dynamics Influence. *ChemMedChem* **14**, 1995–2004 (2019).

147. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
148. Stanzione, F., Giangreco, I. & Cole, J. C. Use of molecular docking computational tools in drug discovery. in *Progress in Medicinal Chemistry* vol. 60 273–343 (Elsevier, 2021).
149. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235–242 (2000).
150. Young, J. Y. *et al.* OneDep: Unified wwPDB System for Deposition, Biocuration, and Validation of Macromolecular Structures in the PDB Archive. *Structure* **25**, 536–545 (2017).
151. Wlodawer, A., Minor, W., Dauter, Z. & Jaskolski, M. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *The FEBS Journal* **275**, 1–21 (2008).
152. Kleywegt, G. J. & Alwyn Jones, T. Model building and refinement practice. in *Methods in Enzymology* vol. 277 208–230 (1997).
153. Cooper, D. R., Porebski, P. J., Chruszcz, M. & Minor, W. X-ray crystallography: Assessment and validation of protein-small molecule complexes for drug discovery. *Expert Opin Drug Discov* **6**, 771–782 (2011).
154. Halgren, T. New Method for Fast and Accurate Binding-site Identification and Analysis. *Chemical Biology & Drug Design* **69**, 146–148 (2007).
155. Halgren, T. A. Identifying and Characterizing Binding Sites and Assessing Druggability. *J. Chem. Inf. Model.* **49**, 377–389 (2009).
156. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410 (1990).
157. Lesk, A. *Introduction to Bioinformatics*. (OUP Oxford, 2014).

158. Roessler, C. G. *et al.* Transitive homology-guided structural studies lead to discovery of Cro proteins with 40% sequence identity but different folds. *Proceedings of the National Academy of Sciences* **105**, 2343–2348 (2008).
159. Bagaria, A., Jaravine, V., Huang, Y. J., Montelione, G. T. & Güntert, P. Protein structure validation by generalized linear model root-mean-square deviation prediction: Protein Structure Validation by GLM-RMSD. *Protein Science* **21**, 229–238 (2012).
160. *Molecular Modeling of Proteins*. vol. 1215 (Springer New York, 2015).
161. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology* **7**, 95–99 (1963).
162. Xu, R. X. *et al.* A Structural Basis for Constitutive Activity in the Human CAR/RXR α Heterodimer. *Molecular Cell* **16**, 919–928 (2004).
163. Fiser, A. & Šali, A. Modeller: Generation and Refinement of Homology-Based Protein Structure Models. in *Methods in Enzymology* vol. 374 461–491 (Academic Press, 2003).
164. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* **26**, 283–291 (1993).
165. Wiederstein, M. & Sippl, M. J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* **35**, W407–W410 (2007).
166. Hooft, R. W. W., Vriend, G., Sander, C. & Abola, E. E. Errors in protein structures. *Nature* **381**, 272–272 (1996).
167. Lovell, S. C. *et al.* Structure validation by C α geometry: ϕ , ψ and C β deviation. *Proteins: Structure, Function, and Bioinformatics* **50**, 437–450 (2003).
168. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

169. Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R. & Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des* **27**, 221–234 (2013).
170. Vanommeslaeghe, K., Guvench, O. & MacKerell, A. D. Molecular Mechanics. *Curr Pharm Des* **20**, 3281–3292 (2014).
171. Tuzun, R. E., Noid, D. W. & Sumpter, B. G. Efficient treatment of out-of-plane bend and improper torsion interactions in MM2, MM3, and MM4 molecular mechanics calculations. *Journal of Computational Chemistry* **18**, 1804–1811 (1997).
172. Brooks, B. R. *et al.* CHARMM: The Biomolecular Simulation Program. *J Comput Chem* **30**, 1545–1614 (2009).
173. CASE, D. A. *et al.* The Amber Biomolecular Simulation Programs. *J Comput Chem* **26**, 1668–1688 (2005).
174. Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **118**, 11225–11236 (1996).
175. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
176. Jing, Z. *et al.* Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications. *Annual Review of Biophysics* **48**, 371–394 (2019).
177. Kamenik, A. S. *et al.* Polarizable and non-polarizable force fields: Protein folding, unfolding, and misfolding. *J. Chem. Phys.* **153**, 185102 (2020).
178. Schmid, N. *et al.* Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur Biophys J* **40**, 843–856 (2011).
179. Huang, J. *et al.* CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat Methods* **14**, 71–73 (2017).

180. Lewars, E. G. *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*. (Springer Science & Business Media, 2010).
181. Henkelman, G., Uberuaga, B. P. & Jónsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *The Journal of Chemical Physics* **113**, 9901–9904 (2000).
182. Chatzieftheriou, S., Adendorff, M. R. & Lagaros, N. D. Generalized Potential Energy Finite Elements for Modeling Molecular Nanostructures. *J. Chem. Inf. Model.* **56**, 1963–1978 (2016).
183. Peng, C., Ayala, P. Y., Schlegel, H. B. & Frisch, M. J. Using redundant internal coordinates to optimize equilibrium geometries and transition states. *Journal of Computational Chemistry* **17**, 49–56 (1996).
184. Henkelman, G. & Jónsson, H. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *The Journal of Chemical Physics* **111**, 7010–7022 (1999).
185. Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R. & Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology* **161**, 269–288 (1982).
186. Rarey, M., Kramer, B., Lengauer, T. & Klebe, G. A Fast Flexible Docking Method using an Incremental Construction Algorithm. *Journal of Molecular Biology* **261**, 470–489 (1996).
187. Friesner, R. A. *et al.* Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* **47**, 1739–1749 (2004).
188. Halgren, T. A. *et al.* Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem* **47**, 1750–1759 (2004).

189. Friesner, R. A. *et al.* Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein–Ligand Complexes. *J. Med. Chem.* **49**, 6177–6196 (2006).
190. Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry* **30**, 2785–2791 (2009).
191. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. & Taylor, R. D. Improved protein–ligand docking using GOLD. *Proteins: Structure, Function, and Bioinformatics* **52**, 609–623 (2003).
192. Khamis, M. A., Gomaa, W. & Ahmed, W. F. Machine learning in computational docking. *Artificial Intelligence in Medicine* **63**, 135–152 (2015).
193. Jorgensen, W. L. & Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **110**, 1657–1666 (1988).
194. Kaminski, G. A., Friesner, R. A., Tirado-Rives, J. & Jorgensen, W. L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **105**, 6474–6487 (2001).
195. Robertson, M. J., Tirado-Rives, J. & Jorgensen, W. L. Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *J Chem Theory Comput* **11**, 3499–3509 (2015).
196. Harder, E. *et al.* OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **12**, 281–296 (2016).
197. Roos, K. *et al.* OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *J. Chem. Theory Comput.* **15**, 1863–1874 (2019).

198. Lu, C. *et al.* OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. *J. Chem. Theory Comput.* **17**, 4291–4300 (2021).
199. Sherman, W., Beard, H. S. & Farid, R. Use of an Induced Fit Receptor Structure in Virtual Screening. *Chemical Biology & Drug Design* **67**, 83–84 (2006).
200. Farid, R., Day, T., Friesner, R. A. & Pearlstein, R. A. New insights about HERG blockade obtained from protein modeling, potential energy mapping, and docking studies. *Bioorganic & Medicinal Chemistry* **14**, 3160–3173 (2006).
201. Fischer, A., Smieško, M., Sellner, M. & Lill, M. A. Decision Making in Structure-Based Drug Discovery: Visual Inspection of Docking Results. *J. Med. Chem.* **64**, 2489–2500 (2021).
202. Kolb, P. *et al.* Structure-based discovery of β_2 -adrenergic receptor ligands. *Proc Natl Acad Sci U S A* **106**, 6843–6848 (2009).
203. Bauer, M. R. & Mackey, M. D. Electrostatic Complementarity as a Fast and Effective Tool to Optimize Binding and Selectivity of Protein–Ligand Complexes. *J. Med. Chem.* **62**, 3036–3050 (2019).
204. Xu, H. E. *et al.* Molecular Recognition of Fatty Acids by Peroxisome Proliferator–Activated Receptors. *Molecular Cell* **3**, 397–403 (1999).
205. Satpathy, R. Application of Molecular Docking Methods on Endocrine Disrupting Chemicals: A Review. *Journal of Applied Biotechnology Reports* **7**, 74–80 (2020).
206. Chen, Y.-C. Beware of docking! *Trends in Pharmacological Sciences* **36**, 78–95 (2015).
207. Durrant, J. D. & McCammon, J. A. Molecular dynamics simulations and drug discovery. *BMC Biology* **9**, 71 (2011).
208. Hollingsworth, S. A. & Dror, R. O. Molecular dynamics simulation for all. *Neuron* **99**, 1129–1143 (2018).

209. Alder, B. J. & Wainwright, T. E. Studies in Molecular Dynamics. I. General Method. *The Journal of Chemical Physics* **31**, 459–466 (1959).
210. Mark, P. & Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **105**, 9954–9960 (2001).
211. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
212. Berendsen, H. J. C., Grigera, J. R. & Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **91**, 6269–6271 (1987).
213. Izadi, S., Anandakrishnan, R. & Onufriev, A. V. Building Water Models: A Different Approach. *J. Phys. Chem. Lett.* **5**, 3863–3871 (2014).
214. Emperador, A., Crehuet, R. & Guàrdia, E. Effect of the Water Model in Simulations of Protein–Protein Recognition and Association. *Polymers* **13**, 176 (2021).
215. Xiong, Y., Izadi, S. & Onufriev, A. *A fast polarizable water model for atomistic simulations*. <https://chemrxiv.org/engage/chemrxiv/article-details/6225a7a6daa4fb438e8268d8> (2022) doi:10.26434/chemrxiv-2022-v80r5.
216. Wang, L.-P., Martinez, T. J. & Pande, V. S. Building Force Fields: An Automatic, Systematic, and Reproducible Approach. *J. Phys. Chem. Lett.* **5**, 1885–1891 (2014).
217. Boonstra, S., Onck, P. R. & van der Giessen, E. CHARMM TIP3P Water Model Suppresses Peptide Folding by Solvating the Unfolded State. *J. Phys. Chem. B* **120**, 3692–3698 (2016).
218. Sajadi, F. & Rowley, C. N. Simulations of lipid bilayers using the CHARMM36 force field with the TIP3P-FB and TIP4P-FB water models. *PeerJ* **6**, e5472 (2018).
219. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **26**, 1781–1802 (2005).

220. Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013).
221. Dunkel, J. & Hilbert, S. Phase transitions in small systems: Microcanonical vs. canonical ensembles. *Physica A: Statistical Mechanics and its Applications* **370**, 390–406 (2006).
222. Banci, L. & Comba, P. *Molecular Modeling and Dynamics of Bioinorganic Systems*. (Springer Science & Business Media, 2012).
223. Werner, T., Morris, M. B., Dastmalchi, S. & Church, W. B. Structural modelling and dynamics of proteins for insights into drug interactions. *Advanced Drug Delivery Reviews* **64**, 323–343 (2012).
224. Knapp, B., Ospina, L. & Deane, C. M. Avoiding False Positive Conclusions in Molecular Simulation: The Importance of Replicas. *J. Chem. Theory Comput.* **14**, 6127–6138 (2018).
225. Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters* **314**, 141–151 (1999).
226. Qi, R., Wei, G., Ma, B. & Nussinov, R. Replica Exchange Molecular Dynamics: A Practical Application Protocol with Solutions to Common Problems and a Peptide Aggregation and Self-Assembly Example. *Methods Mol Biol* **1777**, 101–119 (2018).
227. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc Natl Acad Sci U S A* **99**, 12562–12566 (2002).
228. Theobald, D. L. Rapid calculation of RMSDs using a quaternion-based characteristic polynomial. *Acta Cryst A* **61**, 478–480 (2005).
229. Martínez, L. Automatic Identification of Mobile and Rigid Substructures in Molecular Dynamics Simulations and Fractional Structural Fluctuation Analysis. *PLoS One* **10**, e0119264 (2015).

230. Jolliffe, I. T. & Cadima, J. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **374**, 20150202 (2016).
231. Husic, B. E. & Pande, V. S. Markov State Models: From an Art to a Science. *J. Am. Chem. Soc.* **140**, 2386–2396 (2018).
232. Pande, V. S., Beauchamp, K. & Bowman, G. R. Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* **52**, 99–105 (2010).
233. Shukla, D., Hernández, C. X., Weber, J. K. & Pande, V. S. Markov State Models Provide Insights into Dynamic Modulation of Protein Function. *Acc. Chem. Res.* **48**, 414–422 (2015).
234. Dickson, A. Mapping the Ligand Binding Landscape. *Biophys J* **115**, 1707–1719 (2018).
235. Gu, S., Silva, D.-A., Meng, L., Yue, A. & Huang, X. Quantitatively Characterizing the Ligand Binding Mechanisms of Choline Binding Protein Using Markov State Model Analysis. *PLoS Computational Biology* **10**, e1003767 (2014).
236. Pavliotis, G. A. STOCHASTIC PROCESSES AND APPLICATIONS. (2015).
237. Scherer, M. K. *et al.* PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **11**, 5525–5542 (2015).
238. *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation.* vol. 797 (Springer Netherlands, 2014).
239. Genheden, S. & Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov* **10**, 449–461 (2015).
240. Moreira, I. S., Fernandes, P. A. & Ramos, M. J. Unravelling Hot Spots: a comprehensive computational mutagenesis study. *Theor Chem Acc* **117**, 99–113 (2007).

241. Homeyer, N. & Gohlke, H. Free Energy Calculations by the Molecular Mechanics Poisson–Boltzmann Surface Area Method. *Molecular Informatics* **31**, 114–122 (2012).
242. Sun, H. *et al.* Assessing the performance of MM/PBSA and MM/GBSA methods. 5. Improved docking performance using high solute dielectric constant MM/GBSA and MM/PBSA rescoring. *Phys Chem Chem Phys* **16**, 22035–22045 (2014).
243. Hou, T., Wang, J., Li, Y. & Wang, W. Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. The accuracy of ranking poses generated from docking. *J Comput Chem* **32**, 866–877 (2011).
244. Hou, T., Wang, J., Li, Y. & Wang, W. Assessing the performance of the MM/PBSA and MM/GBSA methods: II. The accuracy of ranking poses generated from docking. *J Comput Chem* **32**, 866–877 (2011).
245. Kollman, P. A. *et al.* Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **33**, 889–897 (2000).
246. Foloppe, N. & Hubbard, R. Towards predictive ligand design with free-energy based computational methods? *Curr Med Chem* **13**, 3583–3608 (2006).
247. Wang, J., Hou, T. & Xu, X. Recent Advances in Free Energy Calculations with a Combination of Molecular Mechanics and Continuum Models. *Current Computer - Aided Drug Design* **2**, 287–306 (2006).
248. Homeyer, N. & Gohlke, H. Free Energy Calculations by the Molecular Mechanics Poisson–Boltzmann Surface Area Method. *Molecular Informatics* **31**, 114–122 (2012).
249. Genheden, S. & Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov* **10**, 449–461 (2015).

250. Forouzesh, N. & Mishra, N. An Effective MM/GBSA Protocol for Absolute Binding Free Energy Calculations: A Case Study on SARS-CoV-2 Spike Protein and the Human ACE2 Receptor. *Molecules* **26**, 2383 (2021).
251. Homeyer, N., Stoll, F., Hillisch, A. & Gohlke, H. Binding Free Energy Calculations for Lead Optimization: Assessment of Their Accuracy in an Industrial Drug Design Context. *J. Chem. Theory Comput.* **10**, 3331–3344 (2014).
252. Cheng, W. & Ng, C. A. Predicting Relative Protein Affinity of Novel Per- and Polyfluoroalkyl Substances (PFASs) by An Efficient Molecular Dynamics Approach. *Environ Sci Technol* **52**, 7972–7980 (2018).
253. McEwan, I. J. Nuclear Receptors: One Big Family. in *The Nuclear Receptor Superfamily: Methods and Protocols* (ed. McEwan, I. J.) 3–18 (Humana Press, 2009). doi:10.1007/978-1-60327-575-0_1.
254. Meyvis, T. K., De Smedt, S. C., Van Oostveldt, P. & Demeester, J. Fluorescence recovery after photobleaching: a versatile tool for mobility and interaction measurements in pharmaceutical research. *Pharm Res* **16**, 1153–1162 (1999).
255. Ishikawa-Ankerhold, H. C., Ankerhold, R. & Drummen, G. P. C. Advanced Fluorescence Microscopy Techniques—FRAP, FLIP, FLAP, FRET and FLIM. *Molecules* **17**, 4047–4132 (2012).
256. Aparicio, O., Geisberg, J. V. & Struhl, K. Chromatin Immunoprecipitation for Determining the Association of Proteins with Specific Genomic Sequences In Vivo. *Current Protocols in Cell Biology* **23**, 17.7.1-17.7.23 (2004).
257. York, B. & O'Malley, B. W. Steroid Receptor Coactivator (SRC) Family: Masters of Systems Biology. *J Biol Chem* **285**, 38743–38750 (2010).
258. Ritter, M. J. *et al.* Nuclear Receptor CoRepressors, NCOR1 and SMRT, are required for maintaining systemic metabolic homeostasis. *Mol Metab* **53**, 101315 (2021).

259. Zhou, M., Li, Q. & Wang, R. Current Experimental Methods for Characterizing Protein–Protein Interactions. *ChemMedChem* **11**, 738–756 (2016).
260. Negishi, M., Kobayashi, K., Sakuma, T. & Sueyoshi, T. Nuclear receptor phosphorylation in xenobiotic signal transduction. *J Biol Chem* **295**, 15210–15225 (2020).
261. Shizu, R. *et al.* Phosphorylated Nuclear Receptor CAR Forms a Homodimer To Repress Its Constitutive Activity for Ligand Activation. *Molecular and Cellular Biology* **37**, e00649-16 (2017).
262. Yeh, S. *et al.* Generation and characterization of androgen receptor knockout (ARKO) mice: An in vivo model for the study of androgen functions in selective tissues. *Proceedings of the National Academy of Sciences* **99**, 13498–13503 (2002).
263. Kronenberger, T. *et al.* Nuclear Receptor Modulators — Current Approaches and Future Perspectives. in *Drug Discovery and Development - From Molecules to Medicine* (IntechOpen, 2015). doi:10.5772/59666.
264. Berney, C. & Danuser, G. FRET or No FRET: A Quantitative Comparison. *Biophys J* **84**, 3992–4010 (2003).
265. Piedade, R. *et al.* Carboxymefloquine, the Major Metabolite of the Antimalarial Drug Mefloquine, Induces Drug-Metabolizing Enzyme and Transporter Expression by Activation of Pregnane X Receptor. *Antimicrobial Agents and Chemotherapy* **59**, 96–104 (2015).
266. Tyree, C. M. & Klausning, K. The mammalian two-hybrid assay for detection of coactivator-nuclear receptor interactions. *Methods Mol Med* **85**, 175–183 (2003).
267. Motta, S., Callea, L., Giani Tagliabue, S. & Bonati, L. Exploring the PXR ligand binding mechanism with advanced Molecular Dynamics methods. *Sci Rep* **8**, 16207 (2018).

268. Wang, W. *et al.* Construction and characterization of a fully active PXR/SRC-1 tethered protein with increased stability. *Protein Eng Des Sel* **21**, 425–433 (2008).
269. Teotico, D. G., Bischof, J. J., Peng, L., Kliewer, S. A. & Redinbo, M. R. Structural basis of human pregnane X receptor activation by the hops constituent colupulone. *Mol Pharmacol* **74**, 1512–1520 (2008).
270. Sadybekov, A. V. & Katritch, V. Computational approaches streamlining drug discovery. *Nature* **616**, 673–685 (2023).
271. Kenda, M. & Sollner Dolenc, M. Computational Study of Drugs Targeting Nuclear Receptors. *Molecules* **25**, 1616 (2020).

6. List of figures

Fig. 1. Overview of the nuclear receptor superfamily	2
Fig. 2. Illustration of nuclear receptor structure	4
Fig. 3. Crystal structure of PXR homodimer	10
Fig. 4. An example of typical workflow of structure-based drug discovery	17
Fig. 5. Ramachandran plot of CAR3-LBD model	21
Fig. 6. Histidine 407 (H407) protonation state	23
Fig. 7. Comparison of time scale for protein motion	34
Fig. 8. Hypothetical scheme of Markov State Models	37

VIII. Result and Discussion

7. Publication I

Discrepancy in interactions and conformational dynamics of pregnane X receptor (PXR) bound to an agonist and a novel competitive antagonist.

Rashidian A, Mustonen EK, Kronenberger T, Schwab M, Burk O, Laufer SA, Pantsar T.

^aDepartment of Internal Medicine VIII, University Hospital Tuebingen, Otfried-Müller-Strasse 14, Tuebingen DE 72076, Germany

^bDr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart and University of Tuebingen, Tuebingen, Germany

^cDepartment of Pharmaceutical and Medicinal Chemistry, Institute of Pharmaceutical Sciences, Eberhard-Karls-Universität, Tuebingen, Auf der Morgenstelle 8, Tuebingen 72076, Germany

^dDepartments of Clinical Pharmacology, and Pharmacy and Biochemistry, University of Tuebingen, Tuebingen, Germany

^eCluster of Excellence iFIT (EXC 2180) “Image-Guided and Functionally Instructed Tumor Therapies”, University of Tuebingen, Tuebingen 72076, Germany

^fTuebingen Center for Academic Drug Discovery & Development (TüCAD2), Tuebingen 72076, Germany

^gSchool of Pharmacy, Faculty of Health Sciences, University of Eastern Finland, Yliopistonranta 1, Kuopio 70210, Finland

Comput Struct Biotechnol Journal.

Volume 13, Page 3004-3018, doi:10.1016/j.csbj .2022.06.020.

Hypothesis: Compound 100 (C-100), a kinase inhibitor, demonstrates antagonistic character when bound to PXR. Our hypothesis is that this compound induces the antagonist conformation upon binding to the PXR. This could be tested by obtaining multiple conformations of the structure to be studied.

Aims: Interpreting antagonistic character of compound 100, a novel molecule, acting on the Pregnane-X receptor (PXR) having SR12813 as full agonist. We aim at using *in silico* methods (molecular docking and MD simulations) with an appropriate simulation time for generating tremendous conformations and employing the two known methods, Principal component analysis (PCA) and MSMs, to detect conformational changes.

Results: Utilizing Principal Component Analysis (PCA), we have successfully illustrated a distinct ligand-dependent conformational behavior of PXR-LBD. In the case of compound 100 (C-100), notably lower Root Mean Square Fluctuations (RMSF) values have been observed in the $\alpha 1$ - $\alpha 2'$ loop, $\beta 4$ - $\alpha 6$ loop, in comparison to the agonist systems. This indicates a more constrained and stable conformation in these specific regions when C-100 is bound. In contrast, the α AF-2 region exhibits higher RMSF values in the presence of C-100, suggesting increased fluctuations compared to the agonist systems. In contrast, when SR12813 and coactivator are present, the α AF-2 region displays the least fluctuation among all the investigated systems, implying a more rigid and stable conformation upon SR12813 binding. Moreover, it has been observed that the closed conformation of the $\beta 4$ - $\alpha 6$ loop plays a crucial role in stabilizing the secondary structure of the $\alpha 6$ helix. Compound 100 prominently promotes the helical configuration, but in the presence of agonists, the secondary structure becomes less stable. It is evident that the configuration of the $\alpha 6$ region is influenced by the ligand it interacts with, leading to distinct intramolecular interactions within the PXR-LBD. In addition, Investigation of the interactions within the $\alpha 6$ region reveals that compound 100 and SR12813 exhibit distinct behaviours in the presence of H407. Unlike SR12813, where H407 takes part in water-mediated interactions between the $\alpha 6$ - $\alpha 7$ loop and $\alpha 10/11$, no such involvement was observed with compound 100. The stabilization of compound 100 and SR12813 in the PXR-LBP is influenced by different polar interactions. While both ligands show limited polar interactions overall, a notable contrast lies in their H407 interactions. Specifically, H407 forms a stable hydrogen bond with SR12813, whereas with compound 100, only a water-mediated interaction is present. Moreover, in the case of compound 100, H407

tends to adopt a conformation that involves a water bridge with N404, a residue located on α 10/11. Compound 100 and SR12813 exhibit similar hydrophobic interactions, excluding the α AF-2 region. In both agonist systems, F420 establishes interactions with I414, while its interactions are considerably reduced with compound 100. Interestingly the RMSF of F420 is greater in the presence of C-100 compared to SRL, highlighting the pronounced flexibility of this specific amino acid residue in the context of compound 100.

Conclusions: Utilizing MD simulations in conjunction with MSM, PCA, and interaction analysis, we conducted an in-depth examination of the conformational dynamics of PXR-LBD while interacting with our novel compound 100. Our investigation reveals differences between the conformational profiles of PXR-LBD bound to compound 100 and the PXR-LBD agonist complex. Notably, the α AF-2 region and H6 displayed prominent alterations, thus providing substantial evidence in support of our initial hypothesis of partial agonism. This study sheds light on the molecular mechanisms underlying the interaction between PXR-LBD and compound 100. Further exploration of these distinct behaviours can potentially unlock novel avenues for drug design and development targeting the PXR-LBD system.



Discrepancy in interactions and conformational dynamics of pregnane X receptor (PXR) bound to an agonist and a novel competitive antagonist



Azam Rashidian^a, Enni-Kaisa Mustonen^b, Thales Kronenberger^{a,c,e,f}, Matthias Schwab^{b,d,e}, Oliver Burk^b, Stefan A. Laufer^{c,e,f}, Tatu Pantsar^{c,g,*}

^a Department of Internal Medicine VIII, University Hospital Tuebingen, Otfried-Müller-Strasse 14, Tuebingen DE 72076, Germany

^b Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart and University of Tuebingen, Tuebingen, Germany

^c Department of Pharmaceutical and Medicinal Chemistry, Institute of Pharmaceutical Sciences, Eberhard-Karls-Universität, Tuebingen, Auf der Morgenstelle 8, Tuebingen 72076, Germany

^d Departments of Clinical Pharmacology, and Pharmacy and Biochemistry, University of Tuebingen, Tuebingen, Germany

^e Cluster of Excellence iFIT (EXC 2180) "Image-Guided and Functionally Instructed Tumor Therapies", University of Tuebingen, Tuebingen 72076, Germany

^f Tuebingen Center for Academic Drug Discovery & Development (TüCAD2), Tuebingen 72076, Germany

^g School of Pharmacy, Faculty of Health Sciences, University of Eastern Finland, Yliopistonranta 1, Kuopio 70210, Finland

ARTICLE INFO

Article history:

Received 18 March 2022

Received in revised form 9 June 2022

Accepted 9 June 2022

Available online 13 June 2022

Keywords:

Pregnane X receptor

Molecular dynamics simulation

PXR ligand binding domain

SR12813 (PubChem CID: 446313)

BAY-1797 (PubChem CID: 124125214)

ABSTRACT

Pregnane X receptor (PXR) is a nuclear receptor with an essential role in regulating drug metabolism genes. While the mechanism of action for ligand-mediated PXR agonism is well-examined, its ligand-mediated inhibition or antagonism is poorly understood. Here we employ microsecond timescale all-atom molecular dynamics (MD) simulations to investigate how our newly identified dual kinase and PXR inhibitor, compound 100, acts as a competitive PXR antagonist and not as a full agonist. We study the PXR ligand binding domain conformational changes associated with compound 100 and compare the results to the full agonist SR12813, in presence and absence of the coactivator. Furthermore, we complement our research by experimentally disclosing the effect of eight key-residue mutations on PXR activation. Finally, simulations of P2X4 inhibitor (BAY-1797) in complex with PXR, which shares an identical structural moiety with compound 100, provide further insights to ligand-induced PXR behaviour. Our MD data suggests ligand-specific influence on conformations of different PXR-LBD regions, including $\alpha 6$ region, $\alpha AF-2$, $\alpha 1-\alpha 2'$, $\beta 1'-\alpha 3$ and $\beta 1-\beta 1'$ loop. Our results provide important insights on conformational behaviour of PXR and offers guidance how to alleviate PXR agonism or to promote PXR antagonism.

© 2022 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Pregnane X receptor (PXR), also known as nuclear receptor (NR) subfamily 1 group I member 2, encoded by the gene *NR1I2*, is a ligand-dependent transcriptional factor that is activated by a structurally diverse set of small molecules [1]. PXR binds various xenobiotic compounds, such as endocrine-disrupting chemicals and pharmaceutical drugs, and endogenous ligands, such as hormones. Ligand-bound PXR regulates the transcription of genes encoding phase I and phase II drug metabolizing enzymes [2] as well as uptake and efflux transporters [3,4]. PXR activation has an impor-

tant role in drug-drug interactions (DDIs) [5], adverse drug reactions [6] and drug treatment efficacy [4,7,8]. In this regard, regulatory agencies, including the European Medicines Agency (EMA) [9] and the United States Food and Drug Administration (FDA) [10], have introduced *in vitro* assays for PXR activation and *in vivo* CYP expression levels in their pipelines for evaluation of drug safety. In addition to its role in small molecule metabolism, PXR is involved in regulation of diverse cellular processes including energy homeostasis, cell proliferation and inflammation [11,12]. PXR expression adapts to (patho) physiological [13] and environmental stimuli [14].

PXR structure comprises three domains: DNA binding domain (DBD), hinge region and ligand-binding domain (LBD). PXR interacts with the gene promoter region of target genes in DNA via its N-terminal DBD [15]. The hinge region is reported to be target of

* Corresponding author at: School of Pharmacy, Faculty of Health Sciences, University of Eastern Finland, Yliopistonranta 1, Kuopio 70210, Finland.

E-mail address: tatu.pantsar@uef.fi (T. Pantsar).

post-translational modifications, which affect PXR mediated gene regulation. For instance, acetylation and deacetylation of K109 modulates PXR transcription activity [16]. LBD comprises eleven α -helices, in addition to the α AF-2 helix (also known as activation function-2), and five-stranded β -sheet (Fig. 1A). The eleven α -helices form three aligned groups: α 1/ α 3, α 4/ α 5/ α 8/ α 9 and α 7/ α 10/ α 11. PXR lacks the typical stable α 2 and α 6 helices that are found in other NRs [17]. With other NRs the ordered α 6 helix results in more tightly packed smaller LBD [17,18]. Moreover, while many NRs exhibit a three-stranded β -sheet in their LBD, PXR comes with two additional strands. These extra β -sheets, together with the additional α 2', contribute to a larger and more flexible LBD, when compared to other NR-LBDs [19–21]. This increased flexibility allows accommodation of diverse set of ligands to the ligand binding pocket (LBP) (Fig. 1B), which is a unique characteristic of PXR.

PXR heterodimerizes with retinoid X receptor alpha (RXR α) in cell cytoplasm and this complex (PXR-RXR α) is transported to the nucleus [23]. Binding an activator ligand to PXR-RXR α heterodimer in nucleus allows the exchange of cofactor (release of a corepressor and recruitment of a coactivator) [24]. Subsequently, this activated PXR complex regulates expression of the target gene.

While the knowledge of PXR ligand-induced activation is well established, the mechanism of action for ligand-mediated inhibitory/antagonistic effects on PXR is less understood [25]. Despite of the numerous PXR agonists that have been reported, a very limited number of antagonists exist. Antagonists binding to the LBP include sulforaphane [26], and SPA70 [27] as shown by *in vitro*

ligand-binding competition assays. Other PXR antagonists, such as ketoconazole or FLB-12 [28] can reduce the endogenous PXR activation without directly or exclusively binding to the LBP, whereas coumestrol [29,30] shown to bind to the LBP and AF-2 domain. What renders a PXR ligand an agonist or an antagonist, and their respective structural triggers, remains poorly understood. For instance, Lin et al. discovered that a close analogue of SPA70, SJB7, is a PXR agonist [27]. Structural differences between SPA70 and SJB7 are minimal, it only exists in their terminal aromatic ring substituents (SI Fig. S1). Lin et al. hypothesized that SJB7 interacts through its *p*-methoxy group with a hydrophobic spot on α AF-2 (residues L428 and F429) and stabilizes the α AF-2 to enable interaction with a coactivator, while SPA70 fails in this due to the lack of this group. More recently, Li et al. reported a set of SPA70 analogues, revealing diverse biological activities of these ligands, ranging from agonists to antagonists and partial agonists [31]. This exemplifies how subtle structural changes may completely shift PXR ligand function and it highlights the promiscuity of PXR-LBD.

No co-crystal structures of PXR and antagonist are currently available to elucidate the details of the PXR-antagonist interactions. Also, for instance, docking approaches are limited as they are unable to capture PXR's characteristic flexibility [21,32–36] and the effect of water [37]. Therefore, molecular dynamics (MD) simulations have been utilized for better understanding of transition state of active to inactive in nuclear receptors [38]. Previous PXR-related MD simulations have mainly focused on PXR agonists. Chandran et al. studied the dynamic behavior of PXR-LBD apo

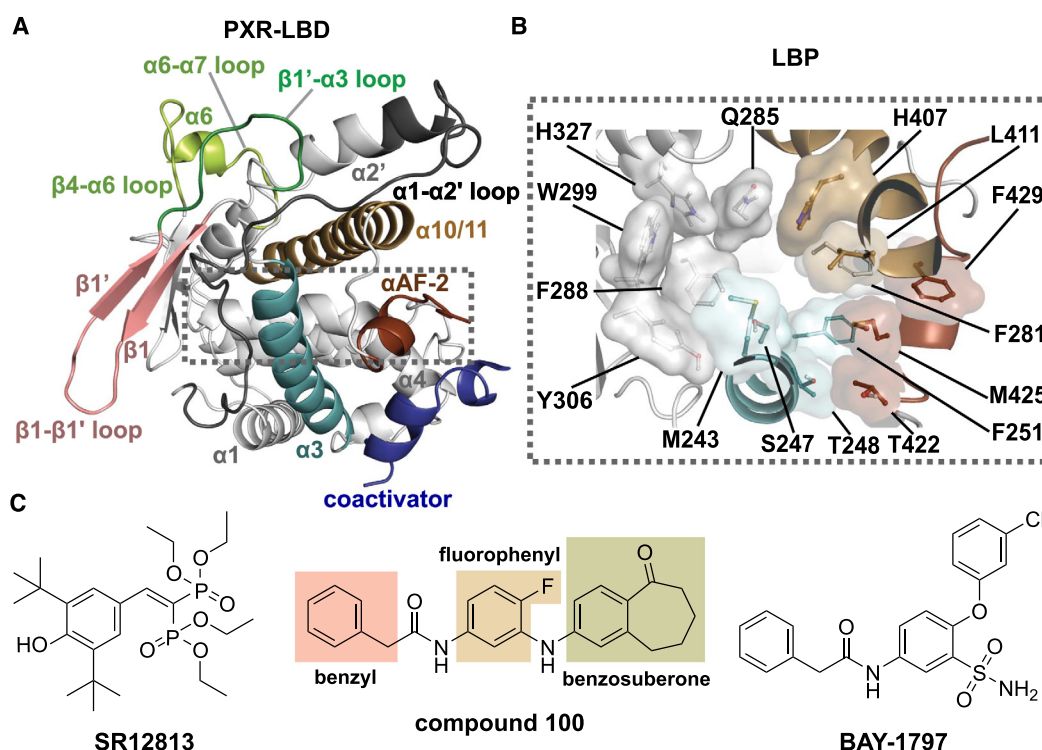


Fig. 1. Overview of the PXR-ligand binding domain (LBD) structure and the small molecule ligands included in this study. (A) Crystal structure of PXR-LBD (PDB ID: 1NRL [22]). The regions of interest are highlighted with the following colour scheme that is used throughout this article: α 1- α 2' loop (residues 177–198), dark grey; β - β 1' (residues 211–225), pink; β 1'- α 3 loop (residues 226–234), dark green; α 3-helix (residues 240–260), cyan; β 4- α 6 loop (residues 309–314), α 6 (residues 315–318) and α 6- α 7 loop (residues 319–323), light green; α 10/ α 11 (residues 389–417) light brown; α AF-2 (residues 423–434), dark brown; coactivator peptide (SRC-1), blue. The dashed rectangular area denotes the location of the ligand binding pocket. (B) Ligand-binding pocket (LBP) of PXR. The main residues forming the LBP and participating in ligand binding are depicted in stick model with transparent molecular surface. Residues are coloured according to their respective regions (see A). (C) 2D structures of PXR agonist SR12813 (SRL), our in-house kinase inhibitor (compound 100), which was found to act also as a competitive PXR antagonist [42], and P2X4 antagonist BAY-1797 that is a PXR agonist [43]. Structure of compound 100 includes a benzosuberone moiety, a fluorophenyl ring and a benzyl group (structural moieties that are discussed in the results). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

structure in comparison to the agonist-bound state, using short MD simulations of 100 ns [39]. They could identify several conformational states for apo PXR-LBD with different volume through MD simulation. Their study revealed that the SR12813 agonist binding events restrict the LBD in a conformation relevant to the size and shape of the ligand. Further, Motta et al. employed the MD-binding method to simulate the SR12813 entry into the LBP. Their result suggested that the ligand would enter to the LBP via a channel between $\alpha 2$ and $\alpha 6$ helices [40]. They also implemented scaled MD (SMD) simulations to extend the sampling of the bound conformations of SR12813 within PXR-LBD (with a total of 2 μ s). They predicted that the binding mode of SR12813 observed in the crystal structure PDB ID: 1NRL [22] is the most stable one among other available PXR-SR12813 complex structures. In addition, Huber et al. performed 200 ns MD simulations of wild type (WT) PXR-LBD and W299A mutant, without ligand and with T090131713 (agonist), SPA70, and SJB7 [41]. They suggested that the extra space conferred by the W299A is the reason for the observed antagonist-to-agonist switch with this mutant for SPA70. This extra space let SPA70 to reside deeper in the pocket, preventing the α AF-2 dislocation and maintaining PXR active.

We recently discovered a novel competitive PXR antagonist from the Tübingen kinase inhibitor collection (TüKIC) compound library [42]. This competitive antagonist, compound 100 (Fig. 1C), suppresses both rifampicin- and SR12813-induced PXR activation, and it does not induce recruitment of SRC-1 to PXR in coactivator recruitment mammalian 2-hybrid assay [42]. Here we aimed to disclose why this kinase inhibitor acts also as an antagonist when in complex with PXR and how it differs from a typical agonist (SR12813). To this end, we applied microsecond timescale MD simulations to understand how these ligands influence PXR's conformational dynamics. Furthermore, we experimentally tested a selection of PXR-LBD mutants and their influence on PXR activity. Finally, we simulated PXR-LBD in complex with BAY-1797 [43], a weak PXR agonist, that shares structural similarity to our competitive antagonist. Our results highlight ligand-specific dynamical behaviour of PXR-LBD and suggest key-changes in ligand-induced antagonism.

2. Results

2.1. Microsecond timescale molecular dynamics simulations reveal discrepancy in motions of compound 100 and SR12813 bound PXR-LBD

To investigate and to compare the interactions and conformational dynamics of the novel competitive antagonist and the classical agonist, we conducted a total of 60 μ s unbiased all-atom MD simulations. These simulations comprised three systems: PXR-LBD in complex with compound 100 (C-100; total simulation time of 30 μ s), SR12813 (SRL; 20 μ s) and SR12813 together with SRC-1 coactivator peptide (SRL + Co; 10 μ s) (SI Fig. S13).

First, to gain a better understanding of the PXR-LBD dynamics, we conducted principal component analysis (PCA) to identify the most essential motions of the protein. The first and the second principal components (PCs) describe together 56% of the data (PC1: 33% and PC2: 23%), while other PCs exhibit individual contributions below 10% (SI Table S1); thus, we focused our analysis on PC1 and PC2 (Fig. 2). The most extensive motions of PC1 occur in $\alpha 1$ - $\alpha 2'$ loop (Fig. 2B; SI Movie M1). In addition, minor movement is observed in the $\beta 4$ - $\alpha 6$ loop (N-terminus of $\alpha 6$ region) and $\beta 1'$ - $\alpha 3$ loop. Of note, these regions, excluding the $\beta 4$ - $\alpha 6$ loop, are part of the novel insert of PXR-LBD that is not found in other NRs. This feature enables PXR-LBD to bind to a wide range of ligands [19]. PC2 displays extensive movement in $\beta 4$ - $\alpha 6$ loop, α AF-2 region,

and the $\alpha 1$ - $\alpha 2'$ loop (Fig. 2B; SI Movie M2). In addition, minor movement is observed in $\beta 1$ - $\beta 1'$ loop and $\beta 1'$ - $\alpha 3$ loop. Importantly, these regions identified by PCA agree with the overall dynamical behaviour of the protein, as demonstrated by backbone root-mean-square fluctuations (RMSFs) (Fig. 2C). Overall, the trend in these fluctuations agrees with the B-factors of the PXR crystal structures (SI Fig. S3). The highest fluctuations appear in $\alpha 1$ - $\alpha 2'$ loop, $\beta 1$ - $\beta 1'$ loop, $\beta 1'$ - $\alpha 3$, $\beta 4$ - $\alpha 6$ loop and α AF-2 region. Here, system specific differences are evident. In PCA highlighted regions, C-100 exhibits lower RMSF values in $\alpha 1$ - $\alpha 2'$ loop, $\beta 4$ - $\alpha 6$ loop and $\beta 1'$ - $\alpha 3$ loop, and higher values in α AF-2 region, compared to the other systems (Fig. 2C; SI Table S2). The least fluctuation in the α AF-2 region is observed in SRL + Co.

Next, we shifted our focus on the differences in the PC scores among the systems. PC1 scores of the compound 100 and SR12813 bound systems are distributed in a wide range (Fig. 2A). However, higher values of PC1 are observed with C-100 (Fig. 2A, SI Fig. S2). Even greater difference is seen with PC2, where C-100 displays clearly higher values. Furthermore, a joint comparison of the PC scores between C-100 and combined agonist systems exemplifies the observed differences (Fig. 2D). Generally, higher PC1 scores are observed for the C-100 compared to the agonist systems (medians of 1.56 Å and -1.77 Å for C-100 and agonists, respectively). Differences in PC2 are more evident (medians of 1.75 Å and -2.77 Å for C-100 and agonists, respectively). Interestingly, PC2 represents the dislocation of α AF-2 from $\alpha 3$ -helix (Fig. 2B; SI Movie M2), a movement which is associated with PXR antagonism [41], and $\beta 4$ - $\alpha 6$ loop association to $\alpha 2'$ (Fig. 2B; SI Movie M2). Overall, PCA exemplifies a clear ligand-dependent conformational behaviour of PXR-LBD.

2.2. Conformational behaviour of the $\alpha 6$ region is ligand-dependent

We next pursued for a more detailed analysis of these PCA-highlighted dynamic regions of PXR-LBD. First, we focused on the $\alpha 6$ region (residues 309–323), comprising $\beta 4$ - $\alpha 6$ loop (N-terminus of the region), $\alpha 6$ helix and $\alpha 6$ - $\alpha 7$ loop (C-terminus of the region) (Fig. 3A). Movement of $\beta 4$ - $\alpha 6$ loop was associated to both PC1 and PC2, appearing even more extensive with PC2 (Fig. 2B, SI Movie M2). To inspect the conformation of this loop, we calculated the distance between A312 (located on $\beta 4$ - $\alpha 6$ loop) and C207 (located on C-terminus of $\alpha 2'$) (Fig. 3A). Clearly smaller distances are observed in the presence of compound 100 (median of 8.2 Å), whereas both agonist-bound systems display significantly longer distances between these two residues (medians of 17.8 Å and 18.6 Å for SRL + Co and SRL, respectively) (Fig. 3A). This indicates that the $\beta 4$ - $\alpha 6$ loop favours an open configuration with SR12813, where this loop resides far from $\alpha 2'$ (Fig. 3B). Conversely, a closed conformation is preferred with compound 100, where the $\beta 4$ - $\alpha 6$ loop is close to $\alpha 2'$. This closed conformation appears to be stabilized via a H-bond between A312 and C207, which is not observed with the agonist (Fig. 3C). Furthermore, we noted that the closed conformation of the $\beta 4$ - $\alpha 6$ loop appears to stabilize the secondary structure of $\alpha 6$ helix (Fig. 3D). While the helical configuration is dominated with compound 100, in agonist-bound systems this secondary structure is more unstable.

Motta et al. predicted that the water channel between $\alpha 2$ and $\alpha 6$ would be an entry pathway for SR12813 [40]. Therefore, we next investigated water-mediated interactions in these regions. Again, notable differences among the systems with these interactions appeared (Fig. 3B; SI Fig. S4). For instance, in C-terminus of $\alpha 6$ region, a water-bridged interaction between E321 and H407 (located on $\alpha 10/11$) is frequent in both agonist systems (60–74%), while it is relatively absent in C-100 (below 6%). Moreover, H407 forms a water-bridged interaction with M323 with agonist (56% and 71%, SRL and SRL + Co, respectively), which is again

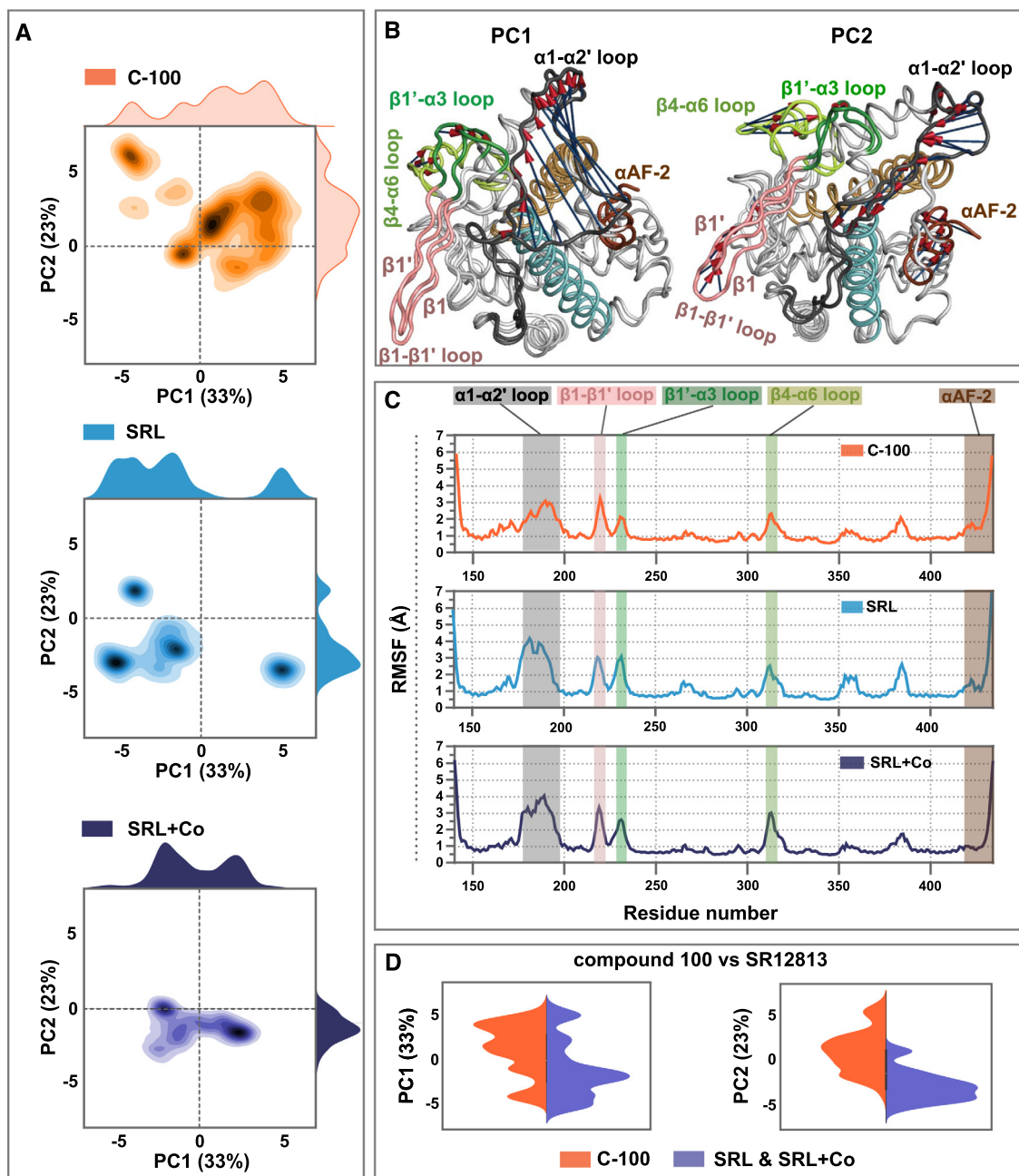


Fig. 2. Conformational dynamics of PXR-LBD is different in compound 100 and SR12813 bound systems. (A) Principal component analysis (PCA) score plot of PC1 and PC2. Each system of the joint PCA is depicted separately: PXR in complex with compound 100 (C-100), orange; PXR in complex with SR12813 (herein called SRL), light blue; PXR in complex with SR12813 and SRC-1 coactivator peptide (herein called SRL + Co), dark blue. (B) The extreme movements of PC1 (left) and PC2 (right). Protein is colour coded as in Fig. 1A. Extreme movements related to each principal component are highlighted with purple arrows with red cones that indicate the direction of the movement. (C) Root-mean-square fluctuation (RMSF) of the protein backbone. The regions associated with the PCA extreme movements are highlighted with their respective colours as in B. (D) A joint comparison of the distributions of PC1 and PC2 scores between C-100 and combined agonist systems (SRL and SRL + Co) shown in violin plot. A kernel density estimation is applied to display the distribution of the data. C-100 is shown in light orange and combined agonist systems shown in purple colour. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

almost inexistent with the compound 100 (Fig. 3B). N-terminus of $\alpha 6$ region is connected to the C-terminus of $\alpha 2'$ helix (residue K210) with system specific water-bridged interactions (SRL and SRL + Co, E309; C-100, E309 and D310) (SI Fig. S4). Regardless of the differences observed for water bridged and H-bond interactions in $\alpha 6$ region, a comparable interaction profile among systems appears between D205 (located on $\alpha 2'$) and R410 or R413 (located on $\alpha 10/11$) (SI Fig. S4). Overall, the $\alpha 6$ region exhibits a ligand-dependent configuration, which is associated with specific intramolecular interactions within the PXR-LBD.

2.3. Different polar interactions contribute to stabilization of compound 100 and SR12813 in PXR-LBP

The analysis of $\alpha 6$ region interactions revealed that in the presence of compound 100, unlike with SR12813, H407 is not involved in water-mediated interactions between $\alpha 6$ - $\alpha 7$ loop and $\alpha 10/11$. As H407 is one of the key residues of PXR-LBP commonly participating in interactions with ligands [44,45], we next shifted our attention to protein-ligand interactions (Fig. 4). Based on ligand RMSF values, both ligands are relatively stable throughout the

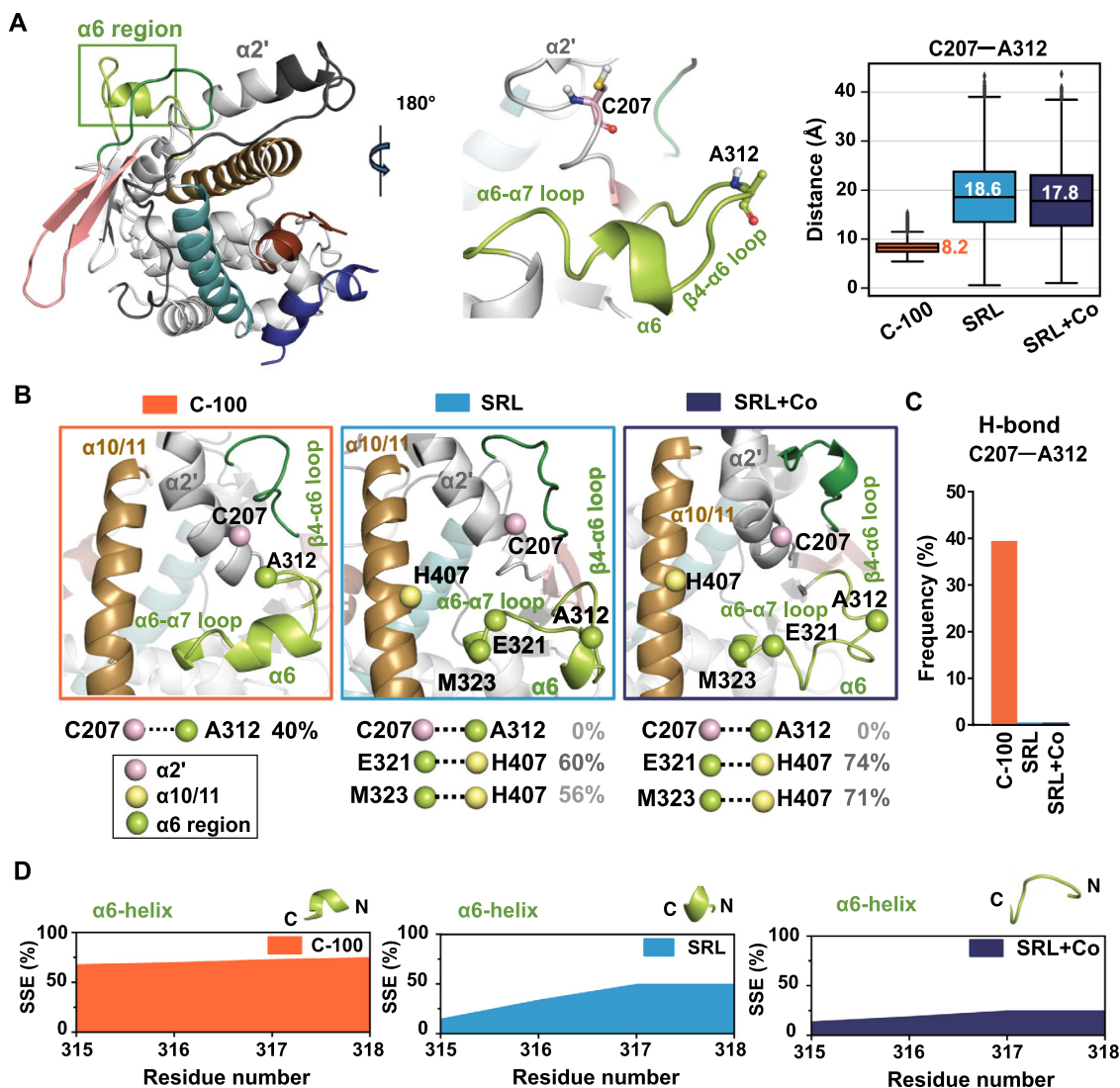


Fig. 3. Compound 100 and SRL12813 induce distinct $\alpha 6$ region conformations. (A) Location of the $\alpha 6$ region in the PXR-LBD. $C\alpha$ - $C\alpha$ distance between C207 (of $\alpha 2'$) and A312 (of $\beta 4$ - $\alpha 6$ loop), which are shown in stick representation in the middle panel, illustrate that the compound 100 promotes a closed configuration of $\beta 4$ - $\alpha 6$ loop, while more open conformations are observed with the agonist. In the boxplots, box represents interquartile range (IQR: 25–75%); the black line represents the median (values are also displayed next to each box); shown in black vertical lines, whiskers represent the data within $1.5 \times$ IQR; outliers are indicated with diamond symbols. Distances were monitored each ns, i.e., there are: $\sim 30,000$ individual datapoints for C-100; $\sim 20,000$ for SRL; and $\sim 10,000$ for SRL + Co. (B) Representative snapshots of the $\alpha 6$ region configuration of each system, water bridge and H-bond interaction differences. Closed configuration of $\beta 4$ - $\alpha 6$ loop in C-100 (left, orange box), opened configurations in agonist systems SRL (middle, light blue box) and SRL + Co (left, dark blue box). Selected water bridge interaction frequencies of the $\alpha 6$ region C-terminus residues (location of the $C\alpha$ -atoms of each residue are indicated with spheres) are shown for each system (see also SI Fig S4). Spheres belonging to the same helix/region are coloured similarly. The colours of the helices are same as in Fig. 1A. C207-A312 interaction is H-bond and other interactions are water bridges. (C) C207-A312 H-bond interaction frequencies (0% for SRL and SRL + Co). (D) Secondary structure of $\alpha 6$ helix appears more stable with compound 100 than with SRL12813. Area plots represent the observed secondary structure element (SSE) of the $\alpha 6$ helix in percentage throughout the simulation. Representative snapshots of the $\alpha 6$ secondary structure are displayed above related plot.

simulations (SI Fig. S5); thus, the data is suitable for comparing interactions between the ligands and among systems. While H407 displays a stable H-bond with SRL12813 (~ 80 – 100%), only a water-mediated interaction occurs with compound 100 ($\sim 15\%$) (Fig. 4A–B). With compound 100, H407 prefers a conformation where it forms a water bridge ($\sim 46\%$) to N404 (located on $\alpha 10/11$). This conformational preference shifts H407 away from the $\alpha 6$ region and beyond the reach of the compound 100 (Fig. 4C). Interestingly, our simulations display a stable H-bond interaction between N404 and G278 (on $\alpha 4$) with SRL12813 (~ 93 – 95%) and this interaction appears only with 23% frequency in C-100. This discrepancy could be the result of the N404 involve-

ment in the water mediated interaction with H407. Compound 100 displays H-bond and water mediated interactions to S247 and Q285 from its amide (15–20%). From these residues, SRL12813 exhibits a direct H-bond interaction only with S247. The missing H-bond to Q285 is compensated with water-mediated interactions. Similarly, increased frequency of water-mediated interaction (58%) appears in SRL, where diminished H-bond interaction to S247 exists (Fig. 4B). Furthermore, compound 100 displays an additional water-mediated interaction to T248 – albeit with low frequency ($\sim 10\%$) – from its NH linker between benzosuberone and fluorophenyl. Overall, limited polar interactions are observed for both ligands, with a clear difference in their H407 interactions.

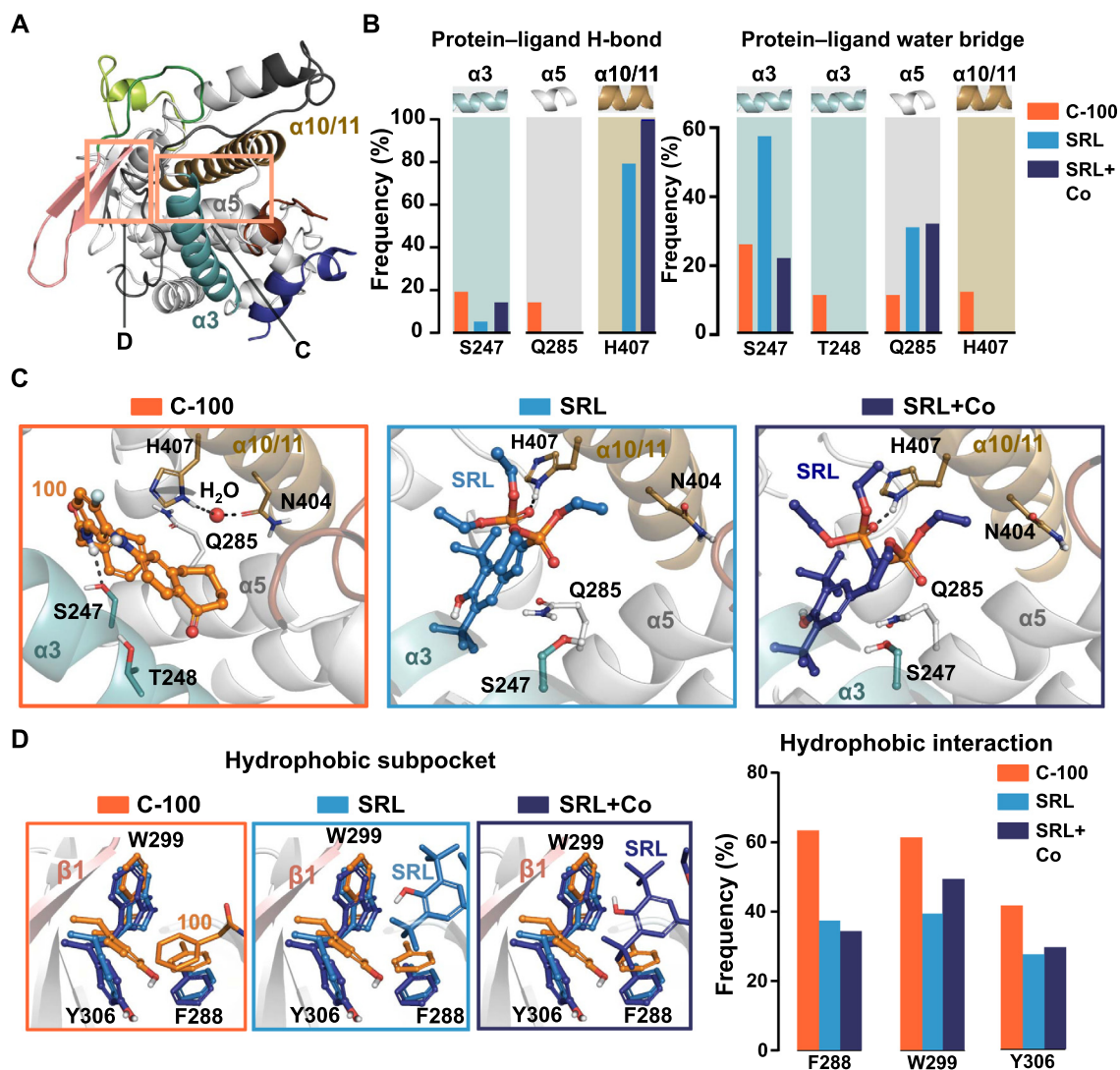


Fig. 4. The key interactions between PXR-LBD and ligands. (A) Locations of the binding site and hydrophobic subpocket in the PXR-LBD, which are shown in more detail in C and D. (B) Protein–ligand H-bond and water-mediated interaction frequencies in individual systems. (C) Representative snapshots of LBP in C-100, SRL and SRL + Co systems. The secondary structure is coloured as in A. H-bonds are displayed with dashed black lines. Water molecule that forms bridge between H407 and N404 in C-100 is illustrated with a red sphere. (D) Superposition of hydrophobic subpocket of both agonist systems over C-100. Residues displayed in dark blue, SRL + Co; light blue, SRL; orange, C-100. Frequencies of hydrophobic interactions between ligands are illustrated in the bar plot. Hydrophobic interactions include: $\pi - \pi$ interactions (face-to-face and face-to-edge) and hydrophobic interactions (see details in methods). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.4. Compound 100 and SR12813 share similar hydrophobic interactions, except in α AF-2 region

The most important interaction region for different PXR ligands is the hydrophobic subpocket, which is formed by a triad of hydrophobic residues: F288, W299 and Y306 [45]. Here, we observed a similar protein–ligand interaction pattern in the presence of SR12813 (with and without coactivator) or compound 100 (Fig. 4D). SR12813 forms hydrophobic interactions to F288 and Y306 for about 30–40% and to W299 for ~40–50% in both systems. Similarly, compound 100 displays hydrophobic interactions between benzyl group and F288, W299 (60%) and Y306 (approximately 45%). The slightly increased frequency of hydrophobic interactions for compound 100 could be attributed to the structural characteristics of the compounds. Compound 100 does not contain any polar atoms in its terminal group that binds to the hydrophobic subpocket, while SR12813 contains an aromatic hydroxyl group in its structure binding to this region which can

explain the observed shift of the Y306 and F288 side chain in C-100 compared to agonist systems (Fig. 4D; SI Fig. S6).

Along with the hydrophobic subpocket, other regions of LBP also contribute to hydrophobic interactions with the ligands (SI Fig. S7A). Namely, α 3, α 5, α 10/11 and the loop connecting α 11 to α AF-2 contain hydrophobic residues that displayed hydrophobic interactions. Clearly higher interaction frequencies with F420 and F429 (located on α AF-2) are observed for SR12813 (20–45%), while compound 100 displays 0–10% interaction frequencies with these residues. Moreover, compound 100 exhibits hydrophobic interaction with F251 (located on α 3-helix), which is not present with SR12813.

Collectively, both ligands show similar hydrophobic interaction patterns within the PXR-LBP, with the exceptions of F251 (located on α 5) and α AF-2 residues F420 and F429. Since F420 is located on the loop connecting α 11 to α AF-2 and this loop has a critical role in the α AF-2 localization [46,47], we explored F420 interactions to its neighbouring residues (SI Fig. S7B). With SR12813, F420 displays

hydrophobic interactions to L411 (located on α_{11}) for roughly 28–40% in both agonist systems. This interaction is infrequent with compound 100 (~8%). F420 also interacts with I414 for 18–30% in both agonist systems, while diminished interactions (only ~3%) appear with compound 100. Interestingly, RMSF of F420 is higher (2.4 Å) with C-100 than with SRL (1.2 Å), demonstrating a high flexibility of this amino acid residue with compound 100.

2.5. α AF-2 is stabilized with SRL12813 and destabilized with compound 100

Compound 100 exhibited diminished interactions with α AF-2. All known ligand-dependent nuclear receptors require α AF-2 domain for an effective interaction with a coactivator [48]. This domain plays a crucial role in the formation of a suitable platform for the coactivator binding on the LBD surface. Therefore, we next investigated more closely the behaviour of α AF-2 in different systems.

First, we monitored the distance between α AF-2 and α_3 -helix (Fig. 5A–B). As α_3 -helix (residues 240–260) is stable in all simula-

tions (RMSF < 1 Å), this distance enables the assessment of the relative position of α AF-2 to the LBD. In the presence of compound 100, this distance is increased (median of 13.6 Å) compared to what is observed for SRL12813 (medians ~11 Å). Furthermore, we noticed that the unfolded loop-like conformation of α AF-2 is slightly preferred with compound 100, in comparison to SRL12813 (Fig. 5C). The presence of the coactivator further stabilizes the alpha-helical secondary structure of α AF-2 based on SRL + Co. Based on the crystal structure of PXR-LBD-SRL12813 with coactivator, a H-bond interaction of T248 and T422 stabilizes α AF-2 closer to α_3 -helix (Fig. 5D). In the simulations, distance between the hydroxyl-oxygens of these residues is increased with compound 100 (median of 7.2 Å) (Fig. 5D), decreasing the H-bond frequency between these residues to 5%. In SRL + Co system, the median value of this distance is 2.9 Å and H-bond contact appears with approximately 80% frequency. Without the coactivator, SRL12813 displays values between these two systems (median of 5.4 Å, H-bond frequency 30%). Finally, we analysed the spatial orientation of α AF-2 relative to LBP using angle calculations (Fig. 5E). To this end, we selected F281 (located on α_5) as the apex (see details of angle selection in methods). The angle between N404 and F429 vectors

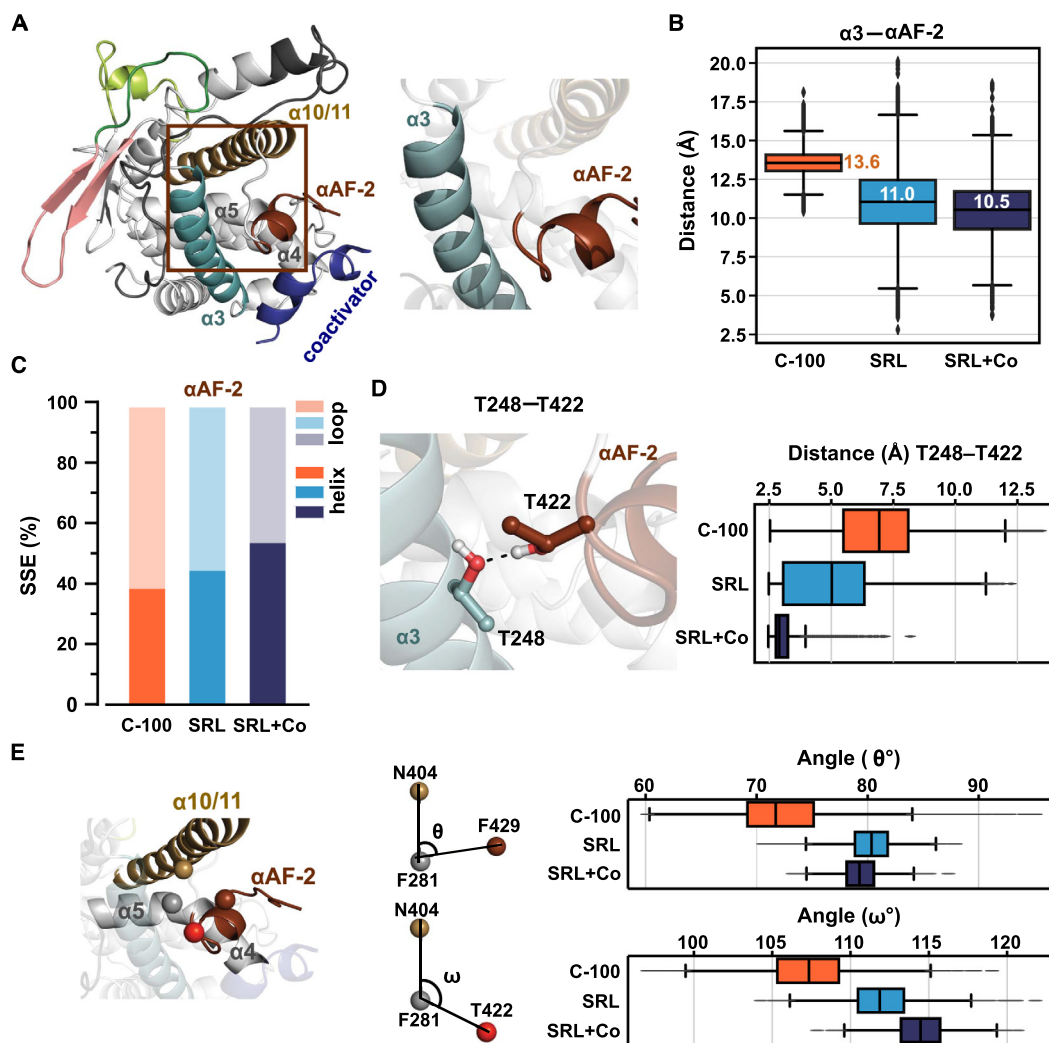


Fig. 5. Destabilization of α AF-2 with compound 100. (A) PXR-LBD crystal structure and close view of α_3 -helix and α AF-2. (B) Box plot represents the distribution of distance between centre of mass of α_3 -helix (residues 240–260) and α AF-2 (residues 423–430). (C) Secondary structure element (SSE) of α AF-2 throughout the simulation. Proportion of helix shown in darker colour, loop-like structure shown in lighter colour. (D) H-bond interacting residues T248 and T422 (PDB ID: 1NRL [22]). Distance between sidechain oxygen atoms of T248 and T422 are shown in box plot. (E) θ and ω define the angle consisting of N404, F281, F429 and N404, F281, T422, respectively where F281 is the apex. Box plots display the distribution of these angles among the three systems.

(θ) is the smallest in the presence of compound 100 (median of $\sim 71^\circ$), while larger values are observed for SRL12813 (~ 79 – 80°). This trend applies with the angle between N404 and T422 vectors (ω), where the smallest angle appears with compound 100 ($\sim 107^\circ$) and agonist systems display larger angles (~ 112 – 115°). Overall, the configuration of α AF-2 is affected by the bound ligand, and compound 100 appears to destabilize the LBD surface associated active conformation.

To evaluate the impact of the SRC-1 in the stabilization of the α AF-2 region with compound 100, we established an arbitrary system including PXR-LBD, compound 100 and SRC-1 (C-100 + Co; total simulation time of 10 μ s). Also in this system, the configuration of α AF-2 is destabilized by compound 100 and its behaviour reflects that of C-100 (SI Fig. S8). The observed distance between α AF-2 and α 3-helix with C-100 + Co (median of 13.7 Å) is not far from C-100. Furthermore, the distance between the hydroxyl-oxygens of T248 and T422 (median of 6.9 Å) is close to the C-100 than that of SRL and SRL-Co. Regarding the spatial orientation of the α AF-2 relative to LBP, the angles θ and ω of C-100 + Co fall in somewhere between C-100 and SRL systems. Finally, increased RMSF values of SRC-1 in C-100 + Co compared to SRL + Co, demonstrate the coactivator instability with compound 100. Therefore, even in the presence of SRC-1, compound 100 appears to destabilize the PXR-LBD surface on the α AF-2 region, encumbering the stable binding of SRC-1. This MD simulation data agrees with the experimental data that demonstrates the failure of compound 100 to recruit SRC-1 to PXR-LBD in coactivator recruitment mammalian 2-hybrid assay [42].

2.6. Markov state modelling reveals compound 100 specific PXR-LBD conformations

We next aimed for a deeper understanding of PXR-LBD conformational dynamics when in complex with compound 100. To this end, we conducted Markov state modelling (MSM) approach that enables the study of long timescale statistical dynamics of a protein by identifying relevant kinetic states (metastable states) and the probability distribution among these states [49,50]. MSM identified five metastable states (S_{I-V}) for the PXR-LBD bound to compound 100 (Fig. 6; SI Fig. S9). The two most dominant metastable states, S_{IV} and S_V , appear with $\sim 29\%$ and $\sim 30\%$ equilibrium probabilities, respectively. Other states (S_I , S_{II} , S_{III}) display lower probabilities in the range of 10–17%. Overall, conformations of LBD subregions in metastable state derived structures are distinct from the agonist associated. Moreover, specific conformations are preferred in individual metastable states. For the flexible $\alpha 1$ - $\alpha 2'$ loop, which flexibility is reduced by compound 100 compared to SR12813 (Fig. 2C), a state-specific conformation appears in S_I and S_{IV} , while in S_{II} , S_{III} and S_V there exist no clear configuration for this loop (Fig. 6). Of note, in the agonist-bound reference crystal structure this loop is disordered (SI Fig. S10). The $\beta 1$ - $\beta 1'$ loop in S_I , S_{II} , S_{IV} (altogether $\sim 53\%$) adopts a clearly defined folded conformation, where it is tightly packed on the LBD (Fig. 6). Configuration of this region is more ambiguous in S_{III} and S_V . In the agonist bound crystal structure, $\beta 1$ - $\beta 1'$ loop appears in an extended conformation, more distant from the LBD (Fig. 6; SI Fig. S10). Indeed, the distance between D219 (apex of $\beta 1$ - $\beta 1'$ loop) and $\alpha 3$ -helix is smaller with compound 100 than with SR12813 bound systems (SI Fig. S11A). For the $\beta 4$ - $\alpha 6$ loop, almost identical conformation is represented in all states, where it resides close to the $\alpha 2'$ (residue C207) (Fig. 6). Conversely, an open configuration for this loop is observed in the agonist-bound crystal structure (Fig. 6, SI Fig. S10). Again, distance between $\alpha 3$ -helix and A312 is smaller in C-100 compared to SRL and SRL + Co (SI Fig. S11B). The $\beta 1'$ - $\alpha 3$ loop in S_{II} , S_{IV} and S_V (altogether $\sim 73\%$) deviates the most from the agonist-associated conformation, while in S_I and S_{III} this deviation is not that evident

(Fig. 6). Interestingly, the α AF-2 region appears in a quite well-defined conformation in S_I , S_{II} , S_{III} and S_V , (altogether $\sim 71\%$), where it is shifted away from the $\alpha 3$ -helix (Fig. 6). Moreover, in S_{IV} α AF-2 appears with a more disordered conformation. These α AF-2 configurations agree with the calculated α AF-2 related distances and angles (Fig. 5). Altogether, MSM revealed unique conformations for the PXR-LBD bound to compound 100 in four regions, $\beta 1'$ - $\alpha 3$ loop, $\beta 1$ - $\beta 1'$ loop, $\beta 4$ - $\alpha 6$ loop and α AF-2, that are distinct from agonist-associated conformations.

2.7. Mutations provide insights to the PXR activation

We next evaluated experimentally the effect of eight different mutations on PXR activity and its ligand-induced activation (Fig. 7A). The alanine mutations involved selected LBD key residues with diverse locations around PXR-LBD. The competitive antagonist compound 100 induces moderately the reporter gene CYP3A4 expression with WT PXR at the tested concentration, while with the full agonists, rifampicin and SR12813, the inducement of the gene expression is manifold (for more detailed biological characterisation of compound 100 see [42]). Mutations of the hydrophobic subpocket forming residues W299 and Y306 resulted in distinct outcomes. While W299A retains the inducibility by rifampicin and behaves similarly as wild type, Y306A renders PXR inactive without any ligand-inducibility (Fig. 7B). From the mutations of the polar residues participating in H-bond interactions (Q285, S247 and H407), S247A and H407A increased the basal level of PXR activity, transforming it into a constitutive active form. These mutations appear still inducible by the PXR agonist rifampicin, and compound 100 suppresses the activity of S247A. MD simulations displayed hydrophobic interaction between F281 and both ligands (Fig. S5). Nevertheless, F281A did not alter PXR activity or inducibility by the ligands. F429A, located in α AF-2, rendered PXR inactive, and was not inducible by the ligands. Finally, W223A mutation in the putative PXR homodimerization interface [51], also resulted in the loss of PXR activity. Overall, mutation analysis revealed the important role of the key residues in modulating PXR activity and activation.

2.8. Hydrophobic subpocket binding moiety does not explain PXR conformational behaviour associated to compound 100

Werner et al. reported a P2X4 inhibitor BAY-1797, which also activated PXR with a minimum efficacious concentration of 1.7 μ M [43]. We got interested in BAY-1797, as it shares an identical phenylacetamide moiety with compound 100 (Fig. 1C). To investigate and compare the behaviour of BAY-1797 with compound 100, we carried out 10 μ s MD simulations for PXR-LBD-BAY-1797. As a starting configuration for these simulations, we utilized a crystal structure of a close analogue of BAY-1797, which exhibits one additional methyl group compared to BAY-1797 (PDB ID: 6HTY [43]). The overall dynamical behaviour of the protein in the simulations (RMSF) follows a similar trend with B-factors of the crystal structure. (SI Fig. S3C). BAY-1797 is well accommodated in the PXR-LBP and its aromatic benzyl group is oriented into the hydrophobic subpocket (Fig. 8A). A comparable profile with compound 100 is observed for BAY-1797 in π – π interactions to W299, F288 and Y306 (Fig. 8A and B, Fig. 4D). In contrast to compound 100, however, BAY-1797 displays a stable H-bond to Q285 and interaction to H407 (43%), which is closer to the SR12813 interaction profile (Fig. 8C). The secondary structure stability of $\alpha 6$ -helix with BAY-1797 resembles compound 100 (Fig. 8D). Nevertheless, based on the distance of A312–C207 (located on $\beta 4$ - $\alpha 6$ loop and on C-terminus of $\alpha 2'$, respectively) BAY-1797 falls somewhere in between of SRL12813 and compound 100 (Fig. 8E), indicating unique conformation for this region. In addition, distance

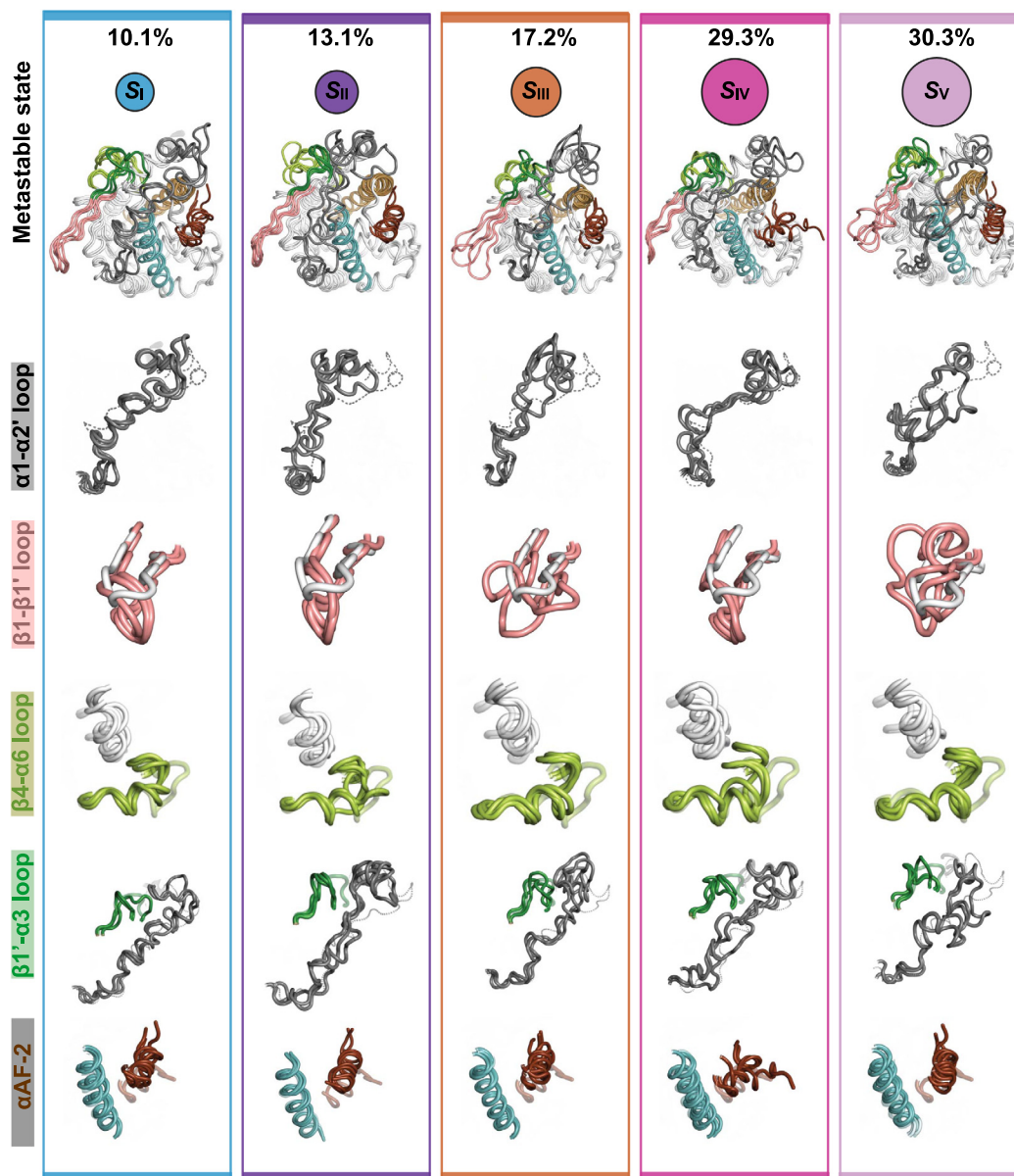


Fig. 6. Markov state modelling reveals five metastable states for the PXR-LBD bound to compound 100. The five metastable states (S_{I-V}) are displayed with three structures. The area of a circle is proportional to the equilibrium probability (π_i) associated to each metastable state (also shown with %). The selected substructure configurations are shown individually, with a reference conformation from an agonist bound crystal structure (PDB ID: 1NRL). Reference crystal substructures are illustrated as follows: $\alpha 1-\alpha 2'$ loop, grey dashed line (disordered in the crystal structure); $\beta 1-\beta 1'$ loop, white; $\beta 4-\alpha 6$ loop, transparent light green, located in the vicinity of $\alpha 2'$ (white helix); $\beta 1'-\alpha 3$ loop, transparent dark green; $\alpha AF-2$, transparent dark brown, located in the vicinity of $\alpha 3$ -helix (cyan helix). The structures of the metastable states are also provided in supplementary PyMOL (v.2.4.2) session-files. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

between $\alpha AF-2$ and $\alpha 3$ -helix in PXR-LBD–BAY-1797 appears in between the SR12813 and compound 100 systems (Fig. 8F). However, the distance between hydroxyl-oxygens of T248 and T422 suggests that with BAY-1797, $\alpha AF-2$ can acquire a stable active configuration that is required for PXR activation (Fig. 8G). This leads hydrophobic interaction with F420 for 30% comparable to SR12813 but no interaction was observed with F251 and F429 in presence of BAY-1797 (SI Fig. S12). The angle between N404 and F429 vectors (θ) is larger (median of $\sim 80^\circ$) compared to that of compound 100 and close to what is observed for SR12813 (Fig. 8H; Fig. 5). This trend applies with the other angle between N404 and T422 vectors (ω), with an angle (median of 114°) close to that of SR12813. Overall, the configuration of $\alpha AF-2$ appears to be stabilized on LBD surface with BAY-1797.

3. Discussion

Since the mechanism of PXR antagonism has not been elucidated and X-ray crystallography is unable to capture the ligand-induced conformational dynamics of PXR-LBD [13], we utilized here in silico approach to disclose the putative ligand-dependent differences in conformational dynamics of PXR-LBD. Our MD data suggest ligand-dependent discrepancy in the conformational preference on different LBD regions. This discrepancy is observed in $\alpha 6$ region, $\alpha AF-2$, $\alpha 1-\alpha 2'$, $\beta 1'-\alpha 3$ and $\beta 1-\beta 1'$ loop.

Earlier studies reported that $\alpha 6$ is less folded or very dynamic in PXR [52,53], enabling a flexible cavity to accommodate ligands of different sizes [19]. We found that compound 100 induced a tightly packed and folded conformation in this region, while a looser con-

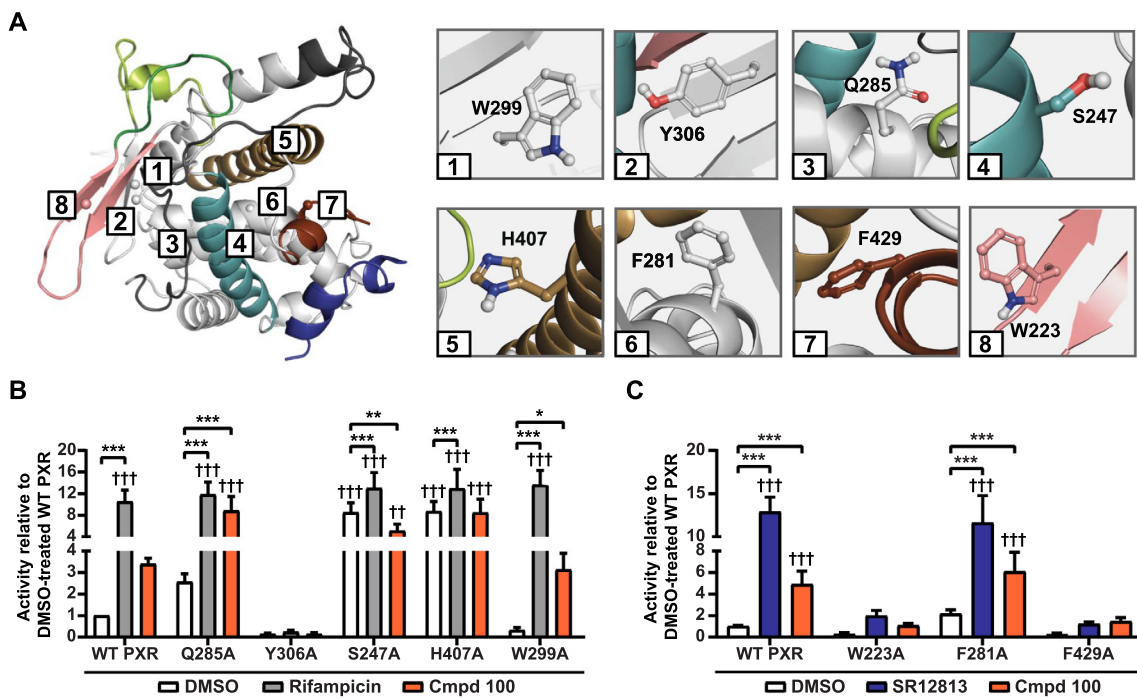


Fig. 7. Ligand-induced effects on activation of PXR mutants. (A) The reference crystal structure (PDB ID: 1NRL) illustrating the locations of the generated PXR mutants (locations of the α -atoms of each mutated residue are indicated with spheres). Each number represents the location of a mutated residue (left). Close-ups of the mutated wild type residues are shown in stick representation (right). (B) Observed ligand induced PXR activation in reporter gene assay. Transfected cells were treated 24 h with 0.1% DMSO, 10 μ M rifampicin or 1 μ M SR12813 (C), or 10 μ M compound 100. Data is presented as mean relative activity \pm SD to DMSO-treated WT PXR from five independent experiments with technical triplicates. ††† $p < 0.001$ compared to DMSO-treated WT PXR analysed with two-way anova with Dunnett's multiple comparisons test. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, compared to DMSO-treated respective mutant analysed with two-way anova with Tukey's multiple comparison test.

figuration was preferred with SR12813. Motta et al. predicted that the water channel between $\alpha 2$ and $\alpha 6$ would be an entry pathway for SR12813 [40]. This region appears as a favourable ligand pathway among nuclear receptors [54]. Interestingly, our simulations suggest that water bridged interactions mediated by E321 and M323 on $\alpha 6$ and H407 on $\alpha 10/11$ appear on this site with SR12813, where no such interaction exists with compound 100. A common feature of a PXR agonist is to engage in direct interactions with H407 of the $\alpha 10/11$ [25,44,45,55], and H407A mutation renders PXR constitutively active. Our simulations suggest that compound 100 does not rely on interactions with H407. This appears to be related to the engagement of H407 in water mediated interaction to N404.

Anami et al. [46] proposed a model for vitamin D receptor (VDR) activation or repression called “folding-door model” to explain VDR-LBD activity. In this model, $\alpha 11$ cooperates with $\alpha AF-2$ in a way that in the presence of an agonist $\alpha AF-2$ is stabilized close to $\alpha 3$ -helix, forming internal interactions with the $\alpha 11$ kink. These interactions close the door ($\alpha 11$). Meanwhile this kink is open in the presence of an antagonist and the $\alpha 11$ - $\alpha AF-2$ loop plays an important role in the unsuitable $\alpha AF-2$ positioning for receptor activation. Here, our simulations display a stable H-bond interaction between N404 (on $\alpha 11$) and G278 (on $\alpha 4$) with SR12813, which is an infrequent interaction with compound 100. Hence, we hypothesize that with PXR the ligand-dependent N404 conformation – and the lack of water bridge interaction between H407 and $\alpha 6$ in presence of compound 100 – may result in a more flexible $\alpha 11$ with compound 100. This would result in the rearrangement of the flexible $\alpha 11$ - $\alpha AF-2$ loop. We also observed a ligand-dependent discrepancy in F420 and $\alpha 11$ interactions, and diminished hydrophobic interaction between compound 100 and this loop. Our findings agree with earlier results reported by Shizu et al. that revealed the important role of F420 (on $\alpha 11$ - $\alpha AF-2$ loop)

in PXR $\alpha AF-2$ stabilization [47]. Moreover, mutation of F429A, located in $\alpha AF-2$, rendered PXR inactive, and was not inducible by the ligands highlighting its relevance in the binding. Altogether, we conclude that in presence of compound 100 these motifs play a role in dislocation of $\alpha AF-2$ from the vicinity of $\alpha 3$ -helix, which is demonstrated by the distance and angle calculations of $\alpha AF-2$ as well as by the MSM. This dislocation of $\alpha AF-2$ was also observed in the simulations of compound 100 in the presence of SRC-1 (C-100 + Co). With SR12813 a stable conformation of $\alpha AF-2$ is maintained, providing a suitable platform for the co-activator binding and subsequent PXR activation. Of note, not only is compound 100 impairing binding of coactivator SRC-1 but also binding of corepressor SMRT [42]. This is in contrast to what is observed with the full antagonist SPA70, which enables recruitment of the corepressor [27]. Therefore, the induced conformational changes by compound 100 enable competitive PXR antagonism but not full antagonism, which may require that the ligand induces suitable conformations for the corepressor binding.

The adaptability of the hydrophobic subpocket in PXR is emphasized by the different conformations of Y306 and F288 that exist with SR12813 and compound 100. Meanwhile, mutation of Y306 renders PXR inactive with both ligands. This could be attributed to the loss of PXR-LBD integrity, as in multiple PXR crystal structures there exists a H-bond interaction between Y306 and H242 (located on $\alpha 3$ -helix), highlighting the important role of this residue in PXR structural stability.

The highest flexibility of PXR-LBD was observed in $\alpha 1$ - $\alpha 2'$ loop with both compound 100 and SR12813. This flexibility is well exemplified by structural data, as this region is disordered in all publicly available PXR structures. Earlier studies show also the high degree of flexibility in this region in the presence of agonist [56,57]. Our MD data displayed somewhat lower RMSF for this region with compound 100. MSM suggested that in some meta-

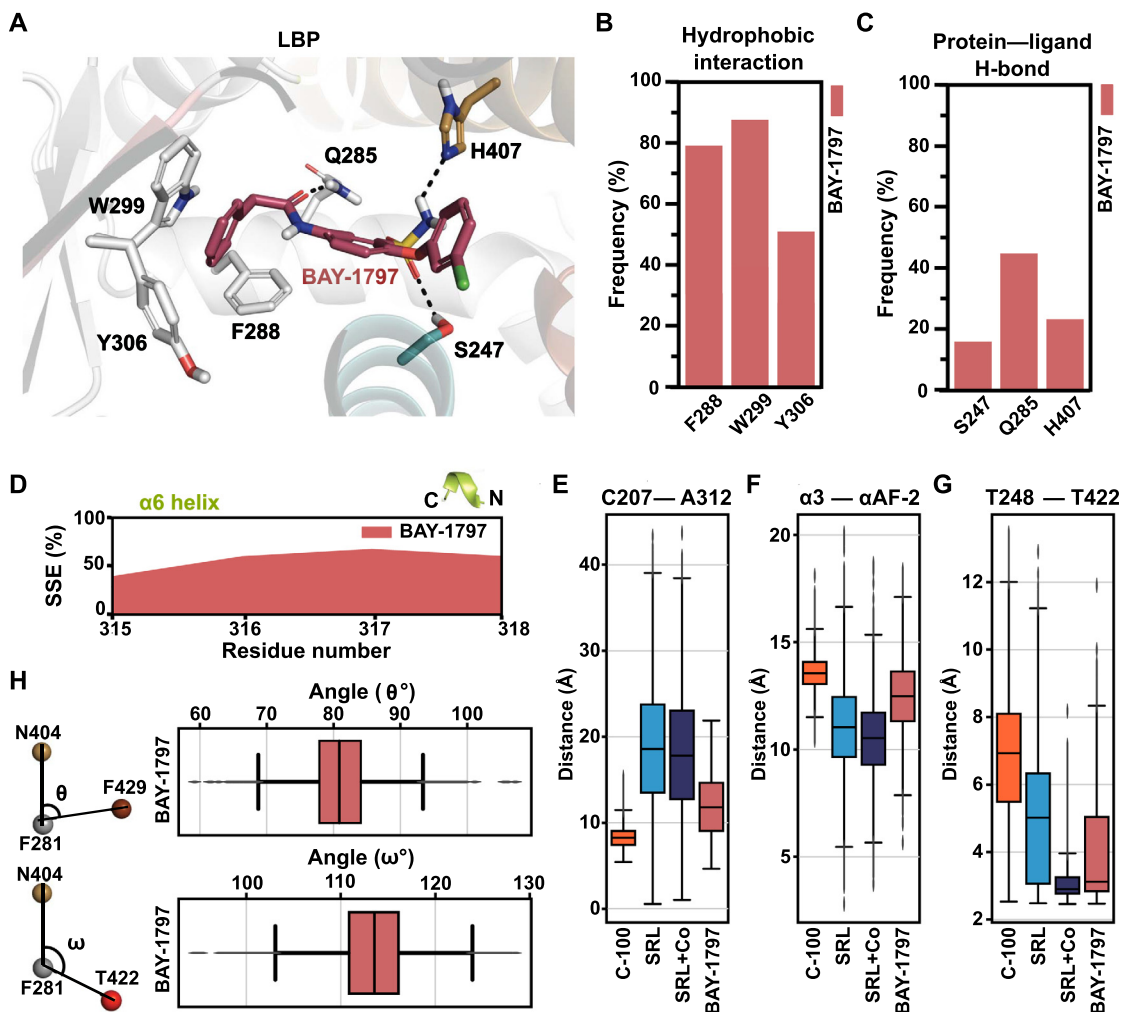


Fig. 8. PXR-LBD-BAY-1797 simulations display behaviour of an agonist. (A) A representative snapshot of LBP with BAY-1797. The colour of the residues is as in PXR-LBP in Fig. 1B. H-bonds are displayed with dashed black lines. (B) Protein-ligand hydrophobic interactions frequencies between ligand and protein in individual systems are illustrated in the bar plot. Hydrophobic interactions include: π – π interactions (face-to-face and face-to-edge) and hydrophobic interactions (see details in methods). (C) Protein-ligand H-bond interactions frequencies between ligand and protein in individual systems are illustrated in the bar. (D) Area plots represent the observed secondary structure element (SSE) of the $\alpha 6$ helix in percentage of BAY-1797 throughout the simulation. (E) $C\alpha$ – $C\alpha$ distance between C207 (of $\alpha 2'$) and A312 (of $\beta 4$ – $\alpha 6$ loop) illustrates that the BAY-1797 promotes a further closed configuration of $\beta 4$ – $\alpha 6$ loop like, while more open conformations are observed with the agonist. (F) Box plot represents the distribution of distance between centre of mass of $\alpha 3$ -helix (residues 240–260) and $\alpha AF-2$ (residues 423–434). (G) Distance between sidechain oxygen atoms of T248 and T422 are shown in box plot. (H) Internal protein interactions of $\alpha AF-2$ region. θ and ω define the angle consists of N404, F281, F429 and N404, F281, T422, respectively where F281 is the apex. Box plots display the distribution of these angles with BAY-1797 (median of $\sim 80^\circ$ for θ and $\sim 113^\circ$ for ω). Respective medians for the other systems are for θ : C-100, $\sim 71^\circ$; SRL, $\sim 80^\circ$; SRL + Co $\sim 79^\circ$ and ω : C-100, $\sim 107^\circ$; SRL, $\sim 112^\circ$; SRL + Co $\sim 115^\circ$.

stable states there exists rather state-specific conformation for this loop, while in others states more disordered configuration appears. It is worth noting that this loop (in proximity of $\alpha 2'$) is part of one of the proposed water channels in PXR [40] that stretches along $\beta 1'$ – $\alpha 3$ loop [58]. For this $\beta 1'$ – $\alpha 3$ loop, MSM displayed a clear deviation from the agonist-associated conformation. Further study is required to disclose the role of different conformations of the ambiguous $\alpha 1$ – $\alpha 2'$ loop together with $\beta 1'$ – $\alpha 3$ loop, to better understand their interrelation on the ligand binding and specificity.

Our simulations suggest that the $\beta 1$ – $\beta 1'$ loop is tightly packed on the LBD with compound 100. Conversely, with SR12813 this substructural element resides farther from the $\alpha 3$ -helix. Interestingly, the W223A mutation on this outer interface renders PXR inactive with both SR12813 and compound 100. The functional relevance of this residue was also revealed by Nobel et al. [51]. Of note, this residue is part of PXR-LBD insertion [58].

For BAY-1797, that shares an identical phenylacetamide moiety with compound 100, we observed an intermediate profile in

between SR12813 and compound 100. It must be noted that SR12813 is a more potent PXR agonist than BAY-1797. In the hydrophobic subpocket BAY-1797 shows a similar behaviour as compound 100. In addition, in the presence of BAY-1797 the distance between $\alpha AF-2$ and $\alpha 3$ -helix is increased compared to SR12813, but not to the extent of compound 100. Nevertheless, the distance between T248 and T422 with BAY-1797 is close to what is observed with SR12813. It was noted by Werner et al. [43] that changing the 3-chlorophenoxy with a larger and more polar substituent alleviated PXR agonism, which could occur either by diminishing PXR binding or by disrupting the $\alpha AF-2$. In this regard, we suggest that along with the classical mechanism of action of a nuclear receptor, the more specific distance of T248 and T422 may be useful when analysing potential risk for PXR agonism by MD simulations.

Overall, our results revealed the adaptability and ligand-specificity in PXR conformational behaviour. More folded $\beta 1$ – $\beta 1'$ loop, compact $\alpha 6$ region, lower flexible $\alpha 1$ – $\alpha 2'$ loop, dislocated

$\beta 1'$ - $\alpha 3$ loop and $\alpha AF-2$ are associated with compound 100. Our study provides more in depth understanding of ligand-induced changes in PXR-LBD substructures. Although our compound is a competitive antagonist [42] and not a full antagonist as SPA70 [27], these results still provide a putative template for designing PXR antagonists. Finally, the identified structural key-regions provide guidance how to potentially avoid PXR agonism.

4. Material and methods

4.1. MD simulations

Modelling was conducted with Maestro (Schrödinger Release 2019-4: Maestro, Schrödinger, LLC, New York, NY, 2019), and OPLS3e force field [59,60], unless otherwise stated. For the simulations of C-100 and C-100 + SRC, we used the PXR crystal structure PDB ID: 4J5W (chain A) [61] and for the simulation of SRL and SRL + Co, we used the PXR crystal structure PDB ID: 1NRL (chain A) [22]. For BAY-1797 simulations, we applied its close analogue PXR co-crystal structure PDB ID: 6HTY (chain B) [43], and the redundant methyl group of the analogue was deleted. The residues missing in the C-terminal (G433 and/or S434) were added to the structures with Maestro tools. For SRL and C-100 systems, SRC-1 was removed. The proteins were prepared using Protein Preparation Wizard (Schrödinger LLC, New York, NY, 2019) [62]. Missing hydrogen atoms were added, bond orders were assigned using CCD database, missing side chains and loops ($\alpha 1$ - $\alpha 2$ loop; residues 178–192 for 4J5W and $\alpha 1$ - $\alpha 2$ loop; residues 178–191 for 1NRL and 6HTY) were filled to the structure using Prime [63], protonation states of amino acids were optimized with PROPKA (Schrödinger, LLC, New York, NY, 2019), the coactivator peptide was capped in both termini and the structures were minimized. To obtain the starting configuration for compound 100, Glide docking was conducted (Glide v. 7.7) [64,65]. Before docking, compound 100, was prepared with LigPrep (Schrödinger, LLC, New York, NY, 2019) to assign the protonation state (Epik; at pH 7.0 +/-2.0) and the partial charges. For the docking of compound 100, default settings were applied, with residues H407, Q285, S247, H327 and F429 selected to define the active site, and the docking was conducted using the standard-precision (SP) and extra-precision (XP) level of accuracy [66]. The docking resulted in an U-shape pose (mainly accommodated in hydrophobic subpocket) from the SP (docking score: -10.542; glide emodel: -86.922) and an extended pose (oriented from hydrophobic subpocket with benzyl moiety, while benzo-suberone oriented towards $\alpha AF-2$ region) from XP (docking score: -13.647, emodel: -98.455). We evaluated these two distinguishable poses in short MD simulations (data not shown) and the pose with extended conformation displayed better stability during the simulation, which was selected as a starting configuration for compound 100 in the production simulations. Of note, based on our later evaluation by the QM Conformer predictor tool, this extended conformation is also proposed for compound 100 as the lowest energy conformation in water [42]. The same pose was also used in C-100 + Co simulation where the SRC-1 peptide was maintained.

For the simulations, we used Desmond MD simulation engine [67]. The prepared systems were solvated in a cubic box with the size of the box set as 15 Å minimum distance from the box edges to any atom of the protein. TIP3P water model [68] was used to describe the solvent and the net charge was neutralized using K^+ ion with final salt concentration of 150 mM. RESPA integrator timesteps of 2 fs for bonded and near, and 6 fs for far were applied. The short-range coulombic interactions were treated using a cut-off value of 9.0 Å. Before the production simulations the systems were relaxed using the default Desmond relaxation protocol. Simulations were run in NPT ensemble, with temperature of 310 K

(Nosé-Hoover thermostat) and pressure of 1.01325 bar (Martyna-Tobias-Klein barostat). For each system, simulations of five replicas with different lengths (2–4 μs) were carried out, resulting in total of 30 μs simulation data for C-100, 20 μs for SRL, 10 μs for SRL + Co, 10 μs for C-100 + SRC and 10 μs for BAY-1797 (SI Table 3; SI Fig. S13). With C-100, additional independent replicas derived from the original simulations were run to obtain sufficient sampling for MSM.

4.2. Analysis of MD simulation data

Principal component analysis. PCA was conducted for the backbone atoms using GROMACS tools (version 2019) (gmx covar and gmx anaeg) [69]. For GROMACS analysis, the Desmond trajectories were aligned and transformed to xtc-format, keeping only backbone atoms. Figures describing the extreme motions were generated and visualized using PyMOL-script Modevectors [70].

RMSD, RMSF, protein secondary structure elements (SSE) and interaction analysis. Maestro simulation interaction analysis tool (Schrödinger, LLC) was used for these analyses. For interaction criteria, default values were used. H-bonds: cut-off of 2.5 Å for donor and acceptor atoms, donor angle of 120° and acceptor angle of 90°. Hydrophobic interactions: cut-off of 3.6 Å between ligand's aromatic or aliphatic carbons and a hydrophobic side chain, π - π interaction was defined as two aromatic groups stacked face-to-face or face-to-edge. Water bridge interactions: default cut-off of 2.8 Å for donor and acceptor atoms, donor angle of 110° and acceptor angle of 90°.

Angle and distance calculations. Maestro event analysis tool (Schrödinger, LLC) was used. Distances between specific secondary structure elements were calculated using their centers of mass with the Maestro script *trj_asl_distance.py* (Schrödinger LLC). For $\alpha 3$ -helix residues 240–260 and $\alpha AF-2$ residues 423–430 were used. For the distance calculation $C\alpha$ atom of each residue was used. The angles were calculated using the $C\alpha$ atom of N404, F281, F429 for θ and $C\alpha$ atom of N404, F281, T422 for ω with the Maestro script *event_analysis.py* and *analyze_simulation.py* (Schrödinger LLC).

Markov state modelling. Bayesian MSM was generated with PyEMMA 2 following the general recommendations [71]. As an input, we used full protein backbone torsion angles. Time-lagged independent component analysis (TICA) was used for dimension reduction [72], using 10 ns as a lag time, and two dimensions. The output of TICA was discretized to microstates using the k-means clustering (number of clusters set as \sqrt{N}), and the microstates were assigned in five macrostates (metastable states) by the Perron-cluster cluster analysis (PCCA++) method [73]. Implied timescales and Chapman-Kolmogorov test suggest a valid model (SI Fig. S14).

Structure and data visualization. Structure visualization was conducted with PyMOL v.2.4 (Schrödinger LLC, New York, NY, USA). Data visualization was completed by Python 3.7, seaborn [74], matplotlib [75] and GraphPad prism (v. 8.0.0 for Windows, GraphPad Software, San Diego, CA, USA).

4.3. Chemicals and reagents

DMSO was purchased from Sigma Aldrich (Taufkirchen, Germany). Rifampicin and SR12813 were obtained from Tocris Bioscience (Bristol, UK). Compound 100 was synthesized in house [42]. Minimum essential medium (MEM) and Trypsin-EDTA were purchased from Thermo Fischer Scientific (Waltham, MA, USA). L-glutamine and penicillin-streptomycin mixture were provided by Biozym (Hessisch Oldendorf, Germany). Fetal bovine serum (FBS) was obtained from Biowest (Nuaille, France).

4.4. Plasmids

Full-length human PXR [76] and CYP3A4 enhancer/promoter reporter gene plasmid pGL4-CYP3A4 (7830Δ7208-364) [77] have been described previously. Metridia luciferase expression plasmid pMetLuc2control was obtained from Takara-Clontech (Mountain View, CA, USA). Site-directed mutagenesis of the full-length PXR expression plasmid with suitable oligonucleotides designed with NEBaseChanger using Q5 Site-Directed Mutagenesis kit (New England Biolabs, Ipswich, MA, USA) was utilized to generate PXR mutants Q285A, Y306A, S247A, H407A, W223A, F281A, W299A, and F429A. The mutations were confirmed by sequencing. Plasmids were purified using PureYield Plasmid Midiprep System (Promega, Madison, WI, USA).

4.5. Cell culture

HepG2 cells (HB-8065, lot number 58341723, ATCC, Manassas, VA) were cultivated at 37 °C, 5% CO₂ in MEM, which was supplemented with 10% FBS, 2 mM glutamine, 100 U/ml penicillin and 100 µg/ml streptomycin. HepG2 cells were originally obtained at passage 74, propagated and used in the experiments between passages 93 and 104. In chemical treatments, regular FBS was replaced by dextran-coated charcoal-treated FBS. Cells were routinely checked for contamination with mycoplasma by PCR (VenorGem Classic, Minerva Biolabs, Berlin, Germany).

4.6. Transient transfections

Transient batch transfection with HepG2 was conducted using 0.6 µl JetPEI transfection reagent per well in a final volume of 25 µl (Polyplus, Illkirch, France). Per well, 0.27 µg pGL4-CYP3A4 (-7830Δ7208-364) luciferase reporter gene plasmid, 0.01 µg Metridia luciferase plasmid pMetLuc2-control and 0.03 µg of expression plasmids encoding human PXR or PXR mutants, were diluted in 150 mM NaCl to a final volume of 25 µl. After at least 24 h incubation, cells were treated for 24 h with 0.1% DMSO, 10 µM rifampicin, 1 µM SR12813 or 10 µM test compounds. Metridia luciferase was measured directly from 10 µl of medium with 100 µl Renilla luciferase assay solution [78] using EnSpire 2300 multimode plate reader (PerkinElmer, Rodgau, Germany) for 0.1 s. For firefly luciferase measurement, cells were lysed with passive lysis buffer (Promega, Madison, WI, USA). 10 µl of lysate was combined with 150 µl firefly luciferase assay solution [76] and activity measured with EnSpire 2300 multimode plate reader for 0.1 s. Results were normalized by dividing firefly luciferase activity by Metridia luciferase activity measured in the same well. Expression of mutants was assessed with Western blot (Supplementary Methods Fig. 1).

Data availability

Raw trajectories of the MD simulations are freely available at <https://doi.org/10.5281/zenodo.6048723>; <https://doi.org/10.5281/zenodo.6355467>; <https://doi.org/10.5281/zenodo.6615454>.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources. A.R. acknowledges European

Union's Horizon 2020 research and innovation program under grant agreement No 825762, EDCMET project. T.P. acknowledges financial support from the Orion Research Foundation sr, European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie (grant no 839230) and Academy of Finland GeneCellNano Flagship (337120). O.B., M.S. Robert Bosch Foundation, Stuttgart, Germany; M.S. Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy-EXC 2180–390900677; E.-K.M., O.B., S.L. Interfaculty Center for Pharmacogenomics and Pharma Research of the University of Tübingen, Tübingen, Germany.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2022.06.020>.

References

- [1] Ihunnah CA, Jiang M, Xie W. Nuclear Receptor PXR, transcriptional circuits and metabolic relevance. *Biochim Biophys Acta* 2011;1812:956–63. <https://doi.org/10.1016/j.bbadis.2011.01.014>.
- [2] Wang Y-M, Lin W, Chai SC, Wu J, Ong SS, Schuetz EG, et al. Piperine activates human pregnane X receptor to induce the expression of cytochrome P450 3A4 and multidrug resistance protein 1. *Toxicol Appl Pharmacol* 2013;272:96–107. <https://doi.org/10.1016/j.taap.2013.05.014>.
- [3] Handschin C, Meyer UA. Induction of Drug Metabolism: The Role of Nuclear Receptors. *Pharmacol Rev* 2003;55:649–73. <https://doi.org/10.1124/pr.55.4.2>.
- [4] Tolson AH, Wang H. Regulation of drug-metabolizing enzymes by xenobiotic receptors: PXR and CAR. *Adv Drug Deliv Rev* 2010;62:1238–49. <https://doi.org/10.1016/j.addr.2010.08.006>.
- [5] Watkins RE, Maglich JM, Moore LB, Wisely GB, Noble SM, Davis-Searles PR, et al. 2.1 Å Crystal Structure of Human PXR in Complex with the St. John's Wort Compound Hyperforin. *Biochemistry* 2003;42:1430–8. <https://doi.org/10.1021/bi0268753>.
- [6] Wang Y-M, Chai SC, Brewer CT, Chen T. Pregnane X receptor and drug-induced liver injury. *Expert Opin Drug Metab Toxicol* 2014;10:1521–32. <https://doi.org/10.1517/17425255.2014.963555>.
- [7] Honkakoski P, Sueyoshi T, Negishi M. Drug-activated nuclear receptors CAR and PXR. *Ann Med* 2003;35:172–82. <https://doi.org/10.1080/07853890310008224>.
- [8] Köhle C, Bock KW. Coordinate regulation of human drug-metabolizing enzymes, and conjugate transporters by the Ah receptor, pregnane X receptor and constitutive androstane receptor. *Biochem Pharmacol* 2009;77:689–99. <https://doi.org/10.1016/j.bcp.2008.05.020>.
- [9] EMA. Guideline on the investigation of drug interactions. European Medicines Agency 2012. Available at: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-investigation-drug-interactions-revision-1_en.pdf.
- [10] FDA. Clinical Drug Interaction Studies – Cytochrome P450 Enzyme- and Transporter-Mediated Drug Interactions Guidance for Industry. U.S. Department of Health and Human Services Food and Drug Administration; 2021. Available at: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-drug-interaction-studies-cytochrome-p450-enzyme-and-transporter-mediated-drug-interactions>
- [11] Ratajowski M, Grzelak I, Wiśniewska K, Ryba K, Gorzkiewicz M, Walczak-Drzewiecka A, et al. Screening of a chemical library reveals novel PXR-activating pharmacologic compounds. *Toxicol Lett* 2015;232:193–202. <https://doi.org/10.1016/j.toxlet.2014.10.009>.
- [12] Chai SC, Cherian MT, Wang Y-M, Chen T. Small-molecule modulators of PXR and CAR. *Biochim Biophys Acta* 2016;1859:1141–54. <https://doi.org/10.1016/j.bbagr.2016.02.013>.
- [13] Chai SC, Wright WC, Chen T. Strategies for developing pregnane X receptor antagonists: Implications from metabolism to cancer. *Med Res Rev* 2020;40:1061–83. <https://doi.org/10.1002/med.21648>.
- [14] Delfosse V, Huet T, Harrus D, Granell M, Bourguet M, Gardia-Parège C, et al. Mechanistic insights into the synergistic activation of the RXR–PXR heterodimer by endocrine disruptor mixtures. *Proc Natl Acad Sci USA* 2021;118. <https://doi.org/10.1073/pnas.2020551118>.
- [15] Wang H, LeCluyse EL. Role of orphan nuclear receptors in the regulation of drug-metabolizing enzymes. *Clin Pharmacokinet* 2003;42:1331–57. <https://doi.org/10.2165/00003088-200342150-00003>.
- [16] Pasquel D, Dorcakova A, Li H, Kortager S, Krasowski MD, Biswas A, et al. Acetylation of lysine 109 modulates pregnane X receptor DNA binding and transcriptional activity. *Biochim Biophys Acta* 2016;1859:1155–69. <https://doi.org/10.1016/j.bbagr.2016.01.006>.
- [17] Dussault I, Lin M, Hollister K, Fan M, Termini J, Sherman MA, et al. A Structural Model of the Constitutive Androstane Receptor Defines Novel Interactions That Mediate Ligand-Independent Activity. *Mol Cell Biol* 2002;22:5270–80. <https://doi.org/10.1128/MCB.22.15.5270-5280.2002>.

- [18] Rochel N, Wurtz JM, Mitschler A, Klaholz B, Moras D. The crystal structure of the nuclear receptor for vitamin D bound to its natural ligand. *Mol Cell* 2000;5:173–9. [https://doi.org/10.1016/s1097-2765\(00\)80413-x](https://doi.org/10.1016/s1097-2765(00)80413-x).
- [19] Watkins RE, Wisely GB, Moore LB, Collins JL, Lambert MH, Williams SP, et al. The human nuclear xenobiotic receptor PXR: structural determinants of directed promiscuity. *Science* 2001;292:2329–33. <https://doi.org/10.1126/science.1060762>.
- [20] Ekins S, Schuetz E. The PXR crystal structure: the end of the beginning. *Trends Pharmacol Sci* 2002;23:49–50. [https://doi.org/10.1016/s0165-6147\(02\)01977-6](https://doi.org/10.1016/s0165-6147(02)01977-6).
- [21] Ekins S, Kortagere S, Iyer M, Reschly EJ, Lill MA, Redinbo MR, et al. Challenges predicting ligand-receptor interactions of promiscuous proteins: the nuclear receptor PXR. *PLoS Comput Biol* 2009;5:. <https://doi.org/10.1371/journal.pcbi.1000594>.
- [22] Watkins RE, Davis-Searles PR, Lambert MH, Redinbo MR. Coactivator binding promotes the specific interaction between ligand and the pregnane X receptor. *J Mol Biol* 2003;331:815–28. [https://doi.org/10.1016/s0022-2836\(03\)00795-2](https://doi.org/10.1016/s0022-2836(03)00795-2).
- [23] Dash AK, Yende AS, Jaiswal B, Tyagi RK. Heterodimerization of Retinoid X Receptor with Xenobiotic Receptor partners occurs in the cytoplasmic compartment: Mechanistic insights of events in living cells. *Exp Cell Res* 2017;360:337–46. <https://doi.org/10.1016/j.yexcr.2017.09.024>.
- [24] Johnson DR, Li C-W, Chen L-Y, Ghosh JC, Chen JD. Regulation and Binding of Pregnane X Receptor by Nuclear Receptor Corepressor Silencing Mediator of Retinoid and Thyroid Hormone Receptors (SMRT). *Mol Pharmacol* 2006;69:99–108. <https://doi.org/10.1124/mol.105.013375>.
- [25] Hall A, Chanteux H, Ménochet K, Ledecq M, Schulze M-S-E-D. Designing Out PXR Activity on Drug Discovery Projects: A Review of Structure-Based Methods, Empirical and Computational Approaches. *J Med Chem* 2021;64:6413–522. <https://doi.org/10.1021/acs.jmedchem.0c02245>.
- [26] Zhou C, Poulton E-J, Grün F, Bammler TK, Blumberg B, Thummel KE, et al. The Dietary Isothiocyanate Sulforaphane Is an Antagonist of the Human Steroid and Xenobiotic Nuclear Receptor. *Mol Pharmacol* 2007;71:220–9. <https://doi.org/10.1124/mol.106.029264>.
- [27] Lin W, Wang Y-M, Chai SC, Lv L, Zheng J, Wu J, et al. SPA70 is a potent antagonist of human pregnane X receptor. *Nat Commun* 2017;8:741. <https://doi.org/10.1038/s41467-017-00780-5>.
- [28] Venkatesh M, Wang H, Cayer J, Leroux M, Salvail D, Das B, et al. In Vivo and In Vitro Characterization of a First-in-Class Novel Azole Analog That Targets Pregnane X Receptor Activation. *Mol Pharmacol* 2011;80:124–35. <https://doi.org/10.1124/mol.111.071787>.
- [29] Wang H, Li H, Moore LB, Johnson MDL, Maglich JM, Goodwin B, et al. The Phytoestrogen Coumestrol Is a Naturally Occurring Antagonist of the Human Pregnane X Receptor. *Mol Endocrinol* 2008;22:838–57. <https://doi.org/10.1210/me.2007-0218>.
- [30] Ekins S, Kholodovych V, Ai N, Sinz M, Gal J, Gera L, et al. Computational Discovery of Novel Low Micromolar Human Pregnane X Receptor Antagonists. *Mol Pharmacol* 2008;74:662–72. <https://doi.org/10.1124/mol.108.049437>.
- [31] Li Y, Lin W, Wright WC, Chai SC, Wu J, Chen T. Building a Chemical Toolbox for Human Pregnane X Receptor Research: Discovery of Agonists, Inverse Agonists, and Antagonists Among Analogs Based on the Unique Chemical Scaffold of SPA70. *J Med Chem* 2021;64:1733–61. <https://doi.org/10.1021/acs.jmedchem.0c02201>.
- [32] Ekins S, Chang C, Mani S, Krasowski MD, Reschly EJ, Iyer M, et al. Human pregnane X receptor antagonists and agonists define molecular requirements for different binding sites. *Mol Pharmacol* 2007;72:592–603. <https://doi.org/10.1124/mol.107.038398>.
- [33] Banerjee M, Chen T. Differential regulation of CYP3A4 promoter activity by a new class of natural product derivatives binding to pregnane X receptor. *Biochem Pharmacol* 2013;86:824–35. <https://doi.org/10.1016/j.bcp.2013.07.023>.
- [34] Sepe V, D'Amore C, Ummarino R, Renga B, D'Auria MV, Novellino E, et al. Insights on pregnane-X-receptor modulation. Natural and semisynthetic steroids from Theonella marine sponges. *Eur J Med Chem* 2014;73:126–34. <https://doi.org/10.1016/j.ejmech.2013.12.005>.
- [35] Kortagere S, Krasowski MD, Reschly EJ, Venkatesh M, Mani S, Ekins S. Evaluation of computational docking to identify pregnane X receptor agonists in the ToxCast database. *Environ Health Perspect* 2010;118:1412–7. <https://doi.org/10.1289/ehp.1001930>.
- [36] Ai N, Krasowski MD, Welsh WJ, Ekins S. Understanding nuclear receptors using computational methods. *Drug Discov Today* 2009;14:486–94. <https://doi.org/10.1016/j.drudis.2009.03.003>.
- [37] Pantsar T, Poso A. Binding Affinity via Docking: Fact and Fiction. *Molecules* 2018;23:E1899. <https://doi.org/10.3390/molecules23081899>.
- [38] Watanabe C, Watanabe H, Tanaka S. An interpretation of positional displacement of the helix12 in nuclear receptors: preexistent swing-up motion triggered by ligand binding. *Biochim Biophys Acta* 2010;1804:1832–40. <https://doi.org/10.1016/j.bbapap.2010.06.009>.
- [39] Chandran A, Vishveshwara S. Exploration of the conformational landscape in pregnane X receptor reveals a new binding pocket. *Protein Sci* 2016;25:1989–2005. <https://doi.org/10.1002/pro.3012>.
- [40] Motta S, Callea L, Giani Tagliabue S, Bonati L. Exploring the PXR ligand binding mechanism with advanced Molecular Dynamics methods. *Sci Rep* 2018;8:16207. <https://doi.org/10.1038/s41598-018-34373-z>.
- [41] Huber AD, Wright WC, Lin W, Majumder K, Low JA, Wu J, et al. Mutation of a single amino acid of pregnane X receptor switches an antagonist to agonist by altering AF-2 helix positioning. *Cell Mol Life Sci* 2021;78:317–35. <https://doi.org/10.1007/s00018-020-03505-y>.
- [42] Mustonen E-K, Pantsar T, Rashidian A, Reiner J, Schwab M, Laufer S, et al. Target Hopping from Protein Kinases to PXR: Identification of Small-Molecule Protein Kinase Inhibitors as Selective Modulators of Pregnane X Receptor from TüKIC Library. *Cells* 2022;11:1299. <https://doi.org/10.3390/cells11081299>.
- [43] Werner S, Mesch S, Hillig RC, Ter Laak A, Klint J, Neagoe I, et al. Discovery and Characterization of the Potent and Selective P2X4 Inhibitor N-[4-(3-Chlorophenoxy)-3-sulfamoylphenyl]-2-phenylacetamide (BAY-1797) and Structure-Guided Amelioration of Its CYP3A4 Induction Profile. *J Med Chem* 2019;62:11194–217. <https://doi.org/10.1021/acs.jmedchem.9b01304>.
- [44] Schuster D, Langer T. The identification of ligand features essential for PXR activation by pharmacophore modeling. *J Chem Inf Model* 2005;45:431–9. <https://doi.org/10.1021/ci049722q>.
- [45] Ngan C-H, Beglov D, Rudnitskaya AN, Kozakov D, Waxman DJ, Vajda S. The Structural Basis of Pregnane X Receptor Binding Promiscuity. *Biochemistry* 2009;48:11572–81. <https://doi.org/10.1021/bj901578n>.
- [46] Anami Y, Shimizu N, Ekimoto T, Egawa D, Itoh T, Ikeguchi M, et al. Apo- and Antagonist-Binding Structures of Vitamin D Receptor Ligand-Binding Domain Revealed by Hybrid Approach Combining Small-Angle X-ray Scattering and Molecular Dynamics. *J Med Chem* 2016;59:7888–900. <https://doi.org/10.1021/acs.jmedchem.6b00682>.
- [47] Shizu R, Nishiguchi H, Tashiro S, Sato T, Sugawara A, Kanno Y, et al. Helix 12 stabilization contributes to basal transcriptional activity of PXR. *J Biol Chem* 2021;297. <https://doi.org/10.1016/j.jbc.2021.100978>.
- [48] Teotico DG, Frazier ML, Ding F, Dokholyan NV, Temple BRS, Redinbo MR. Active Nuclear Receptors Exhibit Highly Correlated AF-2 Domain Motions. *PLoS Comput Biol* 2008;4:. <https://doi.org/10.1371/journal.pcbi.1000111>.
- [49] Chodera JD, Noé F. Markov state models of biomolecular conformational dynamics. *Curr Opin Struct Biol* 2014;25:135–44. <https://doi.org/10.1016/j.sbi.2014.04.002>.
- [50] Husic BE, Pande VS. Markov State Models: From an Art to a Science. *J Am Chem Soc* 2018;140:2386–96. <https://doi.org/10.1021/jacs.7b12191>.
- [51] Noble SM, Carnahan VE, Moore LB, Luntz T, Wang H, Ittoop OR, et al. Human PXR Forms a Tryptophan Zipper-Mediated Homodimer. *Biochemistry* 2006;45:8579–89. <https://doi.org/10.1021/bi0602821>.
- [52] Moore JT, Moore LB, Maglich JM, Klierer SA. Functional and structural comparison of PXR and CAR. *Biochim Biophys Acta* 2003;1619:235–8. [https://doi.org/10.1016/s0304-4165\(02\)00481-6](https://doi.org/10.1016/s0304-4165(02)00481-6).
- [53] Daffern N, Chen Z, Zhang Y, Pick L, Radhakrishnan I. Solution Nuclear Magnetic Resonance Studies of the Ligand-Binding Domain of an Orphan Nuclear Receptor Reveal a Dynamic Helix in the Ligand-Binding Pocket. *Biochemistry* 2018;57:1977–86. <https://doi.org/10.1021/acs.biochem.8b00069>.
- [54] Fischer A, Smieško M. Ligand Pathways in Nuclear Receptors. *J Chem Inf Model* 2019;59:3100–9. <https://doi.org/10.1021/acs.jcim.9b00360>.
- [55] Hillisch A, Gericke KM, Allerheiligen S, Roehrig S, Schaefer M, Tersteegen A, et al. Design, Synthesis, and Pharmacological Characterization of a Neutral, Non-Prodruge Thrombin Inhibitor with Good Oral Pharmacokinetics. *J Med Chem* 2020;63:12574–94. <https://doi.org/10.1021/acs.jmedchem.0c01035>.
- [56] Wang W, Prosser WW, Chen J, Taremi SS, Le HV, Madison V, et al. Construction and characterization of a fully active PXR/SRC-1 tethered protein with increased stability. *Protein Eng Des Sel* 2008;21:425–33. <https://doi.org/10.1093/protein/gzn017>.
- [57] Teotico DG, Bischof JJ, Peng L, Klierer SA, Redinbo MR. Structural basis of human pregnane X receptor activation by the hops constituent colupulone. *Mol Pharmacol* 2008;74:1512–20. <https://doi.org/10.1124/mol.108.050732>.
- [58] Di Masi A, De Marinis E, Ascenzi P, Marino M. Nuclear receptors CAR and PXR: molecular, functional, and biomedical aspects. *Mol Aspects Med* 2009;30:297–343. <https://doi.org/10.1016/j.mam.2009.04.002>.
- [59] Harder E, Damm W, Maple J, Wu C, Reboul M, Xiang JY, et al. OPLS3: A Force Field Providing Broad Coverage of Drug-Like Small Molecules and Proteins. *J Chem Theory Comput* 2016;12:281–96. <https://doi.org/10.1021/acs.jctc.5b00864>.
- [60] Roos K, Wu C, Damm W, Reboul M, Stevenson JM, Lu C, et al. OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *J Chem Theory Comput* 2019;15:1863–74. <https://doi.org/10.1021/acs.jctc.8b01026>.
- [61] Wallace BD, Betts L, Talmage G, Pollet RM, Holman NS, Redinbo MR. Structural and Functional Analysis of the Human Nuclear Xenobiotic Receptor PXR in Complex with RXR α . *J Mol Biol* 2013;425:2561–77. <https://doi.org/10.1016/j.jmb.2013.04.012>.
- [62] Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des* 2013;27:221–34. <https://doi.org/10.1007/s10822-013-9644-8>.
- [63] Jacobson MP, Friesner RA, Xiang Z, Honig B. On the Role of the Crystal Environment in Determining Protein Side-chain Conformations. *J Mol Biol* 2002;320:597–608. [https://doi.org/10.1016/s0022-2836\(02\)00470-9](https://doi.org/10.1016/s0022-2836(02)00470-9).
- [64] Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, et al. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem* 2004;47:1750–9. <https://doi.org/10.1021/jm030644s>.
- [65] Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* 2004;47:1739–49. <https://doi.org/10.1021/jm0306430>.

- [66] Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, et al. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein–Ligand Complexes. *J Med Chem* 2006;49:6177–96. <https://doi.org/10.1021/jm051256a>.
- [67] Bowers KJ, Chow DE, Xu H, Dror RO, Eastwood MP, Gregersen BA, et al. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. SC '06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing, 2006, p. 43–43. <https://doi.org/10.1109/SC.2006.54>.
- [68] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 1983;79:926–35. <https://doi.org/10.1063/1.445869>.
- [69] Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 2015;1–2:19–25. <https://doi.org/10.1016/j.softx.2015.06.001>.
- [70] Law SM. PyMol script: modevectors.py. Available at: <https://raw.githubusercontent.com/Pymol-Scripts/Pymol-script-repo/master/modevectors.py>
- [71] Scherer MK, Trendelkamp-Schroer B, Paul F, Pérez-Hernández G, Hoffmann M, Plattner N, et al. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J Chem Theory Comput* 2015;11:5525–42. <https://doi.org/10.1021/acs.jctc.5b00743>.
- [72] Pérez-Hernández G, Paul F, Giorgino T, De Fabritiis G, Noé F. Identification of slow molecular order parameters for Markov model construction. *J Chem Phys* 2013;139:. <https://doi.org/10.1063/1.4811489>015102.
- [73] Röblitz S, Weber M. Fuzzy spectral clustering by PCCA+: application to Markov state models and data classification. *Adv Data Anal Classif* 2013;7:147–79. <https://doi.org/10.1007/s11634-013-0134-6>.
- [74] Waskom ML. seaborn: statistical data visualization. *Journal of Open Source Software* 2021;6:3021. <https://doi.org/10.21105/joss.03021>.
- [75] Hunter JD. Matplotlib: A 2D Graphics Environment. *Computing in Science Engineering* 2007;9:90–5. <https://doi.org/10.1109/MCSE.2007.55>.
- [76] Geick A, Eichelbaum M, Burk O. Nuclear receptor response elements mediate induction of intestinal MDR1 by rifampin. *J Biol Chem* 2001;276:14581–7. <https://doi.org/10.1074/jbc.M010173200>.
- [77] Burk O, Kuzikov M, Kronenberger T, Jeske J, Keminer O, Thasler WE, et al. Identification of approved drugs as potent inhibitors of pregnane X receptor activation with differential receptor interaction profiles. *Arch Toxicol* 2018;92:1435–51. <https://doi.org/10.1007/s00204-018-2165-4>.
- [78] Piedade R, Traub S, Bitter A, Nüssler AK, Gil JP, Schwab M, et al. Carboxymefloquine, the Major Metabolite of the Antimalarial Drug Mefloquine, Induces Drug-Metabolizing Enzyme and Transporter Expression by Activation of Pregnane X Receptor. *Antimicrobial Agents and Chemotherapy* n.d.;59:96–104. <https://doi.org/10.1128/AAC.04140-14>.

SUPPLEMENTARY INFORMATION

Discrepancy in interactions and conformational dynamics of pregnane X receptor (PXR) bound to an agonist and a novel competitive antagonist

Azam Rashidian ^a, Enni-Kaisa Mustonen ^b, Thales Kronenberger ^{a,c,e,f}, Matthias Schwab ^{b,d,e}, Oliver Burk ^b, Stefan A. Laufer ^{c,e,f}, Tatu Pantsar* ^{c,g}

^a Department of Internal Medicine VIII, University Hospital Tuebingen, Otfried-Müller-Strasse 14, Tuebingen DE 72076, Germany.

^b Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart and University of Tuebingen, Tuebingen, Germany

^c Department of Pharmaceutical and Medicinal Chemistry, Institute of Pharmaceutical Sciences, Eberhard-Karls-Universität, Tuebingen, Auf der Morgenstelle 8, 72076 Tuebingen, Germany.

^d Departments of Clinical Pharmacology, and Pharmacy and Biochemistry, University of Tuebingen, Tuebingen, Germany

^e Cluster of Excellence iFIT (EXC 2180) “Image-Guided and Functionally Instructed Tumor Therapies”, University of Tuebingen, 72076 Tuebingen, Germany

^f Tuebingen Center for Academic Drug Discovery & Development (TüCAD2), 72076 Tuebingen, Germany

^g School of Pharmacy, Faculty of Health Sciences, University of Eastern Finland, Yliopistonranta 1, 70210 Kuopio, Finland.

*Correspondence: <mailto:tatu.pantsar@uef.fi>

Phone: +358 50 388 1062

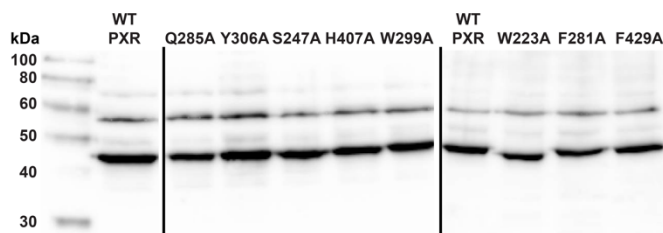
Content

Supplementary Methods	4
CaPO ₄ transfections and protein analysis of mutants.	
Supplementary Figure S1	5
Structures of human PXR agonist SJB7 and antagonist SPA70.	
Supplementary Figure S2	5
PC1 and PC2 comparison between systems.	
Supplementary Table S1	6
The contributions of individual principal components PC1–PC10.	
Supplementary Table S2	6
Root-mean-square fluctuations (RMSFs) of individual residues.	
Supplementary Figure S3	7
Comparison of B-factors from crystal structures to RMSF-values.	
Supplementary Figure S4	7
Additional water-bridged interactions in N-terminus of the α 6 region.	
Supplementary Figure S5	8
Root-mean-square fluctuations (RMSFs) of ligand heavy atoms.	
Supplementary Figure S6	9
Distances of Y306– α 3 and F288– α 3.	
Supplementary Figure S7	9
Hydrophobic interactions of the ligands beyond the hydrophobic subpocket and hydrophobic interaction of F420 with α 10/ α 11.	
Supplementary Figure S8	10
Destabilization of α AF-2 in the presence of SRC-1 with compound 100.	
Supplementary Figure S9	11
The time-lagged independent component analysis (TICA).	
Supplementary Figure S10	11
PXR-LBD conformation in an agonist bound crystal structure (PDB ID: 1NRL).	
Supplementary Figure S11	12
Distances associated to PXR conformation of β 1- β 1' loop and β 4- α 6 loop regions.	

Supplementary Figure S12	12
Hydrophobic interactions of the BAY-1797 beyond the hydrophobic subpocket.	
Supplementary Table S3	13
Overview of the simulated systems.	
Supplementary Figure S13	13
Outline of the conducted simulations.	
Supplementary Figure S14	14
MSM validation.	
References	15

Supplementary Methods — *CaPO₄* transfections and protein analysis of mutants

5x10⁶ HepG2 cells were seeded a day before transfection in a 10 cm diameter dish. 1 hour prior transfection, culture medium was changed. Aqueous DNA solution of 450 µl was prepared, consisting of 2 µg pMetLuc2control plasmid and 10 µg expression plasmids encoding human PXR or PXR mutants. Total amount of DNA was adjusted to 25 µg with pUC18. 50 µl of CaCl₂ was added to each DNA solution, then mixture was added dropwise to an equal volume of 2x HBS while simultaneously mixing before adding dropwise onto cells. Cells were incubated for 5 h before 3 min incubation with 15% glycerol-PBS. Cells were incubated for two days before Metridia luciferase measurement as described in Transient transfections and total protein extraction. Cells were washed with ice-cold PBS and scraped before centrifuging for 5 min at 750 g at 4 °C. Supernatant was discarded and cell pellet resuspended in protein lysis buffer [1] and incubated on ice for 15 min. Lysate was homogenized with ultrasonication (2x30s) using Bioruptor UCD200 (Diagenode, Liège, Belgium). Protein concentration was determined with bicinchoninic acid method. Protein amount was adjusted to the transfection efficiency as determined by measurement of Metridia luciferase activity. Samples were analyzed on 10% SDS-polyacrylamide protein gels, which was followed by Western blotting to nitrocellulose membrane. Blots were incubated overnight at 4 °C with primary antibodies against human PXR (2 µg/ml; RRID AB_2155076; clone H4417, Perseus Proteomics, Tokyo, Japan) followed by incubation for 1 h at room temperature with peroxidase-conjugated secondary rabbit anti-mouse antibody (0.13 µg/ml; P0260, Dako, Glostrup, Denmark). Detection of chemiluminescence was conducted as described previously [2].



Supplementary Methods Fig. 1. Transfected PXR mutants are expressed in HepG2 cells.

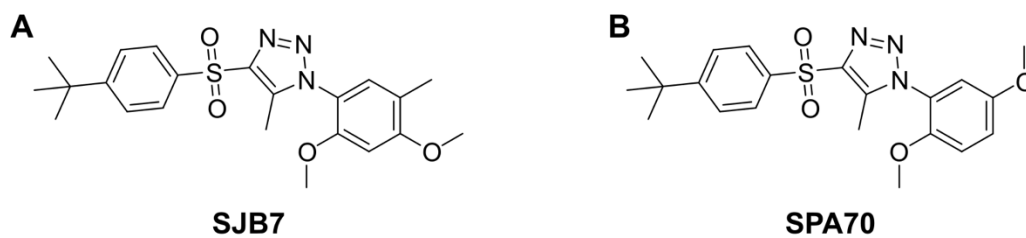


Fig. S1. Structures of human PXR agonist SJB7 (A) and antagonist SPA70 (B).

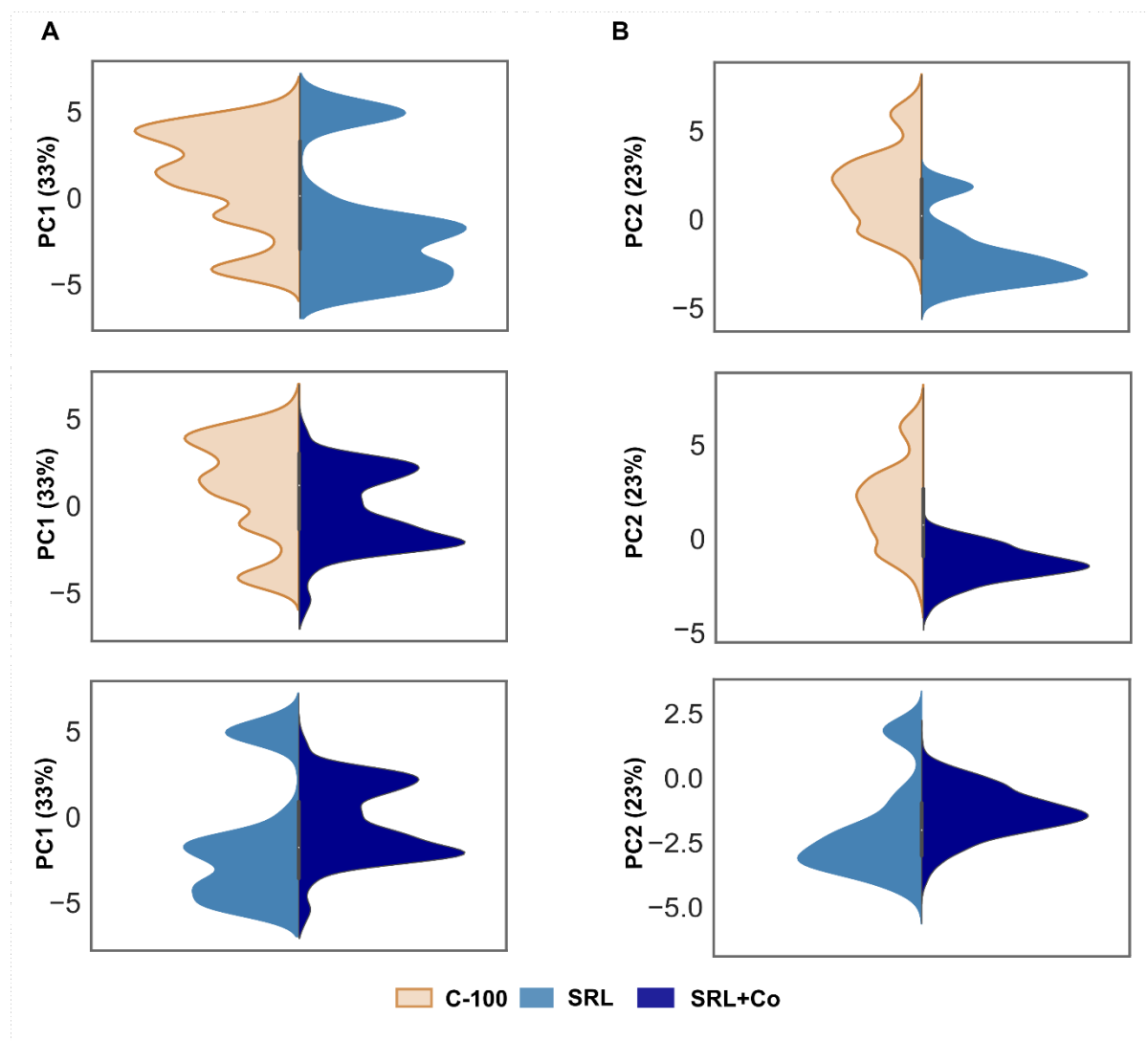


Fig. S2. PC1 and PC2 comparison between systems. In violin plots, a kernel density estimation is used to display the distribution of the data for PC1 (A) and PC2 (B) in each system. Systems are coloured as following: compound 100 (C-100), light orange; SR12813 (SRL), light blue; SR12813 and SRC1 coactivator peptide (SRL+Co), dark blue.

Table S1. The contributions of individual principal components PC1–PC10.

PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
33%	23%	9.6%	9.0%	6.0%	5.2%	4.3%	3.6%	3.5%	2.8%

Table S2. Root-mean-square fluctuations (RMSFs) of individual residues (backbone atoms) located on regions associated with the PCA extreme movements. The regions are highlighted with colours according to the colours in Fig. 2C.

RMSF (Å)							
Residue	SRL+Co	SRL	C-100	Residue	SRL+Co	SRL	C-100
$\alpha 1-\alpha 2'$ loop				$\beta 1'-\alpha 3$ loop			
V177	2.95	2.87	1.79	K226	1.40	1.19	1.04
L178	3.07	3.18	1.78	P227	1.60	1.43	1.16
S179	3.07	3.67	1.98	P228	1.80	1.75	1.46
S180	3.31	3.80	2.21	A229	2.00	2.19	1.67
G181	3.04	3.95	2.40	D230	2.52	2.80	2.08
C182	2.85	4.16	2.16	S231	2.58	2.87	2.02
E183	3.12	3.96	2.09	G232	2.44	3.07	1.93
L184	3.06	3.31	2.09	G233	1.80	2.53	1.43
P185	3.70	3.30	2.31	N234	1.35	1.70	1.04
E186	3.60	3.28	2.61	$\beta 4-\alpha 6$ loop			
S187	3.84	3.90	2.78	D310	1.59	1.57	1.31
L188	3.89	3.82	2.92	T311	2.22	1.93	1.72
Q189	4.02	3.66	3.03	A312	2.94	2.42	2.20
A190	3.52	3.37	2.75	G313	2.99	2.51	2.26
P191	3.33	3.17	2.92	G314	2.66	2.12	1.89
S192	2.72	2.26	2.24	$\alpha AF-2$			
R193	2.04	1.52	1.91	P423	0.93	1.68	1.69
E194	1.82	1.44	1.89	L424	0.80	1.54	1.58
E195	1.68	1.36	1.80	M425	0.77	1.04	1.34
A196	1.18	0.98	1.42	Q426	0.89	1.17	1.41
$\beta 1-\beta 1'$ loop				E427	0.91	1.28	1.42
G217	1.99	1.96	1.97	L428	0.96	1.04	1.38
E218	2.93	2.70	2.61	F429	1.11	1.28	1.76
D219	3.31	3.03	3.18	G430	1.39	1.81	2.09
G220	2.98	2.93	2.81	I431	1.88	2.02	2.41
S221	2.29	2.41	2.23	T432	2.82	2.94	3.24
				G433	4.53	4.69	4.31
				S434	6.12	6.92	5.75

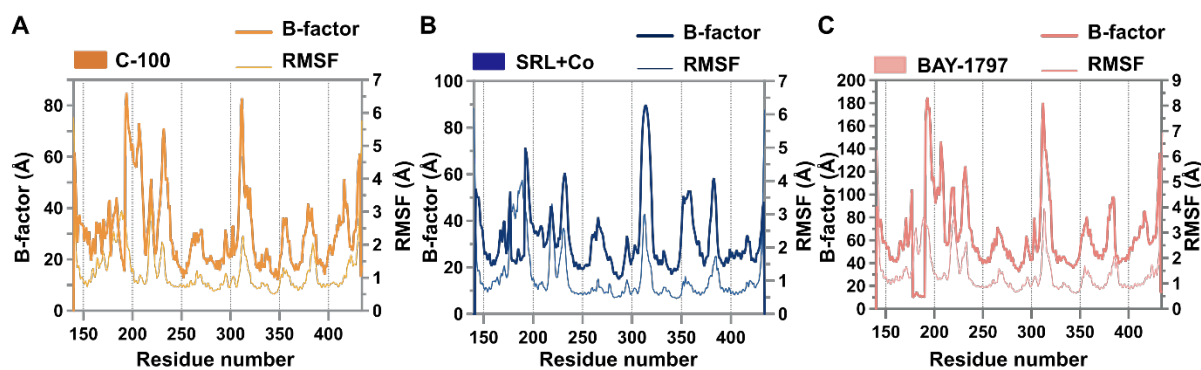


Fig. S3. Comparison of B-factors from crystal structures to RMSF-values. (A) B-factor of PXR crystal structure (PDB ID: 4J5W) aligned with the protein RMSF in C-100 system. (B) B-factor of PXR crystal structure (PDB ID: 1NRL) aligned with the protein RMSF in SRL+Co system. (C) B-factor of PXR crystal structure (PDB ID: 6HTY) aligned with the protein RMSF in BAY-1797 system.

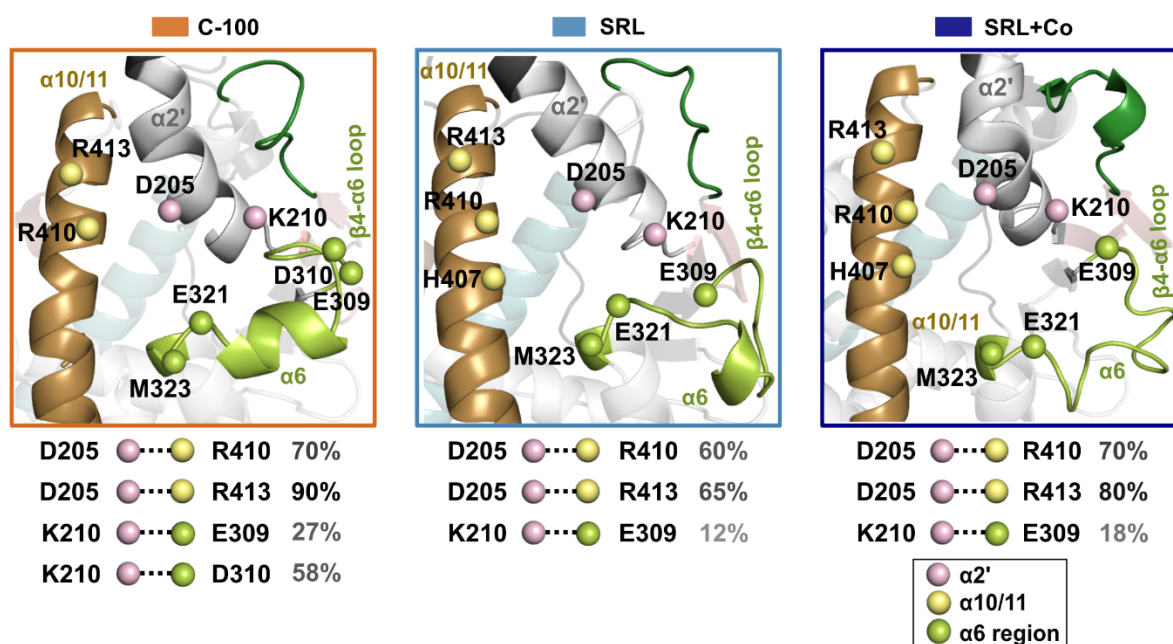


Fig. S4. Additional water-bridged interactions in N-terminus of the $\alpha 6$ region. The additional water bridge interactions in the $\alpha 6$ region (in addition to the observed interactions displayed in Fig. 3B) and their frequencies (cut-off of $\geq 10\%$) are shown for each system. The location of the C α -atom of each interacting residue are indicated with spheres. Spheres belonging to the same helix/region are illustrated with the same colour. The colours of the helices are as in Fig. 1A (main manuscript).

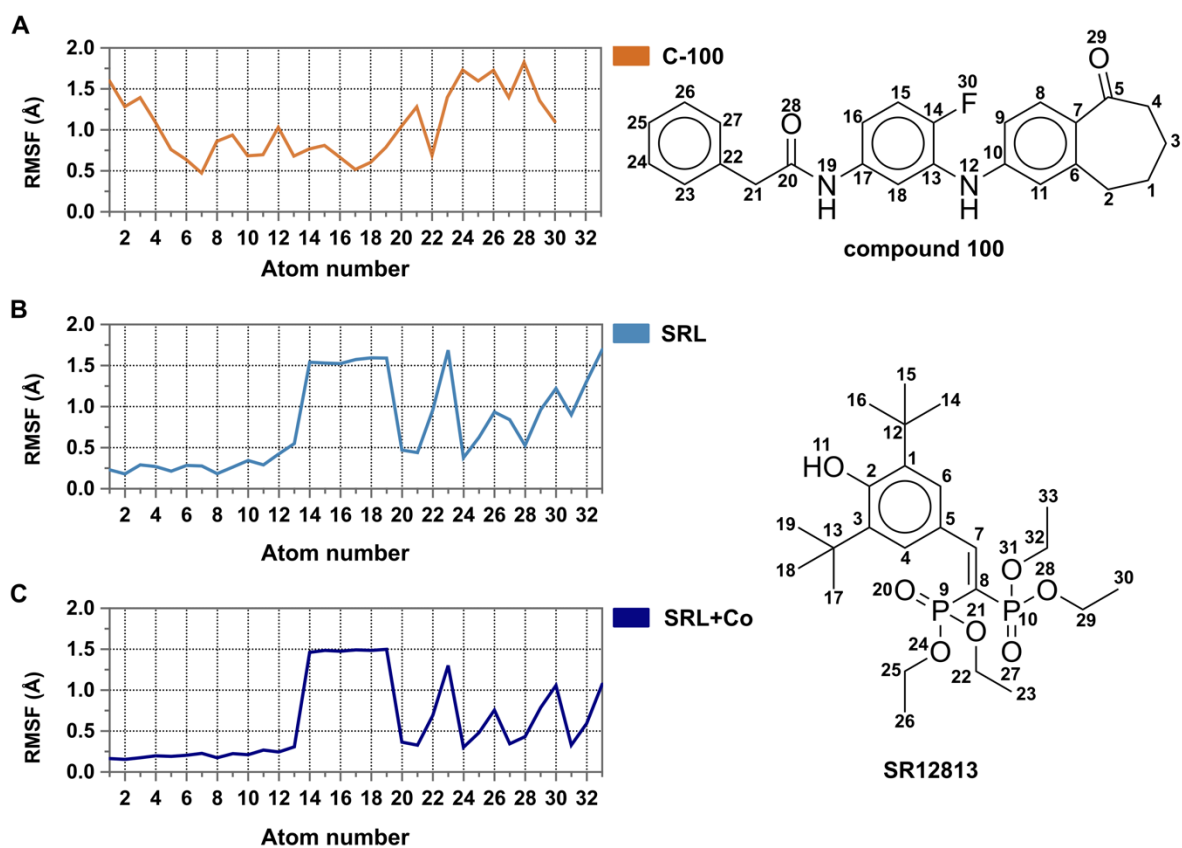


Fig. S5. Root-mean-square fluctuations (RMSFs) of ligand heavy atoms. (A) RMSF of compound 100 (C-100), (B) SR12813 (SRL) and (C) SR12813 with the coactivator (SRL+Co). The atom numbers shown in the plots are illustrated in the 2D structures on the right. Overall, the heavy atoms display low RMSF values (mainly $<1\text{\AA}$; flexible substituents $<1.7\text{\AA}$), demonstrating the stability of the ligands during the simulations.

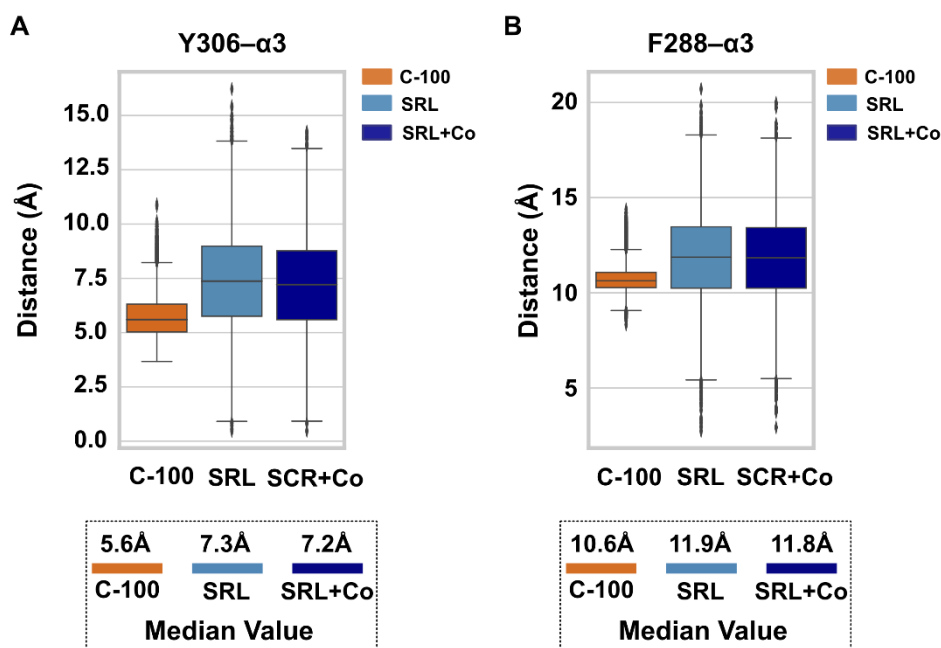


Fig. S6. Distances of Y306- α 3 and F288- α 3. (A) Distance between the centre of mass of α 3 (residues 240–260) and the C α -atom of Y306 (located in β 4). (B) Distance between the centre of mass of α 3 (residues 240–260) and the C α -atom of F288 (located in α 5).

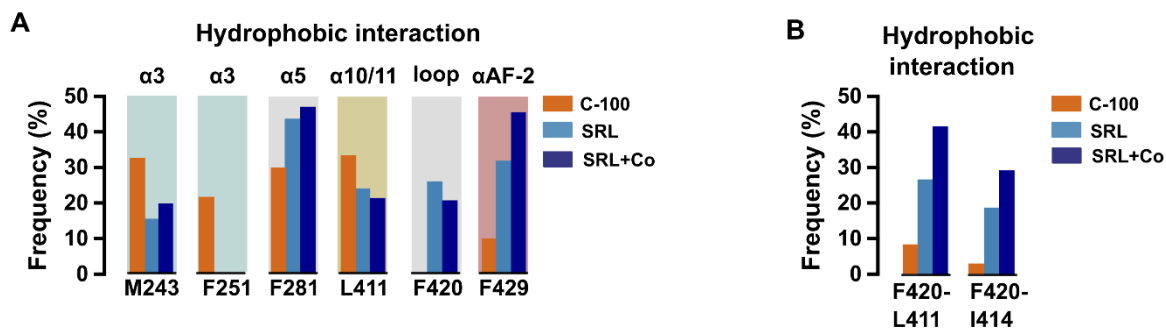


Fig. S7. Hydrophobic interactions of the ligands beyond the hydrophobic subpocket and hydrophobic interaction of F420 with α 10/ α 11. (A) Hydrophobic interactions of the ligands beyond the hydrophobic subpocket. F420 is located on the loop connecting α 11 to α AF-2. The boxes are coloured as in PXR-LBD crystal structure in Fig. 1A. (B) Hydrophobic interaction of F420 with α 10/ α 11.

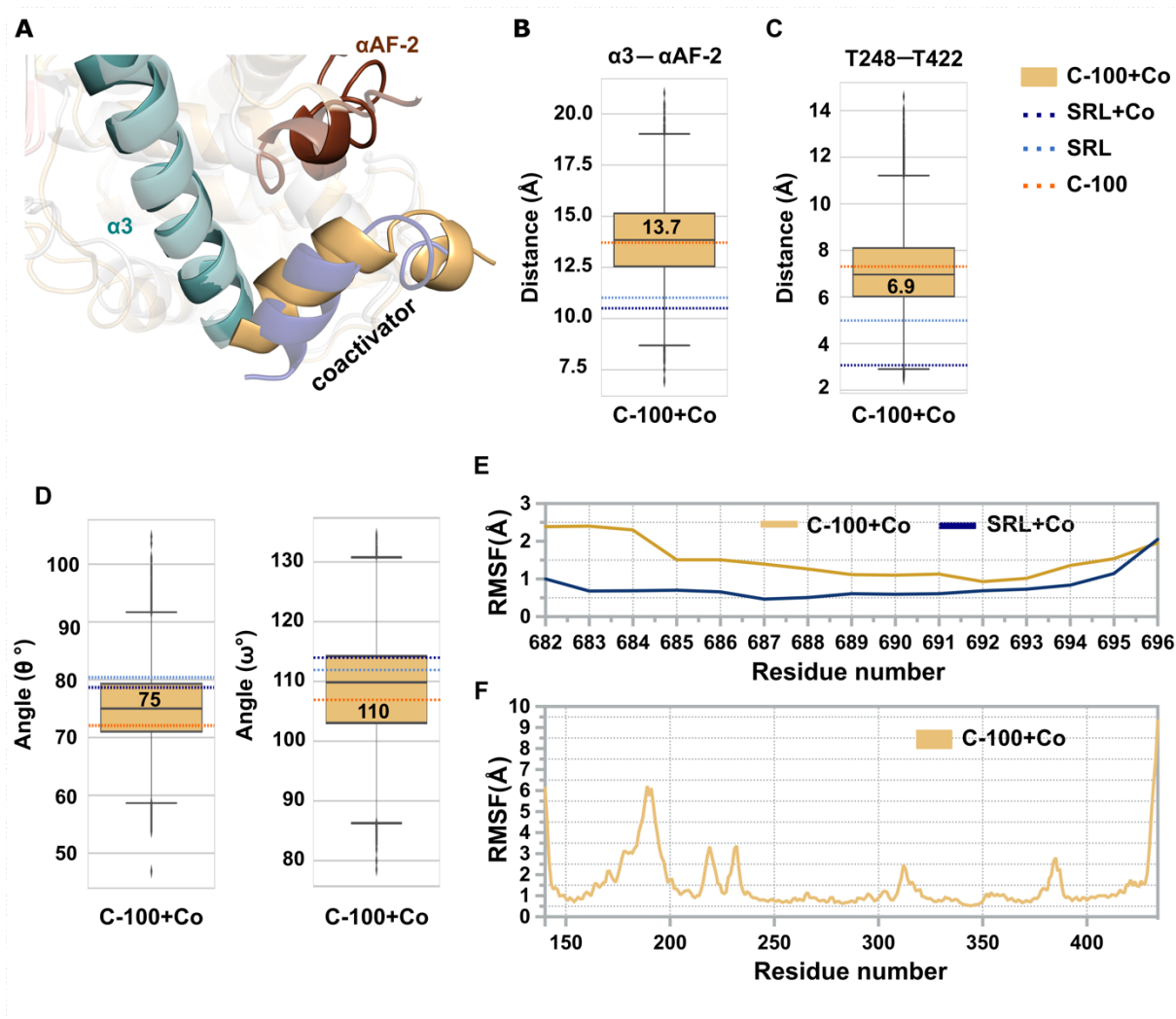


Fig. S8. Destabilization of α AF-2 in the presence of SRC-1 with compound 100. (A) Superposition of representative snapshots from C-100+Co over the SRL+Co system. The non-transparent structure is C-100+Co, and transparent colours represent SRL+Co system. (B) Movement of α AF-2 in C-100+Co resembles C-100 system. Box plot represents the distribution of distance between centre of mass of α 3-helix (residues 240–260) and α AF-2 (residues 423–430). The values in the boxes represent the median of C-100+Co. The dashed lines represent the median value of the of C-100 system (orange dashed line); SRL system (light blue dashed line); SRL+Co system (dark blue dashed line) (see also Fig. 5). (C) Distance between sidechain oxygen atoms of T248 and T422 are shown in box plot. (D) θ and ω define the angle consisting of N404, F281, F429 and N404, F281, T422, respectively where F281 is the apex. Box plots display the distribution of θ and ω angles. (E) Root-mean-square fluctuations (RMSFs) of the SRC-1 backbone atoms display higher values in C-100+Co, especially in its N-terminal part, highlighting its instability with compound 100. (F) Root-mean-square fluctuation of the PXR-LBD backbone in C-100+Co.

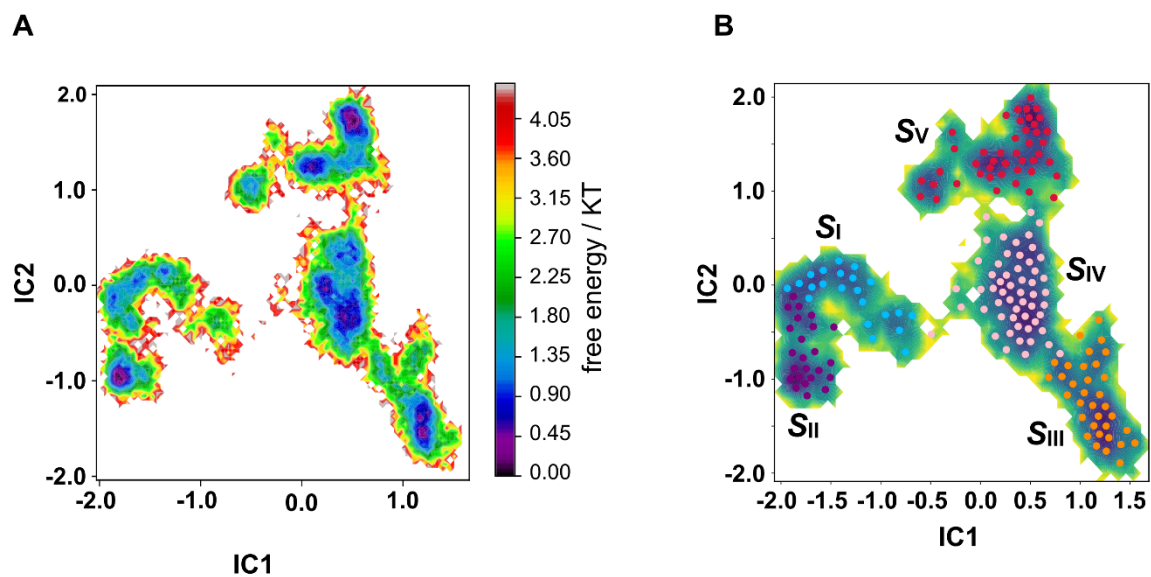


Fig. S9. The time-lagged independent component analysis (TICA). (A) Pseudo free energy map of distribution in C-100 projected on TICA components 1 (IC1) and 2 (IC2). (B) Separation of the five metastable states by PCCA++. Each individually coloured cluster corresponds to one metastable state (S_{I-V}). The colour code for each cluster is as in Fig. 6.

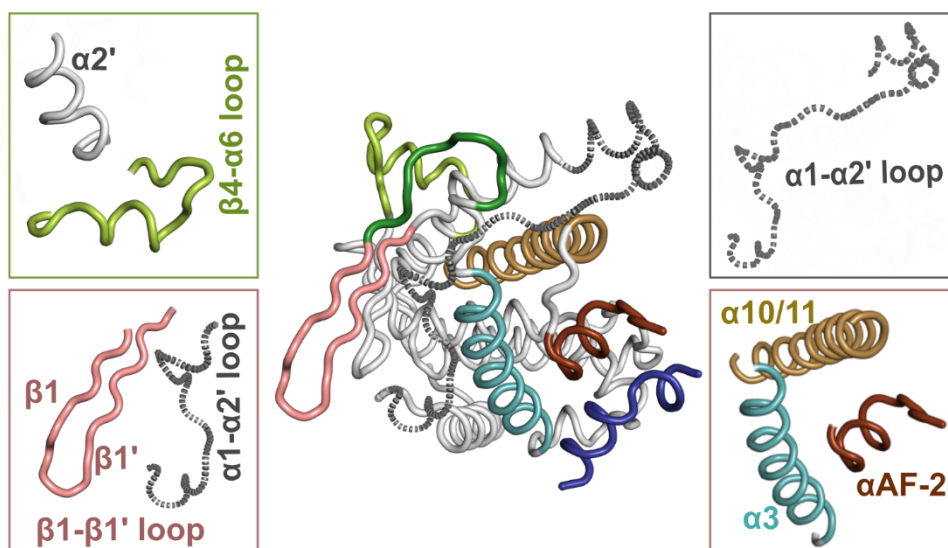


Fig. S10. PXR-LBD conformation in an agonist bound crystal structure (PDB ID: 1NRL). Individual substructures are illustrated as follows: $\alpha1-\alpha2'$ loop, grey dashed line; $\beta-\beta1'$ loop, salmon; $\beta4-\alpha6$ loop, light green; $\beta1'-\alpha3$ loop, dark green; $\alpha-AF-2$, dark brown.

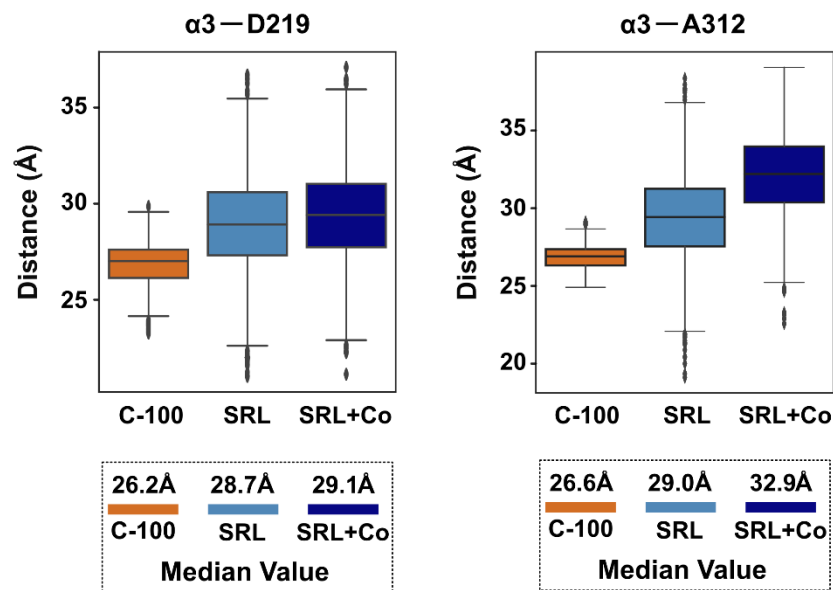


Fig. S11. Distances associated to PXR conformation of $\beta 1$ - $\beta 1'$ loop and $\beta 4$ - $\alpha 6$ loop regions. (A) The distance between D219 (apex $\beta 1$ - $\beta 1'$ loop) and $\alpha 3$ (centre of mass of the residues 240–260). This distance is associated the conformational preference in the $\beta 1$ - $\beta 1'$ loop as $\alpha 3$ -helix is stable in the simulations: longer distances indicate more extended configuration of this loop, more far away from $\alpha 3$ -helix. (B) The distance between A312 (located on $\beta 4$ - $\alpha 6$ loop) and $\alpha 3$ (centre of mass of the residues 240–260). This distance is associated the conformational preference in the in $\beta 4$ - $\alpha 6$ loop.

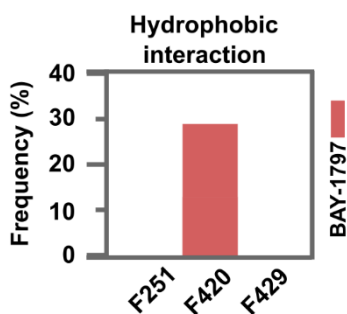


Fig. S12. Hydrophobic interactions of BAY-1797 beyond the hydrophobic subpocket. F420 is located on the loop connecting $\alpha 11$ to $\alpha AF-2$.

Table S3. Overview of the simulated systems.

PDB ID	Compound	Co-activator	MD simulation time
1NRL	SR12813	SRC-1	10 μ s
1NRL	SR12813	-	20 μ s
4J5W	100	-	30 μ s
4J5W	100	SRC-1	10 μ s
6HTY	BAY-1797	-	10 μ s

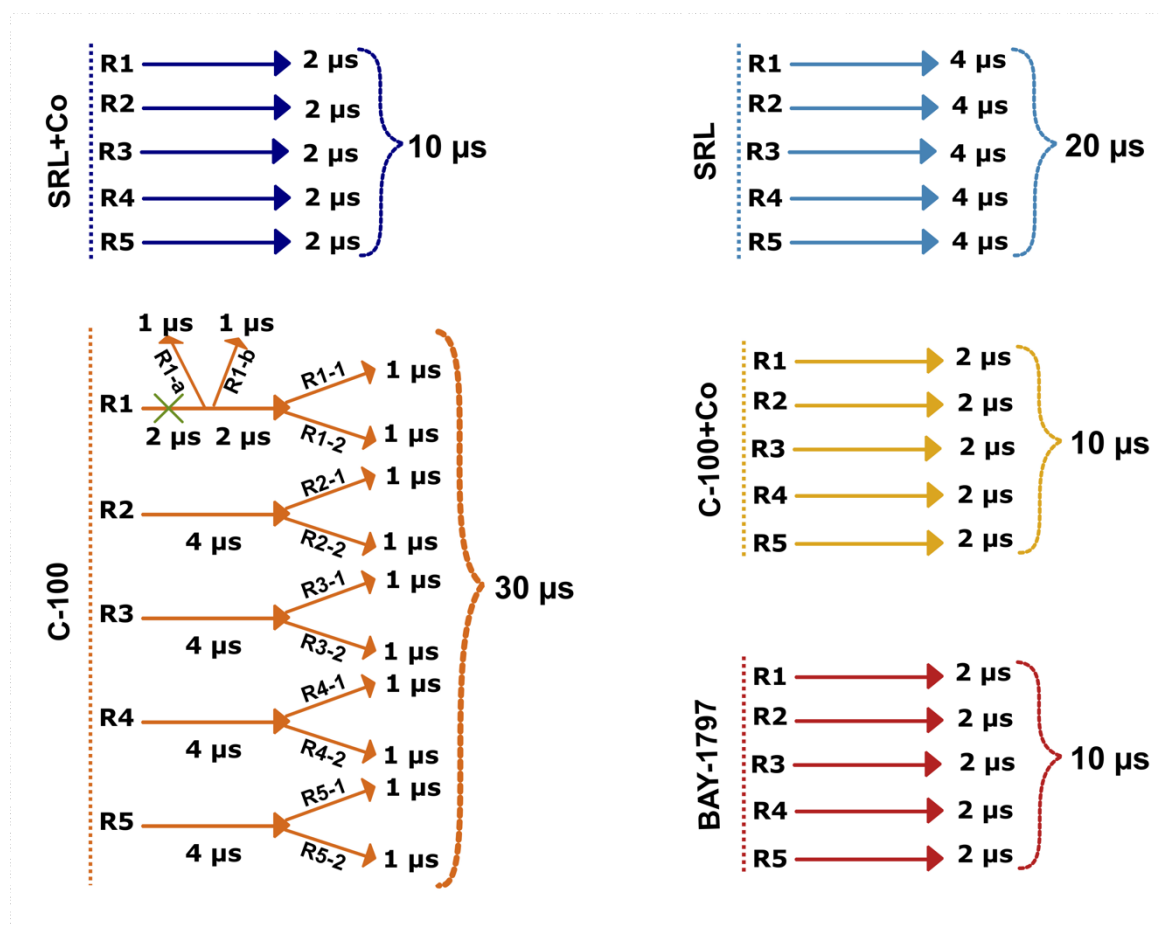


Fig. S13. Outline of the conducted simulations. Fig. S11. Outline of the conducted simulations. This figure illustrates the replicas (R stands for replica) generated and run for each system. In total, seventeen individual replicas were run for C-100 and five replicas for the other systems. Lack of ligand stability in the beginning of the R1 of C-100 led to the generation of more replicas from this trajectory, and the first 2 μ s of R1 was excluded from the analysis (denoted here by the green cross sign). Moreover, to obtain additional sampling for C-100,

additional two 1 μ s replicas were generated from the output conformations of the original replicas for C-100.

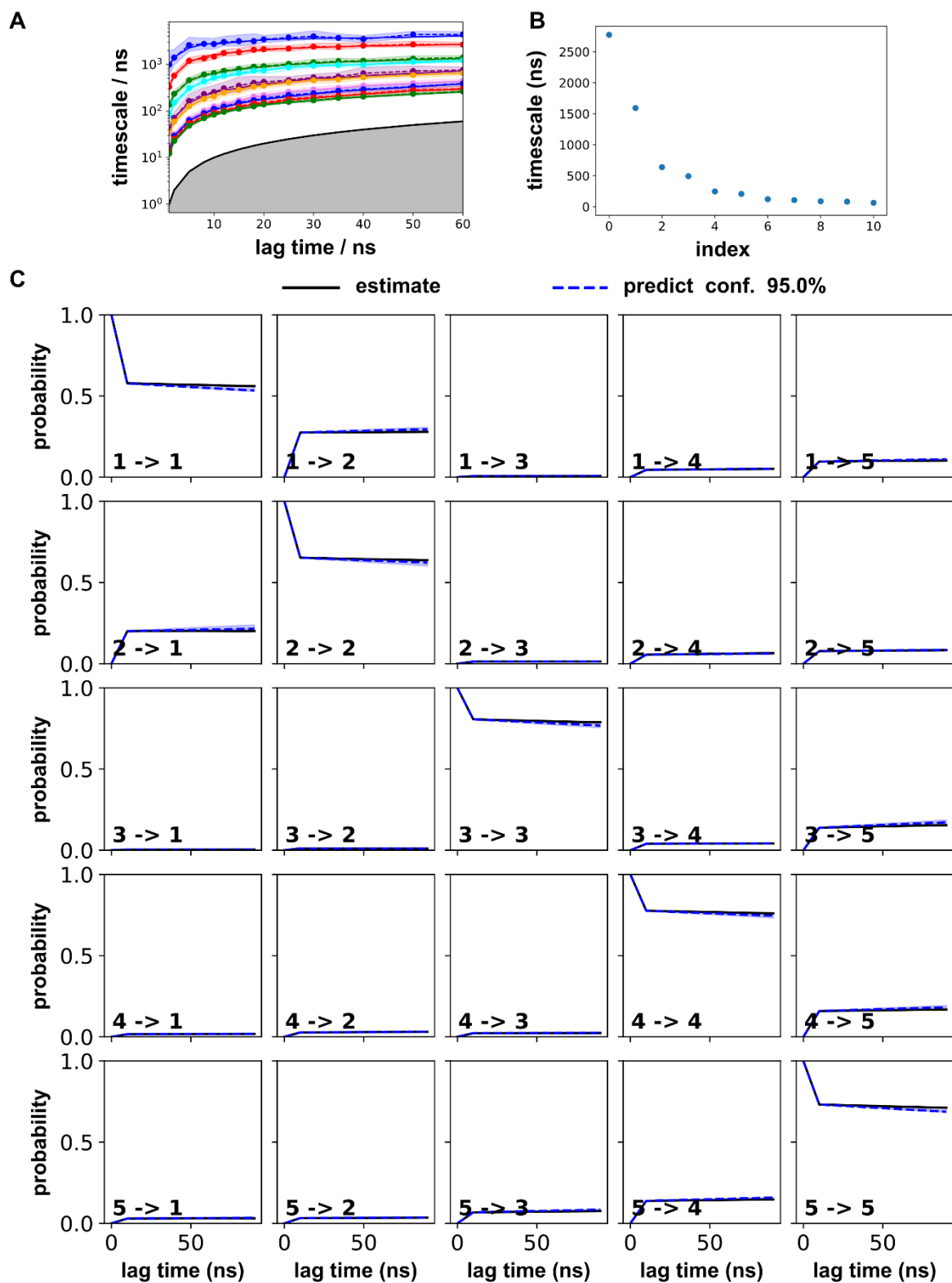


Fig. S14. MSM validation. (A) The implied timescales are converged at the used lag time of 10 ns. (B) The spectral analysis of the timescale separation. A five-state model was selected. (C) The Chapman-Kolmogorov test of the MSM shows that MSM (black line) follows the observed trajectory (blue dashed line, with shaded error estimate).

References

- [1] Sundqvist A, Bengoechea-Alonso MT, Ye X, Lukiyanchuk V, Jin J, Harper JW, et al. Control of lipid metabolism by phosphorylation-dependent degradation of the SREBP family of transcription factors by SCF(Fbw7). *Cell Metab* 2005;1:379–91. <https://doi.org/10.1016/j.cmet.2005.04.010>.
- [2] Jeske J, Bitter A, Thasler WE, Weiss TS, Schwab M, Burk O. Ligand-dependent and -independent regulation of human hepatic sphingomyelin phosphodiesterase acid-like 3A expression by pregnane X receptor and crosstalk with liver X receptor. *Biochem Pharmacol* 2017;136:122–35. <https://doi.org/10.1016/j.bcp.2017.04.013>.

8. Publication II

Target Hopping from Protein Kinases to PXR: Identification of Small-Molecule Protein Kinase Inhibitors as Selective Modulators of Pregnane X Receptor from TüKIC Library

Enni-Kaisa Mustonen ¹, Tatu Pantsar ^{2,3}, Azam Rashidian ⁴, Juliander Reiner ², Matthias Schwab ^{1 5 6}, Stefan Laufer ^{2 6 7}, Oliver Burk ¹

¹Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, and University of Tuebingen, 72074 Tuebingen, Germany.

²Department of Pharmaceutical and Medicinal Chemistry, Institute of Pharmaceutical Sciences, University of Tuebingen, 72076 Tuebingen, Germany.

³School of Pharmacy, Faculty of Health Sciences, University of Eastern Finland, 70210 Kuopio, Finland.

⁴Department of Internal Medicine VIII, University Hospital Tuebingen, 72076 Tuebingen, Germany.

⁵Departments of Clinical Pharmacology and Biochemistry and Pharmacy, University of Tuebingen, 72076 Tuebingen, Germany.

⁶Cluster of Excellence iFIT (EXC 2180) "Image-Guided and Functionally Instructed Tumor Therapies", University of Tuebingen, 72076 Tuebingen, Germany.

⁷Tuebingen Center for Academic Drug Discovery & Development (TüCAD2), 72076 Tuebingen, Germany.

Cells journal

2022;11(8):1299. doi: 10.3390/cells11081299.

Hypothesis: The simultaneous inhibition of PXR and protein kinases by drugs with dual function could represent a promising opportunity to explore novel pathways in cancer treatment.

Aims: To identify small-molecule kinase inhibitors of the Tübingen Kinase Inhibitor Collection (TÜKIC) compound library that would act also as PXR antagonists to conquer the challenges posed by drug resistance in cancer.

Results: In our study, we found four new PXR inhibitors and a closely related full agonist from the in-house TÜKIC compound library. Further analyses confirmed that the most potent inhibitors, 73 and 100, act as PXR ligands. These compounds displayed passive, mixed competitive/noncompetitive antagonism and showed expression of PXR target genes in gene- and tissue- specific manner which makes them selective PXR modulators. By observing the concentration-response curve of the PXR agonist rifampicin, we noticed a dextral shift as we introduced higher concentrations of inhibitors. This observation strongly indicated that the novel compounds can be classified as competitive antagonists. Nevertheless, as the maximum effect was decreasing, it becomes apparent that there is also a contribution of a noncompetitive/allosteric component in the antagonism. Using mammalian two hybrid assays, compound 73 and 100 demonstrates disruption of PXR's coactivators and corepressors interactions. Hence, they can be identified as passive antagonists. On the other hand, rifampicin, and compound 109 induced the interaction of steroid receptor coactivator 1 (SRC1). The result of in vitro limited proteolytic digestion assay indicates that the novel compounds, 73, 100 and 109 are PXR ligands. Further analysis of compounds 100 and 109 by competitive radioligand binding assays demonstrates the inhibition of H-labelled SR12813 (PXR's full agonist) to PXR (data shown in the respective publication). MD simulations suggest the putative binding configuration of compound 100 in LBP which supports the limited proteolytic digestion assay and competitive radioligand binding assay results. While compound 100 did not affect constitutive activity of the LBP-filled PXR triple mutant S208W/S247W/C284W, compound 73 still reduced the respective activity by 40%. This evidence can be attributed to allosteric binding of compound 73. This observation further supports the suggested noncompetitive antagonism for 73. In contrast, the essay did not support the allosteric binding of compound 100 to PXR. Regarding nuclear receptor selectivity of our novel compounds, 109 did not display significant activation of CAR1 but weakly

activated CAR3 and VDR. Both compounds 73 and 100 exhibited a slight inhibitory effect on the constitutive activity of isoform CAR1, resulting in a 30% reduction. Compound 73 demonstrated additional inhibition on the ligand-induced activation of CAR3 by CITCO, leading to a decrease of 46%. It also reduced ligand-induced VDR activation by 51%. Neither CAR3 nor VDR were activated by 73 or 100 on their own. To study the effects of the novel PXR ligands on the expression of endogenous PXR target genes, LS174T colorectal carcinoma cells were chosen due to their comparable PXR expression levels to that of liver cells. We observed the induction of ABCB1 and CYP3A4 gene expression with rifampicin. However, compounds 73 and 100 did not induce the expression of ABCB1. Moreover, these compounds demonstrated the ability to suppress the rifampicin-mediated induction of ABCB1. Despite being classified as PXR antagonists, both compounds induced CYP3A4 expression to a level comparable to that induced by rifampicin. On the other hand, compound 109, as expected from its agonist properties, induced the expression of both genes, with ABCB1 being induced to a lesser extent than observed with rifampicin. Upon examination, compound 100 was found to induce the expression of ABCB1, CYP2B6, and CYP3A4 in HepaRG cells. However, it's worth noting that the induction of these three genes was noticeably milder compared to the response elicited by the rifampicin. Unlike the observations in LS174T cells, compound 100 displayed only minor inhibitory effects on the rifampicin-induced expression of ABCB1 in HepaRG cells. These data suggest that compound 100 might exhibit partial agonist activity in hepatic cells. **Conclusion:** exploration of our in-house kinase inhibitor compound library identified novel selective receptor modulators of PXR. These exceptional compounds bear a common phenylaminobenzosuberone scaffold, represents a previously unknown PXR ligand structure which could serves as a potential foundation for the synthesis of innovative dual PXR and protein kinase inhibitors.

Article

Target Hopping from Protein Kinases to PXR: Identification of Small-Molecule Protein Kinase Inhibitors as Selective Modulators of Pregnane X Receptor from TükIC Library

Enni-Kaisa Mustonen¹, Tatu Pantsar^{2,3} , Azam Rashidian⁴ , Juliander Reiner², Matthias Schwab^{1,5,6}, Stefan Laufer^{2,6,7}  and Oliver Burk^{1,*} 

- ¹ Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, and University of Tuebingen, 72074 Tuebingen, Germany; ennikaisamustonen@gmail.com (E.-K.M.); matthias.schwab@ikp-stuttgart.de (M.S.)
- ² Department of Pharmaceutical and Medicinal Chemistry, Institute of Pharmaceutical Sciences, University of Tuebingen, 72076 Tuebingen, Germany; tatu.pantsar@uef.fi (T.P.); juliander.reiner@uni-tuebingen.de (J.R.); stefan.laufer@uni-tuebingen.de (S.L.)
- ³ School of Pharmacy, Faculty of Health Sciences, University of Eastern Finland, 70210 Kuopio, Finland
- ⁴ Department of Internal Medicine VIII, University Hospital Tuebingen, 72076 Tuebingen, Germany; azam.rashidian@uni-tuebingen.de
- ⁵ Departments of Clinical Pharmacology and Biochemistry and Pharmacy, University of Tuebingen, 72076 Tuebingen, Germany
- ⁶ Cluster of Excellence iFIT (EXC 2180) “Image-Guided and Functionally Instructed Tumor Therapies”, University of Tuebingen, 72076 Tuebingen, Germany
- ⁷ Tuebingen Center for Academic Drug Discovery & Development (TüCAD2), 72076 Tuebingen, Germany
- * Correspondence: oliver.burk@ikp-stuttgart.de; Tel.: +49-711-8101-5091



Citation: Mustonen, E.-K.; Pantsar, T.; Rashidian, A.; Reiner, J.; Schwab, M.; Laufer, S.; Burk, O. Target Hopping from Protein Kinases to PXR: Identification of Small-Molecule Protein Kinase Inhibitors as Selective Modulators of Pregnane X Receptor from TükIC Library. *Cells* **2022**, *11*, 1299. <https://doi.org/10.3390/cells11081299>

Academic Editor: Hiroshi Miyamoto

Received: 18 March 2022

Accepted: 7 April 2022

Published: 12 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Small-molecule protein kinase inhibitors are used for the treatment of cancer, but off-target effects hinder their clinical use. Especially off-target activation of the pregnane X receptor (PXR) has to be considered, as it not only governs drug metabolism and elimination, but also can promote tumor growth and cancer drug resistance. Consequently, PXR antagonism has been proposed for improving cancer drug therapy. Here we aimed to identify small-molecule kinase inhibitors of the Tübingen Kinase Inhibitor Collection (TükIC) compound library that would act also as PXR antagonists. By a combination of in silico screen and confirmatory cellular reporter gene assays, we identified four novel PXR antagonists and a structurally related agonist with a common phenylaminobenzosuberone scaffold. Further characterization using biochemical ligand binding and cellular protein interaction assays classified the novel compounds as mixed competitive/noncompetitive, passive antagonists, which bind PXR directly and disrupt its interaction with coregulatory proteins. Expression analysis of prototypical PXR target genes ABCB1 and CYP3A4 in LS174T colorectal cancer cells and HepaRG hepatocytes revealed novel antagonists as selective receptor modulators, which showed gene- and tissue-specific effects. These results demonstrate the possibility of dual PXR and protein kinase inhibitors, which might represent added value in cancer therapy.

Keywords: pregnane X receptor; protein kinase inhibitor; PXR antagonist; cancer therapy

1. Introduction

Signaling cascades of protein kinases regulate multiple functions in cells, such as cell proliferation and survival. Their dysregulation frequently results in loss of apoptotic cell death and uncontrolled cell growth, which are among the hallmarks of cancer [1]. Thus, inhibition of protein kinases has proven to be an effective therapeutic strategy in cancer treatment; consequently, small-molecule kinase inhibitors have become a clinically important group of molecularly targeted anticancer drugs [1]. At present, approximately 70 small-molecule kinase inhibitors are approved for clinical use by the United States Food

and Drug Administration (FDA) [2]. The large majority of these target tyrosine kinases, such as the anaplastic lymphoma kinase (ALK), epidermal growth factor receptor (EGFR), and vascular endothelial growth factor receptor (VEGFR). Only few inhibitors target serine/threonine kinases, such as B-RAF, or dual specificity kinases such as MEK1/2 [3].

Even if small-molecule kinase inhibitors have been conceived as molecularly targeted drugs, it is now increasingly recognized that they demonstrate pronounced off-target effects, which not only account for adverse drug effects but also may contribute to the desired activity [4]. A special case of off-target effects by protein kinase inhibitors is the modulation of the activity of xenosensing receptors of the chemical defense system, such as aryl hydrocarbon receptor or pregnane X receptor (PXR, NR1I2), a ligand-activated transcription factor of the nuclear receptor family. Activation of PXR induces drug detoxification and/or elimination, which may alter pharmacokinetics of the respective kinase inhibitors and potentially result in loss of efficacy [5]. However, PXR is not only a master regulator of drug detoxification [6], but also it was shown to modulate context-dependent tumor growth [7] and to promote cancer drug resistance [8,9], if activated in cancer cells. Therefore, PXR antagonism is proposed as a potential approach to prevent the formation of cancer drug resistance or even to overcome it [10,11]; although the mechanisms of resistance, which are caused by PXR activation, are still debated [9]. However, this approach is hindered by the limited number of specific and potent PXR antagonists and the challenging design of novel antagonists due to the promiscuous nature of PXR ligand binding [12]. Several protein kinase inhibitors activate PXR, including erlotinib, gefitinib, and sorafenib [13] or are even identified as PXR agonists, such as the MEK1/2 inhibitor U0126 [14] or dabrafenib [15]. So far only pazopanib is disclosed to elicit PXR antagonism [16]. Drugs with the dual function of inhibiting both PXR and protein kinases could open up new avenues for the treatment of cancer and in overcoming cancer drug resistance.

The objective of this study was to identify novel PXR antagonists among kinase inhibitors. To this end, using an *in silico* approach we screened the in-house Tübingen Kinase Inhibitor Collection (TüKIC) compound library, which contains 8500 small-molecule kinase inhibitors, to identify PXR binding compounds. Candidate ligands were subjected to experimental confirmation by PXR-dependent reporter gene assays in agonist and antagonist modes. Overall, this screening strategy resulted in the identification of four novel inhibitors and one strong activator with a common phenylaminobenzosuberone scaffold. Comprehensive characterization using different biochemical assays, assessing direct ligand binding, and cellular assays, addressing the interaction of PXR with coregulatory proteins, as well as gene expression analyses in colorectal cancer cells and differentiated hepatocytes, identified the novel inhibitors as passive mixed competitive/noncompetitive antagonists of PXR, which elicit gene- and tissue-specific modulation of PXR target gene expression.

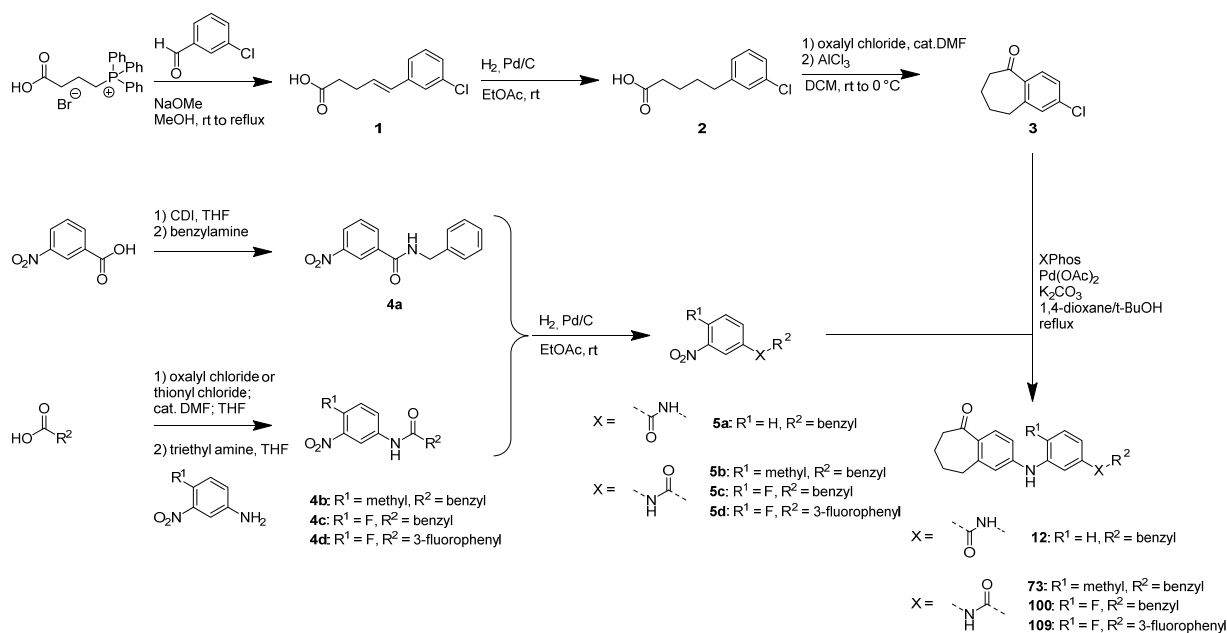
2. Materials and Methods

2.1. Chemicals and Reagents

DMSO and 1 α ,25-dihydroxy vitamin D₃ were purchased from Sigma-Aldrich (Taufkirchen, Germany). CITCO was provided by ENZO Life Sciences (Lörrach, Germany). Rifampicin was purchased from Merck Chemicals (Darmstadt, Germany). SR12813 and T0901317 were obtained from Tocris Bioscience (Bristol, UK). Minimum essential medium (MEM), Dulbecco's modified Eagle's medium (DMEM), William's E medium, and Trypsin-EDTA solution were purchased from Thermo Fischer Scientific (Waltham, MA, USA). L-glutamine, nonessential amino acids, sodium pyruvate, and penicillin–streptomycin mixture were provided by Biozym (Hessisch Oldendorf, Germany). Fetal bovine serum (FBS) was obtained from Biowest (Nuaille, France). Oligonucleotide primers were provided by Biomers (Ulm, Germany). TaqMan probes were purchased from Applied Biosystems (Waltham, MA, USA).

2.2. Origin of Compounds

The hit compounds of the TükIC compound library were synthesized in-house. Synthesis of compound **2** has been reported earlier [17]. The compounds **12**, **73**, **100**, and **109** were synthesized in a convergent synthesis approach, adopted from Martz et al. [17] (Scheme 1). Further details are provided in the Supplementary Information. In brief, the benzosuberone moiety was afforded in three steps starting from 3-chlorobenzaldehyde. The latter one was brought to reaction with the previously in situ formed **4** (triphenylphosphoranylidene)butanoate in a Wittig reaction, and the double bond in the resulting unsaturated carboxylic acid **1** was subsequently reduced by catalytic hydrogenation to achieve **2**. The resulting saturated carboxylic acid was activated as carboxylic acid chloride, and the ring closure to **3** was accomplished by intramolecular Friedel–Crafts acylation. In the synthesis of the side chains only the first step differs, depending on the nature of the amide (anilide or benzamide) and the availability of carboxylic acid chlorides. Due to the relatively higher nucleophilicity of benzylamine compared to anilines, the inverse amide **4a** could be afforded by activating 3-nitrobenzoic acid with carbonyl diimidazole and ensuing addition of benzylamine. For the synthesis of the anilides **4b–d**, the carboxylic acids were activated in situ as carboxylic acid chlorides with thionyl or oxalyl chloride, and then the respective anilines were added together with triethylamine as auxiliary base. The next step was the reduction of the aromatic nitro group to aromatic amines **5a–d**; this succeeded for all structures in the same way as catalytic hydrogenation under palladium/charcoal catalysis. In the final step, the above-described chlorobenzosuberone scaffold and the side chains were coupled to afford the final compounds **12**, **73**, **100**, and **109** in the sense of a Buchwald–Hartwig amination reaction.



Scheme 1. Synthesis of the described compounds; CDI = carbonyldiimidazole; DCM = dichloromethane; t-BuOH = tert-butanol.

2.3. Molecular Modeling

All the modeling was conducted by using OPLS3e force field [18] with Maestro (2018-4; Schrödinger, LLC, New York, NY, USA). The TükIC library of 8475 compounds was first filtered to keep only compounds with heavy atoms between 15–50, and compounds with long aliphatic chains (>6 carbons in a row) were excluded. The remaining compounds were prepared using LigPrep (Schrödinger, LLC, New York, NY, USA) with generating possible ionization states at pH 7.0 ± 2.0 using Epik [19], tautomers, and stereoisomers. Finally, compounds with a molecular weight > 600 and AlogP > 7.5 or < 0 were excluded, resulting

in 14,804 structures that were docked. Virtual screening was conducted by utilizing a PXR structure (PDB ID: 4j5w [20]), which prior to docking was refined by molecular dynamic (MD) simulations in complex with the competitive PXR antagonist SPA70 (data not shown), where three water molecules were kept on the site. Compounds were docked initially with Glide using extra precision (XP) accuracy [21–23]. The top scoring 150 compounds were redocked, using Induced Fit docking [24–26]; the results were visually analyzed, and selected poses underwent a further analysis by short molecular dynamics (MD) simulations, finally resulting in the selection of 56 compounds for experimental confirmation. The analogue searches at the subsequent stages were conducted by 2D structural search from the TÜKIC library.

We used the QM Conformer & Tautomer Predictor tool of Maestro (2021-2; Schrödinger LLC, New York, NY, USA) to generate five conformers for each of the five active compounds with the solvent set as water. This tool uses a stepwise minimization, applying Jaguar [27] in the final stages in the conformation generation.

2.4. Plasmid Constructs

Expression plasmids encoding human nuclear receptors CAR1 [28], PXR [29], RXR α [30], VDR [31], LBP-filled triple mutant PXR(S208W/S247W/C284W) [16], and CAR3 [32] have all been described previously. Expression plasmids encoding fusion proteins of GAL4-DNA binding domain (DBD) and the receptor interaction domains (RID) of steroid receptor coactivator-1 (SRC-1, residues 583–783) [32] and of silencing mediator of retinoid and thyroid hormone receptors (SMRT, residues 1109–1330) as well as the expression plasmid encoding the fusion of GAL4-DBD with PXR LBD helix 1 part (residues 132–188) [31] have been described previously. The expression plasmids encoding fusion proteins of the VP16 activation domain (AD) and PXR LBD (residues 108–434) or helix 2–12 part of it (residues 189–434) have been described previously [31]. The following firefly luciferase reporter gene plasmids have been described previously: CYP3A4 enhancer/promoter reporter gene plasmid pGL4-CYP3A4(7830 Δ 7208-364) [16] and pGL3(DR3)₃Tk, with a trimer of CYP3A23 direct repeat (DR) 3 motif [33]. The pGL4-G5 luciferase reporter gene plasmid was constructed by cloning the 200 bp KpnI/HindIII insert of pGL3-G5 [32], containing the GAL4 binding site pentamer and E1b promoter into pGL4.10(luc2) (Promega, Madison, WI, USA). Renilla luciferase expression plasmid pGL4.75[hRLuc/CMV] (Promega) and Metridia luciferase expression plasmid pMetLuc2control (Takara-Clontech, Mountain View, CA, USA), both under control of the CMV promoter, were used.

2.5. Cell Culture

HepG2 cells (HB-8065, lot number 58341723, ATCC, Manassas, VA) and H-P cells, representing stably transfected HepG2 cells overexpressing PXR [34], were cultivated at 37 °C, 5% CO₂ in MEM, which was supplemented with 10% FBS, 2 mM glutamine, 100 U/mL penicillin, and 100 μ g/mL streptomycin. HepG2 cells were obtained at passage 74, propagated, and used in the experiments between passages 93 and 110. H-P cells were used up to passage 30 after validation of the clone.

LS174T cells (CL-188, ATCC) were cultivated at 37 °C, 5% CO₂ in DMEM supplemented with 10% FBS, 2 mM L-glutamine, 100 U/mL penicillin, 100 μ g/mL streptomycin, 1% nonessential amino acids, and 1 mM sodium pyruvate. LS174T cells were obtained at passage 104, which was then reset as 1 and used in the experiments between passages 9 and 12. In chemical treatments, phenol red-free DMEM was used, and regular FBS was replaced by dextran-coated charcoal-treated FBS. For gene expression analyses, 0.3×10^6 LS174T cells were seeded per well of a 12-well plate. The next day, chemical treatment was initiated for 48–72 h, with daily medium change.

HepaRG cells (Biopredic, Rennes, France) were cultivated in phenol red-free William's E medium, supplemented with 10% FBS, 2 mM glutamine, 100 U/mL penicillin, 100 μ g/mL streptomycin, 5 μ g/mL insulin, and 50 μ M hydrocortisone. For gene expression analyses, 1.0×10^5 cells were seeded per well of a 12-well plate. After reaching confluence, growth

medium was supplemented with 2% DMSO, and cells were further cultivated for 2 weeks for differentiation into hepatocytes, as described previously [35]. Afterward, cells were adapted to induction medium (growth medium with FBS reduced to 2% and 0.2% DMSO) for 48 h. Then chemical treatment was started for another 48 h, with daily medium change.

Cells were routinely checked for contamination with mycoplasma by PCR (VenorGeM Classic, Minerva Biolabs, Berlin, Germany).

2.6. Cell Viability

HepG2 cells were seeded at density of 40,000 cells in 100 μ L per well in white, clear bottom 96-well plates (#655098, Greiner Bio-One, Frickenhausen, Germany). On the following day, cells were treated with 3, 10, or 30 μ M of test compounds for 24 h. Cell viabilities were determined by quantifying ATP content with CellTiter-Glo Luminescent Cell Viability Assay (Promega), according to the manufacturer's instructions. Luminescence was measured using EnSpire 2300 multimode plate reader (PerkinElmer, Rodgau, Germany) for 0.1 s. After subtracting background, relative cell viability was calculated in percent by dividing the value of treated cells by the value of DMSO-treated controls.

2.7. Transient Transfections, Mammalian Two Hybrid, and Reporter Gene Assays

Transient batch transfection of either HepG2 or H-P cells was conducted using JetPEI transfection reagent (Polyplus, Illkirch, France), essentially as recommended by the manufacturer. To investigate PXR activation, per well of a 96-well plate, a plasmid DNA mixture of 0.3 μ g pGL4-CYP3A4(-7830 Δ 7208-364) and 0.01 μ g pGL4.75[hRLuc/CMV] was diluted with 150 mM NaCl to a final volume of 25 μ L. Similarly, 0.6 μ L JetPEI reagent was diluted with 150 mM NaCl to 25 μ L. The diluted jetPEI was added to the diluted DNA mixture and incubated at room temperature for 15 min. In parallel, H-P cells were trypsinized and counted, and the cell number was adjusted to 40,000 cells in 200 μ L, per well. The transfection mixture was added to the cell suspension and pipetted into a 96-well plate (#83.3924.300, Sarstedt, Nümbrecht, Germany). After overnight incubation, cells were treated with chemicals for 24 h before cell lysis with 50 μ L of passive lysis buffer (Promega). Firefly and Renilla luciferase activities were measured from 10 μ L of sample with 150 μ L of firefly luciferase assay solution [29] and 100 μ L Renilla luciferase assay solution [36], respectively, using EnSpire 2300. Results were normalized by dividing Firefly luciferase activity by Renilla luciferase activity measured from the same well.

For testing nuclear receptor selectivity and LBP-filled PXR mutant activation, transient batch transfection was conducted as above, but with HepG2 cells and the following plasmids, with amounts per well of 0.23–0.27 μ g pGL4-CYP3A4(-7830 Δ 7208-364) or pGL3(DR3)₃Tk (as reporter for VDR), 0.01 μ g pMetLuc2control, and 0.03 μ g either CAR1, CAR3, VDR, or PXR(S208W/S247W/C284W) expression plasmids. In addition, 0.03 μ g RXR α expression plasmid was added to CAR3 transfections. Metridia luciferase activity was measured from 10 μ L of medium supernatant after adding 100 μ L Renilla luciferase assay solution.

In mammalian two-hybrid corepressor/coactivator interaction assays, per well 0.24 μ g pGL4-G5, 0.01 μ g pGL4.75[hRLuc/CMV], 0.03 μ g expression plasmid encoding VP16-AD/PXR LBD fusion, and 0.03 μ g expression plasmids encoding GAL4-DBD/SRC1-RID or GAL4-DBD/SMRT-RID fusions, respectively, were used. For the SMRT corepressor interaction assay, additionally 0.015 μ g of RXR α expression plasmid was included.

For the mammalian two-hybrid PXR ligand-binding domain assembly assay, 0.24 μ g pGL4-G5, 0.01 μ g pMetLuc2control, and 0.03 μ g each of expression plasmids encoding GAL4-DBD/PXR(132–188) and VP16-AD/PXR(189–434) fusion proteins were used per well. Otherwise, batch transfection was performed as above with HepG2 cells.

2.8. Limited Proteolytic Digestion

The limited proteolytic digestion assay was performed as described previously [37]. A total of 2.5% DMSO, 30 μ M of T0901317, and 100 and 250 μ M of test compounds were used.

2.9. Competitive Radioligand Binding Assay

The assay was performed by Eurofins Cerep (Celle-Lévescault, France). Compounds **100** and **109** were tested with 7 different concentrations, ranging from 0.01 μM to 10 μM , for IC_{50} determination. Compound binding was calculated as % inhibition of the binding of the radioactively labeled PXR ligand [^3H]SR12813 to recombinant human PXR, as described previously [38].

2.10. Quantitative Real-Time PCR Analysis

Total RNA was prepared from LS174T and HepaRG cells using the NucleoSpin RNA kit (Macherey-Nagel, Düren, Germany). RNA integrity was analyzed by formaldehyde-agarose gel electrophoresis. First strand cDNA was synthesized as described earlier [37]. Relative quantification analyses were conducted with TaqMan RT-PCR utilizing the BioMark HD system and FLEX Six Gene Expression Integrated Fluidic Circuits (Fluidigm, South San Francisco, CA, USA) as described previously [34]. TaqMan gene expression assays were either the commercially available predesigned assays Hs00184500_m1 (ABCB1) and Hs00604506_m1 (CYP3A4) (Thermo Fischer Scientific) or have been described previously: CYP2B6 [31] and 18S rRNA [39]. Data were analyzed as described previously [37]. Gene expression levels were normalized to respective 18S rRNA levels.

2.11. Kinase Inhibition Profiling

Compounds **100** and **109** were profiled for inhibition of kinase activity at concentrations of 1 μM and 10 μM against a panel of 335 wild-type protein kinases with single measurements. The analysis was performed by ProQinase (Freiburg, Germany).

2.12. Statistical Analysis

In quantitative real-time PCR analysis, if the standard deviation of Ct values in technical triplicates exceeded 0.2, the outlier was omitted. In the cell viability experiments, if the coefficient of variation of technical triplicates exceeded 20%, the outlier was omitted. Data are presented as mean \pm SD of at least three independent experiments, each performed in technical triplicates, if not specified otherwise in the respective figure legend. In bar charts, individual experiments are illustrated as dots. In scatter plots, mean and individual samples are shown. Statistical significances were determined using unpaired or paired *t*-test between two groups. Multiple comparisons were performed using ordinary or repeated measures one-way analysis of variance (ANOVA) with Dunnett's multiple comparisons test. Comparisons to a hypothetical value were conducted with one sample *t*-test corrected by the method of Bonferroni. Statistical analyses were performed with GraphPad Prism 9.3.1 (San Diego, CA, USA).

3. Results

3.1. Identification of Novel PXR Inhibitors with Phenylaminobenzosuberone Scaffold

We applied a virtual screening approach to identify potential PXR ligands from the TüKIC compound library (Figure 1A). Based on the in silico evaluation, the candidate ligands were tested in antagonist and agonist modes using the PXR-dependent CYP3A4 reporter gene assay (Supplementary Figure S1). This initial screen, complemented by two subsequent structural analog searches based on the assay results (Figure 1A), resulted in the identification of four novel potential PXR antagonists, which inhibited induction of CYP3A4 reporter activity by the prototypical agonist rifampicin by more than 50% (Figure 1B and Figures S1–S3). The most potent compounds **73** and **100** demonstrated 60–70% inhibition of rifampicin-mediated induction, while only weakly activating the CYP3A4 reporter on their own, at about 10% of rifampicin activity (Supplementary Figures S2 and S3, respectively). Inhibition of agonist-induced PXR activation was not rifampicin-specific, as all novel compounds, except **12**, suppressed SR12813-induced PXR activation to a similar extent as rifampicin-induced activation (Supplementary Figure S4). With the single exception of **12**, the identified candidate ligands elicited pronounced toxicity at 30 μM with cell viability of

$\leq 50\%$. At 10 μM , modest toxicity was observed, with residual viability between 75 to 83% (Supplementary Figure S5). Thus, maximum concentrations were limited to 10 μM in the subsequent concentration–response analyses. Only **12** was used up to 30 μM . Results of the concentration–response analyses, with opposing effects of the compounds in agonist and antagonist modes (Figure 1B,C), further argue against a significant influence of the limited cytotoxicity at 10 μM on PXR-dependent reporter activity. Compounds **73** and **100** showed IC_{50} values of 8.3 μM (95% CI 6.1–11.7 μM) and 2.8 μM (95% CI 2.1–3.8 μM), respectively. Thereby these compounds demonstrated stronger inhibition of rifampicin-induced PXR activation compared to **2** and **12**, which displayed IC_{50} values of 11.2 μM (95% CI 7.7–17.2 μM) and 33.7 μM (95% CI 20.2–61.6 μM), respectively (Figure 1B). As concentrations higher than 10 μM or 30 μM (compound **12** only) could not be used, IC_{50} values were calculated from concentration–response analyses, which did not result in a lower plateau, thereby limiting the significance of the absolute values. Furthermore, **73** and **100** activated PXR only weakly, with maximal effects (E_{max}) of about 2-fold, but with low EC_{50} values of 0.41 μM (95% CI 0.18–0.87 μM) and 0.15 μM (95% CI 0.04–0.45 μM), respectively (Figure 1C). On the other hand, concentration–response curves of **2** and **12** did not reach plateau at the maximum concentration of 10 μM and 30 μM , respectively. Therefore, the observed E_{max} and calculated EC_{50} values must be interpreted with caution. This is also reflected by the inability to calculate the complete confidence interval for the EC_{50} of **2** (EC_{50} of 7.0 μM , 95% CI 1.2–??? μM) and by the large confidence interval with **12** (EC_{50} of 22.3 μM , 95% CI 8.8–110 μM). Interestingly, we identified also a structurally related strong activator, **109** (Figure 1A and Figure S3), which showed 27-fold maximal induction and EC_{50} of 44 μM (95% CI 20–??? μM) (Figure 1C). Due to the incomplete confidence interval calculation, EC_{50} of **109** has to be interpreted with caution.

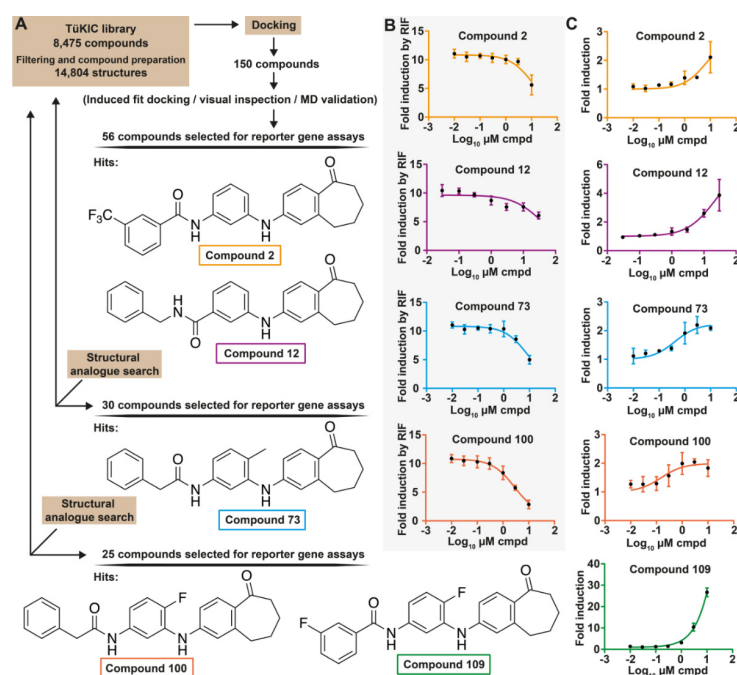


Figure 1. Identification of novel putative PXR ligands from the TüKIC library. (A) Workflow of the in silico screening and structures of the identified candidate ligands. (B) Concentration–response curves of the novel inhibitors for antagonism of PXR activity. (C) Concentration–response curves of the same compounds and **109** for PXR activation. H-P cells were transfected with CYP3A4 promoter reporter and cotreated with 10 μM rifampicin (RIF) and increasing concentrations of the indicated compounds (B) or treated with increasing concentrations of the novel compounds only (C) for 24 h. Data are expressed as mean \pm SD fold induction with respect to the DMSO-treated cells from three independent experiments. IC_{50} and EC_{50} values calculated with nonlinear regression using formula with 3 parameters (constraint bottom =1) using GraphPadPrism.

The identified hit compounds share a common phenylaminobenzosuberone scaffold (Figure 2A). Compound **12** has an inverted amide linking the R^2 and is showing the highest IC_{50} -value. The length of the R^2 seems also important for PXR inhibition, as the shorter phenyl containing **109** does not inhibit PXR. Of note, **2**, which also has a phenyl group in its R^2 , has an additional sterically large trifluoromethyl substituent in its phenyl ring. Furthermore, compounds with the highest inhibitory activity display a heavier substituent than H in their R^1 . Based on the lowest energy conformations in water (Figure 2B, Supplementary Table S1, Supplementary Figures S6–S10), the most potent inhibitors **73** and **100** appear to prefer a more extended conformation, while the other compounds tend to appear in a more folded configuration. The heavier R^1 -substituent found in **73** and **100** seems to promote the extended conformations (100% and 84%, respectively), while in **12**, with an H-atom in this position, this conformation is not preferred (11%) (Supplementary Table S1, Supplementary Figures S6–S10).

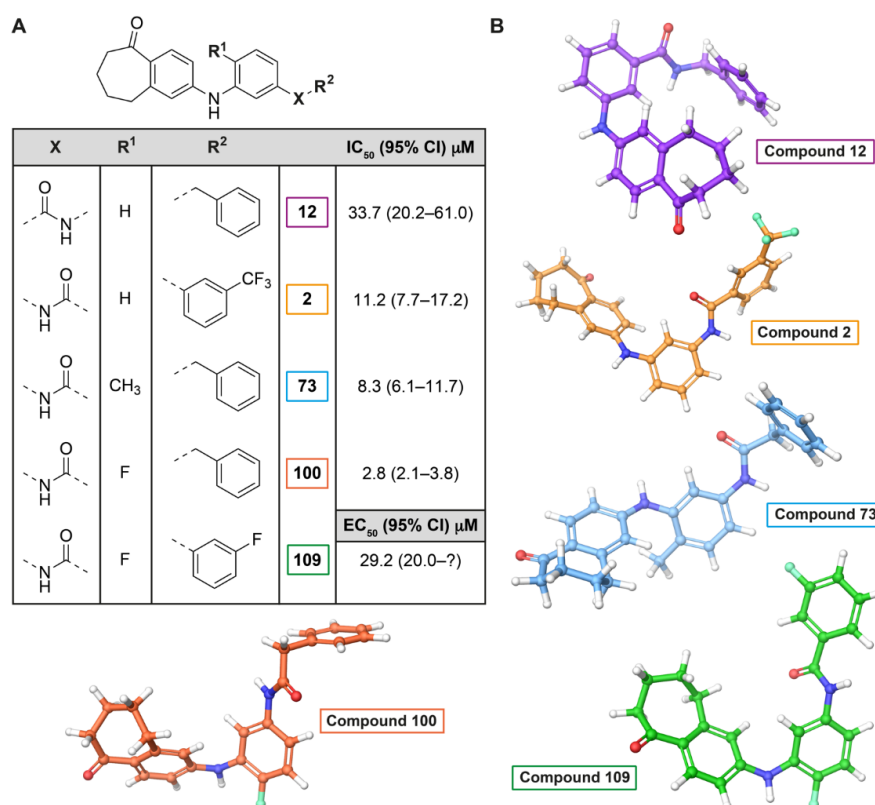


Figure 2. Structure–activity relationship of the identified compounds and their preferred conformations. (A) All compounds share a common phenylaminobenzosuberone core, displaying structural differences in X, R^1 , and R^2 positions. (B) The lowest energy conformation of the compounds in water based on QM Conformer predictor (see details in Methods and Supplementary Table S1 and Supplementary Figures S6–S10).

3.2. Compounds **73** and **100** Demonstrate Competitive Antagonism of PXR

To investigate the mechanism of PXR inhibition by the two strongest inhibitors **73** and **100**, their effects on the concentration–response curve of rifampicin were assessed. Figure 3A shows that at lower concentrations (1 and 3 μM) compound **73** induced a parallel dextral shift of the rifampicin concentration–response curve with corresponding increases in the EC_{50} values of rifampicin. Without compound **73**, EC_{50} of rifampicin was 1.8 μM (95% CI 0.76–4.5 μM), which increased in the presence of 1 μM or 3 μM of compound **73** to 2.9 μM (95% CI 1.0–9.1 μM) or 6.4 μM (95% CI 2.5–19.6 μM), respectively. With 10 μM of compound **73**, rifampicin EC_{50} was further elevated to 7.4 μM (95% CI 2.0–47.7 μM), and reduction of E_{max} was pronounced. Compound **100** also caused a dextral shift of the rifampicin concentration–response curve with corresponding increases in the EC_{50} values

of rifampicin. EC_{50} value of rifampicin raised from 2.0 μM (95% CI 1.6–2.6 μM) in the absence of **100** to 4.8 μM (95% CI 3.4–7.0 μM) with 1 μM of **100**, and further to 9.3 μM (95% CI 5.4–17.0 μM) with 3 μM of **100** (Figure 3B). Rifampicin EC_{50} in the presence of 10 μM **100** could not be calculated reliably. These observations indicate that compounds **73** and **100** act as competitive antagonists of PXR. The reduction in E_{max} , which was observed especially at the highest test concentration of both compounds, may indicate additional noncompetitive antagonism.

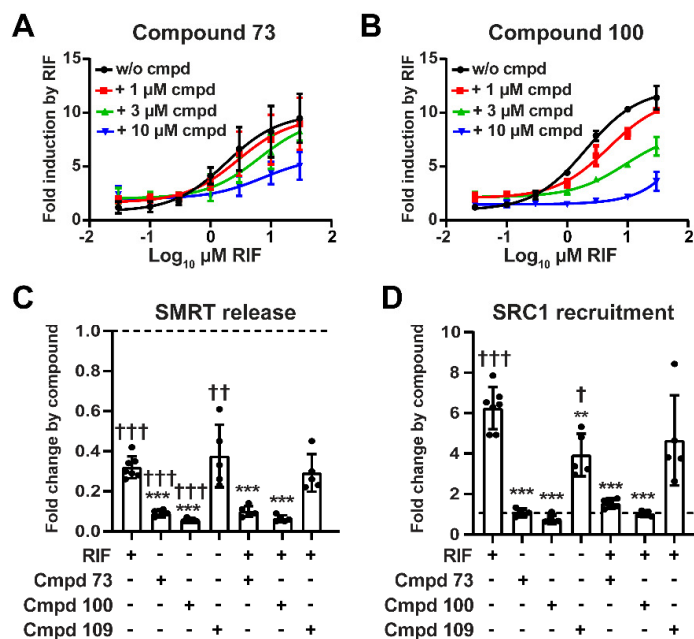


Figure 3. Compounds **73** and **100** act as mixed competitive/noncompetitive passive antagonists of PXR. (A,B) Effects of novel PXR antagonists on the concentration–response curve of rifampicin. H-P cells were transfected with CYP3A4 promoter reporter and cotreated with increasing concentrations of rifampicin (RIF) with or without fixed concentrations of **73** and **100**. Effects of novel compounds on PXR’s interactions with corepressor SMRT (C) and coactivator SRC1 (D). HepG2 cells were cotransfected with GAL4-G5 reporter gene and expression plasmids encoding VP16-PXR LBD(108–434) and (C) GAL4-DBD-SMRT-RID or (D) GAL4-DBD-SRC1-RID fusion proteins and treated with 0.2% DMSO, 10 μM test compounds alone, or cotreated with 10 μM rifampicin. Data are shown as mean \pm SD fold induction with respect to DMSO-treated cells from five independent experiments and individual experiments illustrated with dots. Statistically significant differences are illustrated with asterisks or daggers. ** $p < 0.01$, *** $p < 0.001$, compared to 10 μM rifampicin-treated cells analyzed by one-way ANOVA with Dunnett’s multiple comparisons test. † $p < 0.05$, †† $p < 0.01$, and ††† $p < 0.001$ single treatments compared to DMSO, which was set as 1, analyzed by one sample t -test corrected by the method of Bonferroni.

3.3. Compounds 73 and 100 Disrupt PXR’s Coregulatory Protein Interactions

We assessed the effect of the novel antagonists and, for comparison, of the activator **109** on the ligand-dependent interactions of PXR with coregulatory proteins using respective mammalian two hybrid assays. Rifampicin, the novel antagonists **73** and **100**, as well as the activator **109**, all impaired the constitutive interaction of PXR with the corepressor silencing mediator of retinoic acid and thyroid hormone receptor (SMRT, NCOR2) (Figure 3C). Release of SMRT from PXR was even more pronounced by compounds **73** and **100** than by rifampicin. As expected, rifampicin induced the interaction of PXR with the coactivator steroid receptor coactivator 1 (SRC1, NCOA1). Similarly, PXR’s interaction with SRC1 was induced by the activating compound **109** (Figure 3D). In contrast, compounds **73** and **100** did not induce any interaction of PXR with coactivator SRC1. Furthermore, they abrogated the rifampicin-dependent recruitment of SRC1 by PXR. These results indicate that **73** and

100 can be classified as passive antagonists as they impair the interactions of PXR with both coactivators and corepressors [40], while **109** exhibits agonist properties.

3.4. Phenylaminobenzosuberones Demonstrate Direct Binding to the PXR-LBP

Ligands of nuclear receptors physically interact with the respective ligand binding domain (LBD). To confirm the binding of the antagonists **73** and **100**, as well as of the activator **109**, to the PXR LBD, the respective *in vitro* limited proteolytic digestion assay was applied [37]. The assay relies on conformational changes in the LBD that are induced by the ligand binding to it, which alter the accessibility of proteases to cleavage sites [41]. Figure 4A shows that all compounds, similar to the reference ligand T0901317, resulted in increased protection of three proteolytic fragments from digestion, albeit at relative intensities different from T0901317. These data indicate that the novel compounds act as ligands of PXR. Due to the nature of the PXR limited proteolytic digestion *in vitro* assay, it required higher concentrations of the ligands than PXR activation in cellular assays, which has been described previously [37].

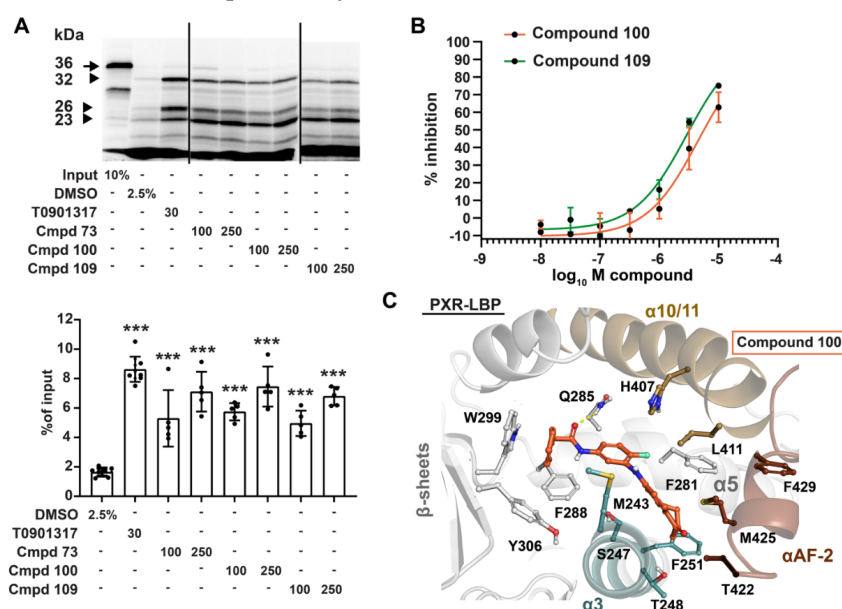


Figure 4. Phenylaminobenzosuberones demonstrate direct binding to PXR. **(A)** Limited proteolytic digestion assay was conducted by preincubating ³⁵S-labeled PXR LBD with 100 μM or 250 μM test compounds, 30 μM T0901317, or 2.5% DMSO. Upper panel shows the radioluminographic scan of a representative experiment. Arrow shows 36 kDa input of PXR LBD, and arrow heads show protected 32, 26, and 23 kDa fragments after limited proteolytic digest with trypsin. Lower panel shows respective densitometric quantifications of the sum of the three protected fragments. Columns show mean ± SD of five independent experiments (each performed as single measurement) and the individual experiments illustrated with dots, calculated with respect to input. Statistically significant differences are illustrated with asterisks. *** *p* < 0.001, compared to respective preincubations with DMSO and analyzed by one-way ANOVA with Dunnett's multiple comparisons test. **(B)** Competitive radioligand binding assay demonstrates direct binding of compounds **100** and **109** to PXR. Data are shown as means ± SD of % inhibition of control specific binding by test compounds from duplicate analysis. **(C)** Putative binding mode of compound **100** in the PXR-LBP. A representative snapshot taken from the microsecond timescale MD simulations [42] is shown in the figure, where compound **100** (orange C-atoms stick model) is shown with the closely located key residues of PXR-LBP (shown with sticks, with colors matching to their secondary structure region: αAF-2, dark brown; α3-helix, cyan; α10/11-helix, light brown; other, white). The benzyl moiety of **100** is occupying the hydrophobic subpocket formed by F288, W299, and Y306, while the benzosuberone is oriented toward the αAF-2 region. H-bond to Q285 from the amide is shown with yellow dashed line.

To corroborate these results, the binding of compounds **100** and **109** was further analyzed by competitive radioligand binding assays. Figure 4B shows that compounds **100** and **109** inhibited the binding of ^3H -labeled SR12813 to PXR with IC_{50} of $4.5\ \mu\text{M}$ (95% CI 1.9–15.4 μM) and $2.9\ \mu\text{M}$ (95% CI 1.9–4.7 μM), respectively. We also investigated the binding of compound **100** by microsecond timescale molecular dynamics (MD) simulations [42]. Figure 4C illustrates one putative binding configuration of compound **100** in the ligand binding pocket (LBP) of PXR based on the simulation data. In conclusion, the results of the limited proteolytic digestion assay and competitive radioligand binding assay, supported by MD simulations, indicate a physical interaction of the novel compounds with PXR and thus confirm ligand binding.

As a cellular assay equivalent of the biochemical ligand binding assays, we applied the mammalian two-hybrid LBD assembly assay, which identifies both agonists and antagonists of nuclear receptors [43]. Figure 5A shows that antagonists **73** and **100** and agonist **109** induced the assembly of the PXR LBD. PXR antagonism can occur either by competitive binding into the PXR ligand binding pocket (LBP), which has been demonstrated with SPA70 [44], or by binding exclusively or additionally to an allosteric site outside LBP. Exclusive binding outside the LBP has been observed with ketoconazole and campothecin [45,46], while coumestrol and pimecrolimus demonstrated binding both to the LBP and to an allosteric site outside it [16,47]. As the previous assays do not distinguish between these possibilities, we next investigated the effects of the novel compounds on the constitutive activity of a LBP-filled PXR mutant. As a result of mutation, ligand binding into the pocket is prevented, and the mutant exhibits high constitutive activity [47]. While **100** did not affect constitutive activity of the PXR mutant, **73** still suppressed the respective activity by 40% (Figure 5B). These data suggest additional binding of **73** outside the LBP, at least with the LBP-filled mutant, which further supports the above suggested noncompetitive antagonism for **73** (see Figure 3A). In contrast, the allosteric binding of compound **100** to PXR outside the LBP is not supported by this assay.

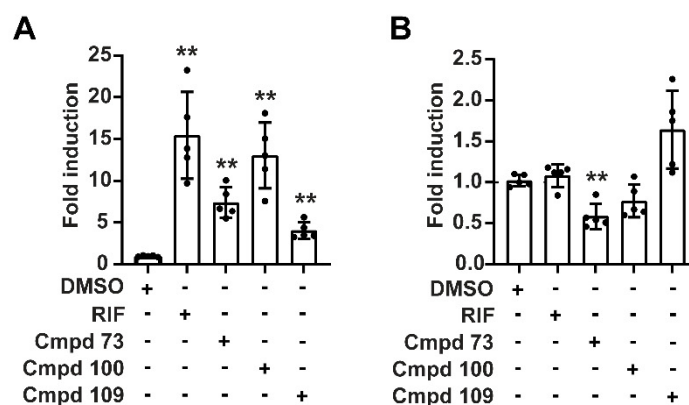


Figure 5. PXR binding of phenylaminobenzosuberones is indicated by cellular PXR LBD assembly assay and inhibition of LBP-filled PXR mutant (A) Novel compounds induce PXR LBD assembly. HepG2 cells were transiently transfected with GL4-G5 reporter, GAL4-DBD/PXR LBD(132–188), and VP16-AD/PXR LBD(189–434) fusion protein expression plasmids and treated with 0.1% DMSO, 10 μM rifampicin, or 10 μM test compounds for 24 h. Data are expressed as mean \pm SD fold induction with respect to DMSO-treated cells from five independent experiments and individual experiments illustrated with dots. (B) Effects of novel compounds on LBP-filled PXR mutant. HepG2 cells were transiently transfected with CYP3A4 reporter and expression plasmid encoding PXR(S208W/S247W/C284W) and treated as described above. Data are shown as mean \pm SD fold induction, with respect to treatment with DMSO only, from five independent experiments. Statistically significant differences are illustrated with asterisks. $** p < 0.01$, compared to respective treatments with DMSO analyzed by repeated measures one-way ANOVA with Dunnett’s multiple comparisons test.

3.5. Kinase Inhibition Profile of 100 and 109

Several kinases have shown to phosphorylate PXR and thereby to modulate the transcriptional activity of the receptor [48–52]. Usually, phosphorylation of human PXR results in impaired transcriptional activity and repression of target gene expression. Thus, kinase inhibition may contribute indirectly to the activation of PXR. Compounds **100** and **109**, as representatives of structurally related antagonists and agonist, were tested against a set of 335 wild-type kinases to disclose their kinase inhibition profile (Supplementary Data S1). At 10 μM , compound **100** inhibited five kinases, namely BRAF, MAPKAPK3, p38 β , PKA, and RAF1 by $\geq 50\%$. At 1 μM , only RAF1 was inhibited considerably. With 10 μM of compound **109**, eight kinases, BRAF, CK1- δ , CK1- ϵ , CK1- $\gamma 3$, p38 α , p38 β , PKA, and RAF1, were inhibited at least by 50%. CK1- δ , p38 α and RAF1 were already inhibited to this extent at 1 μM . Of these kinases, only PKA was previously shown to affect PXR function by inhibitory phosphorylation [49]. Consequently, the inhibition of PKA, which could promote PXR activation, might explain the observed limited PXR activation occurring with **100**, and, on the other hand, it may contribute to the strong PXR activation by **109**.

3.6. Nuclear Receptor Selectivity of Novel PXR Ligands

To assess the selectivity of the novel PXR ligands, we determined their effects on CAR- and VDR-mediated transactivation. CAR and VDR belong to the same NR1I group of nuclear receptors as PXR and share 37–45% sequence similarity in their LBD [53]. Both **73** and **100** slightly suppressed the constitutive activity of isoform CAR1 by 30% (Figure 6A). Compound **73** also impaired the ligand-induced activation of CAR3 by CITCO by 46% (Figure 6B) and reduced the ligand-induced VDR activation by 51% (Figure 6C), while **100** displayed no respective effects. Neither CAR3 nor VDR were activated by **73** or **100** on their own. In contrast, the agonist **109** weakly activated both CAR3 and VDR, but the effects were only 12% and 5% of the effects of the respective prototypical ligands CITCO and 1 α ,25-dihydroxy vitamin D₃. Activation of CAR1 by compound **109** did not reach statistical significance after Bonferroni correction, but a trend was observed ($p = 0.0552$). These results indicate that the novel PXR antagonists, and here especially **100**, demonstrate only minor inhibitory effects on CAR and VDR.

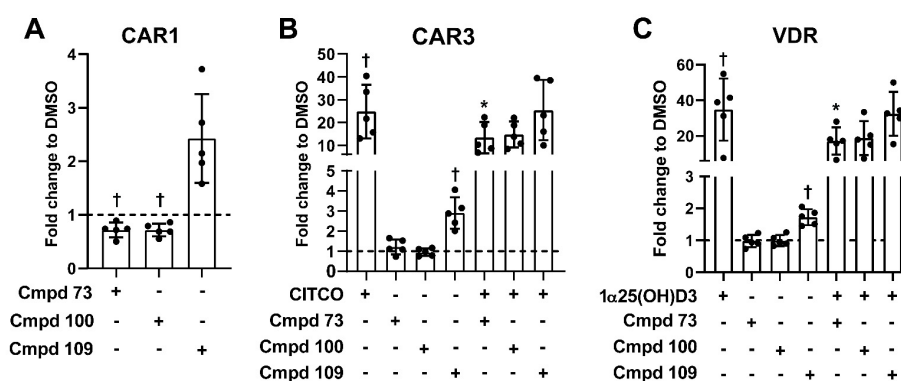


Figure 6. Selectivity of novel PXR ligands within the NR1I group of nuclear receptors. HepG2 cells were transiently transfected with CYP3A4 reporter and expression plasmids encoding (A) CAR1 or (B) CAR3, or (C) direct repeat (DR)₃ reporter and expression plasmid encoding VDR. Cells were treated with 0.2% DMSO or 10 μM chemicals as indicated. 1 α ,25(OH)₂D₃ was used at 1 μM . Data are shown as mean \pm SD fold induction with respect to DMSO-treated cells from five independent experiments and individual experiments illustrated with dots. Statistically significant differences are illustrated with asterisks and daggers. † $p < 0.05$ single treatments compared to DMSO-treated cells set as 1 and analyzed by one sample t -test corrected by the method of Bonferroni. * $p < 0.05$ cotreatments compared to (B) CITCO or (C) 1 α ,25(OH)₂D₃ analyzed by repeated measures one-way ANOVA with Dunnett’s multiple comparisons test.

3.7. Expression of Prototypical Endogenous PXR Target Genes Is Differentially Modulated by the Compounds

To investigate the effects of the novel PXR ligands on the expression of endogenous PXR target genes, we utilized LS174T colorectal carcinoma cells, the PXR expression level of which is comparable to liver [31]. In these cells, rifampicin induced the expression of ABCB1 and CYP3A4. Expression of ABCB1 was not induced by **73** and **100**, and these compounds also suppressed its rifampicin-mediated induction (Figure 7A). Despite being characterized as PXR antagonists, both compounds induced CYP3A4 expression, which was comparable to induction by rifampicin. In accordance with its agonist properties, **109** induced both genes, whereby ABCB1 was induced to a lesser extent than by rifampicin. These results indicate that compounds **73** and **100** antagonize PXR activity in a gene-specific manner in intestinal carcinoma cells.

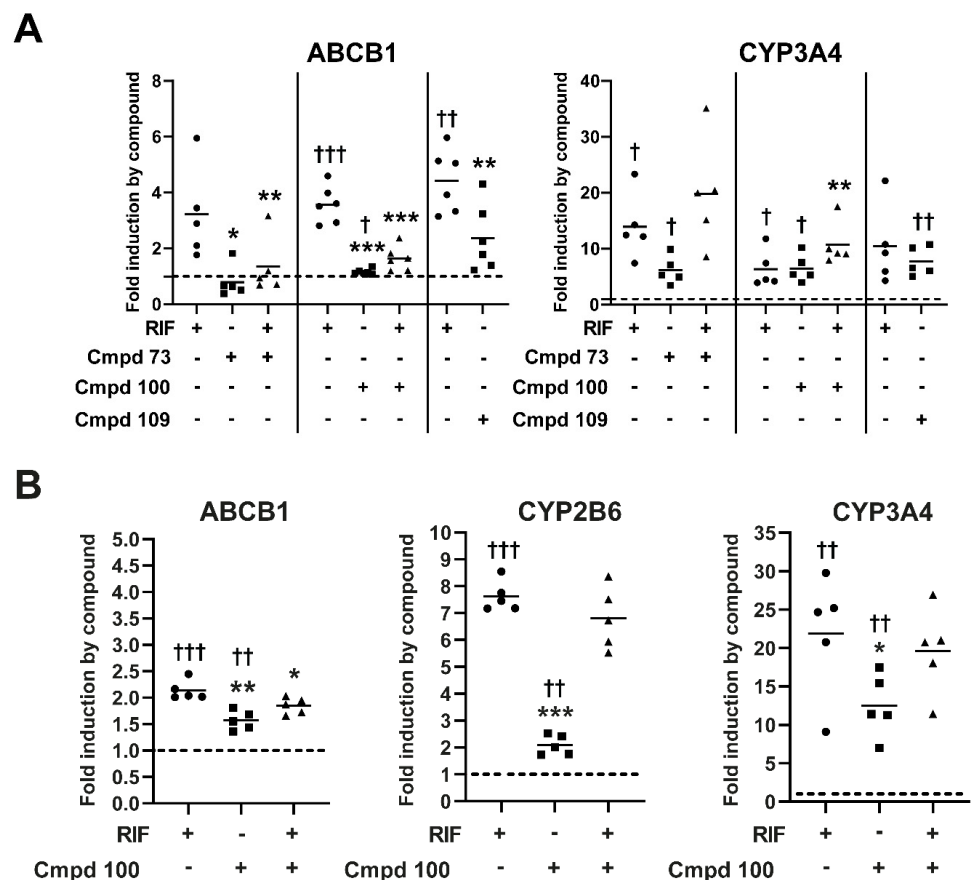


Figure 7. Effects of novel PXR ligands on endogenous PXR target gene expression. **(A)** LS174T cells were treated with 0.1–0.2% DMSO, 10 μ M rifampicin (RIF), or 10 μ M test compounds or cotreated with 10 μ M RIF and compounds **73** and **100** for 48 h (**109**) or 72 h (**73**, **100**). mRNA expression was determined by RT-qPCR and normalized to the expression of 18S rRNA. Data are expressed as mean fold induction with respect to DMSO-treated cells, expression in which was set as 1, from six independent experiments and individual experiments illustrated with dots. **(B)** HepaRG cells were treated with 0.2% DMSO, 10 μ M RIF, or 10 μ M compound **100** or cotreated with 10 μ M RIF and compound **100** for 48 h. mRNA expression was measured as described above. Data are expressed as mean fold induction with respect to DMSO-treated cells from five independent experiments and individual experiments illustrated with dots. Statistically significant differences are illustrated with asterisks and daggers. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$ compared to 10 μ M rifampicin-treated cells analyzed by repeated measures one-way ANOVA with Dunnett’s multiple comparisons test (**73**, **100**) or by paired t -test (**109**). † $p < 0.05$, †† $p < 0.01$, and ††† $p < 0.001$ single treatments compared to DMSO, which was set as 1, analyzed by one sample t -test corrected by the method of Bonferroni.

To investigate whether PXR antagonism can be observed also in other tissues, the effect of compound **100** on PXR target gene expression was analyzed in differentiated HepaRG cells, which closely resemble functional hepatocytes [54]. Compound **100** induced the expression of ABCB1, CYP2B6, and CYP3A4 (Figure 7B). However, the induction of these three genes was considerably weaker than by the prototypical agonist rifampicin. In contrast to LS174T cells, compound **100** demonstrated only minor inhibitory effects on rifampicin-induced expression of ABCB1. In summary, these PXR target gene expression data indicate that compound **100** might possess partial agonist activity in hepatic cells.

4. Discussion

By a combination of in silico molecular modeling and cellular PXR-dependent reporter gene assays, we identified four novel PXR inhibitors and a structurally related full agonist from the in-house TÜKIC compound library. Comprehensive subsequent analyses, addressing key features of nuclear receptor biology, confirmed the most potent inhibitors **73** and **100** as PXR ligands, demonstrating passive, mixed competitive/noncompetitive antagonism, and gene- and tissue-specific modulation of PXR target gene expression, qualifying them as selective PXR modulators.

The lack of experimental structural data for a PXR–antagonist complex, combined with the promiscuous nature of PXR ligand binding, brings challenges to identify antagonists by docking [55], which as a method has limitations on its own [56]. Although, we were able to identify competitive and mixed competitive/noncompetitive antagonists here starting from an in silico screen essentially relying on molecular docking, long timescale MD simulations are required for a proper estimation of the antagonist compound binding mode. A detailed analysis of compound **100** binding mode in the PXR LBD is provided in a complementary publication [42].

The identified antagonists share a common phenylaminobenzosuberone scaffold with the agonist **109**. The high structural similarity of antagonists and the agonist demonstrates that even subtle structural changes have great impact on PXR activation and inhibition. The only obvious conformational difference was that the most potent antagonists **73** and **100** appeared to prefer more extended conformations. We and others have observed previously that subtle structural changes can completely change the activity of PXR ligands. For example, the reduction of artemisinin to its lactol derivative dihydroartemisinin abrogates PXR ligand binding in the PXR-LBD assembly assay [57]. More recently, Li et al. demonstrated for the PXR antagonist SPA70 that small alterations in substituents at the common scaffold change the activity of the compound from antagonistic to full agonistic [58]. Our findings here emphasize the challenging design of PXR antagonists due to the great impact of subtle structural changes to PXR activity.

Regarding characterization of the respective mechanism of antagonism, we focused on the two strongest compounds **73** and **100**. According to the operational definition of antagonism, the dextral shift in the concentration–response curve of PXR agonist rifampicin by increasing concentrations of inhibitors suggested classification of the novel compounds as competitive antagonists. However, as the maximal effect was also decreasing, contribution of a noncompetitive/allosteric component in antagonism appears obvious also. If the compounds act as competitive antagonists, they have to bind into the LBP of PXR. Results from different assays provided independent evidence for LBP binding of the compounds. First, both **73** and **100** demonstrated displacement of LBP-bound agonist SR12813 in the competitive radioligand binding assay. Second, MD simulations enabled the identification of the putative binding mode of compound **100** in the LBP of PXR. Third, the compounds induced the assembly of the PXR LBD, which is not achieved by antagonists with exclusive allosteric binding, such as camptothecin and pazopanib [16]. Independent evidence for noncompetitive or allosteric binding outside the LBP was obtained only for **73**, which still demonstrated limited inhibition of the constitutive activity of the LBP-filled PXR triple mutant S208W/S247W/C284W. Only exclusive allosteric antagonists, such as camptothecin and pazopanib [16] or mixed competitive/allosteric antagonists, such as coumestrol [47]

and pimecrolimus [16], showed inhibition of this mutant's activity. Compounds demonstrating only competitive PXR antagonism, such as nelfinavir, did not [59].

Passive antagonism is suggested by the fact that the novel antagonists abolished the rifampicin-induced interaction of PXR with coactivator SRC1, as well as impairing the constitutive interaction of PXR with corepressor SMRT. Except SPA70, for which recruitment of SMRT was shown [44], previously described PXR antagonists, such as ketoconazole, camptothecin, sulforaphane, pazopanib, and pimecrolimus, all demonstrated the same effects on coregulator interaction as described here for **73** and **100** [16,46]. Molecular dynamic simulations of compound **100** indicated that binding it affects the conformation of the PXR LBD in distinct regions, including the AF-2 region (where coactivators and corepressors bind), supporting the observed biological results [42].

The novel antagonists **73** and **100** demonstrated gene-specific effects in LS174T colorectal cancer cells, not inducing on their own the expression of ABCB1, but inducing CYP3A4 expression to a similar extent as rifampicin. Furthermore, both compounds antagonized exclusively the rifampicin-mediated induction of ABCB1. The strong induction of CYP3A4 and absence of antagonism of the rifampicin-dependent activation is surprising, as the compounds were identified by their inhibition of rifampicin-mediated CYP3A4 enhancer/promoter activation. However, in contrast to the transfected reporter gene construct, the regulatory region of the endogenous gene resides in chromatin. Differences in the promoter context between genes, resulting in altered interaction with or altered activity of coregulators, have been suggested for the explanation of gene-specific effects of nuclear receptor ligands [60]. Alternatively, the observed induction of endogenous CYP3A4 expression in LS174T cells might not result from PXR agonism itself. Given the fact that **100** was shown to inhibit PKA at 10 μ M, it is conceivable that inhibition of the PXR-inhibitory kinase activity of PKA may participate in CYP3A4 induction, especially as it was shown that the activation of PKA resulted in repression of CYP3A4 expression [49]. Even in this scenario, we would have to assume that the effect of PKA inhibition on PXR activity is not relevant for ABCB1. The observed tissue-specific effects of compound **100** might result from divergent coactivator versus corepressor levels, varying PXR levels, or different activities of coactivators or PXR due to modulation of kinase signaling by the compound. These potential cellular variables have been shown previously to determine the activity of selective modulators of nuclear receptors [60].

With BRAF, RAF1, p38 β , and MAPKAPK3, four of the five kinases, which were strongly inhibited by **100**, belong to the MAPK/ERK pathway. Hitherto and in conventional 2D cell culture, the MAPK/ERK pathway has not been associated with regulation of hepatic cytochrome P450 expression, as its inhibition by dominant-negative MEK1 did not induce CYP3A4 in HepG2 cells [14]. However, it was recently shown in 3D spheroid cultures of primary human hepatocytes that pharmacological inhibition of the pathway induced CYP3A4 and CYP2B6 expression [61]. By siRNA-mediated knockdown of PXR, the authors further showed that PXR is involved in the respective CYP3A4 and CYP2B6 induction. Consequently, they suggested inhibition of PXR by the MAPK/ERK pathway. Here we used differentiated HepaRG cells in 2D culture, which are not a pure hepatocyte culture but also contain biliary epithelial cells [54]. If differentiated HepaRG cells resembled 3D hepatocyte spheroids in terms of MAPK/ERK pathway activity, it is conceivable that treatment with **100** might result in inhibition of this pathway, followed by release of PXR inhibition and consequently CYP3A4 and CYP2B6 induction. In this case, the induction of CYP3A4 and CYP2B6 by **100** would not indicate PXR agonism but indirect activation through MAPK/ERK pathway inhibition. Further research is required to distinguish between these possibilities.

Additional inhibition of PXR by protein kinase inhibitors may provide benefit for cancer therapy in several ways. First, many kinase inhibitors are metabolized by cytochrome P450 enzymes and transported by MDR1/P-glycoprotein [62–66], the encoding genes, which are regulated by PXR. In addition, at least 10 of the roughly 70 approved kinase inhibitors used for the treatment of cancer have even been shown to activate

PXR [13,15,63–65,67] (Supplementary Table S2), which may result in autoinduction of drug metabolism and elimination. Thus, generating derivatives, which inhibit the receptor, may result in reduced drug metabolism and/or drug elimination, which may improve drug efficacy and reduce off-target toxicity as allowing lower dosing. The feasibility of a respective structure-based synthesis approach has recently been demonstrated for B-RAF inhibitors structurally related to the PXR activator dabrafenib, which neither bind to nor activate the receptor [68]. Second, activation of PXR is known to promote cancer cell growth and to contribute to the development of cancer drug resistance [8,9]. Strategically, a dual PXR/protein kinase inhibitor is expected to target tumor growth by two different mechanisms and concomitantly will prevent generation of PXR-dependent chemoresistance. The clinical relevance of the concept is illustrated by PXR mediating the chemoresistance of hepatocellular carcinoma to the multikinase inhibitor sorafenib [69].

In conclusion, we have identified and characterized novel selective receptor modulators of PXR from an in-house kinase inhibitor compound library. Their common phenylaminobenzosuberone scaffold represents a previously unknown PXR ligand structure and may be used as a starting point for the synthesis of the suggested dual PXR and protein kinase inhibitors.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/cells11081299/s1>: Supplementary information file: Supplementary methods and compound NMR spectra; Supplementary Figure S1: Effects of 56 in silico screened TüKIC compounds (A) alone or (B) in combination with 10 μ M rifampicin on PXR-mediated transactivation of CYP3A4 reporter gene; Supplementary Figure S2: Effects of first round structural analogues (A) alone or (B) in combination with rifampicin on PXR-mediated transactivation of CYP3A4 reporter gene; Supplementary Figure S3: Effects of second round structural analogues (A) alone or (B) in combination with rifampicin on PXR-mediated transactivation of CYP3A4 reporter gene; Supplementary Figure S4: Effects of potential novel PXR antagonists in combination with 1 μ M SR18213 on PXR-mediated transactivation of CYP3A4 reporter gene; Supplementary Figure S5: Cell viability of HepG2 cells following 24 h treatment with potential novel PXR ligands; Supplementary Figure S6: QM Conformer & Tautomer Predictor output conformations of compound **2**; Supplementary Figure S7: QM Conformer & Tautomer Predictor output conformations of compound **12**; Supplementary Figure S8: QM Conformer & Tautomer Predictor output conformations of compound **73**; Supplementary Figure S9: QM Conformer & Tautomer Predictor output conformations of compound **100**; Supplementary Figure S10: QM Conformer & Tautomer Predictor output conformations of compound **109**; Supplementary Table S1: QM Conformer & Tautomer Predictor output conformations and their energies; Supplementary Table S2: PXR activating protein kinase inhibitors. Supplementary Data File S1: Kinase inhibition profile of compounds **100** and **109**. Supplementary Data File S2: 3D coordinates of the conformations of the small-molecules. Supplementary Data File S3: 3D coordinates of the PXR LBD compound **100** putative binding mode.

Author Contributions: Conceptualization, E.-K.M., T.P., S.L., M.S. and O.B.; methodology, E.-K.M., T.P., A.R., J.R. and O.B.; validation, E.-K.M., T.P., and O.B.; formal analysis, E.-K.M., T.P. and O.B.; investigation, E.-K.M., T.P., A.R. and O.B.; resources, J.R., M.S., S.L. and O.B.; data curation, E.-K.M., T.P. and O.B.; writing—original draft preparation, E.-K.M., T.P., J.R. and O.B.; writing—review and editing, all authors.; visualization, E.-K.M., T.P., A.R. and O.B.; supervision, S.L. and O.B.; project administration, S.L. and O.B.; funding acquisition, S.L. and O.B. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Robert Bosch Stiftung, Stuttgart, Germany (O.B. and M.S.); the Interfaculty Center for Pharmacogenomics and Pharma Research of the University of Tübingen, Tübingen, Germany (E.-K.M., J.R., S.L., and O.B.); and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC 2180–390900677 (M.S. and S.L.). A.R. acknowledges the European Union's Horizon 2020 research and innovation program under grant agreement no 825762, EDCMET project. T.P. acknowledges financial support from the Orion Research Foundation sr, European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie (grant no 839230) and Academy of Finland GeneCellNano Flagship (337120).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data are included in the figures in the manuscript and as supplementary data.

Acknowledgments: We appreciate the expert technical assistance of K. Abuazi-Rincones. Eurofins Cerep (Celle-Lévescault, France) and ProQinase (Freiburg, Germany) are acknowledged for conducting the competitive radioligand binding assay and kinase inhibition profiling, respectively. This work contains parts of the doctoral thesis of E.-K.M.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Kannaiyan, R.; Mahadevan, D. A Comprehensive Review of Protein Kinase Inhibitors for Cancer Therapy. *Expert Rev. Anticancer Ther.* **2018**, *18*, 1249–1270. [[CrossRef](#)] [[PubMed](#)]
2. Attwood, M.M.; Fabbro, D.; Sokolov, A.V.; Knapp, S.; Schiöth, H.B. Trends in kinase drug discovery: Targets, indications and inhibitor design. *Nat. Rev. Drug Discov.* **2021**, *20*, 839–861. [[CrossRef](#)] [[PubMed](#)]
3. Roskoski, R., Jr. Properties of FDA-approved small molecule protein kinase inhibitors: A 2022 update. *Pharmacol. Res.* **2022**, *175*, 106037. [[CrossRef](#)] [[PubMed](#)]
4. Munoz, L. Non-kinase targets of protein kinase inhibitors. *Nat. Rev. Drug Discov.* **2017**, *16*, 424–440. [[CrossRef](#)] [[PubMed](#)]
5. Pollet, M.; Krutmann, J.; Haarmann-Stemann, T. Commentary: Usage of Mitogen-Activated Protein Kinase Small Molecule Inhibitors: More Than Just Inhibition! *Front. Pharmacol.* **2018**, *9*, 935. [[CrossRef](#)]
6. Wang, Y.-M.; Ong, S.S.; Chai, S.C.; Chen, T. Role of CAR and PXR in Xenobiotic Sensing and Metabolism. *Expert Opin. Drug Metab. Toxicol.* **2012**, *8*, 803–817. [[CrossRef](#)]
7. Pondugula, S.R.; Pavek, P.; Mani, S. Pregnane X Receptor and Cancer: Context-Specificity is Key. *Nucl. Recept. Res.* **2016**, *3*, 101198. [[CrossRef](#)]
8. Niu, X.; Wu, T.; Li, G.; Gu, X.; Tian, Y.; Cui, H. Insights into the critical role of the PXR in preventing carcinogenesis and chemotherapeutic drug resistance. *Int. J. Biol. Sci.* **2022**, *18*, 742–759. [[CrossRef](#)]
9. Xing, Y.; Yan, J.; Niu, Y. PXR: A center of transcriptional regulation in cancer. *Acta Pharm. Sin. B* **2020**, *10*, 197–206. [[CrossRef](#)]
10. Masuyama, H.; Nakamura, K.; Nobumoto, E.; Hiramatsu, Y. Inhibition of pregnane X receptor pathway contributes to the cell growth inhibition and apoptosis of anticancer agents in ovarian cancer cells. *Int. J. Oncol.* **2016**, *49*, 1211–1220. [[CrossRef](#)]
11. Staudinger, J.L. Clinical applications of small molecule inhibitors of Pregnane X receptor. *Mol. Cell. Endocrinol.* **2019**, *485*, 61–71. [[CrossRef](#)] [[PubMed](#)]
12. Chai, S.C.; Wright, W.C.; Chen, T. Strategies for developing pregnane X receptor antagonists: Implications from metabolism to cancer. *Med. Res. Rev.* **2020**, *40*, 1061–1083. [[CrossRef](#)] [[PubMed](#)]
13. Harmsen, S.; Meijerman, I.; Maas-Bakker, R.F.; Beijnen, J.H.; Schellens, J.H.M. PXR-mediated P-glycoprotein induction by small molecule tyrosine kinase inhibitors. *Eur. J. Pharm. Sci.* **2013**, *48*, 644–649. [[CrossRef](#)] [[PubMed](#)]
14. Smutny, T.; Bitman, M.; Urban, M.; Dubecka, M.; Vrzal, R.; Dvorak, Z.; Pavek, P. U0126, a mitogen-activated protein kinase kinase 1 and 2 (MEK1 and 2) inhibitor, selectively up-regulates main isoforms of CYP3A subfamily via a pregnane X receptor (PXR) in HepG2 cells. *Arch. Toxicol.* **2014**, *88*, 2243–2259. [[CrossRef](#)]
15. Creusot, N.; Gassiot, M.; Alaterre, E.; Chiavarina, B.; Grimaldi, M.; Boulahtouf, A.; Toporova, L.; Gerbal-Chaloin, S.; Daujat-Chavanieu, M.; Matheux, A.; et al. The Anti-Cancer Drug Dabrafenib Is a Potent Activator of the Human Pregnane X Receptor. *Cells* **2020**, *9*, 1641. [[CrossRef](#)]
16. Burk, O.; Kuzikov, M.; Kronenberger, T.; Jeske, J.; Keminer, O.; Thasler, W.E.; Schwab, M.; Wrenger, C.; Windshügel, B. Identification of approved drugs as potent inhibitors of pregnane X receptor activation with differential receptor interaction profiles. *Arch. Toxicol.* **2018**, *92*, 1435–1451. [[CrossRef](#)]
17. Martz, K.E.; Dorn, A.; Baur, B.; Schattel, V.; Goettert, M.I.; Mayer-Wrangowski, S.C.; Rauh, D.; Laufer, S.A. Targeting the Hinge Glycine Flip and the Activation Loop: Novel Approach to Potent P38 α Inhibitors. *J. Med. Chem.* **2012**, *55*, 7862–7874. [[CrossRef](#)]
18. Roos, K.; Wu, C.; Damm, W.; Reboul, M.; Stevenson, J.M.; Lu, C.; Dahlgren, M.K.; Mondal, S.; Chen, W.; Wang, L.; et al. OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *J. Chem. Theory Comput.* **2019**, *15*, 1863–1874. [[CrossRef](#)]
19. Greenwood, J.R.; Calkins, D.; Sullivan, A.P.; Shelley, J.C. Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. *J. Comput. Aided Mol. Des.* **2010**, *24*, 591–604. [[CrossRef](#)]
20. Wallace, B.D.; Betts, L.; Talmage, G.; Pollet, R.M.; Holman, N.S.; Redinbo, M.R. Structural and Functional Analysis of the Human Nuclear Xenobiotic Receptor PXR in Complex with RXR α . *J. Mol. Biol.* **2013**, *425*, 2561–2577. [[CrossRef](#)]
21. Friesner, R.A.; Murphy, R.B.; Repasky, M.P.; Frye, L.L.; Greenwood, J.R.; Halgren, T.A.; Sanschagrin, P.C.; Mainz, D.T. Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196. [[CrossRef](#)] [[PubMed](#)]

22. Friesner, R.A.; Banks, J.L.; Murphy, R.B.; Halgren, T.A.; Klicic, J.J.; Mainz, D.T.; Repasky, M.P.; Knoll, E.H.; Shelley, M.; Perry, J.K.; et al. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749. [[CrossRef](#)] [[PubMed](#)]
23. Halgren, T.A.; Murphy, R.B.; Friesner, R.A.; Beard, H.S.; Frye, L.L.; Pollard, W.T.; Banks, J.L. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47*, 1750–1759. [[CrossRef](#)] [[PubMed](#)]
24. Farid, R.; Day, T.; Friesner, R.A.; Pearlstein, R.A. New insights about HERG blockade obtained from protein modeling, potential energy mapping, and docking studies. *Bioorg. Med. Chem.* **2006**, *14*, 3160–3173. [[CrossRef](#)] [[PubMed](#)]
25. Sherman, W.; Day, T.; Jacobson, M.P.; Friesner, R.A.; Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* **2006**, *49*, 534–553. [[CrossRef](#)] [[PubMed](#)]
26. Sherman, W.; Beard, H.S.; Farid, R. Use of an induced fit receptor structure in virtual screening. *Chem. Biol. Drug Des.* **2006**, *67*, 83–84. [[CrossRef](#)]
27. Bochevarov, A.D.; Harder, E.; Hughes, T.F.; Greenwood, J.R.; Braden, D.A.; Philipp, D.M.; Rinaldo, D.; Halls, M.D.; Zhang, J.; Friesner, R.A. Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences. *Int. J. Quantum Chem.* **2013**, *113*, 2110–2142. [[CrossRef](#)]
28. Burk, O.; Tegude, H.; Koch, I.; Hustert, E.; Wolbold, R.; Glaeser, H.; Klein, K.; Fromm, M.F.; Nuessler, A.K.; Neuhaus, P.; et al. Molecular mechanisms of polymorphic CYP3A7 expression in adult human liver and intestine. *J. Biol. Chem.* **2002**, *277*, 24280–24288. [[CrossRef](#)]
29. Geick, A.; Eichelbaum, M.; Burk, O. Nuclear receptor response elements mediate induction of intestinal *MDR1* by rifampin. *J. Biol. Chem.* **2001**, *276*, 14581–14587. [[CrossRef](#)]
30. Mathäs, M.; Burk, O.; Qiu, H.; Nusslag, C.; Gödtel-Armbrust, U.; Baranyai, D.; Deng, S.; Römer, K.; Nem, D.; Windshügel, B.; et al. Evolutionary history and functional characterization of the amphibian xenosensor CAR. *Mol. Endocrinol. Baltim. Md* **2012**, *26*, 14–26. [[CrossRef](#)]
31. Burk, O.; Arnold, K.A.; Nussler, A.K.; Schaeffeler, E.; Efimova, E.; Avery, B.A.; Avery, M.A.; Fromm, M.F.; Eichelbaum, M. Antimalarial artemisinin drugs induce cytochrome P450 and *MDR1* expression by activation of xenosensors pregnane X receptor and constitutive androstane receptor. *Mol. Pharmacol.* **2005**, *67*, 1954–1965. [[CrossRef](#)] [[PubMed](#)]
32. Arnold, K.A.; Eichelbaum, M.; Burk, O. Alternative Splicing Affects the Function and Tissue-Specific Expression of the Human Constitutive Androstane Receptor. *Nucl. Recept.* **2004**, *2*, 1. [[CrossRef](#)] [[PubMed](#)]
33. Hustert, E.; Zibat, A.; Presecan-Siedel, E.; Eiselt, R.; Mueller, R.; Fuss, C.; Brehm, I.; Brinkmann, U.; Eichelbaum, M.; Wojnowski, L.; et al. Natural protein variants of pregnane X receptor with altered transactivation activity toward *CYP3A4*. *Drug Metab. Dispos. Biol. Fate Chem.* **2001**, *29*, 1454–1459. [[PubMed](#)]
34. Bitter, A.; Rümmele, P.; Klein, K.; Kandel, B.A.; Rieger, J.K.; Nüssler, A.K.; Zanger, U.M.; Trauner, M.; Schwab, M.; Burk, O. Pregnane X receptor activation and silencing promote steatosis of human hepatic cells by distinct lipogenic mechanisms. *Arch. Toxicol.* **2015**, *89*, 2089–2103. [[CrossRef](#)] [[PubMed](#)]
35. Anthérieu, S.; Chesné, C.; Li, R.; Camus, S.; Lahoz, A.; Picazo, L.; Turpeinen, M.; Tolonen, A.; Uusitalo, J.; Guguen-Guillouzo, C.; et al. Stable expression, activity, and inducibility of cytochromes P450 in differentiated HepaRG cells. *Drug Metab. Dispos. Biol. Fate Chem.* **2010**, *38*, 516–525. [[CrossRef](#)]
36. Piedade, R.; Traub, S.; Bitter, A.; Nüssler, A.K.; Gil, J.P.; Schwab, M.; Burk, O. Carboxymefloquine, the major metabolite of the antimalarial drug mefloquine, induces drug-metabolizing enzyme and transporter expression by activation of pregnane X receptor. *Antimicrob. Agents Chemother.* **2015**, *59*, 96–104. [[CrossRef](#)]
37. Jeske, J.; Windshügel, B.; Thasler, W.E.; Schwab, M.; Burk, O. Human pregnane X receptor is activated by dibenzazepine carbamate-based inhibitors of constitutive androstane receptor. *Arch. Toxicol.* **2017**, *91*, 2375–2390. [[CrossRef](#)]
38. Zhu, Z.; Kim, S.; Chen, T.; Lin, J.-H.; Bell, A.; Bryson, J.; Dubaquié, Y.; Yan, N.; Yanchunas, J.; Xie, D.; et al. Correlation of high-throughput pregnane X receptor (PXR) transactivation and binding assays. *J. Biomol. Screen.* **2004**, *9*, 533–540. [[CrossRef](#)]
39. Hoffart, E.; Ghebreghiorgis, L.; Nussler, A.K.; Thasler, W.E.; Weiss, T.S.; Schwab, M.; Burk, O. Effects of atorvastatin metabolites on induction of drug-metabolizing enzymes and membrane transporters through human pregnane X receptor. *Br. J. Pharmacol.* **2012**, *165*, 1595–1608. [[CrossRef](#)]
40. Schoch, G.A.; D’Arcy, B.; Stihle, M.; Burger, D.; Bär, D.; Benz, J.; Thoma, R.; Ruf, A. Molecular Switch in the Glucocorticoid Receptor: Active and Passive Antagonist Conformations. *J. Mol. Biol.* **2010**, *395*, 568–577. [[CrossRef](#)]
41. Allan, G.F.; Leng, X.; Tsai, S.Y.; Weigel, N.L.; Edwards, D.P.; Tsai, M.J.; O’Malley, B.W. Hormone and antihormone induce distinct conformational changes which are central to steroid receptor activation. *J. Biol. Chem.* **1992**, *267*, 19513–19520. [[CrossRef](#)]
42. Rashidian, A.; Mustonen, E.-K.; Kronenberger, T.; Schwab, M.; Burk, O.; Laufer, S.A.; Pantzar, T. Discrepancy in interactions and conformational dynamics of pregnane X receptor (PXR) bound to an agonist and a novel competitive antagonist. *Comput. Struct. Biotechnol. J.* **2022**, *2022*. *submitted*.
43. Pissios, P.; Tzamelis, I.; Kushner, P.; Moore, D.D. Dynamic stabilization of nuclear receptor ligand binding domains by hormone or corepressor binding. *Mol. Cell* **2000**, *6*, 245–253. [[CrossRef](#)]
44. Lin, W.; Wang, Y.-M.; Chai, S.C.; Lv, L.; Zheng, J.; Wu, J.; Zhang, Q.; Wang, Y.-D.; Griffin, P.R.; Chen, T. SPA70 is a potent antagonist of human pregnane X receptor. *Nat. Commun.* **2017**, *8*, 741. [[CrossRef](#)] [[PubMed](#)]

45. Chen, Y.; Tang, Y.; Robbins, G.T.; Nie, D. Camptothecin attenuates cytochrome P450 3A4 induction by blocking the activation of human pregnane X receptor. *J. Pharmacol. Exp. Ther.* **2010**, *334*, 999–1008. [[CrossRef](#)]
46. Huang, H.; Wang, H.; Sinz, M.; Zoeckler, M.; Staudinger, J.; Redinbo, M.R.; Teotico, D.G.; Locker, J.; Kalpana, G.V.; Mani, S. Inhibition of drug metabolism by blocking the activation of nuclear receptors by ketoconazole. *Oncogene* **2007**, *26*, 258–268. [[CrossRef](#)] [[PubMed](#)]
47. Wang, H.; Li, H.; Moore, L.B.; Johnson, M.D.L.; Maglich, J.M.; Goodwin, B.; Ittoop, O.R.R.; Wisely, B.; Creech, K.; Parks, D.J.; et al. The Phytoestrogen Coumestrol Is a Naturally Occurring Antagonist of the Human Pregnane X Receptor. *Mol. Endocrinol.* **2008**, *22*, 838–857. [[CrossRef](#)]
48. Ding, X.; Staudinger, J.L. Repression of PXR-mediated induction of hepatic CYP3A gene expression by protein kinase C. *Biochem. Pharmacol.* **2005**, *69*, 867–873. [[CrossRef](#)]
49. Lichti-Kaiser, K.; Xu, C.; Staudinger, J.L. Cyclic AMP-dependent protein kinase signaling modulates pregnane x receptor activity in a species-specific manner. *J. Biol. Chem.* **2009**, *284*, 6639–6649. [[CrossRef](#)]
50. Lin, W.; Wu, J.; Dong, H.; Bouck, D.; Zeng, F.-Y.; Chen, T. Cyclin-dependent kinase 2 negatively regulates human pregnane X receptor-mediated CYP3A4 gene expression in HepG2 liver carcinoma cells. *J. Biol. Chem.* **2008**, *283*, 30650–30657. [[CrossRef](#)]
51. Pondugula, S.R.; Brimer-Cline, C.; Wu, J.; Schuetz, E.G.; Tyagi, R.K.; Chen, T. A phosphomimetic mutation at threonine-57 abolishes transactivation activity and alters nuclear localization pattern of human pregnane x receptor. *Drug Metab. Dispos. Biol. Fate Chem.* **2009**, *37*, 719–730. [[CrossRef](#)] [[PubMed](#)]
52. Taneja, G.; Chu, C.; Maturu, P.; Moorthy, B.; Ghose, R. Role of c-Jun-N-Terminal Kinase in Pregnane X Receptor-Mediated Induction of Human Cytochrome P4503A4 In Vitro. *Drug Metab. Dispos. Biol. Fate Chem.* **2018**, *46*, 397–404. [[CrossRef](#)] [[PubMed](#)]
53. Wu, B.; Li, S.; Dong, D. 3D structures and ligand specificities of nuclear xenobiotic receptors CAR, PXR and VDR. *Drug Discov. Today* **2013**, *18*, 574–581. [[CrossRef](#)] [[PubMed](#)]
54. Aninat, C.; Piton, A.; Glaise, D.; Le Charpentier, T.; Langouët, S.; Morel, F.; Guguen-Guillouzo, C.; Guillouzo, A. Expression of Cytochromes P450, Conjugating Enzymes and Nuclear Receptors in Human Hepatoma HepaRG Cells. *Drug Metab. Dispos.* **2006**, *34*, 75. [[CrossRef](#)] [[PubMed](#)]
55. Hall, A.; Chanteux, H.; Ménochet, K.; Ledecq, M.; Schulze, M.-S.E.D. Designing Out PXR Activity on Drug Discovery Projects: A Review of Structure-Based Methods, Empirical and Computational Approaches. *J. Med. Chem.* **2021**, *64*, 6413–6522. [[CrossRef](#)] [[PubMed](#)]
56. Pantzar, T.; Poso, A. Binding Affinity via Docking: Fact and Fiction. *Molecules* **2018**, *23*, 1899. [[CrossRef](#)]
57. Burk, O.; Piedade, R.; Ghebreghiorgis, L.; Fait, J.T.; Nussler, A.K.; Gil, J.P.; Windshügel, B.; Schwab, M. Differential effects of clinically used derivatives and metabolites of artemisinin in the activation of constitutive androstane receptor isoforms. *Br. J. Pharmacol.* **2012**, *167*, 666–681. [[CrossRef](#)]
58. Li, Y.; Lin, W.; Wright, W.C.; Chai, S.C.; Wu, J.; Chen, T. Building a Chemical Toolbox for Human Pregnane X Receptor Research: Discovery of Agonists, Inverse Agonists, and Antagonists Among Analogs Based on the Unique Chemical Scaffold of SPA70. *J. Med. Chem.* **2021**, *64*, 1733–1761. [[CrossRef](#)]
59. Burk, O.; Kronenberger, T.; Keminer, O.; Lee, S.M.L.; Schiergens, T.S.; Schwab, M.; Windshügel, B. Nelfinavir and Its Active Metabolite M8 Are Partial Agonists and Competitive Antagonists of the Human Pregnane X Receptor. *Mol. Pharmacol.* **2021**, *99*, 184–196. [[CrossRef](#)]
60. Johnson, A.B.; O'Malley, B.W. Steroid receptor coactivators 1, 2, and 3: Critical regulators of nuclear receptor activity and steroid receptor modulator (SRM)-based cancer therapy. *Mol. Cell. Endocrinol.* **2012**, *348*, 430–439. [[CrossRef](#)]
61. Hendriks, D.F.G.; Vorrink, S.U.; Smutny, T.; Sim, S.C.; Nordling, Å.; Ullah, S.; Kumondai, M.; Jones, B.C.; Johansson, I.; Andersson, T.B.; et al. Clinically Relevant Cytochrome P450 3A4 Induction Mechanisms and Drug Screening in Three-Dimensional Spheroid Cultures of Primary Human Hepatocytes. *Clin. Pharmacol. Ther.* **2020**, *108*, 844–855. [[CrossRef](#)] [[PubMed](#)]
62. Dohse, M.; Scharenberg, C.; Shukla, S.; Robey, R.W.; Volkmann, T.; Deeken, J.F.; Brendel, C.; Ambudkar, S.V.; Neubauer, A.; Bates, S.E. Comparison of ATP-binding cassette transporter interactions with the tyrosine kinase inhibitors imatinib, nilotinib, and dasatinib. *Drug Metab. Dispos. Biol. Fate Chem.* **2010**, *38*, 1371–1380. [[CrossRef](#)] [[PubMed](#)]
63. *Multi-Disciplinary Review and Evaluation NDA 210496 BRAFTOVI™ (encorafenib)*; U.S. Food & Drug Administration: Silver Spring, MD, USA. Available online: https://www.accessdata.fda.gov/drugsatfda_docs/nda/2018/210496Orig1s000MultidisciplineR.pdf (accessed on 6 April 2022).
64. *Multi-Disciplinary Review and Evaluation NDA 210868 LORBRENA™ (lorlatinib)*; U.S. Food & Drug Administration: Silver Spring, MD, USA. Available online: https://www.accessdata.fda.gov/drugsatfda_docs/nda/2018/210868Orig1s000MultidisciplineR.pdf (accessed on 6 April 2022).
65. *Multi-Disciplinary Review and Evaluation NDA 208772 ALUNBRIG™ (brigatinib)*; U.S. Food & Drug Administration; U.S. Food & Drug Administration: Silver Spring, MD, USA. Available online: https://www.accessdata.fda.gov/drugsatfda_docs/nda/2017/208772Orig1s000MultidisciplineR.pdf (accessed on 6 April 2022).
66. Wang, X.; Zhang, X.; Huang, X.; Li, Y.; Wu, M.; Liu, J. The drug-drug interaction of sorafenib mediated by P-glycoprotein and CYP3A4. *Xenobiotica Fate Foreign Compd. Biol. Syst.* **2016**, *46*, 651–658. [[CrossRef](#)] [[PubMed](#)]
67. MacLeod, A.K.; McLaughlin, L.A.; Henderson, C.J.; Wolf, C.R. Activation status of the Pregnane X Receptor (PXR) influences Vemurafenib availability in humanized mouse models. *Cancer Res.* **2015**, *75*, 4573–4581. [[CrossRef](#)]

68. Schneider, M.; Delfosse, V.; Gelin, M.; Grimaldi, M.; Granell, M.; Heriaud, L.; Pons, J.-L.; Cohen Gonsaud, M.; Balaguer, P.; Bourguet, W.; et al. Structure-Based and Knowledge-Informed Design of B-Raf Inhibitors Devoid of Deleterious PXR Binding. *J. Med. Chem.* **2022**, *65*, 1552–1566. [[CrossRef](#)]
69. Feng, F.; Jiang, Q.; Cao, S.; Cao, Y.; Li, R.; Shen, L.; Zhu, H.; Wang, T.; Sun, L.; Liang, E.; et al. Pregnane X receptor mediates sorafenib resistance in advanced hepatocellular carcinoma. *Biochim. Biophys. Acta Gen. Subj.* **2018**, *1862*, 1017–1030. [[CrossRef](#)]

Supplementary Information

Target hopping from protein kinases to PXR: identification of small-molecule protein kinase inhibitors as selective modulators of pregnane X receptor from TüKIC library

Enni-Kaisa Mustonen (1), Tatu Pantsar (2,3), Azam Rashidian (4), Juliander Reiner (2), Matthias Schwab (1,5,6), Stefan Laufer (2,6,7), Oliver Burk (1)

(1) Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, and University of Tübingen, Tübingen, Germany

(2) Department of Pharmaceutical and Medicinal Chemistry, Institute of Pharmaceutical Sciences, University of Tübingen, Tübingen, Germany

(3) School of Pharmacy, Faculty of Health Sciences, University of Eastern Finland, Kuopio, Finland

(4) Department of Internal Medicine VIII, University Hospital Tübingen, Tübingen, Germany

(5) Departments of Clinical Pharmacology and Biochemistry and Pharmacy, University of Tuebingen, Tübingen, Germany

(6) Cluster of Excellence iFIT (EXC 2180) “Image-Guided and Functionally Instructed Tumor Therapies”, University of Tübingen, Tübingen, Germany

(7) Tuebingen Center for Academic Drug Discovery & Development (TüCAD2), 72076, Tübingen, Germany

Content

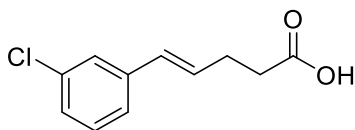
Supplementary methods and compound NMR spectra.....	2
Supplementary figures.....	15
Supplementary tables.....	22

Supplementary Methods: Compound synthesis

General information

Reagents and solvents were purchased from Sigma-Aldrich, VWR international, Merck KGaA, Alfa Aesar, Acros, Roth, Fisher Scientific, Fluka, OxChem, FluoroChem, TCI or abcr and used without further purification. LRMS were determined by DC-MS (Advion expressionS CMS, ESITLC analyses were performed on fluorescent TLC Silica gel 60 F254 aluminum sheets (Merck KGaA) with UV illumination (254/366 nm). The puriflash 430 automated flash chromatography system in combination with standard glass or plastics columns were used for (flash-)chromatography with Davisil LC60A 20-45 micron silica from Grace Davison or Geduran Si60 63-200 micron silica from Merck KGaA as stationary phase. The used mobile phases are indicated in the experimental section. Purity of all compounds was determined via reverse phase HPLC on Hewlett Packard HP 1090 Series II LC equipped with a UV diode array detector (DAD, detection at 230 nm and 254 nm) and was $\geq 95\%$ for tested compounds. The chromatographic separation was performed on a Phenomenex Luna 5u C8 column (150 mm x 4.6 mm, 5 μm) at 35 °C oven temperature with an injection volume of 5 μL (gradient: 0.01 M KH_2PO_4 , pH 2.3 (Solvent A), methanol (Solvent B): 40 % B to 85 % B in 8 min, 85 % B for 5 min, 85% to 40 % B in 1 min, 40 % B for 2 min, flow: 1.5 mL/min, total time 16 min). NMR spectra were measured with a Bruker Avance 200 or Bruker Avance 400 and analyzed with MestReNova v6.0.2-5475 (Mestrelab Research S.L). Residual solvent peaks were used for calibration: ^1H -spectra: 2.55 ppm (DMSO-d_6), 7.27 ppm (CDCl_3). ^{13}C -spectra: 39.51 ppm (DMSO-d_6), 77.00 ppm (CDCl_3). Chemical shifts (δ) are reported in parts per million (ppm), the solvent used is indicated in the experimental section.

5-(3-Chlorophenyl)-pent-4-enoic acid (1)



8,7 mL of a sodium methanolate solution (30 % wt in methanol 46,5 mmol; 2 eq.) were added dropwise to a stirred suspension of 10 g (3-carboxypropyl)triphenylphosphonium bromide (23,29 mmol; 1 eq.) 10 mL dry methanol at rt. After that, the reaction mixture was heated under reflux for further 30 min, before 4,9 g 3-chlorebenzaldehyde (35,94 mmol, 1,5 eq.) were added dropwise and heating to reflux continued overnight. Upon consumption of the starting material, the resulting reddish suspension was allowed to cool to room temperature, poured in 100 mL water, strongly acidified with conc. Aqueous hydrochloric acid, and extracted several times with EtOAc. The combined organic phases were dried over sodium sulfate and the volatiles removed *in vacuo*. The pure product was obtained by flash chromatography (EtOAc/hexanes/ HCOOH 85/15+1%) as yellowish oil (3,74 g, 76 %)

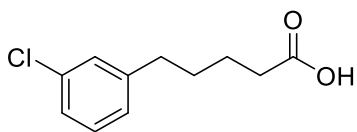
HPLC: 7,59 min

ESI-MS: m/z for $\text{C}_{11}\text{H}_{11}\text{ClO}_2$ $[\text{M}-\text{H}]^- = 208,8$; $\text{M} = 210,66$

Melting point: liquid at rt

^1H -NMR: (200 MHz, DMSO) $\delta = 12.34$ (s, 1H), 7.47 – 7.14 (m, 4H), 6.47 – 6.31 (m, 2H), 2.46 – 2.31 (m, 4H). ppm

5-(3-Chlorophenyl)-pentanoic acid (2)



4,14 g **1** (19,66 mmol; 1 eq.) were dissolved in 20 mL ethyl acetate, 0,208 g palladium on activated coal (10 % wt 0,17 mmol; 0,01 eq.) was added and the resulting suspension thoroughly stirred under hydrogen atmosphere upon complete conversion (TLC-control). The reaction mixture was filtrated over celite and the volatiles removed *in vacuo*. The product was obtained as pale yellow oil and used without further purification. (3,81 g, 91 %)

HPLC: 6,65 min

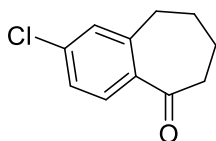
ESI-MS: m/z for $C_{11}H_{13}ClO_2$ $[M-H]^- = 210,8$; $M = 212,67$

Melting point: liquid at rt

IR: 2943, 2855, 1699, 1573, 1403, 1253, 1194, 1078, 911, 778, 699, 541 cm^{-1}

1H -NMR: (200 MHz, DMSO) $\delta = 12.26$ (s, 1H), 7.39 – 7.09 (m, 4H), 2.58 (t, $J = 7.1$ Hz, 2H), 2.22 (t, $J = 6.9$ Hz, 2H), 1.67 – 1.30 (m, 4H) ppm.

2-Chloro-6,7,8,9-tetrahydro-5H-benzo[7]annulen-5-on (3)



6,16 g **2** (28,97 mmol; 1 eq.) and a few drops DMF were dissolved in 30 mL DCM and 4,41 g oxalyl chloride (34,77 mmol; 1,2 eq.) added dropwise whilst stirring at room temperature. Upon completion of the activation indicated by ceasing of gas evolution, the reaction mixture was cooled to 0 °C, 11,53 g aluminium-(III)-chloride (86,92 mmol; 3 eq.) were added in portions and the suspension stirred for another 20-30 minutes. The reaction was quenched by pouring on a mixture of ice and 6N aqueous hydrochloric acid and extracted several times with DCM. The combined organic layers were washed with 5 % aqueous sodium hydroxide solution, dried over sodium sulfate and the volatiles removed *in vacuo*. The product was obtained as pale yellow oil (3,69 g, 60 %) after flash-chromatography (hexanes/DCM 80/20->40/60 within 1 h)

HPLC: 7,51 min (90,65 %)

ESI-MS: m/z for $C_{11}H_{11}ClO$ n. d.; $M = 194,66$

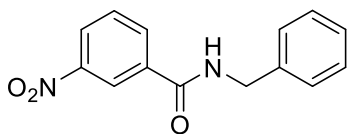
Melting point: liquid at rt

IR: 2938, 2855, 1727, 1677, 1590, 1452, 1282, 1261, 1215, 1086, 961, 816, 782, 749 cm^{-1}

1H -NMR: (200 MHz, DMSO) $\delta = 7.59$ (d, $J = 8.0$ Hz, 1H), 7.44 – 7.34 (m, 2H), 2.99 – 2.80 (m, 2H), 2.75 – 2.55 (m, 2H), 1.71 (dq, $J = 22.7, 6.3$ Hz, 4H) ppm.

^{13}C -NMR: (50 MHz, DMSO) $\delta = 203.85$ (s), 143.69 (s), 137.04 (s), 136.76 (s), 130.07 (s), 129.47 (s), 126.57 (s), 40.12 (s), 31.07 (s), 24.48 (s), 20.12 (s) ppm.

N-Benzyl-3-nitrobenzamide (4a)



200 mg 3-nitrobenzoic acid (1,2 mmol, 1 eq.) and 213 mg CDI (1,32 mmol; 1,1 eq) were dissolved in 5 mL dry THF and stirred until the gas evolution ceased. Following, 128 mg benzylamine (1,2 mmol; 1 eq.) was added and the mixture was stirred at rt until total consumption of the starting material (TLC-control), poured in water and extracted several times with DCM. The combined organic phases were dried over sodium sulfate, filtered, and the volatiles removed *in vacuo*. The remaining off-white solid was used without further purification. (297 mg, 97 %)

HPLC: 6,05 min (97,86 %)

ESI-MS: m/z for $C_{13}H_{12}N_2O_3$ $[M+Na]^+ = 279,0$; $[M+Na+MeOH]^+ = 311,1$; $[M-H]^- = 254,9$ $M = 256,26$

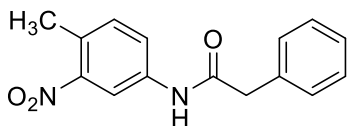
Melting point: 51,9 °C

IR: 3296, 3084, 3029, 1639, 1523, 1444, 1348, 1310, 1048, 932, 836, 740, 657, 611 cm^{-1}

General procedure for preparing 3-nitroanilides

To a stirred solution of the carboxylic acid and catalytic amounts of DMF was dropwise added oxalyl chloride or thionyl chloride and stirring continued at rt until the gas evolution ceased and TLC indicated no remaining starting material. Alternatively, a commercially available carboxylic acid chloride was used. After that, a solution of the respective aniline in THF and triethylamine (4 mmol, 1 eq.) were added and continued in stirring until TLC indicated no further conversion. The reaction mixture was poured in water and extracted several times with EtOAc. The combined organic phases were dried over sodium sulfate, filtered, and the volatiles removed *in vacuo*. The remaining was washed with ice cold ether to obtain the pure title compounds.

N-(4-Methyl-3-nitro-phenyl)-2-phenyl-acetamide (4b)



The title compound was synthesized according to the general procedure above from 750 mg phenylacetic acid (5,51 mmol, 1 eq.), 1,05 g oxalyl chloride (8,26 mmol; 1,5 eq.), 838 mg 4-methyl-3-nitroanilin (5,51 mmol; 1 eq.) and 557mg triethylamine (5,51 mmol; 1eq.) The title compound was obtained as yellowish solid (64 %).

HPLC: 6,89 min (99,55 %)

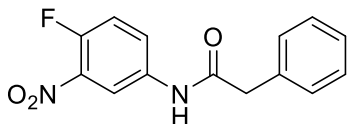
ESI-MS: m/z for $C_{15}H_{14}N_2O_3$ $[M+Na]^+ = 293,1$; $[M-H]^- = 268,9$; $M = 270,29$

Melting point: 125,2 °C

IR: 3234, 315, 3109, 3046, 2972, 2926, 1656, 1606, 1515, 1336, 1286, 1136, 891, 824, 728, 512

¹H-NMR: (200 MHz, DMSO-*d*6) δ = 10.53 (s, 1H), 8.36 (d, *J* = 2.1 Hz, 1H), 7.72 (dd, *J* = 8.3, 2.2 Hz, 1H), 7.40 (d, *J* = 8.4 Hz, 1H), 7.32 – 7.08 (m, 5H), 3.65 (s, 2H), 2.43 (s, 3H) ppm.

N-(4-Fluoro-3-nitro-phenyl)-2-phenyl-acetamid (4c)



The title compound was synthesized according to the general procedure above from 500 mg phenylacetic acid (3,67 mmol; 1 eq.), 436 mg thionyl chloride (3,67 mmol; 1 eq.), 573 mg 4-fluoro-3-nitroanilin (3,67 mmol; 1 eq.) and 371 mg triethylamine (3,67 mmol; 1 eq.). The title compound was obtained as yellowish solid (80 %).

HPLC: 6,57 min (98,65 %)

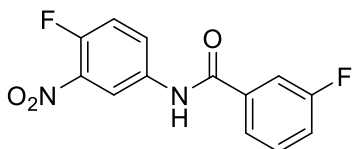
ESI-MS: *m/z* for C₁₄H₁₁FN₂O₃ [M+Na]⁺ = 297,1; [M-H]⁻ = 272,9; M = 274,25

Melting point: 131,5 °C

IR: 3242, 3192, 3138, 3071, 1652, 1606, 1532, 1494, 1403, 1340, 1249, 1140, 886, 828, 762, 724 cm⁻¹

¹H-NMR: (200 MHz, DMSO-*d*6) δ = 10.63 (s, 1H), 8.51 (dd, *J* = 6.9, 2.7 Hz, 1H), 7.86 (ddd, *J* = 9.0, 3.9, 2.9 Hz, 1H), 7.52 (dd, *J* = 11.1, 9.1 Hz, 1H), 7.38 – 7.16 (m, 5H), 3.66 (s, 2H) ppm.

N-(4-Fluoro-3-nitro-phenyl)-3-fluoro-benzamide (4d)



The title compound was synthesized according to the general procedure above from 500 mg 3-fluorobenzoic acid (3,57 mmol; 1 eq.), 467 mg thionyl chloride (3,93 mmol; 1,1 eq.), 557 mg 4-fluoro-3-nitroanilin (3,57 mmol; 1 eq.) and 361 mg triethylamine (3,57 mmol; 1 eq.). The title compound was obtained as yellowish solid (54 %).

Yield: 534,4 mg = 53,8 %

HPLC: 7,29 min (89,22 %)

ESI-MS: *m/z* for C₁₃H₈F₂N₂O₃ [M-H]⁻ = 276,9; M = 278,21

Melting point: 166,0 °C

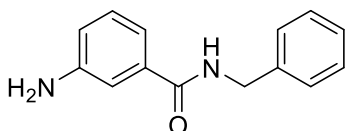
IR: 3321, 3300, 3067, 1656, 1527, 1340, 1219, 886, 841, 803, 749, 512 cm⁻¹

¹H-NMR: (200 MHz, DMSO-*d*6) δ = 10.74 (s, 1H), 8.69 (dd, *J* = 6.9, 2.6 Hz, 1H), 8.22 – 8.06 (m, 1H), 7.82 (m, 2H), 7.69 – 7.37 (m, 3H) ppm.

Reduction of nitroarenes

The respective nitroaryl was dissolved in ethyl acetate, palladium on activated coal (10 % wt, 0,01-0,03 eq.) was added and the resulting suspension thoroughly stirred under hydrogen atmosphere upon complete conversion (TLC-control). The reaction mixture was filtrated over celite and the volatiles removed *in vacuo*. The yielded solid was, depending on its purity, whether used without further purification or purified as indicated.

N-Benzyl-3-aminobenzamid (5a)



The title compound was synthesized according to the general procedure above from 690 mg **4a** (2,69 mmol) and obtained as white solid after washing with hexanes/ether (9/1) (93 %)

HPLC: 3,29 min (99,36 %)

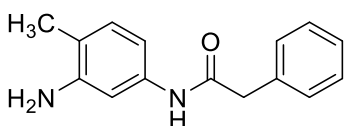
ESI-MS: m/z for $C_{14}H_{14}N_2O$ $[M+Na]^+ = 249,0$; $[M-H]^- = 224,9$; $M = 226,28$

Melting point: 95,9 °C

IR: 3438, 3342, 3288, 3026, 1631, 1573, 1536, 1486, 2323, 1265, 990, 870, 687, 504 cm^{-1}

1H -NMR: (200 MHz, DMSO-*d*₆) $\delta = 8.81$ (t, $J = 5.9$ Hz, 1H), 7.40 – 7.18 (m, 5H), 7.15 – 6.97 (m, 3H), 6.76 – 6.64 (m, 1H), 5.25 (s, 2H), 4.44 (d, $J = 6.0$ Hz, 2H) ppm.

N-(2-Methyl-5-amino-phenyl)-2-phenyl-acetamide (5b)



The title compound was synthesized according to the general procedure above from 870 mg **4b** (3,22 mmol) and obtained as beige solid after washing with ice cold ether (89 %)

HPLC: 3,41 min (100 %)

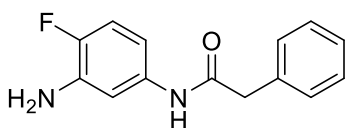
ESI-MS: m/z for $C_{15}H_{16}N_2O$ $[M+Na]^+ = 263,1$; $[M-H]^- = 238,9$; $M = 240,31$

Melting point: 150,3 °C

IR: 3396, 3288, 3055, 3026, 2963, 1673, 1602, 1540, 1511, 1444, 1415, 1344, 1319, 874, 816, 707, 553, 487 cm^{-1}

1H -NMR: (200 MHz, DMSO-*d*₆) $\delta = 9.78$ (s, 1H), 7.37 – 7.14 (m, 5H), 6.93 (d, $J = 1.8$ Hz, 1H), 6.79 (d, $J = 8.0$ Hz, 1H), 6.66 (dd, $J = 8.0, 2.0$ Hz, 1H), 4.80 (s, 2H), 3.35 (s, 2H), 1.98 (s, 3H) ppm.

N-(3-Amino-4-fluorophenyl)-2-phenyl-acetamide (5c)



The title compound was synthesized according to the general procedure above from 760 mg **4c** (2,77 mmol) and obtained as brown solid after washing with ice cold ether (33 %)

HPLC: 4,33 min (98,40 %)

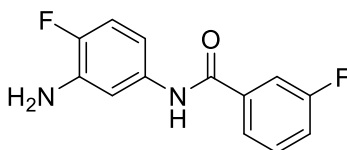
ESI-MS: m/z for $C_{14}H_{13}FN_2O$ $[M+Na]^+ = 267,1$; $[M-H]^- = 242,9$; $M = 244,27$

Melting point: 188,0 °C

IR: 3392, 3275, 3250, 3059, 1669, 1619, 1556, 1507, 1436, 1240, 1194, 866, 807, 762, 699, 562 cm^{-1}

1H -NMR: (200 MHz, DMSO-*d*6) $\delta = 9.89$ (s, 1H), 7.37 – 7.20 (m, 5H), 7.10 (dd, $J = 8.4$, 1.9 Hz, 1H), 6.86 (dd, $J = 10.9$, 9.2 Hz, 1H), 6.72 – 6.61 (m, 1H), 5.12 (s, 2H), 3.57 (s, 2H) ppm.

N-(3-Amino-4-fluoro-phenyl)-3-fluor-benzamide (5d)



The title compound was synthesized according to the general procedure above from 497 mg **4d** (1,79 mmol) and obtained as ochre solid after purification by flash chromatography (hexanes/EtOAc 70/30) (35 %)

HPLC: 4,99 min

ESI-MS: m/z for $C_{13}H_{10}F_2N_2O$ $[M+Na]^+ = 271,0$; $[M+Na+MeOH]^+ = 303,0$; $M = 248,23$

Neg. Mode: m/z $[M-H]^- = 246,9$ (ber.: 247,1)

Melting point: 103,2 °C

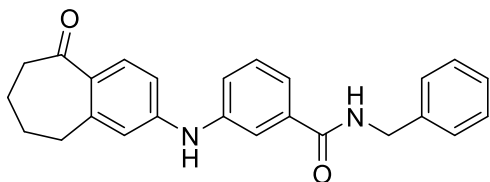
1H -NMR: (200 MHz, DMSO-*d*6) $\delta = 10.08$ (s, 1H), 7.92 – 7.66 (m, 2H), 7.67 – 7.48 (m, 1H), 7.42 (td, $J = 8.7$, 2.5 Hz, 1H), 7.29 (dd, $J = 8.5$, 2.3 Hz, 1H), 7.05 – 6.92 (m, 1H), 6.91 – 6.72 (m, 1H), 5.19 (s, 2H) ppm.

Buchwald-Hartwig-crosscoupling

An argon flushed round bottom flask was charged with chlorobenzosuberone (1,00-1,05 eq), respective arylamine (1,0 eq), caesium carbonate (3 eq), XPhos (0,5 eq) and palladium(II)-acetate (0,1 eq) and equipped with a reflux condenser. The mixture of starting materials was suspended in 10 mL of a mixture of 1,4-dioxane and tert-butanol (4/1) and heated to reflux for about 30 minutes whilst stirring. After consumption of the arylamine (TLC control), the reaction mixture was poured in ammonium chloride solution (aq., sat.) and

extracted three times with EtOAc. The combined organic phases were dried over sodium sulfate and the volatiles removed *in vacuo*. The remaining solid was purified by flash chromatography as indicated.

N-((Phenyl)-methyl)-3-[(5-oxo-5H-6,7,8,9-tetrahydro-benzo[7]annulen-2-yl)amino]-benzamide (12)



According to the general procedure above **compound 12** was synthesized from 81 mg **3** (0,42 mmol; 1,05 eq.), 90 mg **5a** (0,40 mmol; 1 eq.), 95 mg XPhos (0,20 mmol; 0,25 eq.), 9 mg palladium(II)acetate (0,04 mmol; 0,1 eq.) and 390 mg cesium carbonate (1,19 mmol; 3 eq.) and obtained after flash-chromatography (DCM/EA/nHEX 30:15:55 to 30:60:10 within 1,5h) as light yellow solid (56 %).

HPLC: 7,81 min (96,37 %)

ESI-MS: m/z for $C_{25}H_{24}N_2O_2$ $[M+Na]^+ = 407,1$; $[M-H]^- = 383,0$; $M = 384,48$

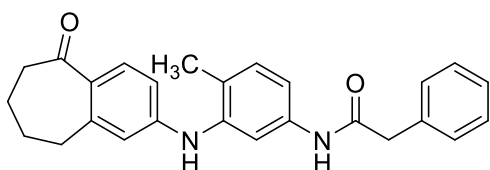
Melting point: 168,4 °C

IR: 3408, 3283, 3059, 2930, 2859, 1647, 1573, 1535, 1481, 1314, 1094, 873, 732, 690, 594, 511 cm^{-1}

1H -NMR: (400 MHz, DMSO-*d*₆) $\delta = 9.03$ (t, $J = 6.0$ Hz, 1H), 8.82 (s, 1H), 7.72 – 7.67 (m, 1H), 7.60 (d, $J = 8.5$ Hz, 1H), 7.50 – 7.45 (m, 1H), 7.40 (t, $J = 7.8$ Hz, 1H), 7.36 – 7.29 (m, 5H), 7.27 – 7.21 (m, 1H), 6.97 (dd, $J = 8.6, 2.3$ Hz, 1H), 6.88 (d, $J = 2.2$ Hz, 1H), 4.47 (d, $J = 6.0$ Hz, 2H), 2.90 – 2.80 (m, 2H), 2.68 – 2.59 (m, 2H), 1.85 – 1.73 (m, 2H), 1.73 – 1.63 (m, 2H) ppm.

^{13}C -NMR: (50 MHz, DMSO-*d*₆) $\delta = 201.91$ (s), 166.15 (s), 147.41 (s), 144.21 (s), 141.76 (s), 139.68 (s), 135.64 (s), 130.62 (s), 129.28 (s), 129.14 (s), 128.24 (s, 2C), 127.14 (s, 2C), 126.70 (s), 121.46 (s), 120.05 (s), 117.84 (s), 115.45 (s), 112.76 (s), 42.59 (s), 40.19 (s), 32.18 (s), 24.63 (s), 20.23 (s) ppm.

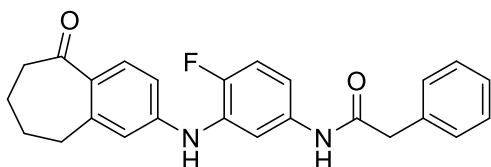
N-(((3-((5-Oxo-6,7,8,9-tetrahydro-5H-benzo[7]annulen-2-yl)-amino)-4-methyl-phenyl)-2-phenyl)-acetamide (73)



According to the general procedure above **compound 73** was synthesized from 100 mg **3** (0,51 mmol; 1 eq.), 123 mg **5b** (0,51 mmol; 1 eq.), 122 mg XPhos (0,26 mmol; 0,5 eq.), 12 mg palladium(II)acetate (0,05 mmol; 0,1 eq.) und 502 mg cesium carbonate (1,54 mmol; 3eq.) and obtained after flash chromatography (DCM/hexanes/EtOAc 3:5:2) as yellow solid (45 %).

HPLC:	9,59 min (96,94 %)
ESI-MS:	m/z for $C_{26}H_{26}N_2O_2$ $[M+Na]^+ = 421,1$; $[M-H]^- = 397,0$; $M = 398,51$
Melting point:	93,8 °C
IR:	3292, 2926, 2855, 1652, 1581, 1519, 1490, 1340, 1319, 1273, 1107, 820, 695, 449 cm^{-1}
1H -NMR:	(400 MHz, DMSO- d_6) $\delta = 10.10$ (s, 1H), 8.05 (s, 1H), 7.59 (s, 1H), 7.55 (d, $J = 8.5$ Hz, 1H), 7.36 – 7.28 (m, 4H), 7.25 (d, $J = 7.5$ Hz, 2H), 7.15 (d, $J = 8.2$ Hz, 1H), 6.72 – 6.66 (m, 1H), 6.65 – 6.61 (m, 1H), 3.60 (s, 2H), 2.86 – 2.75 (m, 2H), 2.64 – 2.58 (m, 2H), 2.13 (s, 3H), 1.81 – 1.71 (m, 2H), 1.71 – 1.59 (m, 2H) ppm.
^{13}C -NMR:	(50 MHz, DMSO- d_6) $\delta = 201.68$ (s), 168.88 (s), 149.34 (s), 144.17 (s), 139.35 (s), 137.76 (s), 136.00 (s), 130.98 (s), 130.58 (s), 129.05 (s, 2C), 128.25 (s, 2C), 127.92 (s), 126.47 (s), 126.25 (s), 114.83 (s), 114.55 (s), 113.71 (s), 111.87 (s), 43.31 (s), 40.19 (s), 32.28 (s), 24.59 (s), 20.24 (s), 17.31 (s) ppm.

N-(((3-((5-Oxo-6,7,8,9-tetrahydro-5H-benzo[7]annulen-2-yl)-amino)-4-fluoro)-phenyl)-2-phenyl)-acetamide (100)



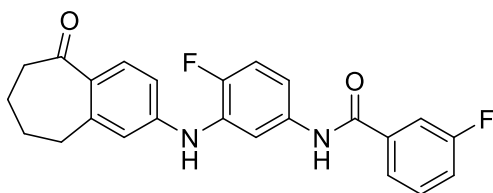
According to the general procedure above **compound 100** was synthesized from 100 mg **3** (0,51 mmol; 1 eq.), 125 mg **5c** (0,51 mmol; 1 eq.), 122 mg XPhos (0,26 mmol; 0,5 eq.), 12 mg palladium(II)acetat (0,05 mmol; 0,1eq.) and 502 mg cesium carbonate (1,54 mmol; 3eq.) and obtained after flash-chromatography (DCM/hexanes/EtOAc 3:5:2) as light brown solid (116 mg, 56 %).

HPLC:	7,61 min (95,73 %)
ESI-MS:	m/z for $C_{25}H_{23}FN_2O_2$ $[M+Na]^+ = 425,1$; $[M-H]^- = 401,0$; $M = 402,47$
Melting point:	65,1 °C
IR:	3280, 2926, 2855, 1652, 1577, 1490, 1344, 1244, 1203, 1107, 803, 687, 445 cm^{-1}
1H -NMR:	(400 MHz, DMSO- d_6) $\delta = 10.19$ (s, 1H), 8.47 (s, 1H), 7.73 (dd, $J = 7.6, 2.2$ Hz, 1H), 7.57 (d, $J = 8.5$ Hz, 1H), 7.35 – 7.29 (m, 4H), 7.28 – 7.22 (m, 2H), 7.22 – 7.15 (m, 1H), 6.82 (dd, $J = 8.6, 1.4$ Hz, 1H), 6.77 – 6.73 (m, 1H), 3.62 (s, 2H),

2.89 – 2.77 (m, 2H), 2.68 – 2.56 (m, 2H), 1.82 – 1.72 (m, 2H), 1.71 – 1.62 (m, 2H). ppm.

¹³C-NMR: (101 MHz, DMSO) δ = 202.01 (s), 169.02 (s), 150.49 (d, J = 240.7 Hz), 147.72 (s), 143.96 (s), 135.87 (s), 135.84 (d, J = 2.5 Hz), 130.36 (s), 129.05 (s), 128.79 (d, J = 12.4 Hz), 128.29 (s), 126.53 (s), 116.09 (d, J = 20.4 Hz), 115.45 (s), 114.18 (d, J = 7.1 Hz), 113.25 (s), 112.69 (s), 43.27 (s), 40.19 (s), 32.19 (s), 24.61 (s), 20.21 (s). ppm.

N-(((3-((5-Oxo-6,7,8,9-tetrahydro-5H-benzo[7]annulen-2-yl)-amino)-4-fluoro)-phenyl)-3-fluorobenzamide (109)



According to the general procedure above **compound 109** was synthesized from 74 mg **3** (0,38 mmol; 1,05 eq.), 90 mg **5d** (0,36 mmol; 1 eq.), 86 mg XPhos (0,18 mmol; 0,5 eq.), 8 mg Palladium(II)acetat (0,04 mmol; 0,1 eq.) und 364 mg cesium carbonate (1,10 mmol; 3 eq.) and obtained after flash chromatography (DCM/hexanes/EtOAc 30:55:15 to 30:10:60 within 1,5h) as light yellow solid (70 %).

HPLC: 8,16 min (98,74 %)

ESI-MS: m/z for C₂₄H₂₀F₂N₂O₂ [M+Na]⁺ = 429,1; [M-H]⁻ = 405,0; M = 406,43

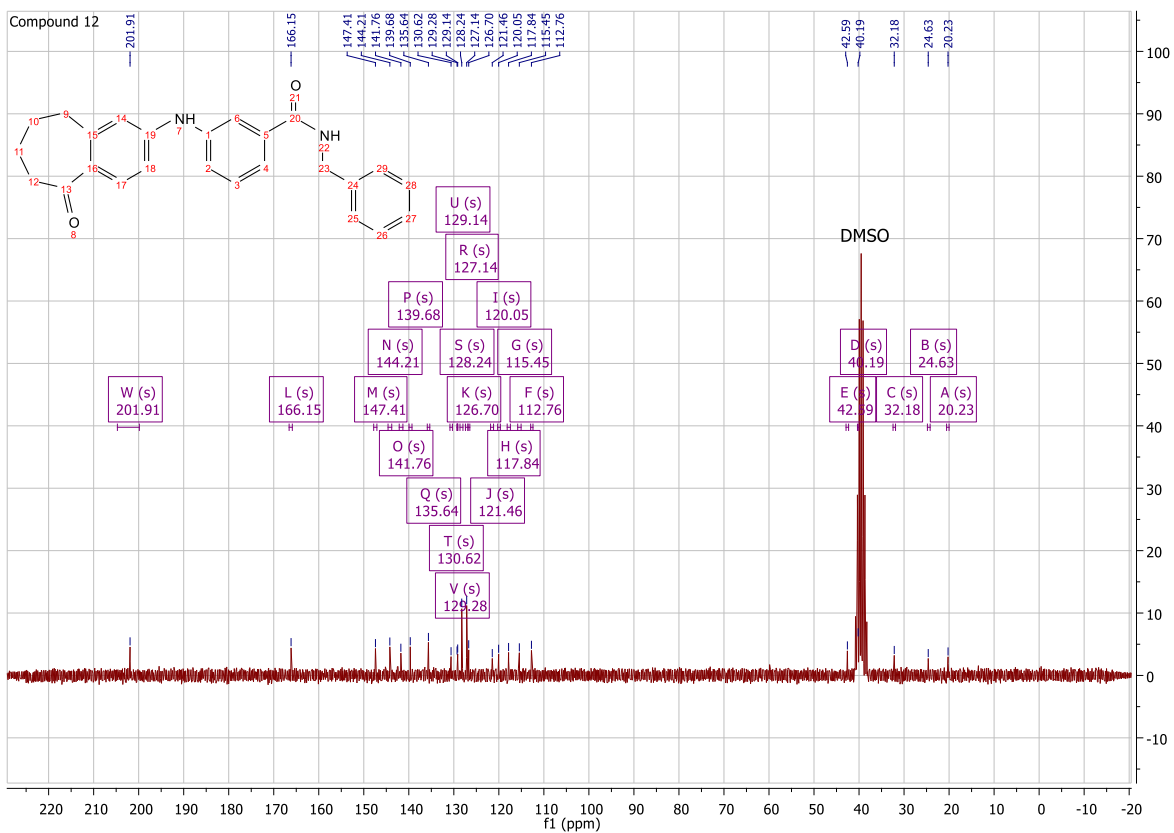
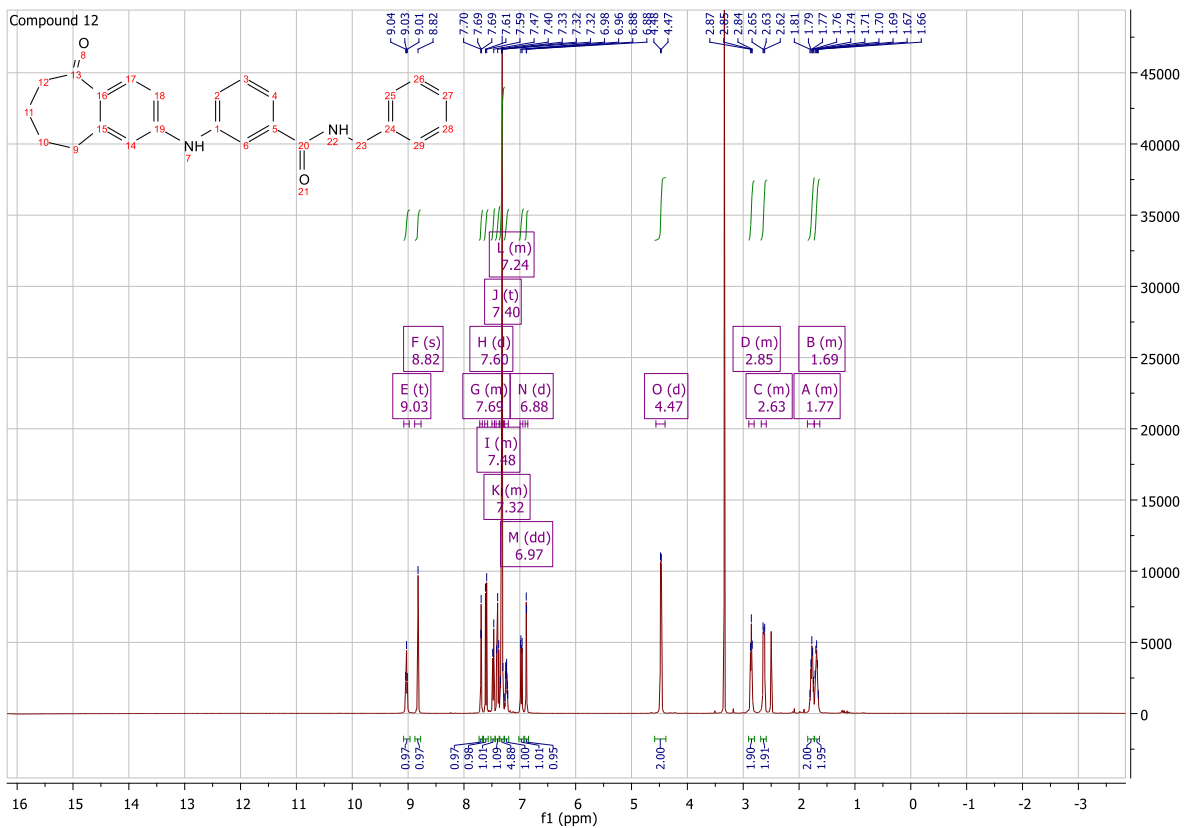
Melting point: 91,5 °C

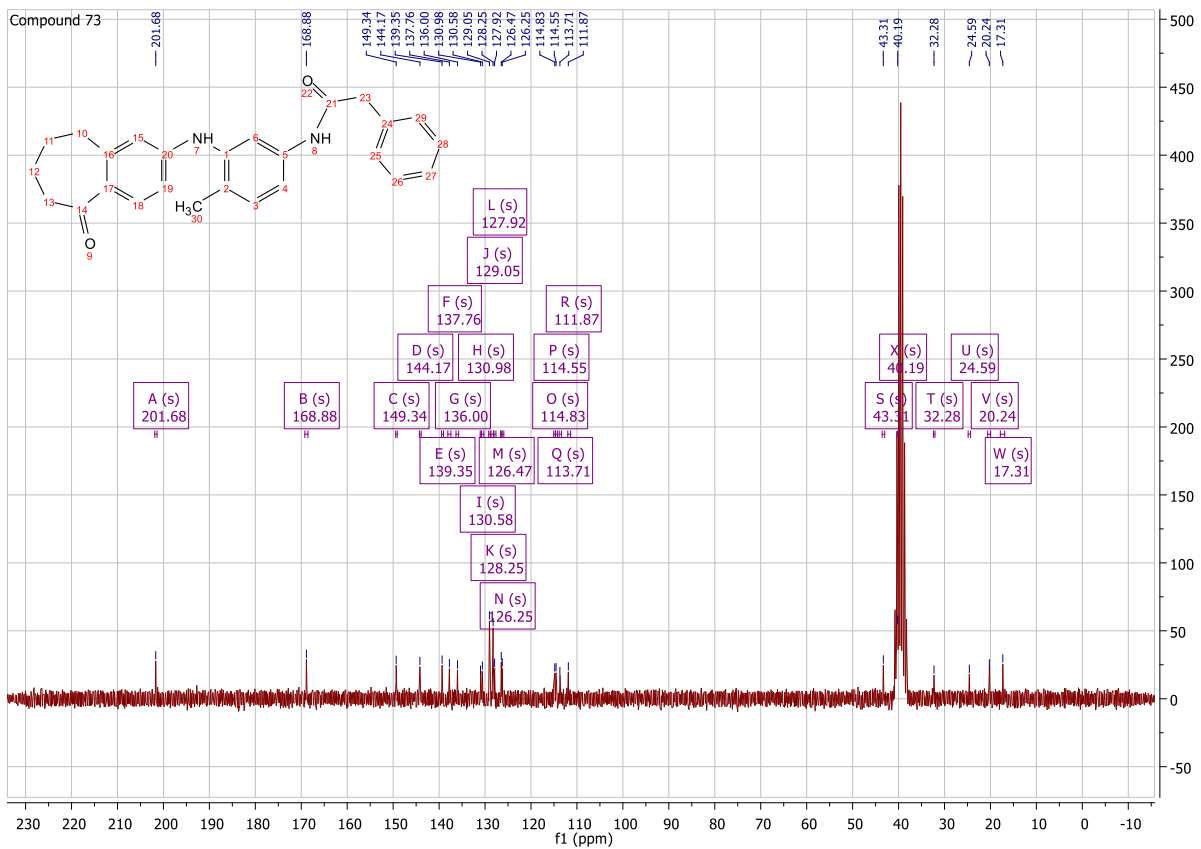
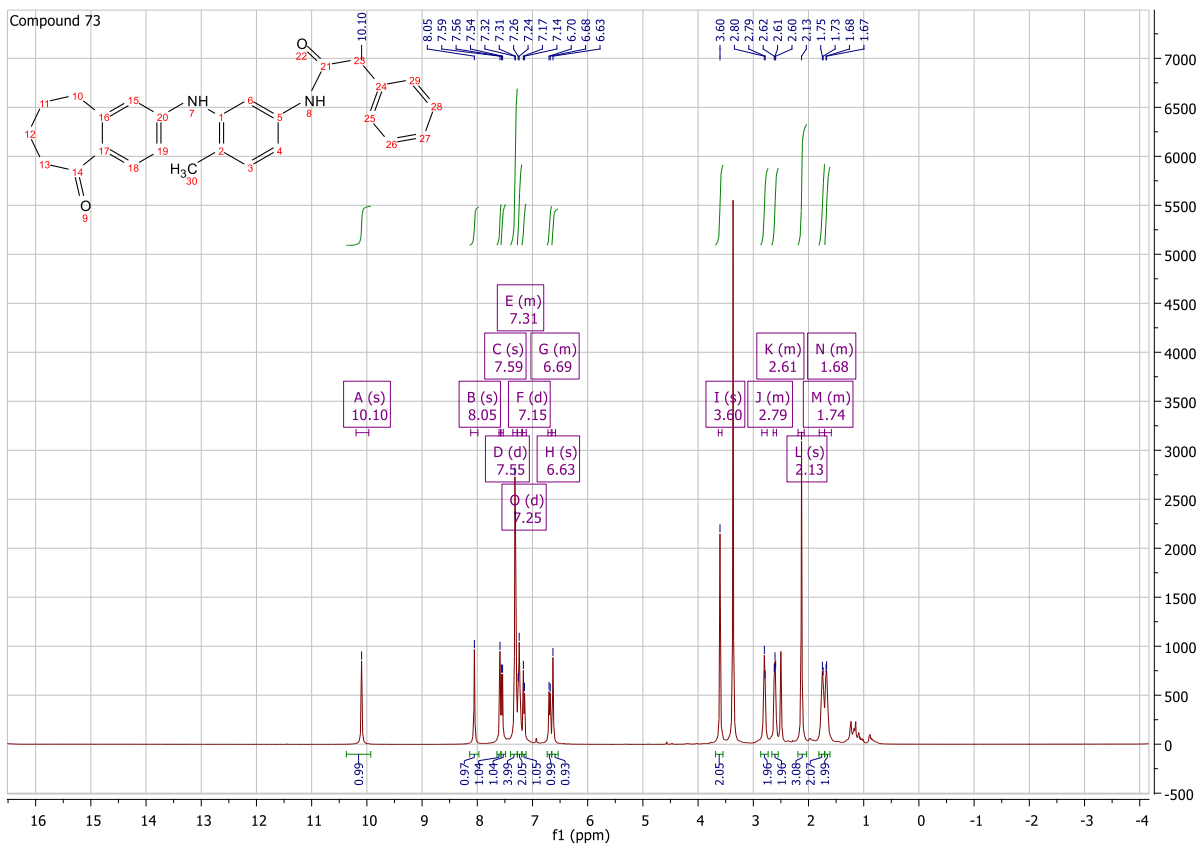
IR: 3296, 2934, 2859, 1644, 1590, 1523, 1436, 1469, 1249, 1095, 791, 741, 449 cm⁻¹

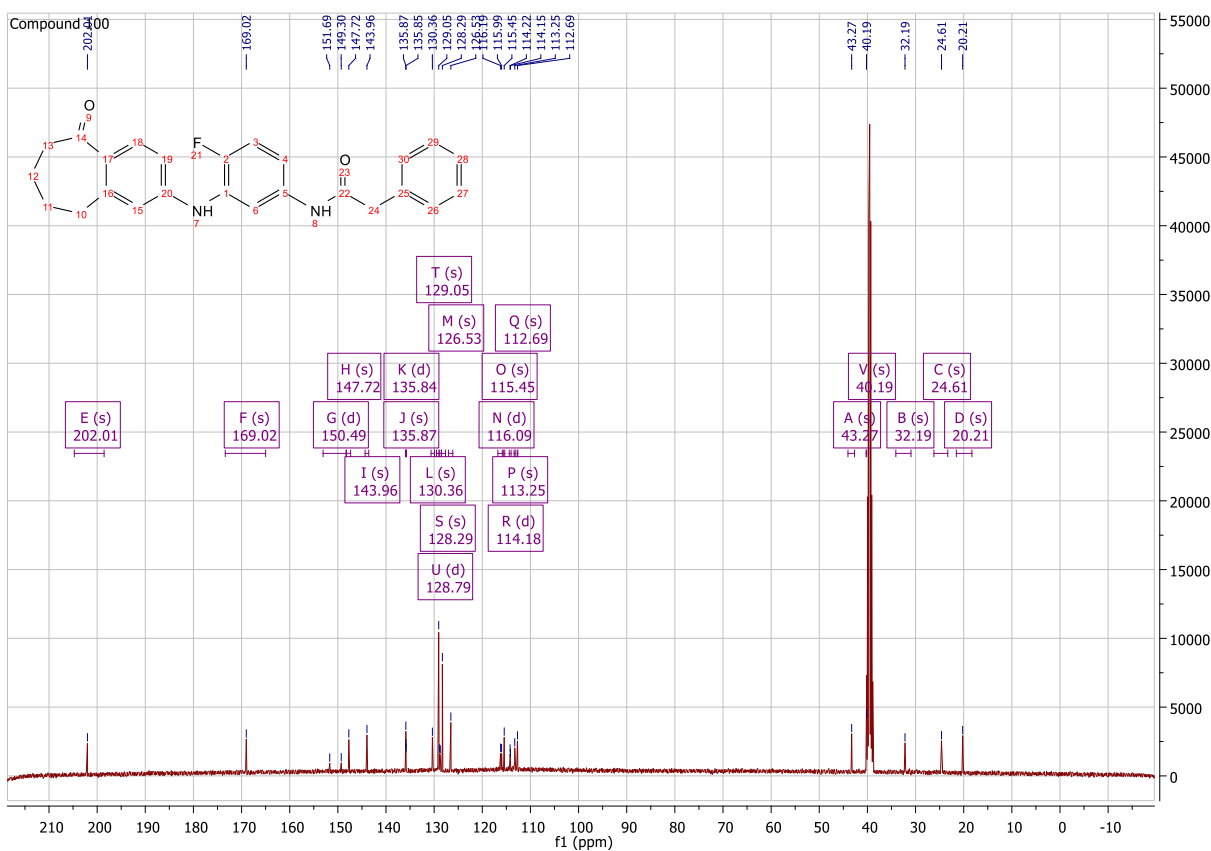
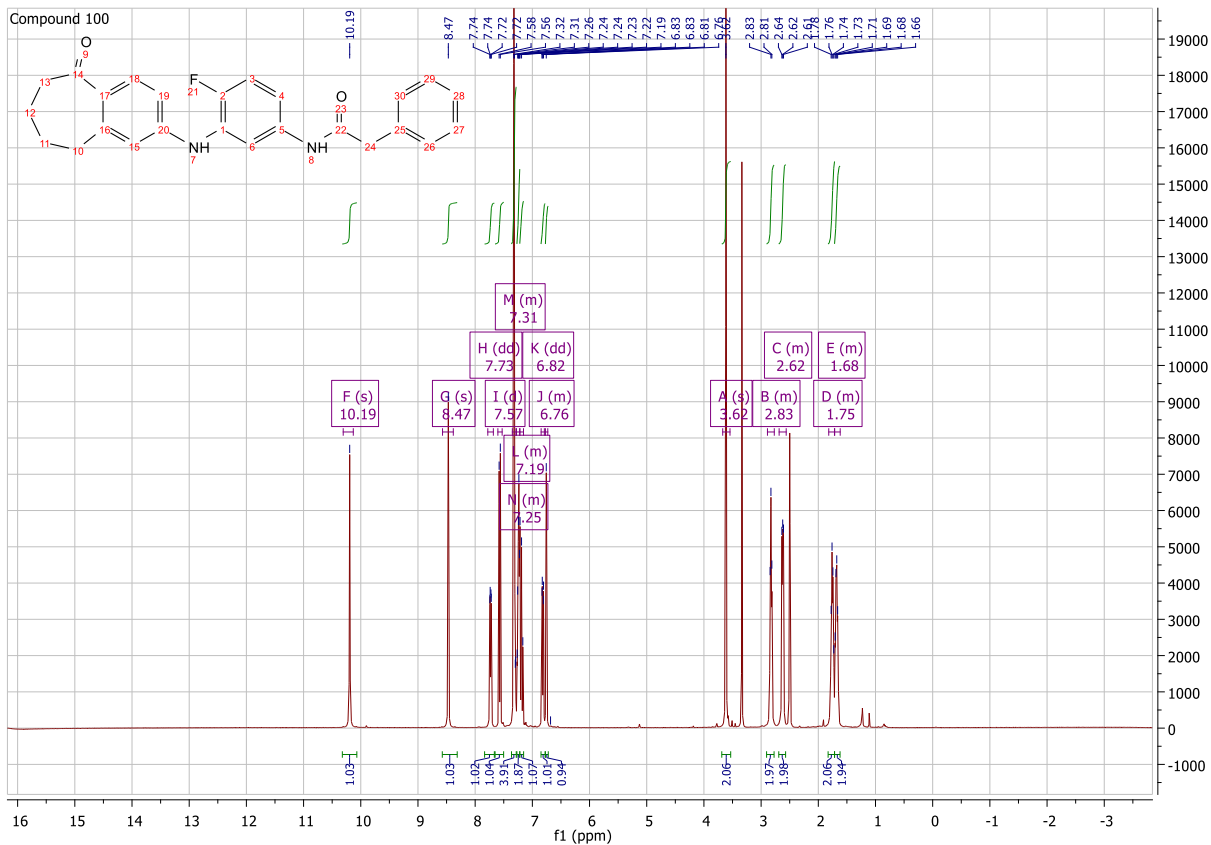
¹H-NMR: (200 MHz, DMSO-*d*₆) δ = 10.33 (s, 1H), 8.53 (s, 1H), 7.91 (dd, J = 7.7, 2.5 Hz, 1H), 7.82 – 7.77 (m, 1H), 7.78 – 7.72 (m, 1H), 7.63 – 7.55 (m, 2H), 7.50 – 7.41 (m, 2H), 7.27 (dd, J = 10.8, 8.9 Hz, 1H), 6.88 (dd, J = 8.5, 1.9 Hz, 1H), 6.82 (d, J = 2.0 Hz, 1H), 2.90 – 2.79 (m, 2H), 2.69 – 2.58 (m, 2H), 1.83 – 1.73 (m, 2H), 1.73 – 1.64 (m, 2H) ppm.

¹³C-NMR: (50 MHz, DMSO-*d*₆) δ = 202.06 (s), 164.10 (d, J = 2.5 Hz), 161.91 (d, J = 244.4 Hz), 150.87 (d, J = 241.4 Hz), 147.70 (s), 144.01 (s), 137.11 (d, J = 6.8 Hz), 135.51 (d, J = 2.6 Hz), 130.61 (d, J = 8.1 Hz), 130.41 (s), 129.11 (s), 128.80 (d, J = 12.6 Hz), 123.88 (d, J = 2.8 Hz), 118.52 (d, J = 21.2 Hz), 116.05 (d, J = 20.6 Hz), 115.51 (s), 115.40 (d, J = 7.1 Hz), 114.47 (d, J = 22.9 Hz), 114.40 (s), 112.73 (s), 40.21 (s), 32.20 (s), 24.63 (s), 20.24 (s) ppm

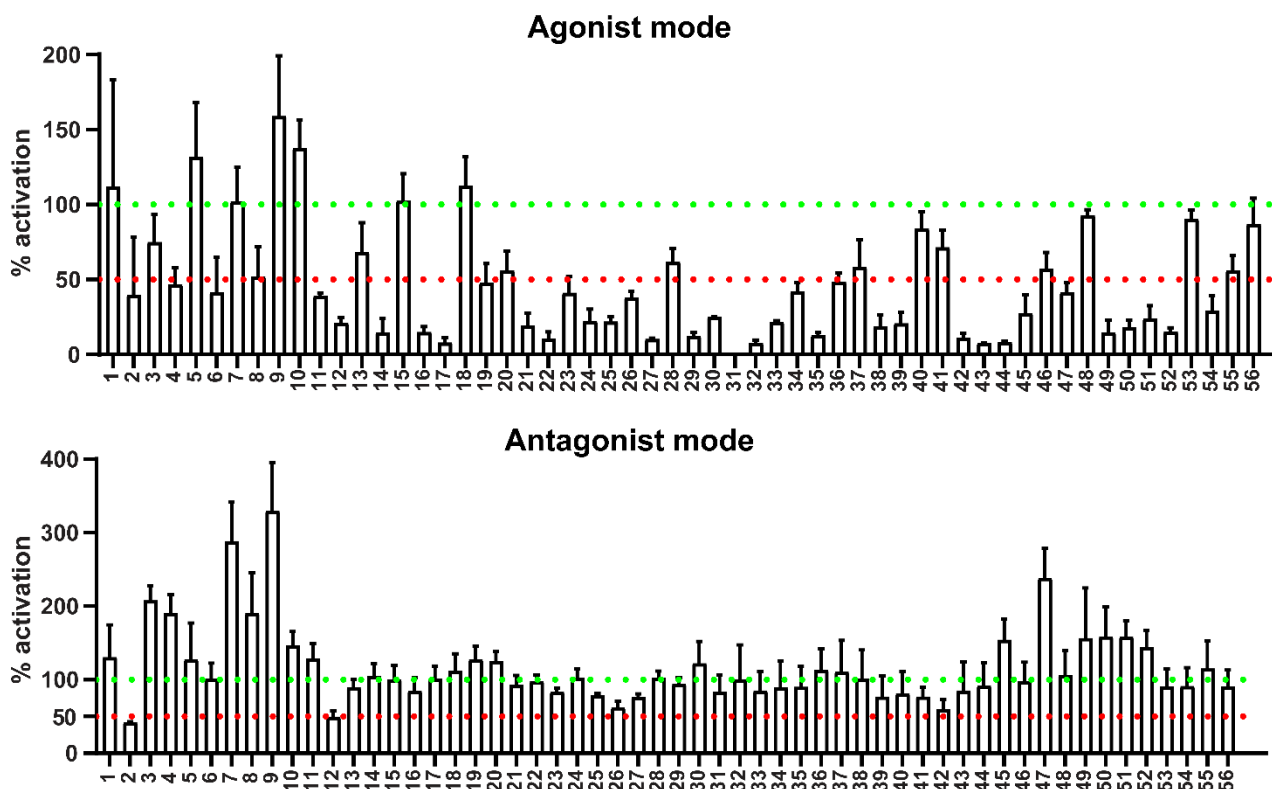
Experimental ¹H- and ¹³C-NMR spectra of compounds 12, 73, 100 and 109



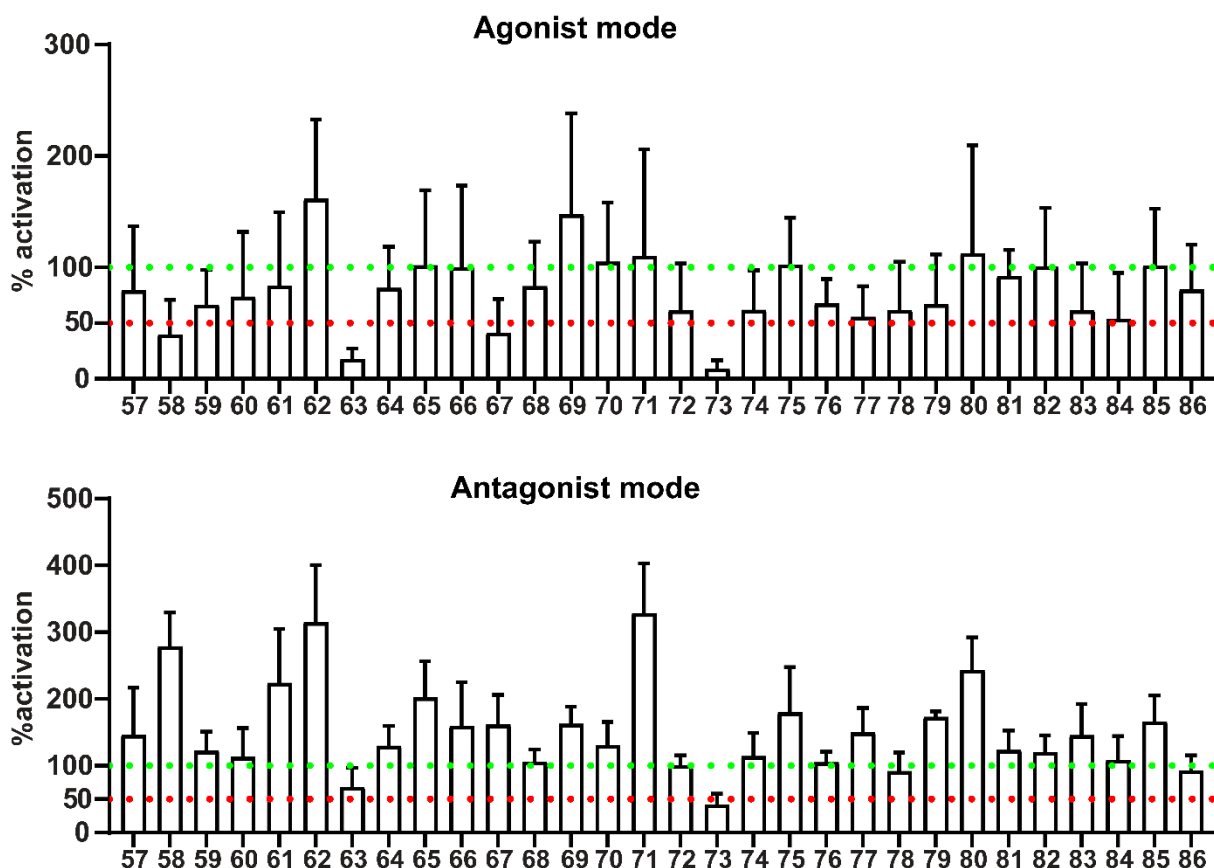




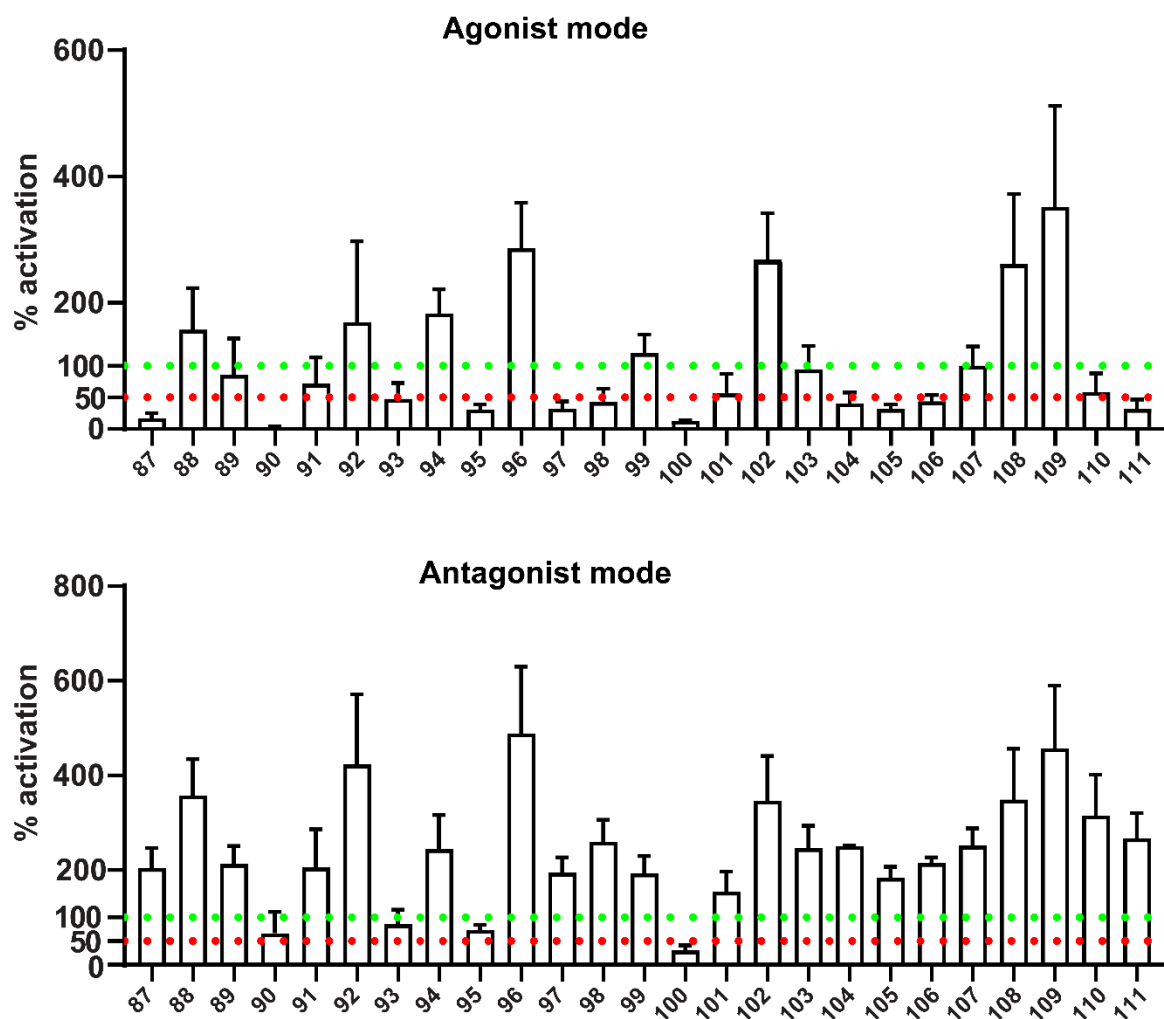
Supplementary Figures



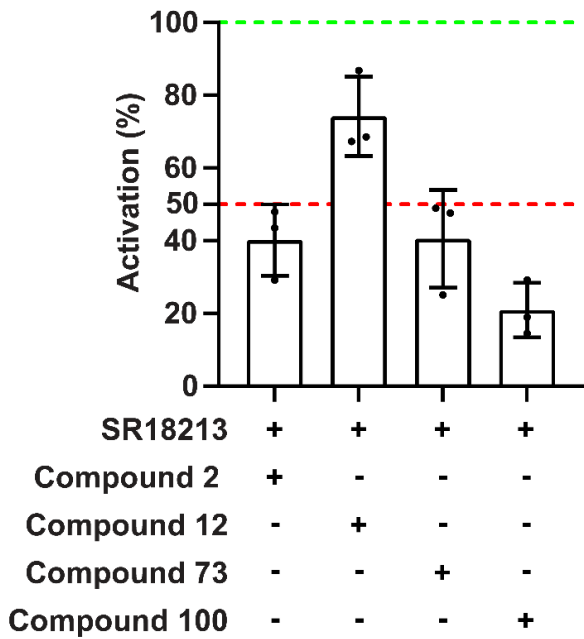
Supplementary Figure S1. Effects of 56 in silico screened TükIC compounds (A) alone or (B) in combination with 10 μM rifampicin on PXR-mediated transactivation of CYP3A4 reporter gene. H-P cells were transiently transfected with CYP3A4 reporter gene, treated with 0.1% (A) or 0.2% (B) DMSO, 10 μM rifampicin, 10 μM test compounds or co-treated with 10 μM rifampicin and 10 μM test compounds. Luciferase activities were measured after 24 h treatment. Data is expressed as mean \pm SD %. Activation was calculated according to Zhu et al. [38] from three independent experiments with technical triplicates. Fold induction achieved by 10 μM rifampicin was set as 100%. Green and red dashed lines represent the 100% and 50% activation, respectively.



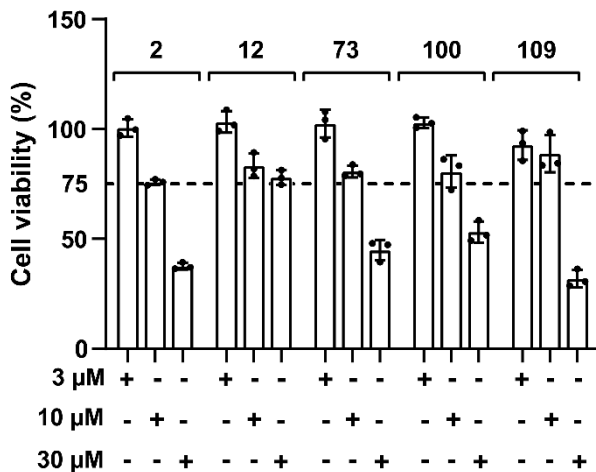
Supplementary Figure S2. Effects of first round structural analogues (A) alone or (B) in combination with rifampicin on PXR-mediated transactivation of CYP3A4 reporter gene. H-P cells were transiently transfected with CYP3A4 reporter gene, treated with 0.1% (A) or 0.2% (B) DMSO, 10 μ M rifampicin, 10 μ M of test compounds or co-treated with 10 μ M rifampicin and 10 μ M test compounds. Luciferase activities were measured after 24 h treatment. Data is expressed as mean \pm SD %. Activation was calculated according to Zhu et al. [38] from three independent experiments with technical triplicates. Fold induction achieved by 10 μ M rifampicin was set as 100%. Green and red dashed lines represent the 100% and 50% activation, respectively.



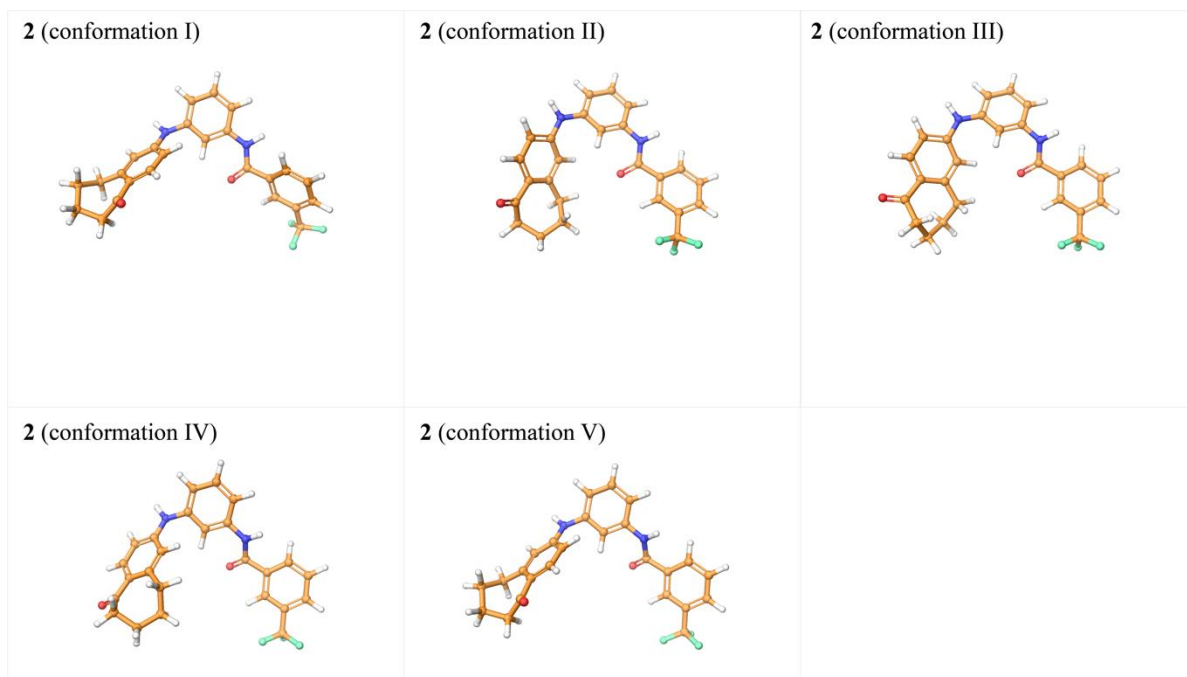
Supplementary Figure S3. Effects of second round structural analogues (A) alone or (B) in combination with rifampicin on PXR-mediated transactivation of CYP3A4 reporter gene. H-P cells were transiently transfected with CYP3A4 reporter gene, treated with 0.1% (A) or 0.2% (B) DMSO, 10 μ M rifampicin, 10 μ M of test compounds or co-treated with 10 μ M rifampicin and 10 μ M test compounds. Luciferase activities were measured after 24 h treatment. Data is expressed as mean \pm SD %. Activation was calculated according to Zhu et al. [38] from three independent experiments with technical triplicates. Fold induction achieved by 10 μ M rifampicin was set as 100%. Green and red dashed lines represent the 100% and 50% activation, respectively.



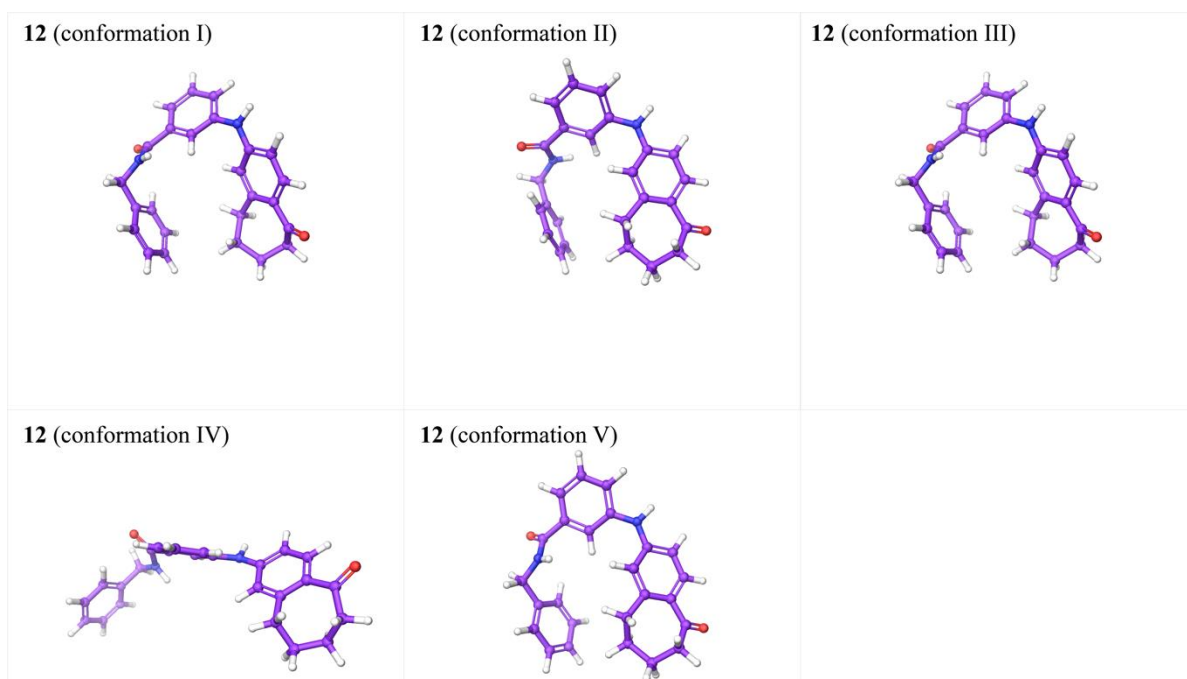
Supplementary Figure S4. Effects of potential novel PXR antagonists in combination with 1 μ M SR18213 on PXR-mediated transactivation of CYP3A4 reporter gene. H-P cells were transiently transfected with CYP3A4 reporter gene, treated with 0.2% DMSO, and co-treated with 1 μ M SR18213 and 10 μ M test compounds. Luciferase activities were measured after 24 h treatment. Data is expressed as mean \pm SD %. Activation was calculated according to Zhu et al. [38] from three independent experiments with technical triplicates individual experiments illustrated with dots. Fold induction achieved by 1 μ M SR12813 was set as 100%. Green and red dashed lines represent the 100% and 50% activation, respectively.



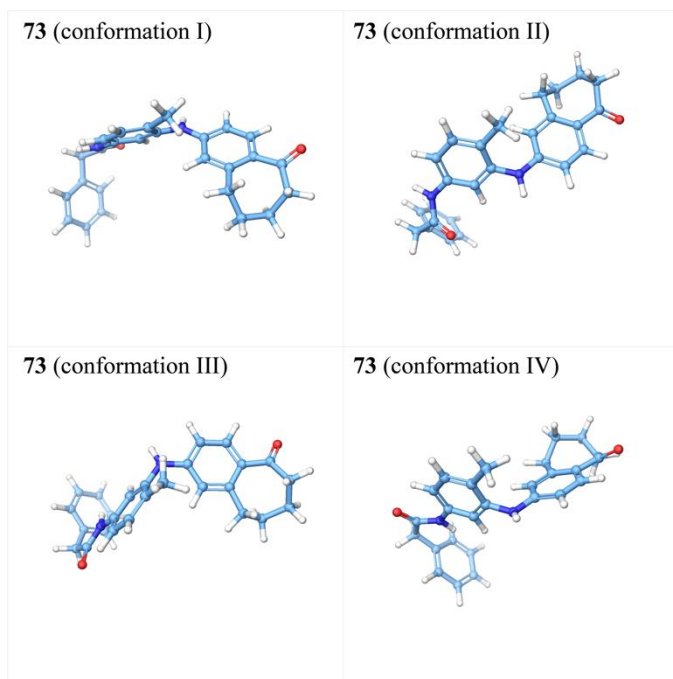
Supplementary Figure S5. Cell viability of HepG2 cells following 24 h treatment with potential novel PXR ligands. HepG2 cells were seeded, and on following day treated with 0.1% DMSO, 3, 10 and 30 μ M of test compounds. Cell viabilities were measured after 24 h incubation with treatments using CellTiter-Glo assay. Cell viability in the presence of vehicle DMSO only was set as 100%. Results are expressed as mean \pm SD from three independent experiments with technical triplicates and individual experiments illustrated with dots.



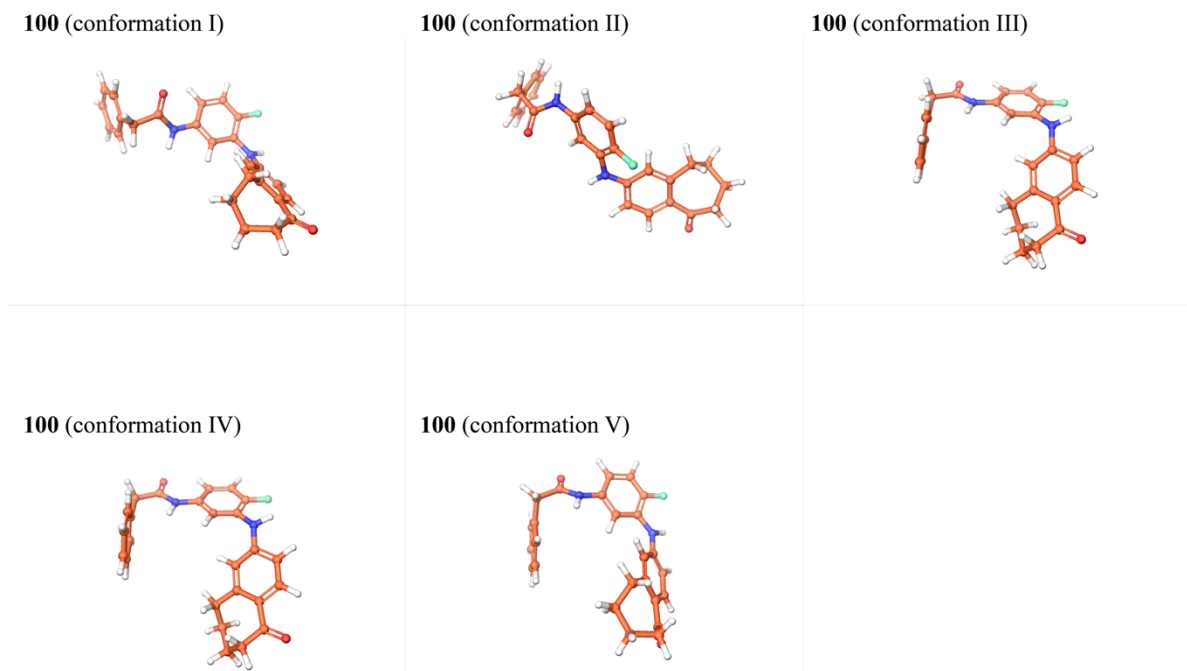
Supplementary Figure S6. QM Conformer & Tautomer Predictor output conformations of compound **2**.



Supplementary Figure S7. QM Conformer & Tautomer Predictor output conformations of compound **12**.

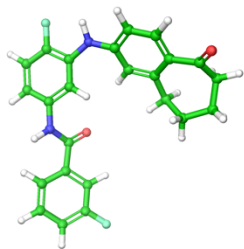


Supplementary Figure S8. QM Conformer & Tautomer Predictor output conformations of compound **73**.

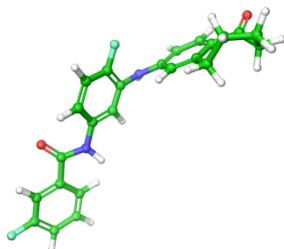


Supplementary Figure S9. QM Conformer & Tautomer Predictor output conformations of compound **100**.

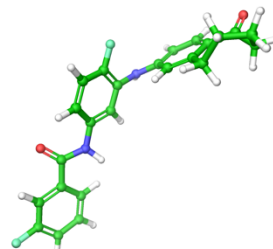
109 (conformation I)



109 (conformation II)



109 (conformation III)



Supplementary Figure S10. QM Conformer & Tautomer Predictor output conformations of compound 109.

Supplementary Tables

Supplementary Table S1. QM Conformer & Tautomer Predictor output conformations and their energies. See SI Figures S6–S10 for the structures.

Compound	Solution phase energy	Boltzmann Population	Conformation extended/folded
2 (conformation I)	-1524.787992	63.152	(Extended ^a)
2 (conformation II)	-1524.786778	17.450	(Folded ^a)
2 (conformation III)	-1524.786163	9.093	(Folded ^a)
2 (conformation IV)	-1524.785900	6.886	(Folded ^a)
2 (conformation V)	-1524.785239	3.419	(Extended ^a)
12 (conformation I)	-1227.019389	27.161	Folded
12 (conformation II)	-1227.019368	26.585	Folded
12 (conformation III)	-1227.019349	26.040	Folded
12 (conformation IV)	-1227.018569	11.406	Extended
12 (conformation V)	-1227.018325	8.808	Folded
73 (conformation I)	-1266.331113	46.333	Extended
73 (conformation II)	-1266.330934	38.341	Extended
73 (conformation III)	-1266.329904	12.872	Extended
73 (conformation IV)	-1266.328339	2.454	Extended
100 (conformation I)	-1326.266051	61.882	Extended
100 (conformation II)	-1326.265065	21.787	Extended
100 (conformation III)	-1326.264220	8.896	Folded
100 (conformation IV)	-1326.263458	3.970	Folded
100 (conformation V)	-1326.263329	3.466	Folded
109 (conformation I)	-1386.203087	38.610	(Folded ^a)
109 (conformation II)	-1386.202939	32.988	(Extended ^a)
109 (conformation III)	-1386.202797	28.402	(Extended ^a)

^aCompounds **2** and **109** have shorter R₂-group which renders their conformational space different from the other compounds.

Supplementary Table S2. PXR activating protein kinase inhibitors.

Protein kinase inhibitor	Reference
Brigatinib	(FDA, 2017) = ref. [65]
Dabrafenib	(Creusot et al., 2020) = ref. [15]
Encorafenib	(FDA, 2018) = ref. [63]
Erlotinib	(Harmsen et al., 2013) = ref. [13]
Gefinitib	(Harmsen et al., 2013) = ref. [13]
Lorlatinib	(FDA, 2018) = ref. [64]
Nilotinib	(Harmsen et al., 2013) = ref. [13]
Sorafenib	(Harmsen et al., 2013) = ref. [13]
Vandetanib	(Harmsen et al., 2013) = ref. [13]
Vemurafenib	(MacLeod et al., 2015) = ref. [67]

9. Publication III

When Two Become One: Conformational Changes in FXR/RXR Heterodimers Bound to Steroidal Antagonists

Alejandro Díaz-Holguín [c, d] A, Azam Rashidian [a, b], Dirk Pijnenburg e], Glaucio Monteiro Ferreira [f], Alzbeta Stefela [g], Miroslav Kaspar [h, i], Eva Kudova [h], Antti Poso [a, b, c], Rinie van Beuningen [e], Petr Pavek [g], Thales Kronenberger [a, b, c]

[a] A. Rashidian,+ Prof. A. Poso, Dr. T. Kronenberger Institute of Pharmacy, Pharmaceutical/Medicinal Chemistry and Tübingen Center for Academic Drug Discovery & Development (TüCAD2) Eberhard Karls University Tübingen Auf der Morgenstelle 8, 72076 Tübingen (Germany)

[b] A. Rashidian,+ Prof. A. Poso, Dr. T. Kronenberger Department of Internal Medicine VIII University Hospital of Tübingen Otfried-Müller-Strasse 14, 72076 Tübingen (Germany) E-mail: thales.kronenberger@uni-tuebingen.de

[c] A. Díaz-Holguín,+ Prof. A. Poso, Dr. T. Kronenberger School of Pharmacy, Faculty of Health Sciences University of Eastern Finland, Kuopio 70211 (Finland)

[d] A. Díaz-Holguín+ Current address: Science for Life Laboratory Department of Cell and Molecular Biology Uppsala University, 75124 Uppsala (Sweden)

[e] D. Pijnenburg, R. van Beuningen PamGene International B.V. Wolvenhoek 10, 5211HH, 's-Hertogenbosch (Netherlands)

[f] Dr. G. Monteiro Ferreira Department of Clinical and Toxicological Analyses School of Pharmaceutical Sciences University of São Paulo Av. Prof. Lineu Prestes 580, 05508-000, São Paulo (Brazil)

[g] A. Stefela, Prof. P. Pavek Department of Pharmacology and Toxicology Charles University Heyrovskeho 1203, 50005, Hradec Kralove (Czechia)

[h] Dr. M. Kaspar, Prof. E. Kudova Institute of Organic Chemistry and Biochemistry Czech Academy of Sciences Flemingovo náměstí 542/2, 16000, Prague (Czechia)

[i] Dr. M. Kaspar Faculty of Sciences, Charles University in Prague Albertov 6, Prague 2, 12843 (Czechia)

[+] These authors contributed equally to this work.

ChemMedChem Journal

Volume18, Issue4, 2023, e202200556, doi:

<https://doi.org/10.1002/cmdc.202200556>

Hypothesis: Our recently disclosed FXR antagonists, namely 2a and 2h, possess the ability to modulate the heterodimerization interface and the α AF-2 conformation.

Aim: Utilizing the long-timescale molecular dynamics simulations to identify the potential binding mode of these intriguing antagonists, focusing on both the heterodimerization interface and the α AF-2 conformation. This can open new possibilities for targeting FXR-related pathways with potential therapeutic implications.

Results: Distinct conformational changes were observed in different regions during our investigation. Specifically, we observed alterations in the FXR/RXR interface (α 9, α 10, and α 11) and in the α AF-2 (α 12– α 3 and L: α 11– α 12) regions, as well as changes in the LBP (ligand-binding pocket) geometry and interaction pattern. Examination of available FXR crystal structures supported these observations by revealing different orientations in loops L: α 1– α 2, L: α 5– α 6, and L: α 11– α 12 depending on the bound ligand. For instance, the CDCA crystal structure displayed an unfolded L: α 5– α 6 loop, and the binding of the partial agonist DM175 destabilized both L: α 5– α 6 and L: α 11– α 12 loops. Our simulations further demonstrated an extensive re-orientation of the LBP, driven by L: α 1– α 2 and L: α 5– α 6 interactions. This structural feature served as a key distinguishing factor between agonists and antagonists. Specifically, CDCA and GW4064 displayed smaller average α 5– α 6 distances promoted by interactions with α 5 residues. On the other hand, compounds 2a and 2h relied solely on α 1 and α 2 interactions. Further, the results from our Free Energy Perturbation (FEP) calculations aligned with previous experimental binding studies, confirming that 2h acted as a stronger antagonist compared to 2a but still displayed weaker binding compared to CDCA. Furthermore, an observation emerged in the context of FXR/RXR geometrical organization by the means of alteration in the dihedral planes and interaction pattern between residues at the heterodimerization interface.

In our antagonist simulations, we noted a destabilization of α 11, which was more pronounced in the monomeric simulations. This destabilization caused displacement in the L: α 11– α 12/ α AF-2 region.

Conclusion: The effects induced by antagonists on the heterodimerization interface appear to be closely linked to the destabilization of α AF-2, resulting in the prevention of both co-activator and co-repressor recruitment. Our findings present novel and valuable insights into the conformational dynamics of FXR, challenging the conventional understanding based solely on crystal structures. This suggests that a more comprehensive assessment of FXR antagonism/agonism shifts is required. To

obtain a more comprehensive and accurate depiction, longer timescales or alternative sampling approaches should be considered.



When Two Become One: Conformational Changes in FXR/RXR Heterodimers Bound to Steroidal Antagonists

Alejandro Díaz-Holguín^{+, [c, d]} Azam Rashidian^{+, [a, b]} Dirk Pijnenburg,^[e]
Glaucio Monteiro Ferreira,^[f] Alzbeta Stefela,^[g] Miroslav Kaspar,^[h, i] Eva Kudova,^[h]
Antti Poso,^[a, b, c] Rinie van Beuningen,^[e] Petr Pavek,^[g] and Thales Kronenberger^{*, [a, b, c]}

Farnesoid X receptor (FXR) is a nuclear receptor with an essential role in regulating bile acid synthesis and cholesterol homeostasis. FXR activation by agonists is explained by an α AF-2-trapping mechanism; however, antagonism mechanisms are diverse. We discuss microsecond molecular dynamics (MD) simulations investigating our recently reported FXR antagonists 2a and 2h. We study the antagonist-induced conformational changes in the FXR ligand-binding domain, when compared to the synthetic (GW4064) or steroidal (chenodeoxycholic acid, CDCA) FXR agonists in the FXR monomer or FXR/RXR hetero-

dimer *r*, and in the presence and absence of the coactivator. Our MD data suggest ligand-specific influence on conformations of different FXR-LBD regions, including the α 5/ α 6 region, α AF-2, and α 9-11. Changes in the heterodimerization interface induced by antagonists seem to be associated with α AF-2 destabilization, which prevents both co-activator and co-repressor recruitment. Our results provide new insights into the conformational behaviour of FXR, suggesting that FXR antagonism/agonism shift requires a deeper assessment than originally proposed by crystal structures.

Introduction

Farnesoid X receptor (FXR) is the main bile acid sensor in the liver and intestine, regulating the transcription of key genes involved in bile acid synthesis and enterohepatic transport.^[1] The FXR functions as a regulator of bile acid homeostasis and in protecting the liver from their toxic overload, but it is also involved in regulation of cholesterol, lipid, glucose and amino acid metabolism, and in controlling cholestasis, insulin sensitivity, and inflammation.^[2,3] Bile acids (BAs) are important signalling steroidal molecules that regulate glucose, lipid, and energy homeostasis via interactions with FXR, the most important bile acid-activated receptor.^[1] Chenodeoxycholic acid (3 α ,7 α -dihydroxy-5 β -cholan-24-oic acid, CDCA, Figure 1) is the most potent endogenous FXR ligand,^[4] while lithocholic acid (3 α -hydroxy-5 β -cholan-24-oic acid, LCA) and its taurine conjugate, tauro-lithocholic acid (TLCA), activate the G protein-coupled bile acid receptor 1 (GPBAR1) with the highest potency among natural BAs.^[1,4]

Huge efforts have been made to discover potent FXR agonists. Clinical trials^[1,5] on nonalcoholic fatty liver show improvements in the liver health, though adverse effects such as lipid accumulation has stimulated research on partial agonists. However, recent animal data attract our attention to the unexplored field of FXR antagonism and the development of efficient and selective FXR antagonists. Beneficial metabolic activities of natural, semisteroidal, or synthetic FXR antagonists in animal models of cholestasis, hypercholesterolemia, and type 2 diabetes mellitus (T2DM) or in pancreatic and colon cancers have been reported.^[6,7]

FXR is a transcription factor from the family of nuclear receptors (NRs). NRs have a conserved fold consisting of an N-terminal DNA-binding domain, which interacts with specific DNA elements, connected to the ligand-binding domain (LBD).

- [a] A. Rashidian,⁺ Prof. A. Poso, Dr. T. Kronenberger
Institute of Pharmacy, Pharmaceutical/Medicinal Chemistry and
Tübingen Center for Academic Drug Discovery & Development (TüCAD2)
Eberhard Karls University Tübingen
Auf der Morgenstelle 8, 72076 Tübingen (Germany)
- [b] A. Rashidian,⁺ Prof. A. Poso, Dr. T. Kronenberger
Department of Internal Medicine VIII
University Hospital of Tübingen
Otfried-Müller-Strasse 14, 72076 Tübingen (Germany)
E-mail: thales.kronenberger@uni-tuebingen.de
- [c] A. Díaz-Holguín,⁺ Prof. A. Poso, Dr. T. Kronenberger
School of Pharmacy, Faculty of Health Sciences
University of Eastern Finland, Kuopio 70211 (Finland)
- [d] A. Díaz-Holguín⁺
Current address: Science for Life Laboratory
Department of Cell and Molecular Biology
Uppsala University, 75124 Uppsala (Sweden)
- [e] D. Pijnenburg, R. van Beuningen
PamGene International B.V.
Wolvenhoek 10, 5211HH, 's-Hertogenbosch (Netherlands)
- [f] Dr. G. Monteiro Ferreira
Department of Clinical and Toxicological Analyses
School of Pharmaceutical Sciences
University of São Paulo
Av. Prof. Lineu Prestes 580, 05508-000, São Paulo (Brazil)
- [g] A. Stefela, Prof. P. Pavek
Department of Pharmacology and Toxicology
Charles University
Heyrovského 1203, 50005, Hradec Kralove (Czechia)
- [h] Dr. M. Kaspar, Prof. E. Kudova
Institute of Organic Chemistry and Biochemistry
Czech Academy of Sciences
Flemingovo náměstí 542/2, 16000, Prague (Czechia)
- [i] Dr. M. Kaspar
Faculty of Sciences, Charles University in Prague
Albertov 6, Prague 2, 12843 (Czechia)

[⁺] These authors contributed equally to this work.

Supporting information for this article is available on the WWW under
<https://doi.org/10.1002/cmdc.202200556>

© 2022 The Authors. ChemMedChem published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

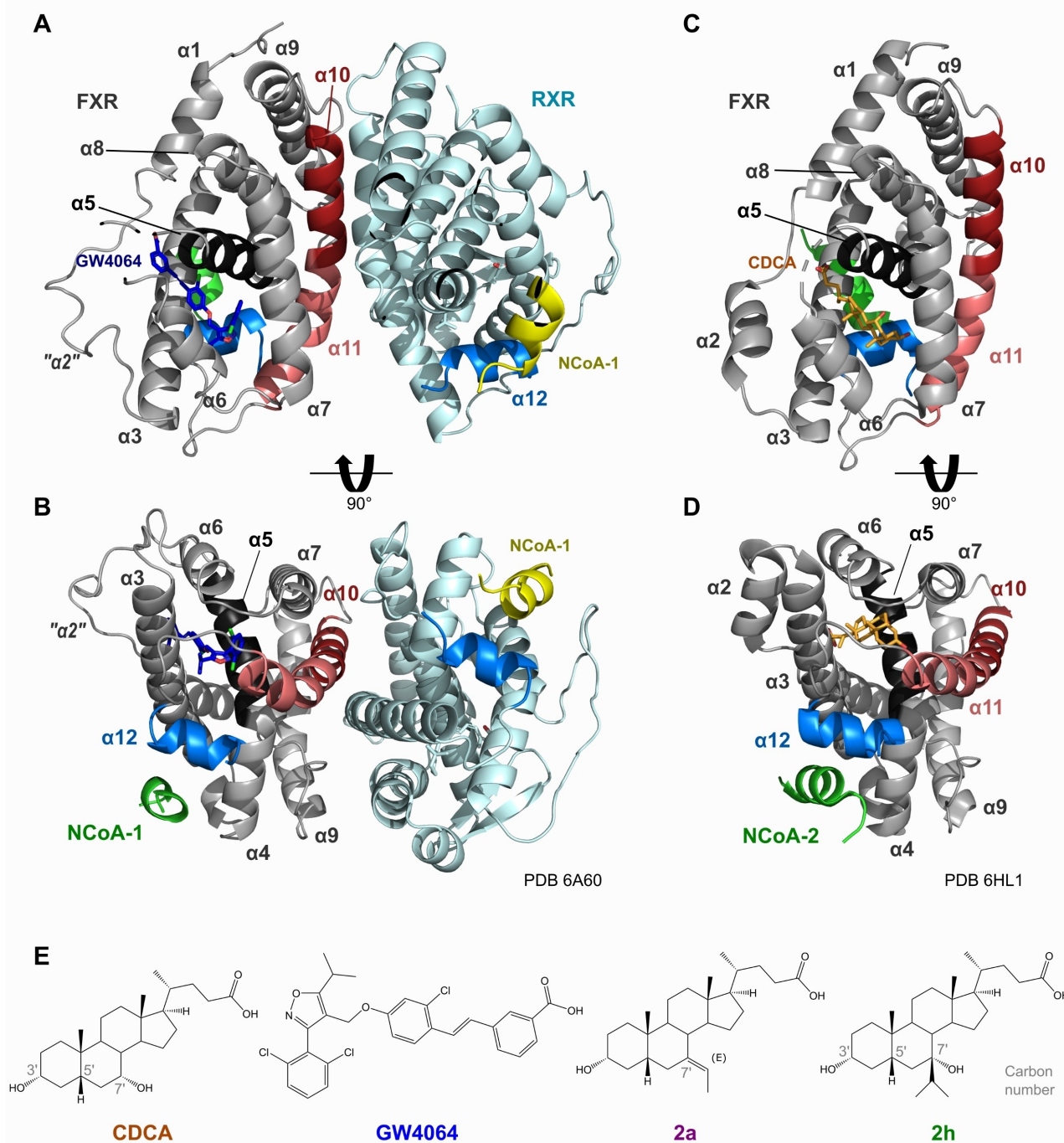


Figure 1. The general architecture of FXR structure. (A) lateral view and (B) top view of an FXR/RXR heterodimer, with FXR represented as grey and RXR as cyan. (C, lateral view) and (D, top view) of an FXR monomeric structure. In all representations, the $\alpha 5$ is displayed in black, $\alpha 10$ in dark red, $\alpha 11$ in light pink, $\alpha 12$ (or α AF-2) in blue and the associated coregulatory peptides in green. (E) chemical structure of FXR binders with CDCA and GW4064 as agonists and (E)-7-ethylidene-lithocholic acid (2a) and 7 β -isopropyl-CDCA (2h) as our studied antagonists. Standard numbering for relevant carbon's is provided in light grey.

The LBD is composed of a three-layered α helical sandwich fold, containing 12 α -helices ($\alpha 1$ to $\alpha 12$, Figure 1A–D). The LBD contains a large lipophilic pocket (ligand binding pocket, LBP), which can bind small ligands, triggering the transcriptional switch in the ligand-dependent modulation. It also displays the activation function-2 (α AF-2), a surface (composed of the

helices 3, 4, 5 and 12) that can recruit co-regulatory proteins. Those co-regulatory proteins can be coactivators or corepressors depending on the ligand interaction in the LBP.

A well-established model proposed to explain vitamin D receptor (VDR) modulation, which can be expanded for other nuclear receptors,^[8] called “folding-door model” focusing on the

LBD conformational changes. In this model, agonist binding stabilizes the $\alpha 12$ is close to $\alpha 3$ -helix, forming internal interactions with the $\alpha 11$ kink. These interactions close the door (the space between $\alpha 11$ and $\alpha 11$ - $\alpha 12$ loop). Alternatively, in presence of antagonists, the $\alpha 11$ - $\alpha 12$ destabilizes leading to an unsuitable α AF-2 positioning for receptor activation. All known ligand-dependent nuclear receptors require this α AF-2 region for coactivator recruitment^[9] Further, the coregulatory proteins mediate the interactions with the transcriptional machinery, resulting in transcription repression/activation.^[10] In its active state, FXR heterodimerizes with another nuclear receptor named retinoid X receptor- α (RXR α , NR2B1).^[11] RXR α , in its turn, is also activated by endogenous ligands, such as 9-cis-retinoic acid (9cRA). The FXR/RXR heterodimer binds to the FXR response element (FXRE), which is mostly a single inverted repeat (IR1) in the promoter region of the target gene.^[12] The work from Merk *et al.* (2019) establishes that FXR activation is a result of an equilibrium between conformational populations, using comparative crystallography and NMR-coregulator binding assays.^[13] Specifically, FXR agonists stabilize the folded and extended helix $\alpha 11$ (heterodimerization interface) and the $\alpha 11$ - $\alpha 12$ loop (composing the α AF-2 region) upon binding. The stabilized α AF-2 region enables coactivator recruitment and therefore FXR activation. In contrast, partial agonists lead to changes in the $\alpha 11$ conformation, which, consequently, destabilizes the $\alpha 11$ - $\alpha 12$ loop and the $\alpha 12$ orientation. Loops regions (L), besides the L: $\alpha 11$ - $\alpha 12$, such as L: $\alpha 1$ - $\alpha 2$ and L: $\alpha 5$ - $\alpha 6$ are differently orientated indicating a dependence on a bound ligand. For instance, the binding of CDCA results in L: $\alpha 5$ - $\alpha 6$ unfolding and binding of the partial agonist DM175 destabilises L: $\alpha 5$ - $\alpha 6$ as well as L: $\alpha 11$ - $\alpha 12$.^[13]

The coregulator binding and dimerization process, together with the fact that steroid receptor co-activators can bind to either FXR or RXR α or both, compose a regulatory mechanism. It is discussed for other NRs, that RXR α dimerization can change the co-recruitment process (enhancing it).^[14] However, it was not until recently that the different interaction motifs and co-regulatory proteins binding to different FXR-ligand combinations started to be discussed.^[15]

Classically, nuclear receptor antagonism has been structurally studied, mostly focusing on the receptor LBD. In the unliganded forms, $\alpha 12$ can assume either an extended conformation, pushed away from the LBD^[16] or adopt multiple conformations in solution.^[17] Typically, antagonist-binding can induce unfavourable $\alpha 12$ conformations for the recruitment of coregulatory proteins by directly interacting with the AF-2 inward pocket with bulkier residues, or by occupying the LBP without establishing the appropriate contacts for activation.^[18,19] This later can lead to perturbation of the $\alpha 11$ and L: $\alpha 11$ - $\alpha 12$, altering the $\alpha 12$ positioning ever so slightly. In FXR-antagonist crystal structures, such as bound to ivermectin (PDB 4WVD^[20]) and N-benzyl-N-(3-(tert-butyl)-4-hydroxyphenyl)-2,6-dichloro-4-(dimethylamino) benzamide (NDB, PDB 4OIV^[21]), the carboxy terminal of the α AF-2 region ($\alpha 12$ and L: $\alpha 11$ - $\alpha 12$) is invisible or the $\alpha 11$ is changed into a β -strand, which highlights this region's high flexibility. It is worth noticing, however, that ivermectin is much larger than other ligands, which distorts $\alpha 2$

and $\alpha 6$, making a poor model to describe the transition from a fully folded FXR-LBD towards a potential new antagonist binding mode. For other receptors, such as PXR, these large conformational changes were shown to be accessible, from the agonist-bound structures, by long timescale molecular dynamics simulations^[22] or directly by analyses of multiple crystals.

Recently, we identified the first-in-class moderately potent steroidal compounds with combined FXR antagonistic and GPCR1 agonistic activities, from a library of 7-alkylated derivatives of 3 α -hydroxy-5 β -cholan-24-oic acid (Figure 1E).^[23] We discovered compounds (E)-7-ethylidene-lithocholic acid (herein, **2a**) and 7 β -isopropyl-CDCA (**2h**) as a novel FXR antagonists with moderate IC₅₀ values against FXR [around 15.1 (**2a**) and 17.5 μ M (**2h**)] in cell-based reporter assays.

The compounds do not activate FXR and **2a** has been found as highly selective, not significantly interacting with other nuclear receptors. Though we generated preliminary models to explain the different ligand potencies, a more complete model discussing their potential binding mode remained to be published. Our work here expands on the potential binding mode of these antagonists and their binding implications both on the heterodimerization interface and on the α AF-2 conformation, using long-timescale molecular dynamics simulations to explain the dynamic behaviour of these components. We started from an agonist/apo structure fully folded FXR-LBD in order to model the initial conformational changes that would happen upon antagonist binding.

Results

Antagonist binding induces large conformational changes beyond the LBP

To investigate and compare the interactions and conformational dynamics of the novel antagonists and two classical agonists, altogether, we conducted a total of 150 μ s simulations. This simulation time was distributed among four FXR-LBD and FXR/RXR α LBDs in a heterodimeric complex, for each inhibitor (Figure 2A illustrates these distributions). Briefly, we simulated FXR-LBD in complex with either agonist (CDCA and GW4064) or antagonist (**2a** or **2h**), for around 15 μ s each (Figure 2A). Antagonist binding mode was proposed according to^[23] and further expanded in our methods section.

We have also simulated the coactivator peptides (herein named as CoA) originally co-crystallized with GW4064 (NCoA-1) and CDCA (NCoA-2) for at least 15 μ s each (Figure S1 displays the RMSD value variation along the merged trajectory). First, to gain a better understanding of the FXR-LBD dynamics, we conducted a principal component analysis (PCA) on the FXR backbone's trajectory to identify the most prominent correlated motions of the protein. This analysis was generated from the merged trajectory of FXR-LBD backbone with all the ligands, from both monomer and heterodimer simulations. PCA revealed that the first and the second principal components (PCs) describe together 26.1% of the data (PC1 14.8% and PC2 11.3%) (Figure 2A, B, respectively).

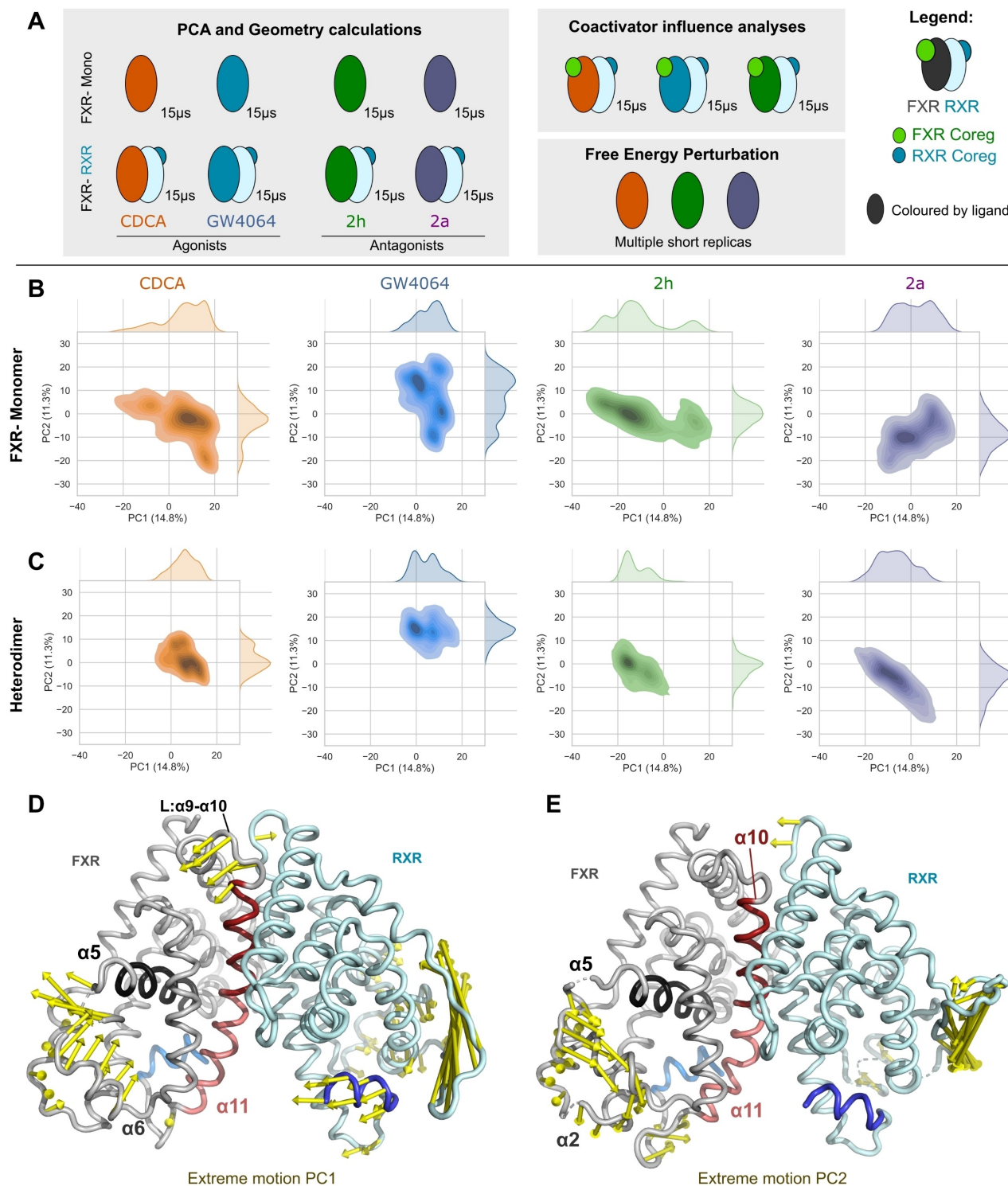


Figure 2. A) Overview of our experimental design depicting all the simulated systems, and in which analyses they were included. FXR-Mono stands for monomeric simulations, FXR/RXR – stands for heterodimer systems without coactivator in the FXR and the last systems represent FXR/RXR heterodimer with coactivators on both receptors. Principal component analysis (PCA) score plot of PC1 and PC2 based on the variance of FXR's backbones in both monomeric simulations (B) and heterodimeric without CoA (C). Each system of the joint PCA is depicted separately: CDCA (orange), GW4064 (blue), 2h (green) and 2a (purple). The extreme movements of PC1 (D) and PC2 (E). Protein is colour coded as in Figure 1. Extreme movements related to each principal component are highlighted with yellow arrows indicating the direction of the movement. RXR extreme motions were calculated only from heterodimeric trajectories (without co-activators) and plotted together to illustrate their movement amplitude.

PC1 scores for the monomeric simulations are distributed in a wider range than the dimeric simulations, with the scores for the simulations with activators CDCA and GW4064 populating the values higher than zero, while the scores for the simulations with antagonists **2a** and **2h** display negative values (Figure 2B). PC1 score values mainly separate between characteristics of agonists and antagonist. Interestingly, monomeric simulations of **2h** display the widest distribution range, suggesting a mixed behaviour, which is not reproduced in the heterodimer systems (Figure 2C). PC1 extreme motions in FXR display the $\alpha 6$ approximation towards the $\alpha 5$ (Figure 2D, in black), concomitantly with the distancing of the $\alpha 9$ - $\alpha 10$ loop from the RXR interface. Smaller contributions of α AF-2 ($\alpha 12$) displacement are also observed, which agrees with the higher RMSF values for antagonists (Figure S2). Agonist-bound dimer simulations displayed a lower median RMSF value in the $\alpha 11$ region when compared to their monomeric counterparts. However, monomerization does not seem to be the only driving force in the conformational changes of antagonistic-bound systems, since $\alpha 11$ also displayed high RMSF values independently.

Surprisingly, PC1 extreme motion on RXR displays an even larger $\alpha 12$ displacement, which was not further investigated in this work. On the other hand, the PC2 score distribution separates the steroid ligands (CDCA, **2a** and **2h**), with the mode distributions populating the negative values, from GW4064. The respective PC2 extreme motion (Figure 2E) displays the $\alpha 2$ displacement and reorganization of $\alpha 6$.

To further study the long-timescale dynamics from a statistical/probabilistic point of view, we conducted a Markov state modelling (MSM) analysis along the merged trajectory of FXR monomers/heterodimers. This allows identification of relevant metastable states and their probability distribution, even if we cannot directly backtrack the conformational changes to the specific simulated systems. MSM revealed five metastable states (S_{1-5}) for the concatenated FXR systems (Supporting Information Figure S3,4). The two most dominant metastable states, S_4 and S_5 , appear with $\sim 29\%$ and $\sim 33\%$ equilibrium probabilities, respectively. Other states (S_1 , S_2 , S_3) display lower probabilities in the range of 12–13%.

Our MSM suggested conformational differences in regions relevant for FXR activation when comparing the metastable states and the original crystal structure (Figures S3 and S4) with a partial agonist (ivermectin; PDB ID 4WVD, Figure S5A). We observed that α AF-2- $\alpha 3$ distances are farther from each other when compared to the original crystals, in three metastates S_2 (12.7%), S_3 (11.9%) and S_5 (33.3%), with a total of 57.9% probability of existing in this conformation. These same states also display conformational shifts in the $\alpha 10/\alpha 11$ (Figure S3), which is consistent with changes induced by partial agonists, such as ivermectin (Figure S5). In contrast, the S_1 (13.1%) and S_4 (28.9%) metastates retained a similar α AF-2- $\alpha 3$ and helices $\alpha 10/\alpha 11$ conformations as the FXR-activator crystals. We suggest S_1/S_4 metastates represent the active conformations, while $S_2/S_3/S_5$ can be used to describe inactive/antagonist-bound systems. Interestingly, recent work from Heering *et al.* (2022)^[15] suggests that partial agonism can be better understood on the differential co-recruitment level, by favouring co-

repressor displacement, an aspect we have not investigated in this work.

However, due to the modest nature of these transitions in the MSM, we decided to closely investigate them along the trajectories. Interestingly, the committor probability graphic (Figure S4C) shows that there are multiple pathways allowing the transitions between these two main clusters, suggesting the description of active/inactive conformations is more nuanced than originally thought. The committor values assigns to each microstate a probability of a structure of being at this state, while the vector of committor probabilities measures the distances among these states.

Based on the PC extreme motions, their respective score distributions, the differences in the RMSF values and regions changing among the different metastable states, we selected relevant regions to further investigate: i) the LBP geometry ($\alpha 2$, $\alpha 5$ and $\alpha 6$ helices); ii) the FXR/RXR interface ($\alpha 9$ - $\alpha 11$ helices) and iii) α AF-2 ($\alpha 12$ and $\alpha 3$ helices), which will be discussed in the following sections.

FXR antagonists alter the ligand-binding pocket geometry

First, we focused on the $\alpha 6$ (residues Glu314-Asn337, in black, Figure 3A) and $\alpha 5$ (residues 346–354) regions and their closing upon ligand binding. Smaller distances are observed in the presence of agonist (medians 14.2–14.7 Å, for CDCA and GW4064), whereas both antagonist-bound systems display a significantly larger gap between these two helices (medians of 17.3 Å and 17.0 Å for **2h** and **2a** heterodimeric simulations, respectively) (Figure 3B). It is noteworthy that simulations with monomers led to bimodal distributions for the antagonistic simulations, with one of the conformation populations (**2a**, median 15.5 Å, however with the average of the lower population around ~ 12 Å) displaying a closer median to the heterodimers. Interestingly, in the CDCA original crystal structure loop region linking $\alpha 4/5$ and $\alpha 6$ (Lys339-Pro341) is not visible,^[13] which already suggests flexibility and dynamics of this loop induced by ligand binding.

This difference between agonists and antagonists can be explained by a few changes in the interaction profile within the LBP. **2a** and **2h**'s carboxylate groups display stable hydrogen bond interactions with Arg264 ($> 50\%$, Figure 3C) in addition to water-mediated interactions with His294 ($\sim 25\%$) and hydrogen bond interactions with Arg331 (common for all ligands). In addition, due to low distance (< 4.5 Å) salt-bridges are likely to be happening, but were not differentially quantified. Since the tested antagonists lack a polar group in the 7' position, hydrogen bonds with Ser332 ($\sim 40\%$) and Tyr361 ($\sim 40\%$) displayed by CDCA (Figure 3A and Figure S6), cannot lock the ligand in the commonly attributed binding mode within the LBP. The tested antagonists rather display relevant interactions with the $\alpha 2$ loop and are pushed far from the heterodimerization interface ($\alpha 11$) or the α AF-2.

Free energy perturbation calculations are proven to be a rigorous tool to validate hypothesized binding modes when in agreement to the experimental data. Nonetheless, the accuracy

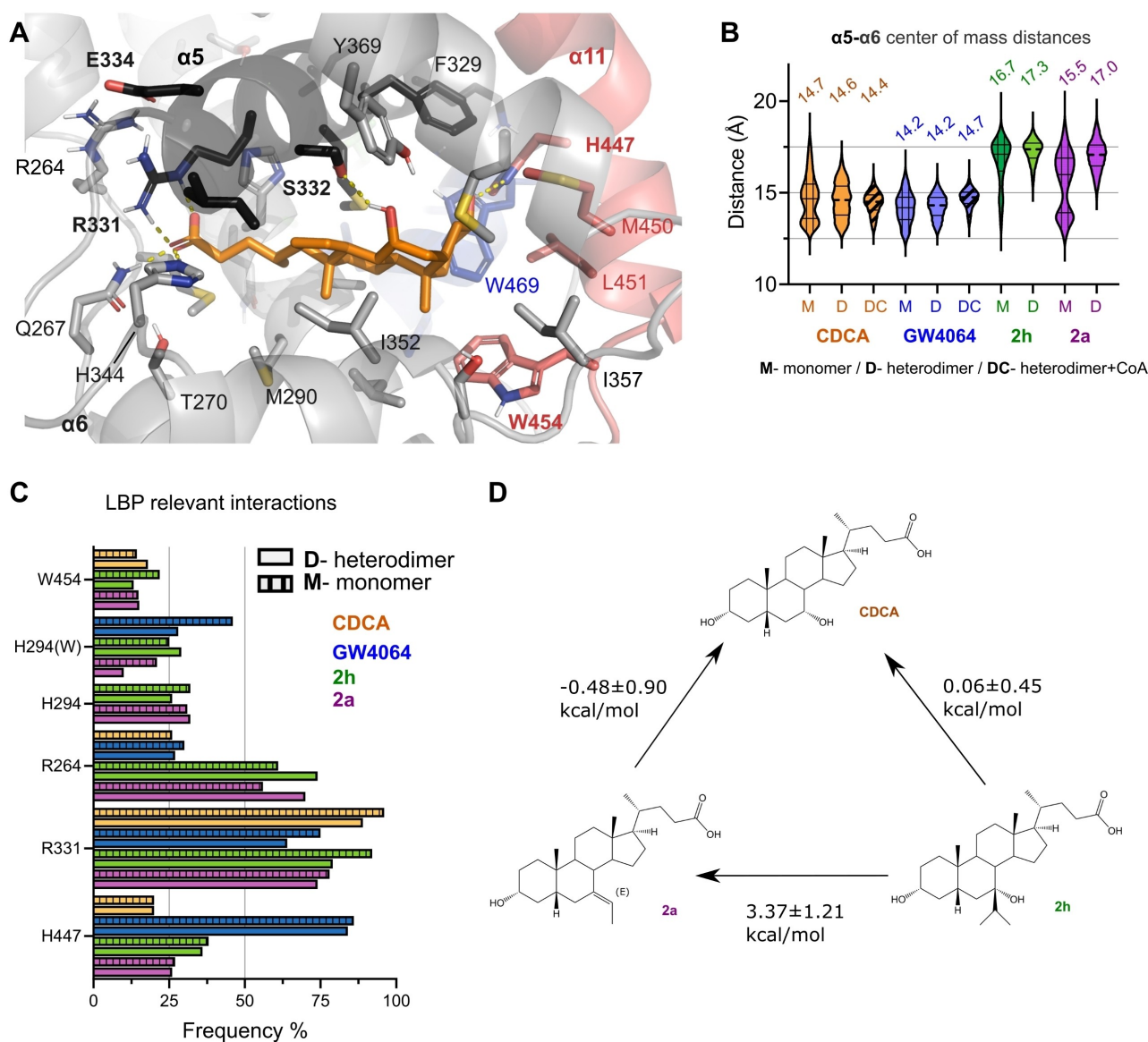


Figure 3. Compounds 2a and 2h induce conformational changes in the FXR's ligand-binding pocket. A) Representation of the ligand-binding pocket of FXR bound to CDCA (orange) and highlighting the amino-acids composing the ligand binding pocket as sticks. B) Distance between the centres of mass for $\alpha 6$ (residues 346–354, grey) and $\alpha 5$ (residues 325–337, in black). Distance plots are depicted as violin plots where the median of the largest continuous distribution is displayed as a number above it. C) Relevant FXR-ligand interaction frequencies (Supporting Information Figure S5 displays all interactions). D) Predicted relative binding free energies ($\Delta\Delta G_{\text{bind}}$, kcal/mol) after closing the thermodynamic cycle for the alchemical transformations involving ligands 2a, 2h and CDCA.

and convergence properties of these type of calculations turn out to be limited when in absence of a high-resolution structure, with short simulation times and when the number of atoms involved in the alchemical transformation is too big. The experimental binding free energies for the FXR ligands of interest (Table 1) were calculated from their reported EC_{50} or IC_{50} values^[23,24] using the approximation $\Delta G_{\text{exp}} \approx -RT \ln(EC_{50} \text{ or } IC_{50})$. The standard error of the mean after 10 replicates is reported. While the alchemical transformations $2a \rightarrow CDCA$ and $2h \rightarrow CDCA$ agree quantitatively (absolute error < 1 kcal/mol) with the experimental values, both antagonists with exception of our simulated nonsteroidal

Table 1. Experimental and calculated relative binding free energies between FXR ligands 2a, 2h and CDCA, as indicated in Figure 4.

Ligand pair	Relative binding free energies $\Delta\Delta G$ (kcal/mol)		
	Experimental	FEP	Error
2a \rightarrow CDCA	-0.45	-0.48 ± 0.90	0.03
2h \rightarrow CDCA	-0.54	0.06 ± 0.45	-0.6
2h \rightarrow 2a	0.09	3.37 ± 1.21	-3.28

agonist does not. Nonetheless, the latter FEP calculation is consistent with our previous results showing that 2h is a stronger antagonist than 2a, but still displays weaker binding when compared to 6 α -ethyl-chenodeoxycholic acid, a highly

potent steroidal FXR agonist.^[23] We believe that longer simulations and/or the usage of different starting points for the antagonists could improve the binding affinity estimation, but as now it stands a crystal structure with the steroidal antagonists would greatly improve this estimation.

Interestingly, with exception of our simulated nonsteroidal agonist (GW4064), which interacts with His447 (> 80%), no other ligand establishes stable polar interactions with residues in the $\alpha 11$ (Figure S6). This is consistent with the high flexibility observed for antagonists (Figure S2) and prompted us to further investigate helix $\alpha 11$ geometry changes along the simulations.

Antagonists alter the heterodimerization interface

The FXR/RXR heterodimerization interface is composed of the $\alpha 10$ and $\alpha 11$ (Figure 4A), with some extended contributions from $\alpha 7$, $\alpha 9$, L: $\alpha 8$ - $\alpha 9$ and L: $\alpha 9$ - $\alpha 10$. We initially investigated the influence of the ligands on the interface dynamics by measuring the distances between FXR's and RXR's $\alpha 10$ s, which are central for heterodimerization (Figure 4B, left). The $\alpha 10$ s distance showed no differences in the medians (Figure 4B, left), suggesting that the centre of this interface is not affected by the ligands. Interestingly, the extremities of the interface display larger movement amplitude with the L: $\alpha 9$ - $\alpha 10$ closing the gap by moving towards the RXR's $\alpha 10$ (Figure 4B, right), where the

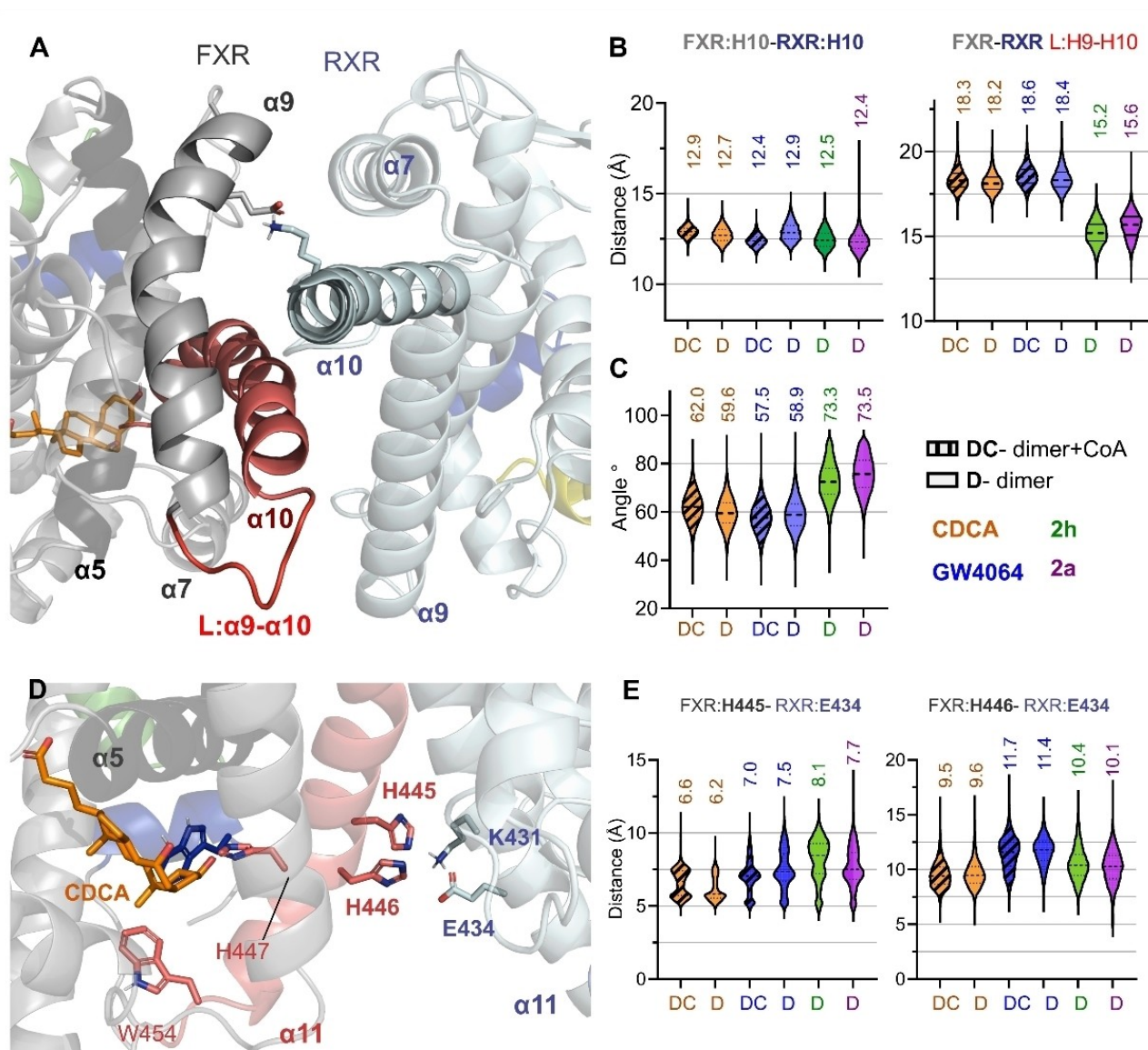


Figure 4. Antagonist binding changes the heterodimerization interface by reorganizing the C-terminus of $\alpha 11$. A) Top view of the FXR/RXR interface highlighting the $\alpha 9$, $\alpha 10$ and the loop connecting both for FXR (grey) and the $\alpha 10$ from RXR (cyan). Distances between RXR's $\alpha 10$ centre of mass (residues 384–407) and the (B, left) FXR's $\alpha 10$ (residues 429–452) and the (B, right) L: $\alpha 9$ - $\alpha 10$ (residues 423–429, in red). C) Dihedral angle variation between the planes composed by $\alpha 7$ vs. $\alpha 9$. D) Overview of the FXR/RXR interface near $\alpha 11$ highlighting the relevant amino acids to sustain the interface as sticks. E) Distances between the centre of mass from RXR's Glu434 and FXR's His445 (left) and His446 (right). All distance plots are depicted as violin plots where the median of the largest continuous distribution is displayed as a number above it.

antagonist simulations display lower median values (median ~ 15 Å) in comparison to the activators (~ 18 Å).

Monitoring the fluctuations in the dihedral angle between the planes composed of $\alpha 7$, $\alpha 9$, $\alpha 10$ and $\alpha 11$ (Figure S7 and Tables S1,2) showed little changes, with exception of the dihedral angle between the planes formed by FXR $\alpha 7$ and RXR $\alpha 9$ (Figure 4A, C). This dihedral variation along the MD simulations enabled us to identify changes in the interface in the presence of agonist, partial agonist and antagonist, suggesting that FXR-bound to activators displays a tighter interaction with RXR.

Complementary to the changes in the upper interface, the distance gaps between the $\alpha 11$ helices of both the receptors are shorter for agonists than for the antagonists, as represented by the distances between the relevant residues such as His445 and His446 in FXR and Glu434 and Lys431 in RXR (Figures 4D, E). In addition, we also observed increased mean distances in residues from the $\alpha 7/\alpha 9$ and $\alpha 10$ (composing the FXR-RXR interface) for antagonistic simulations (Figure S8A, D). These changes are not supported by the loss of helical folding in the $\alpha 11$, as originally proposed by FXR-ivermectin crystal structures, since we have observed only minor folding fluctuations and only in monomeric simulations (Figure S8B, C). Also, no bending between $\alpha 10$ – $\alpha 11$ at the level of Thr442 was observed (*data not shown*). Both behaviours would, possibly, require longer simulations to be fully evaluated.

Interestingly, the role of this histidine-mediated interaction is further supported by the work of Wang *et al.* (2018), who demonstrated that mutations in the $\alpha 10/11$ conferred poorer ligand activation and synergy,^[25] due to the loss of interaction between the protomers (especially where FXR:His445 and RXR:Glu434 mutants were concerned). The presence of a basic amino acid position in His445 of FXR, seems to be exclusive for FXR when compared to other RXR partners, which could explain why **2a/2h** does not affect other members of this family.^[23]

Simulations can recapitulate α AF-2 antagonist-induced conformational change

First, we monitored the distance between α AF-2 (residues 463–472) and $\alpha 3$ -helix (Figure 5A,B). As $\alpha 3$ -helix (residues 281–304) is stable in all simulations (RMSF < 1 Å), this distance enables the assessment of the relative position of α AF-2 to the LBD. Antagonist-bound simulations display moderate changes in distances between $\alpha 3$ – α AF-2 (median ~ 12 Å – antagonist – in comparison to ~ 10 Å – agonists, Figure 5B). These α AF-2 changes are not accentuated in monomer simulations. This suggests that antagonists and agonists would bind similarly to the LBD, without largely affecting heterodimerization, but only the α AF-2 region.

DM175 is a synthetic FXR partial agonist, which similarly to ivermectin, can reduce FXR activation and stimulate its basal levels.^[13] Similar to our compounds, FXR-DM175 promotes the $\alpha 11$ formation and stabilizes the L: $\alpha 3$ – $\alpha 4$ loop.^[6,13] Further, the reported FXR-DM175 structure (PDB ID: 4QE8) displays a Trp454 side chain ~ 12 Å outward of the binding pocket, pushed away

due to the extra tert-butyl moiety. This moiety then occupies a transient pocket composed of the L: $\alpha 11$ – $\alpha 12$. However, the electron density of L: $\alpha 11$ – $\alpha 12$ (Val456–His459, AF-2 loop) is ambiguous, supposedly due to the high ligand-induced flexibility. Our simulations started with Trp454 inwards, and none displayed a flipping outward, contrarily, its sidechain establishes hydrophobic contacts with our antagonists (Figure 5C, Figure S6). The role of Trp454 hydrophobic contacts with the ligand in the L: $\alpha 11$ – $\alpha 12$ stabilization is well-established for steroidal ligands,^[13] as observed by Trp454Ala/Trp454Tyr mutations that prevent CDCA activation (up to 50 μ M), while not affecting GW4064. Our results, however, cannot show Trp454 as a structural determinant for antagonism.

On the other hand, these results suggest that FXR antagonism by **2a/2h** can be explained by changes in the α AF-2 helix conformation, which are mediated by the flexibility of $\alpha 11$ and associated loops, distinctly from the DM175 and other known antagonists/partial agonists.

We also observed that the unfolded loop-like conformation of α AF-2 seems to be preferred for antagonists (Figure 5E, F) in comparison to the agonist counterparts, independently from the dimerization state. The main residue that contributes to this change seems to be Trp469, interacting both with the ligand (Figure 5D) and with His447 (Figure S9). The Trp469 side chain remains in proximity (median distance ~ 2.8 Å) to the ligand for both the agonists and, to a lesser extent ($\sim 40\%$ of the analysed simulation time), the antagonists in the heterodimer simulations. This agrees with the role of the Trp469–His447 interaction in stabilizing the α AF-2 helix conformation, which is less frequent when our antagonists are bound ($\sim 40\%$ of the analysed simulation time, Figure S9).

α AF-2 antagonist-induced conformational changes that impair co-peptide recruitment

We further employed the FXR ligand co-recruitment profiles assessed by the Microarray Assay for Real-Time Coregulator-Nuclear Receptor Interaction (MARCoNI) assay to experimentally verify the recruitment of co-regulatory proteins. The heat-map of FXR ligand interactions generated with the MARCoNI assay represents the Modulation Index (MI, depicted in Log values), where the red colour depicts positive interactions, pointing out the ligand-induced increase in peptide binding, and the blue colour depicts negative interactions (Figure 5G). This suggests that 2 h binding to FXR compromises the integrity of the entire α AF-2 region (Figure 5G, Table S3), given that neither major co-activators nor co-repressors were recruited, when compared to the tested agonists.

Interestingly, the recent work from Rashidian *et al.* (2022) employed long molecular dynamics simulations of PXR-LBD + CoA systems to study antagonist-induced conformational changes.^[22] Given that the antagonist (C100) in question did not experimentally recruit coactivators,^[26] they decided to use an arbitrarily generated PXR–C100 + CoA as a control system to investigate whether their models could recapitulate the experimental results. Indeed, the configuration of α AF-2 is destabi-

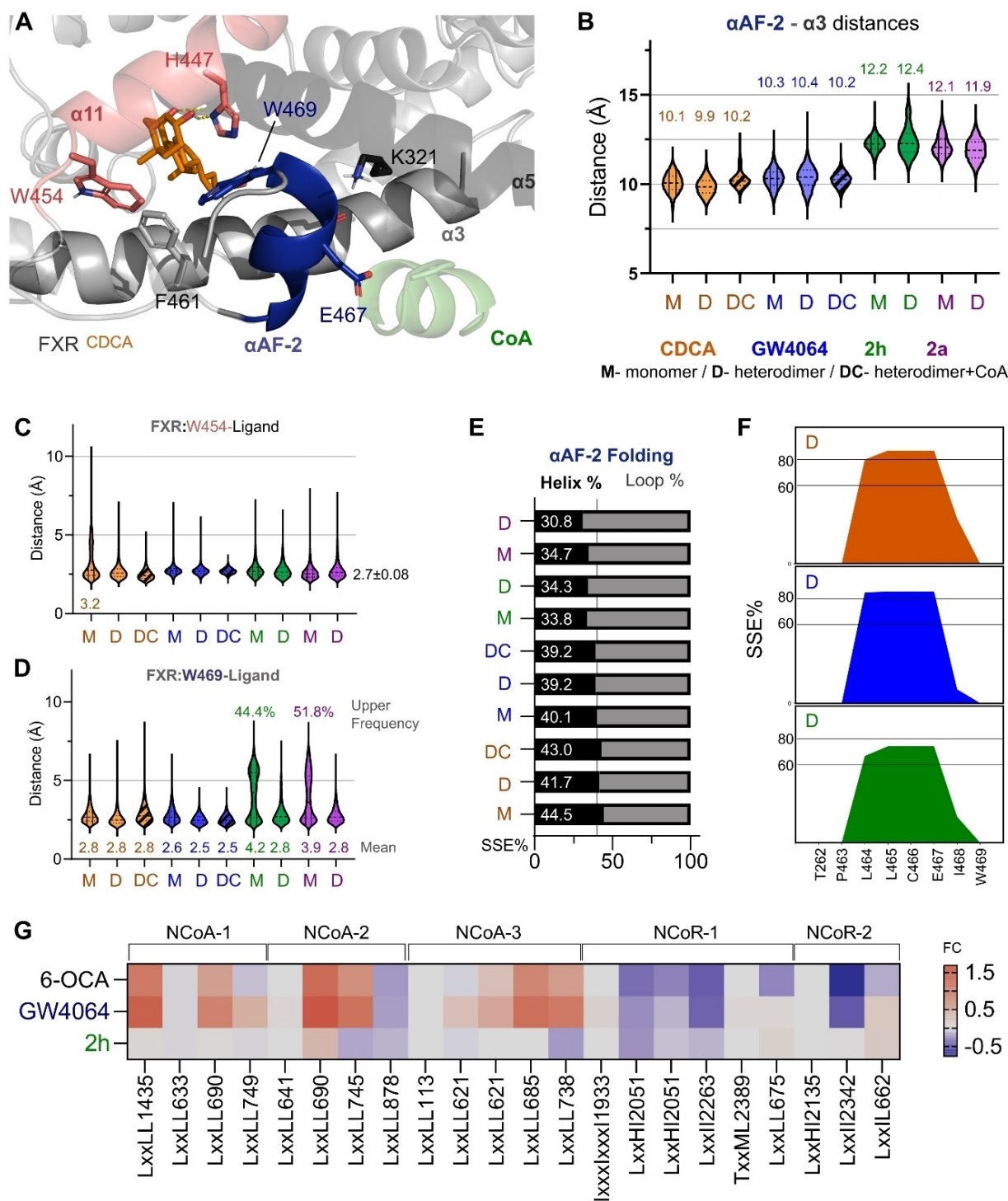


Figure 5. α AF-2 helix displays moderate conformational changes relevant for co-recruitment (A) Location of the α AF-2 helix and surrounding region in the FXR-LBD. (B) Distance between the centres of mass of α AF-2 helix (residues 463–472, α 12) and α 3 (residues 281–304) illustrates that the antagonists promote an open configuration of the α AF-2 helix in both monomeric and dimeric simulations. Minimum distances between one of the ligand's atoms and the sidechain from Trp454 (C) and Trp469 (D). All distance plots are depicted as violin plots, where the median of the largest continuous distribution is displayed as a number above it. For Trp469 antagonistic distances the frequency of the upper distribution is also depicted. (E) Secondary structure of the α 12 helix appear more stable with agonistic compound CDCA and GW4064 than with antagonists 2a and 2h. (F) Area plots represent the observed secondary structure element (SSE%) of the α 12 helix in percentage throughout the simulation. (G) MARCoNI assay heatmap, the red colour shows positive interactions/increased FXR binding, and the blue colour shows negative interactions/decreased FXR bindings of the ligands (6-OCA, stands for obeticholic acid). Modulation Index (MI). FXR-peptide binding was quantified the using Bionavigator software (see methods). MI is the \log_{10} -transformed relative binding value (namely Fold Change, FC), which is calculated as the compound's binding value relative to the vehicle control (DMSO) binding value.

lized by C100 binding. Similarly, inspired by their approach, we generated simulations for FXR-2h+CoA (similarly to our other systems, $5 \times 3 \mu\text{s}$). The cofactor of choice was the same as the CDCA simulations, containing the motif NID-3 (KENALLRYLLDKD, NCOA2), therefore enabling direct comparison.

Our FXR+2 h heterodimer+CoA (Figure 6A–C, labelled as DC) simulations recapitulated most of the key structural features observed with their heterodimer without FXR-CoA (Figure 6A–C, labelled as D), with exception of the $\alpha\text{AF-2}-\alpha\text{3}$ distance whose median values are smaller in comparison to other 2h simulations (11.5 Å, Figure 6C). This is consistent with

the smaller flexibility of the $\alpha\text{AF-2}$ and L: $\alpha\text{11}-\alpha\text{12}$ for 2 h + CoA simulations in comparison to free heterodimers (Figure 6D), suggesting that arbitrarily adding a coactivator peptide stabilizes this motif in the simulations. Interestingly, the NCoA-2 peptide loses folding in its C-terminal portion (Figure 6E–G), 50.3% for CDCA in comparison to 31.6% observed for 2h. This is consistent with the infrequent interactions between this peptide and the α3 and, to a lesser extent, $\alpha\text{AF-2}$ (Table S4 displays the interaction frequency between the coactivator side chains and the closest amino acids from FXR).

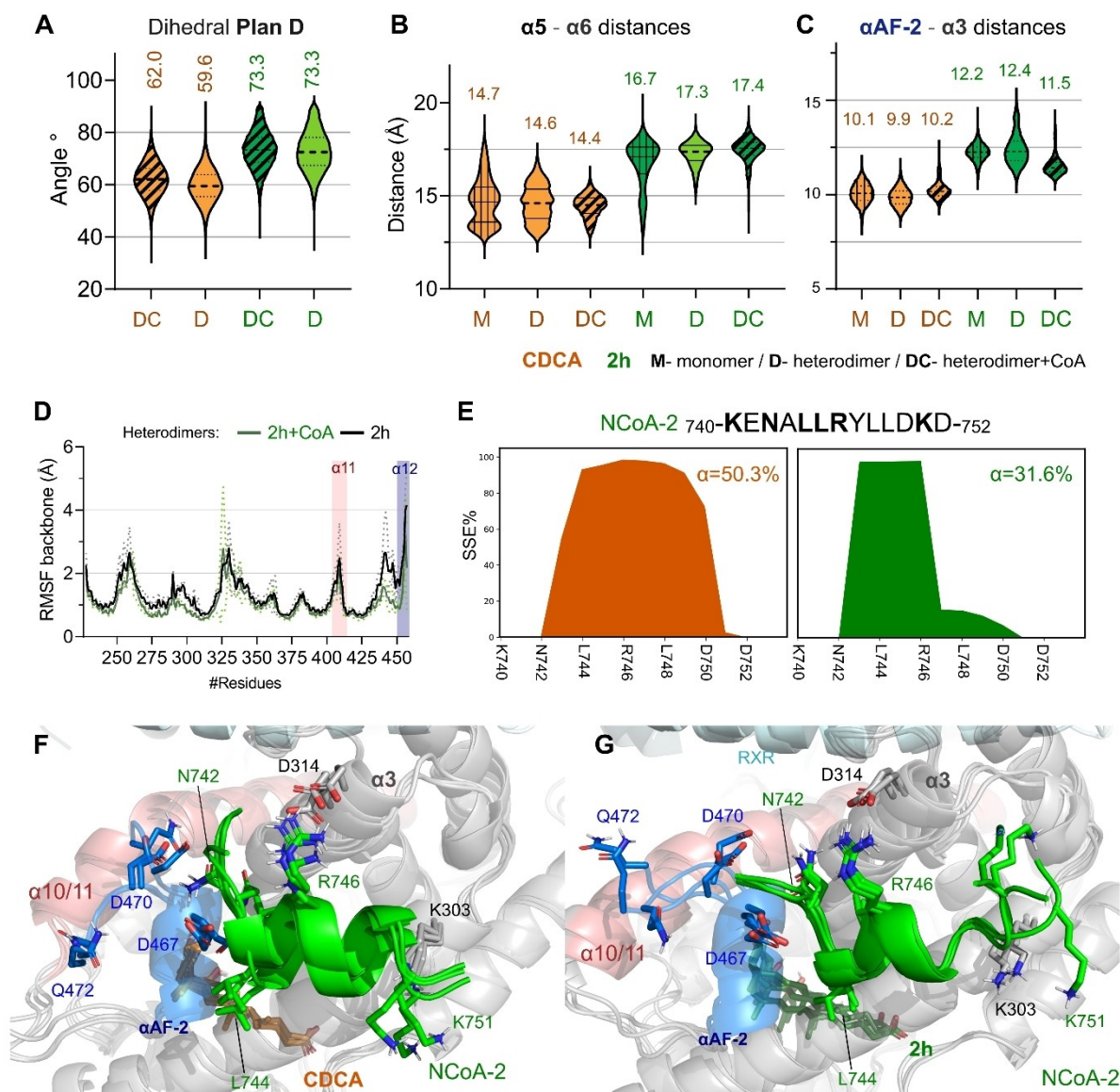


Figure 6. Antagonist binding to FXR-CoA/RXR-CoA complex cannot retain the coactivator's fold. (A–D) quantification of structural changes hallmark for antagonist: A) dihedral angle between the planes composed by α7 vs. α9 . B) Distance between the centres of mass for α6 (residues 346–354, grey) and α5 (residues 325–337, in black). C) Distance between the centres of mass for $\alpha\text{AF-2}$ helix (residues 463–472) and α3 (residues 281–304). All distance plots are depicted as violin plots where the median of the largest continuous distribution is displayed as a number above them. D) Root-mean-square fluctuation (RMSF) of the protein backbone. Comparing heterodimers with co-activators (green) and without (black). For each graphic α11 (salmon) and α12 (blue) helices are highlighted. Bold lines represent the means of five replicas and dashed lines are their respective standard deviations. E) Area plots represent the observed secondary structure element (SSE%) of the coactivator peptide (NCoA-2) in percentage throughout the simulation. Relevant simulation frames (clustered by the coactivator's RMSD values) for CDCA (F) and 2h (G) simulations.

Discussion

FXR ligand-mediated activation by agonists is well understood as a classical α AF-2-trapping mechanism. However, antagonism mechanisms are suggested to be more diverse. Our previous, much shorter, simulations with FXR-antagonists have hinted at the role of L: α 11- α 12 destabilization on monomers.^[23] In this manuscript, we expand on the antagonist binding mode using microsecond timescale all-atom molecular dynamics (MD) simulations. We compared simulations with steroidal antagonists with the synthetic (GW4064) and steroidal (CDCA) FXR agonists. We observed distinct conformational changes in relevant regions, such as the FXR/RXR interface (α 9, α 10 and α 11) and in the α AF-2 (α 12- α 3 and L: α 11- α 12), as well as the LBP geometry/interaction pattern. This is supported by available FXR crystal structures showing that loops L: α 1- α 2, L: α 5- α 6 and L: α 11- α 12 are differently oriented/resolved depending on the ligand bound. CDCA crystal structure, for instance, displays an unfolded L: α 5- α 6 and binding of the partial agonist DM175 destabilises both L: α 5- α 6 and L: α 11- α 12. Our simulations show extensive LBP re-orientation driven by L: α 1- α 2/L: α 5- α 6 as a structural feature distinguishing agonists/antagonists. CDCA and GW4064 display smaller average α 5- α 6 distances, promoted by interactions with α 5's residues, while antagonist compounds **2a/2h** rely only on α 1/ α 2 interactions. This is reflected in the FEP calculations and in agreement with previous experimental binding results showing that **2h** is a stronger antagonist than **2a**, but still displays weaker binding when compared to CDCA.

We also observed changes in the FXR/RXR geometrical organization when comparing agonists/antagonists, by the means of changes in the dihedral planes and, specifically, in the interaction pattern between residues at the heterodimerization interface. Extensive work on nuclear receptors using RXR as a heterodimer partner from the structural perspective has classified them into permissive and nonpermissive heterodimers.^[11,25,27] Structural alignment, using the RXR's α 11 as the reference, shows different α 11 orientations according to the RXR-dimerization partner's permissiveness. Permissive partners (such as PPARs and FXR itself) display a larger bending than nonpermissive partners. This is suggested to allow permissive heterodimers to sense both receptor ligands, whereas nonpermissive heterodimers can respond only to partner ligands, without the influence of RXR.

Changes in the α 11, such as the dramatic unfolding observed in the FXR-ivermectin structure, can lead to the lack of observable helix α 12 and L: α 11- α 12. Similarly, in our antagonist simulations, we observe an α 11 destabilization, accentuated in the monomeric simulations, which leads to L: α 11- α 12 / α AF-2 displacement. Further, our antagonist seems to destabilize the L: α 11- α 12 and consequently the α AF-2 helix active conformation (competent in terms of co-activator recruitment), therefore, preventing both the co-activator and co-repressor recruitment. We propose a dynamic mechanistic interpretation linking the heterodimerization with the co-regulatory protein recruitment, which complements previous crystallographic data.^[25]

Further, the work from Merk *et al.* (2019) establishes that FXR activation is a result of an equilibrium between conformational populations.^[13] Specifically, FXR agonists would stabilize the folded and extended helix α 11 (composing the heterodimerization interface) and the α 11- α 12 loop (composing the α AF-2 region) upon binding. The stabilized α AF-2 region enables coactivator recruitment and therefore FXR activation. In contrast, partial agonist ligands can lead to changes in the helix α 11 conformation, which, consequently, destabilizes the α 11- α 12 loop and the helix α 12 orientation.

In agreement, the CDCA agonist-bound crystal structure has sufficient electron density to accurately represent both the α 12 helix and the entire loop connecting α 11 and α 12, whereas in the DM175 (partial agonist) co-crystal structure, L: α 11- α 12 is invisible, due to destabilisation, and α 12 is shifted to a new position. Our simulations show that the α 12 position is dynamic and fluctuates around the active position site (as shown by the α 12- α 3 distances fluctuation) for agonists. Meanwhile, our simulated antagonists, clearly displace the α AF-2 from the active conformation to a point where no co-regulatory protein was able to stably bind to it. Even modelled Antagonist + CoA showed reduced, but the steady, displacement of the co-activator regulatory motif and further unfolding of the peptide and the α AF-2.

Conclusion

In this work, we started from an agonist/apo structure with a fully folded FXR-LBD in order to simulate the initial conformational changes that would happen upon antagonist binding.. It is important to highlight that longer timescales or alternative sampling approaches could be used to generate a larger picture of the FXR-LBD conformational landscape, especially when considering its high flexibility. Our proposed ligand binding mode would be much closer to that of the partial agonist DM175 than larger ligands, such as ivermectin. However, both ivermectin and DM175 are known to recruit co-repressor proteins upon binding, which was not recapitulated by our MARCoNI assay, suggesting that our antagonists could, indeed, display an alternative binding mechanism. It is difficult to conclude if the antagonists bind to FXR any differently from the agonists (CDCA/GW4064). However, the antagonists fail to stabilize the structural motifs responsible for co-activator recruitment. Therefore, we propose a passive binding as an antagonism mechanism.. Other FXR-antagonist models, such as FXR-F6,^[28] which is based on the oleanane-type triterpenoid agonist-bound structure (PDB 5WZX), proposes a similar binding mode relying on the stable interaction of His and Trp trap of α 12.

Moreover, we understand that *in vivo* modulation of FXR is subject to other layers of complexity, such as distinct FXR isoform expression^[1] or the tissue availability of the different co-regulatory proteins. Noteworthy, FXR is expressed as four different isoforms α 1-4 with modified DBD and hinge regions, but identical LBD domains. These differences were not considered in this current work and the relationship between the DBD

and LBD dynamics would be the scope of future investigation. Finally, further analysis of these allosteric effects on gene transcription profiles, linked to the different ligand-coactivator effects, would be highly relevant from the agonist perspective and could generate new insights for rational drug design.

Experimental Section

Molecular modelling and structure preparation. The Small-Molecule Drug Discovery Suite (v2019.4, Schrödinger, LLC, New York, NY, 2019) was used for all calculations. Protein structures were prepared by adding hydrogen atoms and fixing missing sidechains using the Protein Preparation Wizard (PrepWiz),^[29] missing loops were generated using Prime. Sulphate/crystallization buffer molecules, such as glycerol (GOL) were removed.

The PDB structure of the human monomer ligand-binding domain of FXR co-crystallized with CDCA (PDB ID: 6HL1,^[13] resolution: 1.60 Å) was selected for docking and simulation based on the quality and similarity between the co-crystallized ligand and our series. The structures co-crystallized with the nonsteroidal agonist GW4064 in monomeric (PDB ID: 3DCT,^[30] data not shown, resolution: 2.5 Å) and dimeric states (PDB ID: 6A60, resolution 3.05 Å) were also selected for comparison. FXR/RXR heterodimer structures for CDCA and the studied antagonists were generated by superimposing their FXR monomers against the 6A60 structure.

Comparatively, heterodimers with FXR+coactivators (or called dimer with coactivators, DC) were generated by keeping the original NCoA-1 (for GW4064) or NCoA-2 (for CDCA structure) peptides. N- and C-terminus of these peptides were capped. To generate the heterodimeric models (herein referred to as D), without a co-activator, the peptides were removed from the AF-2 region of FXR but maintained in the RXR structure. Similarly, monomeric systems were generated by removing both RXR and CoA (herein referred to as M). For all structures protonation states of amino acids were optimized with PROPKA (Schrödinger, LLC, New York, NY, 2019), where we selected the most likely ionization state as proposed by the software, and the structures were minimized using steep descent to the cut-off (1 Å of allowed movement on heavy atoms).

Ligand preparation and molecular docking. Potential binding modes for the antagonists **2a** and **2h** were based on our earlier work.^[23] Briefly, all ligands were prepared using LigPrep (v2019.4) with standard options. Molecular docking of the antagonists was performed in a grid encompassing residues around 10 Å from the centroid of the co-crystallized ligand, using the default settings of the Glide program (Glide v7.7, Maestro v2019.4) in precision mode.^[31] Amino acid residues were considered rigid. For each ligand 10 docking poses were generated and visually inspected taking into account key interactions with the Arg331 and the positioning of the steroid ring.

Molecular dynamics simulation. For each ligand (CDCA, GW4064, **2a** and **2h**) monomeric (M) and heterodimeric (D) simulations were generated, for CDCA, GW4064 and **2h** simulations with FXR+CoA were also run (DC). MD simulations were carried out using Desmond,^[32] with the OPLS3e force-field.^[33] The simulated system encompassed the protein-ligand complexes, a predefined water model (TIP3P^[34]) as a solvent and counterions (Na⁺ or Cl⁻ adjusted to neutralize the overall system charge). The system was treated in a cubic box with periodic boundary conditions specifying the shape and the size of the box as 13 Å distance from the box edges to any atom of the protein. In all simulations, we used a time step of 1 fs, the short-range coulombic interactions were treated using a cut-off

value of 9.0 Å using the short-range method, while the Smooth Particle Mesh Ewald method (PME) handled long-range coulombic interactions.^[35] Initially, the relaxation of the system was performed using Steepest Descent and the limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithms in a hybrid manner, according to the established protocol available in the Desmond standard settings. During the equilibration step, the simulation was performed under the NPT ensemble for 5 ns implementing the Berendsen thermostat and barostat methods.^[36] A constant temperature of 310 K was kept throughout the simulation using the Nose-Hoover thermostat algorithm^[37] and Martyna-Tobias-Klein Barostat^[38] algorithm to maintain 1 atm of pressure, respectively. After minimization and relaxation of the system, we continued with a single production step of at least 3 μs, with frames being recorded/saved every 1,000 ps. For each ligand five independent replicas were run totalling around 15 μs per system. Trajectories and interaction data are available on Zenodo repository (under the codes: 10.5281/zenodo.6637313 – dimeric simulations without CoA, 10.5281/zenodo.6637925 – monomeric simulations, and 10.5281/zenodo.6638475 – dimeric simulations with CoA).

The representative structures were selected by inspecting changes in the Root-mean-square deviation (RMSD), meaning for figures a representative frame was selected at random at points of the trajectory where the RMSD for were not fluctuating, after equilibration. Figure S1 represents the variation of the RMSD values along with the simulation, for both template crystal structures and simulations with docking pose. Additionally, the changes in the Root-mean-square fluctuation (RMSF), normalized by residue for the protein backbone are displayed in Figure S2.

Simulation trajectory analyses

Protein-ligand interactions were determined using the Simulation Event Analysis pipeline implemented in Maestro (Maestro v2019.4). The geometric criteria for protein-ligand hydrogen bond is a distance of 2.5 Å between the donor and acceptor atoms (D–H...A); a donor angle of $\geq 120^\circ$ between the donor-hydrogen-acceptor atoms (D–H...A); and an acceptor angle of $\geq 90^\circ$ between the hydrogen-acceptor-bonded atom atoms (H...A–X). Similarly, protein-water or water-ligand hydrogen bond had a distance of 2.8 Å between the donor and acceptor atoms (D–H...A); a donor angle of $\geq 110^\circ$ between the donor-hydrogen-acceptor atoms (D–H...A); and an acceptor angle of $\geq 90^\circ$ between the hydrogen-acceptor-bonded atom atoms (H...A–X). Non-specific hydrophobic interactions are defined for hydrophobic sidechain within 3.6 Å of a ligand's aromatic or aliphatic carbons and π - π interactions required two aromatic groups stacked face-to-face or face-to-edge, within 4.5 Å of distance.

Protein secondary structure elements (SSE) were monitored over the course of the simulation using the Maestro event analysis tool. Angle and distance calculations were performed employing the Maestro event analysis tool (Schrödinger, LLC, New York, NY). Distances between specific secondary structure elements were calculated using their centres of mass, using the script `trj_asl_distances.py` (provided by Schrödinger, LLC, New York, NY).

Distance measurements: The Simulation Interaction Diagram (SID) analyses tool was used for each system and replicate for the binding site and the FXR-RXR interface to calculate distances and interaction frequencies between amino acids. While for the binding site, distance measurements were performed for relevant protein-ligand interactions identified from the SID report, for the heterodimer interface we measured the distances for the interactions reported in the supplementary information by Wang *et al.* (2018)^[25] and additional protein-protein interactions obtained from the SID.

All the measurements were performed using the script `trj_asl_distance.py` having the atom numbers of the atoms involved in the interaction as an argument. Distances between the helices were calculated using `trj_asl_distance.py` script from Schrödinger.

Dihedral angle measurements: The FXR-RXR interface involves helices $\alpha 11$, $\alpha 7$ and $\alpha 9$ from both monomers. Subsequently, four planes for each monomer were defined by selecting the first, third and last $C\alpha$ of each α -helix in the N-terminal to C-terminal direction. Based on these secondary structure-derived planes we measured four dihedral angles (A, B, C and D) along the MD simulation trajectories to study the fluctuations in the interface (Tables S1 and S2). To perform the dihedral angle measurements, the script `analyze_simulation.py` was used having as an argument the six atom numbers defining the two planes (Table S2).

Principal component analysis (PCA): Extreme motions of the protein complexes during the molecular dynamics simulations were analysed using principal component analysis. All the python scripts used in this study were provided by Schrödinger. Analysis was run considering only the variation of the backbone atoms, which were retained in a modified trajectory using `trj_keep_selection_dl.py` script. The entire trajectory was then aligned to frame 0 (initial frame) using `trj_align.py` script and trajectories from all the simulations were merged using the python script `trj_merge.py`. The combined trajectory was used to generate `.xtc` and `.pdb` files (required for the Gromacs software) using `trj_no_virt.py` script, followed by our in-house developed script (`fix_pdb.py`, available upon request) used to fix the `.pdb` file generated in the previous step.

The generated files were used for the PCA analysis using Gromacs. To exclude the extreme motions incurred by loops (of intracellular and extracellular regions), only transmembrane helices were considered for the calculations using the `make_ndx` script from Gromacs. Further, principal component analysis was carried out using `gmx ana eig` following the covariance matrix generation using the `gmx covar` command line script, with standard options. Extreme motion figures were generated and visualized using mode vectors script from PyMOL 2.5.2.

Free energy perturbation (FEP) calculations: Protein-ligand complexes for ligands **CDCA**, **2h** and **2a** were generated using molecular docking with Glide SP to the homology model derived from PDB ID 6HL1, based on our earlier work.^[23] These complexes were used as a starting point for performing FEP/MD simulations with the software Q.^[39] Moreover, we used the QligFEP^[40] pipeline for the generation of all the required input files. The FEP/MD simulations were carried out on spherical boundary conditions (SBC) on a sphere (25 Å radius) where its centre is placed on the centre of geometry of the CDCA agonist ligand. Protein atoms in the outer shell of the sphere (22–25 Å) had a positional restraint of 10 kcal/mol/Å², while solvent atoms were subject to polarization and radial restrains according to the surface-constrained all-atom solvent (SCAAS) model^[41,42] to mimic the properties of bulk water at the sphere surface. Solvent bonds and angles were constrained using the SHAKE algorithm^[43] and all the titratable residues lying outside the sphere were neutralized. Atoms outside the sphere were constrained (200 kcal/mol/Å²) and excluded from the calculation of the nonbonded interactions. Long-range electrostatic interactions beyond a 10 Å cut off were treated with the local reaction field method excluding the atoms undergoing the FEP transformation. While OP2L2005 ligand parameters were obtained from the `ffd_server`,^[44] the OPLS-AA/M force field was used to parametrize protein and solvent (TIP3P). After the Maxwell-Boltzmann velocities were assigned for each of the 10 replicates, the system was heated from 0.1 to 298 K during an initial equilibration stage of 40 ps, where a positional restraint of 10 kcal/

mol/Å² imposed on all the dual-topology heavy atoms was progressively released. Subsequently, a 100 ps unrestrained and unbiased equilibration was performed. Thereafter, the alchemical transformation was divided into 267 intermediate λ states distributed in 5 steps: (A) Introduction of a softcore potential for ligand B atoms, (B) Replacement of the softcore potential of ligand B with a Lennard-Jones potential, (C) Turning off the partial charges of ligand A and turning on the partial charges of ligand B, (D) Introduction of a softcore potential for ligand A atoms, (E) Annihilation of ligand A atoms. At each λ window the MD sampling consists of 20 ps using a 1 fs time step. To close the thermodynamic cycle a FEP/MD simulation was run in a water sphere using same MD parameters (*e.g.*, simulation time, sphere size, etc.) as in the bound system and the Relative Binding Free Energy (RBFE) between each ligand pair was obtained by solving the thermodynamic cycle using the Bennet acceptance ratio (BAR).

Markov state model (MSM) analysis: MSM generation was conducted with PyEMMA 2.^[45] Bayesian MSM was shown following the general recommendations.^[46] The Bayesian MSM are usually used to compute confidence intervals. The individual trajectories of complete FXR-LBD systems (apo and ligands) were used as an input for MSM generation. For featurization, we used the backbone torsions of the FXR-LBD monomer protein domain in the presence of the coactivator, agonist and antagonist. We validated the final model during the decision process by the VAMP-2 backbone torsions score,^[47] which was 3.03 for the final model. VAMP-2 score correspond to more kinetically accurate models and the cross-validated test score is bounded from above by the true kinetic mode, where values lower than 1.5 represented poor/bad features choices. The dimensional reduction was conducted with time-lagged independent component analysis (TICA).^[48] The length of lag time $\tau=800$ ns and 101 dimensions were selected, where the implied timescales were converged. Discretization of the data to microstates was done by k-means clustering (\sqrt{N} used for the number of clusters). Finally, a spectral clustering using the Perron-cluster cluster analysis (PCCA⁺⁺)^[49] assigned the microstates to macrostates. Transition-path theory (TPT) was applied to investigate state transitions and the flux of pathways between metastable states.^[50,51] MSM graphics were generated using the python script, available in the GitHub repository (code: https://github.com/gmf12/FXR-RXR_analysis). The validation of MSM models is shown in Supporting Information Figure S4.

To sum up, by the discrete state decomposition of the trajectories, have been identified, we observed the transition counts matrix (Lag time, free energy and dimension; in this case, the free energy surface is defined by the negative logarithm of the probability computed from the histogram counts; line 147 on GitHub repository: https://github.com/gmf12/NR_MSM/blob/main/MSM_NR-gmf.ipynb), which records how many times the system transitions to the column-indexed state given that it is in the row-indexed state.^[52,53] Finally, the criteria to show the distributions of all loaded features concerning the Lag time, free energy and dimension, however, to construct this depends on your volume of data to load and run the MSM.

Co-peptide recruitment: The ligand-modulated coregulator interactions with the GST-tagged FXR-LBD (Thermo Fisher) and anti-GST antibody labelled with Alexa488 (Invitrogen) were assessed using a PamChip® microarray that contains 154 coregulator-derived binding peptides, including the LXXLL coactivator motif or the LXXXIXXXL corepressor motif from 66 different coregulators (PamGene International B.V.).^[54,55] The reaction was performed in a fully automated microarray processing and fluorescent imaging platform (PamStation12) at 20 °C for 80 cycles (two cycles per minute). After removal of the unbound FXR-LBD receptor by washing each array with 25 μ l Tris-buffered Saline (TBS) buffer, fluorescent images of the PamChip

microarrays were obtained using the PamStation®12, and then were analyzed for quantification of FXR binding using BioNavigator software (PamGene International B.V.). For data and statistical analysis, BioNavigator was used. Modulation index (MI) is the log₁₀-transformed relative binding value, which is calculated by the compound's binding value, relative to the vehicle control (DMSO) binding value. Positive interaction (MI > 0) indicates that ligands increase FXR-coregulator motif binding, while MI < 0 indicates that ligands decrease FXR-coregulator motifs interactions. Overall raw binding values the $v > 50$; the relative binding value < 1; p-value < 0.05 were considered for the analyses.

Acknowledgements

TüCAD2 is funded by the Federal Ministry of Education and Research (BMBF) and the Baden-Württemberg Ministry of Science as part of the Excellence Strategy of the German Federal and State Governments EXC 2180–390900677. A.R. acknowledges the European Union's Horizon 2020 research and innovation program under grant agreement no 825762, EDCMET project. The project EDCMET has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825762. T.K. is supported by the Fortüne initiative (Nr. 2613-0-0) and by the iFIT, which are both initiatives from the Excellence Strategy of the German Federal and State Governments G.M.F. is a recipient of a fellowship from FAPESP (Grant #2021/11205-9, Brazil). Academy of Sciences of the Czech Republic (AS CR) – grant RVO 61388963. The authors would like to thank Mr. Joonas Tervo for providing the fix_pdb.py script and the CSC – Finland for the generous computational resources provided and Prof. Dr. Vinícius Gonçalves Maltarollo for the critical reading. Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are openly available in Zenodo at <https://doi.org/10.5281/zenodo.6637313>, reference number 6637313.

Keywords: bile acids · molecular dynamics simulations · nuclear receptors · farnesoid X receptor

- [1] S. Fiorucci, E. Distrutti, A. Carino, A. Zampella, M. Biagioli, *Prog. Lipid Res.* **2021**, *82*, 101094.
- [2] V. Massafra, R. Pellicciari, A. Gioiello, S. W. C. van Mil, *Pharmacol. Ther.* **2018**, *191*, 162–177.
- [3] C. Y. Han, *Int. J. Mol. Sci.* **2018**, *19*, 2069.
- [4] T. R. Ahmad, R. A. Haeusler, *Nat. Rev. Endocrinol.* **2019**, *15*, 701–712.
- [5] K. V. Kowdley, V. Luketic, R. Chapman, G. M. Hirschfield, R. Poupon, C. Schramm, C. Vincent, C. Rust, A. Parés, A. Mason, H.-U. Marschall, D. Shapiro, L. Adorini, C. Sciacca, T. Beecher-Jones, O. Böhm, R. Pencek, D.

- Jones, *Obeticholic Acid PBC Monotherapy Study Group, Hepatology* **2018**, *67*, 1890–1902.
- [6] L. Jiang, H. Zhang, D. Xiao, H. Wei, Y. Chen, *Comput. Struct. Biotechnol. J.* **2021**, *19*, 2148–2159.
- [7] S. De Marino, C. Festa, V. Sepe, A. Zampella, in *Bile Acids and Their Receptors* (Eds.: S. Fiorucci, E. Distrutti), Springer International Publishing, Cham, **2019**, 137–165.
- [8] Y. Anami, N. Shimizu, T. Ekimoto, D. Egawa, T. Itoh, M. Ikeguchi, K. Yamamoto, *J. Med. Chem.* **2016**, *59*, 7888–7900.
- [9] D. G. Teotico, M. L. Frazier, F. Ding, N. V. Dokholyan, B. R. S. Temple, M. R. Redinbo, *PLoS Comput. Biol.* **2008**, *4*, e1000111.
- [10] B. D. Darimont, R. L. Wagner, J. W. Apriletti, M. R. Stallcup, P. J. Kushner, J. D. Baxter, R. J. Fletterick, K. R. Yamamoto, *Genes Dev.* **1998**, *12*, 3343–3356.
- [11] B. M. Forman, E. Goode, J. Chen, A. E. Oro, D. J. Bradley, T. Perlmann, D. J. Noonan, L. T. Burka, T. McMorris, W. W. Lamph, R. M. Evans, C. Weinberger, *Cell* **1995**, *81*, 687–693.
- [12] E. G. Schuetz, S. Strom, K. Yasuda, V. Lecureur, M. Assem, C. Brimer, J. Lamba, R. B. Kim, V. Ramachandran, B. J. Komoroski, R. Venkataramanan, H. Cai, C. J. Sinal, F. J. Gonzalez, J. D. Schuetz, *J. Biol. Chem.* **2001**, *276*, 39411–39418.
- [13] D. Merk, S. Sreeramulu, D. Kudlinzki, K. Saxena, V. Linhard, S. L. Gande, F. Hiller, C. Lamers, E. Nilsson, A. Aagaard, L. Wissler, N. Dekker, K. Bamberg, M. Schubert-Zsilavecz, H. Schwalbe, *Nat. Commun.* **2019**, *10*, 2915.
- [14] W. Kilu, D. Merk, D. Steinhilber, E. Proschak, J. Heering, *J. Biol. Chem.* **2021**, *297*, 100814.
- [15] J. Heering, N. Jores, W. Kilu, E. Schallmayer, E. Peelen, A. Muehler, H. Kohlhof, D. Vitt, V. Linhard, S. L. Gande, A. Chaikuad, S. Sreeramulu, H. Schwalbe, D. Merk, *ACS Chem. Biol.* **2022**, *17*, 3159–3168.
- [16] R. T. Gampe, V. G. Montana, M. H. Lambert, G. B. Wisely, M. V. Milburn, H. E. Xu, *Genes Dev.* **2000**, *14*, 2229–2241.
- [17] T. S. Hughes, M. J. Chalmers, S. Novick, D. S. Kuruvilla, M. R. Chang, T. M. Kamenecka, M. Rance, B. A. Johnson, T. P. Burris, P. R. Griffin, D. J. Kojetin, *Structure* **2012**, *20*, 139–150.
- [18] A. K. Shiau, D. Barstad, J. T. Radek, M. J. Meyers, K. W. Nettles, B. S. Katzenellenbogen, J. A. Katzenellenbogen, D. A. Agard, G. L. Greene, *Nat. Struct. Biol.* **2002**, *9*, 359–364.
- [19] A. K. Shiau, D. Barstad, P. M. Loria, L. Cheng, P. J. Kushner, D. A. Agard, G. L. Greene, *Cell* **1998**, *95*, 927–937.
- [20] L. Jin, X. Feng, H. Rong, Z. Pan, Y. Inaba, L. Qiu, W. Zheng, S. Lin, R. Wang, Z. Wang, S. Wang, H. Liu, S. Li, W. Xie, Y. Li, *Nat. Commun.* **2013**, *4*, 1937.
- [21] X. Xu, X. Xu, P. Liu, Z. Zhu, J. Chen, H. Fu, L. Chen, L. Hu, X. Shen, *J. Biol. Chem.* **2015**, *290*, 19888–19899.
- [22] A. Rashidian, E.-K. Mustonen, T. Kronenberger, M. Schwab, O. Burk, S. A. Laufer, T. Pantsar, *Comput. Struct. Biotechnol. J.* **2022**, *20*, 3004–3018.
- [23] A. Stefela, M. Kaspar, M. Drastik, T. Kronenberger, S. Micuda, M. Dracinsky, B. Klepetarova, E. Kudova, P. Pavek, *Front. Pharmacol.* **2021**, *12*, 1980.
- [24] M. Makishima, T. T. Lu, W. Xie, G. K. Whitfield, H. Domoto, R. M. Evans, M. R. Haussler, D. J. Mangelsdorf, *Science* **2002**, *296*, 1313–1316.
- [25] N. Wang, Q. Zou, J. Xu, J. Zhang, J. Liu, *J. Biol. Chem.* **2018**, *293*, 18180–18191.
- [26] E.-K. Mustonen, T. Pantsar, A. Rashidian, J. Reiner, M. Schwab, S. Laufer, O. Burk, *Cells* **2022**, *11*, 1299.
- [27] A. I. Shulman, C. Larson, D. J. Mangelsdorf, R. Ranganathan, *Cell* **2004**, *116*, 417–429.
- [28] C. Zhang, Y. Liu, Y. Wang, X. Ge, T. Jiao, J. Yin, K. Wang, C. Li, S. Guo, X. Xie, C. Xie, F. Nan, *J. Med. Chem.* **2022**, DOI 10.1021/acs.jmedchem.2c01394.
- [29] G. Madhavi Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, W. Sherman, *J. Comput.-Aided Mol. Des.* **2013**, *27*, 221–234.
- [30] A. Akwabi-Ameyaw, J. Y. Bass, R. D. Caldwell, J. A. Caravella, L. Chen, K. L. Creech, D. N. Deaton, S. A. Jones, I. Kaldor, Y. Liu, K. P. Madauss, H. B. Marr, R. B. McFadyen, A. B. Miller, F. Navas, D. J. Parks, P. K. Spearing, D. Todd, S. P. Williams, G. B. Wisely, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 4339–4343.
- [31] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, P. S. Shenkin, *J. Med. Chem.* **2004**, *47*, 1739–1749.
- [32] R. O. Dror, M. Ø. Jensen, D. W. Borhani, D. E. Shaw, *J. Gen. Physiol.* **2010**, *135*, 555–62.
- [33] E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J. Y. Xiang, L. Wang, D. Lupyan, M. K. Dahlgren, J. L. Knight, J. W. Kaus, D. S. Cerutti, G. Krilov,

- W. L. Jorgensen, R. Abel, R. A. Friesner, *J. Chem. Theory Comput.* **2016**, *12*, 281–296.
- [34] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926–935.
- [35] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- [36] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- [37] G. J. Martyna, M. L. Klein, M. Tuckerman, *J. Chem. Phys.* **1992**, *97*, 2635–2643.
- [38] G. J. Martyna, M. E. Tuckerman, D. J. Tobias, M. L. Klein, *Mol. Phys.* **1996**, *87*, 1117–1157.
- [39] J. Marelius, K. Kolmodin, I. Feierberg, J. Aqvist, *J. Mol. Graphics Modell.* **1998**, *16*, 213–225, 261.
- [40] W. Jaspers, M. Esguerra, J. Åqvist, H. Gutiérrez-de-Terán, *J. Cheminf.* **2019**, *11*, 26.
- [41] F. S. Lee, A. Warshel, *J. Chem. Phys.* **1992**, *97*, 3100–3107.
- [42] G. King, A. Warshel, *J. Chem. Phys.* **1989**, *91*, 3647–3661.
- [43] J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327–341.
- [44] J. L. Banks, H. S. Beard, Y. Cao, A. E. Cho, W. Damm, R. Farid, A. K. Felts, T. A. Halgren, D. T. Mainz, J. R. Maple, R. Murphy, D. M. Philipp, M. P. Repasky, L. Y. Zhang, B. J. Berne, R. A. Friesner, E. Gallicchio, R. M. Levy, *J. Comput. Chem.* **2005**, *26*, 1752–1780.
- [45] M. K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J.-H. Prinz, F. Noé, *J. Chem. Theory Comput.* **2015**, *11*, 5525–5542.
- [46] C. Wehmeyer, M. K. Scherer, T. Hempel, B. E. Husic, S. Olsson, F. Noé, *Living J. Comput. Mol. Sci.* **2019**, *1*, 5965–5965.
- [47] H. Wu, F. Noé, *arXiv:1707.04659 [math, stat]* **2019**.
- [48] G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, F. Noé, *J. Chem. Phys.* **2013**, *139*, 015102.
- [49] S. Röblitz, M. Weber, *Advances in Data Analysis and Classification* **2013**, *7*, 147–179.
- [50] P. Metzner, C. Schütte, E. Vanden-Eijnden, *Multiscale Model. Simul.* **2009**, *7*, 1192–1219.
- [51] F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, T. R. Weigl, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 19011–19016.
- [52] J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, F. Noé, *J. Chem. Phys.* **2011**, *134*, 174105.
- [53] J.-H. Prinz, J. D. Chodera, F. Noé, in *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation* (Eds.: G. R. Bowman, V. S. Pande, F. Noé), Springer Netherlands, Dordrecht, **2014**, 45–60.
- [54] K. Beekmann, L. H. J. de Haan, L. Actis-Goretta, R. Houtman, P. J. van Bladeren, I. M. C. M. Rietjens, *J. Steroid Biochem. Mol. Biol.* **2015**, *154*, 245–253.
- [55] A. Koppen, R. Houtman, D. Pijnenburg, E. H. Jeninga, R. Ruijtenbeek, E. Kalkhoven, *Mol. Cell. Proteomics* **2009**, *8*, 2212–2226.

Manuscript received: October 11, 2022
Revised manuscript received: November 15, 2022
Accepted manuscript online: November 17, 2022
Version of record online: November 30, 2022

ChemMedChem

Supporting Information

When Two Become One: Conformational Changes in FXR/RXR Heterodimers Bound to Steroidal Antagonists

Alejandro Díaz-Holguín⁺, Azam Rashidian⁺, Dirk Pijnenburg, Glaucio Monteiro Ferreira, Alzbeta Stefela, Miroslav Kaspar, Eva Kudova, Antti Poso, Rinie van Beuningen, Petr Pavek, and Thales Kronenberger*

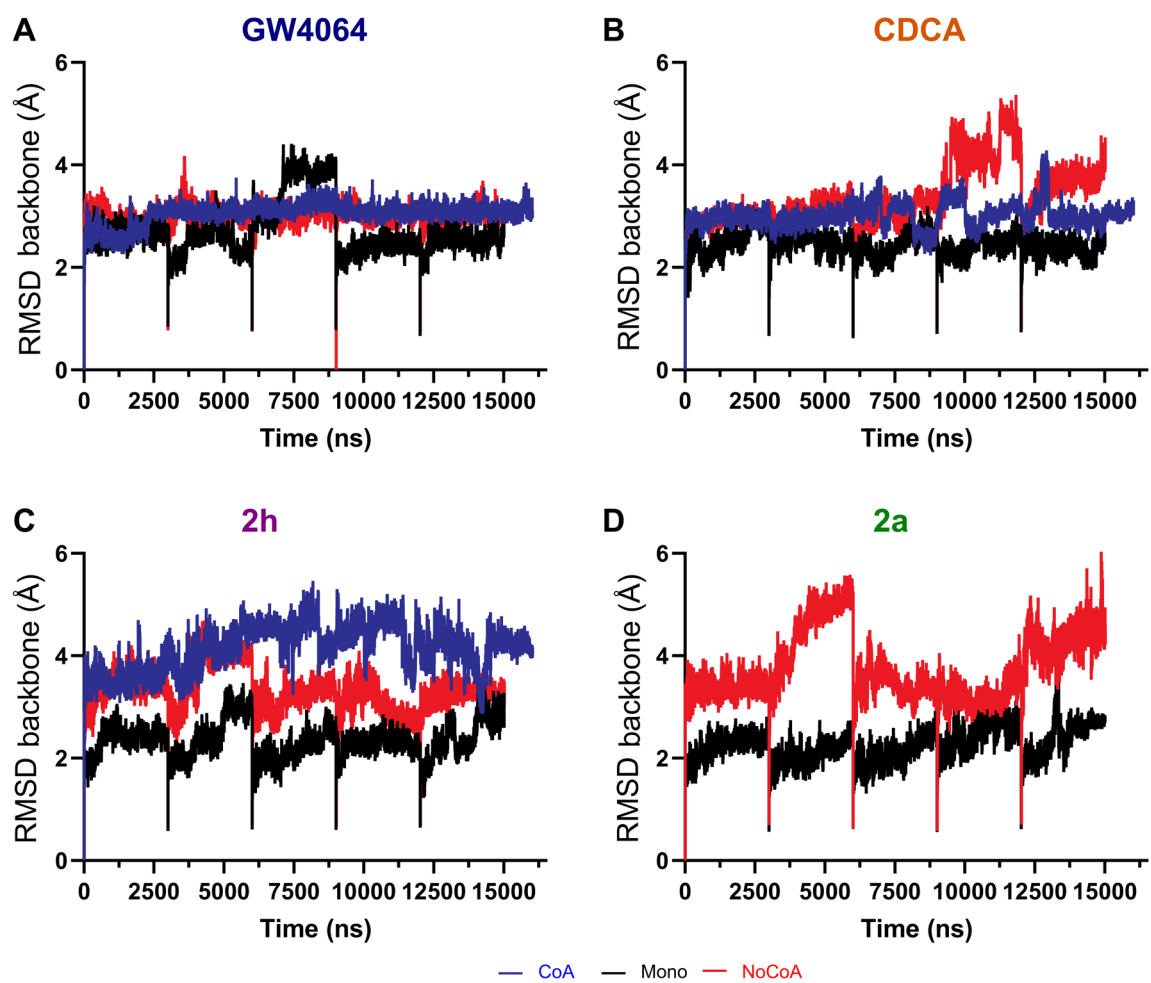


Figure S1. Root-mean-square deviation of the protein backbone along the merged trajectory time (RMSD) for the compounds GW4064 (A), CDCA (B), 2h (C) and 2a (D). Simulations as monomers are displayed as black lines, heterodimers with RXR (as red) and with RXR+CoA as blue lines.

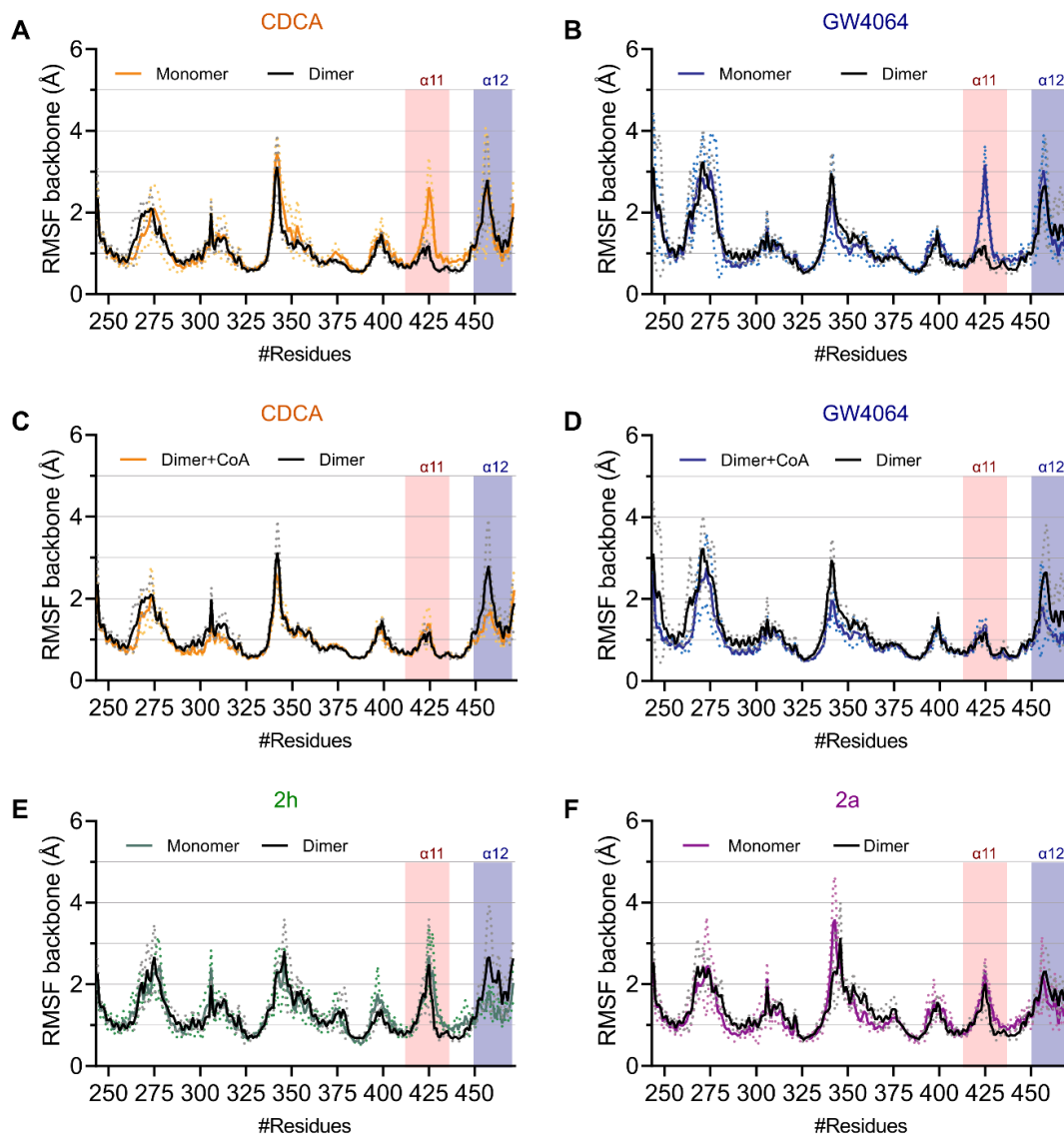


Figure S2. Root-mean-square fluctuation (RMSF) of the protein backbone. Comparing heterodimers with monomers (A,B,E,F, for all tested ligands as highlighted in the labels and colours) and against heterodimers + CoA (C,D, for GW4064 and CDCA). For each graphic $\alpha 11$ (salmon) and $\alpha 12$ (blue) helices are highlighted. Bold lines represent the means of five replicas and dashed lines are their respective standard deviations.

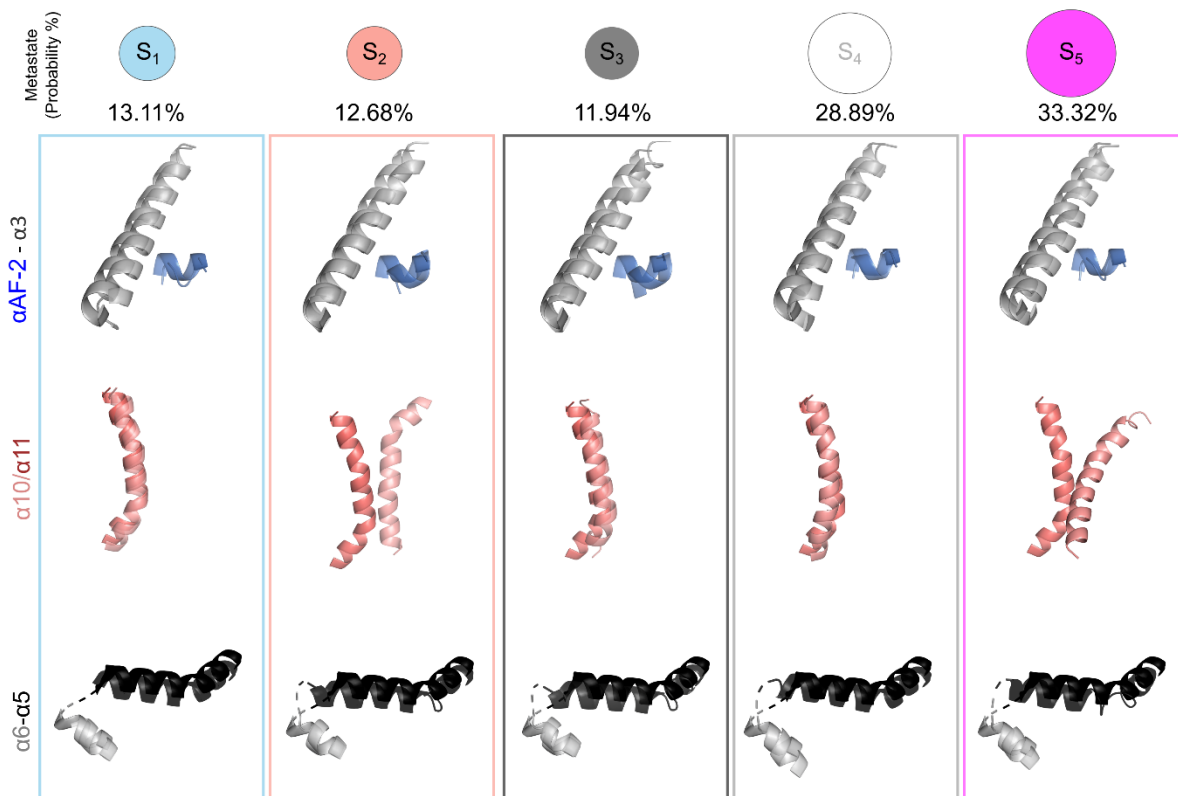


Figure S3. Metastable states of relevant structural parts of FXR, as revealed by Markov state modelling (probability of each state is displayed as a number and further information can be found in Figure S3). Each metastable state (S) is illustrated with a representative structure (solid coloured cartoon), superimposed on the original crystal structure (transparent cartoon).

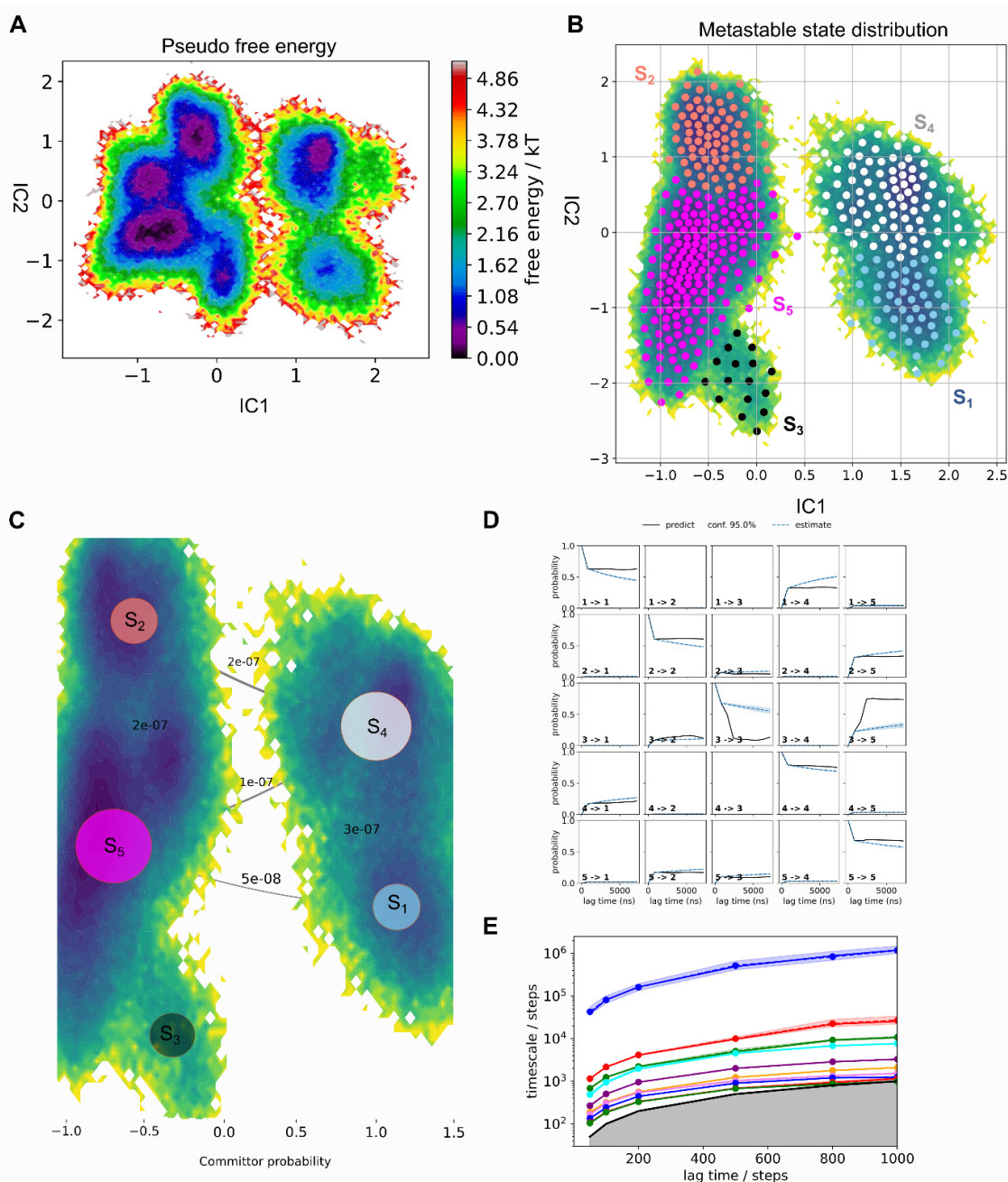


Figure S4. Metastable states of FXR in the presence and absence of the ligands as revealed by Markov state modelling. (A) Pseudo free energy map of distributions along with time-lagged independent components (ICs) 1 and 2. (B) Separation of five metastable states (S1–S5) by PCCA++ analysis. (C) Committor probability of the most representative metastable states. Each metastable state (S) is illustrated with representative structures (coloured ribbons), superimposed to a transparent cartoon with the original crystal structure in the Figure 3 (main manuscript). Equilibrium probability (π_i) of distributions for each state is indicated below the conformations together with circles with an area representing the changes induced by the systems, values near to 1.0 more probable than values close to 0. Validation of Markov State Models. Chapman-Kolmogorov tests demonstrate that the models follow expected estimates to certain extent (D). In the selected lag times the implied timescales are converged (E).

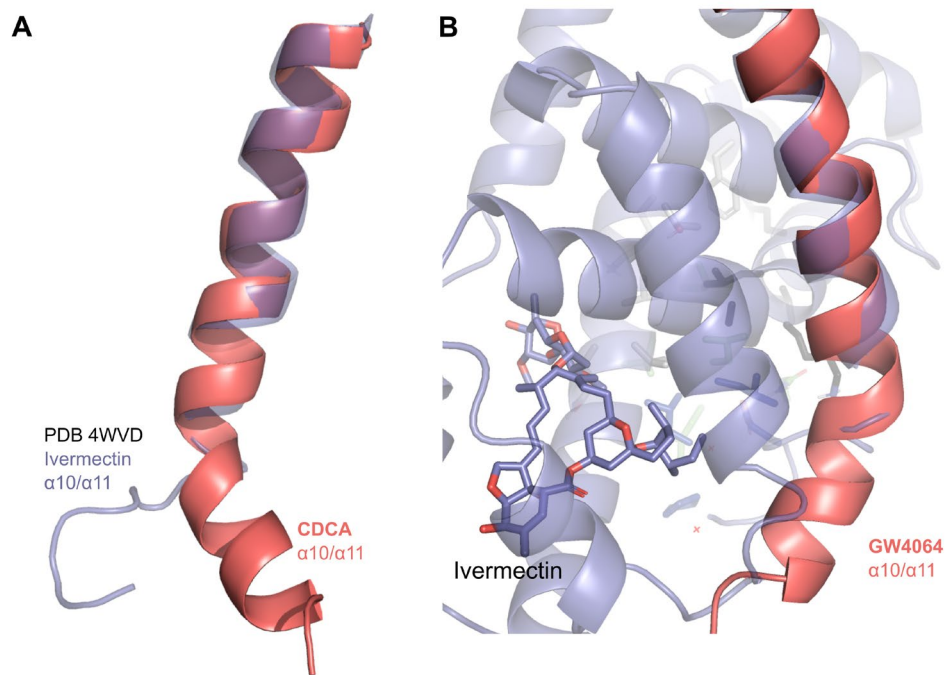


Figure S5. Cartoon representation of the structural diversity on the $\alpha 11$ region, where the ivermectin structure (PDB ID 4WVD), displayed in purple, is compared to representative simulation frames for CDCA (A, salmon) and GW4064 (B, salmon) displays a large unfolding and bending. (B) displays the coordinates of Ivermectin compound as sticks.

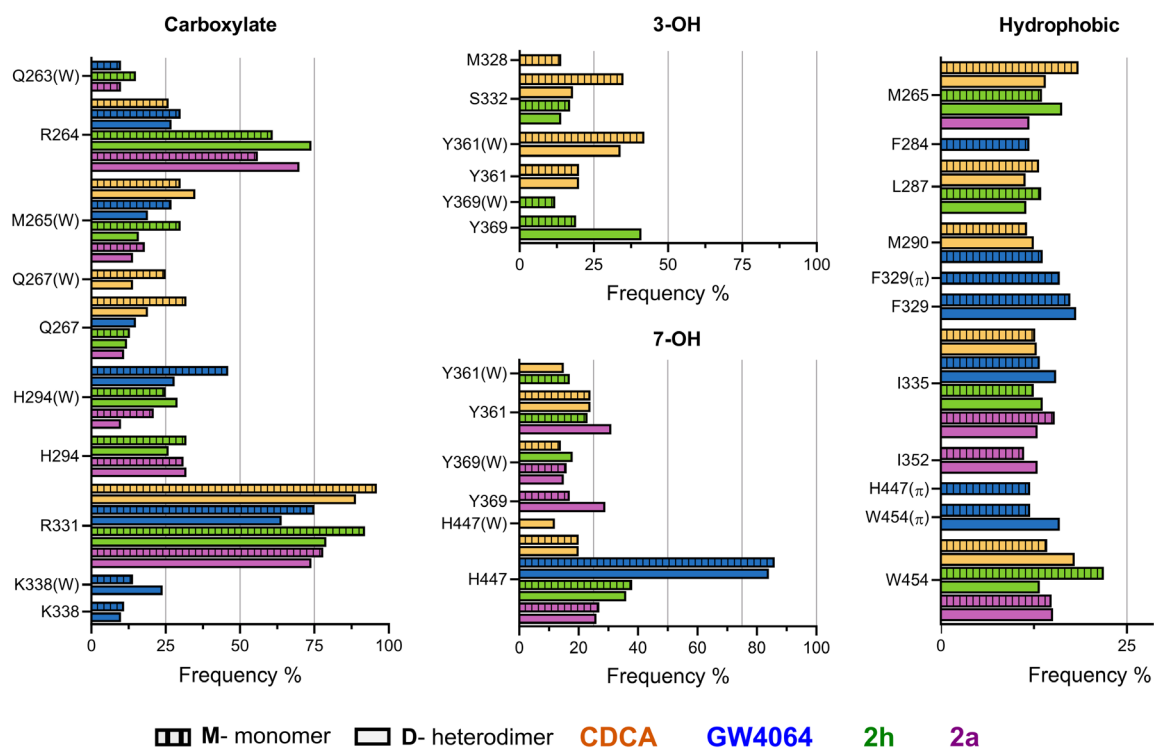
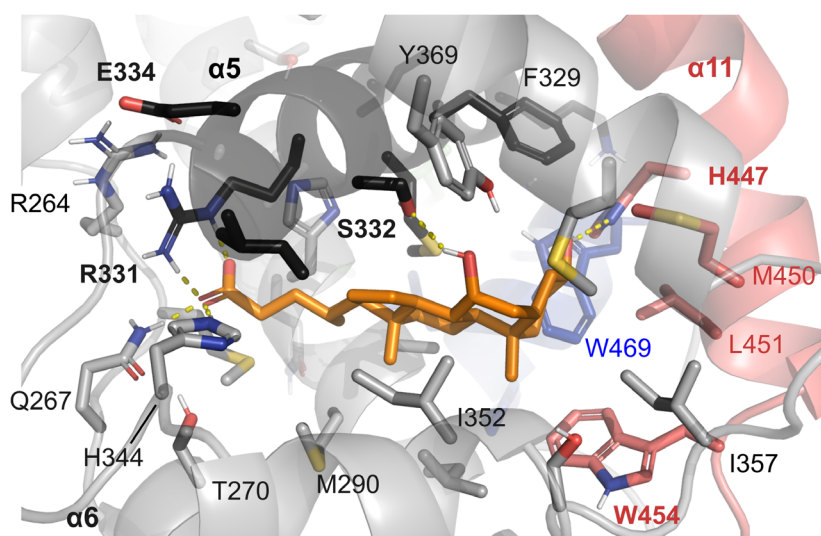


Figure S6. Protein ligand interactions for the different simulated ligands. Upper panel, Representation of the ligand-binding pocket of FXR bound to CDCA (orange) and highlighting as sticks all relevant residues in the vicinity. Down panels, dashed bars represent monomeric simulations and clear colours are their respective heterodimers without CoA.

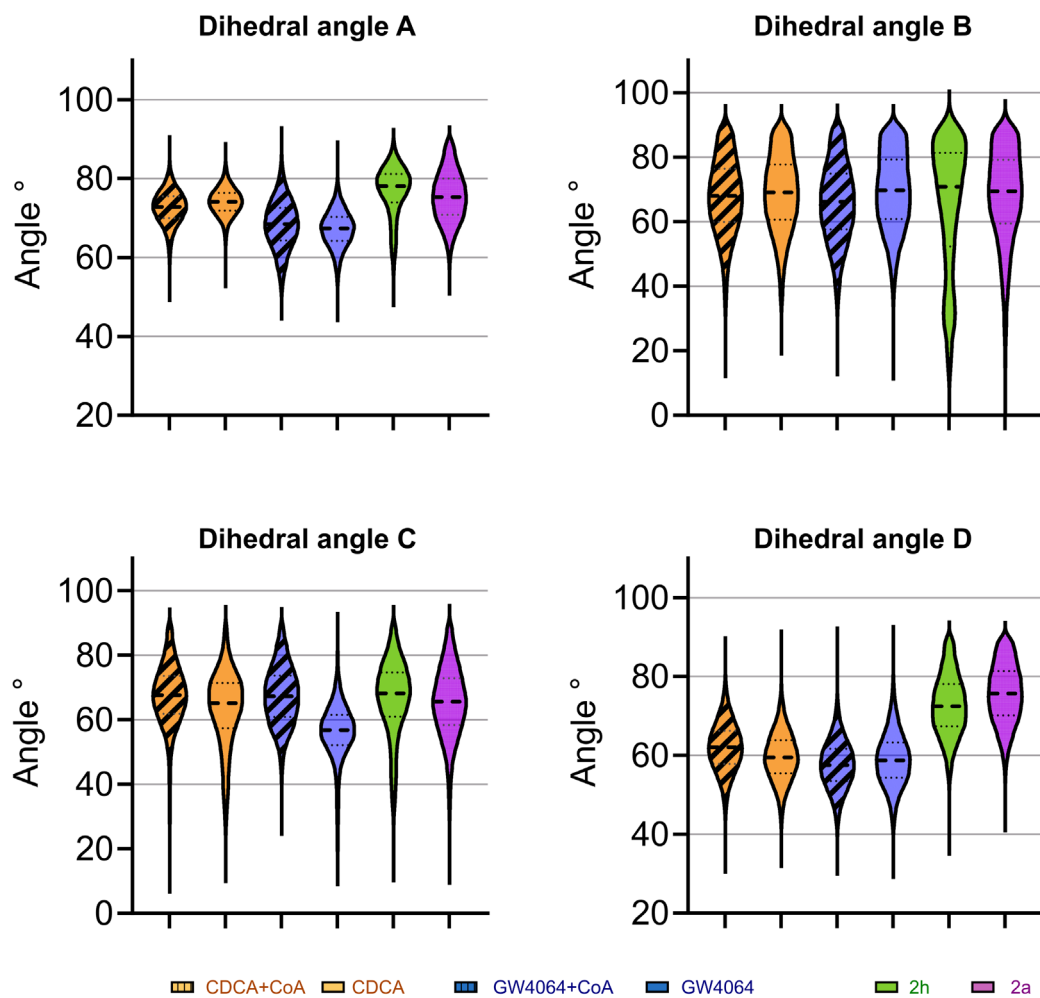


Figure S7. dihedral angle between the planes composed by $\alpha 7$, $\alpha 9$, $\alpha 10$ and $\alpha 11$ in combination (See Tables S1,S2).

Table S1. Planes definition, planes were defined by the carbon alpha of the extremities of residues in the following list order.

Plane number	FXR (Residue numbers)	RXR (Residue numbers)	Secondary structure
1	442-451	426-441	$\alpha 11$
2	427-441	412-425	$\alpha 10$
3	401-421	388-407	$\alpha 9$
4	359-374	342-360	$\alpha 7$

Table S2. Planar angle definition

Dihedral angle	FXR plane number	RXR plane number
A	1	1
B	2	2
C	3	4
D	4	3

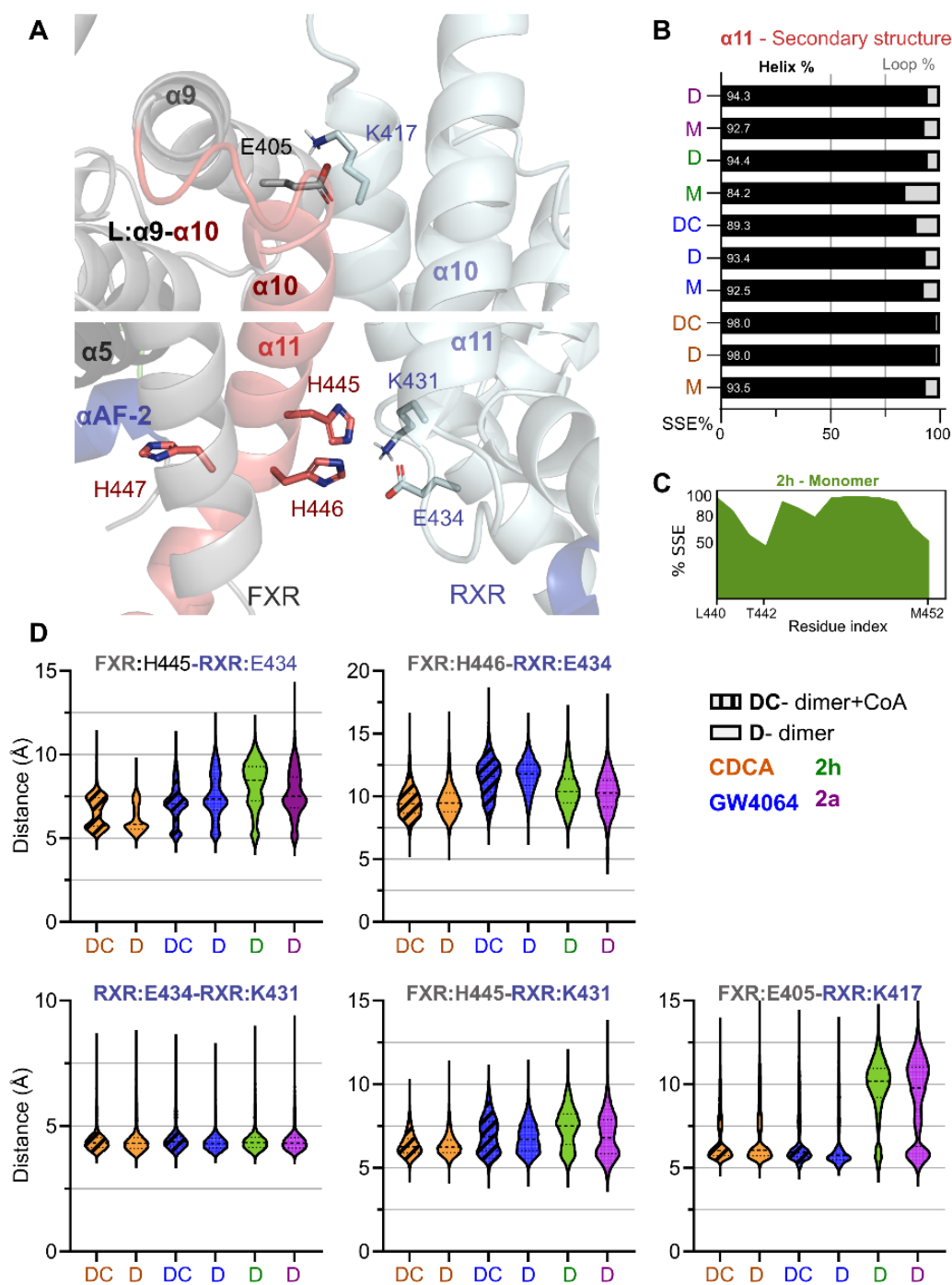


Figure S8. Heterodimerization interface between FXR/RXR. A) overview of the FXR/RXR interface highlighting the relevant amino acids to sustain the interface as sticks (divided into upper and down regions). B) Secondary structure % (SSE% of helical structures) of $\alpha 11$ helix appears to be stable with most of the compounds, with exception of 2h-monomeric simulations. C) Area plots represent the observed secondary structure element (SSE) of the $\alpha 11$ helix in percentage throughout the 2h-monomeric simulation, displaying a short unfolding and therefore a bend associated. D) Distances between centre of mass for relevant residues in the interface.

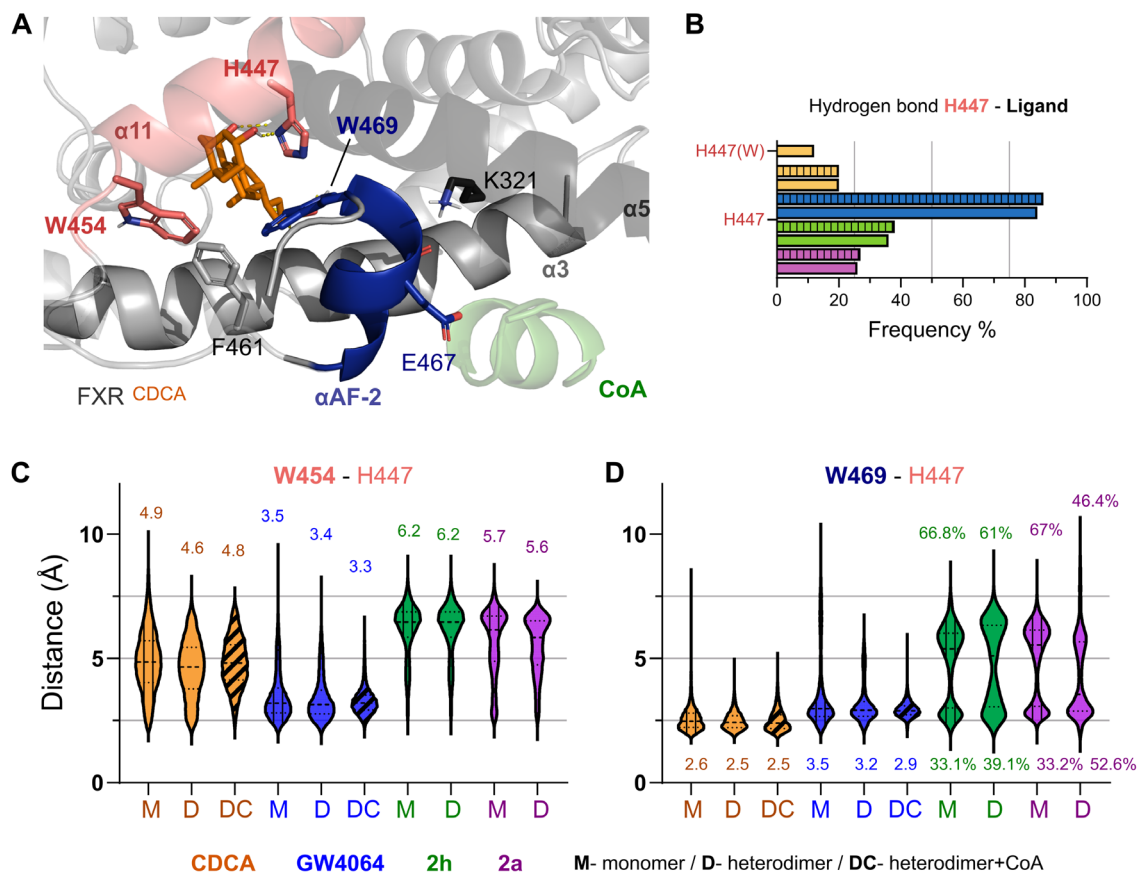


Figure S9. Antagonist induced conformational changes in the α AF-2 region. (A) Location of the α AF-2 helix and surrounding region in the FXR-LBD. (B) His447-Ligand interactions, dashed bars represent the monomeric simulations and clear colours are their respective heterodimers without CoA. Minimum distances between one of the His447 and Trp454 (C) and Trp469 (D) sidechain. All distance plots are depicted as violin plots where the median of the largest continuous distribution is displayed as a number, for Trp469 antagonistic distances the frequency distributions are depicted instead.

Table S3. MARCoNI assay summary table. Modulation Index (MI). FXR-peptide binding was quantified using Bionavigator software (see methods). MI is the \log_{10} -transformed relative binding value, which is calculated by the compound's binding value, relative to the vehicle control (DMSO) binding value. NCOA2_733_755 peptide in bold contains the motif used for GW4064's CoA simulations.

ID	Sequence	Motif	2h	6-OCA	GW4064
NCOA1_1421_1441	TSGPQTPQAQQKSLQLQLLQTE	LxxLL1435	0.06	1.17	1.55
NCOA1_620_643	SDGDSKYSQTSKHLVQLLTTAAEQ	LxxLL633	-0.02	-0.02	-0.02
NCOA1_677_700	PSSHSLTERHKILHRLLQEGSPS	LxxLL690	0.03	0.69	1.09
NCOA1_737_759	ASKKKESKDHQLLRYLLDKDEKD	LxxLL749	-0.02	-0.08	0.47
NCOA2_628_651	GQSRLHDSKGQTKLLQLLTTKSDQ	LxxLL641	0.00	0.00	0.04
NCOA2_677_700	STHGTSLKEKHILHRLLQDSSSP	LxxLL690	0.41	1.38	1.69
NCOA2_733_755	EPVSPKKKENALLRYLLDKDDTK	LxxLL745	-0.16	0.89	1.30
NCOA2_866_888	SQSTFNNPRPQQLGRLLPNQNLP	LxxLL878	-0.12	-0.21	-0.19
NCOA3_104_123_N-KKK	KKKGQGVIDKDSLGPLLLQALDG	LxxLL113	0.00	0.00	0.00
NCOA3_609_631	QRGPLESKGHKLLQLLTCSSDD	LxxLL621	-0.03	-0.03	0.28
NCOA3_609_631_C627S	QRGPLESKGHKLLQLLTSSSDD	LxxLL621	0.00	0.29	0.59
NCOA3_673_695	MHGSLLEKHLRILHKLQNGNSP	LxxLL685	0.00	1.01	1.31
NCOA3_725_747	EQLSPKKKENALLRYLLDRDDP	LxxLL738	-0.21	0.72	1.18
NCOR1_1925_1946	TTITAAANFIDVIITRQIASDKD	lxxxlxxxl1933	0.00	0.00	0.11
NCOR1_2039_2061	MGQVPRTHRLITLADHICQIITQ	LxxHI2051	-0.24	-0.38	-0.26
NCOR1_2039_2061_C2056S	MGQVPRTHRLITLADHISQIITQ	LxxHI2051	-0.08	-0.30	-0.12

NCOR1 2251 2273	GHSFADPASNLGLEDIIRKALMG	LxxII2263	-0.12	-0.46	-0.43
NCOR1 2376 2398	SSTGSTQFPYNPLTMRMLSSTPP	TxxML2389	0.00	0.00	0.05
NCOR1 662 684 C662S	SKNFYFNYKRRHNLNLLQQHKQ	LxxLL675	0.12	-0.29	0.05
NCOR2 2123 2145	APGVKGHQRVVTLAQHISEVITQ	LxxHI2135	0.00	0.00	0.00
NCOR2 2330 2352	QAVQEHASTNMGLEAIIRKALMG	LxxII2342	0.02	-0.76	-0.49
NCOR2 649 671 C649S	SKNFYFNYKKRQNLDEILQQHKL	LxxIL662	0.18	-0.15	0.20

Table S4. Interaction frequency between the coactivator side chains and the closest amino acids from FXR

CDCA			2h		
CoA	FXR	Freq%	CoA	FXR	Freq%
Hydrogen bond side chains – side chains					
Asn742	Glu467	66.3			
Asn742	Lys321	36.7	Asn742	Lys321	45.3
Salt-bridges					
Lys751	Glu300	36.9	Lys740	Gln472	34.7
Arg746	Glu314	83.2	Arg746	Glu314	98.6
Lys740	Glu314	29.2	Lys740	Asp470	72.0
Asp752	Lys303	71.1	Asp750	Lys303	74.6
Hydrogen bond backbone – side chains					
Asn742	Glu467	34.0	Lys740	Gln472	27.0
Leu748	Lys303	63.1	Leu748	Lys303	91.0
			Leu749	Gln309	23.8
Hydrogen bond side chains – backbone					
Leu744	Glu467	49.5	Ala743	Glu467	90.6
Leu745	Glu467	45.0	Asn742	Glu467	65.9
			Glu741	Glu467	82.4

10. Publication IV

Discovery of Receptor Agonists with the Imidazo[1,2-a]pyridine Structure Novel Human Constitutive Androstane

Ivana Mejdrová^{a,‡}, Jan Dušek^{b,‡}, Kryštof Škach^a, Alžbeta Stefela^b, Josef Skoda^b, Karel Chalupská^c, Klara Dohnalová^d, Ivona Pavkova^e, Thales Kronenberger^{f,g}, Azam Rashidian^f, Lucie Smutná^b, Vojtěch Duchoslav^a, Tomas Smutny^b, Petr Pávek^{b*}, Radim Nencka^{a*}

^a Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, Flemingovo nám. 2, 166 10 Prague 6, Czech Republic; ^bDepartment of Pharmacology and Toxicology, Faculty of Pharmacy in Hradec Kralove, Charles University, Akademika Heyrovskeho 1203, 500 05, Hradec Kralove, Czech Republic; ^cCzech Centre for Phenogenomics, Institute of Molecular Genetics of the Czech Academy of Sciences, Vídeňská 1083, 142 20, Prague; Czech Republic; ^d 1 st Medical Faculty, Charles University, Katerinska 32, 112 08, Prague, Czech Republic; ^eFaculty of Military Health Sciences, University of Defense, Trebeska 1575, Hradec Kralove, 500 01, Czech Republic; ^f Department of Internal Medicine VIII, University Hospital of Tübingen, 72076, Tübingen, Germany; ^gSchool of Pharmacy, Faculty of Health Sciences, University of Eastern Finland, 70211, Kuopio, Finland, and Department of Pharmaceutical and Medicinal Chemistry, Institute of Pharmaceutical Sciences, Eberhard Karls Universität, 72076, Tübingen, Germany.

‡- contributed equally; *-corresponding authors

Journal of Medicinal Chemistry

**Permission was granted from American Chemical Society (ACS). Further permissions related to the material excerpted should be directed to the ACS
2023, 66, 4, pages 2422–2456**

doi: <https://pubs.acs.org/doi/10.1021/acs.jmedchem.2c01140>

supporting information:

<https://pubs.acs.org/doi/10.1021/acs.jmedchem.2c01140>

Hypothesis: CAR has emerged as a promising target receptor for metabolic or liver disease therapy. Consequently, harnessing the potential of a selective CAR agonist can enhance CAR's therapeutic relevance while also reducing the significant pharmacological overlap between human CAR and PXR.

Aims: To explore selective human CAR agonists with no activation of PXR or other nuclear receptors, while maintaining favourable ADME characteristics for subsequent investigations in human hepatocyte cellular models or application in humanized CAR mouse models. This research is conducted within a library of kinase inhibitors.

Results: For the modelling analyses, we performed docking studies with compounds 37, 39, 40, and 48 within the CAR-LBD. The simulation analysis reveals that, our novel CAR ligands interacted with His203 and occupied a hydrophobic pocket in the human wtCAR-LBD. Moreover, we found that H203 played a crucial role in stabilizing compound 39, and T225 and D228 also contributed significantly to its stabilization.

Interestingly, all the newly discovered compounds displayed U-shaped conformations similar to CITCO within the wtCAR-LBD. The observed conformations were predominantly influenced by hydrophobic interactions, with compound 39 uniquely interacting with I164 and establishing additional hydrogen bonds with T225 and D228. Notably, H12 was observed to be situated near H3, and interestingly, we noted geometric stabilization facilitated by the N165-Y326 interaction. Our comprehensive analysis, starting from docking data and extending to molecular dynamic (MD) simulations, uncovered that both CITCO and compound 39 engage with wtCAR-LBD, predominantly through hydrophobic contacts. Notably, compound 39 formed stronger polar contacts with wtCAR-LBD than CITCO, owing to the formation of hydrogen bonds between the amide moiety of compound 39 and the backbone oxygen of T225 and D228. Upon analysing the MD trajectories, we observed that compound 39 demonstrated a noteworthy interaction with I164 and exhibited a higher interaction frequency with Y326 in comparison to CITCO. These observations highlight the critical role of H3 and H10/H11 in protein stabilization. While the library of compounds yielded extremely potent CAR agonists, we also considered our assessments of metabolic stability and activity towards PXR. As a result, compound 39 emerged as the chosen candidate for further investigations. While it may not rank among our most potent CAR agonists, its favorable CAR/PXR profile and reasonable metabolic stability render it well-suited for subsequent in vivo experiments. In vitro analysis revealed the selectivity

of compound 39 for other nuclear receptors. Moreover, a pilot single-dose pharmacokinetic study indicated that compound 39 effectively acted as a novel human CAR agonist in animal experiments, demanding further investigation in repeated-dose long-term proof-of-concept studies. Importantly, the chemical tool used in our studies showed no observable toxicity or genotoxic potential, providing additional evidence for the significant activation of human CAR by compound 39.

Conclusion: To summarize, our research has identified a selective CAR receptor agonist, exhibiting activities in relevant in vitro and in vivo models for the human CAR. Notably, Compound 39 shows promise additional preclinical investigations in humanized CAR models or human hepatocyte models. By exploring compound 39 further, we can gain a more holistic comprehension of the distinct functionalities of the human CAR, while mitigating any potential off-target effects on PXR receptor activation.

Discovery of Novel Human Constitutive Androstane Receptor Agonists with the Imidazo[1,2-*a*]pyridine Structure

Ivana Mejdrová,[○] Jan Dušek,[○] Kryštof Škach, Alžbeta Stefela, Josef Skoda, Karel Chalupský, Klára Dohnalová, Ivona Pavkova, Thales Kronenberger, Azam Rashidian, Lucie Smutná, Vojtěch Duchoslav, Tomas Smutny, Petr Pávek,* and Radim Nencka*



Cite This: *J. Med. Chem.* 2023, 66, 2422–2456



Read Online

ACCESS |



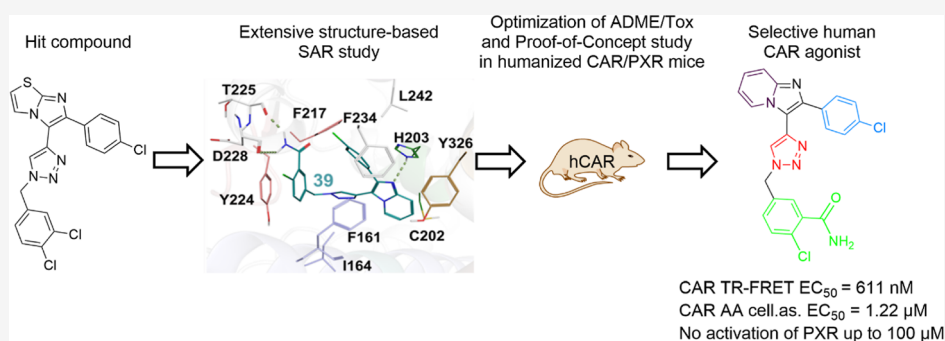
Metrics & More



Article Recommendations



Supporting Information



ABSTRACT: The nuclear constitutive androstane receptor (CAR, NR1I3) plays significant roles in many hepatic functions, such as fatty acid oxidation, biotransformation, liver regeneration, as well as clearance of steroid hormones, cholesterol, and bilirubin. CAR has been proposed as a hypothetical target receptor for metabolic or liver disease therapy. Currently known prototype high-affinity human CAR agonists such as CITCO (6-(4-chlorophenyl)imidazo[2,1-*b*][1,3]thiazole-5-carbaldehyde-*O*-(3,4-dichlorobenzyl)-oxime) have limited selectivity, activating the pregnane X receptor (PXR) receptor, a related receptor of the NR1I subfamily. We have discovered several derivatives of 3-(1*H*-1,2,3-triazol-4-yl)imidazo[1,2-*a*]pyridine that directly activate human CAR in nanomolar concentrations. While compound 39 regulates CAR target genes in humanized CAR mice as well as human hepatocytes, it does not activate other nuclear receptors and is nontoxic in cellular and genotoxic assays as well as in rodent toxicity studies. Our findings concerning potent human CAR agonists with in vivo activity reinforce the role of CAR as a possible therapeutic target.

INTRODUCTION

The constitutive androstane receptor (CAR, NR1I3) is a ligand-activated transcription factor belonging to the nuclear receptor subfamily NR1I.

Human CAR is dominantly expressed in hepatocytes. While the endogenous ligands of human CAR are obscure, a number of naturally occurring steroids such as androstanol, androsteneol, and 5 β -pregnane-3,20-dione have been proposed as endogenous inverse agonists in supraphysiological concentrations.^{1,2} Recent animal studies with a mouse agonist suggest that CAR plays an important role in the metabolism of glucose, lipids, and fatty acids as well as in the endobiotic metabolism of bile acids, cholesterol, bilirubin, and thyroid hormones.³ It has been proposed in several independent animal studies that CAR activation may ameliorate glucose homeostasis and insulin sensibility in the treatment of type 2 diabetes.^{4,5} In addition, since CAR activation affects the expression of lipogenic genes in mice, this might also be a promising therapeutic intervention in the treatment of human obesity, steatosis, or hypercholesterolemia,^{4,6–9} although contradictory

and species-specific reports also exist.^{10–12} CAR activators have been also proposed as a potential therapy for steatohepatitis or liver regeneration.^{13,14}

So far, only two human CAR crystal structures with a human agonist bound have been reported.¹⁵ The CAR ligand-binding domain (LBD) cavity has a mostly hydrophobic and flexible character with a pocket size of 675 Å.^{3,8,16} The hydrophobic cavity suggests that human CAR ligands are mostly highly lipophilic compounds.

Human CAR displays unique properties in comparison with other nuclear receptors as well as its rodent orthologues. CAR variant 1 (wtCAR, CAR1, and wild-type CAR) exhibits strong constitutive activity that can be further activated by agonists or

Received: July 15, 2022

Published: February 9, 2023



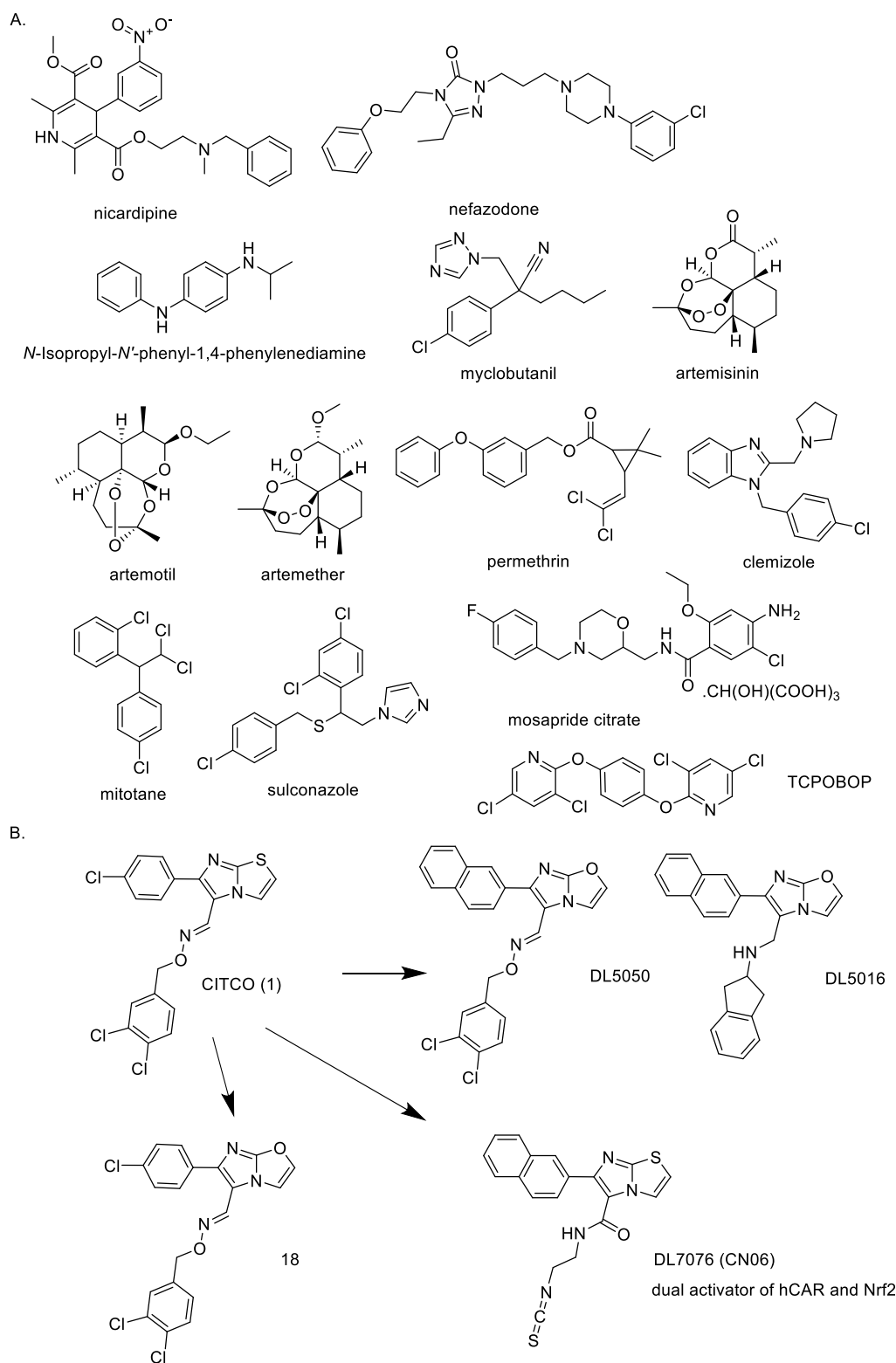


Figure 1. CAR activators discovered by screening chemical libraries (A) or modifications of CITCO as the lead compound (B). TCPOBOP is a mouse CAR ligand.

repressed by inverse agonists. In addition, both direct LBD-dependent and LBD-independent activation are known for CAR.

Human CAR is present in at least three transcript variants (wtCAR, CAR2, and CAR3) in the liver, which differ in their ligand-dependent activation and basal constitutive activities.

The wild-type variant CAR (348 AA, NM_005122.4, and transcript variant 3) features high constitutive activity in the regulation of basal expression of target genes and high sensitivity for inverse agonists. This variant represents about 40% of CAR transcripts in the liver parenchyma. The variant CAR3, also called CAR-SV2 (353 AA, XM_005245697.4,

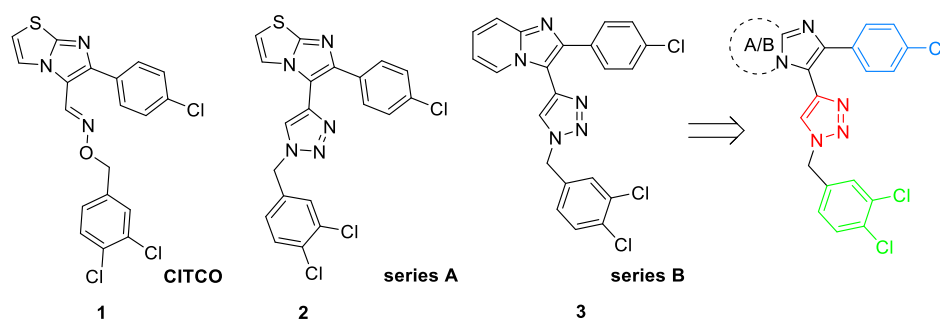


Figure 2. Known human CAR ligand compound **1** (CITCO) and two lead structures (compounds **2** and **3**) with areas of modification in substituted phenyl ring (blue), central heterocyclic linker (red), and substituted benzyl ring (green).

transcript variant X4), which has an insertion of the five amino acids APYLT into the LBD, represents 50% of transcripts. CAR3 has low constitutive activity but is highly inducible by ligands and much more active in the upregulation of CAR target genes in the liver. The transcript variant CAR2 (352 AA, NM_001077480.2) is a minor variant with moderate induction activity. The exact physiological functions of the variants are obscure, but several selective activators of individual variants have been described in the literature.^{8,17–19}

There are no highly potent, specific, and drug-like (with suitable physicochemical and ADME properties) agonists of the human CAR receptor without off-target effects that can be therapeutically used or can serve as a tool in therapeutic intervention with human CAR ligands. The unique properties of human CAR, mainly its hydrophobic pocket and high constitutive activity, make the discovery of specific ligands difficult.²⁰ Therefore, determining suitable drug candidate molecules targeting human CAR and high-affinity endogenous ligands remains problematic.²¹

The only compound known to date is 6-(4-chlorophenyl)imidazo[2,1-*b*]thiazole-5-carbaldehyde *O*-(3,4-dichlorobenzyl)oxime (CITCO, **1**), which is a potent human—but not a mouse—CAR agonist.²² However, this highly lipophilic compound also significantly activates the related pregnane X receptor (NR1I2, PXR) of the same subfamily through π - π interactions with the W299 residue.^{22–24} This may exert an unfavorable effect on glycemia and liver steatosis.²⁵ On the contrary, the prototype mouse CAR ligand 1,4-bis[(3,5-dichloropyridine-2-yl)oxy]benzene (TCPOBOP) does not activate human CAR.²⁶

Different strategies have been used in high-content CAR ligand screenings recently performed, including nuclear translocation assays with an adenoviral-enhanced yellow fluorescent protein-tagged hCAR (Ad/EYFP-hCAR) vector in hepatocytes,^{27,28} mammalian one-hybrid assays using a fusion protein of CAR or its LBD,^{21,23,29–31} and assays employing stable luciferase reporter cell lines expressing wtCAR and treated with an inverse agonist,³² as well as with a CAR3-selective screening method combined with other CAR assays.³³ In addition, studies employing pharmacophore computational modeling and the virtual screening of chemical databases have been performed.^{30,34}

In the past, several CAR activators with various structural features have been discovered in the screened libraries (Figure 1A) or after modification of the lead compound CITCO (**1**), Figure 1B.^{27–29,33–37}

These human CAR ligands, however, still have limited potency to activate human CAR in nanomolar concentrations in comparison with the prototype high-affinity CAR ligand

CITCO. Limited studies are currently being undertaken which explore structure–activity relationship variations by systematic synthesis on the human CAR ligand after the initial hit compound discovery or modification of the human CAR agonist CITCO as a template.

Recently, Liang et al. specifically modified the 4-chlorophenyl, imidazothiazole, and 3,4-dichlorophenyl groups of CITCO.³⁶ Especially, their discovered compound (*E*)-6-(4-chlorophenyl)imidazo[2,1-*b*]oxazole-5-carbaldehyde *O*-(3,4-dichlorobenzyl)oxime and compound DLS050, ((*E*)-6-(naphthalen-2-yl)imidazo[2,1-*b*]oxazole-5-carbaldehyde *O*-(3,4-dichlorobenzyl)oxime) with the imidazoxazole core exert increased potency and selectivity for human CAR activation over human PXR in a human CAR1-expressing reporter cell line and primary human hepatocytes (PHH).³⁶ In an additional study, Liang et al. synthesized a library of CITCO analogues with the 6-(4-chlorophenyl)imidazo[2,1-*b*]oxazole core as well as with modified 4-chlorophenyl or 3,4-dichlorophenyl rings with a variety of substituted arene moieties. In all these novel compounds, the oxime linker of CITCO, which might cause chemical instability, was replaced by groups such as amine, amide, imine, and ether.³⁷ In their study, compound DL5016 (*N*-((6-(naphthalen-2-yl)imidazo[2,1-*b*]oxazol-5-yl)methyl)-2,3-dihydro-1*H*-inden-2-amine), which has an EC₅₀ value of 0.66 μ M in cellular reporter assays, appeared as an efficient and selective human CAR agonist with lower PXR activation than CITCO. In addition, the ligand was shown to induce receptor translocation into the nucleus, to upregulate the expression of the human CAR target gene, and to enhance the efficacy of cyclophosphamide-based cytotoxicity to non-Hodgkin lymphoma cells (Figure 1).³⁷ Very recently, DL7076 (CN06) has been discovered as a dual activator of the CAR and nuclear factor erythroid 2-related factor 2 (Nrf2).³⁸

Recently, we have described 2-(3-methoxyphenyl)quinazoline derivatives modified at position 4 with 4-methoxy, 4-methylthio, or 4(1*H*)-thione moieties as potent but nonspecific human CAR ligands also activating PXR and vitamin D receptors.³⁹ Similarly, a human CAR agonist FL81 (5-(3,4-dimethoxybenzyl)-3-phenyl-4,5-dihydroisoxazole) discovered by another group also activates the PXR receptor to some extent.²³

Interestingly, in recent years, several inverse agonists of human CAR have been discovered with IC₅₀ in submicromolar/nanomolar concentrations such as PK11195 (1-(2-chlorophenyl)-*N*-methyl-*N*-(1-methylpropyl)-3-isoquinoline-carboxamide),⁴⁰ S07662 (1-[(2-methylbenzofuran-3-yl)methyl]-3-(thiophen-2-ylmethyl) urea),²¹ and CINPA1 ([5-

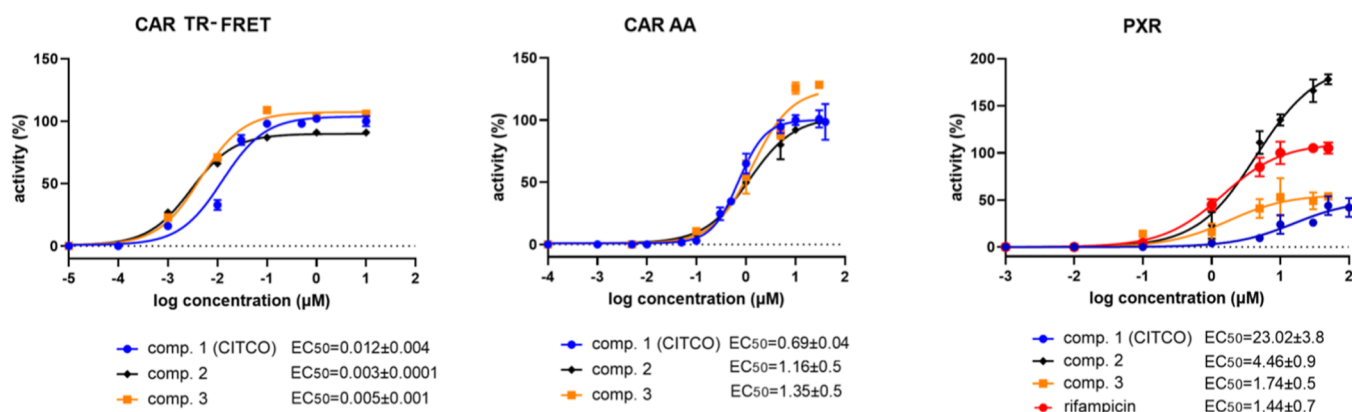
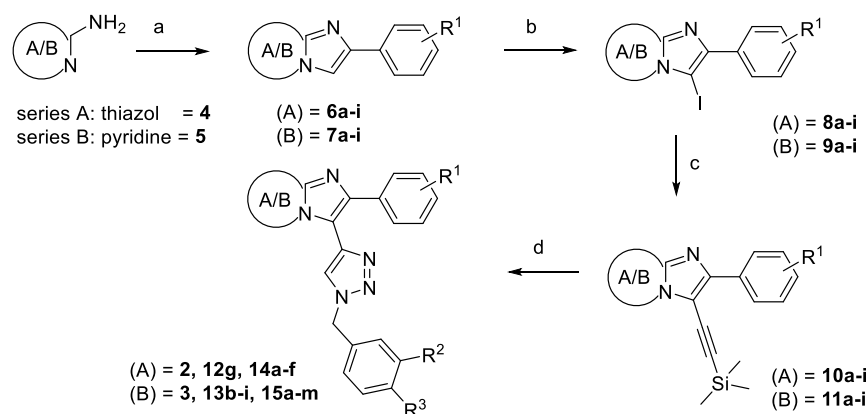


Figure 3. Lead compounds 2 and 3 significantly activate CAR and PXR in the TR-FRET LanthaScreen CAR coactivator assay (CAR TR-FRET), in the CAR LBD assembly assay (CAR AA), or in the PXR-responsive luciferase reporter assay. EC₅₀ (in μM) values were obtained based on sigmoidal dose–response fitting. Activities of CITCO and rifampicin, a PXR agonist, at 10 μM are set to be 100%.

Scheme 1. Preparation of the Lead Compounds and Analogues^a



^aReagents and conditions: (a) for example, 2-bromo-1-(4-chlorophenyl)ethan-1-one, NaHCO₃, EtOH, 70 °C, o.n.; (b) NIS, DCM, 25 °C; (c) TMS-acetylene, CuI, TEA, Pd(PPh₃)₂Cl₂, DMF, 0–25 °C; (d) 4-(azidomethyl)-1,2-dichlorobenzene, CuSO₄·5H₂O, KF, Na-ascorbate, THF/H₂O (1:1), 0–25 °C, 1 h.

[(diethylamino)acetyl]-10,11-dihydro-5*H*-dibenz[*b,f*]azepin-3-yl]carbamic acid ethyl ester).⁴¹

In the present work, we aimed to discover selective human CAR agonists that do not activate PXR or other nuclear receptors but still possess suitable ADME properties for further experiments in human hepatocyte cellular models or application in humanized CAR mouse models. In a library of kinase inhibitors, we found two lead structures (compounds 2 and 3) which appeared to be analogues to the known human CAR ligand CITCO.

RESULTS AND DISCUSSION

We initially synthesized two analogues 2 and 3 of a known yet unspecific CAR agonist, CITCO, by modifying the middle flexible oxime linker to the triazole ring (Figure 2). The oxime moiety is unstable under acidic conditions, which may complicate its use in vivo. The triazole ring offered stability, less flexibility, and good accessibility via an undemanding click reaction. Because of the synthetic feasibility and possibility to expand the variety of derivatives via a single reaction, the CuAAC reaction has been chosen. Different substitution patterns of blue, green, and red areas allowed us to explore the SAR of the compounds regarding the binding site, the bioavailability of the prepared compounds, and selectivity/

specificity toward the key receptors. Our decisions were also based on preliminary docking data (e.g., Figure S-3).

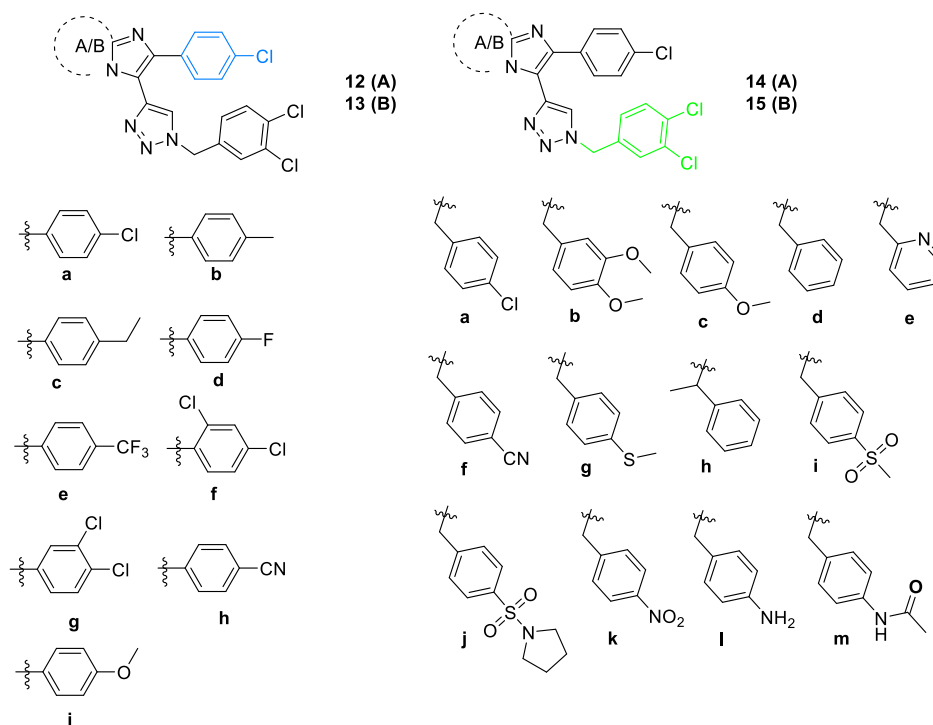
We found that analogues 2 and 3 significantly activate both CAR and PXR. Their activities and affinities toward human CAR were similar to those of CITCO in both the recombinant CAR LBD-dependent TR-FRET assay and cellular luciferase reporter assays. Their potency toward PXR was, however, more significant in comparison with the compound CITCO (Figure 3). Compound 2 displayed less cytotoxicity in COS-1 cells than compound 3 (Table S-1).

Compound 2 with the original imidazo[2,1-*b*]thiazole moiety of CITCO and compound 3 with imidazo[1,2-*a*]pyridine moiety were further modified in three key areas: blue (substituted phenyl ring), red (central heterocyclic linker), and green (substituted benzyl ring) (Figure 2).

Design and Synthesis of the First Generation of Novel CAR Ligands. The first modification of the phenyl and later benzyl ring led to two series A and B based on the compounds 2 and 3, respectively. The synthesis of the key compounds 2 and 3 as well as their modified analogues with a preserved triazole central heterocyclic linker is illustrated in Scheme 1.

The synthesis of both series started from 2-aminothiazole 4 (for series A) or 2-aminopyridine 5 (for series B). Cyclization with appropriate phenylacetyl chloride with various substitu-

Table 1. Modification of the Lead Compounds 2 and 3



comp.	yield ^a (%)	comp	yield ^a (%)	comp	yield ^a (%)	comp	yield ^a (%)
2	90	13f	74	14d	87	15f	82
12g	92	13g	90	14e	80	15g	80
3	57	13h	89	14f	87	15h	80
13b	84	13i	92	15a	92	15i	92
13c	70	14a	88	15b	81	15j	93
13d	74	14b	86	15c	75	15k	88
13e	86	14c	86	15d	78	15l	78
				15e	83	15m	92

^aYields in % correspond to the last cyclization step after purification.

tion patterns (Table 1) in EtOH at 70 °C led to 6-substituted imidazo[2,1-*b*]thiazoles 6a–i or 2-substituted imidazo[1,2-*a*]pyridines 7a–i. Iodination with NIS in DCM led to the iodinated intermediates 8a–i and 9a–i, respectively, in high to quantitative yields. A subsequent Sonogashira reaction with TMS-acetylene under Pd(PPh₃)₂Cl₂ catalysis resulted in compounds 10a–i and 11a–i. Lastly, a triazole ring was formed via a CuAAC click reaction, yielding final compounds 2, 12g, and 14a–f and 3, 13b–i, and 15a–m (Table 1).

Biology. The compounds with a substituted phenyl ring (12g–13i) appeared as potent agonists of the CAR with nanomolar EC₅₀ in a CAR TR-FRET assay. However, these compounds also significantly activated PXR, and most of them decreased the viability of COS-1 or HepG2 cells (Tables 2 and S-1).

Compounds 14a–14f were also found as potent agonists of both CAR and PXR. Similarly, compounds 15a–h displayed significant activation of CAR and PXR with some moderate effects on cellular viability. Among these compounds, 15d was found as a highly efficient CAR agonist, while at the same time, it significantly activated PXR with *E*_{max} higher than that of rifampicin (Figure 4 and Table 2). Showing an opposite result, compounds 15i–m have marginal activities on the CAR and weak activity toward PXR (Table 2).

Compound 15i appeared to be a selective CAR ligand. However, its activity in the TR-FRET CAR coactivator assay was negligible, and its activity in the CAR LBD assembly assay was weaker compared to CITCO (Figure 4). Thus, compound 15i demonstrates the phenomenon that some high-potency compounds in the TR-FRET assay with nanomolar EC₅₀ but less efficacy in cellular assays are, in fact, partial agonists of the CAR. These compounds do not reach the maximal activity (*E*_{max}) of full agonists such as CITCO or compound 15d (Figure 4). We should also consider the possibility that the tested compounds are likely distributed into cell membranes in cellular assays, which results in lower potency (higher EC₅₀ in CAR AA and CAR3 assays) in comparison with the in vitro TR-FRET assay.

In contrast, compound 15j with a sulfonyl pyrrolidine moiety does not possess any activity to the CAR. We suppose that it is too bulky to fit into the CAR LBD domain (Table 2). We can conclude that the substitution of the phenyl ring with a lipophilic moiety increased activities for both the CAR and PXR. Similarly, lipophilic substitution or no substitution on the benzyl ring increased the nonselective activation of CAR and PXR. Compound 15d was found as an efficient dual CAR/PXR agonist (Figure 4). Interestingly, compounds 15f and 15h displayed high potency for wtCAR in TR-FRET and CAR AA assays (with EC₅₀ in the nanomolar range), but they were less

Table 2. Effects of Compounds with Phenyl and Benzyl Ring Modification on the Activation of CAR and PXR^a

Comp.	CAR TR-FRET EC ₅₀ (μM)	CAR AA EC ₅₀ (μM)	CAR3 % CITCO activity ^b	PXR % RIF activity ^b
2	0.003 ± 0.0001	1.16 ± 0.5	78 ± 4	136 ± 8
3	0.005 ± 0.001	1.35 ± 0.5	167 ± 12	53 ± 4
12g	0.016	Nd	411 ± 13	57 ± 5
13b	0.0003	nd ^{tox.}	184 ± 21	47 ± 5
13c	0.0003	nd ^{tox.}	nd ^{tox.}	68 ± 7
13d	0.001	nd ^{tox.}	nd ^{tox.}	55 ± 4
13e	0.001	1.34 ± 0.2	nd ^{tox.}	42 ± 3
13f	0.062	Nd	166 ± 10	128 ± 8
13g	0.002	nd ^{tox.}	323 ± 27	43 ± 4
13h	nd [#]	nd ^{tox.}	nd ^{tox.}	nd ^{tox.}
13i	<0.001	nd ^{tox.}	nd ^{tox.}	89 ± 5
14a	0.007	Nd	143	86 ± 10
14b	0.432	0.15 ± 0.02	171	111 ± 7
14c	0.656	Nd	170	109 ± 10
14d	0.007	Nd	91 ± 9	215 ± 21
14e	0.01	Nd	109 ± 9	170 ± 13
14f	>5	Nd	15 ± 2	87 ± 8
15a	0.002	0.05 ± 0.01	78 ± 4	120 ± 10
15b	0.019	Nd	123 ± 7	89 ± 9
15c	0.040	0.12 ± 0.01	135 ± 11	117 ± 10
15d	0.001	0.46 ± 0.02	176 ± 12	195 ± 12
15e	1.38	Nd	134 ± 7	212 ± 21
15f	0.06	0.12 ± 0.01	33 ± 4	59 ± 8
15g	0.011	2.76 ± 0.2	62 ± 5	25 ± 2
15h	0.08	0.04 ± 0.007	44 ± 7	68 ± 7
15i	0.009	3.05 ± 0.8	18 ± 2	5 ± 0.2
15j	no activity	no activity	3 ± 0.3	16 ± 2
15k	>5	no activity	33 ± 4	21 ± 3
15l	>5	0.50 ± 0.01	73 ± 6	29 ± 4
15m	no activity	no activity	12 ± 0.4	20 ± 4
CITCO	0.012 ± 0.004	0.69 ± 0.04		
10 μM			396 ± 25	27 ± 5
1 μM			100%	8 ± 1
Rifampicin 10 μM				100%

^aTR-FRET LanthaScreen CAR coactivation assay (CAR TR-FRET), CAR LBD assembly assay (CAR AA), luciferase reporter assay with CAR3 variant, or PXR-responsive luciferase reporter assay were used. ^bCompounds were tested at 1 μM (for the CAR3 assay) or at 10 μM for the PXR assay (*n* = 3). nd—not determined due to significant cytotoxicity (nd^{tox.}), due to extensive PXR activation or low CAR activation, nd[#]—not determined due to solubility problem and potential precipitation in solution. EC₅₀ is the concentration required to achieve half-maximum activation in the TR-FRET Lantha Screen CAR Coactivation assay or the CAR AA assay (in μM).

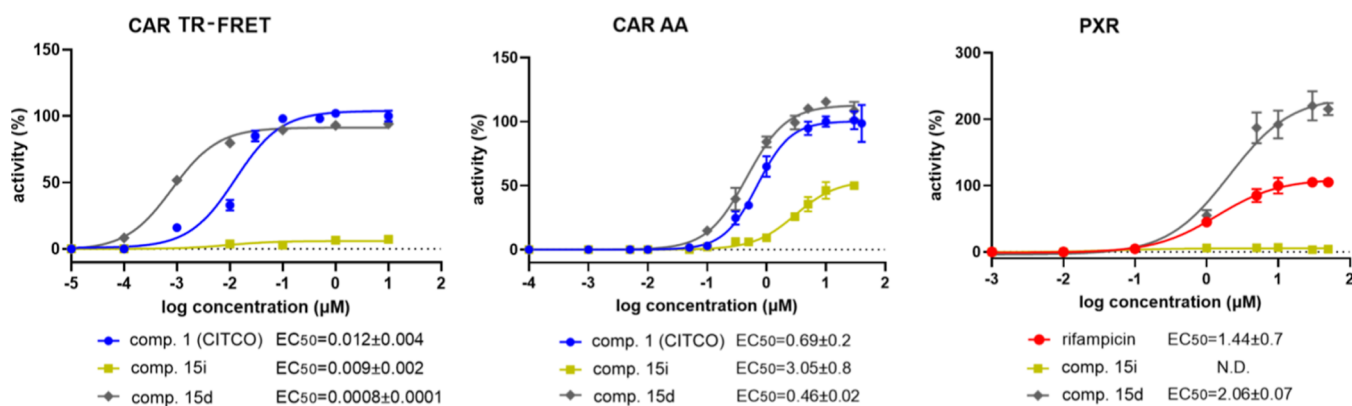
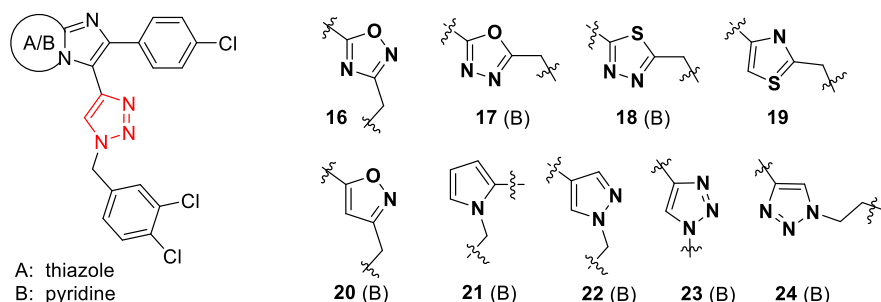


Figure 4. Activities of compounds 15i and 15d to stimulate the CAR in the TR-FRET LanthaScreen CAR Coactivator assay (CAR TR-FRET), in the CAR LBD assembly assay (CAR AA), or in the PXR-responsive luciferase reporter assay. EC₅₀ values (μM) were obtained based on sigmoidal dose–response curve fitting. Activities of CITCO and rifampicin at 10 μM are set to be 100%.

potent in the CAR3 assay, suggesting some selectivity for the wtCAR variant over the CAR3 variant (Table 2).

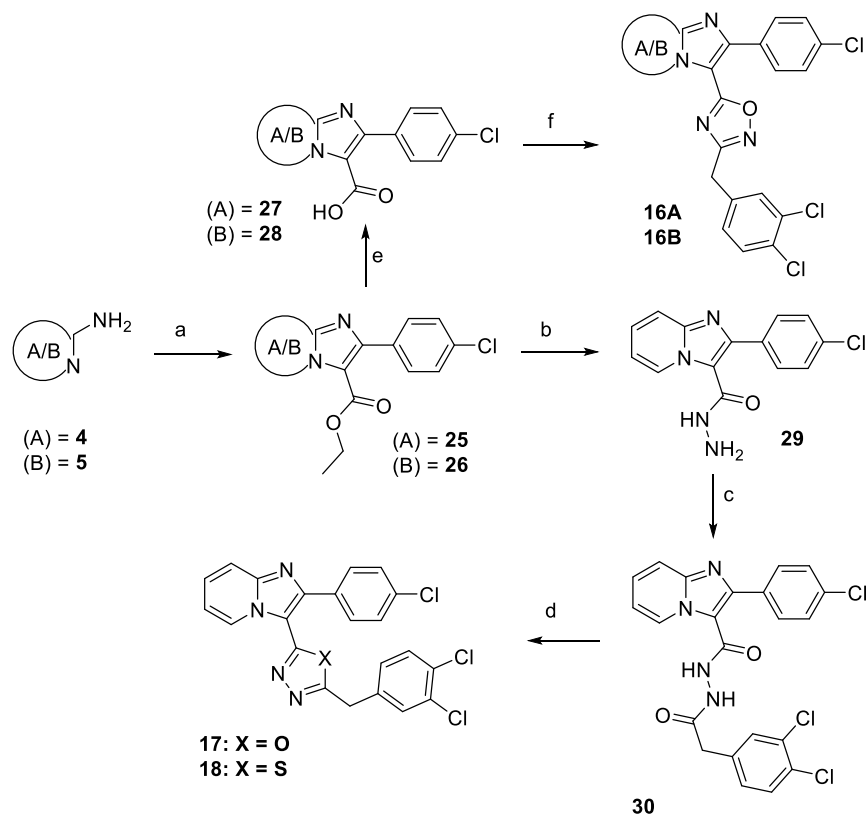
Design and Synthesis of the Second Generation of Novel CAR Ligands. For the next series of compounds, we

Table 3. Triazole Ring Modifications



comp.	yield ^a (%)	comp.	yield ^a (%)
16A	52	20	66
16B	50	21	21
17	6	22	32
18	21	23	73
19A	61	24	95
19B	62		

^aYields in % correspond to the last reaction step after purification.

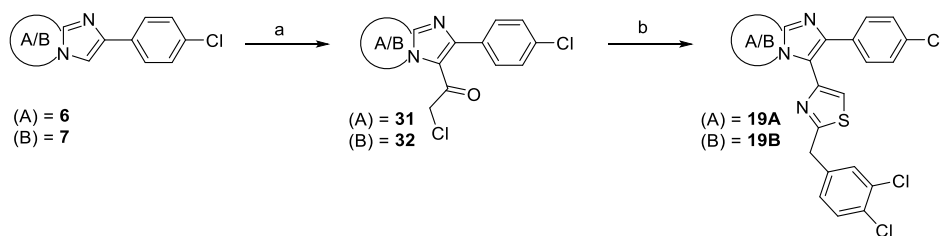
Scheme 2. Preparation of Novel Middle-Ring Heterocyclic Analogues^a

^aReagents and conditions: (a) ethyl 3-(4-chlorophenyl)-3-oxopropanoate, CBr_4 , CH_3CN , 80°C , o.n., 82%; (b) $\text{N}_2\text{H}_4\cdot\text{H}_2\text{O}$ (3 equiv), EtOH, reflux, o.n., 87%; (c) ethyl 2-(4-chlorophenyl)imidazo[1,2-*a*]pyridine-3-carboxylate, HATU, DIPEA, DMF, 25°C , o.n., 90%; (d) tosyl chloride (1.5 equiv), TEA (3 equiv), DCM, 0°C for 17, 6%; Lawesson's reagent (3 equiv), toluene, 100°C , o.n. for 18, 21%; (e) $\text{LiOH}\cdot\text{H}_2\text{O}$, THF/ H_2O 4:1, 25°C , 3 h, quant.; (f) (E)-2-(3,4-dichlorophenyl)-*N'*-hydroxyacetimidamide, EDC, HOBT, DMF, $25\text{--}80^\circ\text{C}$, o.n., 52% resp 50%.

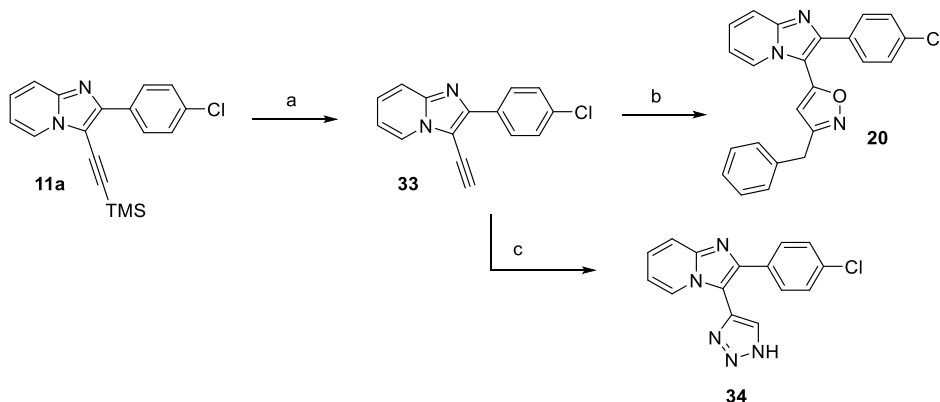
focused on the middle heterocyclic linker. The other rings (phenyl and benzyl) maintained the substitution pattern of compounds 2 and 3. The triazole ring was replaced by several heterocycles with one to three heteroatoms, such as thiadiazol or oxazole. The complete list is shown in Table 3.

Compounds 16A, 16B, 17, and 18 originated from the same precursors 25 or 26, which were synthesized by a condensation

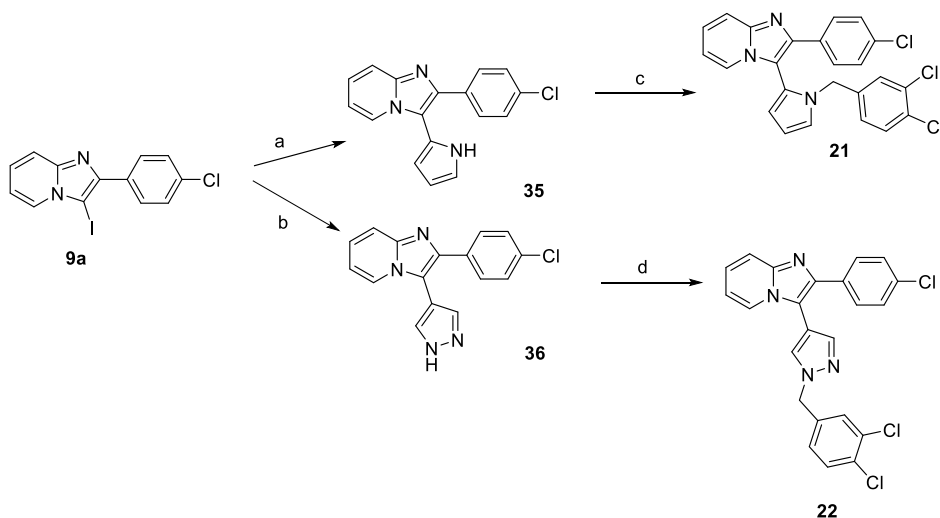
reaction of 4 or 5 with ethyl 3-(4-chlorophenyl)-3-oxopropanoate in the presence of CBr_4 . The ethyl ester moiety of precursors 25 and 26 was hydrolyzed using $\text{LiOH}\cdot\text{H}_2\text{O}$, and the obtained acid derivatives 27 and 28 were treated with EDC and HOBT at 25°C , followed by the addition of substituted *N'*-hydroxyacetimidamide at 80°C to yield compounds 16A and 16B (Scheme 2).

Scheme 3. Preparation of Novel Thiazole Analogues of the Lead Compounds^a

^aReagents and conditions: (a) chloroacetyl chloride (3 equiv), dioxane, 70 °C, 30 min, 100 °C, o.n., 91% resp 76%; (b) 2-(3,4-dichlorophenyl)ethanethioamide, EtOH, reflux, o.n., 61% resp 62%.

Scheme 4. Preparation of Oxazole Derivative 20 and Metabolite M3 (34)^a

^aReagents and conditions: (a) K₂CO₃, MeOH, 25 °C, 2 h, 90%; (b) phenylacetaldehyde, NH₂OH·HCl, NaOH, H₂O/*t*-BuOH, CuI, chloramine T, 25 °C, o.n., 66%; (c) TMSN₃, CuI, DMF/MeOH 10:1, 70 °C, 84%.

Scheme 5. Synthesis of Pyrrole and Pyrazole Derivatives 21 and 22^a

^aReagents and conditions: (a) (1-(*tert*-butoxycarbonyl)-1*H*-pyrrol-3-yl)boronic acid, Pd(PPh₃)₄, Na₂CO₃, dioxane/H₂O, 90 °C, o.n., 35%; (b) (4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-1*H*-pyrazole, Pd(dppf)Cl₂, Na₂CO₃, dioxane/H₂O, 90 °C, o.n., 40%; (c) 1,2-dichloro-4-(chloromethyl)benzene, NaH, DMF, 25 °C, o.n., 21%; (d) 1,2-dichloro-4-(chloromethyl)benzene, CH₃CN, K₂CO₃, 25 °C, o.n., 32%.

In order to synthesize compounds 17 and 18, the ester derivative 26 was reacted with an excess of hydrazine hydrate in EtOH, providing compound 29 and further acylated with 2-(3,4-dichlorophenyl)acetic acid by means of the peptide coupling reagent HATU in DMF. Ring-closing reaction of 30 with tosyl chloride at 25 °C or Lawesson's reagent at 100 °C overnight led to final compounds 17 and 18 respectively, although at very low yields (Scheme 2).

Thiazole analogues 19A and 19B were prepared from intermediates 6a and 7a, which were reacted with an excess of chloroacetyl chloride in dry dioxane at 70 °C for 30 min and then heated up to 100 °C overnight, followed by cyclization with ethanethioamide (1.5 equiv) in EtOH at reflux (Scheme 3).

Sonogashira reaction of 9a with TMS acetylene under Pd(PPh₃)₂Cl₂ catalysis provided intermediate 11a, following

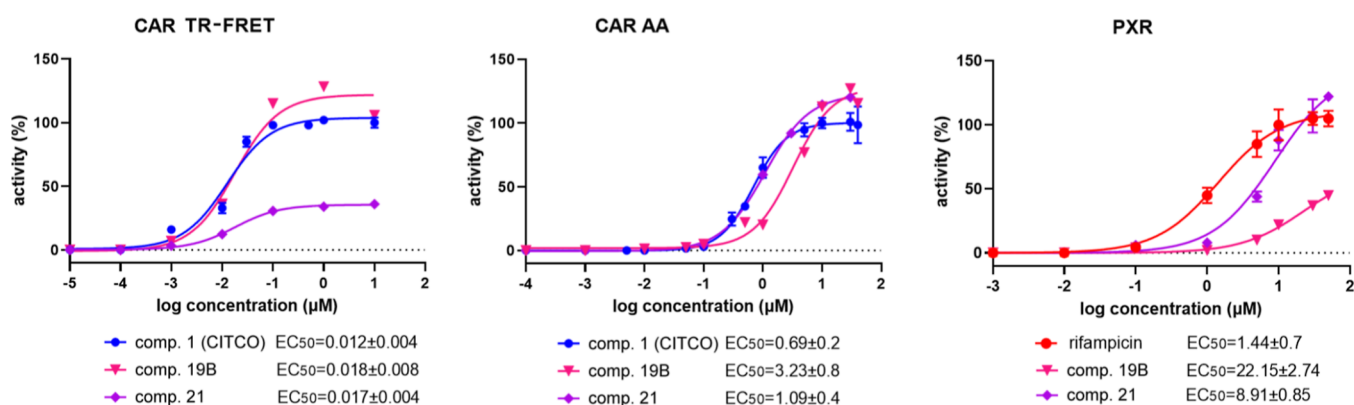


Figure 5. Relative activities of compounds 19B and 21 to activate CAR in TR-FRET CAR coactivator assays (CAR TR-FRET), in the CAR LBD assembly assay (CAR AA), or the PXR-responsive luciferase reporter assay. EC_{50} values (μM) were obtained based on sigmoidal dose–response curve fitting. Activities of CITCO and rifampicin at 10 μM are set to be 100%.

deprotection of the TMS group with K_2CO_3 in MeOH yielded compound 33. Pretreatment of phenylacetaldehyde with hydroxylamine hydrochloride and following reaction with intermediate 33 in the presence of chloramine T and CuI at 25 $^\circ\text{C}$ provided compound 20. The click reaction of compound 33 with TMSN_3 and CuI in a DMF/MeOH mixture under an inert atmosphere led to the unsubstituted triazole derivative 34 (Scheme 4).

Pyrrole and pyrazole derivatives (21 and 22) were obtained in two-step synthesis starting from iodinated precursor 9a, which was coupled in a Suzuki reaction under $\text{Pd}(\text{PPh}_3)_4$ catalysis with (1-(*tert*-butoxycarbonyl)-1*H*-pyrrol-3-yl)boronic acid or 4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-1*H*-pyrazole with $\text{Pd}(\text{dppf})\text{Cl}_2$ as a catalyst, respectively, followed by a substitution reaction with benzyl chloride under basic reaction conditions (Scheme 5). Compounds 23 and 24 were synthesized from intermediate 11a via a click reaction with 4-azido-1,2-dichlorobenzene or 4-(2-azidoethyl)-1,2-dichlorobenzene according to Scheme 1.

Biology. When we tested the compounds with the modified middle heterocyclic linker, we found that the central moiety also contributes to both CAR and PXR activation, although there were no dramatic variations in the effects of different heterocycles. Out of the series of compounds, compound 19B showed higher relative selectivity to the CAR as it significantly activates CAR and CAR3, but it tends to activate PXR from 5 μM (Figure 5). Similarly, compound 21 has high activity toward the CAR in the CAR LBD assembly assay, but it has a significant activity to PXR at a 1 μM concentration. Interestingly, in the CAR TR-FRET assay, the compound seems to be a partial agonist with the E_{max} lower than that of CITCO or compound 19B (Table 4, Figure 5).

No significant cytotoxicity was observed for these compounds in COS-1 and HepG2 cells (Table S-2).

Design and Synthesis of the Third Generation of Novel CAR Ligands. Since the most promising biological results were found with compound 3 derivatives, in the next step, we addressed the modification of the benzyl ring with the main emphasis on the *meta* position while maintaining the triazole ring. The chlorine atom was replaced with a series of acyl molecules (Table 5). This resulted in improved estimated water solubility and bioavailability of the compounds.

The key compound 37 was prepared by the same means as compound 3 (Scheme 1) with a minor modification in the azide coupling partner. Straightforward hydrolysis led to acid

Table 4. Effects of Middle-Ring Heterocyclic Analogues on the Activation of the CAR and PXR^a

comp.	CAR TR-FRET EC_{50} (μM)	CAR AA EC_{50} (μM)	CAR3 % CITCO activity ^b	PXR % RIF activity ^b
16A	0.014	>5	22 \pm 0.4	6 \pm 0.7
16B	0.035	2.61 \pm 0.4	33 \pm 4	7 \pm 0.5
17	0.023	nd	144 \pm 11	56 \pm 4
18	0.015	8.84 \pm 1.15	93 \pm 10	20 \pm 1
19A	0.013	1.16 \pm 0.09	231 \pm 14	122 \pm 10
19B	0.018 \pm 0.08	3.23 \pm 0.8	138 \pm 12	22 \pm 3
20	0.01	0.1 \pm 0.04	263 \pm 27	130 \pm 13
21	0.017 \pm 0.04	1.09 \pm 0.04	29 \pm 0.4	88 \pm 10
22	0.0005	nd ^{tox.}	nd ^{tox.}	124 \pm 13
23	>10	nd	12 \pm 2	5 \pm 0.4
24	0.04	2.8 \pm 0.71	82 \pm 7	141 \pm 10
CITCO	0.012 \pm 0.004	0.69 \pm 0.2		
10 μM			396 \pm 25	27 \pm 5
1 μM			100%	8 \pm 1
Rifampicin				100%
10 μM				

^aTR-FRET LanthaScreen CAR Coactivation assay (CAR TR-FRET), CAR LBD assembly assay (CAR AA), luciferase reporter assay with CAR3 variant, or PXR-responsive luciferase gene assay were used.

^bCompounds were tested at 1 μM (for CAR3 assay) or at 10 μM for PXR assays ($n = 3$). nd—not determined due to significant cytotoxicity (nd^{tox.}), due to extensive PXR activation, or low CAR activation (nd). EC_{50} is the concentration required to achieve half-maximum activation in the TR-FRET LanthaScreen CAR coactivation assay or CAR AA assay (in μM).

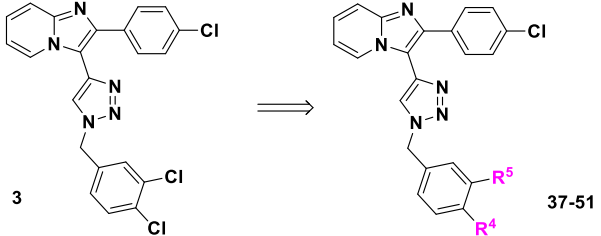
analogue 38, which was converted to amide analogues 39–42 via acylchloride 52 (Scheme 6).
















In order to increase bioavailability, compound 39 was converted to its HCl salt 39 HCl (Scheme 6). The subsequent reaction of *N*-methoxy-*N*-methylbenzamide derivative 42 with a Grignard reagent or LAH at low temperature yielded ketone (43,44) or aldehyde analogues 45, which upon the reaction with hydroxylamine provided derivative 46. Finally, the reduction of the ester derivative with subsequent methylation provided compounds 47 and 48 (Scheme 6).

N-acyl derivative 51 was synthesized in two steps from nitro derivative 49 after reduction and the succeeding acylation reaction (Scheme 7).

Next, we decided to broaden the number of examples of heterocyclic linkers with six-membered heterocycle pyridine

Table 5. Modification of the Benzyl Ring of the Lead Compound 3



Comp.	R4	R5	Yield ^a (%)	Comp.	R4	R5	Yield ^a (%)
37	Cl		84 %	45	Cl		75 %
38	Cl		92 %	46	Cl		90 %
39	Cl		88 %	47	Cl		57 %
40	Cl		89 %	48	Cl		40 %
41	Cl		95 %	49	Cl		83 %
42	Cl		91 %	50	Cl		79 %
43	Cl		72 %	51	Cl		90 %
44	Cl		69 %				

^aYields in % correspond to the last reaction step after purification.

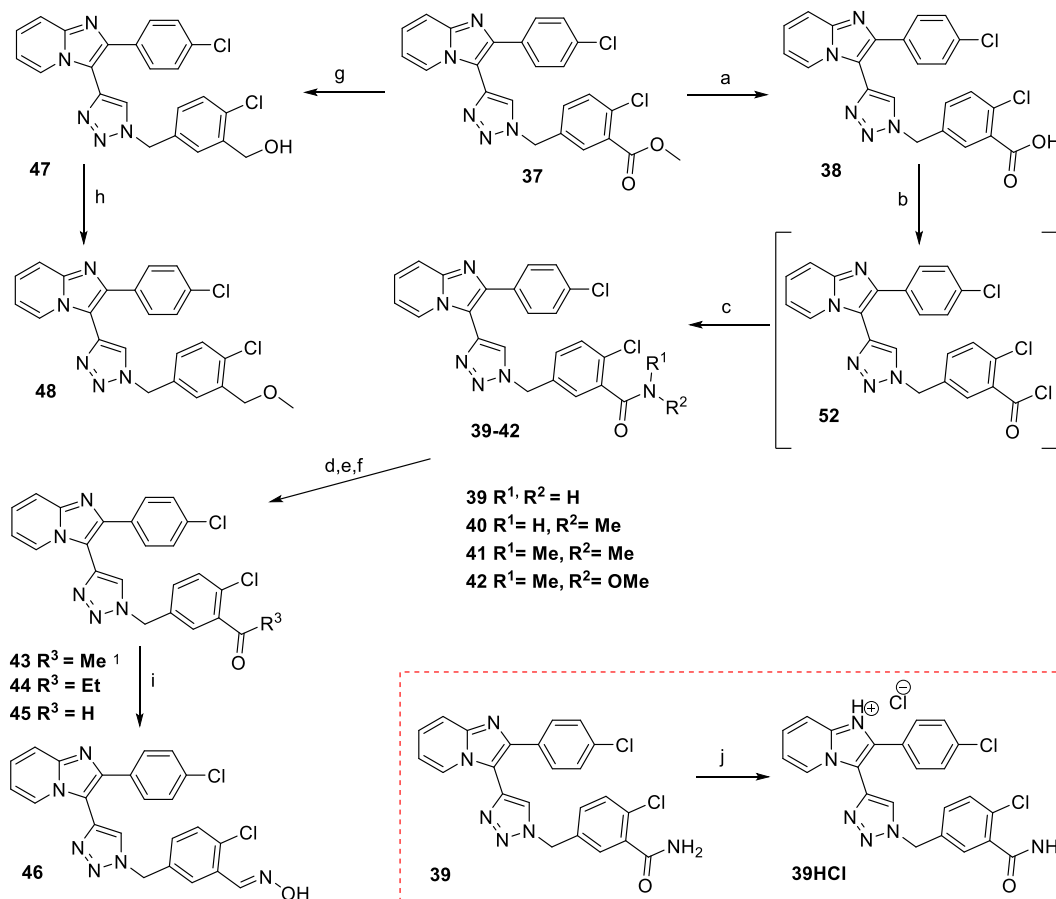
and to replace the original triazole with an aryl ring. Syntheses of these compounds started from iodinated precursor **9a**, which was coupled with aryl/pyridyl boronic acid (Scheme 8), followed by the Negishi coupling reaction with benzylzinc bromide catalyzed by Pd, providing compounds **55** and **56**, respectively. Unfortunately, these compounds were barely soluble, so we did not test them.

Moreover, the methylene part of the linker was exchanged for O and NH. Similarly, to the previously mentioned compounds, intermediate **9a** was coupled with appropriate boronic acid (Scheme 9), providing intermediate compounds **57** and **58** with a free amino group and methoxy group, respectively. Demethylation of compound **58** afforded compound **59** with a free hydroxy group. Both compounds **57** and **59** were coupled with 5-bromo-2-chlorobenzamide by way of Buchwald or Ullmann coupling conditions, yielding compounds **60** and **61** (Scheme 9, Table 6).

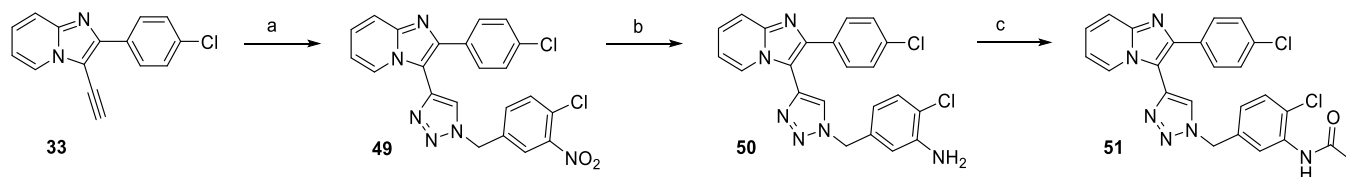
Biology. When we tested derivatives of compound **3** (Series B) with acyl moieties in the meta position of the benzyl ring with the preserved triazole ring as a linker (**37–44**), we found that the moieties significantly contribute to CAR activation but not to PXR activation. Carboxylic acid itself in the meta position (**38**) resulted in a complete loss of CAR activation. However, amides, as well as esters, significantly

activated CAR in all assays. Compound **37** appeared as the most efficient to activate the CAR in the TR-FRET LanthaScreen CAR coactivation assay and highly efficient to activate a CAR LBD assembly assay with an EC₅₀ lower than that of CITCO (EC₅₀ = 0.4 and 152 nM vs 12 and 690 nM, respectively). Importantly, a methylester (compound **37**), amides (**39**, **40**, and **41**) as well as *N*-methoxy-*N*-methylamide (compound **42**) all have minimal (**37** and **41**) or no activity (**39**, **40**, and **42**) to activate PXR at 10 μM (Figure 6 and Table 7). Other compounds from the set also display low activation of PXR. These compounds were also noncytotoxic in viability assays (Table S-3).

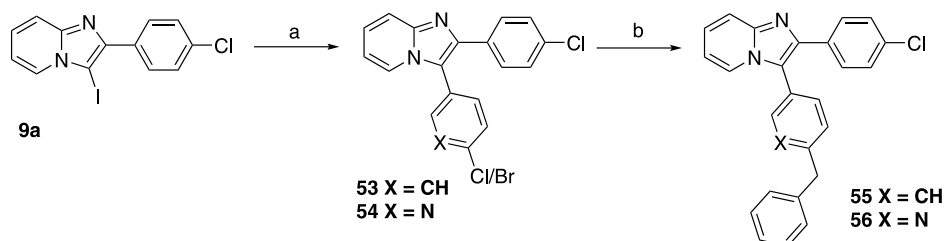
When we looked in detail at the CAR agonists **39**, **40**, **41**, and **42** without significant PXR activation, their potencies in the TR-FRET LanthaScreen CAR coactivation assay were by an order of magnitude lower (and EC₅₀ higher) than that of CITCO and compound **39** seems to be a partial agonist of the CAR in the assay. In the case of CAR LBD assembly and CAR3 variant assays, these compounds activated the CAR LBD with lower but still comparable affinities in comparison with CITCO. This phenomenon may be explained by the different activation of the CAR LBD by these compounds via another coactivator than with PGC1α, which is involved in the TR-FRET CAR coactivation assay. Indeed, SRC-1 (NCOA1)

Scheme 6. Synthesis of Lead Compound 3 Analogues with Benzyl Ring Modification^a

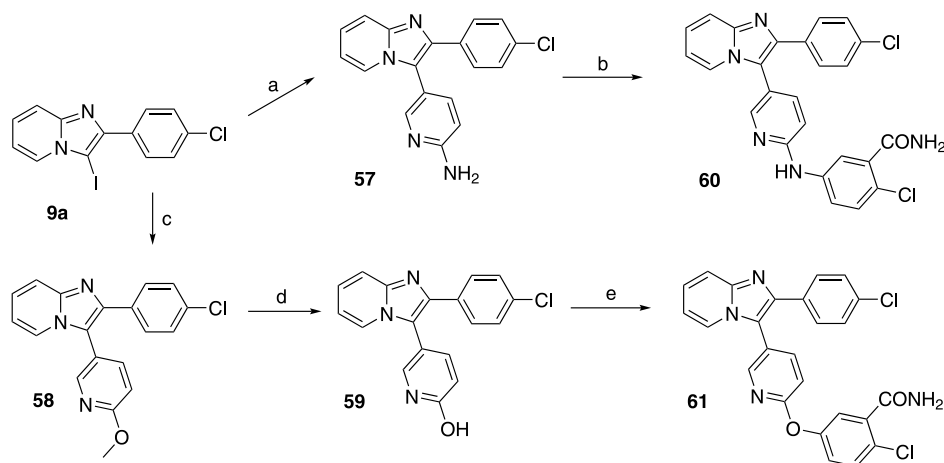
^aReagents and conditions: (a) LiOH·H₂O, THF/H₂O, 25 °C, 2 h, 92%; (b) SOCl₂, toluene; (c) NHR¹R², DIPEA, 88–95%; (d,e) MeMgBr or EtMgBr, THF, 0–25 °C, 72% resp. 69%; (f) LAH (1 equiv), THF, 5 °C, 2 h, 75%; (g) K₂CO₃, MeOH, 25 °C, 2 h; (h) CH₃I, K₂CO₃, DMF, 65%; (i) 45, NH₂OH·HCl, DCM, 2 h, 90%; (j) HCl/diethylether, THF, 20 min, 0 °C, quant.

Scheme 7. Synthetic Pathway of Nitro, Amino, and Acetylamino Derivates^a

^aReagents and conditions: (a) 4-(azidomethyl)-1-chloro-2-nitrobenzene, CuSO₄·5H₂O, KF, Na-ascorbate, THF/H₂O, 25 °C, 1 h, 83% (b) AcOH, Fe, MeOH, reflux, 79%; (c) Ac₂O, pyridine, dioxane, 25 °C, o.n., 90%.

Scheme 8. Preparation of Six-Membered Heterocyclic Analogues^a

^aReagents and conditions: (a) (4-bromophenyl)boronic acid or (6-chloropyridin-3-yl)boronic acid, dioxane/H₂O mixture (4:1), Na₂CO₃, Pd(dppf)Cl₂·DCM, 95 °C, o.n., 59% resp. 63%; (b) Pd₂dba₃, XantPhos, benzylzinc bromide solution 0.5 M in THF, THF, 60 °C, o.n., 72% resp. 75%.

Scheme 9. Synthesis of Pyridine Derivatives with a N/O Linker^a

^aReagents and conditions: (a) 2-aminopyridine-5-boronic acid pinacol ester, dioxane, Na₂CO₃ in 2 mL of H₂O, Pd(dppf)Cl₂, 90 °C, o.n., 53%; (b) 5-bromo-2-chlorobenzamide, dioxane, Na₂CO₃, XantPhos, Pd₂dba₃, 100 °C, o.n., 51%; (c) (6-methoxy-3-pyridinyl)boronic acid, dioxane, Na₂CO₃, H₂O, Pd(dppf)Cl₂, 90 °C, o.n., 53%; (d) 4 M HCl/dioxane, 95 °C, o.n., 73%; (e) 5-bromo-2-chlorobenzamide, BPPO, CuI, K₃PO₄, DMF, 110 °C, o.n., 6%.

Table 6. Modification of the Methylene Linker

comp.	R	yield ^a (%)
60	N bridge	51
61	O bridge	6

^aYields in % correspond to the last reaction step after purification.

along with other coactivators are important for the coactivation of CAR⁸ and we can suppose an array of different coactivators in CAR variants activation in cellular assays. We may also suppose the intracellular accumulation of these compounds, for example, via an uptake mechanism, which may increase their potencies in cellular CAR LBD assembly and CAR3 variant assays, but not in the TR-FRET LanthaScreen CAR coactivation assay.

Interestingly, compounds 39, 40, and 42 moderately deactivated the PXR-responsive construct in concentrations higher than 10 μM.

Compounds with substitution of the meta position of the benzyl ring with other substituents (43–51) and with the

preserved triazole ring as a linker retained efficient CAR activation with high potency (EC₅₀ below 0.1 μM) (Table 7), although these compounds also activate PXR to some degree. Some of these display substantial effects on cellular viability (Table S-3), which may affect cellular assays. With a methoxyethyl moiety, compound 48 was found to activate the CAR LBD assembly assay with the lowest EC₅₀ = 24.9 nM; however, the compound is not selective for the CAR and significantly activates PXR (EC₅₀ = 4.34 ± 1 μM). Compound 48 was also highly potent in the activation of the CAR3 variant in the CAR3 variant assay (Table 7). Interestingly, some compounds such as 45 and 47 had high potency for wtCAR in the TR-FRET and CAR AA assays, but they were less potent in the CAR3 assay, suggesting some selectivity for the wtCAR variant.

Compounds 60–61 with a replaced methylene part of the linker with O and NH (Table 6) lost the activity to the CAR and retained a weak activity to PXR.

Characterization of Induction Properties of Selected Candidates in Human Hepatocyte Models and Their

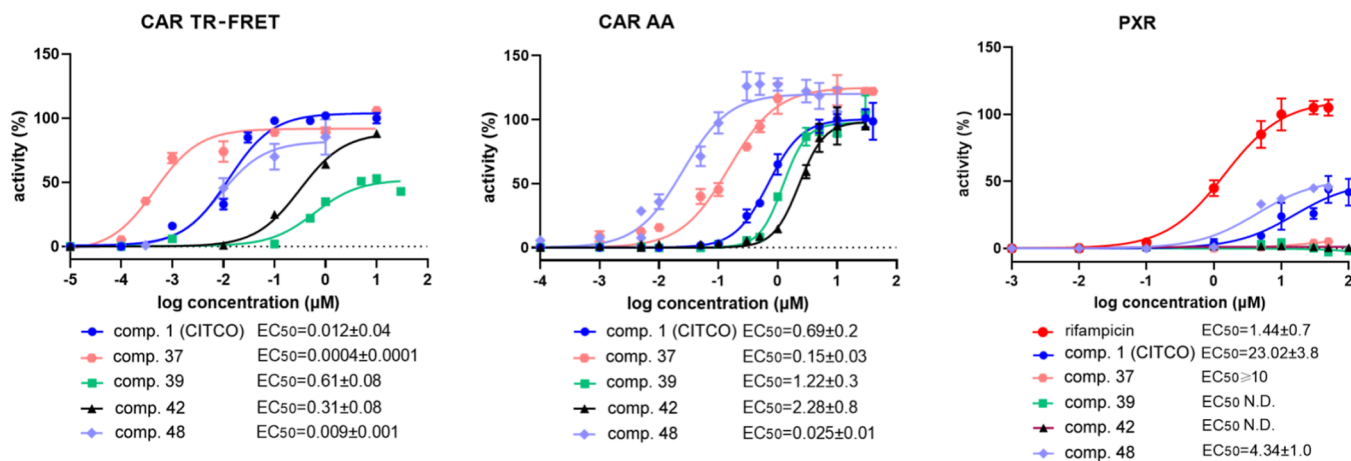


Figure 6. Activation of CAR and PXR in TR-FRET LanthaScreen CAR coactivation assay (CAR TR-FRET), in the CAR LBD assembly assay (CAR AA), or the PXR-responsive luciferase assay. EC₅₀ values (μM) were obtained based on sigmoidal dose–response curve fitting. Activities of CITCO and rifampicin at 10 μM are set to be 100%.

Table 7. Activation of CAR and PXR with Benzyl Ring Modification Analogues of Compound 3 in the TR-FRET LanthaScreen CAR Coactivation Assay (CAR TR-FRET), in the Reporter Assay with the CAR3 Variant, in the CAR LBD Assembly Assay (CAR AA), or the PXR-Responsive Luciferase Gene Assay

comp.	CAR TR-FRET EC ₅₀ (μM)	CAR AA EC ₅₀ (μM)	CAR3 % CITCO activity ^a	PXR % RIF activity ^a
37	0.0004 ± 0.0001	0.148 ± 0.03	137 ± 8	4 ± 0.6
38	no activity	nd	14 ± 1	2 ± 3
39	0.611 ± 0.08	1.22 ± 0.3	68 ± 4	no activation
40	1.686	1.527 ± 0.3	63 ± 4	no activation
41	1.066	1.086 ± 0.1	91 ± 8	4 ± 3
42	0.313 ± 0.08	2.28 ± 0.8	44 ± 2	no activation
43	0.001	nd ^{tox}	333 ± 30	30 ± 2
44	0.011	0.84 ± 0.04	224 ± 21	26 ± 1
45	0.008	<0.1	69 ± 7	43 ± 4
46	0.007	0.160 ± 0.01	82 ± 6	29 ± 3
47	0.001	nd ^{tox}	71 ± 7	77 ± 7
48	0.009 ± 0.001	0.025 ± 0.01	345 ± 32	37 ± 4
49	0.004	0.060 ± 0.008	59 ± 7	35 ± 4
50	0.006	nd ^{tox}	70 ± 8	43 ± 4
51	0.134	nd ^{tox}	97 ± 7	71 ± 6
60	no activity	nd ^{tox}	8 ± 0.7	23 ± 2
61	no activity	no activity	2 ± 0.3	12 ± 1
CITCO	0.012±	0.69 ± 0.2		
10 μM			396 ± 25	27 ± 5
1 μM			100%	8 ± 1
Rifampicin 10 μM				100%

^aCompounds were tested at 1 μM (for CAR3 assay) or at 10 μM for PXR assays (*n* = 3). nd—not determined due to significant cytotoxicity (nd^{tox}), due to extensive PXR activation, or low CAR activation (nd). EC₅₀ is the concentration required to achieve half-maximum activation in the TR-FRET LanthaScreen CAR coactivation assay or CAR LBD assay (in μM).

Interactions with Human CAR Variants. Next, we decided to analyze our novel selective CAR agonists **37**, **39**, **40**, **41**, and **42** to determine whether they could upregulate *CYP2B6* gene mRNA, the typical CAR target gene, in PHH from one donor. We found that all compounds could significantly upregulate *CYP2B6* mRNA. Compounds **39** and **42** tend to be the most potent with a 1 μM concentration in the experiments. These data suggest that the compounds are metabolically stable in metabolically competent hepatocyte cells and that they enter hepatocytes to activate the CAR (Figure 7A).

In translocation experiments with the EGFP-hCAR + Ala chimera, we examined the tested compounds to determine whether they stimulate cytoplasm-to-nuclear translocation of the activated human CAR with extra alanine in the LBD (CAR + A).⁴² We noted that mainly CITCO and compounds **37**, **39**, and **40** significantly decrease the number of cells with specific cytoplasm EGFP-hCAR + Ala localization, and they increase the portion of cells with nuclear localization of the CAR chimera (Figures 7B and S-1). In these experiments, compound **39** appeared as the most promising candidate.

In agreement with cellular assays or induction experiments in PHHs, tested compounds have similar activities in comparison with CITCO (comp. **1**) or compound **37**, which are high-affinity CAR agonists in TR-FRET CAR assays. These data suggest that the cellular environment and signaling have a significant determination on CAR activation.

In the next experiments, we sought to determine whether the discovered selective CAR agonists **37**, **39**, **40**, **41**, and **42** interact with wild-type CAR (wtCAR), human CAR variants 2 (CAR2) and 3 (CAR3), as well as with mouse CAR orthologue in luciferase reporter assays. Efficacy to activate wtCAR was assessed using the CAR LBD assembly assay (CAR AA) or with a wtCAR expression vector that was

inhibited with PK111195 (0.1 μM), a known CAR inhibitor. We found that compound **37** is highly efficient in the stimulation of the variant CAR2 and other variants of CAR in comparison with CITCO (100% activity). Compound **39** significantly activated wtCAR in the CAR LBD assembly assay and the CAR3 variant in the gene reporter assay. Its activity in the assay with wtCAR and its inhibitor PK111195, however, was low, suggesting a weak efficacy to compete with the PK111195 inhibitor in the CAR LBD. Other candidate compounds have lower potency in comparison to CITCO (100% activity) in the activation of CAR variants. Compound **42** appeared as a combined agonist of wtCAR and its variants in all assays. Only compound **37** was found to stimulate the mouse CAR when compared to the mouse ligand TCPOBOP (Figure 7C).

In the follow-up studies, we examined the stability of compound **37** in human and mouse microsomes and in plasma. We found that compound **37** is unstable in both mouse and human microsomes with *t*_{1/2} = 4.78 ± 1.31 min and *t*_{1/2} = 6.38 ± 0.67 min, respectively. Importantly, we found that compound **37** is also unstable in mouse plasma as well with *t*_{1/2} = 22.76 ± 0.03 min (Figure S-2).

Selection of the Candidate for Animal Studies and Detailed Characterization of Compound 39. In the next experiments, we studied the most efficient compound **39** in five PHHs from five different donors to determine whether it could upregulate *CYP2B6*, *CYP3A4*, and *CYP2C9* mRNA. These genes are significantly, but not exclusively, regulated via the CAR in human hepatocytes. Despite high variability in response in different hepatocyte preparations, we found that compound **39** has similar activity to induce these genes in comparison with CITCO (Figure 7D). Western blotting experiments in PHHs (BioIVT) treated with comp. **39**, rifampicin, CITCO, and PXR antagonist SPA70 (10 μM) for

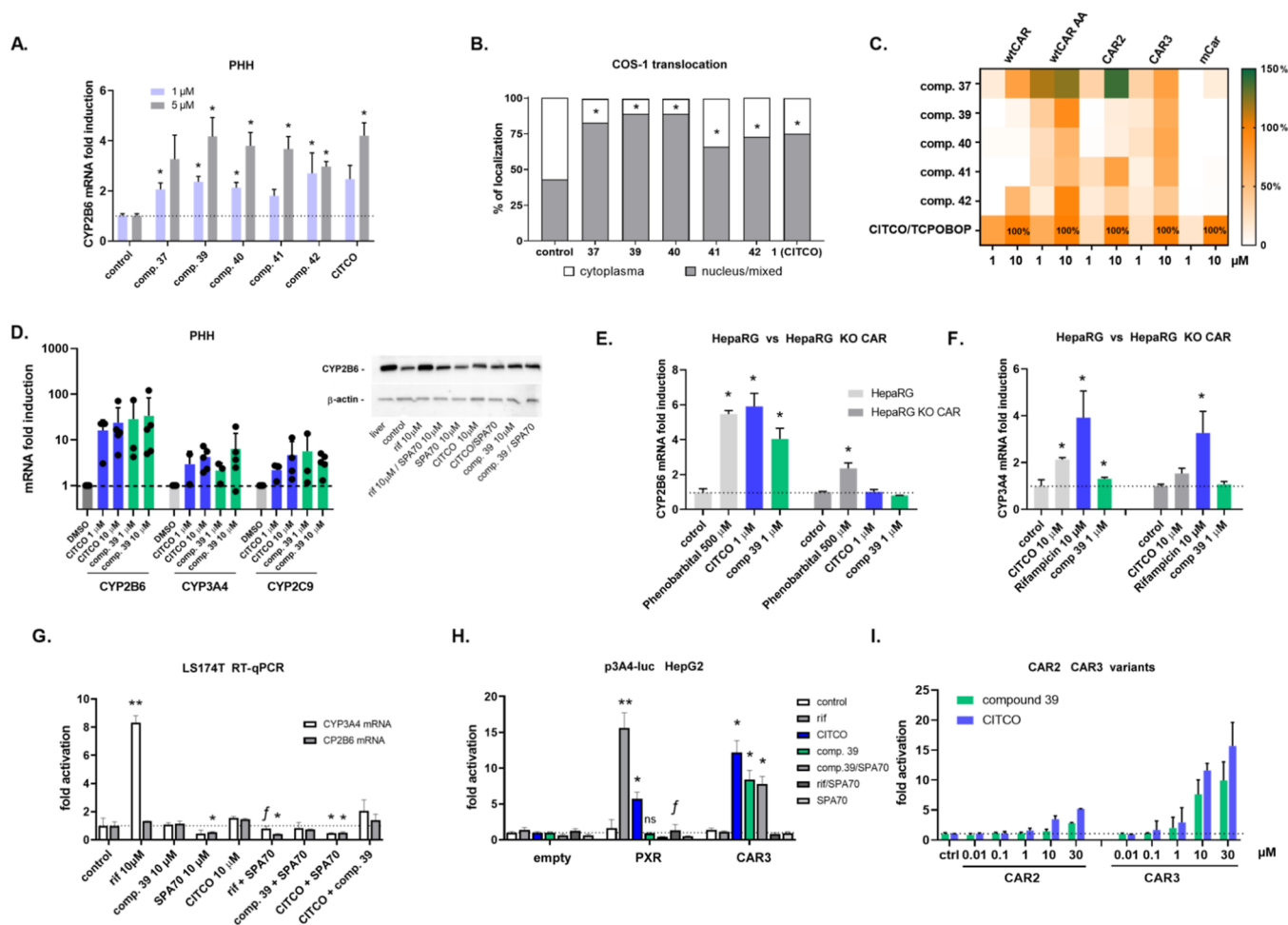


Figure 7. Induction of CAR target genes in primary human hepatocyte models and interactions of selected candidates with CAR transcription variants and mouse CAR. (A) PHHs were treated with compounds 37, 39, 40, 41, and 42 together with CITCO for 24 h. The expression of *CYP2B6* mRNA, a prototype CAR target gene, was analyzed using RT-qPCR in technical triplicates. Data are presented as mRNA fold induction to control (vehicle-treated) samples. (B) Translocation experiments with EGFP-hCAR + Ala chimera in COS-1 cells treated with tested compounds (10 μ M) for 24 h before confocal microscopy. Data are presented as % of cells with specific cytoplasm or mixed/nuclear localization of pEGFP-hCAR + Ala chimeric protein. (C) Interactions of compounds 37, 39, 40, 41, and 42 with human wtCAR in the CAR LBD assembly assay (wtCAR AA) or with wtCAR inhibited with PK11195 (0.1 μ M), with CAR2 or CAR3 variants, or with mouse Car (mCar). (D) PHHs from five donors were treated with compound 39 and CITCO (1 and 10 μ M, respectively) for 48 h. CAR target genes *CYP2B6*, *CYP3A4*, and *CYP2C9* mRNA expression have been studied using RT-qPCR. Data are presented as fold induction to control (vehicle-treated) samples. Western blotting experiments with primary human (BioIVT) treated with comp. 39, rifampicin (rif), CITCO, and PXR antagonist SPA70 (10 μ M) for 48 h. Monoclonal anti-CYP2B6 antibody (PA5-35032) was used to detect CYP2B6 protein. (E,F) HepaRG cells and HepaRG KO CAR cells without functional CAR activity were treated with phenobarbital (500 μ M), CITCO, rifampicin (10 μ M), or compound 39 at a 1 μ M concentration for 48 h. *CYP2B6* and *CYP3A4* mRNA expression have been analyzed using RT-qPCR. (G) LS174T cells expressing PXR, but without functional CAR, were treated with compound 39, rifampicin (rif), CITCO, SPA70, or compound 39 at a 10 μ M concentration for 48 h. *CYP2B6* and *CYP3A4* mRNA expression has been analyzed using RT-qPCR. (H) Luciferase gene reporter assay with the *CYP3A4* gene promoter construct (p3A4-luc) in HepG2 cells transfected with either PXR or CAR3 expression constructs. Cells were treated for 24 h before analysis. (I) Dose–response activation of CAR2 and CAR3 variants with compound 39 in luciferase reporter gene assays. * p < 0.05 and ** p < 0.01-significant *CYP2B6* or *CYP3A4* mRNA upregulation, p3A4-luc activation or EGFP-hCAR + Ala fusion protein nuclear translocation to control samples; f -statistically significant effect of SPA70 on rifampicin-mediated *CYP3A4* mRNA expression or activation of the p3A4-luc luciferase construct.

48 h revealed that compound 39 up-regulates *CYP2B6* protein and that the upregulation is not abolished by the PXR antagonist SPA70 (Figure 7D, inserted panel).

To confirm that compound 39 induces *CYP2B6* mRNA via the activated CAR, we performed experiments with HepaRG and its KO CAR counterpart cell line without CAR expression. We observed the upregulation of *CYP2B6* and *CYP3A4* mRNA only in the HepaRG cells but not in the HepaRG KO CAR cells after treatment with both CITCO and compound 39 (Figure 7E,F).

In the next experiments, *CYP2B6* and *CYP3A4* mRNA expression were analyzed in LS174T cells using RT-qPCR. LS174T cells express endogenous PXR but lack functional CAR.⁴³ We did not observe any significant induction of these genes by compound 39 in these cells (Figure 7G).

Then, we performed luciferase gene reporter assays with the *CYP3A4* gene promoter construct (p3A4-luc) in HepG2 cells transfected with either PXR or CAR3 expression constructs. Compound 39 activated the luciferase construct only in the presence of CAR3, and the PXR antagonist SPA70 had no significant effect on the activation (Figure 7H).

Finally, we examined the dose–response activation of CAR2 and CAR3 variants with compound **39** (Figure 7I). Unfortunately, the profiles of the dose–response curves did not reach the plateau phase and did not allow us to calculate EC_{50} and E_{max} values in the range of concentrations up to 30 μ M (Figure 7I).

Based on the data, we can conclude that compound **37** is the most active ligand for all CAR variants. Compound **39** displayed the most significant activity in the induction experiments in PHH and HepaRG cells irrespective of their lower affinities to wtCAR or CAR3 variants in CAR TR-FRET and cellular assays as well as marginal activity toward the CAR2 variant. In addition, we found that compound **39** does not induce *CYP2B6* or *CYP3A4* mRNA via PXR activation.

Next, we considered the physicochemical properties of selected compounds such as molecular weight (M_w), Log *S* (the solubility of a substance, measured in mol/L), and Log *P* (the partition coefficient is a ratio of concentrations of nonionized compound between water and octanol). Compound **39** is the smallest and less lipophilic candidate compound with better-predicted water solubility among the selected candidate compounds (Table 8).

Table 8. Physicochemical Properties of Selected Selective CAR Ligands

compound	M_w	log <i>P</i>	Log <i>S</i>
37	478.33	5.76	−7.52
39	463.32	4.85	−6.96
40	477.35	5.08	−7.22
41	491.38	5.32	−7.34
42	493.35	5.28	−7.39

Therefore, we decided to use compound **39** in further experiments with other nuclear receptor assays and in humanized CAR mice.

Novel CAR Ligands Interact with His203 and Occupy a Hydrophobic Pocket in Human wtCAR-LBD. For the modeling analyses, we docked the compounds **37**, **39**, **40**, and **48** in human wtCAR-LBD using CITCO as a reference (Figure 8A,B). Furthermore, to explore the wtCAR-LBD conformational dynamics and the interactions of these novel compounds, we conducted 25 μ s of all-atom MD simulations (5 μ s for each system plus CITCO). We studied the differences in the protein–ligand interactions among the systems, in comparison to CITCO. We observed the relevant role of the hydrogen bond interaction of H203 with the phenylimidazole ring in the novel compounds ranging from ~65 to 90%. This interaction was observed in particular for compounds **39** and **40** for ~65 and 90%, respectively (Figure 8C,D; Figures S-4–S-7). In addition to H203, T225 and D228 also have a relevant role in compound **39** stabilization. These interactions were formed between the amide of compound **39** and the T225 and D228 backbone oxygen, (located on H6) for ~35 and 28% of the simulation time, respectively (Figure 8C,D; Figure S-7). It is noteworthy that in the wtCAR/CITCO simulations, these additional polar interactions are not observed.

In addition to the hydrogen bond interactions, all novel compounds show high hydrophobic interaction frequency with F161 (~100%), the H203 imidazole ring (~70 to 100%), and Y224 (90–100%, except for the compound **40**, which is around 20%), and lower interaction with C202, F234, Y326,

and L242 (no interaction with compound **40**) (Figure 8C,D; Figures S-7 and S-8). These interactions were similarly observed with CITCO. Also, some interactions are compound-specific such as I164 with compound **39**, L206 with compounds **37** and **40**, F217 with compounds **37**, **40**, and **48**, and L239 with compounds **40** and **48**. Overall, we observed that all the novel compounds adopted U-shaped conformations similar to CITCO within the wtCAR-LBD (Figure S-6). This conformation is mainly supported by hydrophobic interactions, with an exclusive interaction for compound **39** with I164, and extra T225 and D228 hydrogen bonds for compound **39**, which stabilizes the compound within human wtCAR-LBD.

H12 Positioned in Close Vicinity of H3. MDs revealed no direct interaction between CITCO and residues from H12. In this regard, we then proceeded to investigate the changes in geometry and dynamicity of this region relative to the LBD with novel compounds and CITCO. For this purpose, we calculated the distance between H12 and H3 (center of mass of each helix). The result showed that all novel compounds can stabilize the conformation of H12 in the close vicinity of H3 similar to CITCO (Figure 8E, Figure S-9A). This geometry is known to initiate receptor activation.¹⁵ It has been reported that H12 stays away from the pocket due to the barrier formed by hydrophobic residues in the LBD,⁴⁴ where H11 directs the H12 in this active position.¹⁵ Previous studies also indicate that the free carboxylate of the H12 C-terminus interacts with the K195 side chain (on H4), leading to further H12 stabilization.¹⁵ To assess this phenomenon over the simulation time, we next calculated the distance between the carboxylate group of the H12 C-terminus and the polar group of K195 (Figure 8E; Figure S-9B). The median value for this distance in both wtCAR/CITCO and wtCAR/compounds **37**, **39**, and **40** stands around 3.1 Å, with a further distribution with compound **48**. This geometry enables the hydrogen bond formation between the H3 and H12 regions, providing extra stability to the systems. Taken together, this supports our result in terms of the high binding affinity and potency of our novel compounds.

Further Geometry Stabilization through N165–Y326 Interaction. Along with the closeness of H12 and H3, and the interaction between the H4 and H12 C-terminus, the stabilization of the systems comes through the hydrogen bond interaction between N165 (H3) and Y326 (H10). Both CITCO and **39** show relatively similar rigidity in this region (Figure 8F). The same trends are also observed with other novel compounds (Figure S-8C) with further distribution in the presence of compounds **40** and **48**. Although this interaction has been previously observed in the crystal structure with CITCO,¹⁵ MD data indicates that it is also relevant for our novel compounds.

Taken together, our docking data followed by microsecond timescale all-atom MD simulations revealed that CITCO and compound **39** interact with wtCAR-LBD mainly by hydrophobic contacts and that stronger polar contacts were formed between compound **39** and wtCAR-LBD compared to CITCO due to hydrogen bond interactions between comp. **39** amide moiety and T225 and D228 backbone oxygen. Interestingly, previous findings report that no specific hydrogen bonds are required for CITCO stability inside the CAR.^{18,45} Analyses of the MD trajectories showed that the interaction between compound **39** and I164 besides the higher interaction frequency with Y326 (hydrophobic interaction) compared to that of the CITCO (Figure 8D) could highlight the critical role

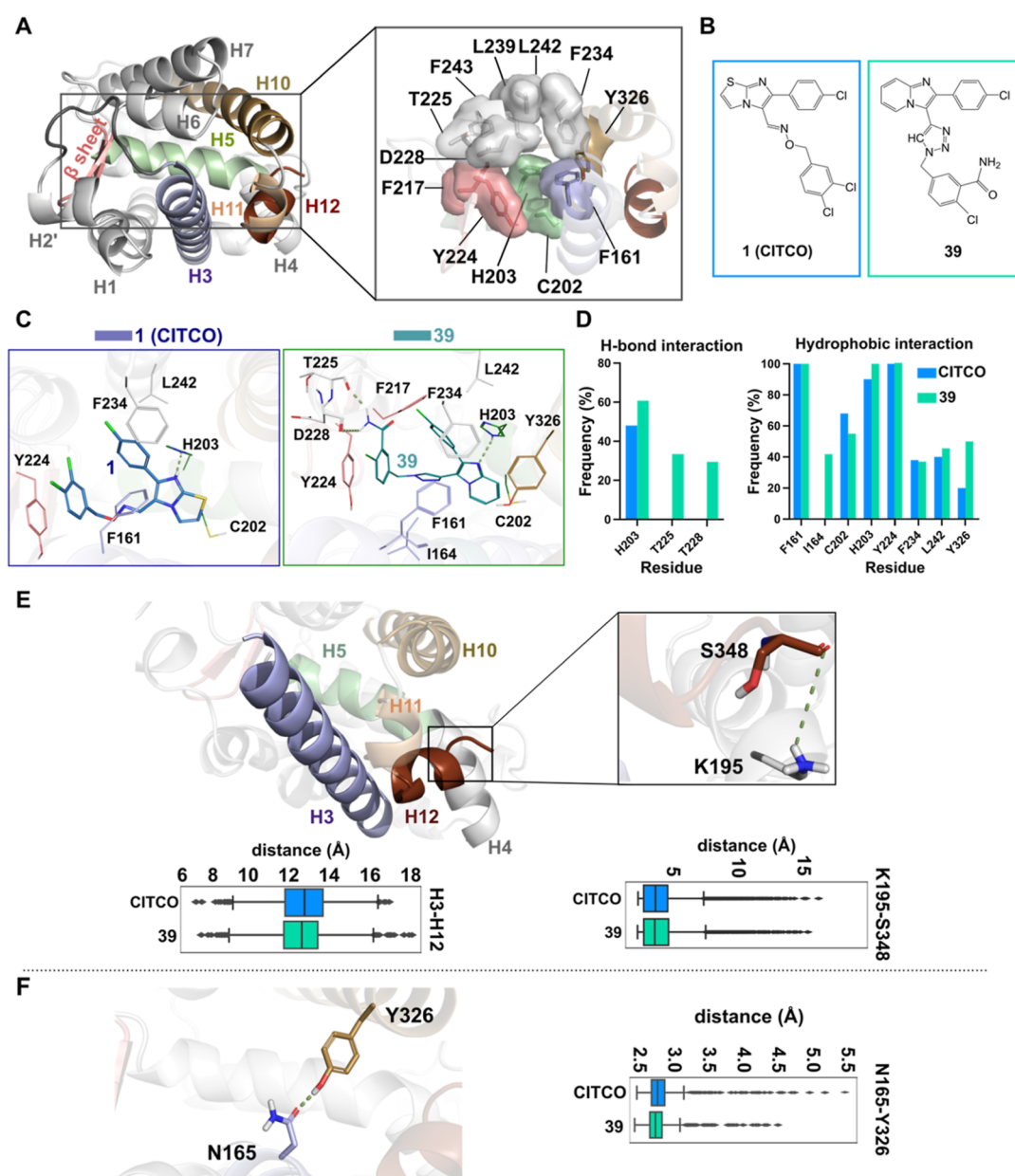


Figure 8. (A) Overview of the CAR LBD structure (wtCAR as the reference structure) and the small-molecule ligand (39) used in this study. The regions of interest are highlighted as follows: H2'-H3 loop (residues 140–153), dark gray; H3 loop (residues 157–178), light blue; H5 (residues 196–209), pale green; β sheets (residues 217–223), pink; H10 (residues 308–333), light brown; H11 (residues 336–339), light orange; H12 (residues 341–348), dark brown. The rectangular area denotes the location of the ligand-binding pocket (LBP) and the residues forming the LBP. The main residues participating in ligand binding are depicted in the stick model with a transparent molecular surface. Residues are colored according to their respective regions (see left structure). (B) 2D structure of CITCO and compound 39. (C) Representative snapshots of LBP with CITCO and compound 39 are shown. The green dashed line represents the hydrogen bond. (D) Frequencies of protein–ligand hydrogen bonds and protein–ligand hydrophobic interactions in percent are shown on the right of the panel. (E) Close view of H3 and H12 zooming in K195 (on H4) and S348 (on H12). The green dashed line represents the hydrogen bond between K195 and S348. Distance between H3 and H12 (center of mass) is represented in the left box plot. Distance between K195 and S348 (oxygen atoms) is represented in the right box plot. The black line in each box represents the median value. (F) Hydrogen bond between the Y326 oxygen atom and N165 polar group is shown as the green dashed line. Color codes are the same as in panel A. Distance between N165 and Y326 (oxygen atoms) is represented in the right box plot.

of H3 and H10 in protein stabilization. Of note, H10 lies on the heterodimerization interface where RXR α binds to the CAR. Our MD data also revealed that the H12 region is ordered and stable upon ligand binding. This event has been earlier reported as a driving force for CAR constitutive activity¹⁵ and, therefore, supports the agonistic effect of compound 39.

Selectivity of Compound 39 to Other Nuclear Receptors. Next, we sought to determine whether compound 39 is selective to the human CAR and whether it activates other nuclear receptors, for which a set of luciferase reporter assays was employed. We confirmed the selectivity of compound 39 for CAR as with no other nuclear receptor or the transcription factor aryl hydrocarbon receptor (AhR) was

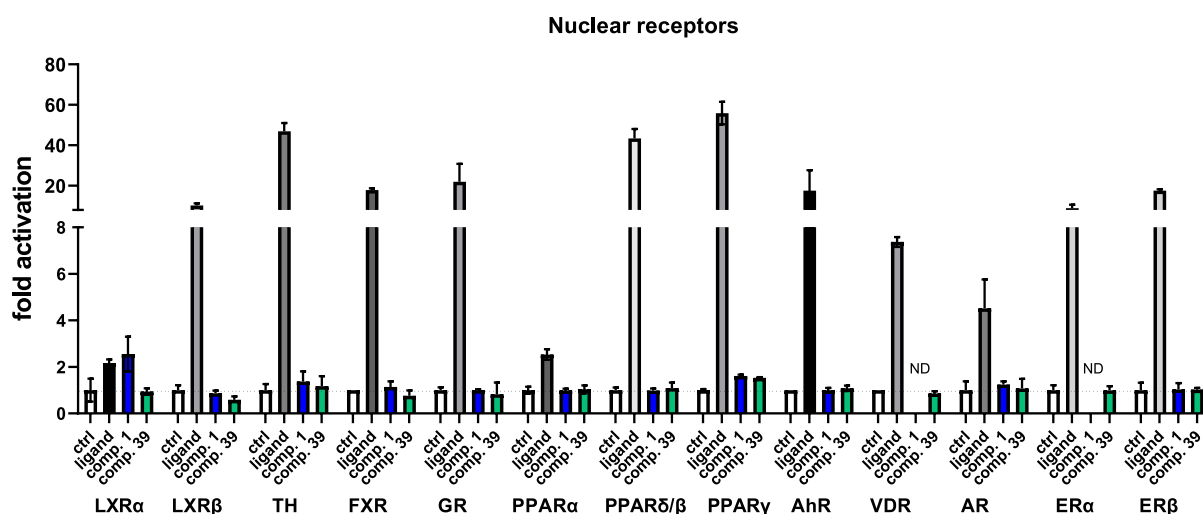


Figure 9. Luciferase reporter assays for human nuclear receptors LXR α , LXR β , TH, FXR, GR, PPAR α , PPAR δ/β , PPAR γ , VDR, AR, ER α , and ER β and for the AhR transcription factor were used to confirm the selectivity of compound 39. Specific ligands (GW3965, thyroxine, obeticholic acid, dexamethasone, fenofibrate, GW501516, rosiglitazone, 3-methylcholanthrene, calcitriol, testosterone, and estradiol) have been used in various luciferase reporter assays. Compound 1 (CITCO) and compound 39 have been tested at 10 μ M in HepG2 cells treated for 24 h.

significantly activated by the compound at 10 μ M concentration (Figure 9).

Microsomal Stability Experiments and Pilot Animal Pharmacokinetic Study. In the following experiments, we evaluated both the plasma and microsomal stability of compound 39 HCl in human plasma, human liver microsomes, as well as liver fraction S9 in time intervals of up to 120 min (Figure 10A,B; Table S-5). We found that compound 39 is highly stable in human plasma ($t_{1/2} \geq 240$ min). However, we observed a significant decline of compound 39 concentration in human microsomes as well as fraction S9 ($t_{1/2} = 38.04$ min and $t_{1/2} = 42.4$ min, respectively) (Figure 10B; Table S-5).

We also evaluated the plasma protein binding of compound 39 in both human and mouse plasma, determining that 98% of compound 39 is bound to human plasma proteins (Table S-6). We observed very similar properties of compound 39 in mouse plasma and mouse hepatic microsomes (Tables S-5–S-7).

In a pilot single-dose pharmacokinetic study, we found fast absorption of compound 39 HCl hydrochloric salt after p.o. application in gavage, although the compound was rapidly eliminated from the plasma (Figure 10C; Table S-9). Significantly, we detected traces of metabolites for compound 39 after i.v. application in plasma. Metabolites M1 and M2 represent compounds 41 and 40. Both compounds are N-methylated derivatives of compound 39 with significant CAR activity. Minor metabolite M3 (compound 34) is 2-(4-chlorophenyl)-3-(1H-1,2,3-triazol-4-yl)imidazo[1,2-a]pyridine, indicating that hepatic metabolic enzymes may attack the methylene bridge between the heterocycle and phenyl rings (Figure 10D). The metabolite is inactive with respect to the CAR activation, and the metabolite was not observed in human liver microsomes with the S9 fraction (data not shown). These data suggest that compound 39 is the main active compound, and it is likely eliminated intact as the parent compound. Nevertheless, further detailed pharmacokinetic studies should focus on the distribution, biliary elimination, and phase II metabolic clearance of the compound.

In addition, we conducted another examination to determine whether compound 39 inhibits the activities of major human cytochrome P450 enzymes. We found that

compound 39 has a minor effect on major cytochrome P450 enzymes. Compound 39 inhibits enzymatic activities of CYP3A4 (with $IC_{50} = 16.08$ μ M) and CYP1A2 ($IC_{50} = 21.07$ μ M) in higher micromolar concentrations, but the compound has no activity on the CYP2B6 enzyme up to 30 μ M concentration (Figure 10E).

Effects of Compound 39 in CAR Humanized Mice. Next, we treated humanized PXR/CAR/CYP3A mice with compound 39 to study the regulation of CAR target genes after a single i.p. application.

We found that compound 39 significantly upregulates *Cyp2b10* mRNA and protein, and human *CYP3A4* mRNA in the humanized model, but significantly decreases the expression of genes *Scd1* and *G6pc* after a single dose of 1 mg/kg. The latter genes are critically involved in triglyceride synthesis and gluconeogenesis in the liver. CITCO appeared more potent to induce *Cyp2b10* mRNA but less potent to upregulate *CYP3A4* mRNA expression, confirming the high efficiency of compound 39 to regulate the key CAR target genes in murine hepatocytes. We also observed the trend of a decrease of *Srebp1* and *Fasn* mRNA expression after compound 39 application (1 mg/kg) (Figure 11A), which agrees with data observed with the mouse CAR ligand TCPOBOP. This suggests that the human CAR ligand 39 recapitulates the significant effect of the murine ligand TCPOBOP on the regulation of lipid metabolism.^{4,5}

We did not observe upregulation of the genes involved in rodent liver proliferation after CAR activation and liver weight gain in the experiments (Figure 11B,C). Nevertheless, long-term studies are needed to examine liver hypertrophy and hyperplasia after repeated treatment with compound 39.

In analyzing blood biochemistry data after the single-dose application of compound 39 (dose 10 mg/kg), we observed a statistically significant decrease in plasma low-density lipoprotein (LDL) levels. This is consistent with results found with the mouse CAR ligand TCPOBOP in wild-type mice, indicating a positive effect of CAR activation on LDL plasma levels.⁹ We also observed a decrease in bile acid and total bilirubin (bilirubin-T) plasma levels after the application of compound 39, although these effects were not statistically

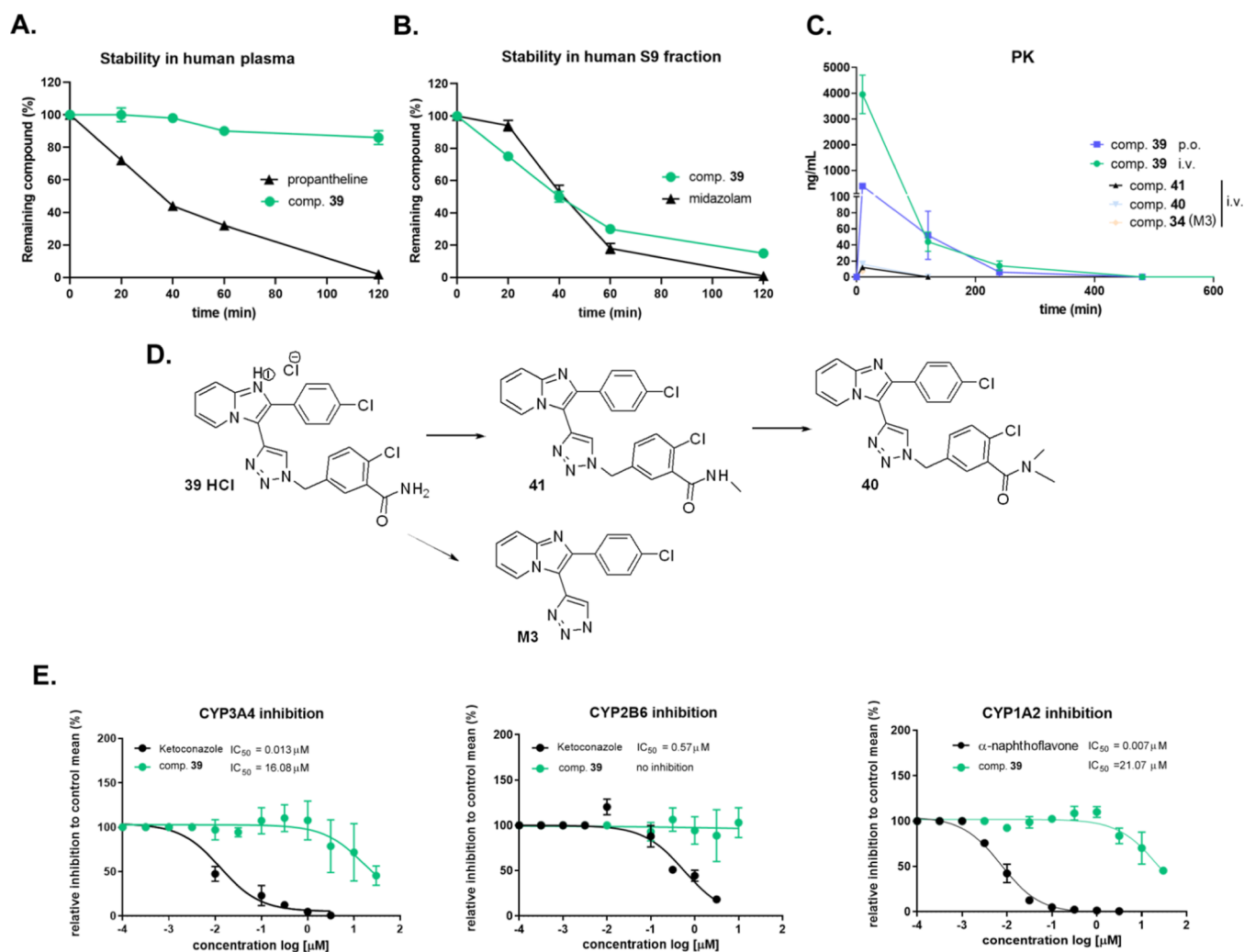


Figure 10. Plasma and microsomal stability experiments and single-dose pharmacokinetics in C57BL/6N mice. The stability of compound 39 in human plasma (A) and human microsomes with S9 fraction (B) were analyzed after 2 h of treatment. (C) Pharmacokinetics (PK) after single-dose application of compound 39 as hydrochloric salt either via i.v. or peroral application (10 mg/kg, $n = 4$) were analyzed in mice over 480 min. (D) Metabolites M1 (comp. 41), M2 (comp. 40), and M3 (comp. 34, 2-(4-chlorophenyl)-3-(1*H*-1,2,3-triazol-4-yl)imidazo[1,2-*a*]pyridine) of compound 39 were observed after i.v. application. Samples have been analyzed using HPLC-MS/MS. (E) Inhibition of CYP3A4, CYP2B6, and CYP1A2 enzymes in microsomes. Compound 39 was tested in the concentration range from 0.1 nM up to 30 μM. Relative activity data were fitted, and dose–response curves were used to obtain IC₅₀.

significant. Neither glucose, plasma triglycerides (TG), HDL lipoproteins, nor liver injury biomarkers (AST, ALT, and LDH) was significantly affected by compound 39 after the single-dose application (Figure 11D).

These results of the pilot single-dose pharmacokinetic study suggest that compound 39 is a novel effective human CAR agonist in animal experiments, a finding which warrants further repeated-dose long-term proof-of-concept studies.

Toxicity Studies of Compound 39. We observed no cytotoxicity in HepG2, COS-1 (Table S-3), HepaRG, HepaRG KO CAR, or in the PHHs after 48 h treatment (data not presented). Furthermore, in the Repeated Dose 7 day Oral Toxicity Study in Rodents (EMA/CPMP/ICH/286/1995, 2009 guidelines), no significant signs of toxicity were observed after the 7 days of oral administration of compound 39 HCl into rats. In particular, no significant changes in body weight, changes in behavior, gross pathology, hematology, and biochemistry parameters were observed after the 7 days of oral administration of the compound 39 HCl in all groups (groups with 1, 10, and 30 mg/kg b.w.) when compared to the

control group. We also tested the cardiotoxicity of compound 39 in a modified hERG fluorescence polarization assay. We did not observe any binding of compound 39 to hERG up to 20 μM (Supporting Information, Chapter 10).

Finally, we did not observe any frame-shift or base-pair substitution mutagenicity of compound 39 in a modified Ames fluctuation assay performed on *Salmonella typhimurium* TA100 and TA98 strains at a concentration of 1 and 10 μM (Table S-10).

CONCLUSIONS

Attempts to delineate the therapeutic implications of the CAR in humans have been hindered by the significant overlap in the pharmacology of human CAR and PXR receptors and the lack of a highly selective and potent human CAR agonist with suitable ADME properties.

In this work, we used a rational design of novel selective human CAR agonists. We applied a bioisosteric approach to the central part of the hit molecules 2 and 3 to prepare new

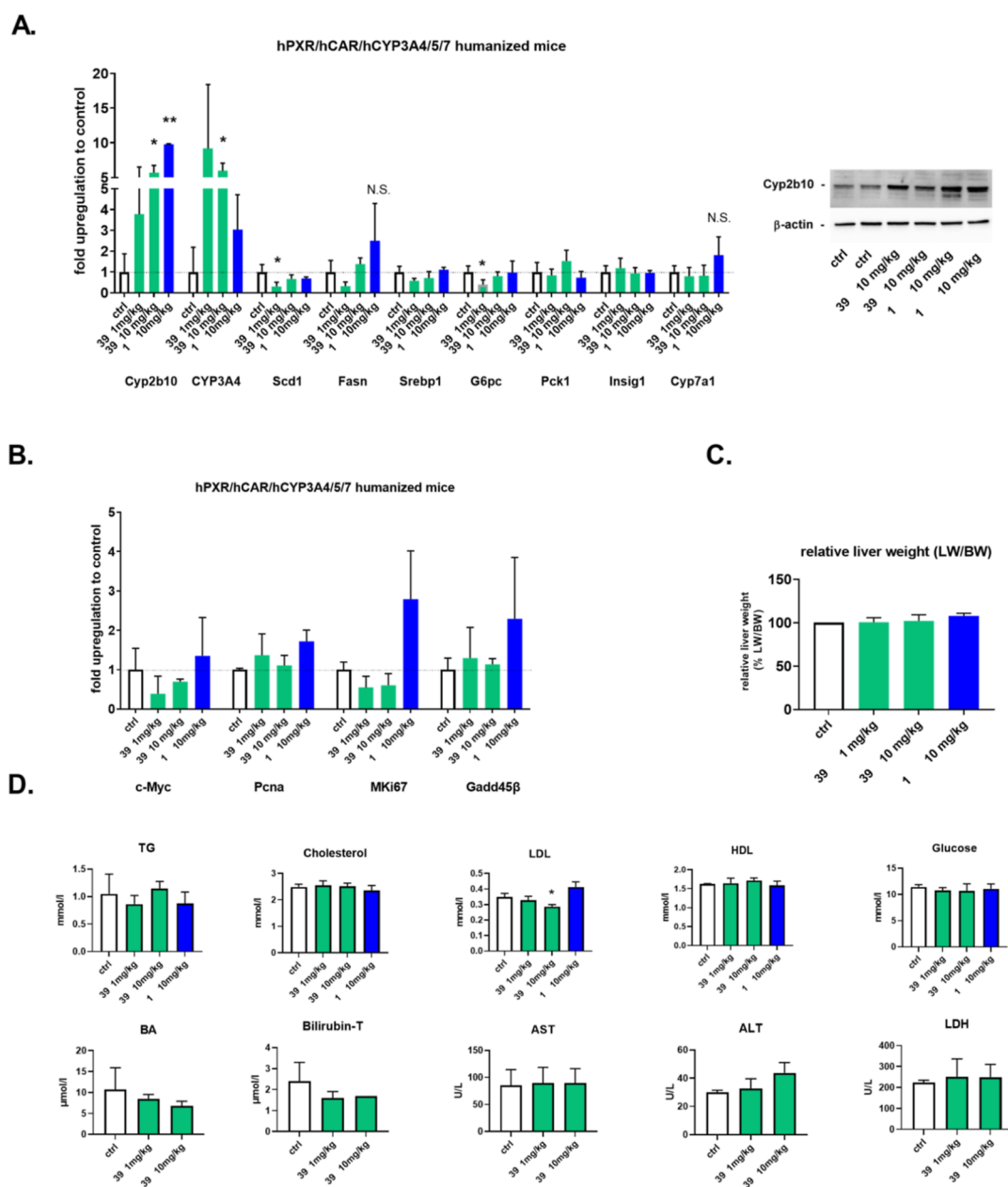


Figure 11. In vivo effects of compound **39** on liver CAR target genes involved in the intermediary metabolism of glucose, lipids and bile acids, hepatocyte proliferation, and apoptosis in humanized PXR/CAR/CYP3A mice ($n = 4$) after single i.p. application of the dose 1 or 10 mg/kg. Mice were sacrificed 36 h after application; livers were subjected to RT-qPCR analysis (A,B), western blotting analysis with anti-Cyp2b10 antibody, or were weighted (C). Blood samples were analyzed for biochemical parameters (D). * $p < 0.05$ -significant effect vs control (vehicle-treated) mice.

ligands for this human nuclear receptor. We were thus able to design a series of novel compounds that differed significantly in both nominal activities as CAR agonists and selectivity toward the PXR receptor as well as enhanced stability in comparison with the model compound CITCO. Based on our results, we performed a careful multiparametric selection of suitable candidates for further pharmacodynamic and pharmacokinetic studies. We found that the imidazo[1,2-*a*]pyridine core with the 1,2,3-triazole linker can be used for the further design of specific human CAR ligands. Replacement of the flexible oxime linker of CITCO with the triazole ring offered stability, less flexibility, and good accessibility via an undemanding click reaction. Modification of the 3,4-dichlorophenyl moiety of the hit compound **3** with amides (analogues **39–42**) resulted in CAR ligands without agonistic activities to PXR (Scheme 6,

Table 7, and Figure 6). Although extremely potent CAR agonists emerged in the resulting library of compounds, we also had to consider their metabolic stability and activity toward PXR. As a result, we decided to use compound **39** for further experiments, which, although not among our most potent CAR agonists, exhibited a desirable CAR/PXR profile and reasonable metabolic stability, allowing subsequent in vivo experiments. Using this chemical tool, which we have shown to have no observable toxicity or genotoxic potential, we were able to prove that compound **39** significantly activates the human CAR, both in vitro in human hepatocyte models and CAR humanized mice. Significantly, we noted that compound **39** regulates typical CAR target genes involved in xenobiotic (*Cyp2b10*), lipid (*Scd1*), or glucose (*G6pc*) metabolism, and it

decreases plasma LDL lipoproteins even after a single dose in humanized PXR/CAR/CYP3A4/3A7 mice.

In summary, our work identifies a selective CAR receptor agonist for which we have demonstrated both in vitro and in vivo activities in relevant models for the human CAR. Compound **39** thus warrants further preclinical studies in humanized CAR models or human hepatocyte models to better understand the unique function of the human CAR without confounding off-target effects on PXR receptor activation.

METHODS

Experimental Methods. Synthesis of Novel Ligands. General chemical procedures: NMR spectra were measured on a Bruker AVANCE II-600 and/or Bruker AVANCE II-500 instruments (600.1 or 500.0 MHz for ^1H and 150.9 or 125.7 MHz for ^{13}C) in hexadeuterodimethyl sulfoxide and referenced to the solvent signal (δ 2.50 and 39.70, respectively). Mass spectra were measured on a LTQ Orbitrap XL (Thermo Fischer Scientific) using electrospray ionization (ESI) and a GCT Premier (Waters) using EI. The elemental analyses were obtained on a Perkin Elmer CHN Analyzer 2400, Series II Sys (PerkinElmer), and X-ray fluorescence spectrometer SPECTRO iQ II (SPECTRO Analytical Instruments, Germany). Column chromatography and thin-layer chromatography (TLC) were performed using Silica gel 60 (Fluka) and Silufol Silica gel 60 F₂₅₄ foils (Merck), respectively. The purity of newly synthesized compounds was >95%, confirmed by UPLC-MS. Solvents were evaporated at 2 kPa and a bath temperature of 30–60 °C. The compounds were dried at 13 Pa and 50 °C.

General Procedure I: Cyclization of Heterocycle. 2-Aminothiazole (3) or 2-aminopyridine (4) was dissolved in EtOH, and substituted or unsubstituted bromoacetophenone derivative (1 equiv) was added, followed by the addition of NaHCO₃ (1 equiv). The reaction mixture was heated at 70 °C overnight. After the completion of the reaction (monitored by TLC or UPLC), the solvent was evaporated to a minimal volume, and the residue was diluted with EtOAc and washed with water. The water phase was extracted twice more with EtOAc, and the combined organic phases were dried over sodium sulfate and evaporated. The residue was purified by flash column chromatography (eluent petrol ether/EtOAc or EtOAc/MeOH).

General Procedure II: Iodination. 2-Substituted imidazo[1,2-*a*]pyridines or imidazo[2,1-*b*]thiazoles were dissolved in CH₃CN (5 mL/mmol) and NIS (1.05 equiv) was added in one portion. The suspension was stirred at 25 °C, and the conversion was monitored by TLC. After the completion of the reaction (1–4 h), the reaction mixture was diluted with EtOAc and washed with a saturated Na₂S₂O₃ solution. The inorganic phase was extracted twice more with EtOAc; combined organic phases were dried over sodium sulfate and evaporated. The residue was purified by flash column chromatography, with the mobile phase petrol ether/EtOAc (10:50%).

General Procedure III: Sonogashira Coupling. 3-Iodoimidazo[1,2-*a*]pyridines or 5-iodoimidazo[2,1-*b*]thiazoles were placed in a dried round-bottom flask, diluted with dry DMF, degassed at 0 °C, and flushed with argon. CuI (10 mol %) and Pd(PPh₃)₂Cl₂ (5 mol %) were added, the mixture was properly degassed, dry TEA (3 equiv) was added, and the mixture was degassed again. Finally, TMS-acetylene (5 equiv) was added in one portion. The reaction mixture was stirred at 25 °C under an argon atmosphere. After the completion of the reaction (monitored by TLC), the mixture was diluted if necessary with CHCl₃ and filtered over Celite. The filtrate was washed with water; the water phase was extracted twice more with CHCl₃; the combined organic phases were dried over sodium sulfate and evaporated. The residue was purified by flash column chromatography (mobile phase petrol ether/EtOAc).

General Procedure IV: Click Reaction. Trimethylsilyl(ethynyl)imidazo[2,1-*b*]thiazole or trimethylsilyl(ethynyl)imidazo[1,2-*a*]pyridine derivatives were dissolved in THF/H₂O mixture (1:1) and an appropriate azido intermediate (1 equiv) was added. The reaction

mixture was degassed at 0 °C, refilled with argon, and CuSO₄·5H₂O (10 mol %), KF (1 equiv), were Na-ascorbate (1 equiv) were added in one portion. The reaction mixture was stirred at 25 °C and monitored by TLC. After the completion of the reaction, the mixture was diluted with EtOAc and washed with water. The water phase was extracted twice more with EtOAc; combined organic phases were dried over sodium sulfate and evaporated. The residue was purified by flash column chromatography (eluent petrol ether/EtOAc or EtOAc/MeOH).

General Procedure V: Ester Hydrolysis. The methyl or ethyl ester derivative was dissolved in THF/H₂O 2:1, and LiOH·H₂O (4 equiv) was added in one portion. The reaction mixture was stirred at 25 °C and monitored by TLC. After the completion of the reaction, the mixture was extracted with EtOAc, and the water phase was acidified to pH 2 and extracted again with EtOAc. The organic phase was dried over sodium sulfate and purified by reverse-phase flash column chromatography.

General Procedure VI: Amide Preparation. 2-Chloro-5-((4-(2-(4-chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-1*H*-1,2,3-triazol-1-yl)-methyl)-benzoic acid (**38**) was placed in a round-bottom flask, and toluene was added (5 mL) followed by the addition of thionyl chloride (0.5 mL, in excess). The reaction mixture was stirred at 90 °C overnight. The reaction mixture was evaporated to dryness, coevaporated with toluene, and used directly for the next step without any purification. The intermediate 2-chloro-5-((4-(2-(4-chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-1*H*-1,2,3-triazol-1-yl)-methyl)benzoyl chloride (**53**) was dissolved in dry DCM and cooled in an ice bath. An appropriate amine (1.2 equiv) was added, followed by the addition of DIPEA (1.5 or 2 equiv in case of amine salts). The reaction mixture was stirred at 25 °C and monitored by TLC or LCMS. After completion of the reaction, the mixture was diluted with DCM, washed with water, and purified by reverse-phase flash CC or flash column chromatography.

General Procedure VII: Ketone Preparation Using the Grignard Reagent. 2-Chloro-5-((4-(2-(4-chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-1*H*-1,2,3-triazol-1-yl)methyl)-*N*-methoxy-*N*-methylbenzamide was dissolved in dry THF (6 mL), cooled in an ice bath, degassed, and refilled with argon. The Grignard reagent (3 M in DEE, 2 equiv) was added in one portion, and the mixture was allowed to warm to 25 °C and stirred overnight. For the LAH reduction in case of an aldehyde, the mixture was stirred at 5 °C for 2 h. The mixture was quenched with a saturated NH₄Cl solution and extracted with EtOAc. Combined organic phases were dried over sodium sulfate, evaporated, and purified if necessary by flash column chromatography (petrol ether/EtOAc 70:100%).

6-(4-Chlorophenyl)-5-(1-(3,4-dichlorobenzyl)-1*H*-1,2,3-triazol-4-yl)imidazo[2,1-*b*]thiazole (2). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (20:50%). Yield: 268 mg (90%). ^1H NMR (500 MHz, DMSO-*d*₆): δ 8.49 (s, 1H), 8.04 (d, *J* = 4.5 Hz, 1H), 7.67–7.70 (m, 4H), 7.42–7.45 (m, 2H), 7.39 (d, *J* = 4.5 Hz, 1H), 7.34 (ddm, *J* = 2.1 Hz, *J* = 8.3 Hz, 1H), 5.69 (br s, 2H). ^{13}C NMR (101 MHz, DMSO): δ 149.63, 142.82, 137.13, 136.99, 133.30, 132.39, 131.49, 131.24, 131.19, 130.42, 129.30, 128.73, 128.66, 123.62, 120.02, 114.52, 113.92, 51.85. HRMS: calcd for [M + H], 459.99518; found, 459.99522.

2-(4-Chlorophenyl)-3-(1-(3,4-dichlorobenzyl)-1*H*-1,2,3-triazol-4-yl)imidazo[1,2-*a*]pyridine (3). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (20:70%). Yield: 130 mg (57%). ^1H NMR (500 MHz, DMSO-*d*₆): δ 8.53 (s, 1H), 8.48 (dt, *J* = 1.1 Hz, *J* = 6.9 Hz, 1H), 7.67–7.71 (m, 5H), 7.41–7.44 (m, 2H), 7.39 (ddd, *J* = 1.2 Hz, *J* = 6.8 Hz, *J* = 9.1 Hz, 1H), 7.35 (dd, *J* = 2.1 Hz, *J* = 8.4 Hz, 1H), 7.00 (td, *J* = 1.2 Hz, *J* = 6.8 Hz, 1H), 5.75 (s, 2H). ^{13}C NMR (101 MHz, CDCl₃): δ 144.95, 142.52, 137.05, 136.40, 133.04, 132.86, 131.53, 131.28, 131.21, 130.34, 129.67, 128.71, 128.59, 126.30, 125.95, 125.45, 117.13, 113.38, 111.57, 51.94. HRMS: calcd for [M + H], 454.03876; found, 454.03886.

6-(4-Chlorophenyl)imidazo[2,1-*b*]thiazole (6a). The title compound was prepared according to General Procedure I. Mobile phase petrol ether/EtOAc (20:60%). Yield: 981 mg (82%). ^1H NMR (401

MHz, DMSO- d_6): δ 8.26 (s, 1H), 7.94 (d, J = 4.4 Hz, 1H), 7.90–7.80 (m, 2H), 7.48–7.39 (m, 2H), 7.28 (d, J = 4.4 Hz, 1H). ^{13}C NMR (101 MHz, DMSO): δ 149.88, 145.54, 133.64, 131.78, 129.09, 126.85, 120.52, 113.85, 110.30. EI MS: calcd for [M + H], 234.0018; found, 234.0020.

6-(3,4-Dichlorophenyl)imidazo[2,1-*b*]thiazole (6g). The title compound was prepared according to General Procedure I. Mobile phase petrol ether/EtOAc (20:50%). Yield: 878 mg (65%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.37 (s, 1H), 8.06 (d, J = 2.0 Hz, 1H), 7.97 (d, J = 4.5 Hz, 1H), 7.82 (dd, J = 8.4, 2.1 Hz, 1H), 7.64 (d, J = 8.4 Hz, 1H), 7.30 (d, J = 4.5 Hz, 1H). ^{13}C NMR (101 MHz, DMSO): δ 149.83, 143.93, 135.18, 131.65, 131.04, 129.21, 126.38, 124.93, 120.26, 113.98, 110.90. HRMS: calcd for [M + Na], 268.97015; found, 268.97029.

2-(4-Chlorophenyl)imidazo[1,2-*a*]pyridine (7a). The title compound was prepared according to General Procedure I. Mobile phase petrol ether/EtOAc (20:60%). Yield: 1.49 g (88%). ^1H NMR (401 MHz, chloroform-*d*): δ 8.12 (dt, J = 1.2 Hz, J = 6.8 Hz, 1H), 7.93–7.87 (m, 2H), 7.84 (d, J = 0.7 Hz, 1H), 7.64 (dq, J = 1.0 Hz, J = 9.2 Hz, 1H), 7.44–7.40 (m, 2H), 7.20 (ddd, J = 1.3 Hz, J = 6.8 Hz, J = 9.2 Hz, 1H), 6.80 (td, J = 1.2 Hz, J = 6.8 Hz, 1H). ^{13}C NMR (101 MHz, CDCl_3): δ 145.86, 144.82, 133.79, 132.45, 129.02, 127.38, 125.73, 125.03, 117.69, 112.72, 108.31. EI MS: calcd for [M + H], 228.0454; found, 228.0456.

2-(*p*-Tolyl)imidazo[1,2-*a*]pyridine (7b). The title compound was prepared according to the described procedure. The identity and purity were confirmed by NMR and HRMS. Mobile phase petrol ether/EtOAc (20:70%). Yield: 853 mg (76%).

^1H NMR (401 MHz, DMSO- d_6): δ 8.50 (dt, J = 6.8, 1.2 Hz, 1H), 8.33 (d, J = 0.7 Hz, 1H), 7.88–7.82 (m, 2H), 7.56 (dq, J = 9.1, 1.0 Hz, 1H), 7.27–7.19 (m, 3H), 6.87 (td, J = 6.7, 1.2 Hz, 1H), 2.33 (s, 3H). ^{13}C NMR (101 MHz, DMSO): δ 145.21, 144.96, 137.43, 131.62, 129.75, 127.24, 125.97, 125.23, 116.99, 112.61, 109.10, 21.33. HRMS: calcd for [M + H], 209.10732; found, 209.10745.

2-(4-Ethylphenyl)imidazo[1,2-*a*]pyridine (7c). The title compound was prepared according to General Procedure I. Mobile phase petrol ether/EtOAc (20:70%). Yield: 1.02 g (74%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.51 (dt, J = 6.8, 1.2 Hz, 1H), 8.34 (d, J = 0.7 Hz, 1H), 7.90–7.85 (m, 2H), 7.56 (dq, J = 9.1, 1.0 Hz, 1H), 7.29–7.26 (m, 2H), 7.23 (ddd, J = 9.1, 6.7, 1.3 Hz, 1H), 6.88 (td, J = 6.7, 1.2 Hz, 1H), 2.63 (q, J = 7.6 Hz, 2H), 1.21 (t, J = 7.6 Hz, 3H). ^{13}C NMR (101 MHz, DMSO): δ 145.21, 144.97, 143.79, 131.88, 128.56, 127.25, 126.04, 125.24, 117.00, 112.63, 109.14, 28.44, 16.01. HRMS: calcd for [M + H], 223.12298; found, 223.12301.

2-(4-Fluorophenyl)imidazo[1,2-*a*]pyridine (7d). The title compound was prepared according to General Procedure I. Mobile phase petrol ether/EtOAc (20:70%). Yield: 988 mg (88%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.52 (dt, J = 1.2, 7.0 Hz, 1H), 8.38 (s, 1H), 8.04–7.96 (m, 2H), 7.57 (d, J = 9.1 Hz, 1H), 7.31–7.21 (m, 3H), 6.89 (td, J = 1.3, 6.8 Hz, 1H). ^{13}C NMR (101 MHz, DMSO): δ 145.30, 143.91, 130.93, 127.37, 125.51, 117.06, 112.80, 109.41. ^{13}C NMR (101 MHz, DMSO- d_6): δ 162.29 (d, J = 244.2 Hz), 130.93, 127.97 (d, J = 8.2 Hz), 116.05 (d, J = 21.5 Hz). HRMS: calcd for [M + H], 213.08225; found, 213.08226.

2-(4-(Trifluoromethyl)phenyl)imidazo[1,2-*a*]pyridine (7e). The title compound was prepared according to General Procedure I. Mobile phase petrol ether/EtOAc (20:70%). Yield: 1.07 g (77%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.57–8.53 (m, 1H), 8.21–8.15 (m, 1H), 7.82–7.77 (m, 1H), 7.61 (dq, J = 9.1, 1.0 Hz, 1H), 7.29 (ddd, J = 9.1, 6.7, 1.3 Hz, 1H), 6.93 (td, J = 6.8, 1.2 Hz, 1H). ^{13}C NMR (101 MHz, DMSO): δ 145.20, 142.84, 138.12, 128.06, 127.75, 127.30, 126.21, 125.89, 125.86, 125.82, 125.78, 125.71, 117.04, 112.84, 110.73. HRMS: calcd for [M + H], 263.07906; found, 263.07907.

2-(2,4-Dichlorophenyl)imidazo[1,2-*a*]pyridine (7f). The title compound was prepared according to General Procedure I. Mobile phase petrol ether/EtOAc (20:70%). Yield: 634 mg (76%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.64 (d, J = 0.7 Hz, 1H), 8.61 (dt, J = 6.8, 1.2 Hz, 1H), 8.29 (d, J = 8.5 Hz, 1H), 7.71 (d, J = 2.2 Hz, 1H), 7.60 (dq, J = 9.1, 1.0 Hz, 1H), 7.54 (dd, J = 8.6, 2.2 Hz, 1H), 7.30 (ddd, J = 9.1, 6.7, 1.3 Hz, 1H), 6.94 (td, J = 6.8, 1.2 Hz, 1H). ^{13}C NMR (101

MHz, DMSO): δ 144.26, 139.80, 132.92, 132.22, 131.75, 131.67, 130.15, 128.08, 127.69, 126.21, 117.15, 113.42, 113.01. HRMS: calcd for [M + H], 263.01373; found, 263.01390.

2-(3,4-Dichlorophenyl)imidazo[1,2-*a*]pyridine (7g). The title compound was prepared according to General Procedure I. Mobile phase petrol ether/EtOAc (20:50%). Yield: 533 mg (87%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.54–8.49 (m, 2H), 8.18 (d, J = 2.0 Hz, 1H), 7.93 (dd, J = 8.4, 2.0 Hz, 1H), 7.68 (d, J = 8.4 Hz, 1H), 7.58 (dq, J = 9.1, 1.0 Hz, 1H), 7.27 (ddd, J = 9.1, 6.7, 1.3 Hz, 1H), 6.91 (td, J = 6.8, 1.2 Hz, 1H). ^{13}C NMR (101 MHz, DMSO): δ 145.10, 141.99, 134.88, 131.74, 131.13, 130.02, 127.25, 127.23, 125.73, 125.70, 116.94, 112.84, 110.47. HRMS: calcd for [M + H], 263.01373; found, 263.01393.

4-(Imidazo[1,2-*a*]pyridin-2-yl)benzonitrile (7h). The title compound was prepared according to General Procedure I. Mobile phase petrol ether/EtOAc (20:50%). Yield: 952 mg (82%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.56 (s, 1H), 8.54 (dt, J = 6.8, 1.2 Hz, 1H), 8.18–8.12 (m, 2H), 7.91–7.85 (m, 2H), 7.60 (dq, J = 9.1, 1.0 Hz, 1H), 7.28 (ddd, J = 9.1, 6.7, 1.3 Hz, 1H), 6.92 (td, J = 6.7, 1.2 Hz, 1H). ^{13}C NMR (101 MHz, DMSO): δ 145.40, 142.63, 138.68, 133.03, 127.44, 126.39, 126.10, 119.33, 117.14, 113.13, 111.41, 110.06. HRMS: calcd for [M + H], 220.08692; found, 220.08686.

2-(4-Methoxyphenyl)imidazo[1,2-*a*]pyridine (7i). The title compound was prepared according to General Procedure I. Mobile phase petrol ether/EtOAc (20:70%). Yield: 200 mg (84%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.49 (dt, J = 6.8, 1.2 Hz, 1H), 7.92–7.87 (m, 2H), 7.54 (dq, J = 9.1, 1.0 Hz, 1H), 7.21 (ddd, J = 9.1, 6.7, 1.3 Hz, 1H), 7.03–6.98 (m, 2H), 6.86 (td, J = 6.7, 1.2 Hz, 1H). ^{13}C NMR (101 MHz, DMSO): δ 159.48, 145.19, 144.90, 127.33, 127.16, 127.00, 125.09, 116.86, 114.59, 112.51, 108.48, 55.60. HRMS: calcd for [M + H], 240.07675; found, 240.07674.

6-(4-Chlorophenyl)-5-iodoimidazo[2,1-*b*]thiazole (8a). The title compound was prepared according to General Procedure II. Mobile phase petrol ether/EtOAc (10:50%). Yield: 421 mg (89%). ^1H NMR (401 MHz, DMSO- d_6): δ 7.98 (m, 2H), 7.87 (d, J = 4.5 Hz, 1H), 7.52 (m, 2H), 7.42 (d, J = 4.5 Hz, 1H). ^{13}C NMR (101 MHz, DMSO): δ 149.88, 145.54, 133.64, 131.78, 129.09, 126.85, 120.52, 113.85, 110.30. HRMS: calcd for [M + H], 360.90577; found, 360.90586.

6-(3,4-Dichlorophenyl)-5-iodoimidazo[2,1-*b*]thiazole (8g). The title compound was prepared according to General Procedure II. Mobile phase petrol ether/EtOAc (10:50%). Yield: 514 mg (96%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.17 (d, J = 2.1 Hz, 1H), 7.99 (dd, J = 8.5, 2.1 Hz, 1H), 7.90 (d, J = 4.5 Hz, 1H), 7.75 (d, J = 8.5 Hz, 1H), 7.46 (d, J = 4.5 Hz, 1H). ^{13}C NMR (101 MHz, DMSO): δ 150.98, 145.37, 134.98, 131.61, 131.21, 130.34, 128.53, 126.97, 120.67, 111.97, 60.96. HRMS: calcd for [M + H], 394.86679; found, 394.86699.

2-(4-Chlorophenyl)-3-iodoimidazo[1,2-*a*]pyridine (9a). The title compound was prepared according to General Procedure II. Mobile phase petrol ether/EtOAc (10:50%). Yield: 352 mg (98%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.42 (dt, J = 6.9, 1.2 Hz, 1H), 8.11–8.06 (m, 2H), 7.63 (dt, J = 9.0, 1.2 Hz, 1H), 7.60–7.55 (m, 2H), 7.38 (ddd, J = 9.0, 6.8, 1.3 Hz, 1H), 7.09 (td, J = 6.8, 1.2 Hz, 1H). ^{13}C NMR (101 MHz, CDCl_3): δ 147.52, 145.52, 132.99, 132.85, 129.73, 128.67, 127.35, 126.47, 117.16, 113.93, 63.82. HRMS: calcd for [M + H], 354.94935; found, 354.94944.

3-Iodo-2-(*p*-tolyl)imidazo[1,2-*a*]pyridine (9b). The title compound was prepared according to General Procedure II. Mobile phase petrol ether/EtOAc (10:50%). Yield: 501 mg (94%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.41 (dt, J = 6.9, 1.2 Hz, 1H), 7.98–7.92 (m, 2H), 7.61 (dt, J = 9.0, 1.1 Hz, 1H), 7.36 (ddd, J = 9.0, 6.7, 1.3 Hz, 1H), 7.31 (d, J = 8.0 Hz, 2H), 7.06 (td, J = 6.8, 1.2 Hz, 1H), 2.36 (s, 3H). ^{13}C NMR (101 MHz, DMSO): δ 147.70, 147.12, 137.95, 131.41, 129.41, 128.32, 127.44, 126.35, 117.29, 113.92, 63.23, 21.36. HRMS: calcd for [M + H], 335.00397; found, 335.00397.

2-(4-Ethylphenyl)-3-iodoimidazo[1,2-*a*]pyridine (9c). The title compound was prepared according to General Procedure II. Mobile phase petrol ether/EtOAc (10:50%). Yield: 1.23 g (86%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.41 (dt, J = 6.9, 1.1 Hz, 1H), 8.00–7.95

(m, 2H), 7.62 (dt, $J = 9.0, 1.1$ Hz, 1H), 7.39–7.36 (m, 3H), 7.35–7.31 (m, 2H), 7.07 (td, $J = 6.8, 1.2$ Hz, 1H), 2.66 (q, $J = 7.6$ Hz, 2H), 1.22 (t, $J = 7.6$ Hz, 3H). ^{13}C NMR (101 MHz, DMSO): δ 147.41, 146.83, 143.95, 131.35, 128.10, 127.93, 127.16, 126.08, 117.00, 113.65, 62.95, 28.15, 15.67. HRMS: calcd for $[\text{M} + \text{H}]$, 349.01962; found, 349.01982.

2-(4-Fluorophenyl)-3-iodoimidazo[1,2-*a*]pyridine (9d). The title compound was prepared according to General Procedure II. Mobile phase petrol ether/EtOAc (10:50%). Yield: 1.42 g (97%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.40 (dt, $J = 6.9, 1.1$ Hz, 1H), 8.12–8.06 (m, 2H), 7.62 (dt, $J = 9.0, 1.1$ Hz, 1H), 7.40–7.30 (m, 3H), 7.07 (td, $J = 6.8, 1.2$ Hz, 1H). ^{13}C NMR (101 MHz, DMSO): δ 162.15 (d, $J = 245.5$ Hz), 147.44, 145.88, 130.42 (d, $J = 3.0$ Hz), 130.15 (d, $J = 8.3$ Hz), 127.25, 126.28, 117.05, 115.48 (d, $J = 21.5$ Hz), 113.78, 63.34. HRMS: calcd for $[\text{M} + \text{H}]$, 338.97980; found, 338.97902.

3-Iodo-2-(4-(trifluoromethyl)phenyl)imidazo[1,2-*a*]pyridine (9e). The title compound was prepared according to General Procedure II. Mobile phase petrol ether/EtOAc (10:50%). Yield: 1.25 g (92%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.46 (d, $J = 6.9$ Hz, 1H), 8.31 (d, $J = 8.0$ Hz, 2H), 7.88 (d, $J = 8.1$ Hz, 2H), 7.67 (d, $J = 9.0$ Hz, 1H), 7.42 (dd, $J = 8.9, 6.9$ Hz, 1H), 7.12 (t, $J = 6.8$ Hz, 1H). ^{13}C NMR (101 MHz, DMSO): δ 147.90, 145.35, 138.31, 128.87, 128.65, 127.74, 126.96, 125.81 ($J = 3.78$ Hz), 124.77, 117.61, 114.38, 65.13. HRMS: calcd for $[\text{M} + \text{H}]$, 388.97570; found, 388.97582.

2-(2,4-Dichlorophenyl)-3-iodoimidazo[1,2-*a*]pyridine (9f). The title compound was prepared according to General Procedure II. Mobile phase petrol ether/EtOAc (10:50%). Yield: 238 mg (89%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.40 (dt, $J = 6.9, 1.2$ Hz, 1H), 7.79 (d, $J = 2.0$ Hz, 1H), 7.64 (dt, $J = 9.0, 1.1$ Hz, 1H), 7.59–7.51 (m, 2H), 7.40 (ddd, $J = 9.0, 6.8, 1.3$ Hz, 1H), 7.12 (td, $J = 6.8, 1.2$ Hz, 1H). ^{13}C NMR (101 MHz, DMSO): δ 147.39, 146.72, 134.57, 134.42, 134.34, 132.87, 129.62, 127.72, 127.46, 126.52, 117.62, 114.32, 67.98. HRMS: calcd for $[\text{M} + \text{H}]$, 388.91037; found, 388.91071.

2-(3,4-Dichlorophenyl)-3-iodoimidazo[1,2-*a*]pyridine (9g). The title compound was prepared according to General Procedure II. Mobile phase petrol ether/EtOAc (10:50%). Yield: 638 mg (98%). ^1H NMR (401 MHz, chloroform-*d*): δ 8.24 (dt, $J = 6.9, 1.1$ Hz, 1H), 8.22 (d, $J = 2.1$ Hz, 1H), 7.96 (dd, $J = 8.4, 2.1$ Hz, 1H), 7.66 (dt, $J = 9.0, 1.2$ Hz, 1H), 7.55 (d, $J = 8.4$ Hz, 1H), 7.33 (ddd, $J = 9.1, 6.8, 1.3$ Hz, 1H), 6.99 (td, $J = 6.9, 1.2$ Hz, 1H). ^{13}C NMR (101 MHz, DMSO): δ 179.90, 147.82, 144.25, 134.93, 131.66, 131.24, 131.09, 129.70, 128.14, 127.74, 127.08, 117.55, 114.43, 64.93. HRMS: calcd for $[\text{M} + \text{H}]$, 388.91037; found, 388.91052.

4-(3-Iodoimidazo[1,2-*a*]pyridin-2-yl)benzotrile (9h). The title compound was prepared according to General Procedure II. Mobile phase petrol ether/EtOAc (10:50%). Yield: 245 mg (90%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.46 (dt, $J = 7.0, 1.1$ Hz, 1H), 8.31–8.26 (m, 2H), 8.00–7.95 (m, 2H), 7.66 (dt, $J = 9.1, 1.1$ Hz, 1H), 7.41 (ddd, $J = 9.1, 6.8, 1.2$ Hz, 1H), 7.11 (td, $J = 6.8, 1.2$ Hz, 1H). ^{13}C NMR (101 MHz, DMSO): δ 147.64, 144.64, 138.54, 132.60, 128.49, 127.49, 126.85, 119.02, 117.37, 114.20, 110.54, 65.39. HRMS: calcd for $[\text{M} + \text{H}]$, 345.98357; found, 345.98370.

3-Iodo-2-(4-methoxyphenyl)imidazo[1,2-*a*]pyridine (9i). The title compound was prepared according to General Procedure II. Mobile phase petrol ether/EtOAc (10:50%). Yield: 238 mg (89%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.39 (dt, $J = 6.9, 1.1$ Hz, 1H), 8.05–7.96 (m, 2H), 7.60 (dt, $J = 9.0, 1.1$ Hz, 1H), 7.35 (ddd, $J = 9.0, 6.7, 1.2$ Hz, 1H), 7.09–7.05 (m, 3H), 3.82 (s, 3H). ^{13}C NMR (101 MHz, DMSO): δ 159.68, 147.67, 146.99, 129.70, 127.38, 126.59, 126.27, 117.16, 114.27, 113.83, 62.62, 55.66. HRMS: calcd for $[\text{M} + \text{H}]$, 350.99888; found, 350.99893.

6-(4-Chlorophenyl)-5-((trimethylsilyl)ethynyl)imidazo[2,1-*b*]thiazole (10a). The title compound was prepared according to General Procedure III (Scheme 1). Mobile phase petrol ether/EtOAc (10:50%). Yield: 458 mg (61%). ^1H NMR (500 MHz, DMSO- d_6): δ 8.10 (m, 2H), 7.92 (d, $J = 4.4$ Hz, 1H), 7.51–7.54 (m, 2H), 7.44 (d, $J = 4.4$ Hz, 1H), 0.31. ^{13}C NMR (101 MHz, DMSO): δ 149.86, 145.03, 133.72, 132.31, 128.81, 127.46, 119.32, 115.60, 106.93, 105.19, 93.51, –0.17. EI MS: calcd for $[\text{M} + \text{H}]$, 330.0414; found, 330.0416.

2-(4-Chlorophenyl)-3-((trimethylsilyl)ethynyl)imidazo[1,2-*a*]pyridine (11a). The title compound was prepared according to General Procedure III. Mobile phase petrol ether/EtOAc (10:50%). Yield: 625 mg (56%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.41 (dt, $J = 6.8, 1.2$ Hz, 1H), 8.28–8.21 (m, 2H), 7.70 (dt, $J = 9.0, 1.1$ Hz, 1H), 7.59–7.53 (m, 2H), 7.46 (ddd, $J = 9.0, 6.8, 1.3$ Hz, 1H), 7.14 (td, $J = 6.8, 1.2$ Hz, 1H), 0.34 (s, 9H). ^{13}C NMR (101 MHz, DMSO): δ 146.17, 144.70, 133.43, 132.04, 128.88, 128.28, 127.82, 125.82, 117.30, 114.28, 108.95, 104.06, 93.27, –0.10. HRMS: calcd for $[\text{M} + \text{H}]$, 325.09223; found, 325.09232.

2-(*p*-Tolyl)-3-((trimethylsilyl)ethynyl)imidazo[1,2-*a*]pyridine (11b). The title compound was prepared according to General Procedure III. Mobile phase petrol ether/EtOAc (10:50%). Yield: 117 mg (64%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.39 (d, $J = 6.8$ Hz, 1H), 8.16 (d, $J = 7.8$ Hz, 2H), 7.69 (d, $J = 9.0$ Hz, 1H), 7.43 (dd, $J = 8.9, 6.9$ Hz, 1H), 7.30 (d, $J = 7.9$ Hz, 2H), 7.12 (t, $J = 6.8$ Hz, 1H), 2.36 (s, 3H), 0.34 (s, 9H). ^{13}C NMR (101 MHz, DMSO): δ 147.99, 144.94, 138.80, 130.71, 129.64, 127.79, 126.95, 125.98, 117.43, 114.32, 108.68, 103.79, 94.13, 21.38, 0.24. HRMS: calcd for $[\text{M} + \text{H}]$, 305.14685; found, 305.14690.

2-(4-Ethylphenyl)-3-((trimethylsilyl)ethynyl)imidazo[1,2-*a*]pyridine (11c). The title compound was prepared according to General Procedure III. Mobile phase petrol ether/EtOAc (10:50%). Yield: 126 mg (63%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.39 (dt, $J = 6.8, 1.2$ Hz, 1H), 8.22–8.16 (m, 2H), 7.69 (dt, $J = 9.0, 1.1$ Hz, 1H), 7.42 (ddd, $J = 9.0, 6.8, 1.3$ Hz, 1H), 7.35–7.29 (m, 2H), 7.11 (td, $J = 6.8, 1.2$ Hz, 1H), 2.65 (q, $J = 7.6$ Hz, 2H), 1.20 (t, $J = 7.6$ Hz, 3H), 0.34 (s, 9H). ^{13}C NMR (101 MHz, DMSO): δ 147.66, 144.72, 144.65, 130.69, 128.12, 127.44, 126.73, 125.64, 117.14, 113.98, 108.33, 103.51, 93.84, 28.18, 15.57, –0.06. HRMS: calcd for $[\text{M} + \text{H}]$, 319.16250; found, 319.16255.

2-(4-Fluorophenyl)-3-((trimethylsilyl)ethynyl)imidazo[1,2-*a*]pyridine (11d). The title compound was prepared according to General Procedure III. Mobile phase petrol ether/EtOAc (10:50%). Yield: 250 mg (69%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.41 (dd, $J = 6.9, 1.4$ Hz, 1H), 8.32–8.25 (m, 2H), 7.70 (dd, $J = 9.1, 1.4$ Hz, 1H), 7.45 (ddd, $J = 9.0, 6.8, 1.3$ Hz, 1H), 7.38–7.31 (m, 2H), 7.13 (tt, $J = 6.9, 1.3$ Hz, 1H), 0.36–0.31 (m, 9H). ^{13}C NMR (101 MHz, DMSO): δ 162.46 (d, $J = 246.2$ Hz), 146.53, 144.70, 129.74 (d, $J = 3.0$ Hz), 128.79 (d, $J = 8.4$ Hz), 127.74, 125.81, 117.24, 117.24, 115.81 (d, $J = 21.6$ Hz), 114.21, 114.21, 108.58, 103.68, 93.46, –0.05. HRMS: calcd for $[\text{M} + \text{H}]$, 309.12178; found, 309.12195.

2-(4-(Trifluoromethyl)phenyl)-3-((trimethylsilyl)ethynyl)imidazo[1,2-*a*]pyridine (11e). The title compound was prepared according to General Procedure III. Mobile phase petrol ether/EtOAc (10:50%). Yield: 123 mg (54%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.45 (tt, $J = 6.8, 1.0$ Hz, 3H), 7.91–7.82 (m, 2H), 7.74 (dt, $J = 9.1, 1.1$ Hz, 1H), 7.49 (ddd, $J = 9.1, 6.8, 1.3$ Hz, 1H), 7.17 (td, $J = 6.8, 1.2$ Hz, 1H), 0.37 (s, 9H). ^{13}C NMR (101 MHz, DMSO): δ 145.80, 145.12, 137.36, 137.35, 129.25, 128.93, 128.38, 127.42, 126.23, 126.08 (q, $J = 3.9$ Hz), 117.80, 114.82, 109.61, 105.20, 93.24, 0.18. HRMS: calcd for $[\text{M} + \text{H}]$, 359.11859; found, 359.11862.

2-(2,4-Dichlorophenyl)-3-((trimethylsilyl)ethynyl)imidazo[1,2-*a*]pyridine (11f). The title compound was prepared according to General Procedure III. Mobile phase petrol ether/EtOAc (10:50%). Yield: 98 mg (64%). ^1H NMR (401 MHz, chloroform-*d*): δ 8.30 (dt, $J = 6.8, 1.2$ Hz, 1H), 7.64 (m, 2H), 7.54 (d, $J = 2.1$ Hz, 1H), 7.38–7.27 (m, 2H), 6.97 (td, $J = 6.8, 1.2$ Hz, 1H), 0.27 (s, 9H). ^{13}C NMR (101 MHz, CDCl₃): δ 146.79, 144.80, 135.05, 134.32, 133.22, 131.39, 130.12, 126.95, 126.52, 125.55, 117.98, 113.42, 108.14, 92.17, 0.04. HRMS: calcd for $[\text{M} + \text{H}]$, 321.14177; found, 321.14179.

2-(3,4-Dichlorophenyl)-3-((trimethylsilyl)ethynyl)imidazo[1,2-*a*]pyridine (11g). The title compound was prepared according to General Procedure III and used as crude in the next step without purification.

4-(3-((Trimethylsilyl)ethynyl)imidazo[1,2-*a*]pyridin-2-yl)-benzotrile (11h). The title compound was prepared according to General Procedure III. Mobile phase petrol ether/EtOAc (10:50%). Yield: 170 mg (65%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.44 (dt, $J = 6.8, 1.2$ Hz, 1H), 8.42–8.38 (m, 2H), 7.99–7.94 (m, 2H), 7.74 (dt, $J = 9.0, 1.1$ Hz, 1H), 7.49 (ddd, $J = 9.1, 6.8, 1.3$ Hz, 1H), 7.17 (td, $J =$

6.8, 1.1 Hz, 1H). ^{13}C NMR (101 MHz, DMSO): δ 145.16, 144.88, 137.51, 132.85, 128.24, 127.06, 125.97, 118.91, 117.56, 114.62, 110.97, 109.67, 105.24, 92.83, -0.14. HRMS: calcd for $[\text{M} + \text{H}]$, 316.12645; found, 316.12656.

2-(4-Methoxyphenyl)-3-((trimethylsilyl)ethynyl)imidazo[1,2-*a*]pyridine (11i). The title compound was prepared according to General Procedure III. Mobile phase petrol ether/EtOAc (10:50%). Yield: 98 mg (64%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.38 (dt, $J = 6.8, 1.2$ Hz, 1H), 8.26–8.18 (m, 2H), 7.67 (dt, $J = 9.0, 1.1$ Hz, 1H), 7.42 (ddd, $J = 9.0, 6.8, 1.3$ Hz, 1H), 7.11 (td, $J = 6.8, 1.2$ Hz, 1H), 7.08–7.03 (m, 2H), 3.82 (s, 3H), 0.34 (s, 9H). ^{13}C NMR (101 MHz, DMSO): δ 159.83, 147.63, 144.60, 128.13, 127.19, 125.75, 125.49, 116.93, 114.07, 113.72, 108.09, 102.92, 94.03, 55.32, -0.04, -0.06, -0.08. HRMS: calcd for $[\text{M} + \text{H}]$, 321.14177; found, 321.14179.

5-(1-(3,4-Dichlorobenzyl)-1H-1,2,3-triazol-4-yl)-6-(3,4-dichlorophenyl)imidazo[2,1-*b*]thiazole (12g). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (30:60%). Yield: 51 mg (92%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.57 (s, 1H), 8.01 (d, $J = 4.5$ Hz, 1H), 7.87 (d, $J = 2.0$ Hz, 1H), 7.71 (d, $J = 2.1$ Hz, 1H), 7.69–7.64 (m, 2H), 7.61 (d, $J = 8.4$ Hz, 1H), 7.40 (d, $J = 4.5$ Hz, 1H), 7.36 (dd, $J = 8.3, 2.1$ Hz, 1H), 5.70 (s, 2H). ^{13}C NMR (101 MHz, DMSO): δ 149.77, 141.34, 136.86, 136.81, 135.00, 131.52, 131.39, 131.24, 131.21, 130.87, 130.44, 130.14, 128.90, 128.65, 127.44, 123.91, 119.85, 114.89, 114.40, 51.92. HRMS: calcd for $[\text{M} + \text{H}]$, 493.95620; found, 493.95625.

3-(1-(3,4-Dichlorobenzyl)-1H-1,2,3-triazol-4-yl)-2-(*p*-tolyl)imidazo[1,2-*a*]pyridine (13b). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (20:70%). Yield: 140 mg (84%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.47 (dd, $J = 7.1, 1.5$ Hz, 1H), 8.45 (d, $J = 1.5$ Hz, 0H), 7.72–7.65 (m, 1H), 7.56–7.51 (m, 1H), 7.39–7.33 (m, 1H), 7.16 (d, $J = 7.8$ Hz, 1H), 6.97 (td, $J = 6.9, 1.7$ Hz, 0H), 5.74 (s, 1H), 2.32 (s, 2H). ^{13}C NMR (101 MHz, DMSO): δ 144.82, 143.87, 137.48, 137.11, 136.79, 131.51, 131.27, 131.23, 131.16, 130.25, 129.17, 128.54, 125.88, 125.74, 125.30, 116.96, 113.06, 110.91, 51.86, 21.01. HRMS: calcd for $[\text{M} + \text{H}]$, 434.09338; found, 434.09355.

3-(1-(3,4-Dichlorobenzyl)-1H-1,2,3-triazol-4-yl)-2-(4-ethylphenyl)imidazo[1,2-*a*]pyridine (13c). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (20:70%). Yield: 350 mg (70%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.48–8.43 (m, 2H), 7.72–7.65 (m, 3H), 7.59–7.54 (m, 2H), 7.38–7.33 (m, 2H), 7.20–7.15 (m, 2H), 6.97 (td, $J = 6.9, 1.2$ Hz, 1H), 5.74 (s, 2H), 2.61 (q, $J = 7.6$ Hz, 2H), 1.19 (t, $J = 7.6$ Hz, 3H). ^{13}C NMR (101 MHz, DMSO): δ 144.85, 143.88, 143.75, 137.12, 136.79, 131.52, 131.24, 131.19, 130.26, 128.56, 127.95, 127.90, 125.90, 125.86, 125.28, 116.98, 113.09, 110.92, 51.91, 28.10, 15.55. HRMS: calcd for $[\text{M} + \text{H}]$, 448.10903; found, 448.10914.

3-(1-(3,4-Dichlorobenzyl)-1H-1,2,3-triazol-4-yl)-2-(4-fluorophenyl)imidazo[1,2-*a*]pyridine (13d). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (30:80%). Yield: 211 mg (74%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.50 (s, 1H), 8.48 (dt, $J = 7.0, 1.2$ Hz, 1H), 7.75–7.65 (m, 6H), 7.41–7.32 (m, 2H), 7.24–7.15 (m, 2H), 6.99 (td, $J = 6.8, 1.2$ Hz, 1H), 5.74 (s, 2H). ^{13}C NMR (101 MHz, DMSO): δ 162.40 (d, $J = 245.1$ Hz), 145.14, 143.14, 137.30, 136.80, 131.80, 131.53, 131.47, 130.91 (d, $J = 3.1$ Hz), 130.62, 130.31 (d, $J = 8.3$ Hz), 128.81, 126.41, 126.10, 125.68, 117.33, 115.82 (d, $J = 21.5$ Hz), 113.53, 111.49, 52.21. HRMS: calcd for $[\text{M} + \text{H}]$, 438.06831; found, 438.06851.

3-(1-(3,4-Dichlorobenzyl)-1H-1,2,3-triazol-4-yl)-2-(4-(trifluoromethyl)phenyl)imidazo[1,2-*a*]pyridine (13e). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (30:80%). Yield: 143 mg (86%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.52 (dt, $J = 7.0, 1.2$ Hz, 1H), 8.38 (s, 1H), 8.04–7.96 (m, 2H), 7.57 (d, $J = 9.1$ Hz, 1H), 7.31–7.21 (m, 3H), 6.89 (td, $J = 6.8, 1.3$ Hz, 1H). ^{13}C NMR (101 MHz, DMSO): δ 162.01 (d, $J = 244.2$ Hz), 145.01, 143.63, 130.65, 127.68 (d, $J = 8.2$ Hz), 127.09, 125.23, 116.78, 115.77 (d, $J = 21.5$ Hz), 112.51, 109.12. HRMS: calcd for $[\text{M} + \text{H}]$, 488.06511; found, 488.06511.

3-(1-(3,4-Dichlorobenzyl)-1H-1,2,3-triazol-4-yl)-2-(2,4-dichlorophenyl)imidazo[1,2-*a*]pyridine (13f). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (40:80%). Yield: 120 mg (74%). ^1H NMR (401 MHz, DMSO- d_6): δ 9.11 (dt, $J = 7.0, 1.1$ Hz, 1H), 8.02 (s, 1H), 7.71 (dt, $J = 9.1, 1.2$ Hz, 1H), 7.68 (d, $J = 2.1$ Hz, 1H), 7.64 (d, $J = 8.3$ Hz, 1H), 7.58 (d, $J = 8.3$ Hz, 1H), 7.54 (d, $J = 2.1$ Hz, 1H), 7.52 (dd, $J = 8.3, 2.1$ Hz, 1H), 7.42 (ddd, $J = 9.1, 6.7, 1.3$ Hz, 1H), 7.25 (dd, $J = 8.3, 2.1$ Hz, 1H), 7.11 (td, $J = 6.8, 1.2$ Hz, 1H), 5.67 (s, 2H). ^{13}C NMR (101 MHz, DMSO): δ 144.69, 140.47, 137.27, 137.05, 134.19, 133.95, 133.64, 132.58, 131.46, 131.09, 130.00, 129.45, 128.30, 127.75, 126.19, 126.02, 123.26, 117.31, 113.72, 113.61, 51.61. HRMS: calcd for $[\text{M} + \text{H}]$, 417.06837; found, 417.06839.

3-(1-(3,4-Dichlorobenzyl)-1H-1,2,3-triazol-4-yl)-2-(3,4-dichlorophenyl)imidazo[1,2-*a*]pyridine (13g). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (30:60%). Yield: 56 mg (90%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.62 (s, 1H), 8.46 (dt, $J = 7.0, 1.2$ Hz, 1H), 7.85 (d, $J = 1.9$ Hz, 1H), 7.74–7.69 (m, 3H), 7.69–7.66 (m, 1H), 7.65 (d, $J = 2.0$ Hz, 1H), 7.62 (d, $J = 8.4$ Hz, 1H), 7.41 (ddd, $J = 9.1, 6.7, 1.3$ Hz, 1H), 7.37 (dd, $J = 8.3, 2.1$ Hz, 1H), 7.02 (td, $J = 6.8, 1.2$ Hz, 1H), 5.76 (s, 2H). ^{13}C NMR (101 MHz, DMSO): δ 144.96, 141.00, 136.95, 136.02, 134.78, 131.55, 131.40, 131.26, 130.91, 130.65, 130.39, 129.30, 128.58, 127.85, 126.58, 126.09, 125.46, 117.20, 113.57, 112.06, 51.97. HRMS: calcd for $[\text{M} + \text{H}]$, 487.99978; found, 487.99980.

4-(3-(1-(3,4-Dichlorobenzyl)-1H-1,2,3-triazol-4-yl)imidazo[1,2-*a*]pyridin-2-yl)benzotrile (13h). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (30:80%). Yield: 189 mg (89%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.59 (s, 1H), 8.44 (dt, $J = 7.0, 1.1$ Hz, 1H), 7.89–7.81 (m, 4H), 7.75–7.69 (m, 3H), 7.42 (ddd, $J = 9.1, 6.7, 1.3$ Hz, 1H), 7.36 (dd, $J = 8.3, 2.1$ Hz, 1H), 7.03 (td, $J = 6.8, 1.2$ Hz, 1H), 5.75 (s, 2H). ^{13}C NMR (101 MHz, DMSO): δ 178.80, 132.59, 131.51, 131.28, 131.21, 130.39, 128.58, 128.51, 126.70, 126.27, 125.49, 117.31, 113.66, 52.00. HRMS: calcd for $[\text{M} + \text{H}]$, 445.07298; found, 445.07299.

3-(1-(3,4-Dichlorobenzyl)-1H-1,2,3-triazol-4-yl)-2-(4-methoxyphenyl)imidazo[1,2-*a*]pyridine (13i). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (30:80%). Yield: 385 mg (92%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.46 (d, $J = 6.5$ Hz, 1H), 7.72–7.64 (m, 1H), 7.63–7.57 (m, 1H), 7.38–7.32 (m, 1H), 6.96 (td, $J = 6.8, 1.2$ Hz, 0H), 6.95–6.90 (m, 1H), 5.74 (s, 1H), 3.78 (s, 2H). ^{13}C NMR (101 MHz, DMSO): δ 159.28, 144.78, 143.76, 137.11, 136.88, 131.49, 131.23, 131.14, 130.23, 129.21, 128.52, 126.50, 125.77, 125.70, 125.23, 116.84, 114.00, 112.96, 110.44, 55.28, 51.88. HRMS: calcd for $[\text{M} + \text{H}]$, 450.08829; found, 450.08826.

5-(1-(4-Chlorophenyl)-1H-1,2,3-triazol-4-yl)-6-(4-chlorophenyl)imidazo[2,1-*b*]thiazole (14a). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (20:60%). Yield: 136 mg (88%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.45 (s, 1H), 8.03 (d, $J = 4.5$ Hz, 1H), 7.75–7.64 (m, 2H), 7.49–7.41 (m, 4H), 7.40–7.36 (m, 3H), 5.68 (s, 2H). ^{13}C NMR (101 MHz, DMSO): δ 149.57, 142.75, 137.05, 135.01, 133.28, 133.07, 132.35, 130.04, 129.25, 128.95, 128.69, 123.45, 119.96, 114.46, 113.94, 52.40. HRMS: calcd for $[\text{M} + \text{H}]$, 426.03415; found, 426.03423.

6-(4-Chlorophenyl)-5-(1-(3,4-dimethoxybenzyl)-1H-1,2,3-triazol-4-yl)imidazo[2,1-*b*]thiazole (14b). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (20:50%). Yield: 120 mg (86%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.40 (s, 1H), 8.01 (d, $J = 4.5$ Hz, 1H), 7.71–7.64 (m, 2H), 7.45–7.40 (m, 2H), 7.38 (d, $J = 4.5$ Hz, 1H), 7.05 (d, $J = 2.0$ Hz, 1H), 6.95 (d, $J = 8.2$ Hz, 1H), 6.90 (dd, $J = 2.0, 8.2$ Hz, 1H), 5.56 (s, 2H), 3.74 (s, 6H). ^{13}C NMR (101 MHz, DMSO): δ 149.80, 149.24, 142.95, 137.18, 133.60, 132.60, 129.53, 128.95, 128.45, 123.44, 121.13, 120.20, 114.75, 114.35, 112.48, 112.30, 56.00, 55.97, 53.45. HRMS: calcd for $[\text{M} + \text{H}]$, 452.09425; found, 452.09438.

6-(4-Chlorophenyl)-5-(1-(4-methoxybenzyl)-1H-1,2,3-triazol-4-yl)imidazo[2,1-*b*]thiazole (14c). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc

(20:50%). Yield: 121 mg (86%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.40 (s, 1H), 8.02 (d, J = 4.5 Hz, 1H), 7.76–7.61 (m, 3H), 7.45–7.41 (m, 3H), 7.38 (d, J = 4.5 Hz, 1H), 7.33 (d, J = 8.7 Hz, 2H), 7.01–6.88 (m, 3H), 5.59 (s, 2H), 3.75 (s, 3H). ^{13}C NMR (101 MHz, DMSO): δ 159.63, 149.81, 142.96, 137.20, 133.60, 132.60, 130.06, 129.52, 128.97, 128.22, 123.42, 120.23, 114.60, 114.34, 55.62, 53.06. HRMS: calcd for [M + H], 422.08369; found, 422.08372.

5-(1-Benzyl-1H-1,2,3-triazol-4-yl)-6-(4-chlorophenyl)imidazo[2,1-b]thiazole (14d). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (20:50%). Yield: 154 mg (87%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.45 (s, 1H), 8.03 (d, J = 4.5 Hz, 1H), 7.75–7.65 (m, 2H), 7.45–7.41 (m, 1H), 7.40–7.32 (m, 8H), 5.68 (s, 2H). ^{13}C NMR (101 MHz, DMSO): δ 149.84, 143.02, 137.27, 136.33, 133.59, 132.62, 129.54, 129.25, 128.97, 128.64, 128.34, 123.75, 120.24, 114.75, 114.29, 53.49. HRMS: calcd for [M + Na], 392.07312; found, 392.07318.

6-(4-Chlorophenyl)-5-(1-(pyridin-2-ylmethyl)-1H-1,2,3-triazol-4-yl)imidazo[2,1-b]thiazole (14e). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (20:50%). Yield: 143 mg (80%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.57 (ddd, J = 4.8, 1.9, 1.0 Hz, 1H), 8.46 (s, 1H), 8.05 (d, J = 4.5 Hz, 1H), 7.85 (td, J = 7.7, 1.8 Hz, 1H), 7.77–7.69 (m, 2H), 7.47–7.42 (m, 2H), 7.41–7.35 (m, 2H), 7.33 (dd, J = 7.8, 1.1 Hz, 1H), 5.81 (s, 2H). ^{13}C NMR (101 MHz, DMSO): δ 154.95, 149.61, 149.53, 142.69, 137.55, 136.85, 133.31, 132.31, 129.26, 128.67, 124.16, 123.44, 122.24, 119.94, 114.47, 114.02, 54.64. HRMS: calcd for [M + H], 393.06837; found, 393.06845.

4-(4-(6-(4-Chlorophenyl)imidazo[2,1-b]thiazol-5-yl)-1H-1,2,3-triazol-1-yl)methyl-benzonitrile (14f). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (20:70%). Yield: 87 mg (87%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.49 (s, 1H), 8.05 (d, J = 4.5 Hz, 1H), 7.92–7.79 (m, 2H), 7.73–7.66 (m, 2H), 7.52–7.47 (m, 2H), 7.47–7.42 (m, 2H), 7.39 (d, J = 4.5 Hz, 1H), 5.80 (s, 2H). ^{13}C NMR (101 MHz, DMSO): δ 149.61, 142.82, 141.48, 137.13, 133.27, 132.93, 132.37, 129.28, 128.83, 128.72, 123.76, 120.00, 118.72, 114.49, 113.87, 111.11, 52.56. HRMS: calcd for [M + H], 487.99978; found, 487.99993.

3-(1-(4-Chlorobenzyl)-1H-1,2,3-triazol-4-yl)-2-(4-chlorophenyl)imidazo[1,2-a]pyridine (15a). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (20:50%). Yield: 145 mg (92%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.49 (s, 1H), 8.45 (dt, J = 7.0, 1.2 Hz, 1H), 7.71–7.66 (m, 3H), 7.51–7.47 (m, 2H), 7.45–7.36 (m, 5H), 7.00 (td, J = 6.8, 1.2 Hz, 1H), 5.73 (s, 2H). ^{13}C NMR (101 MHz, DMSO): δ 149.57, 142.75, 137.05, 135.01, 133.28, 133.07, 132.35, 130.04, 129.25, 128.95, 128.69, 123.45, 119.96, 114.46, 113.94, 52.40. HRMS: calcd for [M + H], 420.07773; found, 420.07765.

2-(4-Chlorophenyl)-3-(1-(3,4-dimethoxybenzyl)-1H-1,2,3-triazol-4-yl)imidazo[1,2-a]pyridine (15b). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (30:60%). Yield: 111 mg (81%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.47–8.42 (m, 2H), 7.71–7.66 (m, 3H), 7.43–7.36 (m, 3H), 7.05 (d, J = 1.9 Hz, 1H), 7.03–6.95 (m, 2H), 6.91 (dd, J = 2.0, 8.2 Hz, 1H), 5.62 (s, 2H), 3.75 (d, J = 3.8 Hz, 6H). ^{13}C NMR (101 MHz, DMSO): δ 149.27, 149.24, 145.12, 142.55, 136.41, 133.26, 133.11, 129.92, 128.94, 128.51, 126.60, 125.78, 125.63, 121.03, 117.34, 113.67, 112.40, 112.33, 112.05, 56.00, 55.97, 53.53. HRMS: calcd for [M + Na], 446.13783; found, 446.13786.

2-(4-Chlorophenyl)-3-(1-(4-methoxybenzyl)-1H-1,2,3-triazol-4-yl)imidazo[1,2-a]pyridine (15c). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (20:50%). Yield: 96 mg (75%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.50–8.35 (m, 2H), 7.71–7.66 (m, 3H), 7.44–7.31 (m, 5H), 7.02–6.93 (m, 1H), 5.63 (s, 2H), 3.76 (s, 3H). ^{13}C NMR (101 MHz, DMSO): δ 159.64, 145.18, 142.68, 136.46, 133.33, 133.07, 129.96, 129.90, 128.95, 128.27, 126.51, 125.76, 125.62, 117.40, 114.64, 113.63, 112.01, 55.63, 53.16. HRMS: calcd for [M + H], 416.12726; found, 416.12730.

3-(1-Benzyl-1H-1,2,3-triazol-4-yl)-2-(4-chlorophenyl)imidazo[1,2-a]pyridine (15d). The title compound was prepared according to

General Procedure IV. Mobile phase petrol ether/EtOAc (20:50%). Yield: 139 mg (78%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.50 (s, 1H), 8.45 (dt, J = 1.1, 6.9 Hz, 1H), 7.69 (m, 3H), 7.44–7.37 (m, 8H), 7.00 (td, J = 1.2, 6.8 Hz, 2H), 5.73 (s, 2H). ^{13}C NMR (101 MHz, DMSO): δ 145.20, 142.73, 136.53, 136.40, 133.32, 133.09, 129.92, 129.29, 128.95, 128.65, 128.24, 126.54, 126.12, 125.63, 117.40, 113.65, 111.95, 53.57. HRMS: calcd for [M + H], 386.11670; found, 386.11681.

2-(4-Chlorophenyl)-3-(1-(pyridin-2-ylmethyl)-1H-1,2,3-triazol-4-yl)imidazo[1,2-a]pyridine (15e). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (20:50%). Yield: 147 mg (83%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.58 (dt, J = 4.7, 1.4 Hz, 1H), 8.52 (s, 1H), 8.47 (dt, J = 6.9, 1.2 Hz, 1H), 7.85 (td, J = 7.7, 1.8 Hz, 1H), 7.77–7.71 (m, 2H), 7.68 (dt, J = 9.1, 1.2 Hz, 1H), 7.44–7.39 (m, 2H), 7.39–7.33 (m, 3H), 6.99 (td, J = 6.8, 1.2 Hz, 1H), 5.87 (s, 2H). ^{13}C NMR (101 MHz, DMSO): δ 155.00, 149.65, 144.91, 142.41, 137.55, 136.14, 133.05, 132.80, 129.64, 128.63, 126.55, 126.16, 125.31, 123.45, 122.21, 117.10, 113.29, 111.70, 54.75. HRMS: calcd for [M + H], 387.11195; found, 387.11198.

4-((4-(2-(4-Chlorophenyl)imidazo[1,2-a]pyridin-3-yl)-1H-1,2,3-triazol-1-yl)methyl)benzonitrile (15f). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (40:80%). Yield: 148 mg (82%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.53 (s, 1H), 8.48 (dt, J = 6.9, 1.2 Hz, 1H), 7.94–7.88 (m, 2H), 7.72–7.67 (m, 3H), 7.53–7.48 (m, 2H), 7.47–7.42 (m, 2H), 7.39 (ddd, J = 9.1, 6.7, 1.3 Hz, 1H), 7.01 (td, J = 6.8, 1.2 Hz, 1H), 5.85 (s, 2H). ^{13}C NMR (101 MHz, DMSO): δ 145.23, 142.82, 141.84, 136.69, 133.31, 133.27, 133.14, 129.94, 129.06, 126.57, 126.42, 125.70, 119.02, 117.41, 113.65, 111.82, 111.44, 52.96. HRMS: calcd for [M + H], 411.11195; found, 411.11215.

2-(4-Chlorophenyl)-3-(1-(4-(methylthio)benzyl)-1H-1,2,3-triazol-4-yl)imidazo[1,2-a]pyridine (15g). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (40:100%). Yield: 72 mg (80%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.46 (s, 1H), 8.44 (dt, J = 6.9, 1.2 Hz, 1H), 7.72–7.65 (m, 3H), 7.44–7.35 (m, 3H), 7.33–7.27 (m, 3H), 6.99 (td, J = 6.8, 1.2 Hz, 1H), 5.67 (s, 2H), 2.47 (s, 3H). ^{13}C NMR (101 MHz, DMSO): δ 144.91, 142.43, 138.60, 136.24, 133.02, 132.81, 132.52, 129.62, 128.76, 128.67, 126.28, 126.24, 125.67, 125.34, 117.10, 113.35, 111.67, 52.88, 14.76. HRMS: calcd for [M + H], 431.09715; found, 431.09723.

2-(4-Chlorophenyl)-3-(1-(1-phenylethyl)-1H-1,2,3-triazol-4-yl)imidazo[1,2-a]pyridine (15h). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (40:100%). Yield: 72 mg (80%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.58 (s, 1H), 8.41 (dt, J = 6.9, 1.2 Hz, 1H), 7.69 (dd, J = 8.4, 1.7 Hz, 3H), 7.44–7.32 (m, 9H), 6.99 (td, J = 6.8, 1.2 Hz, 1H), 6.08 (q, J = 7.1 Hz, 1H), 1.96 (d, J = 7.1 Hz, 3H). ^{13}C NMR (101 MHz, DMSO): δ 144.88, 142.41, 141.14, 136.02, 133.00, 132.75, 129.52, 128.95, 128.58, 128.24, 126.43, 126.20, 125.28, 124.60, 117.08, 113.30, 111.70, 59.89, 21.25. HRMS: calcd for [M + H], 400.13235; found, 400.13214.

2-(4-Chlorophenyl)-3-(1-(4-(methylsulfonyl)benzyl)-1H-1,2,3-triazol-4-yl)imidazo[1,2-a]pyridine (15i). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (40:100%). Yield: 613 mg (92%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.52 (dt, J = 6.8, 1.1 Hz, 1H), 8.07 (s, 1H), 7.98 (m, 2H), 7.72 (dt, J = 9.1, 1.1 Hz, 1H), 7.69 (m, 2H), 7.60 (m, 2H), 7.45 (m, 2H), 7.42 (ddd, J = 9.1, 6.8, 1.1 Hz, 1H), 7.04 (td, J = 6.8, 1.1 Hz, 1H), 5.97 (s, 2H), 3.23 (s, 3H). ^{13}C NMR (101 MHz, DMSO- d_6): δ 145.13, 143.06, 141.54, 140.72, 137.61, 135.29, 133.03, 132.90, 129.83, 128.91, 128.72, 127.68, 126.50, 125.39, 117.20, 113.61, 111.13, 57.56, 43.65. HRMS: calcd for [M - H], 462.07970; found, 462.07935.

2-(4-Chlorophenyl)-3-(1-(4-(pyrrolidin-1-ylsulfonyl)benzyl)-1H-1,2,3-triazol-4-yl)imidazo[1,2-a]pyridine (15j). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (40:100%). Yield: 144 mg (93%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.49 (d, J = 6.6 Hz, 1H), 8.03 (s, 1H), 7.85 (m, 2H),

7.67–7.73 (m, 3H), 7.57 (m, 2H), 7.43 (m, 2H), 7.40 (m, 1H), 7.02 (t, $J = 6.6$ Hz, 1H), 5.94 (s, 2H), 3.17 (m, 4H), 1.66 (m, 4H). ^{13}C NMR (101 MHz, DMSO- d_6): δ 144.97, 142.99, 140.4, 137.50, 136.50, 134.97, 132–132.89 (m), 129.62, 128.55, 127.63, 126.13, 125.07, 116.98, 113.26, 110.97, 57.43, 47.71, 24.65. HRMS: calcd for $[\text{M} - \text{H}]$, 517.12190; found, 517.12134.

2-(4-Chlorophenyl)-3-(1-(4-nitrobenzyl)-1H-1,2,3-triazol-4-yl)imidazo[1,2-*a*]pyridine (15k). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (40:100%). Yield: 253 mg (88%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.56 (s, 1H), 8.49 (dt, $J = 6.9, 1.2$ Hz, 1H), 8.31–8.21 (m, 2H), 7.73–7.67 (m, 3H), 7.62–7.54 (m, 2H), 7.47–7.41 (m, 2H), 7.39 (ddd, $J = 9.0, 6.7, 1.3$ Hz, 1H), 7.00 (td, $J = 6.8, 1.3$ Hz, 1H), 5.92 (s, 2H). ^{13}C NMR (101 MHz, DMSO): δ 147.73, 145.24, 143.78, 142.85, 136.75, 133.30, 133.14, 129.95, 129.39, 129.00, 126.55, 126.44, 125.70, 124.43, 117.40, 113.64, 111.80, 52.70. HRMS: calcd for $[\text{M} - \text{H}]$, 430.09; found, 429.555.

4-((4-(2-(4-Chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-1H-1,2,3-triazol-1-yl)methyl)aniline (15l). 2-(4-Chlorophenyl)-3-(1-(4-nitrobenzyl)-1H-1,2,3-triazol-4-yl)imidazo[1,2-*a*]pyridine (15k) was dissolved in MeOH and AcOH (7 equiv), and Fe (3.5 equiv) was added. The reaction mixture was stirred at reflux until the completion of the reaction. After cooling to 25 °C, the mixture was extracted with EtOAc, washed with NaHCO_3 solution, and the organic phase was dried over sodium sulfate and evaporated. The residue was purified by flash column chromatography, mobile phase cyclohexane/EtOAc (40–100%). Yield: 123 mg (78%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.41 (dt, $J = 7.0, 1.3$ Hz, 1H), 8.39 (s, 1H), 7.72–7.65 (m, 3H), 7.43–7.32 (m, 3H), 7.10–7.04 (m, 2H), 6.98 (td, $J = 6.8, 1.3$ Hz, 1H), 6.59–6.53 (m, 2H), 5.48 (s, 2H), 5.20 (s, 2H). ^{13}C NMR (101 MHz, DMSO): δ 149.31, 145.14, 142.57, 136.35, 133.35, 133.06, 129.86, 129.63, 128.90, 126.44, 125.58, 125.47, 122.83, 117.37, 114.23, 113.57, 112.10, 53.72. HRMS: calcd for $[\text{M} + \text{H}]$, 401.12760; found, 401.12735.

***N*-(4-((4-(2-(4-Chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-1H-1,2,3-triazol-1-yl)methyl)phenyl)acetamide (15m).** 4-((4-(2-(4-Chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-1H-1,2,3-triazol-1-yl)methyl)aniline was dissolved in dioxane, and Ac_2O (1.5 equiv) was added followed by pyridine (1.5 equiv). The reaction mixture was stirred at 25 °C overnight. After the completion of the reaction, the mixture was evaporated, diluted with EtOAc, and washed with water. The organic phase was dried over sodium sulfate and evaporated. The residue was purified by flash column chromatography. Mobile phase cyclohexane/EtOAc (40–100%). Yield: 63 mg (92%). ^1H NMR (401 MHz, DMSO- d_6): δ 10.02 (s, 1H), 8.45 (s, 1H), 8.45–8.42 (m, 1H), 7.72–7.66 (m, 3H), 7.63–7.58 (m, 2H), 7.39 (s, 2H), 7.39–7.34 (m, 1H), 7.33–7.28 (m, 2H), 6.99 (td, $J = 6.8, 1.2$ Hz, 1H), 5.65 (s, 2H), 2.04 (s, 3H). ^{13}C NMR (101 MHz, DMSO): δ 168.54, 144.88, 142.37, 139.48, 136.20, 133.04, 132.78, 130.37, 129.59, 128.65, 128.62, 126.18, 125.58, 125.34, 119.31, 117.08, 113.30, 111.69, 66.52, 53.00, 24.17. HRMS: calcd for $[\text{M} + \text{H}]$, 443.13816; found, 443.13773.

5-(6-(4-Chlorophenyl)imidazo[2,1-*b*]thiazol-5-yl)-3-(3,4-dichlorobenzyl)-1,2,4-oxadiazole (16A). 6-(4-Chlorophenyl)imidazo[2,1-*b*]thiazole-5-carboxylic acid was dissolved in dry DMF, degassed, and refilled with argon. EDC (1 equiv) and HOBt (1 equiv) were added in one portion, and the mixture was stirred at 25 °C for 30 min. Then, a solution of (*E*)-2-(3,4-dichlorophenyl)-*N'*-hydroxyacetimidamide in dry DMF was added, and the reaction mixture was stirred at 80 °C overnight. After cooling to 25 °C, the mixture was diluted with EtOAc, washed with NaHCO_3 solution, and water, and the organic phase was dried over sodium sulfate. The product was isolated by flash column chromatography, mobile phase petrol ether/EtOAc (50:70%). Yield: 250 mg (52%). ^1H NMR (500 MHz, DMSO- d_6): δ 8.33 (d, $J = 4.4$ Hz, 1H), 7.94–7.91 (m, 2H), 7.71 (d, $J = 2.1$ Hz, 1H), 7.63 (d, $J = 4.0$ Hz, 1H), 7.62 (s, 1H), 7.57–7.53 (m, 2H), 7.42–7.39 (m, 1H), 4.23 (s, 2H). ^{13}C NMR (126 MHz, DMSO): δ 168.78, 167.94, 153.92, 150.51, 137.29, 134.31, 132.08, 131.65, 131.45, 131.25, 131.09, 130.18, 130.10, 128.70, 121.58, 117.20,

109.75, 30.64. HRMS: calcd for $[\text{M} + \text{H}]$, 460.97919; found, 460.97921.

5-(2-(4-Chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-3-(3,4-dichlorobenzyl)-1,2,4-oxadiazole (16B). 2-(4-Chlorophenyl)imidazo[1,2-*a*]pyridine-3-carboxylic acid was dissolved in dry DMF, degassed, and refilled with argon. EDC (1 equiv) and HOBt (1 equiv) were added in one portion, and the mixture was stirred at 25 °C for 30 min. Then, a solution of (*E*)-2-(3,4-dichlorophenyl)-*N'*-hydroxyacetimidamide in dry DMF was added, and the reaction mixture was stirred at 80 °C overnight. After cooling to 25 °C, the mixture was diluted with EtOAc and washed with NaHCO_3 solution and water, and the organic phase was dried over sodium sulfate. The product was isolated by flash column chromatography, mobile phase petrol ether/EtOAc (50:70%). Yield: 252 mg (50%). ^1H NMR (401 MHz, DMSO- d_6): δ 9.41 (dt, $J = 6.9, 1.2$ Hz, 1H), 7.90 (dt, $J = 9.0, 1.2$ Hz, 1H), 7.88–7.84 (m, 2H), 7.71 (d, $J = 2.0$ Hz, 1H), 7.67 (ddd, $J = 9.0, 6.9, 1.3$ Hz, 1H), 7.63 (d, $J = 8.3$ Hz, 1H), 7.60–7.54 (m, 2H), 7.42 (dd, $J = 8.3, 2.1$ Hz, 1H), 7.36 (td, $J = 7.0, 1.3$ Hz, 1H), 4.27 (s, 2H). ^{13}C NMR (101 MHz, DMSO): δ 168.27, 167.90, 150.18, 147.19, 137.02, 134.29, 132.16, 131.54, 131.33, 131.19, 130.82, 129.90, 129.77, 129.38, 128.39, 128.01, 117.58, 115.47, 30.46. HRMS: calcd for $[\text{M} + \text{H}]$, 455.02277; found, 455.02288.

2-(2-(4-Chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-5-(3,4-dichlorobenzyl)-1,3,4-oxadiazole (17). 2-(4-Chlorophenyl)-*N'*-(2-(3,4-dichlorophenyl)acetyl)imidazo[1,2-*a*]pyridine-3-carbohydrazide was dissolved in dry DCM, and tosyl chloride (1.5 equiv) was added, followed by TEA (3 equiv) at 0 °C. The reaction mixture was stirred at 25 °C overnight. The mixture was diluted with water and extracted with EtOAc, and the organic phase was washed with a saturated NaHCO_3 solution. The organic phase was dried over sodium sulfate. The product was isolated by flash column and RP-flash column chromatography; mobile phase hexane/EtOAc (30:60%), and ($\text{H}_2\text{O}/\text{CH}_3\text{CN}$ 10:80%). Yield: 16 mg (6%). ^1H NMR (500 MHz, DMSO- d_6): δ 9.30 (dt, $J = 7.0, 1.2$ Hz, 1H), 7.82 (dt, $J = 9.0, 1.2$ Hz, 1H), 7.79–7.75 (m, 2H), 7.61–7.57 (m, 3H), 7.43–7.39 (m, 2H), 7.31–7.25 (m, 2H), 4.35 (s, 2H). ^{13}C NMR (126 MHz, DMSO): δ 163.81, 158.08, 147.83, 146.98, 135.75, 134.31, 132.41, 131.77, 131.62, 131.32, 131.12, 130.70, 129.92, 128.65, 128.52, 127.96, 117.70, 115.12, 106.70, 30.17. HRMS: calcd for $[\text{M} + \text{H}]$, 455.02277; found, 455.02282.

2-(2-(4-Chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-5-(3,4-dichlorobenzyl)-1,3,4-thiadiazole (18). A round-bottom flask was charged with 2-(4-chlorophenyl)-*N'*-(2-(3,4-dichlorophenyl)acetyl)imidazo[1,2-*a*]pyridine-3-carbohydrazide, degassed, and refilled with argon. Dry toluene was added, and the mixture was degassed once more and refilled with argon. Lawesson's reagent (3 equiv) was added, and the mixture was stirred at 100 °C overnight. After cooling to 25 °C, the mixture was diluted with water and extracted with EtOAc. An organic phase was dried over sodium sulfate and evaporated. A residue was purified by flash column chromatography. Mobile phase $\text{H}_2\text{O}/\text{MeOH}$ (30:100%). Yield: 44 mg (21%). ^1H NMR (500 MHz, DMSO- d_6): δ 9.35 (dt, $J = 7.0, 1.2$ Hz, 1H), 7.85 (dt, $J = 9.0, 1.2$ Hz, 1H), 7.78–7.73 (m, 2H), 7.65–7.57 (m, 3H), 7.42–7.38 (m, 2H), 7.31 (ddd, $J = 8.2, 4.5, 1.7$ Hz, 2H), 4.35 (s, 2H). ^{13}C NMR (126 MHz, DMSO): δ 163.75, 158.07, 147.63, 146.86, 135.84, 134.23, 132.26, 131.72, 131.62, 131.43, 131.16, 130.56, 130.10, 128.84, 128.55, 128.08, 117.70, 115.31, 106.65, 29.99. HRMS: calcd for $[\text{M} + \text{Na}]$, 477.00472; found, 477.00467.

6-(4-Chlorophenyl)-5-(2-(3,4-dichlorobenzyl)thiazol-4-yl)imidazo[2,1-*b*]thiazole (19A). 2-Chloro-1-(6-(4-chlorophenyl)imidazo[2,1-*b*]thiazol-5-yl)ethan-1-one (34) was dissolved in EtOH and 2-(3,4-dichlorophenyl)ethanethioamide (1.5 equiv) was added. A reaction mixture was stirred at reflux overnight. After cooling to 25 °C, the mixture was purified by RP-flash column chromatography. Mobile phase $\text{H}_2\text{O}/\text{CH}_3\text{CN}$ (20:80%). Yield: 213 mg (61%). ^1H NMR (401 MHz, DMSO- d_6): δ 8.01 (d, $J = 4.5$ Hz, 1H), 7.71 (d, $J = 2.0$ Hz, 1H), 7.68–7.61 (m, 4H), 7.42–7.36 (m, 4H), 4.46 (s, 2H). ^{13}C NMR (101 MHz, DMSO): δ 169.35, 149.37, 143.99, 143.09, 139.54, 133.79, 132.56, 131.60, 131.53, 131.24, 130.15, 130.04,

129.84, 128.83, 120.43, 118.35, 117.72, 114.49, 37.55. HRMS: calcd for [M + H], 475.96110; found, 475.96130.

4-(2-(4-Chlorophenyl)imidazo[1,2-a]pyridin-3-yl)-2-(3,4-dichlorobenzyl)thiazole (19B). 2-Chloro-1-(2-(4-chlorophenyl)imidazo[1,2-a]pyridin-3-yl)ethan-1-one (35) was dissolved in EtOH, and 2-(3,4-dichlorophenyl)ethanethioamide (1.5 equiv) was added. A reaction mixture was stirred at reflux overnight. After cooling to 25 °C, the mixture was purified by RP-flash column chromatography. Mobile phase: H₂O/CH₃CN (20:80%). Yield: 112 mg (62%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 8.42 (d, *J* = 6.9 Hz, 1H), 7.83 (s, 1H), 7.74 (d, *J* = 2.1 Hz, 1H), 7.67 (q, *J* = 8.1 Hz, 4H), 7.45–7.35 (m, 4H), 6.99 (t, *J* = 6.9 Hz, 1H), 4.51 (s, 2H). ¹³C NMR (101 MHz, DMSO): δ 169.57, 144.48, 143.11, 142.15, 139.27, 133.16, 132.66, 131.29, 131.25, 130.96, 129.86, 129.73, 129.68, 128.56, 126.10, 125.41, 121.39, 117.09, 115.60, 113.16, 37.34. HRMS: calcd for [M + H], 470.00468; found, 470.00488.

3-Benzyl-5-(2-(4-chlorophenyl)imidazo[1,2-a]pyridin-3-yl)-isoxazole (20). Phenylacetaldehyde (0.045 mL, 1 equiv) was dissolved in an H₂O/*t*-BuOH mixture (4 mL) and NH₂OH·HCl (30 mg, 1 equiv), followed by NaOH (20 mg, 1 equiv) addition. The reaction mixture was stirred at rt for 3 h; then, 2-(4-chlorophenyl)-3-ethynylimidazo[1,2-a]pyridine 33 (100 mg, 0.39 mmol), chloramine T (120 mg, 1 equiv), and CuI (8 mg, 10 mol %) were added; and the reaction mixture was stirred overnight. The reaction mixture was diluted with water and extracted with EtOAc. Combined organic phases were dried over sodium sulfate and evaporated. The residue was purified by RP-flash column chromatography. Mobile phase H₂O/CH₃CN (10:100%). Yield: 99 mg (66%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 8.48 (dt, *J* = 6.9, 1.1 Hz, 1H), 7.76 (dt, *J* = 9.1, 1.2 Hz, 1H), 7.68–7.63 (m, 2H), 7.51–7.46 (m, 3H), 7.35 (m, 4H), 7.26 (ddd, *J* = 6.9, 4.6, 3.1 Hz, 1H), 7.12 (td, *J* = 6.9, 1.2 Hz, 1H), 6.82 (s, 1H), 4.11 (s, 2H). ¹³C NMR (101 MHz, DMSO): δ 163.90, 160.02, 146.13, 145.18, 137.73, 133.78, 132.59, 130.33, 129.23, 129.09, 129.04, 127.82, 127.16, 126.15, 117.66, 114.67, 105.32, 31.94. Anal. (C₂₃H₁₆ClN₃O·0.75H₂O): C, H, N. HRMS: calcd for [M + H], 386.10547; found, 386.10564. EA: C, 69.17; H, 4.42; N, 10.52. Found: C, 68.94; H, 4.11; N, 10.41.

2-(4-Chlorophenyl)-3-(1-(3,4-dichlorobenzyl)-1H-pyrrol-2-yl)-imidazo[1,2-a]pyridine (21). 2-(4-Chlorophenyl)-3-(1H-pyrrol-2-yl)-imidazo[1,2-a]pyridine 35 (76 mg, 0.26 mmol) was dissolved in dry DMF (2 mL), degassed, and refilled with argon. NaH (10 mg, 1.3 equiv, 60% in mineral oil) was added, and the mixture was stirred at 25 °C for 30 min. 1,2-Dichloro-4-(chloromethyl)benzene (0.05 mL, 1.3 equiv) was added, and the reaction mixture was stirred at 25 °C overnight. Then, the mixture was diluted with water and extracted with EtOAc. Combined organic phases were dried over sodium sulfate and evaporated. The residue was purified by flash column chromatography, mobile phase petrol ether/EtOAc (15:50%). Yield: 25 mg (21%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 7.57 (m, 3H), 7.47 (d, *J* = 6.8 Hz, 1H), 7.39 (dd, *J* = 2.7, 1.7 Hz, 1H), 7.38–7.33 (m, 2H), 7.27 (ddd, *J* = 9.0, 6.7, 1.2 Hz, 1H), 7.17 (d, *J* = 8.2 Hz, 1H), 6.74 (td, *J* = 6.8, 1.1 Hz, 1H), 6.64 (d, *J* = 2.0 Hz, 1H), 6.58 (dd, *J* = 8.3, 2.0 Hz, 1H), 6.51 (dd, *J* = 3.6, 1.7 Hz, 1H), 6.43–6.38 (m, 1H), 4.80 (d, *J* = 15.4 Hz, H), 4.53 (d, *J* = 15.4 Hz, 1H). ¹³C NMR (101 MHz, DMSO): δ 144.60, 142.39, 138.69, 132.84, 132.59, 130.86, 130.34, 129.80, 128.70, 128.62, 127.99, 127.04, 126.00, 125.48, 123.94, 118.69, 116.79, 113.98, 112.70, 112.19, 109.16, 49.78. HRMS: calcd for [M + H], 452.04826; found, 452.04842.

2-(4-Chlorophenyl)-3-(1-(3,4-dichlorobenzyl)-1H-pyrazol-4-yl)-imidazo[1,2-a]pyridine (22). 2-(4-Chlorophenyl)-3-(1H-pyrazol-4-yl)imidazo[1,2-a]pyridine 36 (30 mg, 0.1 mmol) was dissolved in CH₃CN (2 mL), and the solution was degassed and refilled with argon. K₂CO₃ (15 mg, 1.1 equiv) was added, followed by the addition of 1,2-dichloro-4-(chloromethyl)benzene (0.02 mL, 1 equiv). The reaction mixture was refluxed overnight under an argon atmosphere. The product was isolated from preparative TLC; mobile phase: petrol ether/EtOAc 3:2. Yield: 15 mg (32%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 8.26 (s, 0H), 8.09 (dd, *J* = 6.9, 1.5 Hz, 0H), 7.80 (s, 0H), 7.73–7.68 (m, 1H), 7.68–7.60 (m, 1H), 7.58 (d, *J* = 2.1 Hz, 1H), 7.41 (dd, *J* = 8.5, 1.7 Hz, 1H), 7.36–7.24 (m, 1H), 6.97–6.90 (m, 1H), 5.49 (s,

1H). ¹³C NMR (101 MHz, DMSO): δ 144.74, 141.23, 140.90, 139.00, 133.78, 132.57, 132.37, 131.36, 130.87, 130.02, 129.79, 129.36, 128.82, 128.38, 125.81, 124.78, 117.28, 113.42, 113.20, 109.08, 54.16. HRMS: calcd for [M + H], 453.04351; found, 453.04356.

2-(4-Chlorophenyl)-3-(1-(3,4-dichlorophenyl)-1H-1,2,3-triazol-4-yl)imidazo[1,2-a]pyridine (23). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (30:100%). Yield: 225 mg (73%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 9.31 (s, 1H), 8.49 (dt, *J* = 6.9, 1.1 Hz, 2H), 8.37 (d, *J* = 2.5 Hz, 1H), 8.07 (dd, *J* = 8.8, 2.5 Hz, 1H), 7.95 (d, *J* = 8.8 Hz, 1H), 7.85–7.78 (m, 3H), 7.73 (dt, *J* = 9.1, 1.0 Hz, 2H), 7.48–7.44 (m, 3H), 7.44–7.39 (m, 1H), 7.02 (td, *J* = 6.8, 1.2 Hz, 1H). ¹³C NMR (101 MHz, DMSO): δ 145.10, 142.84, 136.98, 136.19, 132.88, 132.84, 132.55, 132.06, 131.49, 129.65, 128.73, 126.53, 125.45, 124.34, 122.30, 120.57, 117.14, 113.39, 110.89. HRMS: calcd for [M + H], 440.02311; found, 440.02321.

2-(4-Chlorophenyl)-3-(1-(3,4-dichlorophenethyl)-1H-1,2,3-triazol-4-yl)imidazo[1,2-a]pyridine (24). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (30:100%). Yield: 138 mg (95%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 8.33 (s, 0H), 8.25 (d, *J* = 6.9 Hz, 1H), 7.69 (d, *J* = 9.0 Hz, 1H), 7.64 (d, *J* = 8.5 Hz, 1H), 7.58–7.49 (m, 1H), 7.39 (d, *J* = 8.5 Hz, 2H), 7.21–7.13 (m, 1H), 6.99 (t, *J* = 6.6 Hz, 1H), 4.78 (t, *J* = 6.7 Hz, 1H), 3.26 (t, *J* = 6.7 Hz, 1H). ¹³C NMR (101 MHz, DMSO): δ 144.84, 142.21, 139.06, 135.60, 132.98, 132.77, 131.08, 131.05, 130.64, 129.49, 129.46, 129.40, 128.61, 126.20, 125.77, 125.00, 117.13, 113.27, 111.64, 50.51, 34.70. HRMS: calcd for [M + H], 468.05441; found, 468.05450.

Ethyl 6-(4-Chlorophenyl)imidazo[2,1-*b*]thiazole-5-carboxylate (25). 2-Aminothiazole (3 equiv) was dissolved in CH₃CN, and ethyl 3-(4-chlorophenyl)-3-oxopropanoate (1 equiv) was added as a solution in CH₃CN, followed by the addition of CBr₄ (2 equiv). The reaction mixture was stirred at 80 °C overnight. After the completion of the reaction, the mixture was evaporated to a minimal volume, diluted with water, and extracted with EtOAc. Combined organic phases were dried over sodium sulfate and purified by flash column chromatography. Mobile phase petrol ether/EtOAc (25:50%). Yield: 365 mg (82%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 8.22 (d, *J* = 4.5 Hz, 1H), 7.89–7.85 (m, 2H), 7.54 (d, *J* = 4.5 Hz, 2H), 7.52–7.48 (m, 2H), 4.30 (q, *J* = 7.1 Hz, 2H), 1.27 (t, *J* = 7.1 Hz, 3H). ¹³C NMR (101 MHz, DMSO): δ 159.18, 152.65, 151.43, 133.50, 132.49, 131.49, 127.90, 121.94, 116.12, 114.53, 60.83, 14.12. HRMS: calcd for [M + H], 307.03025; found, 307.03034.

Ethyl 2-(4-Chlorophenyl)imidazo[1,2-a]pyridine-3-carboxylate (26). 2-Aminopyridine (3 equiv) was dissolved in CH₃CN, and ethyl 3-(4-chlorophenyl)-3-oxopropanoate (1 equiv) was added as a solution in CH₃CN followed by the addition of CBr₄ (2 equiv). A reaction mixture was stirred at 80 °C overnight. After the completion of the reaction, the mixture was evaporated to a minimal volume, diluted with water, and extracted with EtOAc. Combined organic phases were dried over sodium sulfate and purified by flash column chromatography. Spectral characteristics match those described in the literature.⁴⁶

6-(4-Chlorophenyl)imidazo[2,1-*b*]thiazole-5-carboxylic Acid (27). The title compound was prepared according to General Procedure V with a minor modification. The product was filtered as a precipitate after acidification. Yield: 286 mg (90%). Spectral characteristics matched those described in the literature.⁴⁷

2-(4-Chlorophenyl)imidazo[1,2-a]pyridine-3-carboxylic Acid (28). The title compound was prepared according to General Procedure V with a minor modification. The product was filtered as a precipitate after acidification. Yield: 548 mg (quant.). ¹H NMR (401 MHz, DMSO-*d*₆): δ 9.40–9.32 (m, 1H), 7.83–7.75 (m, 3H), 7.59–7.52 (m, 1H), 7.50 (dd, *J* = 9.0, 2.5 Hz, 2H), 7.21 (td, *J* = 6.9, 1.3 Hz, 1H). ¹³C NMR (101 MHz, DMSO): δ 162.11, 151.23, 146.75, 133.77, 133.70, 132.35, 129.28, 128.92, 128.80, 128.04, 127.77, 117.53, 115.02, 112.51. HRMS: calcd for [M + H], 273.04253; found, 273.04262.

2-(4-Chlorophenyl)imidazo[1,2-*a*]pyridine-3-carbohydrazide (29). Ester derivative was dissolved in absolute EtOH and N₂H₄·H₂O (10 equiv) was added. The reaction mixture was refluxed overnight. The reaction mixture was cooled to 25 °C, and the formed precipitate was filtered, washed with EtOH, and dried. Yield: 412 mg (87%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 9.70 (s, 1H), 8.57 (d, *J* = 7.0 Hz, 1H), 7.92–7.77 (m, 2H), 7.66 (d, *J* = 9.1 Hz, 1H), 7.56–7.43 (m, 2H), 7.40 (t, *J* = 8.0 Hz, 1H), 7.04 (t, *J* = 6.9 Hz, 1H), 4.67 (s, 2H). ¹³C NMR (101 MHz, DMSO): δ 160.38, 144.68, 143.00, 132.96, 132.64, 129.97, 128.57, 126.85, 126.37, 117.04, 115.62, 113.53. HRMS: calcd for [M + H], 287.06942; found, 287.06950.

2-(4-Chlorophenyl)-*N'*-(2-(3,4-dichlorophenyl)acetyl)imidazo[1,2-*a*]pyridine-3-carbohydrazide (30). 2-(4-Chlorophenyl)imidazo[1,2-*a*]pyridine-3-carbohydrazide was dissolved with dry DMF and ethyl 2-(4-chlorophenyl)imidazo[1,2-*a*]pyridine-3-carboxylate (1 equiv) was added. The reaction mixture was degassed and refilled with argon. HATU (1.2 equiv) was added, followed by the addition of DIPEA (1.2 equiv). The reaction mixture was stirred at 25 °C and monitored by UPLC. After the completion of the reaction, the solution was evaporated to a minimal volume and purified by reverse-phase flash column chromatography. Mobile phase petrol ether/EtOAc (40:80%). Yield: 298 mg (90%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 10.49 (s, 1H), 10.43 (s, 1H), 8.80 (dt, *J* = 7.0, 1.2 Hz, 1H), 8.06–7.98 (m, 2H), 7.95 (s, 1H), 7.71 (dt, *J* = 9.0, 1.2 Hz, 1H), 7.65–7.60 (m, 2H), 7.49 (dd, *J* = 9.0, 2.3 Hz, 2H), 7.44 (ddd, *J* = 9.0, 6.8, 1.3 Hz, 1H), 7.34 (dd, *J* = 8.3, 2.1 Hz, 1H), 7.10 (td, *J* = 6.9, 1.2 Hz, 1H), 3.63 (s, 2H). ¹³C NMR (101 MHz, DMSO): δ 169.21, 160.31, 145.02, 144.01, 136.78, 133.29, 132.28, 131.35, 130.97, 130.62, 130.18, 129.86, 129.58, 128.59, 127.33, 126.68, 117.16, 114.63, 113.89, 35.98. HRMS: calcd for [M + H], 473.03334; found, 473.03327.

2-Chloro-1-(6-(4-chlorophenyl)imidazo[2,1-*b*]thiazol-5-yl)ethan-1-one (31). 6-(4-Chlorophenyl)imidazo[2,1-*b*]thiazole (1 mmol) was dissolved in dry dioxane (4 mL), and chloroacetyl chloride (3 equiv) was added in one portion. The reaction mixture was stirred at 70 °C under an argon atmosphere for 30 min and then at 100 °C overnight. After cooling to 25 °C, a precipitate was formed. The suspension was diluted with a saturated NaHCO₃ solution and extracted with EtOAc. The organic phase was dried over sodium sulfate and purified by flash column chromatography, mobile phase petrol ether/EtOAc (30:60%). Yield: 273 mg (91%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 8.46 (d, *J* = 4.4 Hz, 1H), 7.73–7.68 (m, 2H), 7.64 (d, *J* = 4.4 Hz, 1H), 7.62–7.57 (m, 2H), 4.41 (s, 2H). ¹³C NMR (101 MHz, DMSO): δ 180.76, 154.65, 153.23, 134.82, 133.38, 132.00, 129.02, 122.62, 122.19, 117.57, 47.18. HRMS: calcd for [M + H], 310.98072; found, 310.98089.

2-Chloro-1-(2-(4-chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)ethan-1-one (32). 2-(4-Chlorophenyl)imidazo[1,2-*a*]pyridine (1 mmol) was dissolved in dry dioxane (4 mL) and chloroacetyl chloride (3 equiv) was added in one portion. The reaction mixture was stirred at 70 °C under an argon atmosphere for 30 min and then at 100 °C overnight. After cooling to 25 °C, a precipitate was formed. The suspension was diluted with a saturated NaHCO₃ solution and extracted with EtOAc. The organic phase was dried over sodium sulfate and purified by flash column chromatography; mobile phase petrol ether/EtOAc (30:60%). Yield: 492 mg (76%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 9.63 (d, *J* = 6.9 Hz, 1H), 7.94 (d, *J* = 8.9 Hz, 1H), 7.85 (t, *J* = 7.9 Hz, 1H), 7.73 (d, *J* = 8.2 Hz, 2H), 7.65 (d, *J* = 8.2 Hz, 2H), 7.45 (t, *J* = 6.9 Hz, 1H), 4.36 (s, 2H). ¹³C NMR (101 MHz, DMSO): δ 182.37, 151.37, 145.94, 135.44, 132.49, 132.19, 132.15, 129.23, 129.22, 119.42, 117.34, 116.76, 47.93. HRMS: calcd for [M + H], 305.02429; found, 305.02434.

2-(4-Chlorophenyl)-3-ethynylimidazo[1,2-*a*]pyridine (33). 2-(4-Chlorophenyl)-3-((trimethylsilyl)ethynyl)imidazo[1,2-*a*]pyridine (11, 513 mg, 1.58 mmol) was dissolved in MeOH and K₂CO₃ (435 mg, 2 equiv) was added in one portion. The reaction was stirred at 25 °C and monitored by TLC. After the completion of the reaction, the mixture was diluted with DCM and washed with water. Combined organic phases were dried over sodium sulfate and evaporated, and the residue was purified by flash column chromatography. Mobile

phase petrol ether/EtOAc (30:50%). Yield: 362 mg (90%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 8.46 (dt, *J* = 6.8, 1.2 Hz, 1H), 8.28–8.20 (m, 2H), 7.71 (dt, *J* = 9.0, 1.1 Hz, 1H), 7.60–7.52 (m, 2H), 7.45 (ddd, *J* = 9.0, 6.8, 1.3 Hz, 1H), 7.12 (td, *J* = 6.8, 1.2 Hz, 1H), 5.44 (s, 1H). ¹³C NMR (101 MHz, DMSO): δ 145.95, 144.71, 133.40, 131.99, 128.97, 128.32, 128.28, 127.70, 125.77, 117.27, 114.16, 103.47, 94.39, 72.47. HRMS: calcd for [M + Na], 253.05270; found, 253.05273.

2-(4-Chlorophenyl)-3-(1*H*-1,2,3-triazol-4-yl)imidazo[1,2-*a*]pyridine (34). 2-(4-Chlorophenyl)-3-ethynylimidazo[1,2-*a*]pyridine 33 (100 mg, 0.396 mmol) was dissolved in DMF/MeOH (10:1) mixture, degassed, and purged with argon. To this solution, CuI (10 mol %) and TMSN₃ (1 equiv) were added and the mixture was stirred at 70 °C overnight. After the completion of the reaction, the mixture was diluted with EtOAc and washed with water, and the combined organic phases were dried over sodium sulfate and evaporated. The residue was purified by flash column chromatography, mobile phase cyclohexane/EtOAc (10:70%). Yield: 98 mg (84%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 8.42 (dt, *J* = 7.0, 1.2 Hz, 1H), 8.13 (s, 1H), 7.76–7.66 (m, 3H), 7.47–7.41 (m, 2H), 7.37 (ddd, *J* = 9.0, 6.7, 1.2 Hz, 1H), 6.99 (td, *J* = 6.8, 1.2 Hz, 1H). ¹³C NMR (101 MHz, DMSO): δ 145.21, 142.71, 135.51, 133.42, 133.08, 130.22, 129.90, 129.90, 128.95, 126.44, 125.60, 117.35, 113.58, 112.24. HRMS: calcd for [M + H], 296.06975; found, 296.06964.

2-(4-Chlorophenyl)-3-(1*H*-pyrrol-2-yl)imidazo[1,2-*a*]pyridine (35). 2-(4-Chlorophenyl)-3-iodoimidazo[1,2-*a*]pyridine (430 mg, 1.21 mmol) was dissolved in dioxane (12 mL) and (1-(*tert*-butoxycarbonyl)-1*H*-pyrrol-3-yl)boronic acid (260 mg, 1 equiv) was added, followed by a solution of Na₂CO₃ (392 mg, 3 equiv) in 3 mL of H₂O. A reaction mixture was degassed and refilled with argon. Pd(PPh₃)₄ (72 mg, 5% mol) was added, and the mixture was degassed again and refilled with argon. The reaction mixture was stirred at 90 °C overnight. After cooling to 25 °C, the mixture was diluted with water and extracted with EtOAc. Combined organic phases were dried over sodium sulfate and evaporated. The residue was purified by flash column chromatography; mobile phase petrol ether/EtOAc (20:70%). Yield: 103 mg (35%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 11.28 (s, 1H), 7.99 (dt, *J* = 6.9, 1.2 Hz, 1H), 7.69–7.62 (m, 3H), 7.42–7.37 (m, 2H), 7.32 (ddd, *J* = 9.1, 6.7, 1.3 Hz, 1H), 7.07 (td, *J* = 2.7, 1.5 Hz, 1H), 6.93 (td, *J* = 6.8, 1.2 Hz, 1H), 6.40–6.37 (m, 1H), 6.34 (dt, *J* = 3.3, 2.5 Hz, 1H). ¹³C NMR (101 MHz, DMSO): δ 144.27, 141.30, 133.34, 132.24, 128.58, 128.51, 125.77, 124.66, 120.79, 117.38, 116.91, 114.57, 112.91, 111.12, 109.48. HRMS: calcd for [M + H], 294.07925; found, 294.07926.

2-(4-Chlorophenyl)-3-(1*H*-pyrazol-4-yl)imidazo[1,2-*a*]pyridine (36). 2-(4-Chlorophenyl)-3-iodoimidazo[1,2-*a*]pyridine (334 mg, 0.94 mmol) was dissolved in dioxane (8 mL) and 4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-1*H*-pyrazole (183 mg, 1 equiv) was added, followed by a solution of Na₂CO₃ (3 equiv) in 2 mL of H₂O. A reaction mixture was degassed and refilled with argon. Pd(dppf)Cl₂ (40 mg, 5% mol) was added, and the mixture was degassed again and refilled with argon. The reaction mixture was stirred at 90 °C overnight. After cooling to 25 °C, the mixture was diluted with water and extracted with EtOAc. Combined organic phases were dried over sodium sulfate and evaporated. The residue was purified by flash column chromatography; mobile phase petrol ether/EtOAc (20:70%). Yield: 112 mg (40%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 8.54 (d, *J* = 6.7 Hz, 1H), 8.45 (s, 1H), 8.00 (d, *J* = 8.3 Hz, 2H), 7.59 (d, *J* = 9.1 Hz, 1H), 7.51 (d, *J* = 8.3 Hz, 2H), 7.29–7.22 (m, 1H), 6.92 (t, *J* = 6.7 Hz, 1H). HRMS: calcd for [M + H], 295.07450; found, 295.07467.

Methyl 2-Chloro-5-((4-(2-(4-chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-1*H*-1,2,3-triazol-1-yl)-methyl)benzoate (37). The title compound was prepared according to General Procedure IV. Mobile phase petrol ether/EtOAc (60:90%). Yield: 456 mg (84%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 8.54 (s, 1H), 8.46 (dt, *J* = 7.0, 1.1 Hz, 1H), 7.84–7.81 (m, 1H), 7.72–7.67 (m, 3H), 7.65 (d, *J* = 8.3 Hz, 1H), 7.56 (dd, *J* = 8.3, 2.3 Hz, 1H), 7.44–7.40 (m, 2H), 7.40–7.36 (m, 1H), 7.00 (td, *J* = 6.8, 1.2 Hz, 1H), 5.79 (s, 2H), 3.87 (s, 3H). ¹³C NMR (101 MHz, DMSO): δ 165.39, 144.95, 142.46, 136.39, 135.56,

133.01, 132.84, 132.82, 131.82, 131.52, 130.65, 130.46, 129.62, 128.68, 126.29, 125.94, 125.39, 117.12, 113.37, 111.55, 52.89, 52.07. HRMS: calcd for [M + H], 478.08328; found, 478.08321.

2-Chloro-5-((4-(2-(4-chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-1*H*-1,2,3-triazol-1-yl)methyl)-benzoic Acid (38). The title compound was prepared according to General Procedure V. Mobile phase: H₂O/MeOH (10:80%). Yield: 528 mg (92%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 8.56 (s, 1H), 8.46 (d, *J* = 6.9 Hz, 1H), 7.78 (d, *J* = 2.3 Hz, 1H), 7.73–7.67 (m, 3H), 7.62 (d, *J* = 8.3 Hz, 1H), 7.52 (dd, *J* = 8.3, 2.3 Hz, 1H), 7.41 (dd, *J* = 18.9, 8.6 Hz, 3H), 7.01 (t, *J* = 6.8 Hz, 1H), 5.79 (s, 2H). ¹³C NMR (101 MHz, DMSO): δ 166.58, 144.94, 142.42, 136.36, 135.43, 133.00, 132.85, 132.14, 131.83, 131.66, 131.37, 130.37, 129.57, 128.70, 126.28, 125.98, 125.36, 117.11, 113.35, 111.54, 52.11. HRMS: calcd for [M + H], 430.10653; found, 430.10615.

2-Chloro-5-((4-(2-(4-chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-1*H*-1,2,3-triazol-1-yl)methyl)benzamide (39). The title compound was prepared according to General Procedure VI. Mobile phase: H₂O/MeOH (30:100%). Yield: 251 mg (88%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 8.56 (s, 1H), 8.51–8.43 (m, 1H), 8.00–7.84 (m, 1H), 7.70 (td, *J* = 6.8, 2.1 Hz, 4H), 7.53 (t, *J* = 7.3 Hz, 1H), 7.49–7.34 (m, 4H), 7.00 (qd, *J* = 6.8, 6.1, 1.1 Hz, 1H), 5.76 (s, 2H). ¹³C NMR (101 MHz, DMSO): δ 167.85, 140.71, 137.56, 135.11, 134.72, 134.37, 133.41, 133.00, 130.31, 130.17, 129.63, 129.36, 128.39, 127.21, 127.06, 126.55, 117.19, 113.25, 113.19, 52.31. HRMS: calcd for [M + H], 463.08354; found, 463.08359.

3-(1-(3-Carbamoyl-4-chlorobenzyl)-1*H*-1,2,3-triazol-4-yl)-2-(4-chlorophenyl)imidazo[1,2-*a*]pyridin-1-ium Chloride (39HCl). 2-Chloro-5-((4-(2-(4-chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-1*H*-1,2,3-triazol-1-yl)methyl)benzamide was dissolved in THF and cooled in an ice bath, and HCl 1 M in ether was added while stirring vigorously. The formed salt was stirred at 25 °C for 20 min, diluted with dry diethylether, filtered, and washed with more diethylether. EA: C, 54.96%; H, 3.51%; N, 16.54%; Cl, 21.48%.

2-Chloro-5-((4-(2-(4-chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-1*H*-1,2,3-triazol-1-yl)methyl)-*N*-methylbenzamide (40). The title compound was prepared according to General Procedure VI. Mobile phase: H₂O/MeOH (30:100%). Yield: 87 mg (89%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 8.57 (s, 1H), 8.47 (dt, *J* = 7.0, 1.1 Hz, 1H), 8.41 (q, *J* = 4.4 Hz, 1H), 7.72–7.67 (m, 3H), 7.55 (d, *J* = 8.2 Hz, 1H), 7.47–7.41 (m, 4H), 7.39 (ddd, *J* = 9.1, 6.7, 1.2 Hz, 1H), 7.00 (td, *J* = 6.8, 1.2 Hz, 1H), 5.75 (s, 2H), 2.77 (d, *J* = 4.6 Hz, 3H). ¹³C NMR (101 MHz, DMSO): δ 166.58, 144.96, 142.43, 137.62, 136.35, 135.20, 133.01, 132.89, 130.40, 130.24, 129.86, 129.60, 128.75, 128.44, 126.31, 125.94, 125.40, 117.13, 113.38, 111.57, 52.22, 26.17. HRMS: calcd for [M + H], 477.09919; found, 477.09927.

2-Chloro-5-((4-(2-(4-chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-1*H*-1,2,3-triazol-1-yl)methyl)-*N,N*-dimethylbenzamide (41). The title compound was prepared according to General Procedure VI. Mobile phase: H₂O/MeOH (30:100%). Yield: 228 mg (95%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 8.54 (s, 1H), 8.46 (dt, *J* = 6.9, 1.2 Hz, 1H), 8.31 (s, 1H), 7.71–7.66 (m, 3H), 7.58 (d, *J* = 8.3 Hz, 1H), 7.43 (dd, *J* = 8.6, 2.2 Hz, 3H), 7.38 (ddd, *J* = 9.1, 6.7, 1.3 Hz, 1H), 7.35 (d, *J* = 2.2 Hz, 1H), 6.99 (td, *J* = 6.8, 1.2 Hz, 1H), 5.76 (s, 2H), 3.02 (s, 3H), 2.75 (s, 3H). ¹³C NMR (101 MHz, DMSO): δ 166.79, 144.94, 142.44, 136.80, 136.38, 135.78, 133.01, 132.85, 130.03, 130.02, 129.57, 129.06, 128.70, 127.44, 126.26, 125.92, 125.37, 117.11, 113.34, 111.55, 52.23, 37.65, 34.22. HRMS: calcd for [M + H], 491.11484; found, 491.11493.

2-Chloro-5-((4-(2-(4-chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-1*H*-1,2,3-triazol-1-yl)methyl)-*N*-methoxy-*N*-methylbenzamide (42). The title compound was prepared according to General Procedure VI with a minor modification. 2-Chloro-5-((4-(2-(4-chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-1*H*-1,2,3-triazol-1-yl)methyl)-benzoic acid (528 mg, 1.13 mmol) was dissolved in dry DCM (8 mL) and cooled in an ice bath. Oxalyl chloride (0.2 mL, 2 equiv) was added followed by a catalytic amount of DMF. The reaction mixture was stirred at 25 °C overnight. The solvent was evaporated, and crude acyl chloride was used in the next step without further purification. Crude acyl chloride was dissolved in dry DCM (10 mL), and *N,O*-

dimethylhydroxylamine hydrochloride (112 mg, 1 equiv) was added followed by TEA (0.32 mL, 2 equiv). The reaction mixture was stirred at 25 °C for 3 h. Then the mixture was diluted with DCM, washed with water, and purified by flash column chromatography, mobile phase petrol ether/EtOAc (60:100%). Yield: 521 (91%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 8.55 (s, 1H), 8.50–8.39 (m, 1H), 7.69 (dd, *J* = 8.7, 2.0 Hz, 3H), 7.58 (d, *J* = 8.1 Hz, 2H), 7.48–7.40 (m, 4H), 7.38 (ddd, *J* = 9.0, 6.7, 1.3 Hz, 1H), 6.99 (td, *J* = 6.8, 1.2 Hz, 1H), 5.78 (s, 2H), 3.39 (s, 3H), 3.29 (s, 3H). ¹³C NMR (101 MHz, DMSO): δ 166.91, 144.93, 142.44, 136.35, 135.93, 135.22, 133.01, 132.85, 130.22, 129.83, 129.57, 129.40, 128.69, 127.15, 126.24, 125.90, 125.34, 117.11, 113.32, 111.55, 61.20, 52.22, 31.99. HRMS: calcd for [M + H], 507.10976; found, 507.10983.

1-(2-Chloro-5-((4-(2-(4-chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-1*H*-1,2,3-triazol-1-yl)methyl)-phenyl)ethan-1-one (43). The title compound was prepared according to General Procedure VII. Mobile phase: H₂O/MeOH (30:100%). Yield: 68 mg (72%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 8.54 (s, 1H), 8.47 (dt, *J* = 6.9, 1.1 Hz, 1H), 7.73–7.67 (m, 4H), 7.61 (d, *J* = 8.3 Hz, 1H), 7.50 (dd, *J* = 8.3, 2.2 Hz, 1H), 7.45–7.41 (m, 2H), 7.38 (ddd, *J* = 9.0, 6.8, 1.3 Hz, 1H), 7.00 (td, *J* = 6.8, 1.1 Hz, 1H), 5.78 (s, 2H), 2.59 (s, 3H). ¹³C NMR (101 MHz, DMSO): δ 199.81, 144.92, 142.46, 139.01, 136.36, 135.53, 133.01, 132.82, 131.95, 131.18, 129.76, 129.62, 129.14, 128.68, 126.25, 125.89, 125.38, 117.11, 113.32, 111.56, 52.20, 30.64. HRMS: calcd for [M + H], 462.08829; found, 462.08840.

1-(2-Chloro-5-((4-(2-(4-chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-1*H*-1,2,3-triazol-1-yl)methyl)phenyl)propan-1-one (44). The title compound was prepared according to General Procedure VII. Mobile phase: H₂O/MeOH (30:100%). Yield: 51 mg (69%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 8.54 (s, 1H), 8.49–8.42 (m, 1H), 7.72–7.67 (m, 3H), 7.65 (d, *J* = 2.1 Hz, 1H), 7.60 (d, *J* = 8.3 Hz, 1H), 7.50–7.46 (m, 1H), 7.44–7.36 (m, 3H), 7.00 (td, *J* = 6.8, 1.1 Hz, 1H), 5.77 (s, 2H), 2.93 (q, *J* = 7.2 Hz, 2H), 1.12–1.02 (m, 3H). ¹³C NMR (101 MHz, DMSO): δ 203.21, 144.92, 142.44, 139.56, 136.35, 135.51, 133.01, 132.82, 131.52, 130.94, 129.60, 129.30, 128.67, 128.46, 126.26, 125.91, 125.38, 117.11, 113.32, 111.56, 52.23, 35.75, 8.03. HRMS: calcd for [M + H], 476.10394; found, 476.10407.

2-Chloro-5-((4-(2-(4-chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-1*H*-1,2,3-triazol-1-yl)methyl)benzaldehyde (45). The title compound was prepared according to General Procedure VII with LAH. Crystallized from ACN. Yield: 158 mg (75%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 10.34 (s, 1H), 8.56 (s, 1H), 8.47 (dt, *J* = 6.9, 1.2 Hz, 1H), 7.83 (t, *J* = 1.4 Hz, 1H), 7.72–7.67 (m, 5H), 7.45–7.41 (m, 2H), 7.38 (ddd, *J* = 9.1, 6.7, 1.3 Hz, 1H), 6.99 (td, *J* = 6.8, 1.3 Hz, 1H), 5.83 (s, 2H). ¹³C NMR (101 MHz, DMSO): δ 189.72, 144.93, 142.46, 136.39, 136.33, 136.13, 135.17, 133.00, 132.83, 132.36, 131.43, 129.59, 128.79, 128.70, 126.24, 125.95, 125.35, 117.11, 113.33, 111.52, 52.06. HRMS: calcd for [M + H], 448.07264; found, 448.07224.

(*E*)-2-Chloro-5-((4-(2-(4-chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-1*H*-1,2,3-triazol-1-yl)methyl)benzaldehyde Oxime (46). 2-Chloro-5-((4-(2-(4-chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-1*H*-1,2,3-triazol-1-yl)methyl)benzaldehyde (45) was dissolved in DCM, and NH₂OH·HCl (1.2 equiv) was added, followed by the addition of TEA (1.2 equiv). The reaction mixture was stirred at 25 °C and washed with water, and the organic phase was dried over sodium sulfate and evaporated. The residue was purified by flash column chromatography. Mobile phase petrol ether/EtOAc (50:100%). Yield: 75 mg (90%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 11.75 (s, 1H), 8.51 (s, 1H), 8.46 (dt, *J* = 7.0, 1.2 Hz, 1H), 8.36 (s, 1H), 7.78 (d, *J* = 2.2 Hz, 1H), 7.71–7.66 (m, 4H), 7.56 (d, *J* = 8.3 Hz, 1H), 7.43–7.35 (m, 4H), 7.00 (td, *J* = 6.8, 1.3 Hz, 1H), 5.77 (s, 2H). ¹³C NMR (101 MHz, DMSO): δ 145.22, 144.72, 142.75, 136.69, 136.00, 133.29, 133.13, 132.37, 131.02, 130.78, 130.63, 129.89, 128.97, 126.54, 126.35, 126.16, 125.63, 117.41, 113.67, 111.85, 52.64. HRMS: calcd for [M + H], 463.08354; found, 463.08301.

2-Chloro-5-((4-(2-(4-chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-1*H*-1,2,3-triazol-1-yl)methyl)-phenyl)methanol (47). Methyl 2-chloro-5-((4-(2-(4-chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-1*H*-1,2,3-triazol-1-yl)-methyl)benzoate (160 mg, 0.334 mmol) was

dissolved in dry THF (6 mL), cooled in an ice bath, and degassed. A flask was refilled with argon, and LAH (0.35 mL, 1 equiv) was added. A reaction mixture was allowed to warm to 25 °C and stirred overnight. After the completion of the reaction, the mixture was carefully quenched with ice and extracted with EtOAc. Combined organic phases were dried over sodium sulfate and evaporated. The residue was purified by flash column chromatography (cyclohexane/EtOAc 40:70%). Yield: 85 mg (57%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 8.52 (s, 1H), 8.46 (dt, *J* = 6.9, 1.2 Hz, 1H), 7.73–7.67 (m, 4H), 7.51 (d, *J* = 2.3 Hz, 1H), 7.47–7.41 (m, 3H), 7.38 (ddd, *J* = 9.1, 6.7, 1.3 Hz, 1H), 7.30–7.26 (m, 1H), 7.00 (td, *J* = 6.8, 1.2 Hz, 1H), 5.76 (s, 2H), 5.51 (t, *J* = 5.4 Hz, 1H), 4.57 (dt, *J* = 5.4, 0.9 Hz, 2H). ¹³C NMR (101 MHz, DMSO): δ 144.93, 142.41, 140.22, 136.29, 135.16, 133.00, 132.82, 130.72, 129.59, 129.31, 128.71, 127.79, 127.22, 126.24, 125.87, 125.33, 117.11, 113.34, 111.60, 60.22, 52.73. HRMS: calcd for [M + H], 450.08829; found, 450.08835.

3-(1-(4-Chloro-3-(methoxymethyl)benzyl)-1H-1,2,3-triazol-4-yl)-2-(4-chlorophenyl)imidazo[1,2-*a*]pyridine (**48**). 2-Chloro-5-((4-(2-(4-chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-1H-1,2,3-triazol-1-yl)-methyl)phenyl)methanol (85 mg, 0.188 mmol) was dissolved in dry THF (4 mL) and NaH (8 mg, 1.1 equiv) was added. After 10 min, CH₃I (0.013 mL, 1 equiv) was added and the mixture was stirred at 25 °C overnight. The mixture was purified by flash column chromatography (cyclohexane/EtOAc 30:80%). Yield: 35 mg (40%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 8.51 (s, 1H), 8.46 (dt, *J* = 7.0, 1.3 Hz, 1H), 7.71–7.65 (m, 3H), 7.51–7.45 (m, 2H), 7.40 (dd, *J* = 8.6, 2.0 Hz, 2H), 7.35 (d, *J* = 1.3 Hz, 0H), 7.32 (dd, *J* = 8.2, 2.3 Hz, 1H), 6.99 (td, *J* = 6.8, 1.3 Hz, 1H), 5.75 (s, 2H), 4.48 (s, 2H), 3.36 (d, *J* = 1.8 Hz, 3H, water overlapping). ¹³C NMR (101 MHz, DMSO): δ 144.92, 142.42, 136.41, 136.33, 135.26, 133.01, 132.81, 131.72, 129.69, 129.59, 128.71, 128.64, 128.46, 126.20, 125.79, 125.30, 117.10, 113.31, 111.60, 70.77, 58.30, 52.58. HRMS: calcd for [M + H], 464.10394; found, 464.10396.

3-(1-(4-Chloro-3-nitrobenzyl)-1H-1,2,3-triazol-4-yl)-2-(4-chlorophenyl)imidazo[1,2-*a*]pyridine (**49**). The title compound was prepared according to General Procedure IV (Scheme 1). Mobile phase petrol ether/EtOAc (40:100%). Yield: 835 mg (83%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 8.58 (s, 1H), 8.50 (dt, *J* = 6.9, 1.2 Hz, 1H), 8.15 (d, *J* = 2.1 Hz, 1H), 7.82 (d, *J* = 8.4 Hz, 1H), 7.73–7.65 (m, 4H), 7.43–7.38 (m, 2H), 7.36 (ddd, *J* = 9.1, 6.7, 1.3 Hz, 1H), 6.98 (td, *J* = 6.8, 1.2 Hz, 1H). ¹³C NMR (101 MHz, DMSO): δ 147.65, 144.93, 142.50, 136.94, 136.53, 133.62, 132.99, 132.85, 132.31, 129.63, 128.62, 126.15, 125.94, 125.41, 125.39, 125.10, 117.06, 113.23, 111.50, 51.70. HRMS: calcd for [M – H], 465.06281; found, 465.06250.

2-Chloro-5-((4-(2-(4-chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-1H-1,2,3-triazol-1-yl)methyl)-aniline (**50**). 3-(1-(4-Chloro-3-nitrobenzyl)-1H-1,2,3-triazol-4-yl)-2-(4-chlorophenyl)imidazo[1,2-*a*]pyridine (**49**) was dissolved in MeOH and AcOH (7 equiv), and Fe (3.5 equiv) was added. The reaction mixture was stirred at reflux until the completion of the reaction. After cooling to 25 °C, the mixture was extracted with EtOAc, washed with a NaHCO₃ solution, and the organic phase was dried over sodium sulfate and evaporated. The residue was purified by flash column chromatography. Mobile phase cyclohexane/EtOAc (40:100%). Yield: 490 mg (79%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 8.46 (dt, *J* = 7.0, 1.2 Hz, 1H), 8.44 (s, 1H), 7.73–7.66 (m, 3H), 7.44–7.39 (m, 2H), 7.37 (ddd, *J* = 9.1, 6.7, 1.2 Hz, 1H), 7.21 (d, *J* = 8.2 Hz, 1H), 6.99 (td, *J* = 6.8, 1.2 Hz, 1H), 6.71 (d, *J* = 2.1 Hz, 1H), 6.52–6.47 (m, 1H), 5.59 (s, 2H), 5.50 (s, 2H). ¹³C NMR (101 MHz, DMSO): δ 145.11, 144.88, 142.37, 136.29, 135.75, 132.96, 132.81, 132.80, 129.65, 129.60, 129.49, 128.65, 128.61, 126.19, 126.14, 125.74, 125.33, 117.07, 116.94, 116.02, 114.31, 113.30, 111.65, 52.95. HRMS: calcd for [M + H], 435.08863; found, 435.08815.

N-(2-Chloro-5-((4-(2-(4-chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-1H-1,2,3-triazol-1-yl)methyl)-phenyl)acetamide (**51**). 2-Chloro-5-((4-(2-(4-chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-1H-1,2,3-triazol-1-yl)methyl)-aniline (**50**) was dissolved in dioxane and Ac₂O (1.5 equiv) was added followed by pyridine (1.5 equiv). The reaction mixture was stirred at 25 °C overnight. After the completion of the

reaction, the mixture was evaporated, diluted with EtOAc, and washed with water. The organic phase was dried over sodium sulfate and evaporated. The residue was purified by flash column chromatography. Mobile phase cyclohexane/EtOAc (30:100%). Yield: 75 mg (90%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 9.58 (s, 1H), 8.54 (s, 1H), 8.46 (dt, *J* = 6.9, 1.2 Hz, 1H), 7.73–7.67 (m, 4H), 7.54 (d, *J* = 8.3 Hz, 1H), 7.46–7.41 (m, 2H), 7.38 (ddd, *J* = 9.0, 6.7, 1.3 Hz, 1H), 7.17 (dd, *J* = 8.3, 2.2 Hz, 1H), 6.99 (td, *J* = 6.8, 1.2 Hz, 1H), 5.74 (s, 2H), 2.11 (s, 3H). ¹³C NMR (101 MHz, DMSO): δ 169.27, 145.22, 142.68, 136.54, 135.83, 135.80, 133.25, 133.11, 130.25, 129.85, 129.00, 127.90, 126.58, 126.28, 125.78, 125.66, 125.41, 117.35, 114.34, 113.64, 111.84, 52.79, 23.86. HRMS: calcd for [M + H], 477.09919; found, 477.09866.

2-Chloro-5-((4-(2-(4-chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)-1H-1,2,3-triazol-1-yl)methyl)benzoyl Chloride (**52**). The title compound was prepared according to General Procedure V and used in the next step without further purification.

3-(4-Bromophenyl)-2-(4-chlorophenyl)imidazo[1,2-*a*]pyridine (**53**). 2-(4-Chlorophenyl)-3-iodoimidazo[1,2-*a*]pyridine (300 mg, 0.846 mmol) was combined with (4-bromophenyl)boronic acid (1 equiv) and diluted with dioxane/H₂O mixture (4:1; 8 mL). Na₂CO₃ (3 equiv) was added, and the mixture was degassed and refilled with argon. Pd(dppf)Cl₂·DCM (5 mol %) was added, and the mixture was degassed again and refilled with argon. The mixture was stirred at 95 °C overnight (Scheme 8). After the completion of the reaction, the mixture was diluted with water and extracted with EtOAc. The organic phases were dried over sodium sulfate and evaporated. The residue was purified by flash column chromatography. Mobile phase: cyclohexane/EtOAc (20:60%). Yield: 193 mg (59%). ¹³C NMR (101 MHz, DMSO): δ 144.73, 140.86, 133.34, 133.20, 133.17, 132.72, 129.61, 129.22, 128.92, 128.75, 126.10, 124.32, 123.04, 117.42, 113.42.

2-(4-Chlorophenyl)-3-(6-chloropyridin-3-yl)imidazo[1,2-*a*]pyridine (**54**). 2-(4-Chlorophenyl)-3-iodoimidazo[1,2-*a*]pyridine (300 mg, 0.846 mmol) was combined with (6-chloropyridin-3-yl)boronic acid (1.3 equiv) and diluted with a dioxane/H₂O mixture (4:1; 8 mL). Na₂CO₃ (3 equiv) was added, and the mixture was degassed and refilled with argon. Pd(dppf)Cl₂·DCM (5 mol %) was added and the mixture was degassed again and refilled with argon. The mixture was stirred at 95 °C overnight (Scheme 8). After the completion of the reaction, the mixture was diluted with water and extracted with EtOAc. The organic phases were dried over sodium sulfate and evaporated. The residue was purified by flash column chromatography. Mobile phase: cyclohexane/EtOAc (15:60%). Yield: 183 mg (63%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 8.53 (dd, *J* = 2.5, 0.8 Hz, 1H), 8.17 (dt, *J* = 6.9, 1.2 Hz, 1H), 8.06 (dd, *J* = 8.2, 2.5 Hz, 1H), 7.74 (dd, *J* = 8.2, 0.8 Hz, 1H), 7.68 (dt, *J* = 9.1, 1.2 Hz, 1H), 7.58–7.52 (m, 2H), 7.42–7.33 (m, 3H), 6.93 (td, *J* = 6.8, 1.2 Hz, 1H). ¹³C NMR (101 MHz, DMSO): δ 151.85, 151.07, 145.19, 142.40, 142.01, 133.04, 132.95, 129.75, 129.07, 126.53, 125.76, 125.37, 124.70, 117.40, 116.95, 113.59.

3-(4-Benzylphenyl)-2-(4-chlorophenyl)imidazo[1,2-*a*]pyridine (**55**). 3-(4-Bromophenyl)-2-(4-chlorophenyl)imidazo[1,2-*a*]pyridine (**53**) was dissolved in anhydrous THF (6 mL) under an argon atmosphere. Pd₂dba₃ (5 mol %) was added followed by the addition of XantPhos (15 mol %). To the mixture, a benzylzinc bromide solution 0.5 M in THF (2 equiv) was added and the mixture was stirred at 60 °C overnight (Scheme 8). The mixture was evaporated to a minimal volume, adsorbed onto silica, and purified by flash column chromatography followed by reverse-phase flash column chromatography. Mobile phase: cyclohexane/EtOAc (30:100%) and H₂O/ACN (30:100%). Yield: 143 mg (72%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 7.95 (d, *J* = 6.8 Hz, 1H), 7.63 (d, *J* = 9.0 Hz, 1H), 7.61–7.55 (m, 2H), 7.42 (d, *J* = 8.0 Hz, 2H), 7.37 (d, *J* = 8.2 Hz, 2H), 7.35–7.25 (m, 6H), 7.24–7.17 (m, 1H), 6.83 (td, *J* = 6.8, 1.2 Hz, 1H), 4.04 (s, 2H). ¹³C NMR (101 MHz, DMSO): δ 144.17, 142.47, 140.85, 140.15, 133.33, 132.22, 130.77, 130.13, 129.17, 129.03, 128.71, 128.50, 126.78, 126.30, 125.51, 123.96, 120.97, 117.06, 112.93, 66.52.

3-(6-Benzylpyridin-3-yl)-2-(4-chlorophenyl)imidazo[1,2-*a*]pyridine (**56**). 2-(4-Chlorophenyl)-3-(6-chloropyridin-3-yl)imidazo-

[1,2-*a*]pyridine (**54**) was dissolved in anhydrous THF (6 mL) under an argon atmosphere. Pd₂dba₃ (5 mol %) was added, followed by the addition of XantPhos (15 mol %). To the mixture, a benzylzinc bromide solution 0.5 M in THF (2 equiv) was added and the mixture was stirred at 60 °C overnight (Scheme 8). The mixture was evaporated to a minimal volume, adsorbed onto silica, and purified by flash column chromatography followed by reverse-phase flash column chromatography. Mobile phase: cyclohexane/EtOAc (30:100%) and H₂O/ACN (30:100%). Yield: 158 mg (75%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 8.57 (dd, *J* = 2.3, 0.9 Hz, 1H), 8.06 (dt, *J* = 6.9, 1.2 Hz, 1H), 7.89 (dd, *J* = 8.0, 2.3 Hz, 1H), 7.67 (dt, *J* = 9.1, 1.1 Hz, 1H), 7.57–7.52 (m, 2H), 7.48 (dd, *J* = 8.1, 0.9 Hz, 1H), 7.39–7.30 (m, 7H), 7.23 (ddt, *J* = 8.6, 6.2, 1.8 Hz, 1H), 6.88 (td, *J* = 6.8, 1.2 Hz, 1H), 4.21 (s, 2H). ¹³C NMR (101 MHz, DMSO): δ 161.60, 151.00, 144.97, 141.61, 139.81, 139.36, 133.35, 132.78, 129.65, 129.58, 128.97, 126.78, 126.27, 124.52, 124.11, 123.37, 118.26, 117.41, 113.49, 66.82.

5-(2-(4-Chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)pyridin-2-amine (**57**). 2-(4-Chlorophenyl)-3-iodoimidazo[1,2-*a*]pyridine (0.2 g, 0.56 mmol) was combined with 2-aminopyridine-5-boronic acid pinacol ester (1.1 equiv) and dissolved in dioxane (8 mL) followed by a solution of Na₂CO₃ in 2 mL of H₂O. A reaction mixture was degassed and refilled with argon. Pd(dppf)Cl₂ (5 mol %) was added, and the mixture was degassed again and refilled with argon. The reaction mixture was stirred at 90 °C overnight (Scheme 9). After cooling to 25 °C, the mixture was diluted with water and extracted with EtOAc. Combined organic phases were dried over sodium sulfate and evaporated. The residue was purified by flash column chromatography; mobile phase petrol ether/EtOAc (20:70%). Yield: 100 mg (53%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 8.00 (dt, *J* = 6.9, 1.2 Hz, 1H), 7.96 (d, *J* = 2.3 Hz, 1H), 7.71–7.65 (m, 2H), 7.62 (dt, *J* = 9.1, 1.2 Hz, 1H), 7.47 (dd, *J* = 8.5, 2.4 Hz, 1H), 7.44–7.37 (m, 2H), 7.30 (ddd, *J* = 9.0, 6.7, 1.3 Hz, 1H), 6.88 (td, *J* = 6.8, 1.2 Hz, 1H), 6.64 (d, *J* = 8.5 Hz, 1H), 6.38 (s, 2H). ¹³C NMR (101 MHz, DMSO): δ 160.23, 149.97, 144.19, 140.26, 139.39, 133.53, 132.09, 128.97, 128.58, 125.43, 124.19, 119.37, 117.01, 112.82, 112.32, 108.79. HRMS: calcd for [M + H], 321.0907; found, 321.0909.

2-(4-Chlorophenyl)-3-(6-methoxypyridin-3-yl)imidazo[1,2-*a*]pyridine (**58**). 2-(4-Chlorophenyl)-3-iodoimidazo[1,2-*a*]pyridine (0.2 g, 0.56 mmol) was combined with (6-methoxypyridin-3-yl)boronic acid (1.1 equiv) and dissolved in dioxane (8 mL), followed by a solution of Na₂CO₃ in 2 mL of H₂O. The reaction mixture was degassed and refilled with argon. Pd(dppf)Cl₂ (5 mol %) was added, and the mixture was degassed again and refilled with argon. The reaction mixture was stirred at 90 °C overnight. After cooling to 25 °C, the mixture was diluted with water and extracted with EtOAc. Combined organic phases were dried over sodium sulfate and evaporated. The residue was purified by flash column chromatography; mobile phase petrol ether/EtOAc (20:70%). Yield: 100 mg (53%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 8.29 (dd, *J* = 2.5, 0.8 Hz, 1H), 8.03 (dt, *J* = 6.9, 1.2 Hz, 1H), 7.85 (dd, *J* = 8.5, 2.4 Hz, 1H), 7.66 (dt, *J* = 9.0, 1.2 Hz, 1H), 7.63–7.57 (m, 2H), 7.42–7.36 (m, 2H), 7.33 (ddd, *J* = 9.0, 6.7, 1.3 Hz, 1H), 7.05 (dd, *J* = 8.6, 0.8 Hz, 1H), 6.90 (td, *J* = 6.8, 1.3 Hz, 1H), 3.95 (s, 3H). ¹³C NMR (101 MHz, DMSO): δ 164.28, 149.39, 144.77, 142.09, 141.25, 133.43, 132.65, 129.46, 128.97, 126.10, 124.57, 118.88, 118.25, 117.33, 113.36, 112.12, 53.94. HRMS: calcd for [M + H], 336.0904; found, 336.0902.

5-(2-(4-Chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)pyridin-2-ol (**59**). 2-(4-Chlorophenyl)-3-(6-methoxypyridin-3-yl)imidazo[1,2-*a*]pyridine **58** was treated with 4M HCl/dioxane at 95 °C overnight. After completion of the reaction, the solvent was evaporated and the residue neutralized with a saturated NaHCO₃ solution and extracted with EtOAc. The organic phase was dried over sodium sulfate, evaporated, and purified by RP-flash column chromatography. Mobile phase H₂O/MeOH 15:90%. Yield: 72 mg (73%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 12.04 (s, 1H), 8.12 (dt, *J* = 6.9, 1.2 Hz, 1H), 7.74–7.69 (m, 2H), 7.67–7.59 (m, 2H), 7.48–7.41 (m, 3H), 7.32 (ddd, *J* = 9.0, 6.7, 1.2 Hz, 1H), 6.92 (td, *J* = 6.8, 1.2 Hz, 1H), 6.52

(dd, *J* = 9.4, 0.7 Hz, 1H). ¹³C NMR (101 MHz, DMSO): δ 162.36, 144.57, 143.64, 140.96, 138.82, 133.48, 132.60, 129.35, 129.01, 126.01, 124.89, 121.73, 117.63, 117.22, 113.19, 106.40. HRMS: calcd for [M + H], 322.07417; found, 322.07397.

2-Chloro-5-((5-(2-(4-chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)pyridin-2-yl)amino)benzamide (**60**). 5-(2-(4-Chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)pyridin-2-amine was combined with 5-bromo-2-chlorobenzamide (1 equiv) and diluted with dioxane, and the mixture was degassed and refilled with argon. To the solution, Na^tBuO (1.5 equiv), XantPhos (10 mol %), and Pd₂dba₃ (5 mol %) were added and the mixture was stirred at 100 °C overnight. The mixture was filtered through Celite, and the solvent was evaporated to a minimal volume and purified by flash chromatography followed by RP-flash column chromatography. Mobile phase EtOAc, then H₂O/MeOH (20:100%). Yield: 45 mg (51%). ¹H NMR (401 MHz, DMSO-*d*₆): δ 9.67 (s, 1H), 8.30 (dd, *J* = 2.3, 0.8 Hz, 1H), 8.10 (dt, *J* = 6.9, 1.2 Hz, 1H), 7.90–7.82 (m, 3H), 7.71 (dd, *J* = 8.6, 2.4 Hz, 1H), 7.68–7.63 (m, 3H), 7.58–7.55 (m, 1H), 7.43–7.39 (m, 2H), 7.38 (d, *J* = 8.8 Hz, 1H), 7.33 (ddd, *J* = 9.1, 6.7, 1.3 Hz, 1H), 7.04 (dd, *J* = 8.7, 0.8 Hz, 1H), 6.91 (td, *J* = 6.8, 1.2 Hz, 1H). ¹³C NMR (101 MHz, DMSO): δ 168.53, 155.68, 149.29, 144.40, 140.74, 140.26, 139.88, 137.53, 133.32, 132.27, 129.89, 129.16, 128.68, 125.71, 124.33, 120.63, 119.95, 118.64, 117.93, 117.06, 115.77, 113.00, 112.08. HRMS: calcd for [M + H], 474.08829; found, 474.08786.

2-Chloro-5-((5-(2-(4-chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)pyridin-2-yl)oxy)benzamide (**61**). 5-(2-(4-Chlorophenyl)imidazo[1,2-*a*]pyridin-3-yl)pyridin-2-ol (1.5 equiv) was combined with 5-bromo-2-chlorobenzamide (50 mg, 1 equiv) followed by BPO (2 mol %), CuI (2 mol %), and K₃PO₄ (2 equiv). The mixture was degassed and refilled with argon, and dry DMF (1 mL) was added. The mixture was degassed and refilled with argon and heated up to 110 °C overnight. The mixture was filtrated over Celite, dissolved in MeOH, and purified by RP-flash column chromatography. Mobile phase: H₂O/CH₃CN (20:100%). The product was crystallized from THF. Yield: 4 mg (6%). ¹H NMR (500 MHz, DMSO-*d*₆): δ 8.36 (d, *J* = 6.9 Hz, 1H), 8.05 (d, *J* = 2.5 Hz, 1H), 7.96 (s, 1H), 7.85–7.80 (m, 2H), 7.71 (s, 1H), 7.67–7.62 (m, 4H), 7.50 (dd, *J* = 9.4, 2.6 Hz, 1H), 7.48–7.43 (m, 2H), 7.33 (ddd, *J* = 8.9, 6.8, 1.2 Hz, 1H), 6.94 (td, *J* = 6.8, 1.2 Hz, 1H), 6.69 (d, *J* = 9.4 Hz, 1H). ¹³C NMR (126 MHz, DMSO): δ 167.32, 160.73, 144.41, 143.38, 141.21, 140.86, 139.01, 137.62, 133.12, 132.42, 130.28, 129.69, 129.48, 129.23, 128.81, 127.32, 125.87, 125.10, 121.99, 116.95, 116.82, 112.87, 107.21. HRMS: calcd for [M + H], 474.06503; found, 474.06512.

Chemicals for Biological Experiments. Compound 1 (CITCO, (6-(4-chlorophenyl)imidazo [2,1-*b*] [1,3]thiazole-5-carbaldehyde-*O*-(3,4-dichloro-benzyl)oxime)), rifampicin (rif, TCPOBOP, and PK11195 were obtained from Sigma-Aldrich (St. Louis, Missouri, United States, now Merck), which is now known as Merck (Darmstadt, Germany). Phenobarbital (Luminal 200 mg/mL injection) was manufactured by Desitin Pharma spol. s.r.o. (Prague, Czech Republic). Ligands for nuclear receptors (GW3965, thyroxin, obeticholic acid, dexamethasone, fenofibrate, GWS01516, rosiglitazone, 3-methylcholanthrene, calcitriol, testosterone, and estradiol) were purchased from Sigma-Aldrich (now Merck). The prototype ligands were used at 100 nM (dexamethasone, calcitriol), 1 μM (thyroxin), or 10 μM concentrations.

The compounds were dissolved in DMSO, and the final concentration of DMSO in the entire reaction mixture or cultivation media was 0.1%.

Cell Culture. Human hepatocellular carcinoma HepG2 and monkey fibroblast-like COS-1 cell lines were cultured as we have described before.⁴⁸ All experiments were performed between passages 5–13 after thawing. CAR Knockout HepaRG and parent HepaRG cells were obtained from Sigma-Aldrich, now Merck (Darmstadt, Germany). The cell lines were cultivated and differentiated in the same manner on the 12-well plates. For each experiment, the HepaRG cells were seeded at a density of 26,600 cells/cm² and kept in William's medium supplemented with 5 μg/mL insulin, 50 μM hydrocortisone, 10% HyClone fetal serum (GE Healthcare Life Sciences, Pittsburgh, USA). 14 days after seeding, the HepaRG cells

were differentiated into hepatocyte-like cells using 1.5% DMSO in culture media for another 14 days.⁴⁹

LS174T, an epithelial Caucasian colon adenocarcinoma cell line, was obtained from (Merck Life Science spol. s r.o., Prague, 87060401-1VL). The line has functional PXR but very low CAR nuclear receptor expression.⁴³ The cell line has been cultivated for induction experiments as we described before.⁵⁰

Primary Human Hepatocytes. The PHHs (human hepatocytes in monolayer-long-term cultures) were obtained from Biopredic (Rennes, France) (batch HEP220965, 45-year-old female, Caucasian; HEP220966, 53-year-old female, Caucasian; HEP220969, 78-year-old female, Caucasian; HEP220971, 46-year-old male, Caucasian, HEP220976, 73-year-old male, African, HEP220980, 84-year-old male, Caucasian). The cells were cultivated according to the manufacturer's protocol. The PHHs were treated with selected test compounds for 24 or 48 h in the use medium at the concentrations of 1, 5, or 10 μ M. Cryopreserved human hepatocytes (HJK) were purchased from BioIVT (Westbury, New York, USA). PHHs were cultured in William's E medium (ThermoFisher Scientific, Waltham, Massachusetts, USA) supplemented with insulin (10 μ g/mL)–transferrin (5.5 μ g/mL)–sodium selenite (6.7 ng/mL) (Thermo Fisher Scientific), L-glutamine (2 mM)–penicillin (100 U/mL)–streptomycin (100 μ g/mL) (Sigma-Aldrich), 100 nM dexamethasone, and 10% fetal bovine serum (HyClone).

Western blotting experiments have been performed with total cellular or tissue lysates (20 μ g) with polyclonal anti-CYP2B6 (PA5-35032, dilution 1:1500), antibeta actin recombinant rabbit monoclonal antibody (MA5-32540, clone JF53-10), and CYP3A4 polyclonal antibody (PA1-343, 1:2000) (all from Thermo Fisher). For protein analysis, PHHs have been treated for 48 h.

RT-qPCR. RT-qPCR was used to examine CAR target gene expression in PHH, HepaRG cells, or in mouse liver samples. Total RNA, reverse transcription, and qPCR were performed, and mRNA expression data was analyzed as we have described before.⁵¹

All RT-qPCR experiments were performed in triplicate samples, and data are presented as fold induction to vehicle-treated cells with the same reagents we have described before.^{48,52} PCR TaqMan probes for murine genes have been listed in our previous reports.⁵² TaqMan probes for *CYP3A4*, *CYP2C9*, and *CYP2B6* human genes as well as for reference genes *B2M* and *GADPH* genes were obtained from Thermo Fisher: *CYP3A4* (Hs00604506_m1), *CYP2C9* (Hs02383631_s1), *CYP2B6* (Hs04183483_g1), *GADPH* (Hs02758991_g1), and *B2M* (Hs00984230_m1).

Luciferase Assays. A human CAR LBD assembly assay (CAR AA) was performed according to protocols published by Carazo and Pavek with two hybrid expression constructs encoding helices 3–12 (pCAR-C/VP16) and helix 1 (pCAR-N/GAL4) parts of human CAR LBD.⁵³ Cells were treated for 24 h with tested compounds (range of concentration from 1 nM up to 30 μ M). Data are presented as relative activity (%) to compound 1 (CITCO) at 10 μ M. The half-maximal effective concentration (EC_{50}) to activate CAR in the assay was calculated from at least six points of dose–response curves using the GraphPad Prism software.

Luciferase gene reporter assays to determine interactions with CAR and its variants or with PXR were performed as we have described before in HepG2 or COS-1 cells.^{39,48} Using these assays, the relative activation of the CAR3 variant in comparison with the activity of compound 1 (CITCO) at 1 μ M concentration or the relative value of PXR activation (% of rifampicin-mediated PXR activation at 10 μ M concentration) were determined (Tables 2, 4, and 7).

The CYP2B6-luc reporter plasmid (originally entitled B-1.6k/PB/XREM) was kindly donated by Dr. Hongbing Wang (University of Maryland School of Pharmacy, Baltimore, MD, USA) and was used in assays with all human CAR variants as well as with the mouse Car expression vector. Expression vectors (based on pcDNA3.1+/C-(K)-DYK vector) for CAR variant 3 (CAR3, 353 AA, CloneID OHu34914, XM_005245697.4, transcript variant X4, mRNA), CAR variant 2 (CAR2, 352 AA, Clone ID OHu10438, NM_001077480.2), and CAR wild type (wtCAR, 348 AA, Clone ID OHu09315, NM_005122.4, transcript variant 3) were purchased from Genscript

(Piscataway, NJ, USA). The mouse Car expression vector pCMV6-mCar (NM_009803) was obtained from OriGene Technologies, Rockville, MD, USA). Empty expression vectors were used in control experiments.

In addition, other expression constructs for a ligand-activated CAR transcription variant 3 (CAR3, pTracer-CMV2-CAR3) was a kind gift from Dr. C. J. Omiecinski (Pennsylvania State University, State College, PA, USA), which was used to validate our results with the commercial construct.

The CYP3A4 promoter luciferase reporter construct (p3A4-luc) and PXR expression vector for transient transfection luciferase assays have been described before.⁵⁰ The p3A4-luc plasmid bears a distal XREM (–7836/–7208) and a basal promoter sequence (prPXRE, –362/+53) from the *CYP3A4* gene promoter region.

Transient transfection experiments with various nuclear receptor-responsive luciferase assays have been performed as we described before with the same protocol and plasmids.^{54,55}

CYP Enzymatic Activity Assays. Human recombinant CYP3A4, CYP2B6, and CYP1A2 enzymes expressed from cDNA using baculovirus-infected insect cells with human CYP450 reductase and cytochrome b5 in a microsomal fraction (CYP450-Glo CYP3A4 Assay, CYP450-Glo CYP2B6 Assay, and CYP450-Glo CYP1A2 Assay, Promega, Hercules, CA) were used to evaluate the interaction of compound 39 with these enzymes in vitro according to protocols we published before.⁵⁶

TR-FRET CAR Coactivator Binding Assay. The LanthaScreen TR-FRET CAR Coactivator Binding Assay Kit (Thermo Fisher Scientific, Catr. No PV4836) with GST-tagged human CAR LBD and a fluorescein-labeled PGC1 α coactivator peptide was used with slight modifications of the manufacturer's protocol as we have reported before.⁵³ The half-maximal effective concentration (EC_{50}) to activate CAR LBD in the assay was calculated from at least six points (range of 10 pM to 10 μ M) from at least two experiments ($n = 2–3$) using the GraphPad Prism software.

Translocation Assay. Nuclear translocation of pEGFP-hCAR + Ala chimera in COS-1 SV40-transformed African green monkey kidney cells was performed as we have described before with the construct generated in the same report.⁴⁸ The method is a modification of the method originally described by Chen et al.⁴²

Animal Experiments. Humanized PXR-CAR-CYP3A4/3A7 mice (model 11585) were obtained from Taconic (Rensselaer, NY) and kept in a temperature-controlled and light-controlled facility with a 12 h light–dark cycling. All animals had free access to a commercially available laboratory chow diet (Velaz, Prague, Czech Republic). Male 9–14 week-old animals ($n = 4$ per group) were randomized into four groups (control; compound 39 1 mg/kg; compound 39 10 mg/kg; compound 1 10 mg/kg), and these compounds were administered as a single application intraperitoneally in a 5% glycerol formulation in saline. Animals were sacrificed 36 h after the administration, and livers were removed, weighted, and snap-frozen in liquid nitrogen for further total RNA isolation. All animal studies were performed in accordance with the European Directive 86/609/EEC, and they were approved by the Czech Central Commission for Animal Welfare.

Human and Mouse Plasma Protein Binding, Metabolic Stability in Human or Mouse Liver Microsomes, and Human Liver S9 Fraction. Protocols for the plasma protein binding assay and metabolic stability testing in human and mouse liver microsomes or the S9 fraction are described in the Supporting Information, Chapters 6 and 7.

Pharmacokinetic Study after Single-Dose Application. PK studies were performed in C57BL/6N male mice after 10 mg/kg application of compound 39 via either i.v. or gavage application ($n = 4$ per group) using HPLC-MS/MS analysis. Blood samples were taken in the following intervals: 10, 120, 240, 480, 720, and 1440 min. Detailed protocols and PK parameters calculation are described in the Supporting Information, Chapter 8.

Genotoxicity Testing and 7 Day Oral Toxicity Study in Rats. The protocols for the assays are described in the Supporting Information, Chapters 9 and 11.

hERG Fluorescence Polarization Assay. The hERG fluorescence polarization assay was performed as described in the [Supporting Information](#), Chapter 10.

In Silico Molecular Dynamics Analysis. *Molecular Modeling, Receptor and Ligand Preparation:* The crystal structure of the hCAR model was retrieved from the RCSB Protein Data Bank (www.rcsb.org) (PDB code: 1XVP).¹⁵

All ligands for docking were drawn using Maestro (2020.2) and prepared using LigPrep to generate the three-dimensional conformation, adjust the protonation state to physiological pH (7.4), and calculate the partial atomic charges, with the force-field OPLS3e. We employed a standard docking to accommodate the compounds 37, 39, 40, and 48 within the CAR's LBD (PDB ID: 1XVP; resolution: 2.0 Å, cocrystallized with compound 1,¹⁵ amino acid numbering follows the crystal structure), using Glide.⁵⁷ Ligands were docked within a grid around 12 Å from the centroid of the cocrystallized ligand generating 10 poses per ligand. To validate the docking obtained for test ligands, and also to evaluate the capability of the docking algorithm to locate the ligands within the LBD, we redocked the cocrystal ligand (compound 1, a full agonist) inside the CAR LBD. Next, the seven systems (four test compounds plus compound 1) were prepared and minimized by adding hydrogens, adjusting the protonation states of amino acids, and fixing missing side-chain atoms and protein loops using Maestro PrepWizard 2020.2. The molecular dynamics simulation protocol and respective analyses can be found in the [Supporting Information](#), Chapter 2. For each ligand, simulations of five 1 μs independent replicas were carried out, resulting in 25 μs worth of simulations for all five systems.

Statistical Analysis. Data are presented as the means and SD from at least three independent experiments ($n = 3$). A one-way analysis of variance (ANOVA) with Dunnett's *post hoc* test was applied. GraphPad Prism ver. 9.3.1. Software (GraphPad Software, Inc., San Diego, CA, United States) was used to perform statistical analysis.

EC₅₀ indicates the xenobiotic concentration required to achieve half-maximum activation, and relative E_{max} represents the overall maximal calculated activation produced by the tested compound (i.e., maximal efficacy). The activities of compound 1 and rifampicin at 10 μM were set to be 100% in the dose–response calculations. IC₅₀ represents the half-maximal inhibitory concentration in the viability MTT assay or in cytochrome P450 inhibition assays. A p -value of <0.05 was considered to be statistically significant.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jmedchem.2c01140>.

Cytotoxicity analysis of target compounds; in vitro and in vivo pharmacokinetic details of selected compounds; docking and molecular dynamics supporting data and procedures; additional toxicity studies of compound 39; supplementary experimental procedures; elemental analyses; and NMR and HPLC data ([PDF](#))

Molecular formula strings ([CSV](#))

■ AUTHOR INFORMATION

Corresponding Authors

Petr Pávek – Department of Pharmacology and Toxicology, Faculty of Pharmacy in Hradec Kralove, Charles University, 500 05 Hradec Kralove, Czech Republic;
Email: radim.nencka@uochb.cas.cz

Radim Nencka – Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, 166 10 Prague 6, Czech Republic; orcid.org/0000-0001-6167-0380;
Email: pavek@faf.cuni.cz

Authors

Ivana Mejdrová – Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, 166 10 Prague 6, Czech Republic

Jan Dušek – Department of Pharmacology and Toxicology, Faculty of Pharmacy in Hradec Kralove, Charles University, 500 05 Hradec Kralove, Czech Republic

Kryštof Skach – Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, 166 10 Prague 6, Czech Republic; orcid.org/0000-0002-7558-6961

Alžběta Stefela – Department of Pharmacology and Toxicology, Faculty of Pharmacy in Hradec Kralove, Charles University, 500 05 Hradec Kralove, Czech Republic

Josef Skoda – Department of Pharmacology and Toxicology, Faculty of Pharmacy in Hradec Kralove, Charles University, 500 05 Hradec Kralove, Czech Republic

Karel Chalupský – Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, 166 10 Prague 6, Czech Republic; Czech Centre for Phenogenomics, Institute of Molecular Genetics of the Czech Academy of Sciences, 142 20 Prague, Czech Republic

Klára Dohnalová – Czech Centre for Phenogenomics, Institute of Molecular Genetics of the Czech Academy of Sciences, 142 20 Prague, Czech Republic; 1st Medical Faculty, Charles University, 112 08 Prague, Czech Republic

Ivona Pavkova – Faculty of Military Health Sciences, University of Defense, 500 01 Hradec Kralove, Czech Republic; orcid.org/0000-0001-8241-479X

Thales Kronenberger – Department of Internal Medicine VIII, University Hospital of Tübingen, 72076 Tübingen, Germany; School of Pharmacy, Faculty of Health Sciences, University of Eastern Finland, 70211 Kuopio, Finland; Department of Pharmaceutical and Medicinal Chemistry, Institute of Pharmaceutical Sciences, Eberhard Karls Universität, 72076 Tübingen, Germany

Azam Rashidian – Department of Internal Medicine VIII, University Hospital of Tübingen, 72076 Tübingen, Germany

Lucie Smutná – Department of Pharmacology and Toxicology, Faculty of Pharmacy in Hradec Kralove, Charles University, 500 05 Hradec Kralove, Czech Republic

Vojtěch Duchoslav – Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, 166 10 Prague 6, Czech Republic

Tomas Smutny – Department of Pharmacology and Toxicology, Faculty of Pharmacy in Hradec Kralove, Charles University, 500 05 Hradec Kralove, Czech Republic

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.jmedchem.2c01140>

Author Contributions

○I.M. and J.D. contributed equally.

Funding

The research was supported by the Technology Agency of the Czech Republic (The Council of the National Centre of Competence-Personalized Medicine-Diagnostics and Therapy TN01000013), the Ministry of Defence of the Czech Republic (“Long Term Development Plan—Medical Aspects of Weapons of Mass Destruction” of the Faculty of Military Health Sciences, University of Defence, Hradec Kralove, Czech Republic), by InoMed project (CZ.02.1.01/0.0/0.0/18_069/0010046), Czech Scientific Agency (22-05167S to P.P.), and the Czech Academy of Sciences (RVO: 61388963). This work

was supported by the Czech Centre for Phenogenomics (CCP, project no. LM2015040) and OP RDI CZ.1.05/2.1.00/19.0395 (higher quality and capacity for transgenic models).

Notes

The authors declare no competing financial interest. Results of simulations, in the form of trajectory and interaction data, are available on the Zenodo repository (codes: 10.5281/zenodo.5772317).

ABBREVIATIONS

ACN, acetonitrile; ADME, absorption, distribution, metabolism, excretion; CAR AA, CAR ligand-binding domain assembly assay; CYP2B6, cytochrome P450 family two subfamily B member six; CYP3A4, cytochrome P450 family three subfamily A member four; DCM, dichloromethane; DMF, *N,N*-dimethylformamide; EC_{50} , concentration required to achieve half-maximum activation; EGFP, enhanced green fluorescent protein; E_{max} , maximal calculated activation; hERG, human Ether-à-go-go-Related Gene; LBD, ligand-binding domain; LBP, ligand-binding pocket; NIS, *N*-iodosuccinimide; ON, overnight; PHH, primary human hepatocyte; PXR, pregnane X receptor; SRC-1, steroid receptor coactivator one; TEA, triethylamine; THF, tetrahydrofuran; TLC, thin-layer chromatography; TMS, trimethylsilyl; TR-FRET, time-resolved fluorescence energy transfer

REFERENCES

- (1) Forman, B. M.; Tzamelis, I.; Choi, H. S.; Chen, J.; Simha, D.; Seol, W.; Evans, R. M.; Moore, D. D. Androstane metabolites bind to and deactivate the nuclear receptor CAR-beta. *Nature* **1998**, *395*, 612–615.
- (2) Moore, L. B.; Parks, D. J.; Jones, S. A.; Bledsoe, R. K.; Consler, T. G.; Stimmel, J. B.; Goodwin, B.; Liddle, C.; Blanchard, S. G.; Willson, T. M.; Collins, J. L.; Kliewer, S. A. Orphan nuclear receptors constitutive androstane receptor and pregnane X receptor share xenobiotic and steroid ligands. *J. Biol. Chem.* **2000**, *275*, 15122–15127.
- (3) Mackowiak, B.; Hodge, J.; Stern, S.; Wang, H. The Roles of Xenobiotic Receptors: Beyond Chemical Disposition. *Drug Metab. Dispos.* **2018**, *46*, 1361.
- (4) Dong, B.; Saha, P. K.; Huang, W.; Chen, W.; Abu-Elheiga, L. A.; Wakil, S. J.; Stevens, R. D.; Ilkayeva, O.; Newgard, C. B.; Chan, L.; Moore, D. D. Activation of nuclear receptor CAR ameliorates diabetes and fatty liver disease. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 18831–18836.
- (5) Gao, J.; He, J.; Zhai, Y.; Wada, T.; Xie, W. The constitutive androstane receptor is an anti-obesity nuclear receptor that improves insulin sensitivity. *J. Biol. Chem.* **2009**, *284*, 25984–25992.
- (6) Gao, J.; Xie, W. Targeting xenobiotic receptors PXR and CAR for metabolic diseases. *Trends Pharmacol. Sci.* **2012**, *33*, 552–558.
- (7) Jiang, M.; Xie, W. Role of the constitutive androstane receptor in obesity and type 2 diabetes: a case study of the endobiotic function of a xenobiotic receptor. *Drug Metab. Rev.* **2013**, *45*, 156–163.
- (8) Molnár, F.; Kublbeck, J.; Jyrkkärinne, J.; Prantner, V.; Honkakoski, P. An update on the constitutive androstane receptor (CAR). *Drug Metabol. Drug Interact.* **2013**, *28*, 79–93.
- (9) Režen, T.; Tamasi, V.; Lovgren-Sandblom, A.; Bjorkhem, I.; Meyer, U. A.; Rozman, D. Effect of CAR activation on selected metabolic pathways in normal and hyperlipidemic mouse livers. *BMC Genom.* **2009**, *10*, 384.
- (10) Marmugi, A.; Lukowicz, C.; Lasserre, F.; Montagner, A.; Polizzi, A.; Ducheix, S.; Goron, A.; Gamet-Payrastra, L.; Gerbal-Chaloin, S.; Pascussi, J. M.; Moldes, M.; Pineau, T.; Guillou, H.; Mselli-Lakhal, L. Activation of the Constitutive Androstane Receptor induces hepatic lipogenesis and regulates Pnpla3 gene expression in a LXR-independent way. *Toxicol. Appl. Pharmacol.* **2016**, *303*, 90–100.
- (11) Breuker, C.; Moreau, A.; Lakkhal, L.; Tamasi, V.; Parmentier, Y.; Meyer, U.; Maurel, P.; Lumbroso, S.; Vilarem, M. J.; Pascussi, J. M. Hepatic expression of thyroid hormone-responsive spot 14 protein is regulated by constitutive androstane receptor (NR113). *Endocrinology* **2010**, *151*, 1653–1661.
- (12) Maglich, J. M.; Lobe, D. C.; Moore, J. T. The nuclear receptor CAR (NR113) regulates serum triglyceride levels under conditions of metabolic stress. *J. Lipid Res.* **2009**, *50*, 439–445.
- (13) Baskin-Bey, E. S.; Anan, A.; Isomoto, H.; Bronk, S. F.; Gores, G. J. Constitutive androstane receptor agonist, TCPOBOP, attenuates steatohepatitis in the methionine choline-deficient diet-fed mouse. *World J. Gastroenterol.* **2007**, *13*, 5635–5641.
- (14) Tschuor, C.; Kachaylo, E.; Limani, P.; Raptis, D. A.; Linecker, M.; Tian, Y.; Herrmann, U.; Grabliauskaite, K.; Weber, A.; Columbano, A.; Graf, R.; Humar, B.; Clavien, P. A. Constitutive androstane receptor (Car)-driven regeneration protects liver from failure following tissue loss. *J. Hepatol.* **2016**, *65*, 66–74.
- (15) Xu, R. X.; Lambert, M. H.; Wisely, B. B.; Warren, E. N.; Weinert, E. E.; Waitt, G. M.; Williams, J. D.; Collins, J. L.; Moore, L. B.; Willson, T. M.; Moore, J. T. A structural basis for constitutive activity in the human CAR/RXRalpha heterodimer. *Mol. Cell* **2004**, *16*, 919–928.
- (16) Ingraham, H. A.; Redinbo, M. R. Orphan nuclear receptors adopted by crystallography. *Curr. Opin. Struct. Biol.* **2005**, *15*, 708–715.
- (17) Mackowiak, B.; Wang, H. Mechanisms of xenobiotic receptor activation: Direct vs. indirect. *Biochim. Biophys. Acta* **2016**, *1859*, 1130–1140.
- (18) Chai, S. C.; Cherian, M. T.; Wang, Y. M.; Chen, T. Small-molecule modulators of PXR and CAR. *Biochim. Biophys. Acta* **2016**, *1859*, 1141–1154.
- (19) Ross, J.; Plummer, S. M.; Rode, A.; Scheer, N.; Bower, C. C.; Vogel, O.; Henderson, C. J.; Wolf, C. R.; Elcombe, C. R. Human constitutive androstane receptor (CAR) and pregnane X receptor (PXR) support the hypertrophic but not the hyperplastic response to the murine nongenotoxic hepatocarcinogens phenobarbital and chlordane in vivo. *Toxicol. Sci.* **2010**, *116*, 452–466.
- (20) Chai, S. C.; Lin, W.; Li, Y.; Chen, T. Drug discovery technologies to identify and characterize modulators of the pregnane X receptor and the constitutive androstane receptor. *Drug Discov. Today* **2019**, *24*, 906–915.
- (21) Küblbeck, J.; Jyrkkärinne, J.; Molnár, F.; Kuningas, T.; Patel, J.; Windshügel, B.; Nevalainen, T.; Laitinen, T.; Sippl, W.; Poso, A.; Honkakoski, P. New in vitro tools to study human constitutive androstane receptor (CAR) biology: discovery and comparison of human CAR inverse agonists. *Mol. Pharm.* **2011**, *8*, 2424–2433.
- (22) Maglich, J. M.; Parks, D. J.; Moore, L. B.; Collins, J. L.; Goodwin, B.; Billin, A. N.; Stoltz, C. A.; Kliewer, S. A.; Lambert, M. H.; Willson, T. M.; Moore, J. T. Identification of a novel human constitutive androstane receptor (CAR) agonist and its use in the identification of CAR target genes. *J. Biol. Chem.* **2003**, *278*, 17277–17283.
- (23) Küblbeck, J.; Laitinen, T.; Jyrkkärinne, J.; Rousu, T.; Tolonen, A.; Abel, T.; Kortelainen, T.; Uusitalo, J.; Korjamo, T.; Honkakoski, P.; Molnár, F. Use of comprehensive screening methods to detect selective human CAR activators. *Biochem. Pharmacol.* **2011**, *82*, 1994–2007.
- (24) Lin, W.; Bwayi, M.; Wu, J.; Li, Y.; Chai, S. C.; Huber, A. D.; Chen, T. CITCO Directly Binds to and Activates Human Pregnane X Receptor. *Mol. Pharmacol.* **2020**, *97*, 180–190.
- (25) Hakkola, J.; Bernasconi, C.; Coecke, S.; Richert, L.; Andersson, T. B.; Pelkonen, O. Cytochrome P450 Induction and Xeno-Sensing Receptors Pregnane X Receptor, Constitutive Androstane Receptor, Aryl Hydrocarbon Receptor and Peroxisome Proliferator-Activated Receptor alpha at the Crossroads of Toxicokinetics and Toxicodynamics. *Basic Clin. Pharmacol. Toxicol.* **2018**, *123*, 42.
- (26) Tzamelis, I.; Pissios, P.; Schuetz, E. G.; Moore, D. D. The xenobiotic compound 1,4-bis[2-(3,5-dichloropyridyloxy)]benzene is

- an agonist ligand for the nuclear receptor CAR. *Mol. Cell. Biol.* **2000**, *20*, 2951–2958.
- (27) Li, H.; Chen, T.; Cottrell, J.; Wang, H. Nuclear translocation of adenoviral-enhanced yellow fluorescent protein-tagged-human constitutive androstane receptor (hCAR): a novel tool for screening hCAR activators in human primary hepatocytes. *Drug Metab. Dispos.* **2009**, *37*, 1098–1106.
- (28) Mackowiak, B.; Li, L.; Lynch, C.; Ziman, A.; Heyward, S.; Xia, M.; Wang, H. High-content analysis of constitutive androstane receptor (CAR) translocation identifies mosapride citrate as a CAR agonist that represses gluconeogenesis. *Biochem. Pharmacol.* **2019**, *168*, 224–236.
- (29) Imai, J.; Yamazoe, Y.; Yoshinari, K. Novel cell-based reporter assay system using epitope-tagged protein for the identification of agonistic ligands of constitutive androstane receptor (CAR). *Drug Metab. Pharmacokinet.* **2013**, *28*, 290–298.
- (30) Jyrkkärinne, J.; Windshügel, B.; Rönkkö, T.; Tervo, A. J.; Küblbeck, J.; Lahtela-Kakkonen, M.; Sippl, W.; Poso, A.; Honkakoski, P. Insights into ligand-elicited activation of human constitutive androstane receptor based on novel agonists and three-dimensional quantitative structure-activity relationship. *J. Med. Chem.* **2008**, *51*, 7181–7192.
- (31) Dring, A. M.; Anderson, L. E.; Qamar, S.; Stoner, M. A. Rational quantitative structure-activity relationship (RQSAR) screen for PXR and CAR isoform-specific nuclear receptor ligands. *Chem. Biol. Interact.* **2010**, *188*, 512–525.
- (32) Lynch, C.; Zhao, J.; Wang, H.; Xia, M. Quantitative High-Throughput Luciferase Screening in Identifying CAR Modulators. *Methods Mol. Biol.* **2016**, *1473*, 33–42.
- (33) Keminer, O.; Windshügel, B.; Essmann, F.; Lee, S. M. L.; Schiergens, T. S.; Schwab, M.; Burk, O. Identification of novel agonists by high-throughput screening and molecular modelling of human constitutive androstane receptor isoform 3. *Arch. Toxicol.* **2019**, *93*, 2247–2264.
- (34) Lynch, C.; Pan, Y.; Li, L.; Ferguson, S. S.; Xia, M.; Swaan, P. W.; Wang, H. Identification of novel activators of constitutive androstane receptor from FDA-approved drugs by integrated computational and biological approaches. *Pharm. Res.* **2013**, *30*, 489–501.
- (35) Burk, O.; Piedade, R.; Ghebregiorghis, L.; Fait, J. T.; Nussler, A. K.; Gil, J. P.; Windshügel, B.; Schwab, M. Differential effects of clinically used derivatives and metabolites of artemisinin in the activation of constitutive androstane receptor isoforms. *Br. J. Pharmacol.* **2012**, *167*, 666–681.
- (36) Liang, D.; Li, L.; Lynch, C.; Diethelm-Varela, B.; Xia, M.; Xue, F.; Wang, H. DL5050, a Selective Agonist for the Human Constitutive Androstane Receptor. *ACS Med. Chem. Lett.* **2019**, *10*, 1039–1044.
- (37) Liang, D.; Li, L.; Lynch, C.; Mackowiak, B.; Hedrich, W. D.; Ai, Y.; Yin, Y.; Heyward, S.; Xia, M.; Wang, H.; Xue, F. Human constitutive androstane receptor agonist DL5016: A novel sensitizer for cyclophosphamide-based chemotherapies. *Eur. J. Med. Chem.* **2019**, *179*, 84–99.
- (38) Stern, S.; Liang, D.; Li, L.; Kurian, R.; Lynch, C.; Sakamuru, S.; Heyward, S.; Zhang, J.; Kareem, K. A.; Chun, Y. W.; Huang, R.; Xia, M.; Hong, C. C.; Xue, F.; Wang, H. Targeting CAR and Nrf2 improves cyclophosphamide bioactivation while reducing doxorubicin-induced cardiotoxicity in triple-negative breast cancer treatment. *JCI Insight* **2022**, *7*, No. e153868.
- (39) Smutny, T.; Nova, A.; Drechslerová, M.; Carazo, A.; Hyrsova, L.; Hrušková, Z. R.; Kuneš, J.; Pour, M.; Špulák, M.; Pavek, P. 2-(3-Methoxyphenyl)quinazoline Derivatives: A New Class of Direct Constitutive Androstane Receptor (CAR) Agonists. *J. Med. Chem.* **2016**, *59*, 4601–4610.
- (40) Li, L.; Chen, T.; Stanton, J. D.; Sueyoshi, T.; Negishi, M.; Wang, H. The peripheral benzodiazepine receptor ligand 1-(2-chlorophenyl-methylpropyl)-3-isoquinoline-carboxamide is a novel antagonist of human constitutive androstane receptor. *Mol. Pharmacol.* **2008**, *74*, 443–453.
- (41) Cherian, M. T.; Lin, W.; Wu, J.; Chen, T. CINPA1 is an inhibitor of constitutive androstane receptor that does not activate pregnane X receptor. *Mol. Pharmacol.* **2015**, *87*, 878–889.
- (42) Chen, T.; Tompkins, L. M.; Li, L.; Li, H.; Kim, G.; Zheng, Y.; Wang, H. A single amino acid controls the functional switch of human constitutive androstane receptor (CAR) 1 to the xenobiotic-sensitive splicing variant CAR3. *J. Pharmacol. Exp. Ther.* **2010**, *332*, 106–115.
- (43) Burk, O.; Arnold, K. A.; Nussler, A. K.; Schaeffeler, E.; Efimova, E.; Avery, B. A.; Avery, M. A.; Fromm, M. F.; Eichelbaum, M. Antimalarial artemisinin drugs induce cytochrome P450 and MDR1 expression by activation of xenosensors pregnane X receptor and constitutive androstane receptor. *Mol. Pharmacol.* **2005**, *67*, 1954–1965.
- (44) Shan, L.; Vincent, J.; Brunzelle, J. S.; Dussault, I.; Lin, M.; Ianculescu, I.; Sherman, M. A.; Forman, B. M.; Fernandez, E. J. Structure of the murine constitutive androstane receptor complexed to androstenol: a molecular basis for inverse agonism. *Mol. Cell* **2004**, *16*, 907–917.
- (45) Suino, K.; Peng, L.; Reynolds, R.; Li, Y.; Cha, J. Y.; Repa, J. J.; Kliewer, S. A.; Xu, H. E. The nuclear xenobiotic receptor CAR: structural determinants of constitutive activation and heterodimerization. *Mol. Cell* **2004**, *16*, 893–905.
- (46) Trapani, G.; Franco, M.; Ricciardi, L.; Latrofa, A.; Genchi, G.; Sanna, E.; Tuveri, F.; Cagetti, E.; Biggio, G.; Liso, G. Synthesis and binding affinity of 2-phenylimidazo[1,2- α]pyridine derivatives for both central and peripheral benzodiazepine receptors. A new series of high-affinity and selective ligands for the peripheral type. *J. Med. Chem.* **1997**, *40*, 3109–3118.
- (47) Palagiano, F.; Arenare, L.; Luraschi, E.; Caprariis, P.; Abignente, E.; D'Amico, M.; Filippelli, W.; Rossi, F. Research on heterocyclic compounds. XXXIV. Synthesis and SAR study of some imidazo[2,1-b]thiazole carboxylic and acetic acids with antiinflammatory and analgesic activities. *Eur. J. Med. Chem.* **1995**, *30*, 901–909.
- (48) Skoda, J.; Dusek, J.; Drastik, M.; Stefela, A.; Dohnalova, K.; Chalupsky, K.; Smutny, T.; Micuda, S.; Gerbal-Chaloin, S.; Pavek, P. Diazepam Promotes Translocation of Human Constitutive Androstane Receptor (CAR) via Direct Interaction with the Ligand-Binding Domain. *Cells* **2020**, *9*, 2532.
- (49) Hyrsova, L.; Smutny, T.; Carazo, A.; Moravcik, S.; Mandikova, J.; Trejtnar, F.; Gerbal-Chaloin, S.; Pavek, P. The pregnane X receptor down-regulates organic cation transporter 1 (SLC22A1) in human hepatocytes by competing for (“squelching”) SRC-1 coactivator. *Br. J. Pharmacol.* **2016**, *173*, 1703–1715.
- (50) Pavek, P.; Pospeschova, L.; Svecova, L.; Syrova, Z.; Stejskalova, L.; Blazkova, J.; Dvorak, Z.; Blahos, J. Intestinal cell-specific vitamin D receptor (VDR)-mediated transcriptional regulation of CYP3A4 gene. *Biochem. Pharmacol.* **2010**, *79*, 277–287.
- (51) Skoda, J.; Dusek, J.; Drastik, M.; Stefela, A.; Dohnalova, K.; Chalupsky, K.; Smutny, T.; Micuda, S.; Gerbal-Chaloin, S.; Pavek, P. Diazepam Promotes Translocation of Human Constitutive Androstane Receptor (CAR) via Direct Interaction with the Ligand-Binding Domain. *Cells* **2020**, *9*, 2532.
- (52) Dusek, J.; Skoda, J.; Holas, O.; Horvatova, A.; Smutny, T.; Linhartova, L.; Hirsova, P.; Kucera, O.; Micuda, S.; Braeuning, A.; Pavek, P. Stilbene compound trans-3,4,5,4-tetramethoxystilbene, a potential anticancer drug, regulates constitutive androstane receptor (Car) target genes, but does not possess proliferative activity in mouse liver. *Toxicol. Lett.* **2019**, *313*, 1–10.
- (53) Carazo, A.; Pávek, P. The Use of the LanthaScreen TR-FRET CAR Coactivator Assay in the Characterization of Constitutive Androstane Receptor (CAR) Inverse Agonists. *Sensors* **2015**, *15*, 9265–9276.
- (54) Stefela, A.; Vrzal, R.; Pavek, P. (E)-7-ethylidene-lithocholic acid (7-ELCA) is a potent dual farnesoid X receptor (FXR) antagonist and GPBAR1 agonist inhibiting FXR-induced gene expression in hepatocytes and stimulating glucagon-like peptide-1 secretion from enteroendocrine cells. *Front. Pharmacol.* **2021**, *12*, 1980.
- (55) Dvořák, Z.; Vrzal, R.; Pavek, P.; Ulrichova, J. An evidence for regulatory cross-talk between aryl hydrocarbon receptor and

glucocorticoid receptor in HepG2 cells. *Physiol. Res.* **2008**, *57*, 427–435.

(56) Dusek, J.; Carazo, A.; Trejtnar, F.; Hyrsova, L.; Holas, O.; Smutny, T.; Micuda, S.; Pavek, P. Steviol, an aglycone of steviol glycoside, sweeteners, interacts with the pregnane X (PXR) and aryl hydrocarbon (AHR) receptors in detoxification regulation. *Food Chem. Toxicol.* **2017**, *109*, 130–142.

(57) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196.

Recommended by ACS

Imidazo[1,2-*a*]Pyridine Derivatives as Novel Dual-Target Inhibitors of ABCB1 and ABCG2 for Reversing Multidrug Resistance

Hui Li, Xiao-Li Xu, *et al.*

FEBRUARY 13, 2023
JOURNAL OF MEDICINAL CHEMISTRY

READ 

Discovery of 4-(1,2,4-Oxadiazol-5-yl)azepan-2-one Derivatives as a New Class of Cannabinoid Type 2 Receptor Agonists for the Treatment of Inflammatory Pain

Jinshan Nan, Shengyong Yang, *et al.*

FEBRUARY 23, 2023
JOURNAL OF MEDICINAL CHEMISTRY

READ 

Development and Characterization of Fluorescent Probes for the G Protein-Coupled Receptor 35

Lai Wei, Xinmiao Liang, *et al.*

FEBRUARY 22, 2023
ACS MEDICINAL CHEMISTRY LETTERS

READ 

Development of a Covalent Inhibitor of c-Jun N-Terminal Protein Kinase (JNK) 2/3 with Selectivity over JNK1

Wenchao Lu, Nathanael S. Gray, *et al.*

FEBRUARY 24, 2023
JOURNAL OF MEDICINAL CHEMISTRY

READ 

Get More Suggestions >

11. Publication V [Submitted Manuscript]

Filling the blank space: Branched 4-nonylphenols isomers are responsible for 2 robust constitutive androstane receptor (CAR) activation by nonylphenol

Azam Rashidian^{1,‡}, Jan Dusek^{2,‡}, Martin Drastik², Lucie Smutná², Kristin Fritsche³,
Albert Braeuning³, Antti Poso^{1,4}, Thales Kronenberger^{4*}, Petr Pavek^{2*}

¹ Department of Internal Medicine VIII, University Hospital of Tübingen, 72076, Tübingen, Germany.

² Department of Pharmacology and Toxicology, Faculty of Pharmacy in Hradec Kralove, Charles University, Akademika Heyrovskeho 1203, 500 05, Hradec Kralove, Czech Republic.

³ German Federal Institute for Risk Assessment, Department Food Safety, Max-Dohrn-Str. 8-10, 10589 11 Berlin, Germany.

⁴ **a.** School of Pharmacy, Faculty of Health Sciences, University of Eastern Finland, 70211, Kuopio, Finland. **b.** Department of Pharmaceutical and Medicinal Chemistry, Institute of Pharmaceutical Sciences, Eberhard-Karls-Universität, Tuebingen, Auf der Morgenstelle 8, 72076 Tuebingen, Germany.

c. Tuebingen Center for Academic Drug Discovery & Development (TüCAD2), 72076 Tuebingen, Germany. **d.** Excellence Cluster "Controlling Microbes to Fight Infections" (CMFI), 72076 Tübingen, Germany.

‡ These authors equally contributed to this work

Content not peer-reviewed by ChemRxiv. License: CC BY-NC-ND 4.0

<https://doi.org/10.26434/chemrxiv-2023-6jvnf> ORCID: <https://orcid.org/0000-0001-6933-7590>

Hypothesis: The high potency of mixed NPs on CAR isoforms could be attributed to branched 4-NPs.

Aims: We aim to enhance the safe application of 4-nonylphenols (4-NPs) compounds across diverse industries by understanding the mechanistic aspects underlying CAR activation by these compounds.

Results: In our study, we have demonstrated that branched 4-nonylphenols (4-NPs) exhibit more stable binding to activate both the wild-type (wt)CAR1 and the CAR3 variant LBDs during molecular dynamic (MD) simulations. Interestingly, compared to the linear 4-NP or tert-nonylphenol (22NP), the branched 4-NP exhibited a higher efficacy in activating both CAR3 and wtCAR1 LBDs. To further investigate the effects of these compounds in a cellular context, we conducted experiments using HepaRG cells. Remarkably, all tested NP compounds exhibited a noteworthy up-regulation of CYP2B6, which serves as a relevant hallmark indicative of CAR activation. The findings from our simulation analysis revealed the critical roles of Helices $\alpha 3$ and $\alpha 5$ in the interaction with wtCAR1 and CAR3. Furthermore, it is worth noting that none of the investigated ligands directly interact with α AF-2, the region associated with receptor activation and agonism. Nevertheless, upon analyzing the protein-ligand interaction data, we find strong evidence suggesting that all tested compounds demonstrate favorable interactions in both CAR3-LBD and CAR1-LBD. These results are consistent with the assembly assay data. Furthermore, our extended simulations unveil insights into the AF-2 region's behavior. Regardless of the specific ligand, the helix within the AF-2 region has a significant potential for conformational changes and dynamic shifts between different states. These observations highlight the importance of considering longer timescales in our analyses to capture the complete representation of NP binding dynamics, as relying solely on short conformational changes may lead to an incomplete understanding of the process. Furthermore, the assessment of free binding energy corroborates the high binding affinity to NP, which is comparable to CITCO.

Conclusion: Previously, NP, a mixture of isomers, has demonstrated potent endocrine-disrupting properties, capable of binding to ER, PXR, wtCAR1, and rodent CAR. Our research aims to delve into the intricate mechanisms underlying CAR activation by 4-NP. In this study, we conducted molecular dynamic (MD) simulations

and cellular experiments, utilizing various CAR constructs, CAR ligand-binding domain (LBD) mutants, and differentiated HepaRG hepatocyte cellular models. The goal was to investigate the interactions of individual branched (22NP, 33NP, and 353NP) and linear 4-NPs with CAR and its variants. This work presents the first comprehensive description of the interactions between individual 4-NP isomers and the human CAR receptor, including its dominant variant CAR3. By gaining a deeper understanding of the molecular mechanisms behind CAR activation by specific 4-NP isomers, we hope to contribute to the development of safer practices involving NP.

Filling the blank space: Branched 4-nonylphenols isomers are responsible for robust constitutive androstane receptor (CAR) activation by nonylphenol

Azam Rashidian^{1,‡}, Jan Dusek^{2,‡}, Martin Drastik², Lucie Smutná², Kristin Fritsche³, Albert Braeuning³, Antti Poso^{1,4}, Thales Kronenberger^{4*}, Petr Pavěk^{2*}

¹ Department of Internal Medicine VIII, University Hospital of Tübingen, 72076, Tübingen, Germany.

² Department of Pharmacology and Toxicology, Faculty of Pharmacy in Hradec Kralove, Charles University, Akademika Heyrovskeho 1203, 500 05, Hradec Kralove, Czech Republic.

³ German Federal Institute for Risk Assessment, Department Food Safety, Max-Dohrn-Str. 8-10, 10589 Berlin, Germany.

⁴ **a.** School of Pharmacy, Faculty of Health Sciences, University of Eastern Finland, 70211, Kuopio, Finland. **b.** Department of Pharmaceutical and Medicinal Chemistry, Institute of Pharmaceutical Sciences, Eberhard-Karls-Universität, Tuebingen, Auf der Morgenstelle 8, 72076 Tuebingen, Germany. **c.** Tuebingen Center for Academic Drug Discovery & Development (TüCAD₂), 72076 Tuebingen, Germany. **d.** Excellence Cluster "Controlling Microbes to Fight Infections" (CMFI), 72076 Tübingen, Germany

‡ These authors equally contributed to this work

* corresponding authors:

T.K.: thales.kronenberger@uni-tuebingen.de, Auf der Morgenstelle 8, DE72076, Tübingen (Germany).
P.P.: pavek@faf.cuni.cz, Akademika Heyrovskeho 1203, 500 05, Hradec Kralove, Czech Republic.

ABSTRACT

4-Nonylphenol (4-NP), a *para*-substituted phenolic compound mixture comprising a straight or branched carbon chain, is a widespread ubiquitous environmental pollutant and food contaminant. 4-NP, particularly the branched form, has been identified as an endocrine disruptor with potent activities on estrogen nuclear receptors. Constitutive Androstane Receptor (CAR, NR1H3) is another crucial nuclear receptor that regulates hepatic lipid, glucose, and steroid metabolism and is involved in endocrine disruption mechanism. The nonylphenol mixture has been described as an extremely potent activator of both human and rodent CAR. However, detailed mechanistic aspects of CAR activation by NP are enigmatic and it is not known if 4-NP can directly interact with the CAR ligand binding domain (LBD).

Here we examined interactions of individual branched (22NP, 33NP, and 353NP) and linear 4-NPs with CAR and its variants using molecular dynamics (MD) simulations, cellular experiments with various CAR constructs or CAR ligand binding domain (LBD) mutants, or in differentiated HepaRG hepatocyte cellular model.

Our results demonstrate that branched 4-NPs display more stable poses to activate both wild-type (wt)CAR1 as well as CAR3 variant LBDs in MD simulations. Consistently, branched 4-NP activated CAR3 and wtCAR1 LBD more efficiently than linear 4-NP or *tert*-nonylphenol (22NP). Furthermore, in HepaRG cells, we observed that all NP significantly up-regulated CYP2B6, a relevant hallmark for CAR activation. This is the first study to provide detailed insights into the direct interaction between individual 4-NPs and the human CAR-LBD, as well as its dominant variant CAR3. The work could contribute to the safer use of individual NPs in many areas of industry.

Keywords: constitutive androstane receptor, CAR ligand binding domain, Endocrine disrupting chemicals, Molecular Dynamic simulations, 4-Nonylphenol.

1. INTRODUCTION

Endocrine disruptors (or EDs) are structurally diverse, often lipophilic substances capable of bioaccumulation (e.g. persistent organic pollutants), which due to our

54 industrialized world are present in most of our daily products. Examples include
55 plasticizers, such as alkylphenol ethoxylates (Hernandez et al., 2007), pesticides
56 (Heindel et al., 2015), fungicides, and various polyhalogenated organic compounds,
57 which are present in consumer products, the environment, or exposures from industrial
58 sources (Kublbeck et al., 2020b). EDs may interfere with our homeostasis by mimicking
59 endocrine hormones. The consequence of these environmental contaminants on
60 human health is well described: disruption of the endocrine system function, which
61 leads to adverse hormonal or metabolic health effects on an organism level (Casals-
62 Casas and Desvergne, 2011; Gore et al., 2015; Heindel et al., 2015; Papalou et al.,
63 2019; Yilmaz et al., 2020).

64 Since the liver and adipose tissue are the main organs where the EDs accumulation
65 takes place and many EDs are nuclear receptor ligands, insulin resistance and
66 alteration of lipid, cholesterol, and bile acid metabolism are common consequences of
67 their intoxication (Heindel et al., 2017). Exposure to EDs can cause metabolic disorders
68 such as type 2 diabetes and non-alcoholic fatty liver disease (Shin et al., 2013; Yang
69 et al., 2016). In this regard, European Commission has funded the European Cluster
70 projects to improve the identification of EDs (Kublbeck et al., 2020a).

71
72 Among one of the most common plasticizers currently in use are those from the
73 nonylphenol (or NP) series. NP is derived mainly from the hydrolytic degradation of
74 alkyl phenol ethoxylate compounds, which are used as non-ionic surfactants in a range
75 of industrial agricultural, and domestic applications, principally as emulsifying, wetting,
76 dispersing, surfactant, or stabilising agents. Nonylphenol is widely used in food
77 packaging, hair dyes, other personal care products, textile printing and dyes, pesticides
78 and other agricultural-related substances, and non-ionic surfactants. Owing to its low
79 solubility, NP is prone to accumulate in environmental matrices, such as in water, soils
80 and sediments (Li et al., 2019).

81 Nonylphenol's widespread presence in the environment is illustrated in a study that
82 found NP to be one of three groups of organic wastewater contaminants that together
83 contributed to almost 80% of the total measured concentration of organic wastewater
84 contaminants in U.S. streams (Kolpin et al., 2002).

85
86 NP's high lipophilicity allows its persistence in the environment and NP thus
87 accumulates in adipose tissues affecting both human and animal health (Muller et al.,
88 1998). Chemically, NP is a mixture of *para*-, *ortho*-, and *meta*-isomers of substituted
89 phenolic compounds with a straight or branched carbon chain, when *para*-NP (or 4-
90 NP) comprising approximately 90% of the isomers.

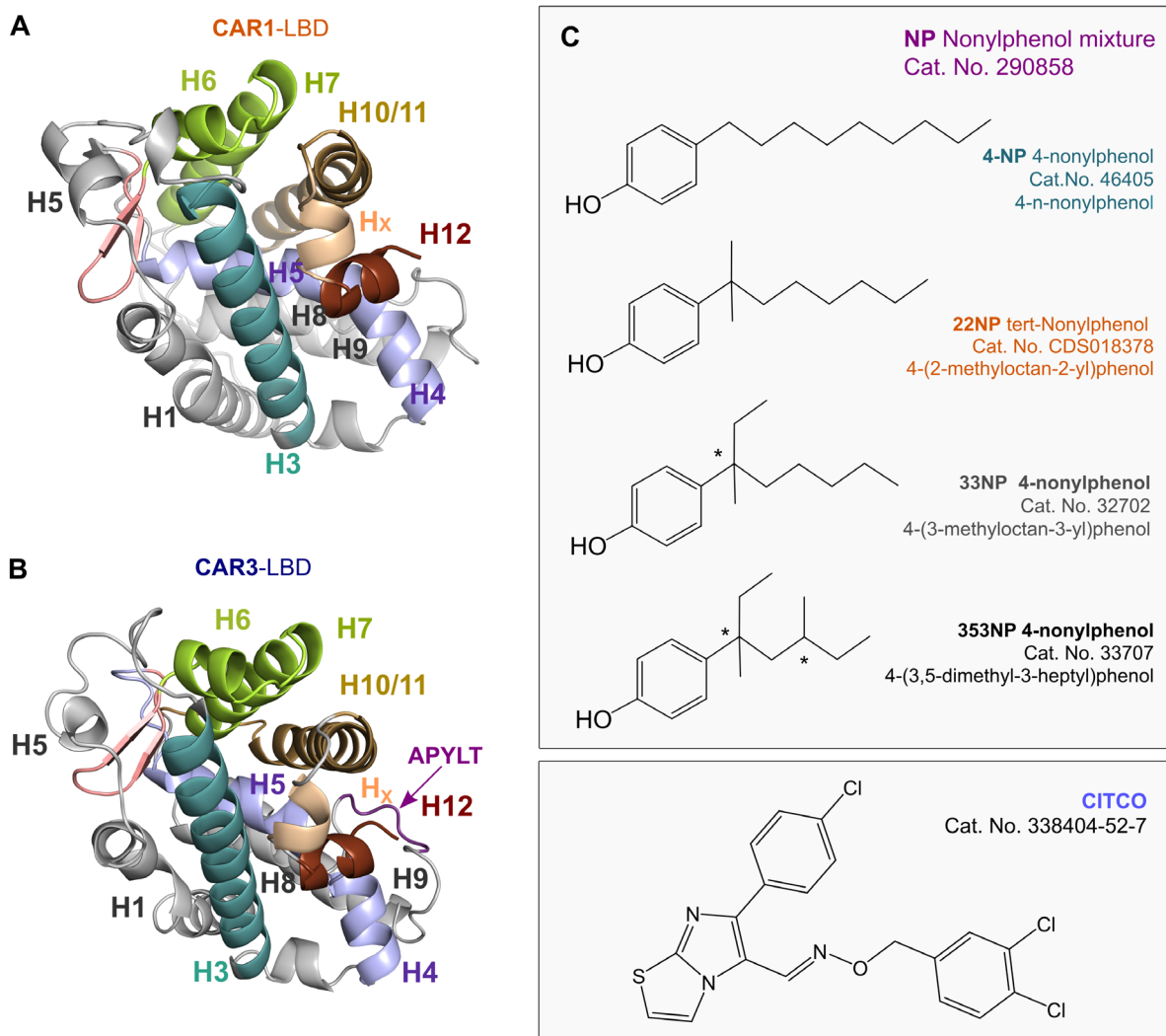
91 A recent review on infant food, cereal and nuts showed 4-NP quantities greater than
92 10-100 µg/kg (Bhandari et al., 2021). However, analysis of the number and effect of
93 the individual isomers are not systematic. This can be due to the lack of reference
94 standards to quantitatively assess these isomers. In terms of reference standards of
95 branched 4-NP, only a few of branched 4-NP are commercially available namely:
96 22NP, 33NP, 353NP, 363NP, 3E2 and 44NP (Bhandari et al., 2021; Lu et al., 2021b).
97 The NP has been reported as a highly effective activator of rodent Constitutive
98 Androstane Receptor (CAR, NR1I3), but also of human and rodent Pregnane X
99 Receptor (PXR, NR1I2) and ERs (Kretschmer and Baldwin, 2005). NP interaction with
100 ER and its consequences are well known. NP increases the risk of estrogen-dependent
101 neoplasms such as ovarian cancer, and breast and uterine cancer (Bhandari et al.,
102 2021; Lu et al., 2021b). In males, NP decreases fertility and sperm viability (Bhandari
103 et al., 2021; Lu et al., 2021b). Aquatic organisms, such as fishes and phytoplankton
104 also suffer from NP acute toxicity (Bhandari et al., 2021)(Hong and Li, 2007) in which

105 NP can be incorporated into the food chain. NP is also known to act in the hippocampus
106 and frontal cortex resulting in impaired cognitive function due to induced inflammatory
107 and apoptotic processes and interfering with neurotransmission secretion (Lu et al.,
108 2021b). Despite NP's overall relevance, a health-based guideline value has yet to be
109 internationally established (Chung, 2021; Li et al., 2019).

110
111 CAR nuclear receptor is one of the main drug- and xenobiotic-responsive transcription
112 factors regulating the expression of detoxification enzymes, conjugation enzymes and
113 transporters involved in the elimination of endogenous and exogenous substances,
114 such as bilirubin, bile acids, and xenobiotics (di Masi et al., 2009; Kublbeck et al.,
115 2020a). In addition to this, CAR has an important role in the metabolism of glucose,
116 and lipids, in energy homeostasis as well as in the endobiotic metabolism of steroid
117 hormones (Mackowiak et al., 2018). Many environmental toxicants or industrial
118 chemicals as well as some drugs have been identified as CAR activators (Mejdova et
119 al., 2023). Considering the significant role of CAR in hepatic metabolism regulation, it
120 has been intensively studied as the molecular target involved in endocrine disruption
121 and metabolic disturbances (Kublbeck et al., 2020a).

122 CAR has a typical nuclear receptor structure consisting of the DNA binding domain
123 (DBD) connected to the ligand binding domain (LBD, Fig. 1A,B) via a hinge region.
124 CAR interacts with the promoter response elements of target genes with its N -terminal
125 DBD. Human CAR has at least three transcript variants in the liver, which differ in their
126 ligand-dependent activation and basal activity. The wild-type variant of CAR
127 (wtCAR1/CAR1, accession number: NM_005122.4, transcript variant 3) features high
128 constitutive activity in the regulation of target genes. This variant represents about 40%
129 of CAR transcripts in the liver. The variant CAR3 (CAR-SV2, XM_005245697.4,
130 transcript variant X4), which has an insertion of the five amino acids APYLT into the
131 loop between H8-H9, represents 50% of transcripts. CAR3 has low constitutive activity
132 but is highly inducible by ligands. Finally, the transcript variant CAR2
133 (NM_001077480.2) is a minor variant with moderate induction activity (Chai et al.,
134 2016; Mackowiak and Wang, 2016; Molnar et al., 2013; Ross et al., 2010). Both direct
135 LBD-dependent and LBD-independent activation are known for CAR, highlighting a
136 complex response pathway (Mackowiak and Wang, 2016).

137 Only two human CAR crystal structures bound to agonists have been reported (Xu et
138 al., 2004), displaying a large hydrophobic and flexible binding pocket (Ingraham and
139 Redinbo, 2005; Molnar et al., 2013). CAR-LBD consists of approximately 250 amino
140 acid residues (Chai et al., 2016), comprised of three β -strands and eleven α -helices
141 (Xu et al., 2004). Of note, CAR contains the activation function-2 (AF-2), where
142 coregulatory proteins can bind, and the CAR-specific single-turn helix (H_x), which
143 restricts the H12 position resulting in the constitutive activity of this receptor (Xu et al.,
144 2004). In its active state, CAR-LBD also heterodimerizes with RXR α -LBD (Suino et al.,
145 2004).



146
147
148 **Fig. 1.** (A) Overview of the wtCAR1-ligand binding domain (CAR1-LBD) and (B) CAR3-LBD. The
149 insertion loop is highlighted in black colour shown by the magenta arrow (APYLT). The regions of interest
150 are highlighted as follows: H3 (residues 157–178), pale green; H5 (residues 196–209), light blue; β
151 sheets (residues 217–233), pink; H10/H11 (residues 311–333 in CAR1 and 316-338 in CAR3), light
152 brown; Hx-helix (residues 336–339 in CAR1 and 341–344 in CAR3), light orange; H12 (residues 341–
153 348 in CAR1 and 346–348 in CAR3), dark brown. (C) small molecule ligands used in this study, upper:
154 nonylphenol and isomers, bottom: CITCO control.

155
156 Hernandez *et al.*, (2007) (Hernandez *et al.*, 2007) showed that NP (a mixture of
157 nonylphenols) is a potent mouse CAR activator, and this was demonstrated *in vitro* in
158 transactivation assays and *in vivo* in wild-type and transgenic CAR knockout mice.
159 Later, the same group reported that 4-NP (technical grade, ~85% *para*-isomers)
160 regulates key CAR-target genes *Cyp2b10*, *Cyp2c29* and *Cyp3a11* in a gender-specific
161 manner in wild-type females, but not in wild-type males or CAR-null mice
162 demonstrating male mice resistant to CAR-mediated effects of 4-NP (Hernandez *et al.*,
163 2009). The gender-specific effects of NP mixture in the regulation of main cytochrome
164 P450 enzymes in males and females have been described before as feminizing profile
165 in males and masculinizing profile in females in testosterone hydroxylation (Hernandez
166 *et al.*, 2006).
167 Meanwhile, Dring *et al.* (2010) (Dring *et al.*, 2010) found that 4-NP (mixture of *para*
168 isomers), significantly activates human wild-type CAR as well as CAR3 variant.
169 However, DeKeyser *et al.* (2011) (DeKeyser *et al.*, 2011) reported that the mixture of

170 ring and chain NP isomers activates only wild-type CAR variant in luciferase reporter
171 assays with androstanol (to suppress high basal activity), but not human CAR2 or
172 CAR3 variants.

173 Recently, Kamata *et al.*, found in yeast cells transduced with the human CAR that an
174 unspecified 4-NP mixture of isomers (CAS 84852-15-3) is by two orders of magnitude
175 more potent to activate CAR than linear 4-NP (CAS 104-40-5 or also known 4-*n*-NP)
176 (Kamata *et al.*, 2018). Importantly, 4-*tert*-octylphenol was the most potent compound
177 in the study (Kamata *et al.*, 2018).

178 Nowadays it is clear that CAR1 and CAR3 can be activated by different ligands
179 (Keminer *et al.*, 2019; Mejdrova *et al.*, 2023). Due to the low availability of crystal
180 structures of human CAR1 and, to date, no crystal structures of human CAR3, the
181 discussion on this chemically diverse set of ligands remains a challenge. Meanwhile,
182 docking approaches are limited as they are unable to capture CAR's characteristic LBD
183 flexibility. Therefore, molecular dynamics (MD) simulations were successfully
184 employed to study the binding of newly developed human CAR ligands (Mejdrova *et*
185 *al.*, 2023).

186 In this study, we focus on elucidating the binding mechanism of four 4-NP isomers into
187 CAR1- and CAR3-LBDs using extended MD simulations. By analysing the
188 conformational dynamics, we compare these changes to the well-known CAR agonist
189 CITCO. We provide evidence that aliphatic 4-NPs are direct CAR agonist. Our findings
190 also reveal that branched 4-NPs adopt more favourable poses for activating both
191 wtCAR1 and CAR3 LBDs in MD simulations. Moreover, consistently with the simulation
192 results, branched 4-NPs exhibits more efficient activation of CAR3 and wtCAR1 LBDs
193 compared to linear 4-NP. We thus demonstrate that 4-NP are novel environmental
194 ligands of CAR which interact directly with CAR LBDs.

195

196 **2. MATERIALS AND METHODS**

197 **2.1. Model generation and structure preparation**

198 We modelled the systems with Maestro (Schrödinger Release 2022.4 Maestro,
199 Schrödinger, LLC, New York, NY, 2022), and OPLS4 force-field (Lu *et al.*, 2021a),
200 unless otherwise stated. CAR3 homology model was generated using the CAR1
201 crystal structure (PDB ID: 1XVP (Xu *et al.*, 2004), chain D) as a template. CAR3-LBD
202 model was generated by inserting the five amino acids APYLT in a loop connecting
203 H8- and H9-helices (so called L:H8-H9) utilizing the Advanced homology modelling
204 tool in Schrödinger suite (v2022.4). The missing side chain of inserted residues was
205 placed using Prime, followed by loop refinement using the same software (Jacobson
206 *et al.*, 2004).

207 The proteins were prepared using Protein Preparation Wizard (Schrödinger LLC, New
208 York, NY, 2022). Missing hydrogen atoms were added, bond orders were assigned
209 using CCD database, and protonation states of amino acids were optimized with
210 PROPKA (Schrödinger, LLC, New York, NY, 2022) at pH 7.4. Both CAR1 and CAR3-
211 LBDs were optimized in terms of hydrogen positions, utilizing the H-bond assignment
212 algorithm PROPKA in the Protein Preparation Wizard tool of Maestro, to select the
213 most likely protonation states and tautomer for the Histidine residues. We agreed with
214 the software suggestions, followed by optimizing the generated H-bonding species.
215 Finally, the structure was globally minimized using the steep descent method.

216 Ligands were prepared using LigPrep (Schrödinger, LLC, New York, NY, 2022) to
217 assign the protonation state (Epik; at pH 7.0±2.0) and the partial charges. Isomers'
218 chiral centre configurations were retrieved from the literature using their respective
219 CAS numbers (Fig. 1C).

220

221 **2.2. Molecular docking and pose selection**

222 To obtain the starting configuration for CAR1 systems, Glide docking was conducted
223 (Glide v. 7.7)(Friesner et al., 2004). For docking, we used default settings and defined
224 residues with 10 Å around CITCO crystallographic structure for the binding site.
225 Docking was conducted using standard precision (SP). Redocking of CITCO was
226 conducted as the reference pose. The docking in CAR1-LBD resulted in poses mainly
227 accommodated in hydrophobic subpocket and the representative pose was selected
228 based on the glide docking score and glide energy (Table S1). A similar docking
229 method was implemented for CAR3-LBD systems as well and the representative poses
230 were selected accordingly (Table S2). The docking results from the SP generated two
231 poses for 4-NP within CAR3-LBD with similar scores and energy but in different
232 orientations, namely 4-NP₁ and 4-NP₂. We evaluated these two distinguishable poses
233 in short MD simulations, and both poses displayed good stability during the simulation
234 (data not shown). Hence, we included both poses in our study. For all the other
235 isomers, highest scoring docking poses were visually inspected and the ones sharing
236 similarities with the CITCO complex were selected to undergo MDs.

237

238 **2.3. Molecular dynamics simulations**

239 We simulated the monomeric CAR without the coactivator peptide in the AF-2 region,
240 with a similar protocol as previously described (Mejdrova et al., 2023). We used the
241 Desmond MD simulation engine (Bowers et al., 2006) and the OPLS4 force-field (Lu
242 et al., 2021a). The prepared systems were solvated in a cubic box with the size of the
243 box set as 13 Å minimum distance from the box edges to any atom of the protein.
244 TIP3P water model (Jorgensen et al., 1983) was used to describe the solvent and the
245 net charge was neutralized using Na⁺ ion with a final salt concentration of 150 mM.
246 RESPA integrator timesteps of 2 fs for bonded and near and 6 fs for far were applied.
247 The short-range coulombic interactions were treated using a cut-off value of 9.0 Å,
248 whereas long-range coulombic interactions were estimated using the Smooth Particle
249 Mesh Ewald (PME) method (Darden et al., 1993). Before the production simulations,
250 the systems were relaxed using the default Desmond relaxation protocol. Simulations
251 were run in NPT ensemble, with a temperature of 310 K (using the Nosé-Hoover
252 thermostat, (Hoover, 1985; Nosé, 1984)) and pressure of 1.01325 bar (Martyna-
253 Tobias-Klein barostat, (Martyna et al., 1996)). For each system, simulations of five
254 were carried out, resulting in 5 μs simulation data for each system (Fig. S1).

255

256 **2.4. Analysis of MD simulation trajectories**

257 **2.4.1. Protein-ligand interactions and protein properties.** Maestro simulation
258 interaction analysis tool (Schrödinger, LLC) was used for the analysis of RMSD, RMSF,
259 and interaction analysis. We used default values for interaction which are H-bonds:
260 cut-off of 2.5 Å for donor and acceptor atoms, donor angle of 120° and acceptor angle
261 of 90°. Hydrophobic interactions: cut-off of 3.6 Å between ligand's aromatic or aliphatic
262 carbons and a hydrophobic side chain, π-π interaction was defined as two aromatic
263 groups stacked face-to-face or face-to-edge. Water bridge interactions: default cut-off
264 of 2.8 Å for donor and acceptor atoms, donor angle of 110° and acceptor angle of 90°.
265 For angle and distance calculations, the Maestro event analysis tool (Schrödinger,
266 LLC) was used. Distances between specific secondary structure elements were
267 calculated using their centres of mass with the Maestro
268 script *trj_asl_distance.py* (Schrödinger LLC). For the helix H3 centre of mass of
269 residues 157–178 and H12 centre of mass of residues 341–348 in CAR1 and 346–353
270 in CAR3 were used.

271

272 **2.4.2. MM-GBSA binding energy calculations.** Molecular mechanics with
273 generalized Born and surface area (MM-GBSA) predicts the binding free energy of
274 protein-ligand complexes (Li et al., 2011). The ligands' ranking based on the free
275 energy could be correlated to the experimental binding affinities, especially in a
276 congeneric series. Every 10th frame from the simulations was considered for the
277 calculations. These were used as input files for the MM-GBSA calculations with
278 thermal_mmgsa.py script from Schrödinger package. Calculated free-binding
279 energies (kcal/mol) are represented by the MM-GBSA and normalized by the number
280 of heavy atoms (HAC), according to the following formula: ligand efficiency =
281 $\text{Ln}(\text{binding energy}) / (1 + \text{Ln}(\text{HAC}))$.
282

283 **2.5. Reporter gene assays**

284 **2.5.1. Cells and cell cultivation.** HepG2 cells were purchased from the European
285 Collection of Authenticated Cell Cultures (Salisbury, UK) and cultured at 37 °C in a 5%
286 CO₂ atmosphere in antibiotic-free Dulbecco's Modified Eagle Medium (ThermoFischer
287 Scientific, Waltham, MA, United States) supplemented 10% foetal bovine serum (FBS,
288 Merck, Darmstadt, Germany) and sodium-pyruvate (Merck, Darmstadt, Germany).
289

290 **2.5.2. Plasmids.** CYP2B6-luc reporter plasmid (originally entitled B-1.6k/PB/XREM)
291 was kindly donated by Dr. Hongbing Wang (University of Maryland School of
292 Pharmacy, Baltimore, MD, USA). The luciferase construct was used with all human
293 CAR variants and with the mouse Car expression vector. Expression vectors (based
294 on pcDNA3.1+/C-(K)-DYK vector) for CAR variant 3 (CAR3, 353 amino acids, CloneID
295 OHu34914, XM_005245697.4, transcript variant X4 mRNA), CAR variant 2 (CAR2,
296 352 AA, Clone ID OHu10438, NM_001077480.2), and CAR wild type (CAR1, wtCAR1,
297 348 AA, Clone ID OHu09315, NM_005122.4, transcript variant 3) were purchased from
298 Genscript (Piscataway, NJ, USA). The mouse Car expression vector pCMV6-mCar
299 (NM_009803) were obtained from OriGene Technologies, Rockville, MD, USA).
300 pGAL4-CAR+AAA construct encoding for human CAR-LBD with three extra alanine
301 residues and L206S mutant have been described before (Skoda et al., 2020a).
302

303 **2.5.3. Transient Transfection and Luciferase Gene Reporter Assays.** All transient
304 transfection gene reporter assays were carried out using Lipofectamine™ 3000
305 (ThermoFischer Scientific, Waltham, MA, United States), according to the
306 manufacturer's protocol. The HepG2 or COS-1 cells were seeded at a density of
307 40,000 cells/cm² on 48-well plates. Usually, 100 ng/well of an expression vector and
308 150ng/well of the p2B6-luc luciferase reporter construct were used in the experiments
309 (Skoda et al., 2020b). Human CAR LBD assembly assay has been performed
310 according to protocols published by Carazo and Pavék with two hybrid expression
311 constructs encoding helices 3–12 (pCAR-C/VP16) and helix 1 (pCAR-N/GAL4) parts
312 of human CAR LBD (Carazo and Pavék, 2015). Cells were treated for 24 h with tested
313 compounds at concentrations 10 and 30 μM for the following 24 h in normal media with
314 FBS or with charcoal-stripped FBS (in the case of CAR2 variant experiments). CITCO
315 was used as the prototype ligand for human CAR variants (10 μM) and TCPOBOP as
316 a murine CAR ligand (10 μM). All transient transfection assays were normalized with
317 the *Renilla reniformis* luciferase transfection control plasmid (pRL-TK, 30 ng/well).
318 After the 24 h incubation with the tested compounds, the cells were washed with PBS
319 and lysed with the passive lysis buffer (Promega). Luminescence activity was analysed
320 using the Dual luciferase detection kit (Promega). Relative activation of the luciferase
321 construct was expressed as a percentage of the vehicle (0.1% DMSO)-treated

322 samples (100%). In all experiments, results are presented as means and SD from at
323 least three independent experiments (n=3) performed in triplicates.

324
325 **2.5.4. HepaRG cultivation.** HepaRG cell cultivation started 4 weeks before the
326 experiments took place. It was divided into 2 weeks of proliferation and 2 weeks of
327 differentiation. The cells were seeded onto 12-wells plates at the density of 1,000,000
328 cells/well. Proliferation was achieved in Williams E medium with 2 mM L-Glutamine,
329 10% Foetal bovine serum, 100 U/ml penicillin and 100 µg/ml streptomycin, 0.05%
330 human insulin and 50 µM hydrocortison-hemisuccinate. The differentiation medium
331 consisted proliferation medium with an additional DMSO of 1.7%. Two days before the
332 treatment the medium was changed to a treatment medium that consisted of 2% FBS
333 and 0.5% DMSO as we described before (Hyrsova et al., 2016). HepaRG cells were
334 treated with NP mixture or its isomers for 24 h for RT-PCR experiments.

335
336 **2.5.5. AdipoRed assays.** The triglyceride accumulation was measured with AdipoRed
337 assay (AdipoRed Adipogenesis Assay Reagent, PT-7009, Lonza, Basel, Switzerland)
338 as described previously by Lasch *et al.*, (2021) (Lasch et al., 2021). We used model
339 LXR agonist GW3965 (Merck) at 10 µM as a positive control inducing lipid
340 accumulation in HepaRG cells, which were treated for 72h.

341
342 **2.6. Statistical analyses and image generations.** Unless stated otherwise in the
343 figure legends, data are presented as the mean ± standard deviation (SD). A one-way
344 analysis of variance (ANOVA) with a Dunnett's *post hoc* test or Bonferroni test was
345 applied. GraphPad Prism 9.5.1 Software (GraphPad Software, Inc., San Diego, CA,
346 USA) was used to perform statistical analysis. A *p*-value of < 0.05 was considered to
347 be statistically significant.

348 Structural data visualization was conducted with PyMOL v.2.5.2 (Schrödinger LLC,
349 New York, NY, USA). Data visualization was also completed by Python 3.7, seaborn
350 (v0.12.2), matplotlib (Hunter, 2007) and GraphPad Prism (v. 9.5.2 for Windows,
351 GraphPad Software, San Diego, CA, USA).

352 353 **3. RESULTS**

354 **3.1. Nonylphenol and its isomer derivatives can activate CAR1 and CAR3 by** 355 **direct interaction with the ligand binding domain**

356 We assessed the interaction of NP, 4-NP and its isomers with wtCAR1, inducible CAR3
357 variant, and CAR2 variant in several assays. Additionally, the wtCAR1 ligand-binding
358 domain (LBD) assembly assay and mouse CAR (mCAR) assay in transiently
359 transfected HepG2 cells were also used. Indeed, we found that NP is a highly potent
360 agonist of CAR in all assays that we used (Fig. 2).

361 The assembly assay relies on the idea that a compound binding to the ligand binding
362 pocket would be able to stabilize the two portions of CAR LBD into a single functional
363 entity (Fig. 2A). In wtCAR1-LBD assembly assay, the NP mixture activated CAR at
364 10 µM more efficiently than the model agonist CITCO, at 10 µM (Fig. 2B), which
365 correlates with the dose-response experiments (Fig. 2C), where CITCO was slightly
366 less potent (EC₅₀: 0.67 µM, CI 95%: 0.41 – 1.54) than nonylphenol is (EC₅₀: 0.20 µM,
367 0.03 – N.D., upper limit not defined due to lack of saturation). Interestingly, all NP
368 isomers, except the linear 4-NP, can induce the wtCAR1 assembly (at 1 µM), with the
369 more branched isomer (353NP) or larger (22NP) inducing the higher fold changes in
370 comparison to the control (FC: 295.8-fold and 377.6, respectively).

371 In next experiments, we employed wtCAR1 expression vector to overexpress wtCAR1
372 in HepG2 cells together with the p2B6-luc construct with natural responsive CYP2B6

373 gene promotor sequences. This assay better reflect the wtCAR1-mediated
374 transcriptional machinery; however, it is less sensitive due to spontaneous
375 translocation and high basal activation of wtCAR1 in the nucleus of HepG2 cells. In
376 terms of the overall activity of the wtCAR1 variant, monitored by the reporter gene
377 assay (RGA), both CITCO and NP mixture can activate the receptor, at 10 μ M to a
378 similar extent, but not at 1 μ M (Fig. 2D). In terms of specific isomers, only 22NP
379 significantly activates wtCAR1 at 10 μ M.

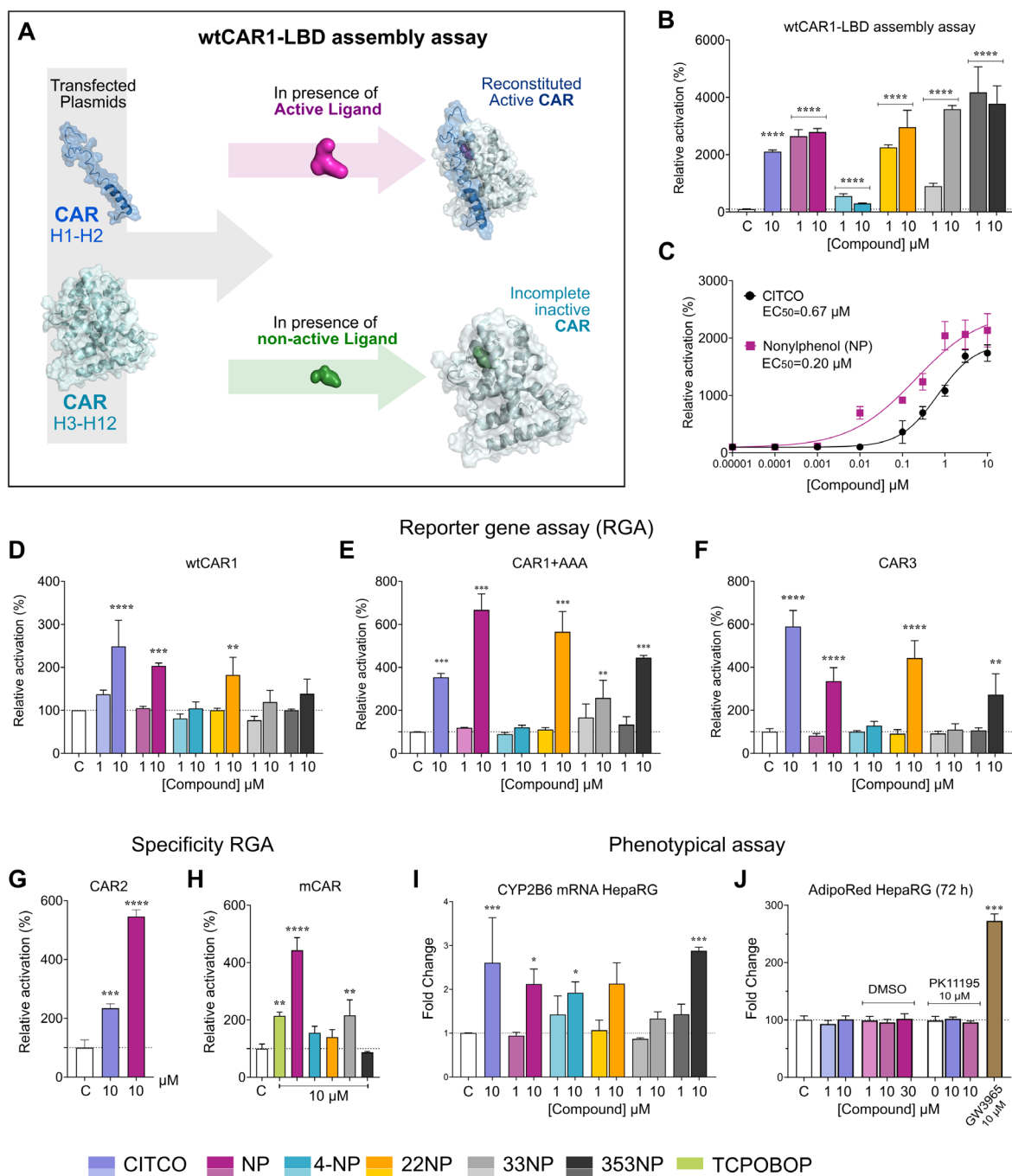
380 Alternatively, when assessing the activation of highly inducible CAR1+AAA (an artificial
381 mutant, Fig. 2E) and CAR3 (a naturally occurring, Fig. 2F) variants, we observe that a
382 higher activation by 22NP (56.5-fold for CAR1+AAA in comparison to control and 44.3-
383 fold for CAR3) followed by modest activation by 353NP (FC: 44.5-fold for CAR1+AAA
384 and 27.3 for CAR3). Overall, CAR3 variant showed similar results to wtCAR1, but the
385 activation was 10-fold by CITCO (Fig. 2F). This is unsurprising, given that CAR3 is the
386 non-constitutively active variant of CAR, and it is therefore well inducible by its
387 agonists.

388 CAR2 variant was activated by CITCO 10 μ M 2.34-fold and by nonylphenol 10 μ M
389 5.46-fold. With this variant, nonylphenol is a more potent agonist than CITCO (Fig. 2G).
390 Interestingly, mouse CAR was activated by NP mixture and 33NP at 10 μ M
391 concentration (Fig. 2H). Therefore, we can conclude that NP mixture and 33NP are
392 agonists of both human and murine CAR, despite interspecies differences, whereas
393 22NP and 353NP are strong human CAR agonists.

394
395 In next induction experiments, we analysed if NP, 4-NP's isomers and CITCO up-
396 regulate the typical CAR target CYP2B6 gene mRNA in differentiated HepaRG cells.
397 We found that there is significant induction of CYP2B6 mRNA expression in
398 differentiated HepaRG cells by NP mixture, 4-NP and 353NP (at 10 μ M, Fig. 2I), which
399 is consistent with CAR activation in RGA assay.

400 In a pilot AdipoRed experiment, we observed no intracellular triglyceride accumulation
401 after treatment with CAR activators, while cells treated with GW3965, an LXR agonist,
402 at 10 μ M had a 2.5-fold increase in lipid accumulation (Fig. 2J). Monitoring liver
403 triglyceride accumulation *in vitro* can function as an indicator of hepatic fatty acid
404 changes (Lasch et al., 2021). Such an endpoint could summarize the adverse outcome
405 induced by an endocrine disruptor acting on multiple nuclear receptors. This result
406 suggests that NP mixture would not induce/increase the lipid accumulation within the
407 tested timescale, however, we cannot disregard effects from longer incubation times.

408
409
410
411



412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427

Fig. 2. Nonylphenol and its derivatives active wtCAR1/CAR1 and CAR3 by directly interacting with the ligand binding domain. A) Illustration of the assembly assay principle where two independent fragments of CAR (H1-H2 and H3-H12) are co-transfected, allowing for active ligands to reconstitute the full CAR, in its active state. (B-C) The CAR-LBD assembly assay was performed in COS-1 cells treated with NP and its isomers. (D-H) The reporter gene luciferase assay of different CAR constructs was assessed in HepG2 cells. Luciferase reporter gene assays were performed in HepG2 cells transiently transfected with either human wtCAR1 (D), CAR+AAA (E), CAR3 (F), CAR2 (G) variants or mice CAR (H) expression vectors and RXR α construct and an appropriate responsive luciferase reporter promoter construct (p2B6-luc) together with *Renilla* expression construct for transfection normalization (see methods). Cells were treated with NP mixture, 4-NP, 22NP, 33NP, 353NP (1 or 10 μ M) together with CAR direct agonist CITCO (10 μ M) and, for mice CAR agonist TCPOBOP (10 μ M), for 24 h after transfection. HepaRG cells were used to monitor the expression levels of CYP2B6 mRNA using RT-PCR expression analysis after 48 h treatment. (J) In the AdipoRed triglycerides accumulation study (J), cells were treated with vehicle or (DMSO, 0.01%) co-treated with PK11195 (10 μ M) together with NP

428 (from 1 to 30 μM) and CITCO (from 1 to 10 μM) for 72 h. Relative activation (%) is depicted as relative
429 fold activation to control the samples (DMSO 0.1%) and data are presented as the means \pm SD from
430 three independent experiments ($n = 3$) performed in triplicates. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$,
431 statistically significant effects of CITCO or the tested NPs to control the cells.

432
433 These RGA and expression data altogether suggest that NP, 22NP and 353NP are
434 direct activators of wtCAR1 and CAR3. Therefore, herein, we discuss the binding
435 mechanism of four 4-NP isomers into wtCAR1- and CAR3-LBDs using a total of 45 μs
436 worth of simulations to disclose the conformational dynamics, comparing these
437 changes to the well-known CAR agonist CITCO (10 μs worth of simulations). These
438 trajectories were analysed in terms of ligand stability within the pocket, protein-ligand
439 interactions and binding energy, as well as geometrical changes between relevant
440 parts of the receptor.

441 442 **3.2 Both CAR3 and wtCAR1 NP-bound complexes display binding stability.**

443 Our initial docking approaches for the linear 4-NP yielded one main conformation in
444 CAR1 (Fig. 3A, highlighted by orange squares) and two potential conformations on
445 CAR3, namely 4-NP₁ and 4-NP₂, in the CAR3 binding site (Fig. 3B, blue squares). To
446 assess the stability of the ligands within the ligand binding pocket (LBP), we conducted
447 MD simulations for each pose and examined movement indicators.

448 First, we employed root mean square deviation (RMSD) as an indicator of ligand
449 stability relative to the protein. RMSD is a valid indicator where it indicates how stable
450 the ligand is with respect to the protein. Our result of RMSD calculation over the MD
451 trajectories revealed the stability of all studies ligands inside the LBP (Supporting
452 information, Fig. S2-S5). Notably, CAR-CITCO shows the highest stability
453 (CAR1/CAR3's protein RMSD: 1.6/2.3 Å, supporting information, Fig. S2, S3), while
454 the linear 4-NP shows the highest RMSD values and amplitude (CAR1/CAR3's protein
455 RMSD: 2.03/2.7 Å). The branched 4-NPs systems demonstrate comparable stability
456 as CAR-CITCO does (22NP, CAR1/CAR3's protein RMSD: 1.8/2.5; 33NP,
457 CAR1/CAR3's protein RMSD: 1.9/2.3 and 353 NP, CAR1/CAR3: 1.97/2.4).

458
459 Additionally, we considered two other properties, the solvent-accessible surface area
460 (SASA) and the radius of gyration (R_g), as indicators of ligand stability and molecular
461 compactness/extendedness, respectively. The ligands exhibited comparable ranging
462 from 3.4 to 3.8 Å (Fig. S6, S7) except for CITCO which displayed a larger R_g value of
463 around 4.4 Å in both CAR isoforms. This could be related to the molecular
464 weight of the ligand since R_g is directly related on it. In terms of SASA, all tested ligands
465 exhibit higher values (ranging from 5.5 to 7.2 Å²), in comparison to CITCO (~0.3 to
466 0.6 Å², Supporting information, Fig. S8). This finding could be related to nonylphenol's
467 higher flexibility within the pocket, instead of its reportedly apolar character. The
468 analysis of ligand stability within the pocket prompted us to further investigate the
469 protein-ligand interactions.

470 471 **3.3 Helices $\alpha 3$ and $\alpha 5$ play important role in NP interaction to wtCAR1 and CAR3.**

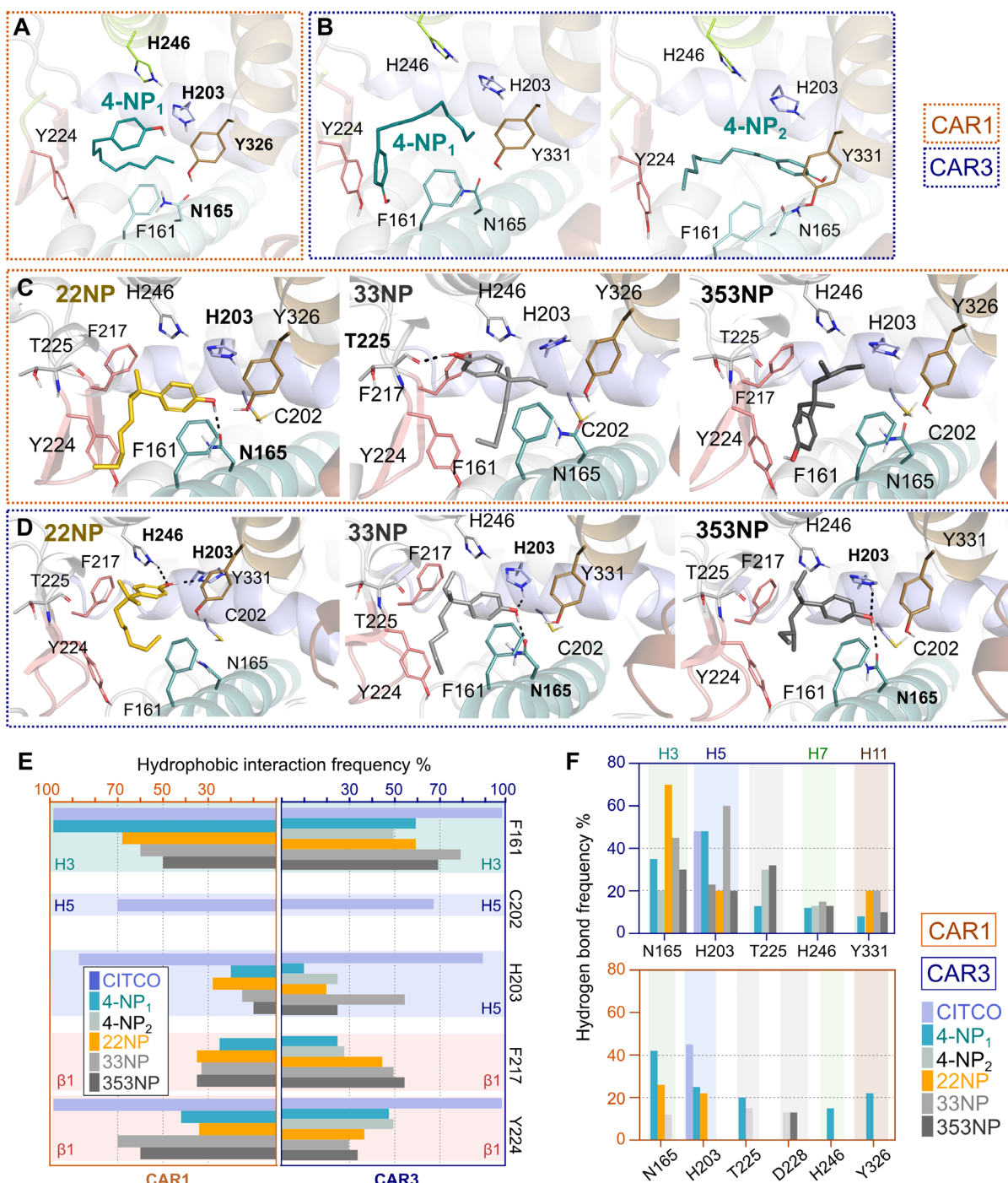
472 As mentioned, a representative snapshot of the 4-NP, proposing its binding mode,
473 reveals a single relevant conformation for CAR1 (Fig. 3A, highlighted by orange
474 squares) and two potential binding modes, namely 4-NP₁ and 4-NP₂, in the CAR3
475 binding site (Fig. 3B, blue squares). On the other hand, the most populated cluster for
476 the NP isomers, namely 22NP, 33NP and 353NP exhibits a variety of conformations
477 in CAR1 (Fig. 3C) and a more consistent scenario for CAR3, with most of the poses
478 resembling 4-NP₂ (Fig. 3D).

479 Furthermore, analysis of protein-ligand interactions suggests that CAR1- and CAR3
480 simulations share similar key residues for the binding, F161 (H3), H203 (H5), F217 and
481 Y224 (β -sheet), which exhibit all interaction frequencies of >60-100% of the analysed
482 simulation time (Fig. 3E). Interestingly, Y224 interaction seems to stabilize linear NP
483 isomers in CAR3, while displaying a preference for branched NPs in CAR1. Several
484 non-polar residues, including L206, F217, F234, and L242 also contribute to ligand
485 bindings but to a lesser extent (Supporting information, Table. S3).

486 In general, the binding mode of 4-NPs' isomers is further influenced by polar residues:
487 N165, H203, T225, H246 and Y326 (Y331 in CAR3), which are involved in interactions
488 with all nonylphenol compounds. In contrast, CITCO only forms H-bond with H203
489 about 45% of the simulation time (Fig. 3F). The contribution of H203 for 33NP binding
490 is the highest (60%) comparable to the linear 4-NP (50%). In addition, N165 shows the
491 highest contribution for 22NP stability.

492 Overall, the stability of CITCO relies on hydrophobic interactions and to a lesser extent
493 the H203 hydrogen bond interaction. Both linear and branched nonylphenols
494 maintained stability through a combination of hydrophobic interactions and some
495 hydrogen bonds (Fig. 3; Table S3, S4). Hydrophobic interactions with F161 and F217
496 are the most frequent, followed by hydrogen bonds to a lesser degree. There are few
497 system-specific relevant hydrogen bond interactions, such as the contribution of T225
498 in 4-NP₂ and 353NP, and H246 and Y331 in 4-NP₁ in CAR3.

499 It is noteworthy that none of the studied ligands directly interacts with the α AF-2, which
500 is associated with receptor activation and agonism. However, protein-ligand interaction
501 data supports the idea that all tested compounds would favourably interact in CAR3-
502 LBD as well as in CAR1-LBD, which is consistent with the assembly assay data.
503



504
505
506
507
508
509
510
511
512
513
514
515
516
517
518

Fig. 3. The representative snapshots of the binding mode of compounds complex with CAR1 (orange squares) and CAR3 (blue) in individual systems and their related protein interactions. 4-NP generated a single relevant conformation for CAR1 (A) and two potential binding modes 4-NP₁ and 4-NP₂ in the CAR3 binding site (B). Most populated cluster for the NP isomers 22NP, 33NP and 353NP in CAR1 (C) and CAR3 (D). Hydrogen bonds are depicted as dashed lines and the secondary structure is coloured as in Fig. 1A, B. Frequencies of protein-ligand interactions both hydrophobic interactions (E) including π - π interactions (face to face and face to edge) and protein-ligand hydrogen bond (F).

3.4 NP stabilizes the AF-2 active conformation in CAR1 and CAR3 isoforms.

The AF-2 region is critical for coactivator recruitment. In most NR, H12 is positioned via a bound ligand. Though the high basal activity of CAR is independent of H12 position, several mutations in this region abolished CAR activation demonstrating its importance for activity (Dussault and Forman, 2002; Frank et al., 2004). The AF-2 region is stabilized by CAR-specific structural elements, such as the helix X (H_x,

519 residues 336–339), which is a two-turn helix tightly packed between H3 and H10/H11
520 (Fig. 4A and insets). Dusselt *et al.*, 2002 showed that the insertion of three amino acids
521 into the CAR H_x abolished CAR basal activity demonstrating the importance of this
522 helix's configuration for AF-2 stabilization (Dussault and Forman, 2002).

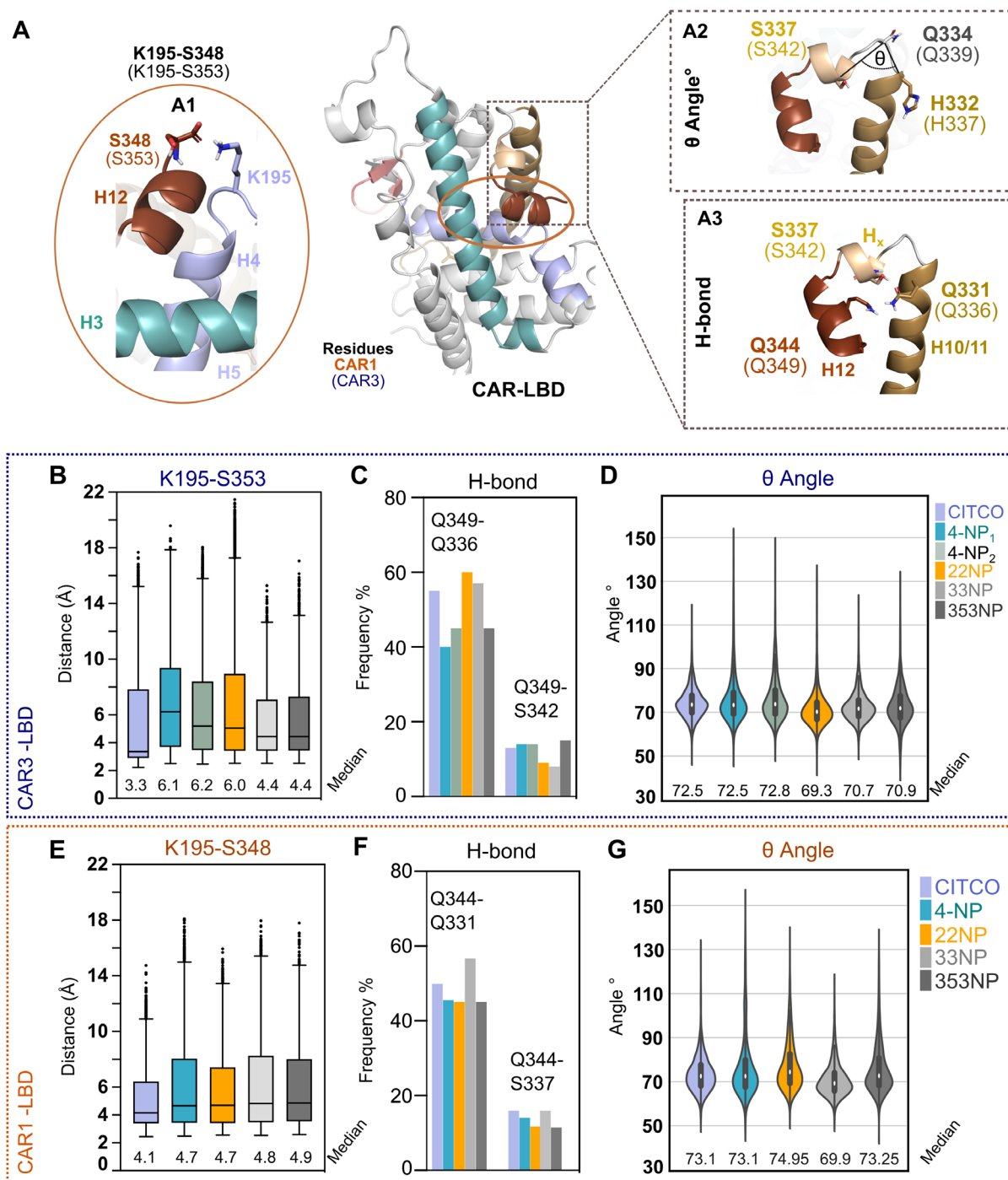
523
524 The H-bond interaction between the side chain of K195 (on H4) and C-terminal
525 carbonyl of S348 (S353 in CAR3) plays a role in locking the AF-2 in an active
526 conformation (Fig. 4A, Inset A₁). To investigate this interaction, we compared the
527 distance between these residues in the presence of linear and branched NPs in both
528 wtCAR1 and CAR3 isoforms. CAR3-CITCO exhibited the smallest distance,
529 approximately 3.3 Å (Fig. 4B). This was, closely followed by 33NP and 353NP (median
530 distance ~4.4 Å). The largest distances belong to 4-NP₁ (6.2 Å) and 4-NP₂ (5.2 Å).

531 This geometry enables hydrogen bond formation between H4 and H12 regions in
532 CAR3, which can be monitored for discussing agonism. Suino *et al.* 2004) reported
533 that in the presence of an inverse agonist, androstanol, H10/H11 dissociates and forms
534 a short helix (H_x) (Suino *et al.*, 2004). Consistently, TCPOBOP stabilizes H12 in the
535 active conformation (Wright *et al.*, 2011), where the two segments of H10 associates
536 together. Accordingly, mutations in this region, such as alanine substitutions of E339,
537 L340 and L343 decrease the CAR basal activity level (Wright *et al.*, 2011).

538
539 To further explore the conformation of H10/H11 and H_x, we calculated the angle (θ)
540 between these two helices formed by H332, Q334 and S337 (H337, Q339 and S342
541 in CAR3) (Fig. 4A, inset A₂, and Fig. 4D,G). The result revealed that branched
542 nonylphenols, 22NP, 33NP and 35NP, had angles of 69.3°, 70.7° and 70.9°,
543 respectively. On the other hand, both poses of linear NP, 4NP₁ and 4NP₂, displayed
544 angles with 72.5° and 72.8°, respectively, with highest distribution. The lowest
545 fluctuation was for CITCO, exhibiting an angle of 72.5°, which is close to that of 4NP₂,
546 and thus confirming its stronger affinity.

547 We also investigated the behaviour of Q344 (Q349 in CAR3) during the simulation, as
548 this interaction was previously proposed to stabilize the H11-H12 conformation
549 (Jyrkkarinne *et al.*, 2012). We studied the difference in the interaction of this residue
550 with the neighbour helices, H_x and H10/H11, among the systems, and compared it with
551 CITCO (Fig. 4C). Our findings revealed a high frequency of H-bond interaction
552 between Q349/Q344 (CAR3/CAR1 numbering) and the side chains of
553 Q336/Q331(H10/H11). This interaction was particularly frequent for CITCO, 22NP and
554 33NP for ~60% of simulation time. In contrast, it was less frequent for the 353NP and
555 linear NPs (~40–45%), consistently across both CAR3 and CAR1. However, no
556 frequent interaction (<20%) was established between the side-chain Q349/Q344
557 (CAR3/CAR1 numbering) and the backbone of S342/S337 in H_x (~10%, Fig. 4C, F).

558



559
 560 **Fig. 4.** Investigation of AF-2 region stabilization in wtCAR1 and CAR3 isoforms (A). CAR-LBD snapshot
 561 in the middle of the panel, location of K195-S348 (S353 in CAR3) denotes in brown circle zooming in
 562 left side (inset A1), the right-side boxes represent angle (θ) between H332, Q334 and S337 (H337, Q339
 563 and S342 in CAR3), (inset A2) and H-bond interaction between Q344, S337 and Q331 (inset A3). (B)
 564 K195-S353 distance in CAR3-LBD systems. (C) Q349-Q336 and Q349-S342 H-bond frequency in
 565 CAR3-LBD systems. (D) measured (θ) angles between H10/H11 and H_x in CAR3-LBD systems. (E)
 566 K195-S353 distance in CAR1-LBD systems. (F) Q344-Q331 and Q344-S337 H-bond frequency in
 567 CAR1-LBD systems. (G) measured (θ) angles between H10/H11 and H_x in CAR1-LBD systems. Median
 568 values for the distances and angles distributions are depicted near their respective analysed groups.
 569

570 Similar experiments were also conducted for wtCAR1 systems (Fig. 4E – G). In terms
 571 of K195-S348 interaction, the trend is comparable to that observed in CAR3 systems.
 572 Again, CITCO displayed the shortest distance (4.1 Å) and other compounds behave
 573 more or less similarly (4.7–4.9 Å, Fig. 4E). In terms of Q344-Q331 interaction, CITCO

574 and 33NP induce a slight increase to (~52% and 60% respectively) compared to that
575 of other compounds in CAR1. This resulted in a tighter angle (θ), between H332, Q334
576 and S337, in presence of 33NP and 353NP (Fig. 4G).

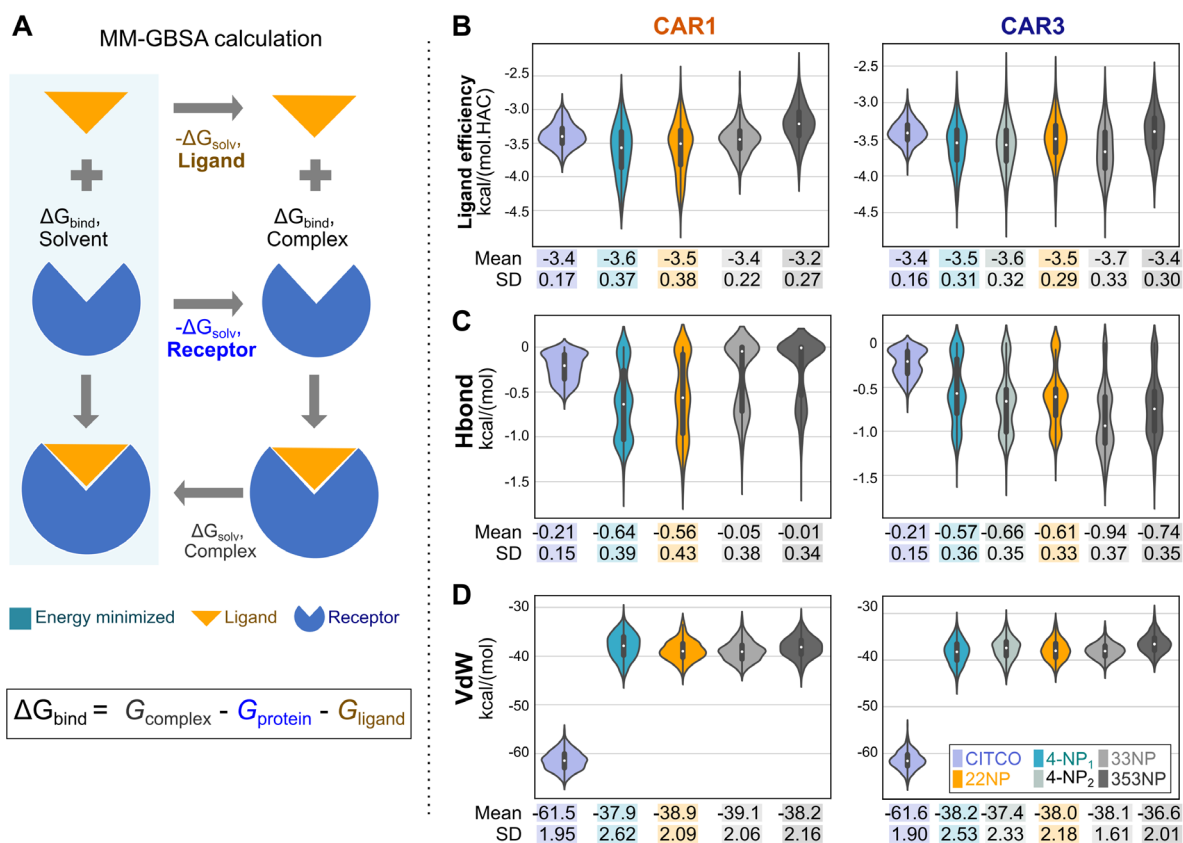
577 Interestingly, previous simulations including the corepressor peptide on the AF-2
578 region showed a clear shift of H12 position (Jyrkkarinne et al., 2012), induced by the
579 binding of inverse agonists. these shorter simulations (10 ns) were sufficient to show
580 the H12 movement toward H10, rather than away from the ligand binding domain, as
581 typically described for other NRs. This movement appears to be induced by the binding
582 of the peptide to accommodate its longer length compared to the shorter coactivator
583 peptide, rather than being solely ligand-induced.

584 Our extended simulations reveal that, independently from the ligand, the AF-2 region's
585 helix has a significant potential for conformational changes and dynamic shift between
586 states (Fig. S9). This suggests that analysing only short conformational changes would
587 lead to an incomplete representation of the NP binding.
588

589 **3.5 Free binding energy supports high binding affinity to NP comparable with** 590 **CITCO**

591 We further explored the CAR-NP interactions by predicting the binding energy using
592 free energy calculations with the molecular mechanics-generalized Born surface area
593 (MM-GBSA) method (Li et al., 2011). The MM-GBSA calculations start with the ligand
594 being extracted from the optimized complex and an energy calculation is run on it
595 without minimization (Fig. 5A), to get the energy of the ligand as optimized in the
596 binding pocket. Next, energy minimization is run on the ligand outside of the receptor.
597 Both calculations are being done with the ligand alone in the solution. The energy
598 difference is the ligand strain energy. The same procedure is performed for the
599 receptor counterpart, generating the receptor energy. The differences between those
600 two calculations generate the ligand binding energy, which can be either decomposed
601 by residues or by properties.
602

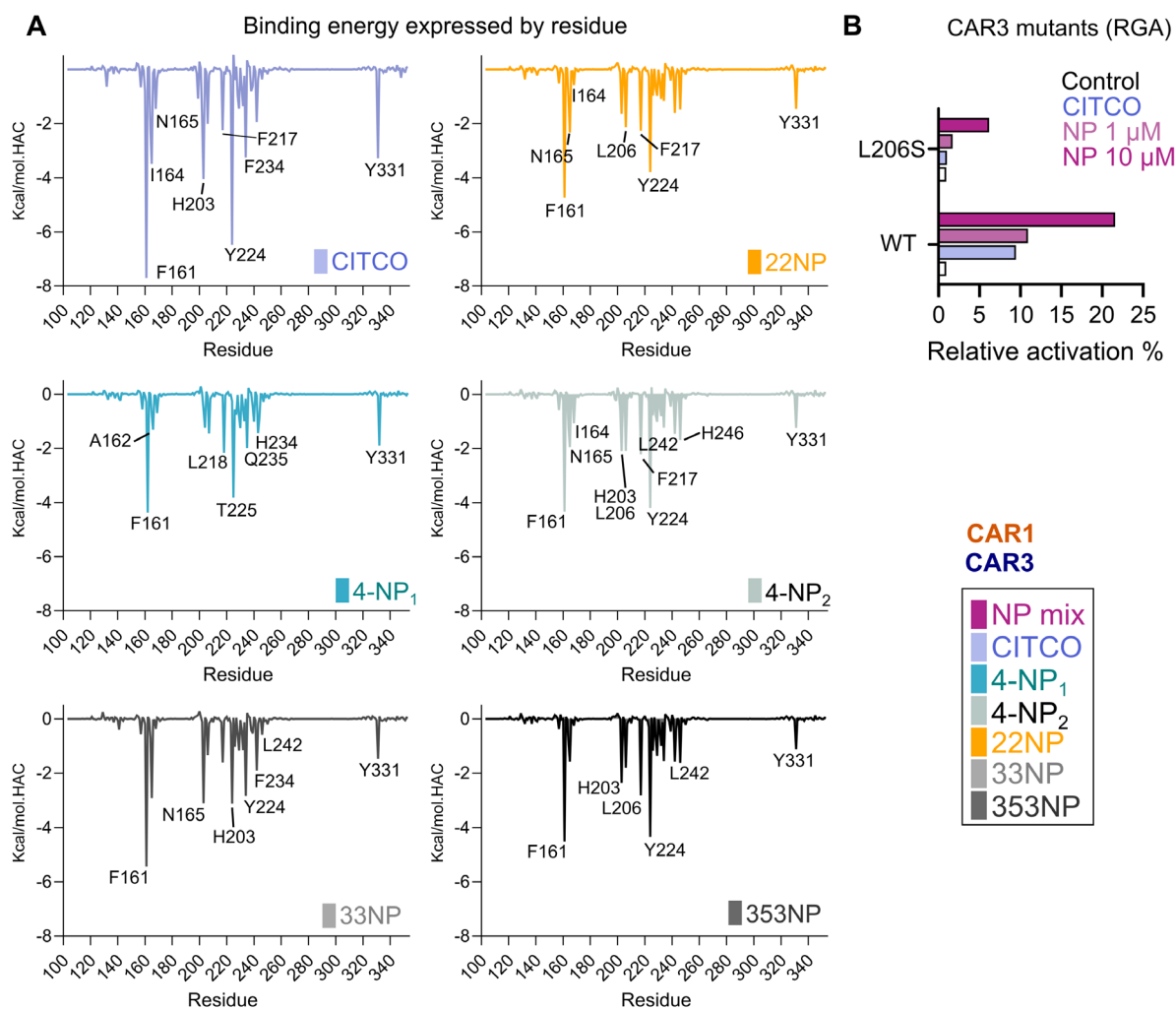
603 Our study showed that wtCAR1 and CAR3 exhibited comparable binding energies for
604 both CITCO and the NP isomers with a binding energy difference (ΔG_{bind}) of
605 approximately ~ -3.5 kcal/mol.HAC (Fig. 5B). However, there were notable differences
606 in the distribution of binding energies. CITCO displayed a narrower range of binding
607 energies, ranging from -3.1 – -3.6 kcal/mol.HAC. In contrasts, the NPs display boarder
608 range of energies, which could extend up to -4.5 kcal/mol.HAC. The decomposition of
609 this binding energy among hydrogen bond (Fig. 5C) and hydrophobic (Fig. 5D)
610 contributing terms suggests that wtCAR1 relies more on polar contacts to define the
611 binding energy profile. Moreover, 22-NP would have a higher polar binding affinity to
612 wtCAR1 than CITCO, and overall higher than CAR3. The higher hydrophobic
613 contribution for CITCO, in comparison to the NPs, can be explained by its larger apolar
614 surface. These findings suggest that the NP compounds possess a binding affinity that
615 is comparable to CITCO, indicating their potential as ligands for CAR receptors.



616
 617 **Fig. 5.** A) schematic representation of MM-GBSA calculations protocol and respective equation for
 618 binding energy calculation. Ligand efficiency binding energy (normalizing binding energy by the number
 619 of heavy atoms) (B), and their decompositions into hydrogen bond (B) and hydrophobic (D) terms, for
 620 both CAR1 and CAR3. In all violin plot graphics, the median of the calculated energies is displayed
 621 below, together with its standard deviation, and free energy binding calculation (Kcal/mol normalized by
 622 the Heavy Atoms Count, HAC), is decomposed as the average per residue of CAR sequences.
 623

624 Previous studies have attempted to correlate the activity of agonists with predicted
 625 binding energies, using the MM-GBSA method with 1 ns MDs (Kublbeck et al., 2011).
 626 Despite the small number of tested ligands (n=9) and short simulations, these studies
 627 found a strong correlation between the activation potential measured by mammalian
 628 one- and two-hybrid assays and the calculated binding energies ($r^2=0.88$ and $r^2=0.75$,
 629 respectively). However, the model used in these studies could not distinguish between
 630 agonists and inverse agonists.

631 More recent studies have conducted longer simulations and calculated binding
 632 energies based on two relevant binding mode conformations: one at the beginning (first
 633 nanosecond) and one at the end (9 ns) of the simulations (Jyrkkarinne et al., 2012).
 634 Interestingly, the addition of corepressor binding did not improve the binding energy
 635 predictions, and the overall correlation decreased when considering this additional
 636 factor.
 637



638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660

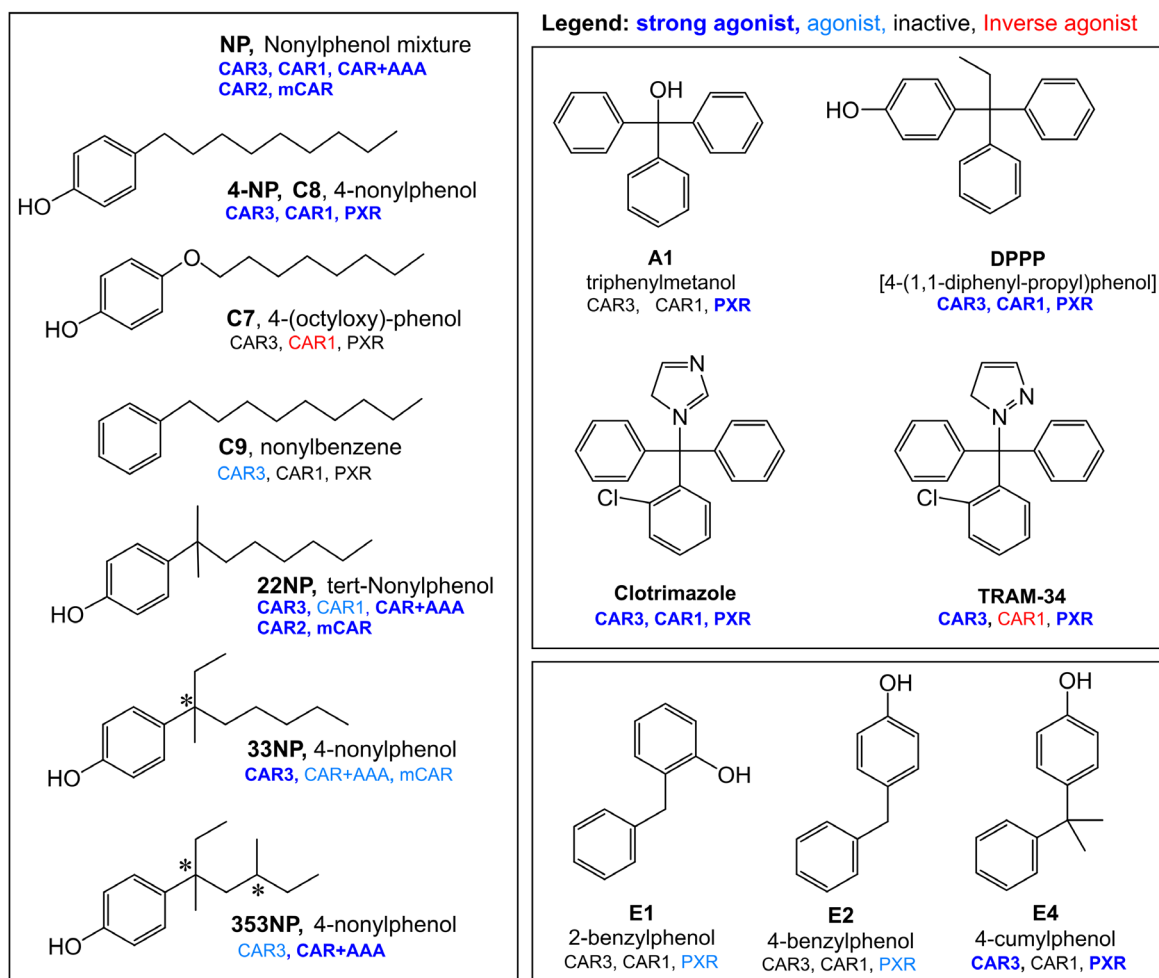
Fig. 6. A) free energy binding calculation (Kcal/mol normalized by the Heavy Atoms Count, HAC), decomposed as the average per residue of CAR3 sequences. B) CAR3 L206S-mutant abolished CAR3 activation by 4-NP mixture in reporter gene assay (see methods).

Interestingly, ΔG bind decompositions in terms of amino acids for CAR3 (Fig. 6A), reveals that F161 and Y224 contribute to most of the NP binding affinity, closely followed by N165 and L206 in branched NPs. In mutation experiments with L206S CAR3 mutant, 4-NP mixture lost its activity confirming its relevance (Fig. 6B). Previously, CAR1+F161A mutant decreased CITCO (at 1 μ M) activation by 10-fold (Kublbeck et al., 2011). Indeed, none of their tested ligands managed to activate the CAR1+F161A mutant, apart from clotrimazole. Only clotrimazole activation of the F161A mutant was attributed to its rigid structure and stable pocket occupancy, which is in line with our proposal for the NPs, where stable poses yielded higher activation. Further, their models showed no relevant interactions between CITCO/clotrimazole with N165. Indeed, N165A mutation increased both CITCO and clotrimazole activation levels (Kublbeck et al., 2011). In this sense, we hypothesize that the studied NPs would behave similarly as CITCO on the F161A variant, however, the N165A change would be relevant to change the NPs' binding specificity.

DISCUSSION

661 Together with PXR, CAR interacts with many estrogen-like endocrine disruptor
662 chemicals among their large range of xenobiotic binders. The NP mixture has been
663 described as a potent activator of both human and rodent CAR. 4-NP, especially
664 branched 4-NPs, is a strong endocrine disruptor activating many NRs. However,
665 detailed molecular mechanistic aspects of CAR activation by NP compounds have not
666 been fully elucidated. Also, since CAR3 variant is highly responsive to ligands and
667 abundant in the human liver, therefore, identifying its agonists and activators is
668 pharmacologically valuable (Lynch et al., 2014). Over the years, several experimental
669 screenings have been conducted to detect CAR3 ligands particularly and CAR
670 generally (Keminer et al., 2020). Herein, we investigate the binding mechanism of four
671 4-NP isomers to wtCAR1 and CAR3 and compared their conformational changes to
672 the well-known CAR agonist CITCO.

673 We observed that NP mixture and its isomers can directly interact with the CAR-LBD,
674 as shown by their ability to reconstitute full CAR structure in the assembly assay.
675 Indeed, branched individual 4-NP isomers bind to CAR more effectively than CITCO,
676 while linear (4-NP) is significantly less potent. These data well correlate with the data
677 done in CAR-LBD assembly assays (Fig. 2B) and in RGA with CAR+AAA mutant (Fig.
678 2E). Interestingly, however, just the NP mixture and 22NP can significantly activate
679 wtCAR1 in RGA (Fig. 2D), using a transfected wtCAR1 expression construct.
680 Additionally, CAR3 and murine CAR also display significant activation by 22NP/353NP
681 and 33NP, respectively (Fig. 2F,H and Fig. 7). Consistently, NP mixture was previously
682 described as a murine CAR activator in transient transfection assay (EC_{50} : 2.58 μ M)
683 (Baldwin and Roling, 2009).
684



685
 686 **Fig. 7.** Summary of our results in the context of other phenolates interacting with CAR and PXR. All
 687 literatures values are derived from this manuscript or the works by Hernandez *et al.*, (2007) and/or Dring
 688 *et al.*, (2010).
 689

690 In terms of target genes that are dominantly regulated, PXR primarily induces CYP3A4
 691 mRNA expression, while CAR primarily regulates CYP2B6 transcription upon
 692 activation. In the study, we confirmed that NP, 4-NP and 353NP significantly up-
 693 regulates CYP2B6 mRNA expression in differentiated HepaRG cells (Fig. 2I). In
 694 addition, 22NP also up-regulated CYP2B6 mRNA levels, even though the effect was
 695 not statistically significant. These data indicate that NPs and its isomers activate
 696 CAR1/CAR3 in HepaRG cells to regulate transcription of CYP2B6 gene, although PXR
 697 may also contribute to this up-regulation.
 698

699 Mechanistically, we studied conformation changes of CAR1/3 upon binding of
 700 individual NP isomers. We generated a competent model for CAR3, considering the
 701 effect of APYLT insertion (L:H8–H9) in comparison to CAR1 in our monomeric
 702 simulations. Our study did not detect special differences related to the APYLT insertion,
 703 the higher residues' RMSF (Table S5, S6) indicates the non-interactive residues in
 704 monomer, which agrees with previous studies (Keminer *et al.*, 2019; Omiecinski *et al.*,
 705 2011). Moreover, our CAR3 simulations did not show the H12 dislocation, as proposed
 706 for the classical NR-trapping mechanism. However, in contradiction to Keminer *et al.*,
 707 (2019), we observed an interaction between H10 (R325) and L:H8–H9 loop (D276)
 708 (data not shown). This might be derived from the absence of the heterodimeric partner

709 in our simulations since the presence of RXR would promote reorientation and its
710 interaction with H8–H9. However, this phenomenon needs further investigation.
711 Our initial docking poses for 4-NP, showed a single relevant conformation for wtCAR1
712 and two in the CAR3 binding sites (Fig. 3A,B). Hence, we included both docking poses
713 in our study. Though the starting configuration had opposite orientation, each ligand
714 discovered each other's orientation along simulation. This observation is confirmed by
715 transition point in RMSD plot (Fig. S4, S5). Multiple binding modes for congener ligands
716 are not unheard for nuclear receptors. Some examples such as the nonanoic acid
717 binding to PPAR γ , even display multiple binding modes for the same ligand (Shang et
718 al., 2018). This could be explained by CAR' large ligand binding pocket, which allows
719 such freedom for ligand movement.

720 Our simulations revealed all tested compounds are favourably stabilized in CAR3 as
721 they do in wtCAR1, which is supported by the assembly assay results, where all ligands
722 reconstitute full wtCAR1. The linear 4-NP proposed binding suggests a single relevant
723 conformation for wtCAR1 and two in the CAR3 binding site, however, their positions
724 are interchangeable and unstable during the trajectories. Alternatively, branched 4-NP
725 isomers, namely 22NP, 33NP and 353NP show a smaller conformation variety, mostly
726 being stabilized by hydrophobic contacts with F161 (H3), H203 (H5), F217 and Y224
727 (β -sheet), all displaying an interaction frequency of >60-100% of the analysed
728 simulation time.

729 Remarkably, F161 (H3) and Y224 (β -sheet) play a role in stabilizing CAR3-branched
730 4-NP interactions and to a lesser extent CAR1-linear 4-NPs and both greatly contribute
731 to most of the NP binding affinity, pointing out that occupying the hydrophobic pocket
732 composed by H3 and the β -sheet is key for the CAR3 potency and selectivity.

733 Predicting relative binding affinity can provide insights into the bioaccumulation
734 potential of novel EDs. MM-GBSA, for instance, is more computationally efficient than
735 rigorous alchemical perturbation methods (e.g., free energy perturbation), but still more
736 robust when compared to docking scoring functions (Genheden and Ryde, 2015).
737 Another advantage is the incorporation of the explicit solvent dynamics' influence in
738 the ligand binding, therefore accurately estimating their entropic contribution, which is
739 particularly relevant highly hydrophobic nature of the EDs. Similar approaches studied
740 PFAS binding dynamics using short MD simulations (Cheng and Ng, 2018), where they
741 observed that the predicted absolute PFAS binding energies were lower than
742 corresponding experimental values, however, still displayed great correlation.
743 Therefore, they emphasized the use of relative binding affinities rather than their
744 absolute binding strengths to rank the compounds. One has to bear in mind, however,
745 that directly correlating binding affinity with the results of the reporter gene assay can
746 be misleading since it is a cell-based assay and the compound's solubility, organelle
747 binding and transporting rate factors may determine their activity levels.

748

749 Despite the relevance of hydrophobic interactions stabilizing linear 4-NPs, NP
750 analogues such as nonylbenzene (e.g. **C9**, without a hydroxyl group, Fig. 7) being
751 moderately active (<10-fold at 10 μ M) on CAR3 and inactive on both CAR1 and PXR
752 (Dring et al., 2010), highlights the relevance of polar contacts, especially for CAR1. To
753 note, the same group reports CAR3 activation on the ~100-fold range with 22NP at
754 similar concentrations (10 μ M). Our simulation data supports CAR1's greater polar
755 requirement, showing the N165 (H3) and H203 (H5) pair sandwiching the NPs'
756 hydroxyl group, most expressively with 22NP.

757 Other ED compounds with similar polarity distribution and displaying phenolate
758 moieties (Fig. 7), such as 2-benzylphenol and 4-benzylphenol (two rings connected by
759 an ethyl linker) or triphenylmethanol (3 rings and a single hydroxyl), had no activity on

760 CAR3, wtCAR1 or PXR (up to 10 μ M, (Dring et al., 2010). Meanwhile 4-cumylphenol
761 and DPPP [3-(1,1-diphenylpropyl) phenol] displayed >100-fold and ~50-fold induction
762 on CAR3 activation. These compounds have a branched aliphatic linker and an
763 increased sp³ ratio in comparison to their inactive counterparts and, therefore, would
764 chemically resemble our branched NPs.

765 Finally, though the study of individual contaminant exposure is important,
766 environmental exposure happens as a combination of multiple isomers or a complex
767 chemical mixture. In CAR response to xenobiotics, the additive effects of such
768 chemicals have been extensively described (Baldwin and Roling, 2009). Our study
769 determines that NP mixture has a comparatively high potency to individual branched
770 NPs, and these values can be used to further expand on their additive's properties, as
771 well as assessing them in daily used products or environmental samples.

772

773 CONCLUSIONS

774 NP, as a mixture of isomers, was previously shown as a potent endocrine disruptor
775 binding to ER, PXR, wtCAR1 and rodent CAR. Our work expanded on the detailed
776 mechanistic aspects of CAR activation by 4-NP. Here we examined interactions of
777 individual branched (22NP, 33NP, and 353NP) and linear 4-NPs with CAR and its
778 variants using molecular dynamics (MD) simulations, cellular experiments with various
779 CAR constructs or CAR ligand binding domain (LBD) mutants or differentiated
780 HepaRG hepatocyte cellular model.

781 We found that branched 4-NPs display better poses to activate both wild-type wtCAR1
782 as well as CAR3 LBDs in MD simulations. Consistently, branched 4-NPs activated
783 CAR3 and wtCAR1 LBD more efficiently than linear 4-NP or *tert*-nonylphenol (22NP).
784 In HepaRG cells we demonstrate that all tested NP and 4-NP have some capacity to
785 up-regulated CYP2B6 RNA expression, which is the target gene dominantly controlled
786 by CAR.

787 This is the first work describing interactions of individual 4-NP isomers in detail with
788 the human CAR receptor and its dominant variant CAR3. We believe this new
789 approach, combining the validation of individual isomers with extensive molecular
790 modelling calculations, can contribute to the safer use of NP.

791

792 ASSOCIATED CONTENT

793 **Supporting Information.** The Supporting Information is available free of charge at the
794 same website from this manuscript. Supplementary figures, data collection, and further
795 information are available. All molecular dynamics trajectories and raw data related to
796 the protein-ligand interactions within the simulations are available in the repository:
797 10.5281/zenodo.8113303

798

799 AUTHOR INFORMATION

800 Corresponding authors: T.K. thales.kronenberger@uni-tuebingen.de, Auf der
801 Morgenstelle 8, DE72076, Tübingen (Germany), P.P. pavek@faf.cuni.cz, Faculty of
802 Pharmacy, Charles University, Akademika Heyrovskeho 1203, 500 05, Hradec
803 Kralove (Czech Republic).

804

805 ORCID

806 Azam Rashidian: 0000-0002-5628-3439
807 Jan Dusek: 0000-0001-7685-1483
808 Martin Drastik: 0000-0001-6492-3566
809 Lucie Smutná: 0000-0003-0153-0057
810 Kristin Fritsche: 0000-0002-8177-5822
811 Albert Braeuning: 0000-0003-3810-027X
812 Antti Poso: 0000-0003-4196-4204
813 Thales Kronenberger: 0000-0001-6933-7590
814 Petr Pavek: 0000-0001-8769-4196

815
816 Author Contributions. A.R. and J.D. equally contributed to this work. A.R., with the
817 support of M.D., performed the *in silico* experiments, using docking and MD simulations
818 and their respective data analyses. A.P. provided resources and helped with
819 manuscript writing. J.D.; L.S; K.F; contributed to the execution of biological assays.
820 A.B. contributed to the manuscript revision.

821
822 P.P. and T.K. conceived the original idea, initiated the project, oversaw all experiments,
823 data analysis, and wrote the manuscript with assistance from other authors. All authors
824 contributed to the manuscript review, provided comments, and suggestions.

825
826 **Conflict of interest**
827 The authors declare that they have no known competing financial interests or personal
828 relationships that could have appeared to influence the work reported in this paper.

829
830 **Funding**
831 A.R., K.F., A.B., A.O., T.K. and P.P. acknowledge European Union's Horizon 2020
832 research and innovation program under grant agreement No 825762, EDCMET
833 project Metabolic effects of Endocrine Disrupting Chemicals: novel testing METHods
834 and adverse outcome pathways (EDCMET). T.K. is funded by the fortune initiative
835 and from TüCAD₂ and CMIF. TüCAD₂ and CMIF are funded by the Federal Ministry
836 of Education and Research (BMBF) and the Baden-Württemberg Ministry of Science
837 as part of the Excellence Strategy of the German Federal and State Governments.
838 P.P. is funded by Czech Scientific Foundation (22-05167S).

839
840 **Acknowledgement**
841 The authors wish to acknowledge CSC – IT Center for Science, Finland, for the very
842 generous computational resources. The CYP2B6-luc reporter plasmid (originally
843 entitled B-1.6k/PB/XREM) was kindly donated by Dr. Hongbing Wang (University of
844 Maryland School of Pharmacy, Baltimore, MD, USA), which were are grateful for.

845
846 **ABBREVIATIONS**
847 BSA, bovine serum albumin; CAR, Constitutive Androstane Receptor; CYP450,
848 cytochrome P450; DMSO, dimethyl sulfoxide; ER, Estrogen Receptor; FDA, Food and
849 Drug Administration; HPLC, high-performance liquid chromatography; HRMS, high-
850 resolution mass spectrometry; IC₅₀, half-maximal inhibitory concentration; MD,
851 molecular dynamics simulations; NAD, nicotinamide adenine dinucleotide; NADPH,
852 nicotinamide adenine dinucleotide phosphate; NMR, nuclear magnetic resonance; NP,
853 nonylphenol; PXR, Pregnane X Receptor.

854
855 **References**
856

857 Baldwin WS and Roling JA (2009) A concentration addition model for the activation of the
858 constitutive androstane receptor by xenobiotic mixtures. *Toxicol Sci* **107**:93-105.

859 Bhandari G, Bagheri AR, Bhatt P and Bilal M (2021) Occurrence, potential ecological risks,
860 and degradation of endocrine disrupter, nonylphenol, from the aqueous environment.
861 *Chemosphere* **275**:130013.

862 Bowers KJ, Chow E, Xu H, Dror RO, Eastwood MP, Gregersen BA, Klepeis JL, Kolossvary I,
863 Moraes MA, Sacerdoti FD, Salmon JK, Shan Y and Shaw DE (2006) Scalable
864 Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In 2006.
865 <https://doi.org/10.1145/1188455.1188544>. *Proceedings of the 2006 ACM/IEEE*
866 *Conference on Supercomputing; SC '06; ACM: New York, NY, USA, .*

867 Carazo A and Pavek P (2015) The Use of the LanthaScreen TR-FRET CAR Coactivator Assay
868 in the Characterization of Constitutive Androstane Receptor (CAR) Inverse Agonists.
869 *Sensors* **15**:9265-9276.

870 Casals-Casas C and Desvergne B (2011) Endocrine disruptors: from endocrine to metabolic
871 disruption. *Annu Rev Physiol* **73**:135-162.

872 Darden T, York D and Pedersen L (1993) Particle Mesh Ewald: An N·log(N) Method for Ewald
873 Sums in Large Systems. *J Chem Phys* **98**:10089–10092. .

874 DeKeyser JG, Laurenzana EM, Peterson EC, Chen T and Omiecinski CJ (2011) Selective
875 phthalate activation of naturally occurring human constitutive androstane receptor
876 splice variants and the pregnane X receptor. *Toxicol Sci* **120**:381-391.

877 di Masi A, De Marinis E, Ascenzi P and Marino M (2009) Nuclear receptors CAR and PXR:
878 Molecular, functional, and biomedical aspects. *Molecular aspects of medicine* **30**:297-
879 343.

880 Dring AM, Anderson LE, Qamar S and Stoner MA (2010) Rational quantitative structure-
881 activity relationship (RQSAR) screen for PXR and CAR isoform-specific nuclear
882 receptor ligands. *Chem Biol Interact* **188**:512-525.

883 Dussault I and Forman BM (2002) The nuclear receptor PXR: a master regulator of "homeland"
884 defense. *Crit Rev Eukaryot Gene Expr* **12**:53-64.

885 Frank C, Molnar F, Matilainen M, Lempiainen H and Carlberg C (2004) Agonist-dependent
886 and agonist-independent transactivations of the human constitutive androstane receptor
887 are modulated by specific amino acid pairs. *The Journal of biological chemistry*
888 **279**:33558-33566.

889 Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll
890 EH, Shelley M, Perry JK, Shaw DE, Francis P and Shenkin PS (2004) Glide: a new
891 approach for rapid, accurate docking and scoring. 1. Method and assessment of docking
892 accuracy. *J Med Chem* **47**:1739-1749.

893 Genheden S and Ryde U (2015) The MM/PBSA and MM/GBSA methods to estimate ligand-
894 binding affinities. *Expert Opin Drug Discov* **10**:449-461.

895 Gore AC, Chappell VA, Fenton SE, Flaws JA, Nadal A, Prins GS, Toppari J and Zoeller RT
896 (2015) EDC-2: The Endocrine Society's Second Scientific Statement on Endocrine-
897 Disrupting Chemicals. *Endocr Rev* **36**:E1-E150.

898 Heindel JJ, Blumberg B, Cave M, Machtinger R, Mantovani A, Mendez MA, Nadal A, Palanza
899 P, Panzica G, Sargis R, Vandenberg LN and Vom Saal F (2017) Metabolism disrupting
900 chemicals and metabolic disorders. *Reprod Toxicol* **68**:3-33.

901 Heindel JJ, Vom Saal FS, Blumberg B, Bovolini P, Calamandrei G, Ceresini G, Cohn BA,
902 Fabbri E, Gioiosa L, Kassotis C, Legler J, La Merrill M, Rizzir L, Machtinger R,
903 Mantovani A, Mendez MA, Montanini L, Molteni L, Nagel SC, Parmigiani S, Panzica
904 G, Paterlini S, Pomatto V, Ruzzin J, Sartor G, Schug TT, Street ME, Suvorov A, Volpi
905 R, Zoeller RT and Palanza P (2015) Parma consensus statement on metabolic disruptors.
906 *Environ Health* **14**:54.

- 907 Hernandez JP, Huang W, Chapman LM, Chua S, Moore DD and Baldwin WS (2007) The
908 environmental estrogen, nonylphenol, activates the constitutive androstane receptor.
909 *Toxicol Sci* **98**:416-426.
- 910 Hernandez JP, Chapman LM, Kretschmer XC and Baldwin WS (2006) Gender-specific
911 induction of cytochrome P450s in nonylphenol-treated FVB/NJ mice. *Toxicol Appl*
912 *Pharmacol* **216**:186-196.
- 913 Hernandez JP, Mota LC, Huang W, Moore DD and Baldwin WS (2009) Sexually dimorphic
914 regulation and induction of P450s by the constitutive androstane receptor (CAR).
915 *Toxicology* **256**:53-64.
- 916 Hong L and Li MH (2007) Acute toxicity of 4-nonylphenol to aquatic invertebrates in Taiwan.
917 *Bull Environ Contam Toxicol* **78**:445-449.
- 918 Hoover WG (1985) Canonical dynamics: Equilibrium phase-space distributions. *Phys Rev A*
919 *Gen Phys* **31**:1695-1697.
- 920 Hunter JD (2007) Matplotlib: A 2D Graphics Environment. *Computing in Science Engineering*
921 **9**:90-95. .
- 922 Hyrsova L, Smutny T, Carazo A, Moravcik S, Mandikova J, Trejtnar F, Gerbal-Chaloin S and
923 Pavek P (2016) The pregnane X receptor down-regulates organic cation transporter 1
924 (SLC22A1) in human hepatocytes by competing for ("squelching") SRC-1 coactivator.
925 *British journal of pharmacology* **173**:1703-1715.
- 926 Chai SC, Cherian MT, Wang YM and Chen T (2016) Small-molecule modulators of PXR and
927 CAR. *Biochim Biophys Acta* **1859**:1141-1154.
- 928 Cheng W and Ng CA (2018) Predicting Relative Protein Affinity of Novel Per- and
929 Polyfluoroalkyl Substances (PFASs) by An Efficient Molecular Dynamics Approach.
930 *Environ Sci Technol* **52**:7972-7980.
- 931 Chung SWC (2021) The development of isomer-specific analysis of branched 4-nonylphenol
932 in food for dietary exposure - a critical review of analytical methods and occurrence in
933 foodstuffs. *Food Addit Contam Part A Chem Anal Control Expo Risk Assess* **38**:842-
934 855.
- 935 Ingraham HA and Redinbo MR (2005) Orphan nuclear receptors adopted by crystallography.
936 *Current opinion in structural biology* **15**:708-715.
- 937 Jacobson MP, Pincus DL, Rapp CS, Day TJ, Honig B, Shaw DE and Friesner RA (2004) A
938 hierarchical approach to all-atom protein loop prediction. *Proteins* **55**:351-367.
- 939 Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML and (1983) Comparison of
940 Simple Potential Functions for Simulating Liquid Water. . *J Chem Phys* **79**:926-935. .
- 941 Jyrkkarinne J, Kublbeck J, Pulkkinen J, Honkakoski P, Laatikainen R, Poso A and Laitinen T
942 (2012) Molecular dynamics simulations for human CAR inverse agonists. *Journal of*
943 *chemical information and modeling* **52**:457-464.
- 944 Kamata R, Nakajima D and Shiraishi F (2018) Agonistic effects of diverse xenobiotics on the
945 constitutive androstane receptor as detected in a recombinant yeast-cell assay. *Toxicol*
946 *In Vitro* **46**:335-349.
- 947 Keminer O, Teigeler M, Kohler M, Wenzel A, Arning J, Kassner F, Windshugel B and
948 Eilebrecht E (2020) A tiered high-throughput screening approach for evaluation of
949 estrogen and androgen receptor modulation by environmentally relevant bisphenol A
950 substitutes. *Sci Total Environ* **717**:134743.
- 951 Keminer O, Windshugel B, Essmann F, Lee SML, Schiergens TS, Schwab M and Burk O
952 (2019) Identification of novel agonists by high-throughput screening and molecular
953 modelling of human constitutive androstane receptor isoform 3. *Archives of toxicology*
954 **93**:2247-2264.
- 955 Kolpin DW, Furlong ET, Meyer MT, Thurman EM, Zaugg SD, Barber LB and Buxton HT
956 (2002) Pharmaceuticals, hormones, and other organic wastewater contaminants in U.S.
957 streams, 1999-2000: a national reconnaissance. *Environ Sci Technol* **36**:1202-1211.

958 Kretschmer XC and Baldwin WS (2005) CAR and PXR: xenosensors of endocrine disrupters?
959 *Chem Biol Interact* **155**:111-128.

960 Kublbeck J, Laitinen T, Jyrkkarinne J, Rousu T, Tolonen A, Abel T, Kortelainen T, Uusitalo J,
961 Korjamo T, Honkakoski P and Molnar F (2011) Use of comprehensive screening
962 methods to detect selective human CAR activators. *Biochemical pharmacology*
963 **82**:1994-2007.

964 Kublbeck J, Niskanen J and Honkakoski P (2020a) Metabolism-Disrupting Chemicals and the
965 Constitutive Androstane Receptor CAR. *Cells* **9**.

966 Kublbeck J, Vuorio T, Niskanen J, Fortino V, Braeuning A, Abass K, Rautio A, Hakkola J,
967 Honkakoski P and Levonen AL (2020b) The EDCMET Project: Metabolic Effects of
968 Endocrine Disruptors. *Int J Mol Sci* **21**.

969 Lasch A, Marx-Stoelting P, Braeuning A and Lichtenstein D (2021) More than additive effects
970 on liver triglyceride accumulation by combinations of steatotic and non-steatotic
971 pesticides in HepaRG cells. *Archives of toxicology* **95**:1397-1411.

972 Li J, Abel R, Zhu K, Cao Y, Zhao S and Friesner RA (2011) The VSGB 2.0 model: a next
973 generation energy model for high resolution protein structure modeling. *Proteins*
974 **79**:2794-2812.

975 Li X, Huo J, Liu Z, Yue Q, Zhang L, Gong Y, Chen J and Bao H (2019) An updated weight of
976 evidence approach for deriving a health-based guidance value for 4-nonylphenol. *J Appl*
977 *Toxicol* **39**:87-100.

978 Lu C, Wu C, Ghoreishi D, Chen W, Wang L, Damm W, Ross GA, Dahlgren MK, Russell E,
979 Von Bargaen CD, Abel R, Friesner RA and Harder ED (2021a) OPLS4: Improving Force
980 Field Accuracy on Challenging Regimes of Chemical Space. *J Chem Theory Comput*
981 **17**:4291-4300.

982 Lu D, Yu L, Li M, Zhai Q, Tian F and Chen W (2021b) Behavioral disorders caused by
983 nonylphenol and strategies for protection. *Chemosphere* **275**:129973.

984 Lynch C, Pan Y, Li L, Ferguson SS, Xia M, Swaan PW and Wang H (2013) Identification of
985 novel activators of constitutive androstane receptor from FDA-approved drugs by
986 integrated computational and biological approaches. *Pharmaceutical research* **30**:489-
987 501.

988 Lynch C, Pan Y, Li L, Heyward S, Moeller T, Swaan PW and Wang H (2014) Activation of
989 the constitutive androstane receptor inhibits gluconeogenesis without affecting
990 lipogenesis or fatty acid synthesis in human hepatocytes. *Toxicol Appl Pharmacol*
991 **279**:33-42.

992 Lynch C, Zhao J, Wang H and Xia M (2016) Quantitative High-Throughput Luciferase
993 Screening in Identifying CAR Modulators. *Methods Mol Biol* **1473**:33-42.

994 Mackowiak B, Hodge J, Stern S and Wang H (2018) The Roles of Xenobiotic Receptors:
995 Beyond Chemical Disposition. *Drug metabolism and disposition: the biological fate of*
996 *chemicals*.

997 Mackowiak B and Wang H (2016) Mechanisms of xenobiotic receptor activation: Direct vs.
998 indirect. *Biochim Biophys Acta* **1859**:1130-1140.

999 Martyna GJ, Tuckerman ME, Tobias DJ and Klein ML (1996) Explicit reversible integrators
1000 for extended systems dynamics. *Mol Phys* **87**:1117-1157.

1001 Mejdrova I, Dusek J, Skach K, Stefela A, Skoda J, Chalupsky K, Dohnalova K, Pavkova I,
1002 Kronenberger T, Rashidian A, Smutna L, Duchoslav V, Smutny T, Pavek P and Nencka
1003 R (2023) Discovery of Novel Human Constitutive Androstane Receptor Agonists with
1004 the Imidazo[1,2-a]pyridine Structure. *J Med Chem*.

1005 Molnar F, Kublbeck J, Jyrkkarinne J, Prantner V and Honkakoski P (2013) An update on the
1006 constitutive androstane receptor (CAR). *Drug metabolism and drug interactions* **28**:79-
1007 93.

1008 Muller S, Schmid P and Schlatter C (1998) Pharmacokinetic behavior of 4-nonylphenol in
1009 humans. *Environ Toxicol Pharmacol* **5**:257-265.

1010 Nosé S (1984) A Unified Formulation of the Constant Temperature Molecular Dynamics
1011 Methods. *Journal of Chemical Physics*, **81**:511-519.

1012 Omiecinski CJ, Coslo DM, Chen T, Laurenzana EM and Peffer RC (2011) Multi-species
1013 analyses of direct activators of the constitutive androstane receptor. *Toxicol Sci*
1014 **123**:550-562.

1015 Papalou O, Kandaraki EA, Papadakis G and Diamanti-Kandarakis E (2019) Endocrine
1016 Disrupting Chemicals: An Occult Mediator of Metabolic Disease. *Front Endocrinol*
1017 (*Lausanne*) **10**:112.

1018 Ross J, Plummer SM, Rode A, Scheer N, Bower CC, Vogel O, Henderson CJ, Wolf CR and
1019 Elcombe CR (2010) Human constitutive androstane receptor (CAR) and pregnane X
1020 receptor (PXR) support the hypertrophic but not the hyperplastic response to the murine
1021 nongenotoxic hepatocarcinogens phenobarbital and chlordane in vivo. *Toxicol Sci*
1022 **116**:452-466.

1023 Shang J, Brust R, Mosure SA, Bass J, Munoz-Tello P, Lin H, Hughes TS, Tang M, Ge Q,
1024 Kamenekca TM and Kojetin DJ (2018) Cooperative cobinding of synthetic and natural
1025 ligands to the nuclear receptor PPARgamma. *Elife* **7**.

1026 Shin JA, Lee JH, Lim SY, Ha HS, Kwon HS, Park YM, Lee WC, Kang MI, Yim HW, Yoon
1027 KH and Son HY (2013) Metabolic syndrome as a predictor of type 2 diabetes, and its
1028 clinical interpretations and usefulness. *J Diabetes Investig* **4**:334-343.

1029 Skoda J, Dusek J, Drastik M, Stefela A, Dohnalova K, Chalupsky K, Smutny T, Micuda S,
1030 Gerbal-Chaloin S and Pavek P (2020a) Diazepam Promotes Translocation of Human
1031 Constitutive Androstane Receptor (CAR) via Direct Interaction with the Ligand-
1032 Binding Domain. *Cells* **9**:2532.

1033 Skoda J, Dusek J, Drastik M, Stefela A, Dohnalova K, Chalupsky K, Smutny T, Micuda S,
1034 Gerbal-Chaloin S and Pavek P (2020b) Diazepam Promotes Translocation of Human
1035 Constitutive Androstane Receptor (CAR) via Direct Interaction with the Ligand-
1036 Binding Domain. *Cells* **9**.

1037 Suino K, Peng L, Reynolds R, Li Y, Cha JY, Repa JJ, Kliewer SA and Xu HE (2004) The
1038 nuclear xenobiotic receptor CAR: structural determinants of constitutive activation and
1039 heterodimerization. *Molecular cell* **16**:893-905.

1040 Wright E, Busby SA, Wisecarver S, Vincent J, Griffin PR and Fernandez EJ (2011) Helix 11
1041 dynamics is critical for constitutive androstane receptor activity. *Structure* **19**:37-44.

1042 Xu RX, Lambert MH, Wisely BB, Warren EN, Weinert EE, Waitt GM, Williams JD, Collins
1043 JL, Moore LB, Willson TM and Moore JT (2004) A structural basis for constitutive
1044 activity in the human CAR/RXRalpha heterodimer. *Molecular cell* **16**:919-928.

1045 Yang KC, Hung HF, Lu CW, Chang HH, Lee LT and Huang KC (2016) Association of Non-
1046 alcoholic Fatty Liver Disease with Metabolic Syndrome Independently of Central
1047 Obesity and Insulin Resistance. *Sci Rep* **6**:27034.

1048 Yilmaz B, Terekci H, Sandal S and Kelestimur F (2020) Endocrine disrupting chemicals:
1049 exposure, effects on human health, mechanism of action, models for testing and
1050 strategies for prevention. *Rev Endocr Metab Disord* **21**:127-147.

1051

Filling the blank space: Branched 4-nonylphenols isomers are responsible for robust constitutive androstane receptor (CAR) activation by nonylphenol

Azam Rashidian^{1,‡}, Jan Dusek^{2,‡}, Martin Drastik², Lucie Smutná², Kristin Fritsche³, Albert Braeuning³, Antti Poso^{1,3,5}, Thales Kronenberger^{3,5*}, Petr Pavek^{2*}

¹ Department of Internal Medicine VIII, University Hospital of Tübingen, 72076, Tübingen, Germany.

² Department of Pharmacology and Toxicology, Faculty of Pharmacy in Hradec Kralove, Charles University, Akademika Heyrovskeho 1203, 500 05, Hradec Kralove, Czech Republic.

³ German Federal Institute for Risk Assessment, Department Food Safety, Max-Dohrn-Str. 8-10, 10589 Berlin, Germany.

⁴ **a.** School of Pharmacy, Faculty of Health Sciences, University of Eastern Finland, 70211, Kuopio, Finland. **b.** Department of Pharmaceutical and Medicinal Chemistry, Institute of Pharmaceutical Sciences, Eberhard-Karls-Universität, Tuebingen, Auf der Morgenstelle 8, 72076 Tuebingen, Germany. **c.** Tuebingen Center for Academic Drug Discovery & Development (TüCAD₂), 72076 Tuebingen, Germany. **d.** Excellence Cluster "Controlling Microbes to Fight Infections" (CMFI), 72076 Tübingen, Germany

‡ These authors equally contributed to this work

* corresponding authors:

T.K.: thales.kronenberger@uni-tuebingen.de, Auf der Morgenstelle 8, DE72076, Tübingen (Germany).

P.P.: pavek@faf.cuni.cz, kademika Heyrovskeho 1203, 500 05, Hradec Kralove, Czech Republic.

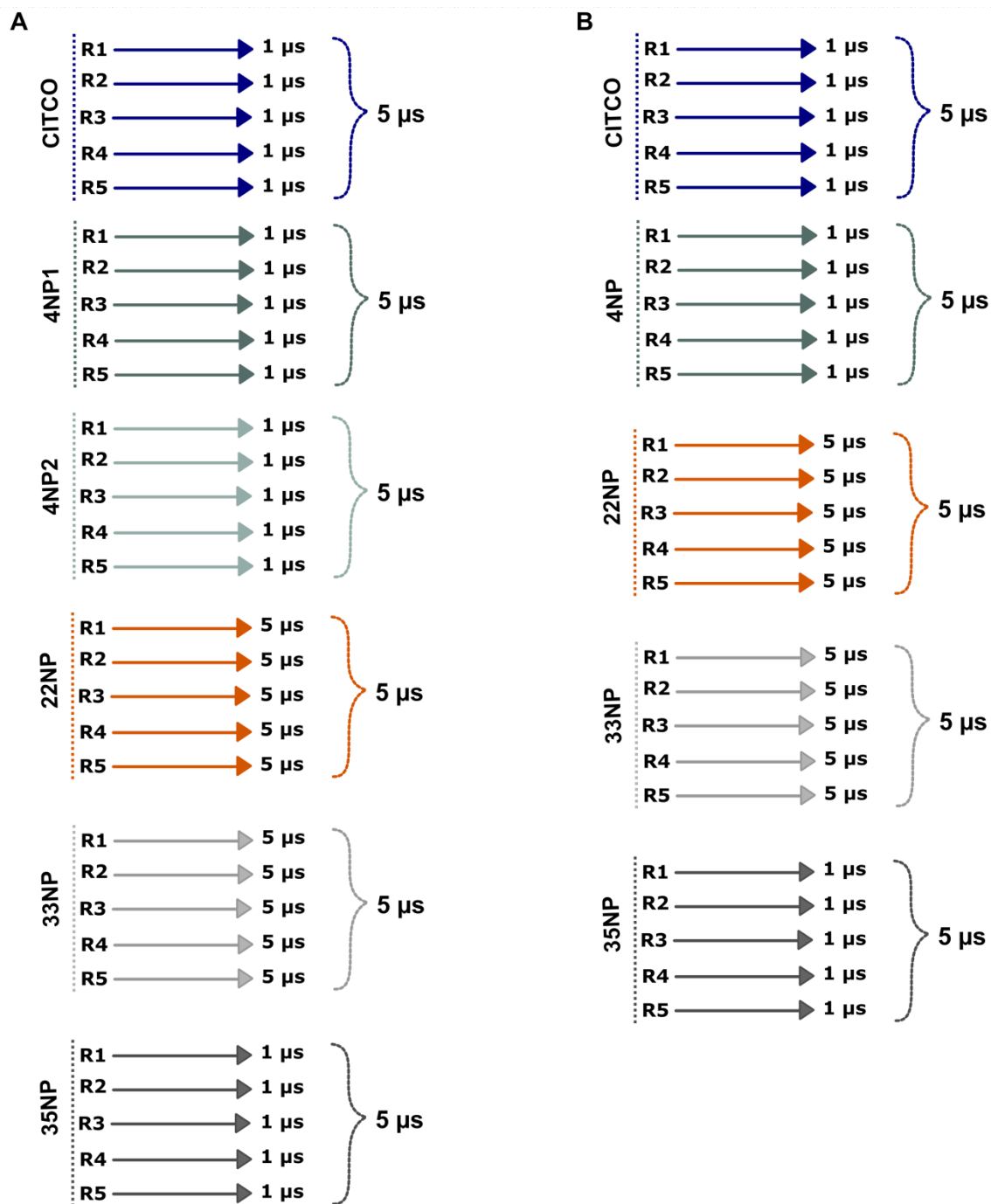


Fig. S1. Outline of the conducted simulations. Figure illustrates the generated replicas (R stands for replica). In total, five individual replicas were run for each system. (A) demonstrates the generated systems with CAR3-LBD in presence of CITCO (purple) and test ligands, namely: linear 4-NP₁ (dark green) and 4-NP₂ (light green), 22NP (orange), 33NP (light grey), 35NP (dark grey). (B) demonstrates the generated systems with CAR1-LBD in presence of CITCO and test ligands (linear 4-NP, 22NP, 33NP, 35NP).

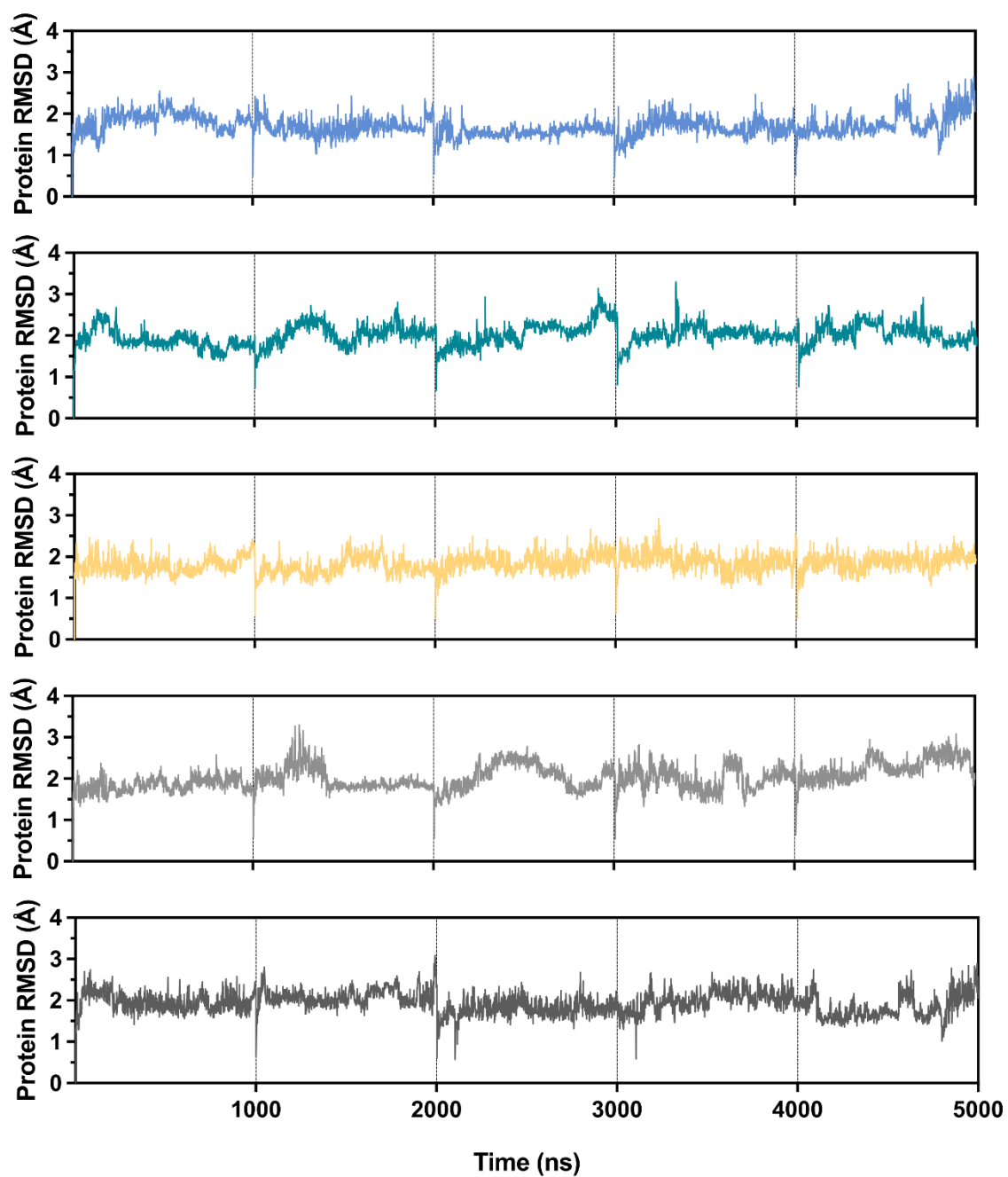


Fig. S2. Protein RMSD (Å) in CAR1-LBD. Colours are described as follow: CITCO (purple) and test ligands, namely: linear 4-NP (dark green), 22NP (orange), 33NP (light grey), 35NP (dark grey).

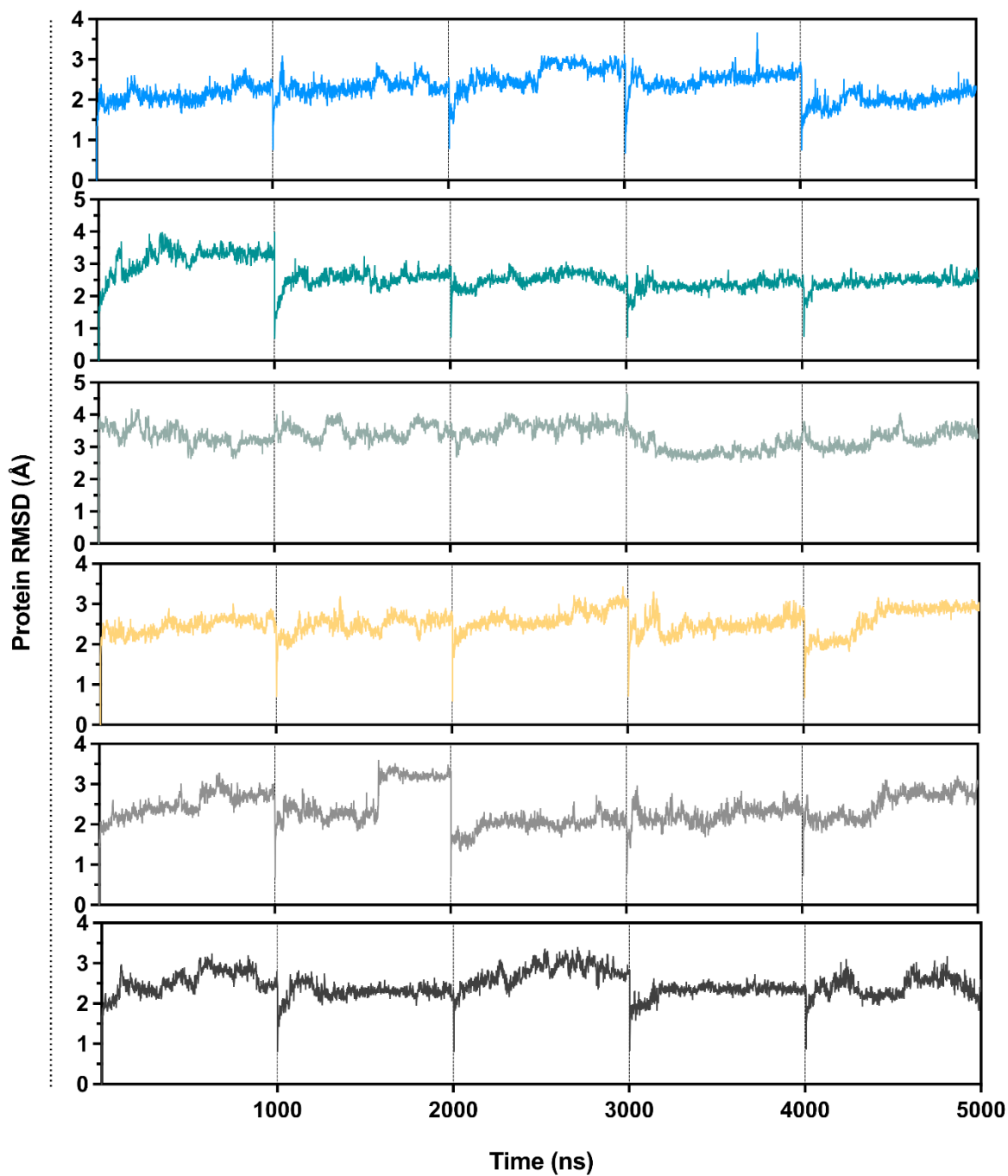


Fig. S3. Protein RMSD (Å) in CAR3-LBD. Colours are described as follow: CITCO (purple) and test ligands, namely: linear 4-NP₁ (dark green) and 4-NP₂ (light green), 22NP (orange), 33NP (light grey), 35NP (dark grey).

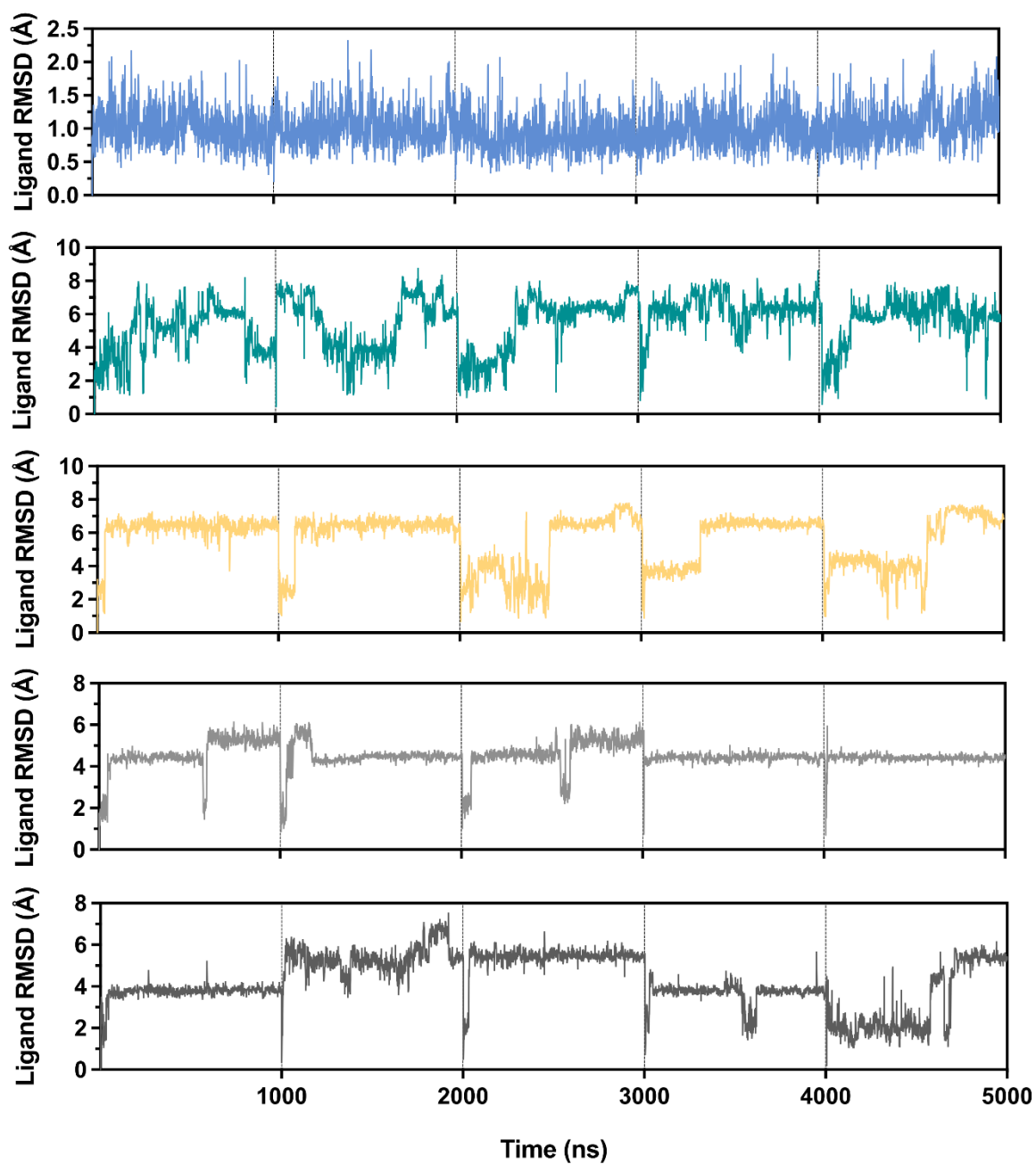


Fig. S4. Ligand RMSD (Å) in CAR1-LBD. Colours are described as follow: CITCO (purple) and test ligands, namely: linear 4-NP (dark green), 22NP (orange), 33NP (light grey), 35NP (dark grey).

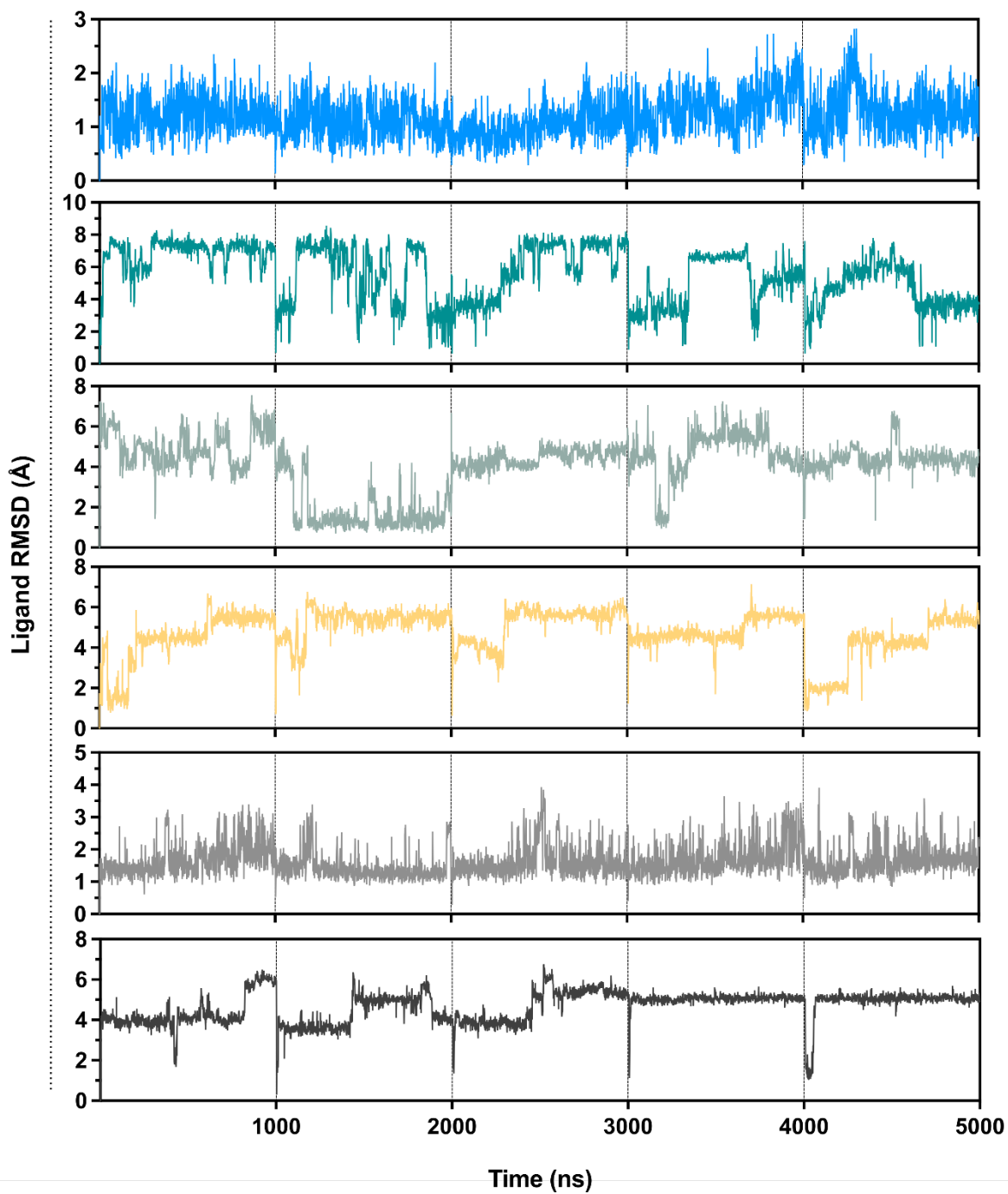


Fig. S5. Ligand RMSD (Å) in CAR3-LBD. Colours are described as follow: CITCO (purple) and test ligands, namely: linear 4-NP₁ (dark green) and 4-NP₂ (light green), 22NP (orange), 33NP (light grey), 35NP (dark grey).

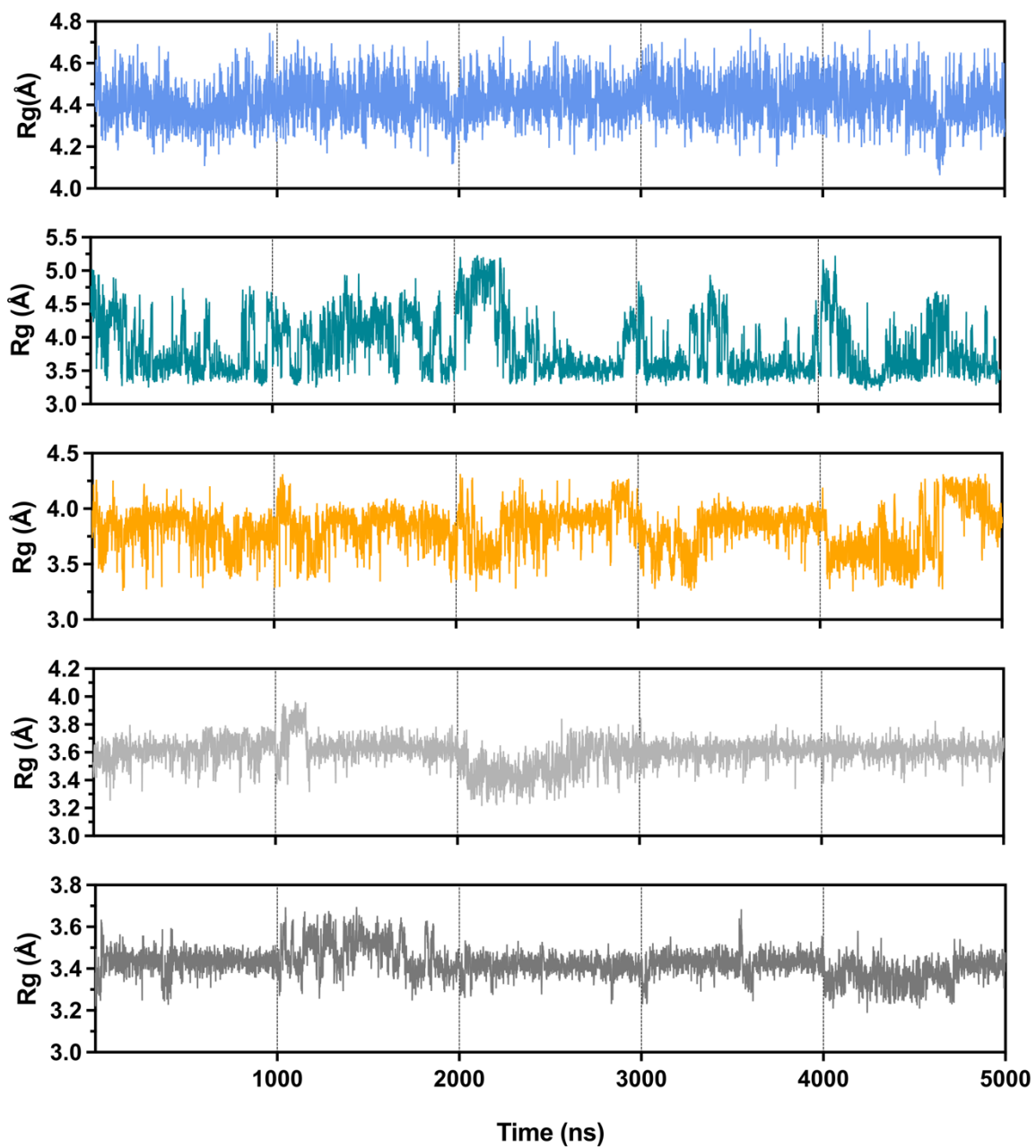


Fig. S6. Radius of Gyration (Rg) of wtCAR1 individual system. Colours are described as follow: CITCO (purple) and test ligands, namely: linear 4-NP (dark green), 22NP (orange), 33NP (light grey), 35NP (dark grey).

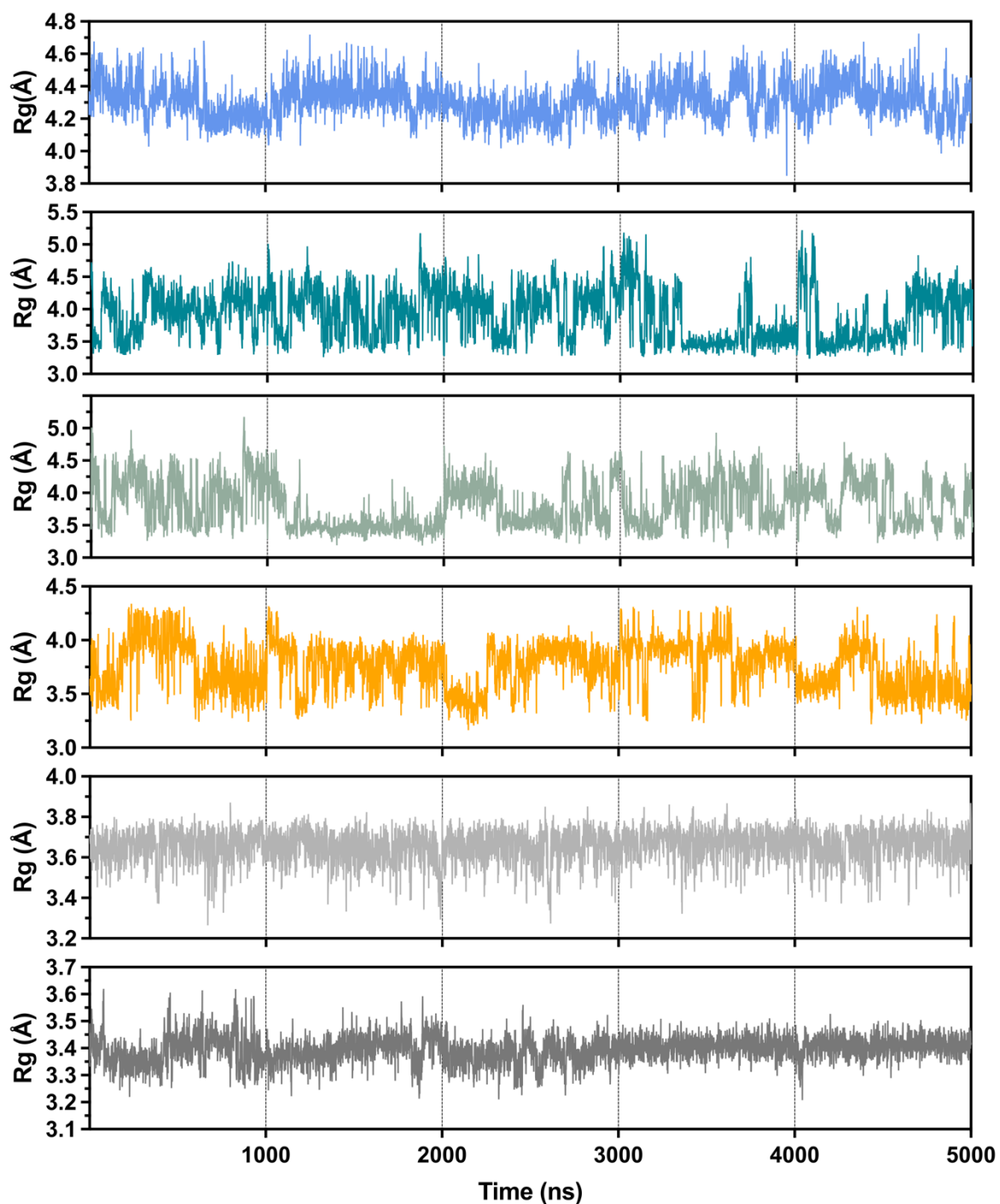


Fig. S7. Radius of Gyration (Rg) of CAR3 individual system. Colours are described as follow: CITCO (purple) and test ligands, namely: linear 4-NP₁ (dark green) and 4-NP₂ (light green), 22NP (orange), 33NP (light grey), 35NP (dark grey).

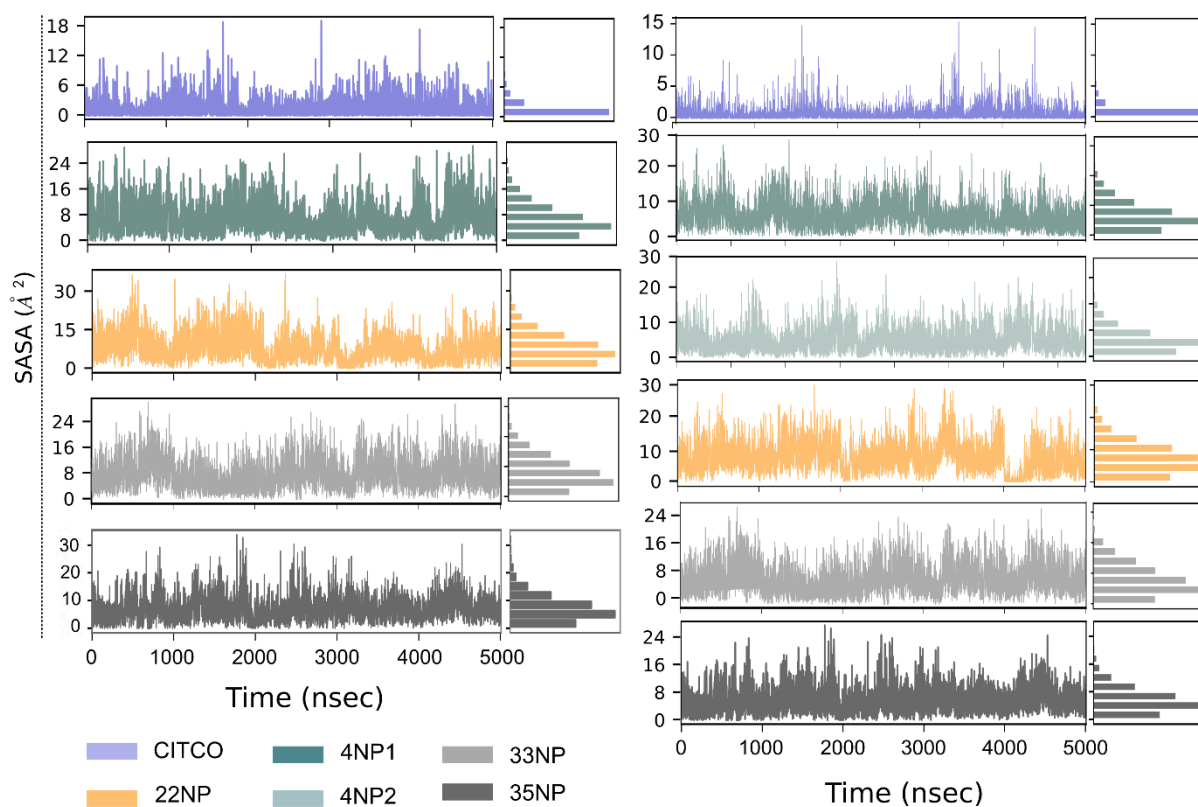


Fig. S8. SASA of individual systems. Colours are described as follow: CITCO (purple) and test ligands, namely: linear 4-NP₁ (dark green) and 4-NP₂ (light green), 22NP (orange), 33NP (light grey), 35NP (dark grey).

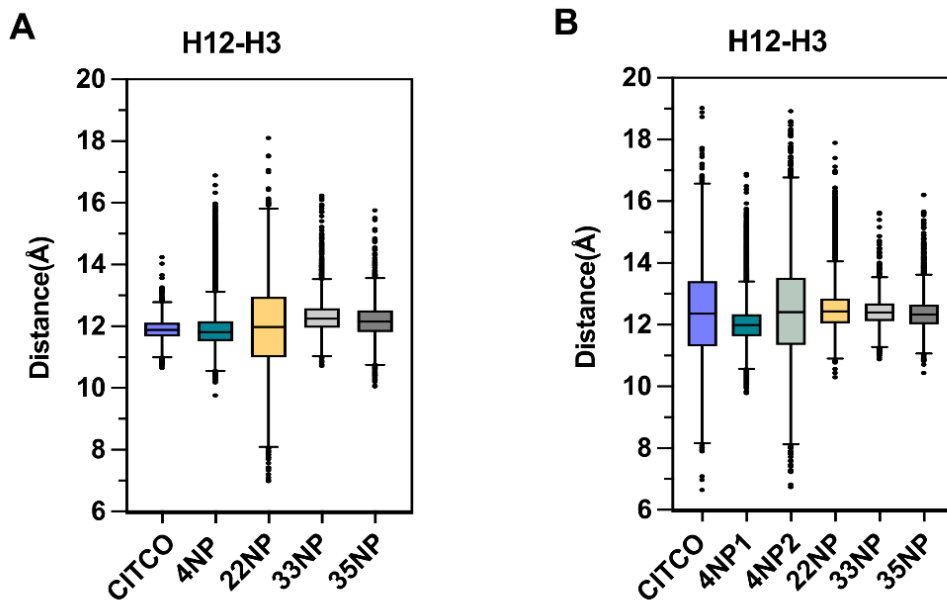


Fig. S9. (A) Distance between α 3-helix (centre of mass of residues 157–178) and α AF-2 (centre of mass of residues 341–348 in CAR1-LBD). **(B)** Distance between α 3-helix (centre of mass of residues 157–178) and α AF-2 (centre of mass of residues 346–353 in CAR3-LBD). Colours are described as follow: CITCO (purple) and test ligands, namely: linear 4-NP₁ (dark green) and 4-NP₂ (light green), 22NP (orange), 33NP (light grey), 35NP (dark grey).

Supplementary Tables

Table S1. Docking poses for CAR1

Compound	docking score	glide emodel	glide energy
35NP	-7.474	-39.654	-27.567
35NP	-7.381	-39.785	-27.557
33NP	-7.377	-40.046	-28.742
33NP	-7.133	-37.898	-28.052
33NP	-7.081	-37.747	-27.902
22NP	-6.911	-42.483	-31.413
33NP	-6.839	-39.237	-28.962
22NP	-6.789	-30.812	-30.813

Table S2. Docking poses for CAR3

Compound	glide score	glide emodel	glide energy
35NP	-7.984	-42.603	-29.767
35NP	-7.758	-41.615	-29.087
33NP	-7.462	-41.074	-30.305
33NP	-7.350	-39.985	-28.854
35NP	-7.283	-41.510	-28.947
33NP	-7.052	-39.300	-29.500
33NP	-7.027	-39.544	-28.705
33NP	-6.927	-39.895	-28.720
22NP	-6.851	-40.891	-29.635
22NP	-6.827	-40.285	-29.348
22NP	-6.712	-38.671	-28.594
33NP	-6.635	-39.326	-29.335

Table S3. Protein-ligand Hydrophobic interactions in CAR1-LBD individual system.

Isoform	CAR1					
	residue					
compound	C202(H5)	L206	F217	F234	L242	Y326
CITCO	60	-	-	40	40	22
4NP1	-	25	22	20	20	30
22NP	-	10	35	15	12	-
33NP	-	10	30	8	11	5
35NP	-	12	35	10	13	-

Table S4. Protein-ligand Hydrophobic interactions in CAR3-LBD individual system.

Isoform	CAR3					
	residue					
compound	C202(H5)	L206	F217	F234	L242	Y331
CITCO	60	-	-	30	40	22
4NP1	-	25	25	15	20	30
4NP2	-	33	25	20	22	-
22NP	-	20	45	12	15	-
33NP	60	-	45	-	-	-
35NP	-	20	53	10	14	-

Table S5. RMSF of α 8- α 9 loop in CAR1-LBD

Residue	Compounds				
	CITCO	4NP	22NP	33NP	35NP
P270	0.696	1.022	0.961	0.985	0.934
D271	0.873	1.199	1.138	1.180	1.109
R272	0.808	1.009	0.979	1.014	0.950
P273	1.162	1.629	1.644	1.635	1.550
G274	0.910	1.117	1.128	1.154	1.077
V275	0.842	1.003	1.020	1.048	0.965
T276	0.899	1.145	1.171	1.174	1.092
Q277	0.822	1.123	1.138	1.133	1.071
R278	0.784	1.043	1.037	1.049	0.984
D279	0.860	1.160	1.174	1.154	1.098
E280	0.756	1.043	1.053	1.014	0.970

Table S6. RMSF of α 8- α 9 loop in CAR3-LBD

Residue	Compounds					
	CITCO	4NP1	4NP2	22NP	33NP	35NP
P270	1.75	2.84	1.35	2.902	1.901	3.315
A271	2.50	3.34	2.37	3.002	2.442	3.633
P272	4.24	3.61	2.77	4.186	4.211	4.913
Y273	5.77	4.77	4.01	5.408	5.320	6.044
L274	6.29	4.80	4.51	5.286	5.118	6.200
T275	4.77	3.95	3.63	4.149	4.672	5.325
D276	4.24	4.72	4.14	3.969	4.142	4.743
R277	3.57	3.57	3.02	2.661	3.197	3.734
P278	3.02	2.86	3.00	3.110	2.728	3.087
G279	3.36	2.41	2.22	2.324	1.943	2.190
V280	2.52	1.51	1.81	1.735	1.438	1.870

IX. Acknowledgements

Here, I would like to express my gratitude to exceptional individuals and institutions who have played pivotal roles in shaping and enriching my learning journey.

I would like to immensely thank my supervisors: Prof. Dr. Antti Poso for providing me with the opportunity to delve into the realm of in silico drug design, which has opened new horizons for personal and professional growth, for his willingness to share his many years of experience with computational chemistry, for his insightful guidance and feedback, and for providing a peaceful environment for carrying out this research. Additionally, I express my thanks to Prof. Dr. Stephan Laufer for his expert supervision and invaluable feedback, collaboration, and his commitment to fostering excellence among his students. It has been an honour and a privilege to work under their guidance.

I extend my heartfelt appreciation to Dr. Thales Kronenberger who patiently supervised from the fundamental stages to more advanced level, whose expertise, feedback, and encouragement have been undoubtedly played a pivotal role in shaping the quality and direction of my research. Special thanks to Dr. Tatu Patsar, whose knowledge, advice and feedback have significantly enriched this thesis.

I am grateful to Prof. Dr. Michael Laemmhofer for providing me with the opportunity to participate in the Erasmus program, broadening my horizons and enhancing my understanding of different cultures.

I would like to acknowledge Prof. Dr. Petr Pavek at Charles University for his support and resources, which have been instrumental in my research endeavours.

I am appreciative to the Finland CSC Centre, for providing the opportunity to work in HPC cluster and providing the computational resources necessary for conducting complex simulations and data analysis.

Furthermore, I express my gratitude to European Union's Horizon 2020 research and innovation programme under grant agreement No. 825762 for their financial assistance, enabling me to pursue my research goals.

I would like to sincerely thank my parents for their unwavering love and support. Their sacrifices and encouragement have been the driving force behind my academic pursuits. I am deeply appreciative of all the individuals and institutions mentioned above. Their impact on my academic journey will always be cherished.