Studying the Conformational Changes of Protein Kinases Using In Silico Structural Approaches

Dissertation

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

vorgelegt von
Ekaterina Shevchenko
Aus Moskau, Russland

Tübingen
2022

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

Tag der mündlichen Qualifikation:	06.02.2023
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter:	Prof. Dr. Stefan Laufer
2. Berichterstatter:	Prof. Dr. Antti Poso

I. Summary

Protein kinases were brought to the scientific community's attention with the remarkable approval of imatinib more than 20 years ago. This approval was not only a breakthrough in targeted cancer therapies but also laid the foundation for in-depth exploration of protein kinases. Since then, enormous efforts have been made to identify, characterise, and investigate the dynamic and function of this broad protein family. Nowadays, protein kinases are associated with numerous human diseases, including the origin of cancer and beyond. In addition, mutations in the highly dynamic catalytic domain of protein kinases can lead to the deregulation of the cell machinery, whilst the key role of protein kinases in the phosphorylation of the downstream protein is crucial for numerous regulation pathways.

Extensive research of protein kinases led not only to the steady development and approval of new inhibitors and therapeutics, but also to the awareness that there are still plenty of questions to be answered about these dynamic proteins. Dimerisation effects, hydrophobic interaction networks, and modest geometrical alterations in the catalytic domain (and beyond) lead to numerous effects that are still to be understood. Worth mentioning is the existence of the so-called dark kinome, which contains understudied kinases. Together, this basis makes protein kinases an intriguing yet demanding research topic. One of the reasons for this challenge is their highly dynamic nature, which is hard to capture with current experimental methods, especially on the atomistic level.

Hence, the In Silico methods come in hand to study their dynamics. Offering a wide range of novel computational techniques, they provide a possibility to lift the veil of secrecy and glance into the minor details of kinase behavior. Therefore, the introduction of this thesis is composed of two chapters, covering the main research focus – protein kinases, and the central methodology – computer-aided drug design. The first chapter is dedicated to protein kinases and offers a broad overview of classification, function, and drug discovery trends, along with a detailed review of the kinase domain's structural features. Finally, the In Silico drug discovery chapter provides a comprehensive overview of computational chemistry methods, with the main emphasis on the application and analyses of molecular dynamics.

This thesis has resulted in three scientific publications, included in the results and discussion section. In the first manuscript, long-scale molecular dynamics is utilized to gain insights into the impact of phosphorylation and mutation on the autoinhibition and dimerization of mitogen-activated protein kinase kinase 4 (MKK4). The second publication investigates the statistical trends and patterns related to protein kinase regulatory hydrophobic spine (R-spine), emphasizing the alfa-helical hydrophobic spine residue three (RS3). Finally, the third

publication presents the long-scale molecular dynamic application for studying the inhibition impact on the tetramerization of the enzyme enoyl-ACP reductase (FabI).

The results of these publications demonstrate the successful application of In Silico methods – specifically molecular dynamics for drug design purposes. Additionally, a few framework modifications are provided in the first and third publications for the exploration of geometrical motions within protein structure throughout molecular dynamics simulation.

II. Zusammenfassung

Proteinkinasen erlangten vor mehr als 20 Jahren durch die bemerkenswerte Zulassung von Imatinib die Aufmerksamkeit der wissenschaftlichen Gemeinschaft. Diese Zulassung war nicht nur ein Durchbruch bei zielgerichteten Krebstherapien, sondern legte auch den Grundstein für eine eingehende Erforschung von Proteinkinasen. Seitdem wurde intensive Forschung betrieben, um die Dynamik und Funktion dieser breiten Proteinfamilie zu identifizieren, zu charakterisieren und zu untersuchen. Heutzutage werden Proteinkinasen mit zahlreichen menschlichen Krankheiten in Verbindung gebracht, einschließlich der Onkologie und auch weit darüber hinausgehend. Weiter führen Mutationen in der hochdynamischen katalytischen Domäne von Proteinkinasen zur Deregulierung der Zellmaschinerie, wobei die Schlüsselrolle von Proteinkinasen bei der Phosphorylierung des nachgeschalteten Proteins für zahlreiche Signalwege entscheidend ist.

Die langjährige Erforschung der Proteinkinasen führte nicht nur zur stetigen Entwicklung und Zulassung neuer Hemmstoffe und Therapeutika, sondern auch zu der Gewißheit, daß es noch viele Fragen über diese dynamischen Proteine zu beantworten gibt. Dimerisierungseffekte, hydrophobe Interaktionsnetzwerke und bescheidene geometrische Veränderungen in der katalytischen Domäne (und auch außerhalb) führen zu zahlreichen Effekten, die es noch zu verstehen gilt. Hervorzuheben ist auch die Existenz des sogenannten Dark Kinome, welches nicht annotierte und nicht erforschte Kinasen enthält. Zusammengenommen macht dies die Proteinkinasen zu einem äußerst interessanten, aber auch herausfordernden Forschungsthema. Einer der Gründe dafür ist ihre hochdynamische Natur, die mit den derzeitigen experimentellen Methoden, insbesondere auf atomistischer Ebene, nur schwer zu erfassen ist.

Daher kommen die In-Silico-Methoden zum Einsatz, um dynamisch verändernde Objekte zu untersuchen. Sie bieten eine breite Palette neuartiger Rechentechniken und ermöglichen einen Blick auf die kleinen Details der Kinase-Lebensdauer.

Daher besteht die Einleitung dieser Dissertation aus zwei Kapiteln, die den Forschungsschwerpunkt der Proteinkinasenund die zentrale Methodikdes computergestützten Wirkstoffdesigns behandeln. Das erste Kapitel ist den Proteinkinasen gewidmet und bietet einen umfassenden Überblick über die Klassifizierung, die Funktion und die Trends in der Arzneimittelentwicklung sowie einen detaillierten Überblick über die strukturellen Merkmale der Kinasedomäne. Das Kapitel über die In-Silico-Wirkstoffentdeckung schließlich bietet einen umfassenden Überblick über die Methoden der Computerchemie, wobei der Schwerpunkt auf der Anwendung und der Analyse der Molekulardynamik liegt.

Diese Dissertation hat zu drei wissenschaftlichen Publikationen geführt, die im Kapitel Ergebnisse und Diskussion vorgestellt werden. Im ersten Manuskript wird die Molekulardynamik eingesetzt, um Einblicke auf die Einflüsse von Phosphorylierungen und Mutationen der Autoinhibition und der Dimerisierung der Mitogen-aktivierten Proteinkinase Kinase 4 (MKK4) zu gewinnen. In der zweiten Publikation werden die statistischen Trends und Muster im Zusammenhang mit dem regulatorischen hydrophoben Spine (R-Spine) der Proteinkinase untersucht, wobei der alfa-helische hydrophobe Spine-Residue drei (RS3) im Vordergrund steht. Schließlich wird in dem dritten Manuskript die Anwendung der molekularen Dynamik im großen Maßstab zur Untersuchung der Auswirkungen der Hemmung auf die Tetramerisierung des Enzyms Enoyl-ACP-Reduktase (FabI) vorgestellt.
Die Ergebnisse dieser Veröffentlichungen zeigen die erfolgreiche Anwendung von In-Silico-Methoden, insbesondere der Molekulardynamik, für die Entwicklung von Arzneimitteln. Darüber hinaus werden in der ersten und dritten Veröffentlichung einige Modifikationen der Rahmenbedingungen für die Erforschung geometrischer Bewegungen innerhalb der Proteinstruktur durch die Simulation der Molekulardynamik vorgestellt.

III. List of publications

Publication I

E. Shevchenko, A. Poso, T. Pantsar, The autoinhibited state of MKK4: Phosphorylation, putative dimerization and R134W mutant studied by molecular dynamics simulations, *Computational and Structural Biotechnology Journal,* Volume 18 (**2020**),
DOI: 10.1016/j.csbj.2020.09.017

Publication II

E. Shevchenko, T. Pantsar, Regulatory spine RS3 residue of protein kinases: a lipophilic bystander or a decisive element in the small-molecule kinase inhibitor binding? *Biochemical Society Transactions* **(2022)**,
DOI: 10.1042/BST20210837

Publication III

V. Gonçalves Maltarollo, E. Shevchenko, I. D. de Miranda Lima, E. A. Cino, G. M. Ferreira, A. Poso, T. Kronenberger, Do Go Chasing Waterfalls: Enoyl Reductase (FabI) in Complex with Inhibitors Stabilizes the Tetrameric Structure and Opens Water Channels. *Journal of Chemical Information and Modeling* **(2022)**,
DOI: 10.1021/acs.jcim.2c01178

IV. Author contributions

Publication I

The autoinhibited state of MKK4: Phosphorylation, putative dimerisation and R134W mutant studied by molecular dynamics simulations.

Ekaterina Shevchenko:

Conceptualisation and investigation
Methodology and formal analysis
Data curation and visualisation
Writing of the manuscript
Discussion of the results and interpretation

Prof. Dr. Antti Poso:

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

Publication II

Regulatory spine RS3 residue of protein kinases: a lipophilic bystander or a decisive element in the small-molecule kinase inhibitor binding?

Ekaterina Shevchenko:

Methodology and formal analysis

Discussion of the results and interpretation

Data curation and visualisation

Writing of the manuscript

Dr. Tatu Pantsar:

Data curation and visualisation

Study supervision

Writing of the manuscript

Discussion of the results and interpretation

Conceptualisation and methodology

Corresponding author

Publication III

Do Go Chasing Waterfalls: Enoyl Reductase (FabI) in Complex with Inhibitors Stabilises the Tetrameric Structure and Opens Water Channels.

Prof. Dr. Vinicius Gonçalves Maltarollo:

Conceptualisation and investigation

Discussion of the results and interpretation

Ekaterina Shevchenko:

Formal analysis

Data curation and automation

Proofreading of the manuscript

Igor Daniel de Miranda Lima:

Water analyses methodology and execution

Proofreading of the manuscript

Prof. Dr. Elio A. Cino:

Water analyses methodology and execution

Proofreading of the manuscript

Dr. Glaucio Monteiro Ferreira:

Markov State Model generation and interpretation

Proofreading of the manuscript

Prof. Dr. Antti Poso:

Proofreading of the manuscript

Study supervision

Corresponding author

Dr. Thales Kronenberger:

Conceptualisation, methodology and investigation

Discussion of the results and interpretation

Study supervision

Corresponding author

V. Poster presentations

DPhG (German Pharmaceutical Society) Annual meeting 2021, virtual event, September 29th - October 1st

The autoinhibited state of MKK4: Phosphorylation, putative dimerisation and R134W mutant studied by molecular dynamics simulations.

E. Shevchenko, A. Poso, T. Pantsar

BioExcel Summer School on Biomolecular Simulations 2022, Science and Technology Park of Sardinia, Pula, Italy**,** June 12th - 27th.

Structural determinants of Liver receptor X modulation.

E. Shevchenko, J.G. Pedreira, T. Kronenberger

VI. List of abbreviations

ABL	ABL proto-oncogene 1
ADME	Absorption, distribution, metabolism, and excretion
AGC	Protein kinase A, G, and C kinase group
AI	Artificial intelligence
ALK	Anaplastic lymphoma receptor tyrosine kinase
APE	Conserved Ala-Pro-Glu kinase motif
ATP	Adenosine triphosphate
AUC	Area under the curve
BCR	Breakpoint cluster region protein
BRAF	v-Raf murine sarcoma viral oncogene homolog B
CADD	Computer-aided drug design
CAMK	Calcium/calmodulin-dependent kinases
CDK	Cyclin-dependent kinase group
CK1	Casein kinase 1 group
CLK	CDC like kinase kinase group
CMGC	CMGC kinase group
CML	Chronic myelogenous leukaemia
DFG	Conserved Asp-Phe-Glu kinase motif
ePKs	Eukaryotic protein kinases
FDA	Food and Drug Administration
FGFR	Fibroblast growth factor receptor
GK	Gatekeeper residue
GPCR	G-protein-coupled receptor
GSK3	Glycogen synthase kinase 3
HTVS	High throughput virtual screening
IRAK4	Interleukin 1 receptor associated kinase 4
IRE1α	Inositol-requiring enzyme 1α
JAK	Janus kinase
KDR	Kinase insert domain receptor
LBDD	Ligand-based drug design
MAPK	The Mitogen-activated protein kinase
MAP2K	The Mitogen-activated protein kinase 2

MAP2K3	Mitogen-activated protein kinase 3
MD	Molecular dynamics
MEK1/2	Mitogen-activated protein kinase kinase 1
MKK4	Mitogen-activated protein kinase kinase 4
mTOR	Mechanistic target of rapamycin
MW	Molecular weight
NMR	Nuclear magnetic resonance
NSCLC	Non-small-cell lung cancer
NVE	Microcanonical ensemble
NpT	Isothermal-Isobaric Ensemble
NVT	Canonical ensemble
PBC	Periodic boundary conditions
PCA	Principal component analysis
PDB	Protein Data Bank
PDGFR	Platelet derived growth factor receptor
PKA	Protein kinase A
PKB	Protein kinase B
PKC	Protein kinase C
PKG	Protein kinase G
Po/w	n-octanol-water partition coefficient
PROTAC	Proteolysis targeting chimera
PSA	Polar surface area
QSAR	Quantitative Structure-Activity Relationships
RAS	Rat sarcoma virus protein family
RET	Ret proto-oncogene
RGC	Receptor Guanylate Cyclases kinase group
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
ROC	Receiver operating characteristics
ROCK1	Rho associated coiled-coil containing protein kinase 1
ROCK2	Rho associated coiled-coil containing protein kinase 2
RS1	Regulatory spine residue 1
RS2	Regulatory spine residue 2
RS3	Regulatory spine residue 3

RS4	Regulatory spine residue 4
SAR	Structure-activity relationship
SASA	Solvent-accessible surface area
SBDD	Structure-based drug discovery
SP	Standard precision docking
STE	Serine/threonine kinase group
SYK	Spleen tyrosine kinase
TIP3P	Three-site rigid water molecule
TK	Tyrosine kinase group
TKL	Tyrosine kinase-like group
VEGFR2	Kinase insert domain receptor
XP	Extra precision docking

1. Introduction

1.1 protein kinases

Eukaryotic protein kinases (ePKs) are a significant family of regulatory enzymes which act as switches in the cell. Kinases catalyse the transfer of ATP γ-phosphate to the target proteins. ePKs are involved in the regulation of the major cell biology events by phosphorylation, which controls the activity of the downstream targets [1–3].

Due to the existence of protein kinases dysregulation in a number of illnesses, such as cancer, inflammation, and autoimmune disorders[2], as well as the fact that function distortion leads to dysregulated cellular signalling and disease development [4,5], this druggable protein family is of great interest. The prominence of ePKs in the signalling pathways that control the traits of malignant cells, taken together with their druggability, underline protein kinases as important drug targets [6,7]. This fact led to the creation of a significant body of sequence, structure, bioactivity, and mutation data by the scientific community. Nevertheless, attempts are still being made to elucidate the biological function of the poorly understood kinases, also referred to as the dark kinome [8]. In recent years, the hunt for additional targets has increased due to the advent of resistance mechanisms to currently available kinase-targeting treatments [9].

1.1.1 The kinome

One of the largest superfamilies of homologous proteins, the kinome, is encoded by approximately 2% of the human genome and comprises over 500 proteins[10]. The majority of protein and atypical kinases were identified as a result of human genome sequencing[10]. Based on the sequence similarity, the human kinome is divided into eight main groups, complemented by 13 atypical families[10]. However, the number of groups may vary based on the chosen classification method and the rapidly increasing pool of data regarding kinases. The names of kinases frequently originate from the modulators that regulate their activity. For instance, the kinases modulated by cyclin are termed CDK or cyclin-dependent kinases. Another example is the kinases regulated by Ca^{2+}–calmodulin, that form CAMK or Calmodulin-dependent kinase group. Sometimes the further subdivision of a kinase group is based on the presence of isoenzymes, as with the type I (PKRAR1) and type II (PRKAR2) cyclic-AMP dependent protein kinases, which share homologous catalytic subunits but distinct regulatory subunits that bind cyclic AMP[11].

In particular cases, the kinase classification follows the biological function rather than relies on the sequence similarity within the group members. Examples of this phenomenon are the Mitogen-activated protein kinases (MAPKs), which form a signalling cascade of sequentially

phosphorylated kinases [12]. Each cascade starts with an extracellular stimulus that leads to activation of MAPK kinase kinase (MAPKK), followed by phosphorylation of MAPK kinase (MAPKK), and finally resulting in the activation of a particular MAPK (i.e. ERK, p38 or JNK kinase) [13]. While the MAPK kinases are involved in the same transducing pathway, the MAPK genes belong to the CMGC group, MAP2K and MAP2K3 to the STE group, and finally, MAP2K4 genes originate from both STE and TKL groups[8].

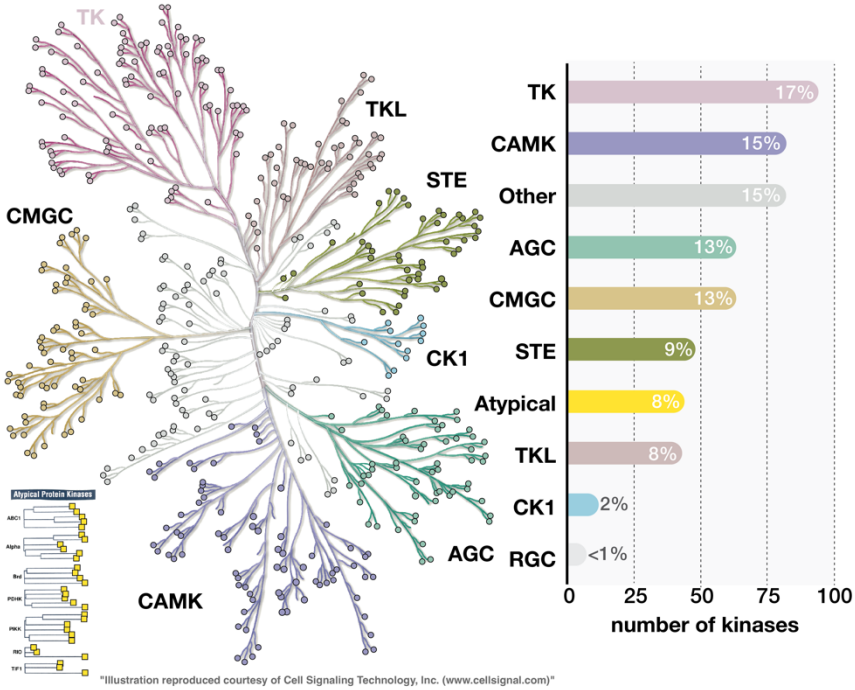


Figure 1. Kinase distribution by groups in the human kinome. The colours of the chart bars are as the groups in the kinome. The human kinome tree illustration was made with the help of KinMap[14]. RGC group is not present on the kinome tree.

Protein kinases are divided into eight main groups: tyrosine kinases (TK), Calcium/calmodulin-dependent kinases (CAMK), AGC, CMGC, serine/threonine kinases (STE), tyrosine kinase-like (TKL), casein kinase 1 (CK1), and receptor Guanylate Cyclases (RGC). ePKs without conserved kinase motifs or altered regulatory and catalytic spines comprise Atypical kinases group [15]. Finally, ePKs that could not be categorised into the above groups form an additional 'Other' group. Further, each kinase group will be described separately.

1. TK group

TK is the most populated kinase group according to KinMap[14], comprising 94 kinases or 17% of the total kinome (Figure 1, pink). In contrast to the majority of other kinases, which are

selective for serine or threonine residues, this group nearly exclusively catalyses the phosphorylation of tyrosine residues in the target protein [10].

2. CAMK

CAMK is the second most populated group, with 82 members or 15% of the total kinome (Figure 1, violet). CAMK kinases phosphorylate either serine or threonine residues in the target proteins and are regulated by the concentration fluctuation of intracellular calcium ions[16].

3. AGC

The term AGC is formed from the names of subsequent kinase families. Namely, cAMP-dependent protein kinase 1(PKA, alternatively referred to as PKAC), protein kinase C (PKC), and cGMP-dependent protein kinase (PKG, alternatively referred to as CGK1α)[17]. The AGC group includes 63 protein kinases, constituting 13% of the total kinome (Figure 1, cyan). The unique feature of the AGC group is in the activation mechanism. The activation of AGC kinases frequently involves phosphorylation in the co-called hydrophobic motif, which is located outside the kinase domain [17].

4. CMGC

CMGC group name originates from a set of member families, where 'c' stands for cyclin-dependent kinase (CDK), 'm' for mitogen-activated protein kinase (MAPK), 'g' for glycogen synthase kinase (GSK3), and 'c' for CDC like kinase kinase (CLK). CMGC consists of 63 protein kinases, comprising 13% of the total kinome (Figure 1, beige). Special features of CMGC kinases include the co-called CMGC-insert region and the CMGD arginine. These regions are found in the activation loop, close to the phosphorylation site [18,19].

5. STE

STE, or serine/threonine kinases with 48 members, comprise 9% of the total kinome (Figure 1, green). One of the main functions of the STE kinases is the MAPK cascade signal transduction [20].

6. TKL

The tyrosine kinase-like (TKL) group, with 43 members or 8% of the total kinome, comprises diverse functions and contains a range of families with low correlation to each other (Figure 1, dusty pink). Furthermore, the TKL and TK groups share high sequence similarity. One example of TKL kinases is the RAF family, a crucial component of the MAPK kinase pathway.

7.CK1

The casein kinase 1 (CK1) group is a small group with 12 members or 2% of the total kinome (Figure 1 blue). CK1 is a group of kinases that share high sequence similarity and differ

significantly from the other kinase groups [21]. The known activation mechanisms of CK1 include autophosphorylation and activation via interaction with cellular proteins [22].

8. RGC

Receptor Guanylate Cyclases is the smallest kinome group, consisting of just seven kinases (Figure 1, light grey). RGC are transmembrane proteins with an intracellular kinase domain that is catalytically inactive[23].

9. Other kinases

Besides kinases that could not be classified into the above groups, this group include several unique kinases [23]. Other kinases form the third largest group in the kinome, with 82 members, comprising 8% of the total kinome (Figure 1, grey).

10. Atypical kinases

Atypical kinases comprise 44 proteins, constituting 8% of the total kinome (Figure 1, indicated with yellow squares). These proteins exhibit conventional kinase function and activity but lack the usual highly conserved kinase domain or possess a low degree of sequence similarity [10,15].

1.1.2 Kinase structural features

Protein kinases can be soluble (nuclear or cytoplasmatic) or transmembrane proteins, containing several domains that manage kinase activity, multimerisation, or serve as a regulatory module to control the recruitment of binding partners[24,25]. The architecture of a protein kinase can be divided into three major parts: the intracellular, the transmembrane, and the extracellular domains. The precise mechanism of the domains' machinery is unknown to date [26]. The intracellular domain is described as responsible for the substrate phosphorylation and formation of multimers. The kinase transmembrane region consists of a single helix and serves as a link for signal transduction between the domains[27]. The extracellular domain contains a ligand's binding site and is primarily known as the kinase catalytic domain. Upon the ligand binding complex, conformational rearrangements occur in other domains, leading to oligomerisation and activating signalling cascades [26,28]. Further information in this thesis is dedicated to the catalytic kinase domain.

The remarkable conservation of kinases' catalytic domain has already been known for several decades, with sequencing analyses identifying prominent motifs (hinge, catalytic Lysine, DFG (Asp-Phe-Glu), and APE (Ala-Pro-Glu) motifs) were identified [10]. A few years later, the first crystal structure of a kinase catalytic domain was solved (cAMP-dependent protein kinase PDB ID: 2cpk [29], 1991), which supported the initial sequence analysis with 3D data and allowed for the first function predictions [11]. The following structure, published in 1993, described the

rigorous process for docking ATP into the active site cleft [30] (cAMP-dependent protein kinase PDB ID: 1amp [30]). Ever since, various unique structural characteristics of protein kinases have been identified, including dynamic regulatory elements, activation mechanisms, and violin models of kinase allostery. A comprehensive historical overview of development in the area of protein kinases is presented by Taylor et al. [31] and R. Roskoski[32]. The kinase exists in an active or inactive state depending on how the conserved structural components are arranged [13–15]. Since eukaryotic kinases are highly dynamic, with numerous conformations occupying the two functional states, rather than having a single defined active and inactive conformation like many enzymes, this categorisation is immensely challenging [16]. The transfer of ATP γ-phosphate is the event that all kinases share with the active state.

1.1.2.1 Kinase catalytic domain

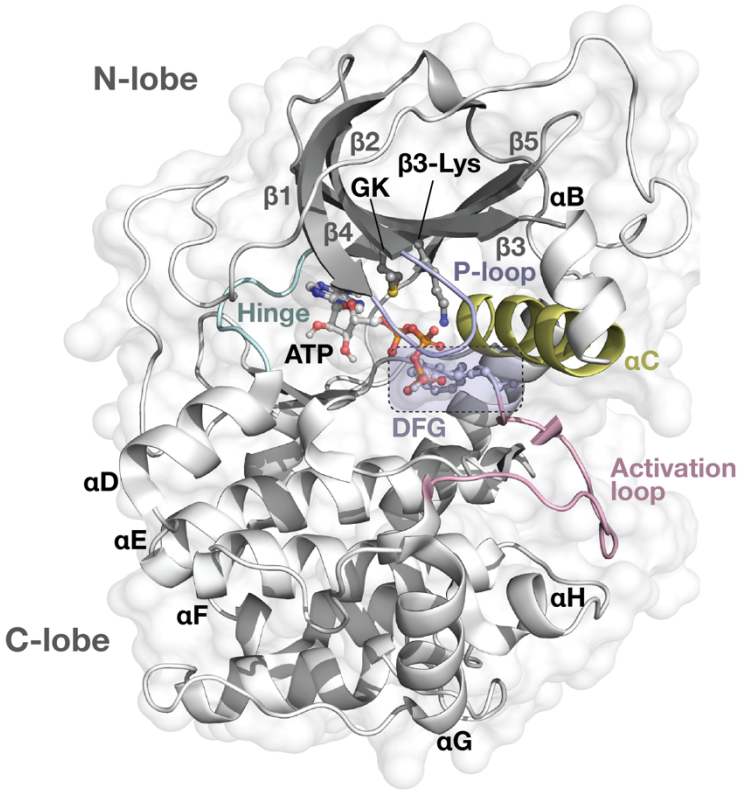


Figure 2. **The structural overview of the conserved kinase catalytic domain** (cAMP-dependent protein kinase, PDB ID: 4wb5[33]). ATP, β3-Lys, GK, and DFG are shown with the stick model. GK- gatekeeper residue. The αC-helix is depicted with a yellow cartoon, the DFG motif with violet, the activation loop with pink, and the hinge region with cyan.

A protein kinase's core is composed of two domains linked together via the hinge region: the highly dynamic N-lobe, which contains five antiparallel β-sheet strands, and the more stable C-lobe, which in most cases contains αD-αI helices connected via loops (Figure 2)[31,34,35]. The role of the C-lobe in substrate phosphorylation is to bind the substrate and move it in close proximity to the ATP, which enables substrate phosphorylation[31,34]. Indeed, the deep cleavage that separates the two lobes serves as the binding site for ATP (Figure 3), which is also the primary site for inhibitor development.

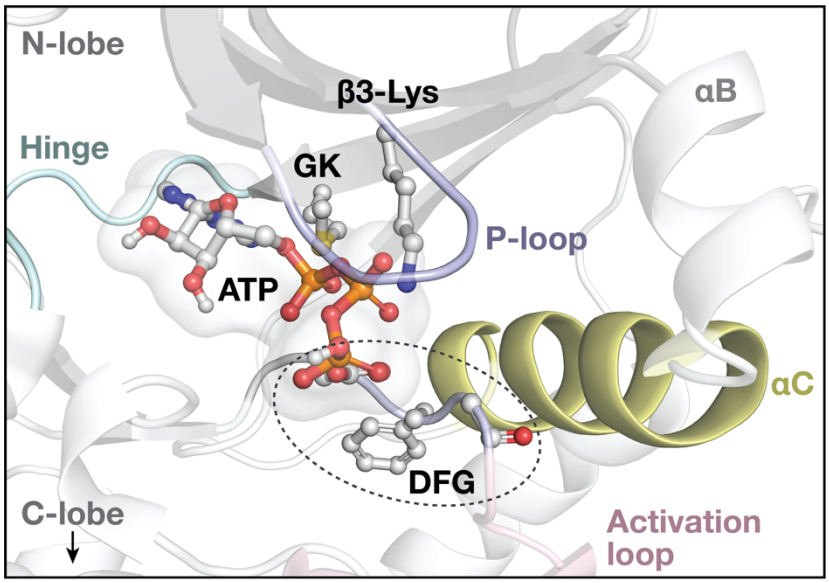


Figure 3. **The ATP binding pocket of protein kinase** (cAMP-dependent protein kinase, PDB ID: 4wb5[33]). ATP, β3-Lys, GK, and DFG are shown with the stick model. GK- gatekeeper residue. The αC-helix is depicted with a yellow cartoon, the DFG motif with violet, the activation loop with pink, and the hinge region with cyan.

The disruption of the interaction network between the lobes is associated with a protruding movement of the αC-helix away from the ATP-binding site, which leads to physiological kinase inactivation. The P-loop, a component of the activation segment, is found next to the αC-helix following the phosphorylation sites (Figure 3). The activation segment is one of the key regulatory elements of the kinase; upon the change of position, it can affect both kinase catalytic performance and substrate binding [36]. The αC-helix performs its regulatory function with the help of a salt bridge between a conserved glutamate residue and the catalytic lysine, another conserved residue in the β3 sheet [37]. When this salt bridge occurs, the αC-helix closes, covering the binding site- this state, known as αC-in conformation, is a sign of the active state of kinase. In the inactive state, the αC-helix moves in the opposite direction, achieving the co-called αC-out state [38].

The activation segment consists of two function-related parts: the activation loop and the DFG motif (Figure 2). The term DFG comes from the three residues that constitute this motif and are consistent in most kinases: aspartic acid, phenylalanine, and glycine. The loop where DFG is found is characterised as the magnesium positioning loop due to the interaction between the

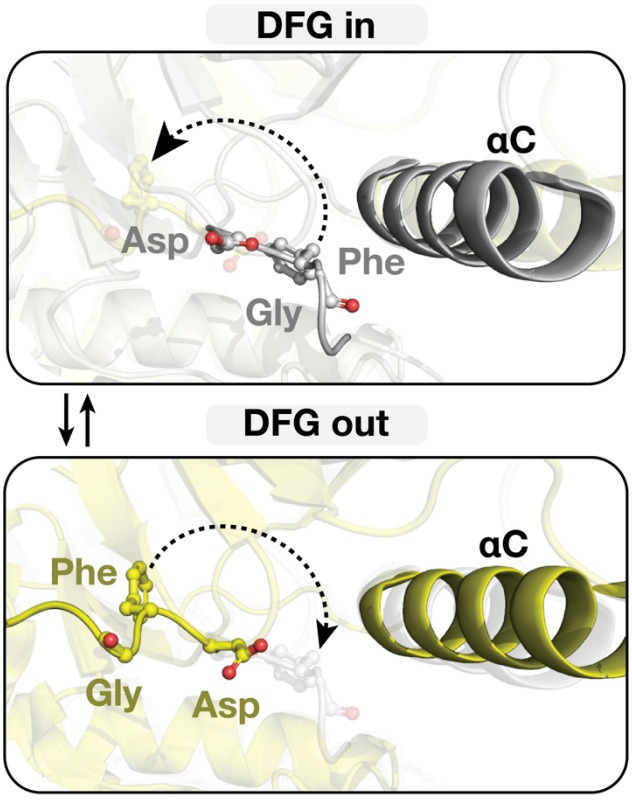


Figure 4. Two key subcategories of the DFG-motif position: DFG-in (BRAF, PDB ID: 2fb8[39]) and DFG-out (BRAF, PDB ID: 4xv9[40]). The outward movement of DFG towards the ATP-binding site is indicated with the dashed line. Asp, Gly, and Phe from DFG are shown with the stick model.

aspartate from the DFG motif and the magnesium inside the binding site[41]. Another feature of the DFG motif is the ability to support the identification of kinase state (active or inactive), depending on its orientation concerning ATP[38] (Figure 4).

Two key subcategories stand out from the numerous classifications based on the locations of residues in DFG: the active, DGF-in and inactive DFG-out[37] (Figure 4). In active DFG-in conformation, the aspartate's side-chain points to the phosphates of ATP, therefore coordinating the magnesium ion [38,42,43]. In inactive DFG-out conformation, the phenylalanine is then switched outwards the ATP-binding site. This outward movement opens a deep

hydrophobic pocket and is followed by a series of events that result in the disorganisation of kinase substructures, such as the activation loop [38,42].

The activation loop comprises the phosphorylation site of the kinase. The phosphorylation site generally refers to either serine and threonine residues or single tyrosine that undergo phosphorylation upon activation. Nevertheless, the phosphorylation mechanisms are highly diverse among the kinases[44], some of which may also employ cis phosphorylation in conjunction with the extra scaffold protein [45]. Moreover, the activation loop encompasses one highly conserved motif consisting of tyrosine or histidine, followed by arginine and aspartate[20]. This motif, termed Y/HRD, participate in the catalytic activation of protein kinases, supporting the ATP's phosphate transfer[29,31,46]. The HRD-Asp is a highly conserved residue that regulates the substrate peptide's correct orientation and acts as a catalytic base[36,47,48]. The end of the activation loop is marked with another motif consisting of alanine, proline, and glutamic acid (APE)[37]. In turn, the APE motif supports the kinase transition to the active state by anchoring the activation loop to the α-F helix[49].

Lastly, the conserved hinge region (Figure 3) harbours the essential gatekeeper, which was early acknowledged as important for inhibitor design[50]. The name 'gatekeeper' originates from the fact that the size and shape of the amino acid that defines the gatekeeper characterise the accessibility to the buried hydrophobic region of the ATP-binding pocket [51]. Important to notice that the variations in gatekeeper amino acid do not influence ATP binding itself but control the ATP analogues (or inhibitors) recognition[52,53]. The gatekeeper residues' control over the buried hydrophobic pocket accessibility is frequently used in drug design approaches for achieving compound selectivity across kinases[54]. Moreover, mutation of the gatekeeper residue frequently results in a kinase that can possess normal functioning but is unable to bind an inhibitor or results in significantly weaker inhibitor-kinase interactions[55].

1.1.2.2 Dynamic regulatory elements

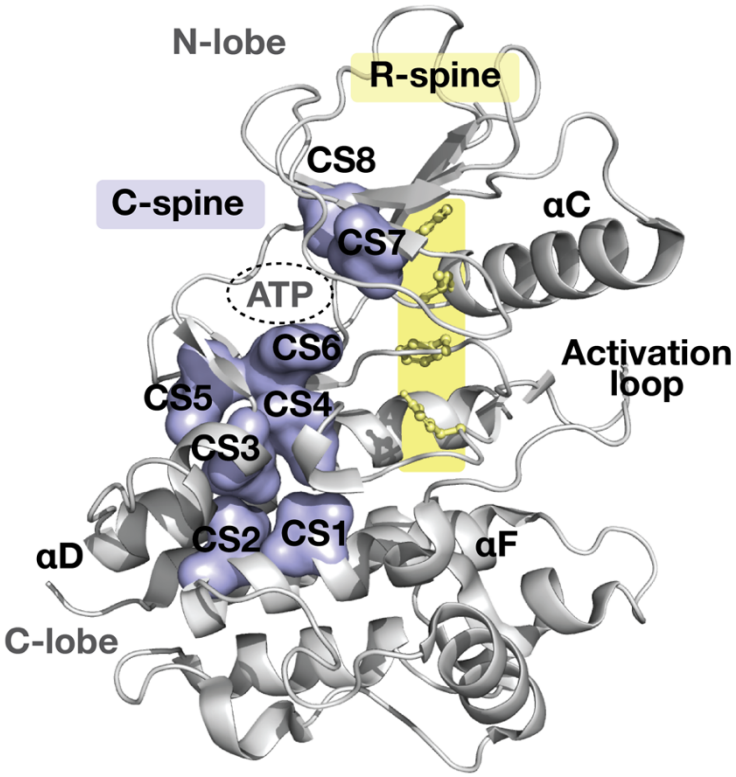


Figure 5. A typical structure of a protein kinase with a highlighted catalytic spine (C-spine) and regulatory spine (R-spine) (BRAF, PDB ID: 4e26[56]). C-spine residues are illustrated with violet surfaces, and R-spine residues are shown with the stick model, coloured in yellow. The ATP binding pocket is highlighted with the dashed oval. CS1–CS8 indicate catalytic spine residue 1–8.

As highly dynamic molecular switches, protein kinases feature a set of conserved motifs that function as a framework to span the entire kinase core [57]. Namely, the regulatory spine (R-spine) and the catalytic spine (C-spine) are the two key dynamic components found in kinases (Figure 5)[58,59]. Located parallel to each other, the spines connect the N and C-lobe of a kinase with a joint point at the α-F helix. The spines' arrangement and machinery have led to frequent comparisons to the human vertebral column, which provides rigidity yet permits multiple independent motions. Though they are attached to the same α-F helix, the two spines serve very different functions. As a component of the kinase dynamic architecture, the R-spine acts as a trigger for the transition into the kinase's active state, while ATP binding then completes the C-spine[60,61]. Even though the spines are conserved structural motifs, they do not share notable sequence similarity among kinases [34].

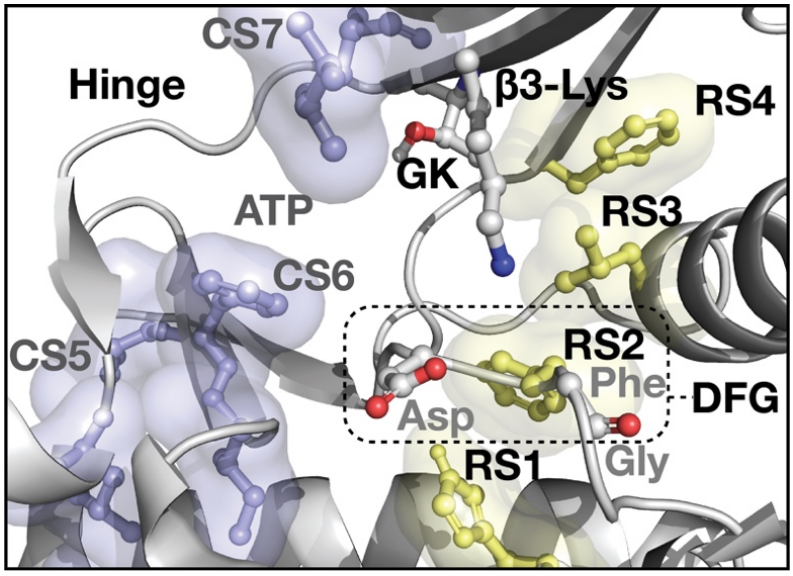


Figure 6. Zoomed image of ATP-binding site in a protein kinase (BRAF, PDB ID: 4e26[56]). The tight and aligned packing of the R-spine indicates the active state of kinase. C-spine residues (violet surfaces) are located next to the R-spine (yellow surfaces), GK – Gatekeeper residue. Spine residues are shown with the stick model.

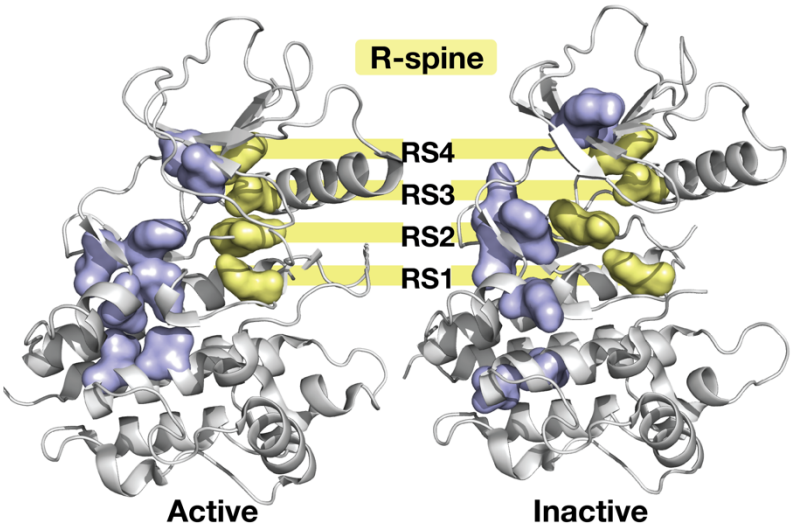


Figure 7. The regulatory spine (R-spine) in active or assembled conformation (BRAF, PDB ID: 4e26[56]) and inactive or dissembled conformation (BRAF, PDB ID: 1uwh[62]). R-spine residues are illustrated with yellow surfaces and C-spine with violet.

Four residues make up the hydrophobic R-spine, two of which are located in the C-lobe (RS1 and RS2), and the other two are found in the N-lobe (RS3 and RS4) (Figure 6). Therefore R-spine provides the connection between the two lobes[34]. It is noteworthy that the RS1 is the first residue of the HRD motif, the RS2 residue is the second residue in DFG, and

the RS3 residue is in the αC-helix, which is crucial for kinase activation. This observation underlines the correlation between conserved structural elements and a consistent sequence pattern of kinases. Coming from an essential activity element of kinase, the R-spine is amassed upon the active conformation of the kinase and is triggered by phosphorylation or kinase interaction with other activating agents, as cyclin in cyclin-dependent kinases [59,63,64] (Figure 7). The R2 residue shifts position in response to activation, supporting the opening and expansion of the activation loop. However, the overall geometry of the R-spine remains unaltered during the phosphoryl transfer process. Moreover, the increased catalytic activity of kinases was associated with mutations in the R-spine, which resulted in its additional stabilisation[64,65].

The C-spine comprises eight hydrophobic residues, with two located in the N-lobe and six in C-lobe and assembled upon the ATP binding [34] (Figure 5). The role of ATP binding for the assembly is related to the ATP's adenine ring, which links the C- and N-lobe parts of the C-spine. When the C-spine is assembled, the kinase aligns both lobes, sealing itself. Summing up, assembled R-spine is a sign of kinase activation, while the assembly of the C-spine indicates kinase preparation for catalysis.

1.1.3 Kinase drug discovery

Protein kinases over decades have been recognised as highly relevant drug targets. Drug design options were first believed to be restricted to ATP displacement in the binding site[37]. The development of ATP competitors is not trivial due to the well-established role of ATP in the cell. Twenty years after the kinases' acknowledgement as drug targets are marked by a series of simultaneous discoveries in a scientific society. Among others are the approval of the ROCK1, ROCK2, and mTOR inhibitors fasudil and sirolimus, along with the discovery of the revolutionary imatinib[37]. Imatinib's clinical success was a driving force behind the outbreak of activity in targeted oncology therapy, where kinase inhibition remains a dominant force to this day.

1.1.3.1 Types of kinase inhibitors

The dynamic nature of protein kinases allowed for the development of inhibitors that can distinguish between various inactive conformations or active state. However, initially should be mentioned that the small molecule kinase inhibitors can be divided into two major classes: reversible binders and covalent inhibitors[66–68]. The complexity of the conformational states and features in the kinase catalytic domain, as well as the range of insights that are continuously

revealed in their dynamics, make further categorization of reversible binders a challenging task [69]. However, the reverse kinase inhibitors can be introduced as following: ATP-mimetic inhibitors, which target the active state of the kinase belong to Type I, to the inactive – Type II; allosteric inhibitors make up the Type III, and surface-pocket binders the Type IV [70]. Several intermediate binding modes do not fit the Type I and II definition. Due to this, type 1.5 inhibition was proposed and is generally used. Further, each type of inhibition is described separately.

Type I. The target of type 1 inhibitors is the active state of a kinase. By interacting with the active conformational site, these kinase inhibitors change the structural conformation that would otherwise be utilised for phosphate transfer[72]. The purine moiety, meant to imitate the ATP hinge-binding, is frequently included in type 1 inhibitors. However, it is important to note that the αC-helix and DFG-motif are in the "in" position with type 1 inhibition (Figure 8). Among various approved type 1 inhibitors in cancer treatment are vemurafenib[73] (BRAF V600E) and gefitinib[74] (EGFR L858R).

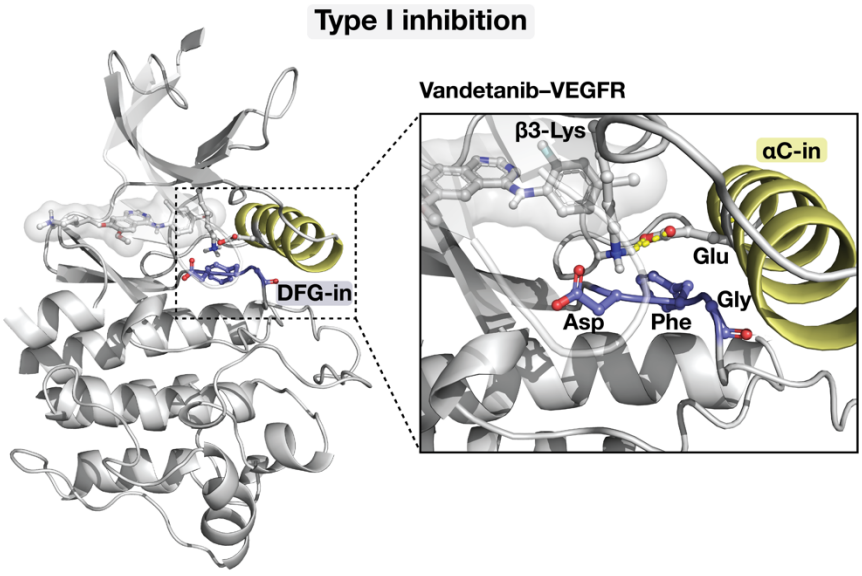


Figure 8. An example of type I kinase inhibitor binding mode: Vandetanib bound to VEGFR (PDB ID: 2ivu [71]). DFG (purple) and αC helix (yellow) exhibit the "in" position. Vandetanib, the DFG residues, αC-Glu, and β3-Lys are shown in the ball and stick representation. Hydrogen bonds are shown as dotted yellow lines.

Type II. The inactive, or DFG-out state of a kinase is targeted by type 2 inhibitors and is connected to the outward movement of DFG moiety, which results in the formation of a large hydrophobic pocket, that may be occupied by the inhibitor. These inhibitors establish additional interactions in the open DFG-out conformation as well as multiple hydrogen bonds in the hinge region of the target kinase[75] (Figure 9). Type 2 inhibitors encompass imatinib and sorafenib, which have become iconic in kinase drug design.

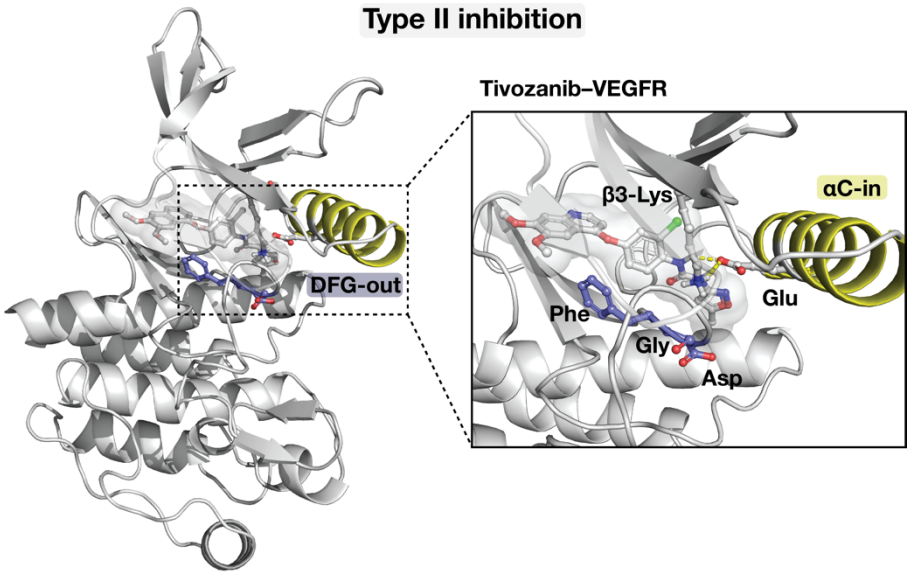


Figure 9. An example of type II kinase inhibitor binding mode: Tivozanib bound to VEGFR (PDB ID: 4ase[76]). DFG (purple) is in the "out" position, while αC helix (yellow) shows the "in" position. Tivozanib, the DFG residues, αC-Glu, and β3-Lys are shown in the ball and stick representation. Hydrogen bonds are shown as dotted yellow lines.

Type 1.5. The concept of type 1.5 inhibition is ambiguous (Figure 10), with several common indications[37]. Namely, the canonical G-loop location or the activation loop may change due to the ligand binding, leading to an abnormal active conformation[55].

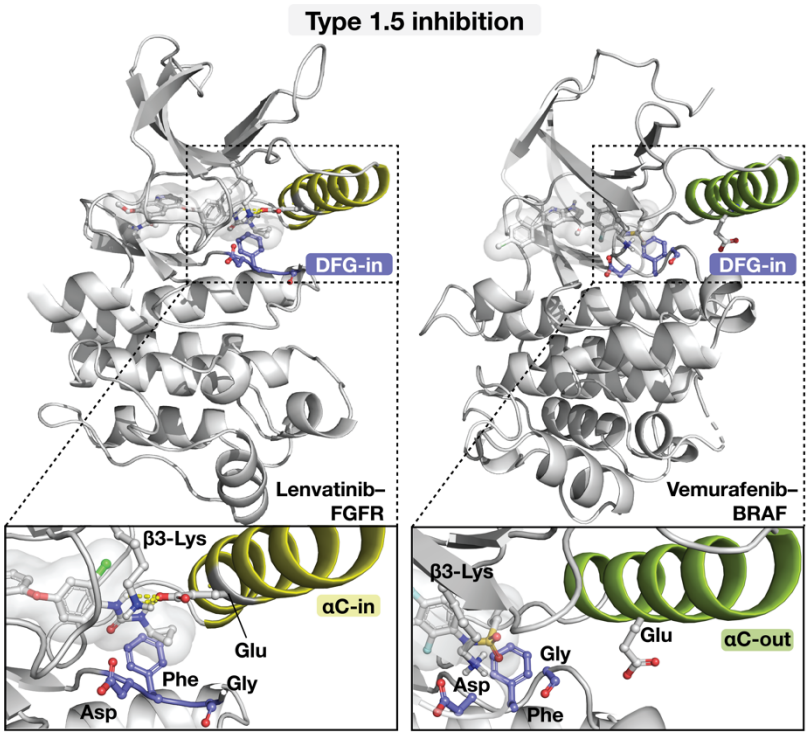


Figure 10. Examples of type 1.5 inhibitor binding modes. Lenvatinib bound to DFG-in, αC-in FGFR1 (PDB ID: 5zv2 [77]), and Vemurafenib bound to DFG-in αC-out BRAF (PDB ID: 4rzv [78]) underline the variation in the binding modes of type 1.5 inhibitors. The ligands and the DFG residues, αC-Glu, and β3-Lys are shown in the ball and stick representation. Hydrogen bonds are shown as dotted yellow lines.

Recently, in the work of Attwood et al.[37], the distortion of the R-spine without occupancy of the hydrophobic DFG-out pocket has been mentioned as a type 1.5 inhibition indicator (Figure 11). The influence or disruption of a particular structural element of a kinase can enhance selectivity and reduce the number of off-targets, making type 1.5 inhibition an appealing strategy for designing novel compounds. For example, the high selectivity of EGFR inhibitor lapatinib is described as connected with the structural rearrangements upon binding to a DFG-in conformation[79].

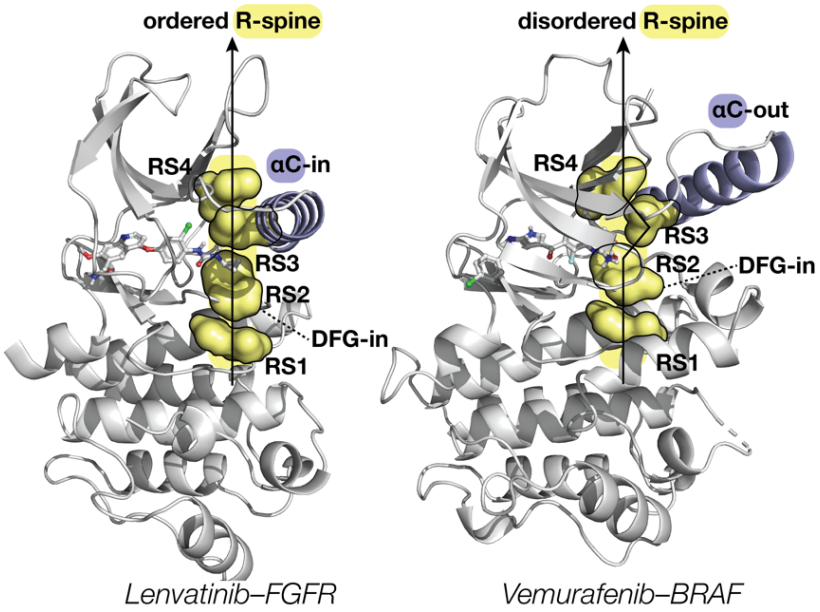


Figure 11. **Type 1.5 inhibitors influence regulatory R-spine differently.** With the Vemurafenib-BRAF complex (PDB ID: 4rzv[78], the αC-helix protrudes outside, which leads to a shift in the RS3 residue position and results in a disordered R-spine. DFG motif maintains the DFG-in position for both complexes. With the Lenvatinib-FGFR1 complex (PDB ID: 5zv2[77]), the R-spine remains in the ordered conformation.

Type III. Type III inhibitors are classified as allosteric, meaning that the inhibitors' binding site lies outside the ATP-binding pocket. Type III inhibitors bind next to ATP cleft, therefore not interfering with ATP binding[80]. Examples of this inhibition type encompass targeting CHEK1 [81], Akt [82], and MEK1[83] kinases.

Type IV. As Type III, Type IV inhibitors are classified as allosteric, with the characteristic feature in binding to remote from ATP sites[80]. An example of a type IV inhibitor is GNF-2 (BCR-Abl inhibitor) which locks kinase in the inactive state without binding to the ATP-cleft [84]. A total of 10 allosteric inhibitors are now undergoing clinical trials; three of these, including the pioneer in allosteric inhibition trametinib (MEK1/2 inhibitor), are already approved[85].

Covalent inhibitors. Another approach of kinase inhibition that has gained interest during the last decade is the covalent targeting [68,86]. Covalent inhibitor opts for targeting nucleophilic amino acids (primarily, but not limited to, cysteines) within the ATP-binding pocket. To date, this approach resulted in six approved kinase inhibitors [68]. Additionally, there are entirely distinct methods of achieving kinase inhibition. For example, macrocycles (ALK inhibitor

lorlatinib [87]), proteolysis-targeting chimeras, or PROTACs, (targeting BRAF V600E mutant [88]).

1.1.3.2 Approved kinase inhibitors in oncology

To date (November 2022), 89 kinase inhibitors targeting over 20 kinase families have received FDA clinical approval[89]. This number represents around 20% of the originally proposed druggable human kinome [32,90]. Tyrosine kinase inhibitors are the most commonly approved kinase inhibitors, although only 30% of all known human protein kinases belong to the tyrosine kinase family[37]. This demonstrates great potential in developing new kinase inhibitors against other understudied families. The driving function of kinases in cancer is supported by comprehensive clinical data related to kinase activation by either activating mutations or translocations [91,92]. One of the aspects of carcinogenesis is genomic instability, which is characterised by a high frequency of mutations and chromosomal alterations across the cellular lineages[93]. The genomic instability frequently results in the appearance of cancer mutations, which are divided into driver and passenger ones. Driver mutations lead the somatic cells to altered, frequently pathologically increased cell proliferation, which in turn affects the tissue microenvironment and contribute to the tumour development and progression[94]. On the contrary, the passenger mutations do not promote this proliferative benefit[94]. According to sequencing data, nearly 97% of cancer mutations are passenger mutations[95]. The development and approval of cancer therapies have been made simpler by identifying and characterising these cancer drivers, starting with the ground-breaking approval of imatinib to treat chronic myelogenous leukaemia (CML) caused by the BCR-ABL translocation, which prompts increased tyrosine kinase activity [97].

Cancer with the leading amount of approved kinase inhibitors is the NSCLC (non-small-cell lung cancer). The most prevalent form of lung cancer, NSCLC, accounts for 82% of all lung cancer diagnoses with a significant mortality rate[98,99]. This correlation exemplifies a pattern in the development of kinase inhibitors and oncology in general. The molecular characterisation of tumours guides subsequent drug discovery to combat the emergence of drug resistance and stimulates the development of small molecule inhibitors. Moreover, 70% of the FDA-approved kinase inhibitors address various resistance mutations[37].

1.1.3.3 Kinase inhibition beyond oncology

Despite kinase inhibition being dominantly applied in the area of cancer, the application is being rapidly and widely spread to other therapeutic areas of interest, such as immune-related and degenerative diseases (Figure 12).

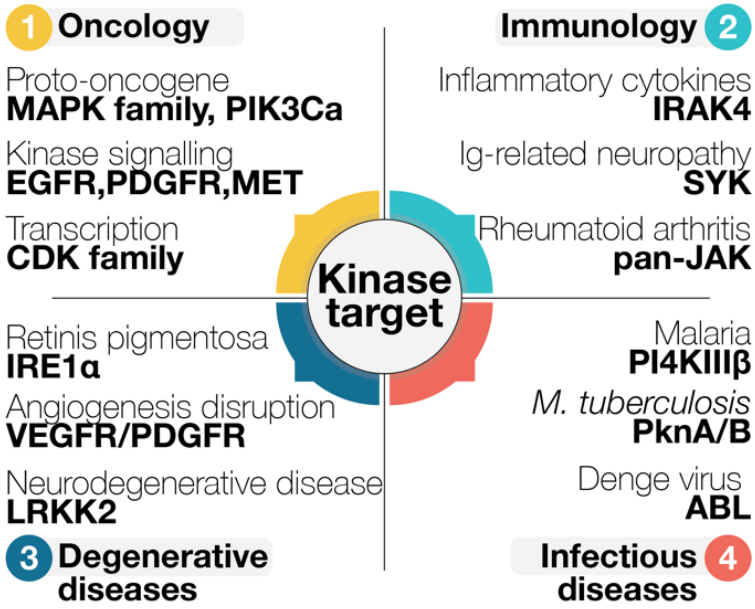


Figure 12. Kinase inhibitors in various therapeutic areas.

Deregulation of the immune system regularly results in the overproduction of inflammatory mediators, which commonly underlies the development of inflammatory and autoimmune disorders, such as arthritis, asthma, psoriasis, or sepsis [100]. One of the strategies against these diseases is targeting protein kinases that either mediate the intracellular activities of cytokines or are crucial for their secretion. For example, the FDA has authorised the pan-JAK inhibitor tofacitinib to treat rheumatoid and psoriatic arthritis, indicating that kinase inhibitors may apply in chronic inflammatory diseases[101]. Targeting protein kinases that restrict the release of inflammatory cytokines is becoming of interest in scientific society. One such target in this field is IRAK4 which is necessary for the generation of pro-inflammatory cytokines and chemokines generated by the activation of Toll-like receptors[102]; the other example is the SYK pathway targeting, which includes fostamatinib that is currently under clinical trials against immunoglobulin related neuropathy [103].

The field of degenerative diseases is the other area where kinase inhibition can contribute to disease suppression. Several kinase inhibitors have entered clinical trials in this field despite the lack of common knowledge of the disease's progression mechanism[104]. Namely, a small molecule inhibition of inositol-requiring enzyme 1α (IRE1α) for treating retinitis pigmentosa[105]

and VEGFR/PDGFR inhibitor for disrupting the angiogenesis in wet age-related macular degeneration[106].

In addition, a few kinase inhibitors were applied in infectious diseases, namely for malaria therapy[107], along with viral diseases: ABL inhibition against dengue virus[108], PkA and PkB against *Mycobacterium tuberculosis*[109]. Moreover, ABL, MAPK, and CDK kinase families were recently studied as potential therapeutic targets against Coronavirus infections [110].

1.2. In silico Drug Discovery

1.2.1 Introduction and perspective

Since its debut in the early 1980s, in silico or computer-aided drug design (CADD), has gained recognition as a methodology that influences almost every step of drug discovery. CADD has become integral to the industry and academic research since the constant development of its techniques brings a layer of structural rationality to the hit identification and development [111]. The relevance of molecular modelling and its fundamentals are found in the understanding of protein-ligand interactions, which is also the foundation of medicinal chemistry. CADD offers a wide range of methods and techniques to gain insights into protein-ligand binding and provide valuable data to support medicinal chemistry efforts. For instance, one of the very fundamental principles of molecular modelling is to achieve a good geometric fit. This concept, introduced by Emil Fischer as the "lock-and-key"[112], remains to be relevant a century after the publishing date. The lock-and-key concept as CADD is not only about a good geometrical fit but also about substantial complementarity within the hydrophobic and polar interactions of the protein binding site and the ligand. Apart from its interactions, molecular modelling often considers ligand and binding site flexibility, binding and distortion energies, solvation effects, entropy, and complementarity of the molecular force field [113]. CADD techniques allow gaining insights into the atomistic details beyond what can be obtained with the current experimental methods. Additionally, molecular modelling works with large pools of chemical data, allowing great chemical diversity to be reasonably analysed and generating new hypotheses capitalising on large datasets. The research & development process is steadily modernised through the rapid adoption of technologies and techniques that have the potential to drastically improve the drug development pipeline. Despite the speed at which these approaches are emerging, a comprehensive understanding of their applicability and limitation is to be achieved [114]. However, this brief description tells only a part of the story, providing a glance into molecular modelling. The overview of sample CADD workflows is presented further in this thesis, emphasising molecular dynamics techniques.

1.2.2 Sample drug discovery workflows

1.2.2.1 Structure-based drug discovery (SBDD)

SBDD is a drug design method which utilizes 3D structures of target proteins and focuses on the design and optimisation of a ligand that fits accurately inside the binding pocket and results in beneficial protein-ligand interactions[115]. Structure-based drug discovery is a rapidly growing area due to steadily increasing structural information available, arising not only from genomics and proteomics data but also powered with AI systems and novel machine learning approaches [116]. The SBDD process is iterative and frequently requires multiple cycles before an optimal initial hit compound can proceed to experimental validation. Figure 13 describes the essential steps that are involved in the structure-based pipeline. Following the initial step of target selection, the appropriate computational model has to be developed. One of the key factors at this stage is to avoid the creation of misleading and biologically irrelevant models that will result in the pipeline being ruined in the early phases. Important to notice that one should consider the possibility of the target protein exploiting dimerisation and/or multimerisation, as well as interaction with other proteins and macromolecules such as nucleic acids and/or membranes. The formation of functional assemblies frequently leads to alterations in the geometry of protein substructures, which can cause the shift of size or interactions near the binding site. Understanding the functional changes that a target protein undergoes upon modulation/activation is crucial for creating an initial representative model that aligns with real-world conditions. In the next step, the target protein 3D structure is generated or obtained upon availability from the Protein Data Bank (PDB). Whether the crystal structure or NMR of the target is not resolved to date, the 3D structure can be predicted by leveraging other structures with the highest sequence similarity to the target. The tools of choice at this step are homology modelling and AI-powered approaches, such as AlphaFold, chosen according to the uniqueness of the target and the system's complexity. In the next step, the binding site of the target needs to be identified to facilitate further optimization of hit compounds down the line. In some cases, the approximate location of the binding site is well-established within a protein family. It provides the structural features or specific residues that can be a starting point for the target binding pocket generation. For instance, in protein kinases, various subcultural elements can be used as landmarks within the binding site, such as the gatekeeper, DFG-motif or G-rich loop. If applicable, the binding site can be obtained with the localisation of natural substrate, for example, ATP binding site, in the event of kinases. Whether the binding site is unknown for the target protein and other members within the protein family of interest, computational

methods can be applied to suggest the probable locations. One can choose between the methods relying on geometrical properties, such as POCKET[117], PASS[118], LIGSITE[119], or combined with the physics approach, such as PocketFinder[120] or SiteMap [121,122]. The software of choice for project utilisation was SiteMap as a part of Schrödinger Maestro.

There are several options to pursue after the binding site is identified. For instance, one can establish a pharmacophore model, which offers an additional opportunity to validate the model before proceeding with pharmacophore-based virtual screening. Another option is to carry out the docking of a substantial chemical library (more on this topic in section 1.2.5. Molecular docking). Once the method is chosen, and the screening database is selected and prepared according to the requirements of the method, the virtual screening can be utilised. Next, when the initial hits are obtained, the compound set should be scored and filtered according to the properties required for project purposes. The essential pharmacokinetic properties for the hits list validation include Physico-chemical parameters (MW, number of heavy atoms, hydrogen bond donors and acceptors, and rotatable bonds), lipophilicity (Log Po/w), water solubility (log S), pharmacokinetics, drug-likeness (Lipinski, Violation), and synthesis accessibility for medicinal chemistry [123]. The cycle between model validation and the top hits choice is frequently repetitive and requires refinement. Once the initial hits satisfy the criteria, the optional short-scale MD simulations (200ns - 500ns) can be performed to validate the ligand stability within the target binding pocket, ensuing with selecting top hit compounds for experimental validation.

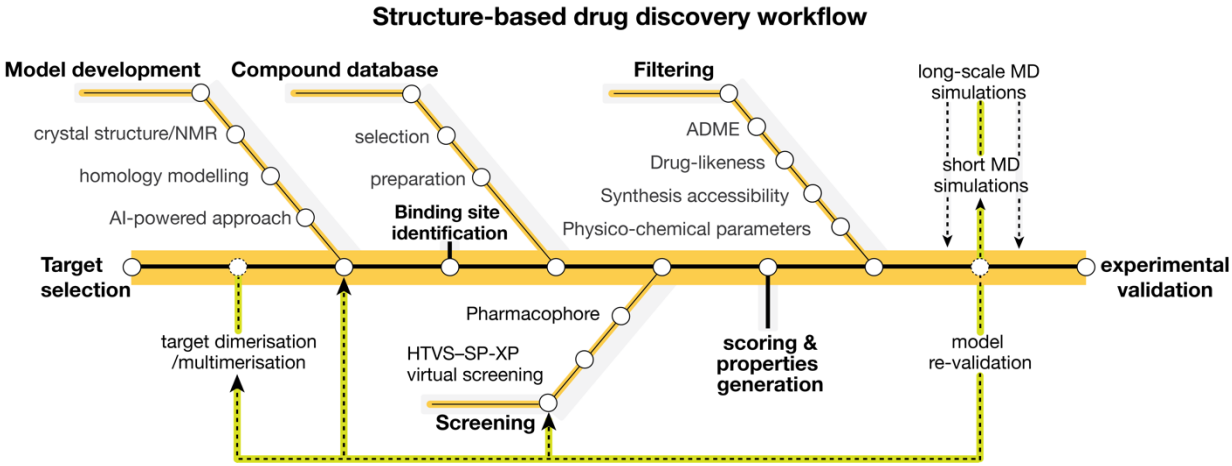


Figure 13: An example of a structure-based drug discovery workflow. The thick yellow line illustrates the path from the target selection to experimental validation, while the brunches highlight the variety of methods that might be used along the journey. The dashed green arrows

indicate that LBDD is an iterative process, emphasizing the necessity of extra computations and revalidation along the way.

1.2.2.2 Ligand-based drug design (LBDD)

The premise of ligand-based drug design originates from the foundation concept of medicinal chemistry, which asserts that molecules with a high degree of structural similarity possess comparable bioactivity[124]. LBDD approaches utilise primary data on active compounds (approved drugs, published reports, etc.) to predict or generate novel drug-like compounds with similar biological effects. The inactive compounds are also beneficial for the predictions, while they can be used to identify undesirable ligand features and validate the accuracy of the computational model. In this sense, the typical workflow starts with an initial set of compounds with known potency that proceeds to a chosen computational method for the similarity search. The similarity search may be conducted using a wide range of molecular descriptors or filters. Molecular descriptors are categorised based on the search's dimensionality. Molecular weight and logP are typical 1D descriptors, whereas topological indications and fingerprints are 2D descriptors. The characteristics covered by 3D descriptors encompass a broad range of properties from electrostatic potential to the 3D geometry of ligand moieties. Furthermore, the 3D descriptors are integrated into pharmacophore modelling. According to IUPAC recommendations, a pharmacophore is "an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response" [125]. The ideal scenario for pharmacophore modelling is when a protein-ligand complex is co-crystallised, and the ligand shows sufficient potency and bioactivity. An additional step using Quantitative Structure-Activity Relationships (QSAR) or machine learning can be added to the pipeline if a comprehensive dataset of active compounds is available. The most popular metric to quantify the similarity of new compounds to the initial set is the Tanimoto coefficient [126,127].

1.2.3 Protein structure preparation

The emerging problem of refining protein crystal structures with initial low resolution is a relevant question in computational chemistry due to the rapidly increasing number of structures in the Protein Data Bank (PDB). The need for protein preparation arises from several factors, such as the absence of hydrogens in most X-ray crystal structures, tautomer uncertainties and the influence of crystal packing. The quality of the upcoming model and the stability of protein-ligand complexes are both significantly influenced by hydrogen bonds, in addition to having

an impact on the binding specificity. The hydrogen bond composition of a molecule affects how small molecules are transported, distributed, metabolized, and absorbed [128]. Nevertheless, hydrogens are absent in most structures solved by X-ray crystallography due to the method limitations [129]. However, some hydrogen atoms may be assigned with high-resolution crystallography[130]. The absent hydrogens are less common in PDB files that are solved via NMR crystallography since hydrogens are typically modelled in before the molecule is folded to meet the requirements of the NMR interatomic distance[128]. Crystal packing may affect the protein side-chain conformations and lead to a misleading interpretation of the protein binding pocket [131]. For instance, the angles of the backbone, which characterise the overall tertiary structure, are one of the most crucial variables governing conformational specification and determining the interaction network in a protein [132]. Therefore, selecting a correct rotamer for residues such as asparagine and glutamine can significantly affect the overall structural stability and define the interaction network [133]. Abnormal residue conformation impacts not only the energy landscape of a protein but also the shape of the binding pocket [133]. Moreover, potential side-chain rearrangements around a single point mutation can be associated with mutation itself or due to variations in crystal packing forces[134–136]. Whether the errors in the original data derived from experiments are left untreated can lead to misleading results and the wrong creation of the initial model. The protein preparation may be carried out using a variety of applications; moreover, it is frequently incorporated into molecular docking software as an essential and initial step in creating reliable protein-ligand interaction models. Examples of such software include GROMACS[137], AutoDock[138], HADDOCK[139,140], and Protein Preparation Wizard[141] (Schrödinger, LLC). Although the exact protocols for protein preparation differ based on the software, several aspects are universal. The first aspect is the correct allocation of missing hydrogens, which addresses the common lack of those in the crystal structures. The second aspect is dedicated to correcting hydrogen bonds following their energy profiles. The hydrogen bond optimization process generally starts with a sampling of the angle rotation in the terminal rotatable bonds of the residues and is followed by an examination of the whole hydrogen bond network. Often protein preparation packages base their hydrogen assignment protocols on cluster combination and Monte Carlo sampling[141]. Further, the protein undergoes the minimisation protocol with a chosen force field.

1.2.4. Homology modelling

Homology modelling is a technique that allows predicting the secondary protein structure based on the sequence similarity to the template protein. The baseline concept stands that

proteins with a high degree of sequence similarity would have comparable folding[142,143]; when homology modelling applied together with molecular dynamics can serve for in silico design of mutagenesis experiments and the study of protein conformational changes as well as allosteric effects.

One can generate the homology models on the template of experimentally established protein structure, complemented with a homologous sequence of sufficient similarity or alternatively with another protein template structure that shares family-specific structural features with the future model. For instance, class A GPCRs partake seven transmembrane helices that traverse the membrane. Therefore, homology modelling of GPCR with unknown crystal structure can be supported by the seven transmembrane helices, which are used as a template, or skeleton for modelling[144]. This way, one can compensate for low sequence similarity between the target protein and existing template crystal structures by finding a well-established feature that is common within the protein family of interest. The other alternative is composite or chimera homology modelling, which allows using multiple structures as a template after the sequence alignment to the target one. The process of homology modelling can be roughly divided into the following steps: selection of the template 3D structure, or structures (in Chimera approach); sequence alignment of the target protein with the template; construction of the model, based on the 3D structures and alignment; Refining and evaluation of the model.

The final evaluation step is as crucial for the quality of the obtained model as the initial selection of the templates. For this reason, various techniques and software are available for this purpose. One example of software for evaluating homology models is PROCHECK, which compares models' stereochemical parameters against high-resolution and previously validated crystal structures to assess the overall quality of the protein model and individual residue geometries [145]. Web-based server examples for structure validations include such platforms as PDBsum [146] and MolProbity [147].

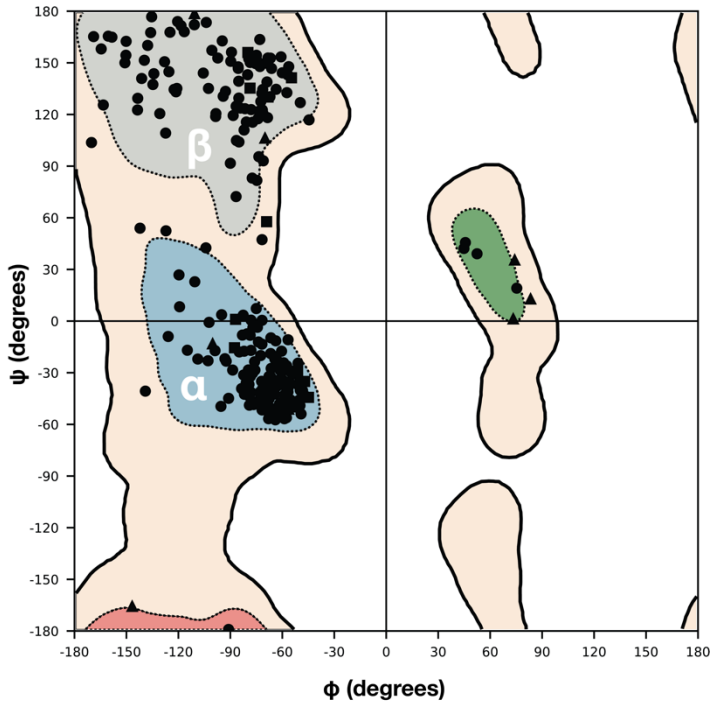


Figure 14. Example of Ramachandran plot of the autoinhibited MKK4 model, adapted from Shevchenko et al.[148]. The model is based on PDB ID: 3alo[149]. The plot area describes the dihedral angles of all residues in MKK4. Glycine residues are plotted as triangles, proline as squares, and other residues as circles. The areas inside the dashed line are considered favoured, the beige area as favoured, and the white as disallowed.

Another method for evaluating the accuracy of an obtained model is the Ramachandran plot [150]. Software packages like those mentioned above typically include the Ramachandran plot, although it is worthwhile to describe separately. Ramachandran plot allows the description of residue geometry by plotting φ and ψ torsional angles of amino acids. The fundamental principle of this approach is based on the finding that computationally calculated angles that prevent steric collisions can be divided into several groups that correspond to the secondary structure elements [150,151]. According to Ramachandran Principe, α-helices, β-strands, and loops are the conformations that polypeptide chains are likely to form because the steric clashes between atoms prevent the formation of other conformations[150,151]. Thus, the Ramachandran plot area is divided into α-helical (Figure 14, blue area), β-strand (Figure 14, gray area), and loop area (Figure 14, green area). Each point on the plot represents the φ–ψ angle combination for one amino acid. The core regions (Figure 14, dashed line areas) represent the most favoured angle combinations, the allowed regions (Figure 14, beige area) are located around the core, while the remaining space is disallowed area (Figure 14, white area) is considered to indicate

improper φ–ψ combinations. This plot can be utilised for on-the-fly modification of the protein geometry by finding residues that are plotted in the forbidden area of protein dihedrals using interactive Ramachandran plots (as implemented, for instance, in Schrödinger Maestro Suite) with further minimisation.

It is important to notice that homology modelling is frequently a recurring procedure: one should ensure that the obtained model is representative and not biased due to the template choice. In other words, different templates yield different models and may arise various errors. Therefore, good practice calls for constructing several homology models based on different template structures, followed by structural comparison and validation. Various tools and online recourses are available for homology modelling, for example, SwissModel[152], MODELER[153], Modweb[154], Phyre2[155], and Prime homology modelling module[131,156]. One of the newest methods is AlphaFold[116], an artificial intelligence that predicts the three-dimensional structure of a protein from its amino acid sequence.

1.2.5 Molecular Docking

The docking methodology, especially in small molecule drug design, remains the pioneer computational tool for ligand orientation prediction within a protein binding site since its invention in the early 1980s[157]. The molecular docking approach includes several stages, roughly divided into the following steps: ligand preparation and generation of ligand conformations; protein preparation with the selection of the docking area; initial pose generation and, at last, the pose scoring[158]. To date, abundant studies have been performed to distinguish the most proper protocol and scoring function, while scoring of the docking poses is a crucial aspect for the evaluation of the results. The scoring functions can be roughly divided into three subclasses: empirical, knowledge-based and force field-based[157].

Diverse approaches exist in the field of molecular docking: one requires beforehand separate preparation of the protein structure and the set of ligands in the relevant force field (Glide[159–161]). At the same time, the other software combines the stages under one protocol (HADDOCK[139], AutoDock Vina[138,162,163]). Nonetheless, ligand preparation protocols generally include the generation of 3D geometry and tautomer and ionisation states prediction. After the ligand poses are generated with the selected algorithm, the validation of the results is to be conducted.

Additionally, if the known active binding compounds are available for the protein of interest, they can be included in a validation round dataset before the docking. In this respect, the known actives hypothetically show a suitable binding mode and are among the high-ranked

compounds according to the docking score. Moreover, the known negatives (or decoys) can be included in the ligand dataset prior to the docking to evaluate the precision of the model and minimise false positives and avoid artificial enrichment[164]. The enrichment factor, therefore, represents the model's ability to detect the true positives, avoiding the inclusion of decoy compounds[165].

Important to notice that the enrichment factor has a restricted use because of its substantial dependency on the number of known actives in the initial dataset. A ROC curve (receiver operating characteristics) is another way of estimating docking efficiency. When combined with plot visualisation, the area under the curve (AUC) determines the degree of enrichment [164]. Besides the need for actives compounds to be present for the target, the application of ROC curves has several restrictions[166]. There are, however, published methods to prevent obtaining biased data from curve[167]. When the docking is carried out against a novel target, with no primary data on active compounds available, the docking validation relies on the scoring function, the prediction of binding energy, and especially the visual evaluation of docking poses. Prior to the in vitro validation of the top compounds, the docking output may be additionally filtered based on other chemical characteristics (such as ADME) with optional protein-ligand stability validation by short MD simulations.

Each docking software has a unique protocol that specifies the sequence and methodology of common docking steps, from ligand preparation to initial pose scoring. The Glide[159–161] protocol is described below as an example of a pipeline within the docking software. In Glide [159–161], docking is further categorized according to its precision. The first category is known as high throughput virtual screening or HTVS and is typically used for screening substantial amounts of compounds, starting at a library size of a few hundred thousand and going up to a few million. The number of intermediate conformations and the accuracy of the final torsional refinement and sampling are reduced by HTVS since it is designed to produce quick screening output, focusing on speed rather than quality. Standard precision (or SP) docking is the next category. The scoring function used by SP and HTVS is identical. However, SP generates more intermediate conformation samplings. SP is frequently applied to a fraction of the top hits from the HTVS output to improve prediction accuracy. Next in the order is the XP or extra precision docking, which is more profound and discriminating but takes significantly more time to execute. Compared to SP, XP does more thorough sampling by starting with SP sampling before initiating its own anchor-and-grow approach. Additionally, XP utilises a more complex scoring algorithm and places higher demands on the complementarity of ligand-receptor shapes. These differences eliminate the false positives that SP-precision permits. To ensure that

the more expensive docking algorithm is carried out on valuable poses, it is advised to run the compound database using HTVS-SP docking first, then take the top 10% to 30% of the resultant hits and dock them using XP precision.

1.2.6 Virtual Screening.

Virtual screening campaigns are broadly used in industry and academic research environments; the number of successful examples of virtual screening applications for drug discovery is steadily rising[157,168–170]. The primary goal of virtual screening is to obtain the maximum amount of potentially active compounds from a chemical database with great chemical diversity that can be further scored, ranked, filtered, and experimentally evaluated. The main benefit of virtual screening is the possibility to access and filter a vast chemical space, which would be challenging and resource-intensive with experimental techniques. By definition, virtual screening is much more than docking an extensive library of compounds. The variety of methods that can be attributed to virtual screening allows for several predictions. For instance, shape screening can be used to explore the geometrical features of the binding pocket. Pharmacophore screening allows finding the optimal geometrical ligand-protein fit while focusing on the desirable protein-ligand interactions. Fragment and fingerprint screening aids in increasing the chemical diversity of compounds. Virtual screening is commonly carried out on a library of compounds that may be purchased and subjected to experimental testing. Utilising the In Silico approach can significantly cut expenses and resource waste during the initial screening phase.

Additionally, the suppliers of compound libraries typically offer classified sets of compounds according to, for instance, therapeutic area or bioactivity. Presumably, the virtual screening can be applied to an in-house database or made entirely virtual with further synthesis. This thesis further describes two broad categories of virtual screening: ligand-based (section 1.2.2.1) and structure-based (section 1.2.2.2).

1.2.7 Molecular Dynamics

Molecular dynamics (MD) simulations empower the studying of the structure within molecular systems at the atomic scale in a dynamic manner. The MD simulation generates a trajectory that allows observing an atom's individual movement as a function of time. The forces between interacting atoms are estimated using a force field, and the system's total energy is computed according to Newton's law of classical mechanics. The integration of Newton's laws of movements produces subsequent configurations of the developing system during MD

simulations, producing trajectories that describe the locations and velocities of the particles throughout time. The applications of MD simulations cover a wide range of possibilities within in silico drug design. Among others are the docking postprocessing of protein-ligand complexes[171,172], drug binding studies[173,174], protein-protein interactions[175], and the discovery of structural states and binding sites[176].

One of the essential aspects that can be carefully considered when the MD simulation is conducted is the timescale, as different types of protein motions occur at distinct timescales. For instance, the side-chain rotamer movements can be observed in the range of ps to μs, followed by the loop motions in the field of ns to μs, with more significant domain movements starting to be evident in μs+ timescale[177]. Thus, one should consider the reasonable timescale of the MD simulation in accordance with the movements to be observed to deliver valuable results. The other crucial point for comprehensive MD simulation is the number of replicas of the single system to be conducted to avoid false positive conclusions[178]. The term 'replica' refers to the simulations of the identical system, sharing the same number of atoms, initial structure and preparation protocol, repeated several times. The difference between replicas is in the velocities generated randomly according to the Maxwell distribution[178]. As a result, the velocities are unique for each replica, which leads to the production of different simulation trajectories. However, even replicas with identical velocities can produce distinct trajectories for various reasons, including machine-specific settings and the specification of the compiling[178]. In theory, the ergodic principle claim that the velocities have no impact on infinite dynamic simulations[179]. While molecular dynamics are not run eternally in real life, random velocities are strictly important, especially for short-scale MDs. The random initial velocity assignment ensures that the results of every simulation are slightly different, even if the other settings are identical. In other words, the random velocities provide the opportunity to observe real-world phenomena happening with the same system at different time points. Moreover, multiple replicas ensure a specific movement or interaction observed not in a single simulation but in a statistically significant amount of replicas is not biased or artificially generated but related to the real-world evidence. The choice between the multiple simulation replicas over a single but long-scale one is frequently a question that should be answered prior to the initiation of the MD according to the aims and resources of the project.

1.2.7.1 System preparation

A basic pipeline for setting up a system for MD simulation requires the protein of interest to be prepared in accordance with the simulation software and chosen force field. While specific

variations and step order may vary, the system preparation step is essential for building an MD system. System preparation follows the structure preparation (1.2.3) with additional examination of crystallisation artefacts, correct bond orders and the presence of ions and unassociated molecules. Next, the solvation model is to be chosen. Many types of water models have been developed regarding the system solvation with water. These models are categorised based on several important factors, including the number of interaction points the molecule can make (site), the inclusion of polarisation effects, and flexibility or rigidity. The water model chosen for the articles in this thesis was TIP3P, a three-site rigid water molecule with assigned Lennard-Jones parameters and potential[100].

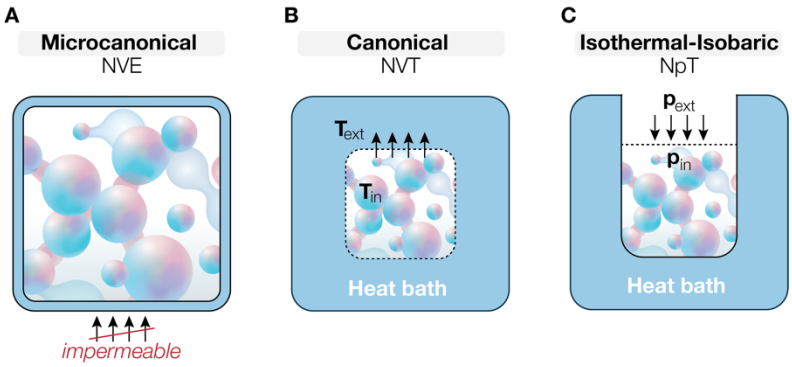


Figure 15. Thermodynamic ensembles in Molecular Dynamics. (A) A Microcanonical (NVE) ensemble is an isolated system with a fixed amount of particles, total volume, and energy. **(B)** Canonical (NVT) ensemble as a system with fixed total volume but with temperature-dependent boundary energy transfer. **(C)** Isothermal-Isobaric (NpT) ensemble as a system allowing volume alteration for the pressure equilibration. Ext – external, environmental; in – internal, model system value.

A thermodynamics ensemble used to characterise the system must be chosen during the system preparation step. The thermodynamics ensemble is an idealisation of the model system composed of multiple replicas of this system, all taken into account simultaneously, and each reflects a potential state in which the actual system may be [180]. In other words, the thermodynamics ensemble is a statistical approximation that allows to extrapolate the fundamental features of the thermodynamic system through classical and quantum mechanics [181,182]. The ensemble classification is based on how the model system is separated from the outer environment. The first category is the microcanonical ensemble, which depicts the model system as entirely isolated from the environment (**Figure 15A**). The microcanonical ensemble refers to the acronym NVE, which means that the total number of particles N, volume V, and

energy E remains constant[183]. As a factor that assumes the thermodynamic interaction of the model system with its surroundings, the temperature cannot be defined for the NVE ensemble. NVE is used for describing states of a system with defined total energy. The next ensemble category is termed canonical and provides the option to determine the system temperature, therefore referring to the acronym NVT, where N stands for the total number of particles, V for volume and T for temperature. NVT allows the energy transfer between the boundaries of the model system and the environment while prohibiting substance exchange **(Figure 15B)**[183]. In other words, this ensemble addresses the model system's potential states that are in thermal equilibrium with the environment, which can be visualised as a heat bath with constant temperature, which is several magnitudes larger than the model system[184]. The difference in size ensures that no amount of heat produced by the model system will cause the heat bath's temperature to rise considerably. As a result of the thermal contact between the model system and the environment, the system will now transmit heat to and across the environment until they reach thermal equilibrium[185]. The last ensemble category is the isothermal-isobaric, which as with the canonical ensemble, allows the energy transfer between the model system and environment, but allows the volume change. The system's volume fluctuates to equilibrate the system's internal pressure with the pressure applied to the system by its environment **(Figure 15C)**[185]. The acronym for the isothermal-isobaric ensemble is NpT, where N stands for the persistent total number of particles, p for pressure and T for temperature.

Moreover, NpT is the most used ensemble. However, prior to the start of MD, the so-called model relaxation step is conducted. The model relaxation is an essential step which ensures the proper system equilibration and removal of the high strain degree from the newly constructed system. The model relaxation involves several phases, where a particular ensemble is combined with specific pressure, temperature, and timestep until the equilibration is achieved. The placement and selection of ions and the specification of temperature and pressure are all components of the system preparation process. It is frequently desirable to have an electrically neutral system for MD simulation. However, it is not strictly necessary, for example, with the Desmond MD engine [186], which applies a uniform background charge distribution to neutralize the system in the Ewald summation [187]. Moreover, the system can be set up in a salt solution rather than a pure solvent.

The periodic boundary conditions (PBC) are then applied to the system by selecting the simulation box's form and cut-off. It is necessary to note that to account for potential protein movements, the distance from the simulation box's edge to the protein must be considerable. If the simulation box size is insufficient, MD-derived findings can frequently be artefactual or

misleading. The force field should be chosen to continue with MD simulation once a system has been solvated and configured for boundary conditions.

1.2.7.2 Force fields

While in classical physics, force fields evaluate a system's potential energy, the fundamental distinction between the two in molecular modelling is that the energy landscape is depicted as an energy gradient dispersed across the particle positions [188]. Reconstructing a realistic simulation of a molecular system on an atomistic level is strongly dependent on force field parameters, which are the core for deriving meaningful structural information and relative energies from MD simulation.

To date, additive or non-polarizable force fields are the most prevalent in small molecule drug design, frequently sharing the basic potential energy function and the parameters comprising the function itself [189]. The name 'additive' originates from Coulomb's equation for electrostatic interactions, which states that a system's potential energy is the sum of all atom-atom individual interactions (Equation 1) [189]. The potential functions, which are essentially an array of equations representing the potential energy, and its components, form the basis of the force field core. The parameters in this set of equations are another component distinguishing a force field.

In the vast majority of condensed-phase simulations, the total potential energy of a cohort of molecules is determined as a sum of inter- and intra- molecular interaction energies between all components of the system (Equation 1)[190,191].

$$E(\vec{X}) = \sum_{a<b} Eab + \sum_{a} E_a^{int}$$

Equation 1. The total potential energy of a set of constituents ab with coordinates$\vec{X}$. ab refers to a set of molecules and/or ions.

The following terms commonly describe the intramolecular potential energy: harmonic bond stretching (Equation 2), angle bending (Equation 3), torsional angle definition as Fourier series (Equation 4), Coulomb electrostatics and Lennard-Jones potential (Equation 5)[191].

$$E_{bond} = \sum_{i} k_{b,i} \left(r_i - r_{0,i}\right)^2$$

Equation 2. The first force field term defines bond stretching energies in harmonic (ideal) conditions, where k_b is the bond-stretching constant, which regulates the rigidity of the bond spring.

$$E_{bend} = \sum_i k_{\vartheta,i} \left(\vartheta_i - \vartheta_{0,i}\right)^2$$

Equation 3. The second force field term defines the angle bending energy, where k_ϑ is the angle bending constant, determining the rigidity of the angle spring.

$$E_{torsion} = \sum_i \{\frac{V_{1,i}(\mathbf{1} + \boldsymbol{cos\varphi_i})}{\mathbf{2}} + \frac{V_{2,i}(\mathbf{1} - \boldsymbol{cos2\varphi_i})}{\mathbf{2}} + \frac{V_{3,i}(\mathbf{1} + \boldsymbol{cos3\varphi_i})}{\mathbf{2}}\}$$

Equation 4. The third force field term represents the torsional energies as a Fourier series.

$$E_{nb} = \sum_{i<j} \{\frac{q_i q_j e^2}{\boldsymbol{r_{ij}}} + \mathbf{4}\boldsymbol{\varepsilon_{ij}}[\frac{\boldsymbol{\sigma_{ij}}^{\mathbf{12}}}{\boldsymbol{r_{ij}}} - \frac{\boldsymbol{\sigma_{ij}}^{\mathbf{6}}}{\boldsymbol{r_{ij}}}]\}$$

Equation 5. The fourth force field term represents the non-bonded energy between all atom pairs, where $\boldsymbol{\sigma_{ij}}$ is the equilibrium distance and $\boldsymbol{r_{ij}}$ is the distance between interacting atoms.

Force fields can be roughly divided into three major classes. One of the defining characteristics of class one is the utilisation of harmonic movements to depict bond stretching and angle bending. According to the assumption, the amount of restoring force is proportional to the displacement from the equilibrium position for class one force fields [192]. The approximation in class one force fields is referred to as quadratic because the square of the displacement energy is linearly associated with the harmonic energy of oscillator [193]. Moreover, the parametrisation of bond stretching and the angle bending often approach harmonic behaviour only close to the equilibrium. The most famous examples of the class one force fields are the Optimized Potentials for Liquid Simulations (OPLS) [194], AMBER [195,196], CHARMM [197], and GROMOS [198]. One of the first techniques to be established with constant parameter optimisation for the propagation of thermodynamic characteristics in the liquid state applied to small molecules was OPLS, which stands for optimised potentials for liquid simulations [194]. On the OPLS core, the next generation of force fields was built: OPLS3[199], OPLS3e[200] and OPLS4[201] with the maintenance of the nonbonded parameters. In general, AMBER refers to a collection of force fields that can be split according to the simulated biomolecular system. For instance, for protein simulations, AMBER suggests ff19SB[202], for lipids or complex membrane simulations –

LIPID21[203], and for nucleic acids – DNA OL15[204], and RNA OL3[205]. As AMBER CHARMM also represents a set of force fields, for instance, the all-atom CHARMM22[206] and extended atom force field CHARMM19.

Class two provides additional anharmonic cubic and quartic components for the potential energy of bonds, resulting in a more detailed geometrical modelling of the vibrations of bonds. In addition, these force fields include cross-terms that describe the interactions between nearby located angles and dihedrals. Class two force fields include MMFF94 [207], commonly referred to as Merck Molecular Force Field), which parameters are primarily derived from quantum calculations rather than experimental data [208]. Another example is UFF [209] (Universal Force Field), which application is somewhat restricted since the parameters, particularly for metals and inorganic substances, have not been sufficiently validated [210].

Class three includes force fields that contain extended parameters applicable to organic chemistry, such as the Jahn-Teller effect or stereoelectronic effects. For example, AMOEBA [211] is a polarisable force field that employs atomic-induced dipole to model polarisation while assuming that averaging polarisation is insufficient [212]. Another example is DRUDE [213], which uses non-polarizable force fields to leverage atom-to-atom Coulomb electrostatic interactions as its core while integrating polarisation effects via NAMD and a dual-Langevin thermostat approach [214]. Important to highlight another class of force fields, known as coarse-grained force fields, which employ a distinct strategy in molecular dynamics simulations.

The idea behind the coarse-grained approach is the reduction of the number of degrees of freedom within a system. This is achieved by parametrising the most significant interactions with the force field while representing a particular set of atoms as a single bead. The definition of the most significant interactions might be intricate depending on the parametrisation method, hence tabulated potentials are frequently employed. The purpose of coarse-grained models is to replicate specific characteristics of a given system, which can encompass an atomistic protein model or experimental data. The properties one intends to replicate in the model determine the classification of the coarse-grained force fields. For instance, free energy conservation is the focus of the MARTINI [215] force field and the simplex method [216]. Another example is inverse Monte Carlo with structure-based coarse-grained modelling, emphasising the radial distribution [217].

1.2.7.3 Molecular Dynamics analysis

The vast spectrum of MD analysis methods can be divided into two primary perspectives, which encompass the ligand and the protein. Further, each subcategory is described with examples of frequently applied techniques.

1.2.7.3.1 Protein perspective

Protein Root mean square deviation (RMSD)

RMSD is a measurement of the average distance between given groups of atoms as Cα atoms or protein backbone commonly used in analyses of MD [218]. RMSD provides insights into the deviations in the protein conformation along the MD trajectory, which correlates with the indication of protein stability. Namely, the values are calculated for atomic coordinates at each frame of the MD trajectory, which allows seeing the value variation along the timescale. One crucial factor in RMSD is the choice of the reference point to which the subsequent RMSD values are compared. This is because the quality and reliability of the reference point determines how accurate the RMSD calculations will be. For instance, the simulation's starting frame is commonly designated as the reference point in MD. Further, if the starting frame contains any poor contacts, clashes or ambiguities in the ligand pose, one will obtain high RMSD values. High values in this regard point back to the poor choice of the reference point rather than necessarily indicating ligand instability or poor model quality. Moreover, RMSD indicates the approximate timeline of system equilibration, which can be observed from higher-than-average RMSD values at the beginning of MD (generally restricted to 200 ns with Desmond MD engine). RMSD can be calculated in the majority of drug discovery software, such as GROMACS [74], Maestro Simulation interaction analysis tool, or for complex selections can be scripted with MDAnalysis.analysis.rms [219] module for Python.

Protein Root mean square fluctuation (RMSF)

RMSF is a measurement of deviation from a selected group of atoms reference position that occurs along the MD trajectory and is averaged over the total amount of atoms [220]. Thus, the RMSF values indicate which structural elements of a protein fluctuate most compared to their reference positions. Conceptually fluctuations of RMSF values in a particular part of the protein align with B-factor values from the experimental x-ray structures, allowing direct comparison between the computational and empirical data. Generally, RMSF and RMSD calculations are bound together as an essential basis for trajectory analysis; therefore, they can be calculated within the same software.

Protein secondary structure analysis

The initial classification of the secondary structure was based on distinct hydrogen bonding modes. It encompassed three elements: helix, strand and coil [221], further this classification was extended to eight states: α-helix, 3_{10} helix, π-helix, β-strand, β-bridge, β-turn, bend, and loop or others [222]. As the physical process of protein folding is tightened up with the various biological events and processes [223], deriving secondary structure information from MD trajectories can provide insights into the structural behaviour of the protein. Although to date, the possibility of computational prediction is limited to helix, strand, or loop (coil), the results can highlight valuable changes in the investigated structure correlated with the impact of the ligand, mutation or normal protein function. The computational results of the secondary structure prediction can be further supported with circular dichroism (CD), which provides an experimental evaluation of proteins' secondary structure. Up to this point, secondary structure calculations along MD trajectories are somewhat rare and frequently require custom coding for a system. However, the starting data for these calculations can be extracted as raw output from the Maestro Simulation interaction analysis tool (Schrödinger, LLC, New York, NY).

Principal component analysis (PCA)

Principal component analysis (PCA) is a prominent approach for evaluating large datasets with several dimensions or variables per observation. PCA enhances the readability of data while maintaining the most information, besides facilitating visualisation. In other words, PCA retains the greatest amount of statistically significant data within a given dataset that results in dimensionality while reducing the "noisy" data that causes valuable statistical trends to be overseen. This is conducted by the linear translation of the dataset to a new coordinate system, which produces a two or three-dimensional plot with data points that allows deriving the clusters of nearby data points visually. The number of principle components chosen for the visualisation is the main difference between 2D and 3D PCA. The construction of principle components in PCA aims to capture the most variety in the dataset: PC1 depicts the greatest variety, PC2 describes the next-greatest variety, and so forth. Most of the variation may therefore be captured by the first two to three PCAs, and the remaining ones can be eliminated without losing valuable for analysis data. The linear translation (linear reduction) is a process of finding new uncorrelated variables that represent linear functions of the initial set and sequentially optimising the variance[224]. In addition, PCAs are widely used in a variety of

disciplines where large datasets must be statistically analysed. Among these disciplines are genomics, astrophysics, machine learning and data science.

Moreover, PCA is a widely used approach for analysing MD trajectories, where it applies the same principle: a linear transformation that diagonalises the covariance matrix and eliminates the spontaneous linear correlations between the coordinates[225,226]. The PCA results can be utilised for the free energy landscape construction, which describes the metastable conformational states and the transition states[227]. These states contribute to understanding functionally relevant motions within the investigated system, which can be hard to derive by visual examination due to numerous other motions simultaneously occurring within the MD trajectory. The PCA analysis can be conducted with GROMACS[74] covariance tool (gmx covar and gmx anaeig scripts).

Markov State Modelling

Markov modelling (MSM) plays a significant role in the modelling and interpretation of MD simulation data, while it allows pinpointing of statistically significant events in the MD trajectories[228]. MSM is a robust framework for analysing long-scale MDs, where it becomes more challenging to visually observe changes in protein dynamics and substructural geometrical variations.

MSM stands for a master equation framework, indicating that the system's complete dynamics can be highlighted using just the MSM. The master equation formalism has also been applied in a wide range of scientific disciplines[229]. By definition, MSM is a $n \times n$ square matrix (often referred to as transition probability matrix), in which the n states represent the total number of configurations that the system may exist in [230]. The dynamical change of the system can be observed by defining the states at time points separated by the lag of time (τ). This separation of the states ensures the lag of time to be a Markov process[230]. The term “Markov process” refers to a random process in which, given the present, the future is independent from the past[228]. In terms of MD, the Markov process means that the probability of the system transition from state x to state y does not depend on where the system was before state x.

The obstacles in building an MSM may be divided into two main categories: the definition of the states in a kinetically relevant order and the effective employment of the state decomposition for building an efficient transition matrix[231]. Additionally, a bias-variance problem related to the choice of the number of states is a dilemma with MSMs. When utilizing a limited number of states, one knows analytically that the predicted value of the relaxation timescales is less than the real value and that this bias gets less as the number of states increases.

On the other hand, with a given data set, as the number of states rises, the statistical error in the MSM grows[231]. Although some sufficient algorithms have been developed, to date, there are no algorithms in the literature that automatically and effectively balance these conflicting sources of errors[232].

Distance calculations

Distance analysis allows measuring the distances between pairs within a selection as a function of time. For example, the selection may encompass two atoms (Cα), two residues or two structural elements of a protein (helix/chain/lobe). These[230] calculations allow promptly visualising the change relevant to the project distances; derived knowledge can be used to understand substructure stability or behaviour related to ligand binding. In addition, with GROMACS[74] gmx distance script, the reference positions for the distance calculations can be defined by the centre of mass or the centre of geometry for both atom, residue or custom selection. This option allows extracting relevant results from dynamically changing reference selection, which is highly important for the MD trajectory analysis.

Angle and plane calculations

Angle or plane calculations are beneficial to support visually perceived motions within the MD trajectory with the numerical values. There are various examples of possible practical applications of this method. For instance, angle calculations can be utilised to define the protruding outward movement of αC-helix in kinases quantitatively, justify the opening or closure of Switch-II in RAS proteins or assist the hit-to-lead identification by defining the degree of pocket opening in MD simulations.

The reference positions of the vertex and apex points can be defined using several methods, as seen with the distance calculations—namely, the centre of mass or geometry for an atom, residue, or substructure. The validation of the residues chosen for the computation of the angle value is an optional step that allows to verify if the calculated value is biased or whether a particular selection has no impact on the average angle value along the trajectories. This is accomplished by selecting the neighbouring to the reference selection residues and repeating the calculation. The average values from the reference and validation calculations should appear to be in the same range. Another way to produce reliable angle calculation between moving helices or domains is to choose the stable residue interval for the apex. The MD-derived RMSD values can be used to support the interval stability. Python is the primary tool for carrying out these computations for an MD trajectory since they are highly dependent on the

investigated system. Additionally, the classes required to access data in the MD trajectories are provided by the MDAnalysis package for Python [233,234].

1.2.7.3.2 Ligand perspective

Ligand Root mean square deviation (RMSD)

The theoretical background of RMSD calculations is the same for both ligand and protein, although the RMSD results are interpreted somewhat differently when a ligand is involved. For instance, if RMSD values of the ligand can be aligned on the ligand itself, the graph will reproduce the ligand's internal fluctuations as a function of time. When a shift of values is observed, the ligand's conformation changes during the trajectory. Alternately, the ligand can be aligned on the protein, in which case the plot will demonstrate how the ligand fluctuates concerning the protein. In this case, a significant shift in values may indicate that the ligand is unstable inside the binding site or moving outside of it. Additionally, RMSD calculations can be used to compare the difference between a docking output to a known binding pose from a crystal structure to validate the docking precision.

Ligand Root mean square fluctuation (RMSF)

Ligand RMSF shares the theoretical background with the protein RMSF and indicates the atom fluctuations in the ligand. Similar to ligand RMSD, ligand RMSF can be aligned either on the ligand itself or on the protein. The data in the first case indicates ligand fluctuations, whereas, in the second, it depicts fluctuations in correspondence to the binding site. The flexibility or conformational alterations of a ligand inside a binding site are often evaluated using the ligand RMSF.

Interaction pattern determination

The analysis of the interaction network that the ligand is forming with the protein is the fundamental part of the simulation analysis as the non-covalent interactions, namely hydrogen bonds and hydrophobic interactions, are the core of the ligand-protein binding [235]. Unlike docking, MD-derived interaction analysis gives not a single-point calculation but the statistical range of the simulation time when the ligand is in contact with a particular protein residue. Moreover, the interaction can be plotted as a function of time. The derived knowledge can be further applied for SAR analysis and hit-to-lead optimisation. Finally, distinct software may have different definitions of distance and angle thresholds that characterise specific interactions, particularly in the case of hydrophobicity.

MM-GBSA Energy calculations

Molecular mechanics energies combined with the generalised Born and surface area continuum solvation (MM-GBSA) allow estimating binding free energy ΔG, which determines the strength of protein-ligand binding[236]. This technique has been effectively utilised to explain experimental results as well as to enhance virtual screening and docking outcomes [237]. When derived from the MD trajectory, MM-GBSA offers data on the interactions the ligand employs to anchor itself inside the binding pocket. These calculations may be used in the hit-to-lead approach to guide synthetic chemists' efforts toward improving the weakest feature of the ligand or to indicate which moiety should be modified to boost binding affinity.

In computational chemistry, ΔG Binding Energy is frequently calculated in accordance with the thermodynamic cycle (Equation 6)[238].

$$\Delta G_{bind,solv} = \Delta G_{bind,vacuum} + \Delta G_{solv,complex} - (\Delta G_{solv,ligand} + \Delta G_{solv,receptor})$$

Equation 6. Binding-free energy in computational approaches

In turn, the solvation-free energy (ΔG_{solv}) is described as a sum of the polar (ΔG_{pol}) and non-polar (ΔG_{nonpol}) components. When MM/GBSA is applied on the MD trajectory, the calculations are performed for every simulation frame in accordance with Equation 7.

$$\Delta G_{bind} = < \Delta G^{complex}(i) - \Delta G^{protein}(i) - \Delta G^{ligand}(i) >_i$$

Equation 7. Binding-free energy applied for MD trajectory, where $<...>_i$ indicates that calculations are conducted over *i* simulation frames.

Next, each ΔG in Equation 7 is decomposed to the terms described in Equation 8.

$$\Delta G = E_{int} + E_{ele} + E_{vdw} + \Delta G_{pol} + \Delta G_{np} - T\Delta S$$

Equation 8. Decomposition of Binding-free energy into physicochemical terms

In equation 8, E_{int} stands for the molecular mechanic internal energy, E_{ele} for electrostatic energy, and E_{vdw} for Van der Waals energies. The last three terms of this equation mean the following: The ΔG_{pol} and ΔG_{np} are the polar and nonpolar solvation-free energies, T is the absolute temperature, and ΔS represents the entropy estimate. While in MD, the protein is

frequently solvated in a water box, before the MM/GBSA computations, the water molecules are removed and replaced with an implicit representation[239]. The generalised Born (GB, reflects term MM/GBSA) or Poisson-Boltzmann (PB, reflects MM/PBSA) can be used as implicit representation[240]. For the GB model, the polar solvation energy is described with Equation 9 [238].

$$\Delta G_{pol} = \sum_{ij} \Delta G_{tj}^{pol} \approx -\frac{1}{2}\left(\frac{1}{\epsilon_{in}} - \frac{1}{\epsilon_{out}}\right)\frac{1}{1+\beta\alpha}\sum_{ij} q_i q_j \left(\frac{1}{f_{ij}^{GB}} + \frac{\alpha\beta}{A}\right)$$

Equation 9. The polar solvation energy for the GB model.

In equation 9, ε_{in} = 1 (dielectric constant of the solute) and ε_{out} = 80 (dielectric constant of the solvent); $\beta = \varepsilon_{in}/\varepsilon_{out}$; α = 0.571412. The A corresponds to the molecule's electrostatic dimensions. The f_{ij}^{GB} functional form (Equation 10) describes the distance between atomic charges (r_{ij}) and effective Born radii (R), which in turn indicates the depth with how each atom is buried in the solvent (Equation 11)[241,242].

$$\boldsymbol{f_{ij}^{GB} = \left[r_{ij}^2 + R_i R_j exp\left(-\frac{r_{ij}^2}{4R_i R_j}\right)\right]^{\frac{1}{2}}}$$

Equation 10.

$$\boldsymbol{R_i^{-3} = \left(-\frac{1}{4\pi}\oint \frac{r - r_i}{|r - r_i|^6} \cdot dS\right)}$$

Equation 11. Effective Born radii. Contour integral is limited to the dielectric boundary *(∂V)* of the molecule, r and r_i describe the molecule's position per the surface vector element *dS.*

The ΔG_{np} from equation 8 is estimated as proportional to the molecules' solvent accessible area (SASA) multiplied by the factor γ[243].

Once conducted, MM-GBSA calculations generate plenty of energy properties. These properties, which are broken down into contributions from different components in the energy expression, report energies for the ligand, receptor, and complex structures, as well as energy changes related to strain and binding. Moreover, MM-GBSA can report the ligand efficiency,

which is calculated as divided by the number of heavy atoms. This, in turn, allows to indicate the energy penalty to adopt the conformation of the ligand into the binding conformation.

Ligand properties

Plenty of ligand properties or quantitative descriptors are routinely applied for compound filtering before or after virtual screenings, and large-scale dockings can be used for calculation on an MD trajectory. However, the number of properties that are worth calculating as a function of time is quite limited and primarily used for purposes in industrial drug development. The typical examples of such descriptors are polar surface area (PSA) or solvent-accessible surface area (SASA). PSA is used to examine such parameters as oral absorption or blood-brain barrier permeation [244]. Not until the very late stages of drug development does the practical application of this knowledge become significant for medicinal chemistry.

1.3. Concluding Remarks and Outlook

For over 30 years, kinases have been sought as therapeutic targets in cancer. These initiatives have transformed cancer therapy and significantly increased the range of available treatments. Kinase inhibitors have the potential to play a significant role in the treatment of inflammatory and autoimmune diseases, while the clinical pipelines are now growing in various directions. Kinase drug design is constantly increasing, driven by the joint effort of medicinal and computational chemistry, supported by clinical practitioners, biophysicists, and many other specialists. The discipline continues to expand at this high rate because of the propensity for ongoing discoveries and innovative insights connected to protein kinases, while innovative *In Silico* approaches continue to facilitate the expansion of possibilities in kinase drug design into a novel target space.

The addressed above In Silico drug design approaches fall into two main categories: one focuses on comprehending the processes and subprocesses involved in disease development, and the other offers brand-new methods for disease intervention. In other words, the first is the use of computing to aid in elucidating the molecular mechanisms underlying abnormal biological processes. Specifically, locating previously undiscovered biomolecular events that might be connected to the pathophysiological mechanisms in potentially curable diseases, determining the ligand binding mode at a particular binding site with potential modulation of a protein's function for a specific circumstance, and investigating the mutational consequences related to disease progression.

It is important to emphasise additional issues that can be accomplished with the In Silico approach while addressing the second point of new molecular strategy identification. These

include the design of novel and highly selective drugs via virtual screening campaigns combined with experimental techniques; prediction and simultaneous optimisation of biopharmaceutical properties, such as solubility and toxicity predictions; discovery of novel and potentially druggable protein pockets. In addition, the stabilisation of a specific protein substructure or residue that encourages system-specific interactions can be used for drug development employing ligand feature identification.

The published manuscripts demonstrate how In Silico methods contribute to the drug discovery area by highlighting minor events occurring in the atomistic level of protein structures. For instance, the first manuscript demonstrates the mechanism and importance of dimerization for MKK4 functioning. Furthermore, it proposes how cancer's most frequent MKK4 mutation influences the protein structure and leads to pathogenic effects. This knowledge provides a basis for further research and development of novel therapeutics and inhibitors. The second manuscript offers a comprehensive overview of the hydrophobic regulatory spine (R-spine), a core element of protein kinases. The deep focus on the RS3 R-spine residue is substantiated by its location deep in the ATP-binding cleft, hence to the chief inhibitors binding site. Finally, the presented data provide an overview and correlation between the FDA-approved inhibitors and the frequently overlooked hydrophobic interactions with the R-spine. The third publication demonstrates extensive molecular dynamics studies, resulting in an understanding of the inhibition effect on tetramerization and providing a glance into the water interaction network of FABI.

Together, these published works underline the importance of the subtle details in the protein function and cell machinery. Providing the knowledge that fills the understanding gaps, In Silico methods emphasize the development of new therapeutics as composed of minor details, each of which contributes to the broader picture.

References

1. Kooistra, A. J. & Volkamer, A. Chapter Six - Kinase-Centric Computational Drug Development. in *Annual Reports in Medicinal Chemistry* (ed. Goodnow, R. A.) vol. 50 197–236 (Academic Press, 2017).
2. Schwartz, P. A. & Murray, B. W. Protein kinase biochemistry and drug discovery. *Bioorganic Chem.* **39**, 192–210 (2011).
3. Roskoski, R. Hydrophobic and polar interactions of FDA-approved small molecule protein kinase inhibitors with their target enzymes. *Pharmacol. Res.* **169**, 105660 (2021).
4. Sever, R. & Brugge, J. S. Signal Transduction in Cancer. *Cold Spring Harb. Perspect. Med.* **5**, a006098 (2015).
5. Bhullar, K. S. *et al.* Kinase-targeted cancer therapies: progress, challenges and future directions. *Mol. Cancer* **17**, 48 (2018).
6. Cohen, P. Protein kinases — the major drug targets of the twenty-first century? *Nat. Rev. Drug Discov.* **1**, 309–315 (2002).
7. Theivendren, P. *et al. Importance of Protein Kinase and Its Inhibitor: A Review*. *Protein Kinases - Promising Targets for Anticancer Drug Research* (IntechOpen, 2021). doi:10.5772/intechopen.98552.
8. Berginski, M. E. *et al.* The Dark Kinase Knowledgebase: an online compendium of knowledge and experimental results of understudied kinases. *Nucleic Acids Res.* **49**, D529–D535 (2021).
9. Alexander, P. B. & Wang, X.-F. Resistance to receptor tyrosine kinase inhibition in cancer: molecular mechanisms and therapeutic strategies. *Front. Med.* **9**, 134–138 (2015).
10. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The Protein Kinase Complement of the Human Genome. *Science* **298**, 1912–1934 (2002).
11. KREBS, E. G. The phosphorylation of proteins: a major mechanism for biological regulation. *Biochem. Soc. Trans.* **13**, 813–820 (1985).
12. Morrison, D. K. MAP Kinase Pathways. *Cold Spring Harb. Perspect. Biol.* **4**, a011254 (2012).
13. Cuevas, B. D., Abell, A. N. & Johnson, G. L. Role of mitogen-activated protein kinase kinase kinases in signal integration. *Oncogene* **26**, 3159–3171 (2007).
14. Eid, S., Turk, S., Volkamer, A., Rippmann, F. & Fulle, S. KinMap: a web-based tool for interactive navigation through human kinome data. *BMC Bioinformatics* **18**, 16 (2017).
15. Kanev, G. K. *et al.* The Landscape of Atypical and Eukaryotic Protein Kinases. *Trends Pharmacol. Sci.* **40**, 818–832 (2019).
16. Islam, Md. S. Calcium Signaling: From Basic to Bedside. in *Calcium Signaling* (ed. Islam, Md. S.) 1–6 (Springer International Publishing, 2020). doi:10.1007/978-3-030-12457-1_1.
17. Pearce, L. R., Komander, D. & Alessi, D. R. The nuts and bolts of AGC protein kinases. *Nat. Rev. Mol. Cell Biol.* **11**, 9–22 (2010).
18. Varjosalo, M. *et al.* The Protein Interaction Landscape of the Human CMGC Kinase Group. *Cell Rep.* **3**, 1306–1320 (2013).
19. Kannan, N. & Neuwald, A. F. Evolutionary constraints associated with functional specificity of the CMGC protein kinases MAPK, CDK, GSK, SRPK, DYRK, and CK2α. *Protein Sci. Publ. Protein Soc.* **13**, 2059–2077 (2004).
20. Dan, I., Watanabe, N. M. & Kusumi, A. The Ste20 group kinases as regulators of MAP kinase cascades. *Trends Cell Biol.* **11**, 220–230 (2001).
21. Schittek, B. & Sinnberg, T. Biological functions of casein kinase 1 isoforms and putative roles in tumorigenesis. *Mol. Cancer* **13**, 231 (2014).
22. Knippschild, U. *et al.* The CK1 Family: Contribution to Cellular Stress Response and Its Role in Carcinogenesis. *Front. Oncol.* **4**, (2014).
23. Duong-Ly, K. C. & Peterson, J. R. The Human Kinome and Kinase Inhibition as a therapeutic strategy. *Curr. Protoc. Pharmacol. Editor. Board SJ Enna Ed.--Chief Al* **0 2**, Unit2.9 (2013).
24. Arter, C., Trask, L., Ward, S., Yeoh, S. & Bayliss, R. Structural features of the protein kinase domain and targeted binding by small-molecule inhibitors. *J. Biol. Chem.* **298**, (2022).
25. Cowan-Jacob, S. W. & Moebitz, H. Ten things you should know about protein kinases: IUPHAR Review 14: Ten things you should know about protein kinases. *Br. J. Pharmacol.* **172**, 2675–2700 (2015).
26. Li, E. & Hristova, K. Receptor tyrosine kinase transmembrane domains. *Cell Adhes. Migr.* **4**, 249–254 (2010).
27. Finger, C., Escher, C. & Schneider, D. The Single Transmembrane Domains of Human Receptor Tyrosine Kinases Encode Self-Interactions. *Sci. Signal.* **2**, ra56–ra56 (2009).
28. Hubbard, S. R. & Miller, W. T. Receptor tyrosine kinases: mechanisms of activation and signaling. *Curr. Opin. Cell Biol.* **19**, 117–123 (2007).
29. Knighton, D. R. *et al.* Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science* **253**, 407–414 (1991).

30. Zheng, J. *et al.* Crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MgATP and peptide inhibitor. *Biochemistry* **32**, 2154–2161 (1993).
31. Taylor, S. S. *et al.* From structure to the dynamic regulation of a molecular switch: A journey over 3 decades. *J. Biol. Chem.* **296**, 100746 (2021).
32. Roskoski, R. A historical overview of protein kinases and their targeted small molecule inhibitors. *Pharmacol. Res.* **100**, 1–23 (2015).
33. Cheung, J. *et al.* Structural insights into mis-regulation of protein kinase A in human tumors. *Proc. Natl. Acad. Sci.* **112**, 1374–1379 (2015).
34. Meharena, H. S. *et al.* Deciphering the Structural Basis of Eukaryotic Protein Kinase Regulation. *PLoS Biol.* **11**, e1001680 (2013).
35. Taylor, S. S., Keshwani, M. M., Steichen, J. M. & Kornev, A. P. Evolution of the eukaryotic protein kinases as dynamic molecular switches. *Philos. Trans. R. Soc. B Biol. Sci.* **367**, 2517–2528 (2012).
36. Adams, J. A. Activation Loop Phosphorylation and Catalysis in Protein Kinases: Is There Functional Evidence for the Autoinhibitor Model? *Biochemistry* **42**, 601–607 (2003).
37. 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.* **20**, 839–861 (2021).
38. Taylor, S. S., Shaw, A. S., Kannan, N. & Kornev, A. Integration of Signaling in the Kinome: Architecture and Regulation of the αC Helix. *Biochim. Biophys. Acta* **1854**, 1567–1574 (2015).
39. King, A. J. *et al.* Demonstration of a Genetic Therapeutic Index for Tumors Expressing Oncogenic BRAF by the Kinase Inhibitor SB-590885. *Cancer Res.* **66**, 11100–11105 (2006).
40. Zhang, C. *et al.* RAF inhibitors that evade paradoxical MAPK pathway activation. *Nature* **526**, 583–586 (2015).
41. Kornev, A. P. & Taylor, S. S. Defining the conserved internal architecture of a protein kinase. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **1804**, 440–444 (2010).
42. Cl, M., Ap, K., Mk, G. & Ss, T. Dynamic architecture of a protein kinase. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E4623-31 (2014).
43. Ung, P. M.-U., Rahman, R. & Schlessinger, A. Redefining the Protein Kinase Conformational Space with Machine Learning. *Cell Chem. Biol.* **25**, 916-924.e2 (2018).
44. Nolen, B., Taylor, S. & Ghosh, G. Regulation of Protein Kinases: Controlling Activity through Activation Segment Conformation. *Mol. Cell* **15**, 661–675 (2004).
45. Lochhead, P. A. Protein Kinase Activation Loop Autophosphorylation in Cis: Overcoming a Catch-22 Situation. *Sci. Signal.* **2**, pe4–pe4 (2009).
46. Hanks, S. K., Quinn, A. M. & Hunter, T. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science* **241**, 42–52 (1988).
47. Johnson, L. N., Noble, M. E. M. & Owen, D. J. Active and Inactive Protein Kinases: Structural Basis for Regulation. *Cell* **85**, 149–158 (1996).
48. Adams, J. A. Kinetic and Catalytic Mechanisms of Protein Kinases. *Chem. Rev.* **101**, 2271–2290 (2001).
49. Steichen, J. M. *et al.* Global Consequences of Activation Loop Phosphorylation on Protein Kinase A. *J. Biol. Chem.* **285**, 3825–3832 (2010).
50. Fox, T. *et al.* A single amino acid substitution makes ERK2 susceptible to pyridinyl imidazole inhibitors of p38 MAP kinase. *Protein Sci.* **7**, 2249–2255 (1998).
51. Romano, V., de Beer, T. A. P. & Schwede, T. A computational protocol to evaluate the effects of protein mutants in the kinase gatekeeper position on the binding of ATP substrate analogues. *BMC Res. Notes* **10**, 104 (2017).
52. Elphick, L. M., Lee, S. E., Gouverneur, V. & Mann, D. J. Using Chemical Genetics and ATP Analogues To Dissect Protein Kinase Function. *ACS Chem. Biol.* **2**, 299–314 (2007).
53. Liu, Y., Shah, K., Yang, F., Witucki, L. & Shokat, K. M. A molecular gate which controls unnatural ATP analogue recognition by the tyrosine kinase v-Src. *Bioorg. Med. Chem.* **6**, 1219–1226 (1998).
54. Blencke, S. *et al.* Characterization of a Conserved Structural Determinant Controlling Protein Kinase Sensitivity to Selective Inhibitors. *Chem. Biol.* **11**, 691–701 (2004).
55. Zuccotto, F., Ardini, E., Casale, E. & Angiolini, M. Through the "Gatekeeper Door": Exploiting the Active Kinase Conformation. *J. Med. Chem.* **53**, 2681–2694 (2010).
56. Qin, J. *et al.* Identification of a Novel Family of BRAFV600E Inhibitors. *J. Med. Chem.* **55**, 5220–5230 (2012).
57. Taylor, S. S., Shaw, A. S., Kannan, N. & Kornev, A. P. Integration of signaling in the kinome: Architecture and regulation of the αC Helix. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **1854**, 1567–1574 (2015).
58. Taylor, S. S. & Kornev, A. P. Protein kinases: evolution of dynamic regulatory proteins. *Trends Biochem. Sci.* **36**, 65–77 (2011).
59. Shevchenko, E. & Pantsar, T. Regulatory spine RS3 residue of protein kinases: a lipophilic bystander or a decisive element in the small-molecule kinase inhibitor binding? *Biochem. Soc. Trans.* **50**, 633–648 (2022).

60. Kornev, A. P., Haste, N. M., Taylor, S. S. & Ten Eyck, L. F. Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proc. Natl. Acad. Sci.* **103**, 17783–17788 (2006).
61. Kornev, A. P., Taylor, S. S. & Ten Eyck, L. F. A helix scaffold for the assembly of active protein kinases. *Proc. Natl. Acad. Sci.* **105**, 14377–14382 (2008).
62. Wan, P. T. C. *et al.* Mechanism of Activation of the RAF-ERK Signaling Pathway by Oncogenic Mutations of B-RAF. *Cell* **116**, 855–867 (2004).
63. Barrett, C. P. & Noble, M. E. M. Molecular motions of human cyclin-dependent kinase 2. *J. Biol. Chem.* **280**, 13993–14005 (2005).
64. Hu, J. *et al.* Kinase regulation by hydrophobic spine assembly in cancer. *Mol. Cell. Biol.* **35**, 264–276 (2015).
65. Yap, J. *et al.* The stability of R-spine defines RAF inhibitor resistance: A comprehensive analysis of oncogenic BRAF mutants with in-frame insertion of αC-β4 loop. *Sci. Adv.* **7**, eabg0390 (2021).
66. Cowan-Jacob, S. W., Jahnke, W. & Knapp, S. Novel approaches for targeting kinases: allosteric inhibition, allosteric activation and pseudokinases. *Future Med. Chem.* **6**, 541–561 (2014).
67. Liu, Q. *et al.* Developing irreversible inhibitors of the protein kinase cysteinome. *Chem. Biol.* **20**, 146–159 (2013).
68. Gehringer, M. Covalent Kinase Inhibitors: An Overview. in *Proteinkinase Inhibitors* (ed. Laufer, S.) 43–94 (Springer International Publishing, 2021). doi:10.1007/7355_2020_103.
69. Roskoski, R. Properties of FDA-approved small molecule protein kinase inhibitors: A 2021 update. *Pharmacol. Res.* **165**, 105463 (2021).
70. Müller, S., Chaikuad, A., Gray, N. S. & Knapp, S. The ins and outs of selective kinase inhibitor development. *Nat. Chem. Biol.* **11**, 818–821 (2015).
71. Knowles, P. P. *et al.* Structure and Chemical Inhibition of the RET Tyrosine Kinase Domain. *J. Biol. Chem.* **281**, 33577–33587 (2006).
72. Zhang, J., Yang, P. L. & Gray, N. S. Targeting cancer with small molecule kinase inhibitors. *Nat. Rev. Cancer* **9**, 28–39 (2009).
73. Oneal, P. A. *et al.* FDA Approval Summary: Vemurafenib for the Treatment of Patients with Erdheim-Chester Disease with the BRAFV600 Mutation. *The Oncologist* **23**, 1520–1524 (2018).
74. Kazandjian, D. *et al.* FDA Approval of Gefitinib for the Treatment of Patients with Metastatic EGFR Mutation–Positive Non–Small Cell Lung Cancer. *Clin. Cancer Res.* **22**, 1307–1312 (2016).
75. Liu, Y. & Gray, N. S. Rational design of inhibitors that bind to inactive kinase conformations. *Nat. Chem. Biol.* **2**, 358–364 (2006).
76. McTigue, M. *et al.* Molecular conformations, interactions, and properties associated with drug efficiency and clinical performance among VEGFR TK inhibitors. *Proc. Natl. Acad. Sci.* **109**, 18281–18289 (2012).
77. Matsuki, M. *et al.* Lenvatinib inhibits angiogenesis and tumor fibroblast growth factor signaling pathways in human hepatocellular carcinoma models. *Cancer Med.* **7**, 2641–2653 (2018).
78. Karoulia, Z. *et al.* An Integrated Model of RAF Inhibitor Action Predicts Inhibitor Activity against Oncogenic BRAF Signaling. *Cancer Cell* **30**, 485–498 (2016).
79. Wood, E. R. *et al.* A unique structure for epidermal growth factor receptor bound to GW572016 (Lapatinib): relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells. *Cancer Res.* **64**, 6652–6659 (2004).
80. Roskoski, R. Classification of small molecule protein kinase inhibitors based upon the structures of their drug-enzyme complexes. *Pharmacol. Res.* **103**, 26–48 (2016).
81. Vanderpool, D. *et al.* Characterization of the CHK1 Allosteric Inhibitor Binding Site. *Biochemistry* **48**, 9823–9830 (2009).
82. BARNETT, S. F. *et al.* Identification and characterization of pleckstrin-homology-domain-dependent and isoenzyme-specific Akt inhibitors. *Biochem. J.* **385**, 399–408 (2005).
83. Ohren, J. F. *et al.* Structures of human MAP kinase kinase 1 (MEK1) and MEK2 describe novel noncompetitive kinase inhibition. *Nat. Struct. Mol. Biol.* **11**, 1192–1197 (2004).
84. Adrián, F. J. *et al.* Allosteric inhibitors of Bcr-abl–dependent cell proliferation. *Nat. Chem. Biol.* **2**, 95–102 (2006).
85. Lu, X., Smaill, J. B. & Ding, K. New Promise and Opportunities for Allosteric Kinase Inhibitors. *Angew. Chem. Int. Ed.* **59**, 13764–13776 (2020).
86. Gehringer, M. Covalent inhibitors: back on track? *Future Med. Chem.* **12**, 1363–1368 (2020).
87. Johnson, T. W. *et al.* Discovery of (10R)-7-amino-12-fluoro-2,10,16-trimethyl-15-oxo-10,15,16,17-tetrahydro-2H-8,4-(metheno)pyrazolo[4,3-h][2,5,11]-benzoxadiazacyclotetradecine-3-carbonitrile (PF-06463922), a macrocyclic inhibitor of anaplastic lymphoma kinase (ALK) and c-ros oncogene 1 (ROS1) with preclinical brain exposure and broad-spectrum potency against ALK-resistant mutations. *J. Med. Chem.* **57**, 4720–4744 (2014).

88. Posternak, G. *et al.* Functional characterization of a PROTAC directed against BRAF mutant V600E. *Nat. Chem. Biol.* **16**, 1170–1178 (2020).
89. List of clinically approved kinase inhibitors | MRC PPU. *MRC Protein Phosphorylation and Ubiquitylation Unit* https://www.ppu.mrc.ac.uk.
90. Fedorov, O., Müller, S. & Knapp, S. The (un)targeted cancer kinome. *Nat. Chem. Biol.* **6**, 166–169 (2010).
91. Essegian, D., Khurana, R., Stathias, V. & Schürer, S. C. The Clinical Kinase Index: A Method to Prioritize Understudied Kinases as Drug Targets for the Treatment of Cancer. *Cell Rep. Med.* **1**, 100128 (2020).
92. Stransky, N., Cerami, E., Schalm, S., Kim, J. L. & Lengauer, C. The landscape of kinase fusions in cancer. *Nat. Commun.* **5**, 4846 (2014).
93. Burrell, R. A., McGranahan, N., Bartek, J. & Swanton, C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* **501**, 338–345 (2013).
94. McFarland, C. D. *et al.* The Damaging Effect of Passenger Mutations on Cancer Progression. *Cancer Res.* **77**, 4763–4772 (2017).
95. Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).
96. Tuna, M., Amos, C. I. & Mills, G. B. Molecular mechanisms and pathobiology of oncogenic fusion transcripts in epithelial tumors. *Oncotarget* **10**, 2095–2111 (2019).
97. Ben-Neriah, Y., Daley, G. Q., Mes-Masson, A. M., Witte, O. N. & Baltimore, D. The chronic myelogenous leukemia-specific P210 protein is the product of the bcr/abl hybrid gene. *Science* **233**, 212–214 (1986).
98. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2022. *CA. Cancer J. Clin.* **72**, 7–33 (2022).
99. de Groot, P. M., Wu, C. C., Carter, B. W. & Munden, R. F. The epidemiology of lung cancer. *Transl. Lung Cancer Res.* **7**, 220–233 (2018).
100. Cohen, P. Targeting protein kinases for the development of anti-inflammatory drugs. *Curr. Opin. Cell Biol.* **21**, 317–324 (2009).
101. Wollenhaupt, J. *et al.* Safety and efficacy of tofacitinib, an oral janus kinase inhibitor, for the treatment of rheumatoid arthritis in open-label, longterm extension studies. *J. Rheumatol.* **41**, 837–852 (2014).
102. Kim, T. W. *et al.* A critical role for IRAK4 kinase activity in Toll-like receptor-mediated innate immunity. *J. Exp. Med.* **204**, 1025–1036 (2007).
103. Bajpai, M. Fostamatinib, a Syk inhibitor prodrug for the treatment of inflammatory diseases. *IDrugs Investig. Drugs J.* **12**, 174–185 (2009).
104. Ferguson, F. M. & Gray, N. S. Kinase inhibitors: the road ahead. *Nat. Rev. Drug Discov.* **17**, 353–377 (2018).
105. Ghosh, R. *et al.* Allosteric Inhibition of the IRE1α RNase Preserves Cell Viability and Function during Endoplasmic Reticulum Stress. *Cell* **158**, 534–548 (2014).
106. TG Therapeutics, Inc. *Phase I Study of Ublituximab and Umbralisib in Combination With Targeted Immunotherapy in Patients With Relapsed-refractory Chronic Lymphocytic Leukemia (CLL) or Richter's Transformation (RT) of CLL*. https://clinicaltrials.gov/ct2/show/NCT02535286 (2022).
107. Lucet, I. S., Tobin, A., Drewry, D., Wilks, A. F. & Doerig, C. Plasmodium kinases as targets for new-generation antimalarials. *Future Med. Chem.* **4**, 2295–2310 (2012).
108. Jm, C., S, M., Cs, B. & Sm, P. Sphingosine kinase 1 in viral infections. *Rev. Med. Virol.* **23**, (2013).
109. Prisic, S. & Husson, R. N. Mycobacterium tuberculosis Serine/Threonine Protein Kinases. *Microbiol. Spectr.* **2**, (2014).
110. Pillaiyar, T. & Laufer, S. Kinases as Potential Therapeutic Targets for Anti-coronaviral Therapy. *J. Med. Chem.* acs.jmedchem.1c00335 (2021) doi:10.1021/acs.jmedchem.1c00335.
111. Prieto-Martínez, F. D., López-López, E., Eurídice Juárez-Mercado, K. & Medina-Franco, J. L. Chapter 2 - Computational Drug Design Methods—Current and Future Perspectives. in *In Silico Drug Design* (ed. Roy, K.) 19–44 (Academic Press, 2019). doi:10.1016/B978-0-12-816125-8.00002-X.
112. Wiley, J. Perspectives in Supramolecular Chemistry Volume I. 30.
113. Schneider, H.-J. Introduction to Molecular Recognition Models. in *Protein-Ligand Interactions* 21–50 (John Wiley & Sons, Ltd, 2003). doi:10.1002/3527601813.ch2.
114. Osakwe, O. Chapter 5 - The Significance of Discovery Screening and Structure Optimization Studies. in *Social Aspects of Drug Discovery, Development and Commercialization* (eds. Osakwe, O. & Rizvi, S. A. A.) 109–128 (Academic Press, 2016). doi:10.1016/B978-0-12-802220-7.00005-3.
115. Klebe, G. Protein Modeling and Structure-Based Drug Design. in *Drug Design: Methodology, Concepts, and Mode-of-Action* (ed. Klebe, G.) 429–448 (Springer, 2013). doi:10.1007/978-3-642-17907-5_20.
116. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
117. Levitt, D. G. & Banaszak, L. J. POCKET: a computer graphics method for identifying and displaying protein cavities and their surrounding amino acids. *J. Mol. Graph.* **10**, 229–234 (1992).

118. Brady, G. P. & Stouten, P. F. Fast prediction and visualization of protein binding pockets with PASS. *J. Comput. Aided Mol. Des.* **14**, 383–401 (2000).
119. Hendlich, M., Rippmann, F. & Barnickel, G. LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *J. Mol. Graph. Model.* **15**, 359–363, 389 (1997).
120. Huang, B. & Schroeder, M. LIGSITEcsc: predicting ligand binding sites using the Connolly surface and degree of conservation. *BMC Struct. Biol.* **6**, 19 (2006).
121. Halgren, T. New Method for Fast and Accurate Binding-site Identification and Analysis. *Chem. Biol. Drug Des.* **69**, 146–148 (2007).
122. Halgren, T. A. Identifying and Characterizing Binding Sites and Assessing Druggability. *J. Chem. Inf. Model.* **49**, 377–389 (2009).
123. Opo, F. A. D. M. *et al.* Structure based pharmacophore modeling, virtual screening, molecular docking and ADMET approaches for identification of natural anti-cancer agents targeting XIAP protein. *Sci. Rep.* **11**, 4049 (2021).
124. Martin, Y. C., Kofron, J. L. & Traphagen, L. M. Do Structurally Similar Molecules Have Similar Biological Activity? *J. Med. Chem.* **45**, 4350–4358 (2002).
125. Wermuth, C. G., Ganellin, C. R., Lindberg, P. & Mitscher, L. A. Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). *Pure Appl. Chem.* **70**, 1129–1143 (1998).
126. Bajusz, D., Rácz, A. & Héberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminformatics* **7**, 20 (2015).
127. Rácz, A., Bajusz, D. & Héberger, K. Life beyond the Tanimoto coefficient: similarity measures for interaction fingerprints. *J. Cheminformatics* **10**, 48 (2018).
128. Williams, M. A. & Ladbury, J. E. Hydrogen Bonds in Protein-Ligand Complexes. in *Protein-Ligand Interactions* 137–161 (John Wiley & Sons, Ltd, 2003). doi:10.1002/3527601813.ch6.
129. Bitencourt-Ferreira, G., Veit-Acosta, M. & de Azevedo, W. F. Hydrogen Bonds in Protein-Ligand Complexes. in *Docking Screens for Drug Discovery* (ed. de Azevedo Jr., W. F.) 93–107 (Springer, 2019). doi:10.1007/978-1-4939-9752-7_7.
130. Kuster, D. J., Liu, C., Fang, Z., Ponder, J. W. & Marshall, G. R. High-Resolution Crystal Structures of Protein Helices Reconciled with Three-Centered Hydrogen Bonds and Multipole Electrostatics. *PLOS ONE* **10**, e0123146 (2015).
131. Jacobson, M. P., Friesner, R. A., Xiang, Z. & Honig, B. On the Role of the Crystal Environment in Determining Protein Side-chain Conformations. *J. Mol. Biol.* **320**, 597–608 (2002).
132. Lovell, S. C., Word, J. M., Richardson, J. S. & Richardson, D. C. Asparagine and glutamine rotamers: *B* -factor cutoff and correction of amide flips yield distinct clustering. *Proc. Natl. Acad. Sci.* **96**, 400–405 (1999).
133. Weichenberger, C. X. & Sippl, M. J. NQ-Flipper: recognition and correction of erroneous asparagine and glutamine side-chain rotamers in protein structures. *Nucleic Acids Res.* **35**, W403–W406 (2007).
134. Kossiakoff, A. A., Randal, M., Guenot, J. & Eignebrot, C. Variability of conformations at crystal contacts in BPTI represent true low-energy structures: Correspondence among lattice packing and molecular dynamics structures. *Proteins Struct. Funct. Bioinforma.* **14**, 65–74 (1992).
135. Eigenbrot, C., Randal, M. & Kossiakoff, A. A. Structural effects induced by mutagenesis affected by crystal packing factors: The structure of a 30–51 disulfide mutant of basic pancreatic trypsin inhibitor. *Proteins Struct. Funct. Bioinforma.* **14**, 75–87 (1992).
136. Zhang, X., Wozniak, J. A. & Matthews, B. W. Protein Flexibility and Adaptability Seen in 25 Crystal Forms of T4 Lysozyme. *J. Mol. Biol.* **250**, 527–552 (1995).
137. Bauer, P., Hess, B. & Lindahl, E. GROMACS 2022.3 Manual. (2022) doi:10.5281/zenodo.7037337.
138. Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009).
139. Dominguez, C., Boelens, R. & Bonvin, A. M. J. J. HADDOCK: A Protein−Protein Docking Approach Based on Biochemical or Biophysical Information. *J. Am. Chem. Soc.* **125**, 1731–1737 (2003).
140. van Zundert, G. C. P. *et al.* The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **428**, 720–725 (2016).
141. Madhavi Sastry, G., 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).
142. Pearson, W. R. An Introduction to Sequence Similarity ("Homology") Searching. *Curr. Protoc. Bioinforma. Ed. Board Andreas Baxevanis Al* **0 3**, 10.1002/0471250953.bi0301s42 (2013).
143. Grigoriev, I. V. & Kim, S.-H. Detection of protein fold similarity based on correlation of amino acid properties. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 14318–14323 (1999).
144. Bayrak, A. *et al.* Discovery and Development of First-in-Class ACKR3/CXCR7 Superagonists for Platelet Degranulation Modulation. *J. Med. Chem.* **65**, 13365–13384 (2022).

145. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
146. Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S. & Thornton, J. M. PDBsum: Structural summaries of PDB entries. *Protein Sci. Publ. Protein Soc.* **27**, 129–134 (2018).
147. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci. Publ. Protein Soc.* **27**, 293–315 (2018).
148. Shevchenko, E., Poso, A. & Pantsar, T. The autoinhibited state of MKK4: Phosphorylation, putative dimerization and R134W mutant studied by molecular dynamics simulations. *Comput. Struct. Biotechnol. J.* **18**, 2687–2698 (2020).
149. Matsumoto, T. *et al.* Crystal structures of MKK4 kinase domain reveal that substrate peptide binds to an allosteric site and induces an auto-inhibition state. *Biochem. Biophys. Res. Commun.* **400**, 369–373 (2010).
150. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95–99 (1963).
151. Ramachandran, G. N. & Sasisekharan, V. Conformation of polypeptides and proteins. *Adv. Protein Chem.* **23**, 283–438 (1968).
152. Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
153. Eswar, N., Eramian, D., Webb, B., Shen, M.-Y. & Sali, A. Protein Structure Modeling with MODELLER. in *Structural Proteomics: High-Throughput Methods* (eds. Kobe, B., Guss, M. & Huber, T.) 145–159 (Humana Press, 2008). doi:10.1007/978-1-60327-058-8_8.
154. Pieper, U. *et al.* ModBase, a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res.* **42**, D336–D346 (2014).
155. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
156. Jacobson, M. P. *et al.* A hierarchical approach to all-atom protein loop prediction. *Proteins Struct. Funct. Bioinforma.* **55**, 351–367 (2004).
157. Kitchen, D. B., Decornez, H., Furr, J. R. & Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* **3**, 935–949 (2004).
158. Brooijmans, N. & Kuntz, I. D. Molecular Recognition and Docking Algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **32**, 335–373 (2003).
159. 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).
160. 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).
161. 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).
162. Eberhardt, J., Santos-Martins, D., Tillack, A. F. & Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **61**, 3891–3898 (2021).
163. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).
164. Verdonk, M. L. *et al.* Virtual Screening Using Protein−Ligand Docking: Avoiding Artificial Enrichment. *J. Chem. Inf. Comput. Sci.* **44**, 793–806 (2004).
165. Huang, N., Shoichet, B. K. & Irwin, J. J. Benchmarking Sets for Molecular Docking. *J. Med. Chem.* **49**, 6789–6801 (2006).
166. Lätti, S., Niinivehmas, S. & Pentikäinen, O. T. Rocker: Open source, easy-to-use tool for AUC and enrichment calculations and ROC visualization. *J. Cheminformatics* **8**, 45 (2016).
167. Clark, R. D. & Webster-Clark, D. J. Managing bias in ROC curves. *J. Comput. Aided Mol. Des.* **22**, 141–146 (2008).
168. Kenyon, V., Chorny, I., Carvajal, W. J., Holman, T. R. & Jacobson, M. P. Novel human lipoxygenase inhibitors discovered using virtual screening with homology models. *J. Med. Chem.* **49**, 1356–1363 (2006).
169. Li, H. *et al.* Identification of Novel Falcipain-2 Inhibitors as Potential Antimalarial Agents through Structure-Based Virtual Screening. *J. Med. Chem.* **52**, 4936–4940 (2009).
170. Shah, F. *et al.* Computer-Aided Drug Design of Falcipain Inhibitors: Virtual Screening, Structure–Activity Relationships, Hydration Site Thermodynamics, and Reactivity Analysis. *J. Chem. Inf. Model.* **52**, 696–710 (2012).
171. Cavalli, A., Bottegoni, G., Raco, C., De Vivo, M. & Recanatini, M. A Computational Study of the Binding of Propidium to the Peripheral Anionic Site of Human Acetylcholinesterase. *J. Med. Chem.* **47**, 3991–3999 (2004).
172. Robustelli, P. *et al.* Molecular Basis of Small-Molecule Binding to α-Synuclein. *J. Am. Chem. Soc.* **144**, 2501–2510 (2022).

173. Dror, R. O., Jensen, M. Ø., Borhani, D. W. & Shaw, D. E. Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations. *J. Gen. Physiol.* **135**, 555–562 (2010).
174. Shan, Y. *et al.* How does a small molecule bind at a cryptic binding site? *PLOS Comput. Biol.* **18**, e1009817 (2022).
175. Mysore, V. P. *et al.* A structural model of a Ras–Raf signalosome. *Nat. Struct. Mol. Biol.* **28**, 847–857 (2021).
176. Wang, Y., Lupala, C. S., Liu, H. & Lin, X. Identification of Drug Binding Sites and Action Mechanisms with Molecular Dynamics Simulations. *Curr. Top. Med. Chem.* **18**, 2268–2277 (2018).
177. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007).
178. 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).
179. Grossfield, A. & Zuckerman, D. M. Quantifying uncertainty and sampling quality in biomolecular simulations. *Annu. Rep. Comput. Chem.* **5**, 23–48 (2009).
180. Oostenbrink, C., van Lipzig, M. M. H. & van Gunsteren, W. F. 4.25 - Applications of Molecular Dynamics Simulations in Drug Design. in *Comprehensive Medicinal Chemistry II* (eds. Taylor, J. B. & Triggle, D. J.) 651–668 (Elsevier, 2007). doi:10.1016/B0-08-045044-X/00268-6.
181. Gao, X., Gallicchio, E. & Roitberg, A. E. The generalized Boltzmann distribution is the only distribution in which the Gibbs-Shannon entropy equals the thermodynamic entropy. *J. Chem. Phys.* **151**, 034113 (2019).
182. Chaudhuri, G. & Gupta, S. D. Specific heat and bimodality in canonical and grand canonical versions of the thermodynamic model. *Phys. Rev. C* **76**, 014619 (2007).
183. Dunkel, J. & Hilbert, S. Phase transitions in small systems: Microcanonical vs. canonical ensembles. *Phys. Stat. Mech. Its Appl.* **370**, 390–406 (2006).
184. Gao, X. The mathematics of the ensemble theory. *Results Phys.* **34**, 105230 (2022).
185. Appendix H - Statistical Mechanics of the Gibbs "Ensemble". in *Understanding Molecular Simulation (Second Edition)* (eds. Frenkel, D. & Smit, B.) 563–571 (Academic Press, 2002). doi:10.1016/B978-012267351-1/50027-4.
186. Kurzweil, R. Keynote Speech: Raymond Kurzweil. in *Proceedings of the 2006 ACM/IEEE conference on Supercomputing* 0-es (Association for Computing Machinery, 2006). doi:10.1145/1188455.1198315.
187. Desmond User Manual. 178.
188. Frenkel, D. & Smit, B. Chapter 4 - Molecular Dynamics Simulations. in *Understanding Molecular Simulation (Second Edition)* (eds. Frenkel, D. & Smit, B.) 63–107 (Academic Press, 2002). doi:10.1016/B978-012267351-1/50006-7.
189. Lin, F.-Y. & MacKerell, A. D. Force fields for small molecules. *Methods Mol. Biol. Clifton NJ* **2022**, 21–54 (2019).
190. Appendix A - Equations of Motion from the Lagrangian or Hamiltonian. in *Understanding Molecular Simulation (Second Edition)* (eds. Frenkel, D. & Smit, B.) 481–494 (Academic Press, 2002). doi:10.1016/B978-012267351-1/50020-1.
191. Jorgensen, W. L. & Tirado-Rives, J. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl. Acad. Sci.* **102**, 6665–6670 (2005).
192. Guvench, O. & MacKerell, A. D. Comparison of Protein Force Fields for Molecular Dynamics Simulations. in *Molecular Modeling of Proteins* (ed. Kukol, A.) 63–88 (Humana Press, 2008). doi:10.1007/978-1-59745-177-2_4.
193. Ponder, J. W. & Case, D. A. Force Fields for Protein Simulations. in *Advances in Protein Chemistry* vol. 66 27–85 (Academic Press, 2003).
194. 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).
195. Salomon-Ferrer, R., Case, D. A. & Walker, R. C. An overview of the Amber biomolecular simulation package. *WIREs Comput. Mol. Sci.* **3**, 198–210 (2013).
196. Case, D. A. *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 (2005).
197. Lee, J. *et al.* CHARMM-GUI supports the Amber force fields. *J. Chem. Phys.* **153**, 035103 (2020).
198. Christen, M. *et al.* The GROMOS software for biomolecular simulation: GROMOS05. *J. Comput. Chem.* **26**, 1719–1751 (2005).
199. 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).
200. Roos, K. *et al.* OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *J. Chem. Theory Comput.* **15**, 1863–1874 (2019).
201. Lu, C. *et al.* OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. *J. Chem. Theory Comput.* **17**, 4291–4300 (2021).

202. Tian, C. *et al.* ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **16**, 528–552 (2020).
203. Dickson, C. J., Walker, R. C. & Gould, I. R. Lipid21: Complex Lipid Membrane Simulations with AMBER. *J. Chem. Theory Comput.* **18**, 1726–1736 (2022).
204. Galindo-Murillo, R. *et al.* Assessing the Current State of Amber Force Field Modifications for DNA. *J. Chem. Theory Comput.* **12**, 4114–4127 (2016).
205. Zgarbová, M. *et al.* Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **7**, 2886–2902 (2011).
206. MacKerell, A. D. *et al.* All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
207. Halgren, T. A. & Nachbar, R. B. Merck molecular force field. IV. conformational energies and geometries for MMFF94. *J. Comput. Chem.* **17**, 587–615 (1996).
208. Halgren, T. A. MMFF VI. MMFF94s option for energy minimization studies. *J. Comput. Chem.* **20**, 720–729 (1999).
209. Rappe, A. K., Casewit, C. J., Colwell, K. S., Goddard, W. A. I. & Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **114**, 10024–10035 (1992).
210. Heinz, H., Koerner, H., Anderson, K. L., Vaia, R. A. & Farmer, B. L. Force Field for Mica-Type Silicates and Dynamics of Octadecylammonium Chains Grafted to Montmorillonite. *Chem. Mater.* **17**, 5658–5669 (2005).
211. Wu, J. C., Chattree, G. & Ren, P. Automation of AMOEBA polarizable force field parameterization for small molecules. *Theor. Chem. Acc.* **131**, 1138 (2012).
212. Jing, Z. *et al.* Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications. *Annu. Rev. Biophys.* **48**, 371–394 (2019).
213. Li, H. *et al.* Drude Polarizable Force Field for Molecular Dynamics Simulations of Saturated and Unsaturated Zwitterionic Lipids. *J. Chem. Theory Comput.* **13**, 4535–4552 (2017).
214. Jiang, W. *et al.* High-Performance Scalable Molecular Dynamics Simulations of a Polarizable Force Field Based on Classical Drude Oscillators in NAMD. *J. Phys. Chem. Lett.* **2**, 87–92 (2011).
215. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **111**, 7812–7824 (2007).
216. Wan, M., Song, J., Li, W., Gao, L. & Fang, W. Development of Coarse-Grained Force Field by Combining Multilinear Interpolation Technique and Simplex Algorithm. *J. Comput. Chem.* **41**, 814–829 (2020).
217. Dunn, W. L. Inverse Monte Carlo analysis. *J. Comput. Phys.* **41**, 154–166 (1981).
218. Sargsyan, K., Grauffel, C. & Lim, C. How Molecular Size Impacts RMSD Applications in Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **13**, 1518–1524 (2017).
219. Theobald, D. L. Rapid calculation of RMSDs using a quaternion-based characteristic polynomial. *Acta Crystallogr. A* **61**, 478–480 (2005).
220. Martínez, L. Automatic Identification of Mobile and Rigid Substructures in Molecular Dynamics Simulations and Fractional Structural Fluctuation Analysis. *PLoS ONE* **10**, e0119264 (2015).
221. Pauling, L., Corey, R. B. & Branson, H. R. The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci.* **37**, 205–211 (1951).
222. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).
223. Lyu, Z., Wang, Z., Luo, F., Shuai, J. & Huang, Y. Protein Secondary Structure Prediction With a Reductive Deep Learning Method. *Front. Bioeng. Biotechnol.* **9**, (2021).
224. Jolliffe, I. T. & Cadima, J. Principal component analysis: a review and recent developments. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **374**, 20150202 (2016).
225. Sittel, F., Jain, A. & Stock, G. Principal component analysis of molecular dynamics: On the use of Cartesian vs. internal coordinates. *J. Chem. Phys.* **141**, 014111 (2014).
226. Principal Component Analysis for Time Series and Other Non-Independent Data. in *Principal Component Analysis* (ed. Jolliffe, I. T.) 299–337 (Springer, 2002). doi:10.1007/0-387-22440-8_12.
227. Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. THEORY OF PROTEIN FOLDING: The Energy Landscape Perspective. *Annu. Rev. Phys. Chem.* **48**, 545–600 (1997).
228. Sarich, M., Prinz, J.-H. & Schütte, C. Markov Model Theory. in *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation* (eds. Bowman, G. R., Pande, V. S. & Noé, F.) 23–44 (Springer Netherlands, 2014). doi:10.1007/978-94-007-7606-7_3.
229. Klippenstein, S. J., Pande, V. S. & Truhlar, D. G. Chemical Kinetics and Mechanisms of Complex Systems: A Perspective on Recent Theoretical Advances. *J. Am. Chem. Soc.* **136**, 528–546 (2014).

230. Konovalov, K. A., Unarta, I. C., Cao, S., Goonetilleke, E. C. & Huang, X. Markov State Models to Study the Functional Dynamics of Proteins in the Wake of Machine Learning. *JACS Au* **1**, 1330–1341 (2021).
231. Pande, V. S., Beauchamp, K. & Bowman, G. R. Everything you wanted to know about Markov State Models but were afraid to ask. *Methods San Diego Calif* **52**, 99–105 (2010).
232. McGibbon, R. T., Schwantes, C. R. & Pande, V. S. Statistical Model Selection for Markov Models of Biomolecular Dynamics. *J. Phys. Chem. B* **118**, 6475–6481 (2014).
233. Gowers, R. J. *et al.* MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. *Proc. 15th Python Sci. Conf.* 98–105 (2016) doi:10.25080/Majora-629e541a-00e.
234. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319–2327 (2011).
235. Patil, R. *et al.* Optimized Hydrophobic Interactions and Hydrogen Bonding at the Target-Ligand Interface Leads the Pathways of Drug-Designing. *PLOS ONE* **5**, e12029 (2010).
236. 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).
237. Genheden, S. & Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **10**, 449–461 (2015).
238. 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).
239. Godschalk, F., Genheden, S., Söderhjelm, P. & Ryde, U. Comparison of MM/GBSA calculations based on explicit and implicit solvent simulations. *Phys. Chem. Chem. Phys.* **15**, 7731–7739 (2013).
240. Still, W. C., Tempczyk, A., Hawley, R. C. & Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **112**, 6127–6129 (1990).
241. Aguilar, B. & Onufriev, A. V. Efficient Computation of the Total Solvation Energy of Small Molecules via the R6 Generalized Born Model. *J. Chem. Theory Comput.* **8**, 2404–2411 (2012).
242. Aguilar, B., Shadrach, R. & Onufriev, A. V. Reducing the Secondary Structure Bias in the Generalized Born Model via R6 Effective Radii. *J. Chem. Theory Comput.* **6**, 3613–3630 (2010).
243. Gohlke, H. & Case, D. A. Converging free energy estimates: MM-PB(GB)SA studies on the protein–protein complex Ras–Raf. *J. Comput. Chem.* **25**, 238–250 (2004).
244. Clark, D. E. What has polar surface area ever done for drug discovery? *Future Med. Chem.* **3**, 469–484 (2011).

2. List of figures

3. Results and Discussion

3.1 Publication I

The autoinhibited state of MKK4: Phosphorylation, putative dimerization and R134W mutant studied by molecular dynamics simulations

Ekaterina Shevchenko [a], Antti Poso [a, b], Tatu Pantsar [b, c]

[a] Department of Internal Medicine VIII, University Hospital Tübingen, Otfried-Müller-Strasse 14, 72076 Tübingen, Germany

[b] School of Pharmacy, University of Eastern Finland, Yliopistonranta 1C, 70210 Kuopio, Finland

[c] Department of Pharmaceutical and Medicinal Chemistry, Institute of Pharmaceutical Sciences, Eberhard Karls Universität Tübingen, Auf der Morgenstelle 8, 72076 Tübingen, Germany

Computational and Structural Biotechnology Journal
Volume 18, 2020, Pages 2687-2698, DOI: 10.1016/j.csbj.2020.09.017

Computational and Structural Biotechnology Journal 18 (2020) 2687–2698

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

The autoinhibited state of MKK4: Phosphorylation, putative dimerization and R134W mutant studied by molecular dynamics simulations

Ekaterina Shevchenko [a], Antti Poso [a,b], Tatu Pantsar [b,c,*]

[a] Dept of Internal Medicine VIII, University Hospital Tübingen, Otfried-Müller-Strasse 14, 72076 Tübingen, Germany
[b] School of Pharmacy, University of Eastern Finland, Yliopistonranta 1C, 70210 Kuopio, Finland
[c] Department of Pharmaceutical and Medicinal Chemistry, Institute of Pharmaceutical Sciences, Eberhard Karls Universität Tübingen, Auf der Morgenstelle 8, 72076 Tübingen, Germany

ARTICLE INFO

Article history:
Received 7 July 2020
Received in revised form 8 September 2020
Accepted 10 September 2020
Available online 20 September 2020

Keywords:
Protein kinases
MAP kinase kinase 4
Molecular dynamics simulation
Protein conformation
Protein dimerization

ABSTRACT

Protein kinases are crucial components of the cell-signalling machinery that orchestrate and convey messages to their downstream targets. Most often, kinases are activated upon a phosphorylation to their activation loop, which will shift the kinase into the active conformation. The Dual specificity mitogen-activated protein kinase kinase 4 (MKK4) exists in a unique conformation in its inactive unphosphorylated state, where its activation segment appears in a stable α-helical conformation. However, the precise role of this unique conformational state of MKK4 is unknown. Here, by all-atom molecular dynamics simulations (MD simulations), we show that this inactive state is unstable as monomer even when unphosphorylated and that the phosphorylation of the activation segment further destabilizes the autoinhibited α-helix. The specific phosphorylation pattern of the activation segment has also a unique influence on MKK4 dynamics. Furthermore, we observed that this specific inactive state is stable as a dimer, which becomes destabilized upon phosphorylation. Finally, we noticed that the most frequent MKK4 mutation observed in cancer, R134W, which role has not been disclosed to date, contributes to the dimer stability. Based on these data we postulate that MKK4 occurs as a dimer in its inactive autoinhibited state, providing an additional layer for its activity regulation.

1. Introduction

One of the key processes in the regulation of complex cellular signalling networks is protein phosphorylation [1]. This phosphorylation is conducted by protein kinases, which transfer a phosphate from an ATP molecule to a target protein, modulating the activity of their downstream target. To maintain the cellular homeostasis, kinase activity is under a strict regulation in the cells. This highly regulated kinase activity is severely distorted in a wide variety of diseases, such as in cancer, resulting in deregulated cellular signalling and disease progression [2,3]. Already several kinase inhibitors are available in the clinical use with the main indication in cancer and oncology, while emerging therapeutic areas include autoimmune and inflammatory diseases [4]. As the therapeutic potential of the protein kinases is enormous, there is a growing need to understand the function and behaviour of these dynamic proteins and their subtle discrepancies in more detail.

Protein kinases not only modulate their target proteins' activity by phosphorylation, but also kinases' activity itself is most often regulated via phosphorylation [5]. Most of the kinases are phosphorylated in their activation loop that transform the kinase from its inactive state to its active state, where the kinase is able to bind and phosphorylate its substrate(s). Interestingly, some kinases have a secondary phosphorylation site in close proximity to the primary site in their activation loop. For instance, two phosphorylation sites are found in ERK2 (THR183/TYR185) with one-residue in between [6], whereas CHK2 (THR383/THR387) exhibits three-residues in between [7,8], while SLK (THR183/SER189) and LOK (THR185/SER191) kinases show a five-residue distance in their activation loop phosphorylation sites [8,9]. Although dual phosphorylation is required, for instance, for full activation of ERK2 [6], the exact result from different phosphorylation patterns of two nearby phosphorylation sites is generally not well understood

* Corresponding author.
E-mail addresses: tatu.pantsar@uni-tuebingen.de, tatu.pantsar@uef.fi (T. Pantsar).

https://doi.org/10.1016/j.csbj.2020.09.017

with most of the kinases. Furthermore, protein kinases may exploit dimerization and/or multimerization as a way to control their kinase activity [10]. For example, kinase dimerization is related to active state with CHK2 [11], DAPK3 [8], RAF kinases and RIPK1 [12–15], whereas inactive dimers exist, *e.g.,* with ligand-free EGFR (inactive symmetric kinase domain dimer) and PDZ-binding kinase (PBK/TOPK) [16–19].

One of the kinases, which has two phosphorylation sites in the close proximity of each other in its activation segment, is the dual specificity mitogen-activated protein kinase kinase 4 (MKK4). MKK4 is encoded by the *MAP2K4* gene, consist of 399 residues and has one isoform with 97.3% identity (410 residues, with an 11 residue insert in the N-terminal part). MKK4 has a serine (SER257) and a threonine (THR261) residues in its activation segment that can be phosphorylated [20]. The phosphorylation of the SER257 is essential for MKK4′s activation, but the THR261 phosphorylation is required for its full activation [21]. The dual specificity name originates from the fact that MKK4 is able to phosphorylate and activate both, c-Jun NH2-terminal kinases (JNKs) and p38 MAP kinases, displaying preference in phosphorylating tyrosine in JNKs activation loop and threonine in the latter [22]. In turn, activated JNKs and p38 MAP kinases are involved in such biological processes like proliferation, apoptosis and cell differentiation [23]. Not much is known about possible dimerization of MKK4.

MKK4 has been suggested to play a crucial role in certain physiological functions and disease development; notably, it has a decisive function in liver regeneration [24]. In tumour development, its role is somewhat controversial, or at least appears to be tissue dependent. Generally, however, MKK4 is considered as a tumour suppressor [25]. Inactivation of MKK4 can exert tumour suppressor activity at both early and late stages of lung tumorigenesis [26]. Also, decreased expression of MKK4 is related to ovarian cancer metastasis and its downregulated phosphorylation levels are associated with a poor prognosis in colorectal cancer patients [27,28]. Conversely, increased activity of MKK4 was shown to promote prostate cancer [29]. In tumours across the primary tissue types, MKK4 is underexpressed in the ovary (84.21%), in the large intestine (23.28%) and in the pancreas (21.23%) [30,31]. Loss of function mutations in the MKK4 gene were reported in approximately 5% of tumours from a variety of tissues [32]. Moreover, it was recently noted as a significantly mutated gene in the colorectal cancer [33]. Based on COSMIC database, 2% of all tumours harbour MKK4 mutation, highlighting two hot-spot mutations: R134W and S184L (COSMIC v.91) [31]. The S184L is located in the ATP-binding site, most likely compromising nucleotide binding, and is an inactivating mutation [34]. The most frequent MKK4 mutation R134W, where an arginine residue is replaced with a tryptophan, is located in the loop between β3-sheet and αC-helix. However, no functional data of the R134W mutation exist to date and its effect on MKK4′s function is unclear.

Currently three MKK4 structures are available in the RCSB Protein Data Bank (PDB IDs: 3alo [35], 3aln [35], 3vut [36]) (SI Fig. S1). All of the structures represent the inactive unphosphorylated form of the protein, where 3vut is the apo-structure and 3aln and 3alo are co-crystallized with the non-hydrolysable ATP analogue, AMP-PNP. In addition, 3alo (resolution 2.6 Å, R_{free} 0.271) is crystallized with a short p38 peptide that is bound on top of the N-terminal lobe. This structure is particularly interesting, as a part of its activation segment (ILE250–ARG264) appears in an ordered autoinhibited conformation, forming a long α-helix that protrudes from the kinase (SI Fig. S1). The other structures, 3aln (resolution 2.3 Å, R_{free} 0.378) and 3vut (resolution 3.5 Å, R_{free} 0.407), are lower quality with two disordered regions: ASP263–GLY283 and GLN316–VAL320. Furthermore, 3vut is missing its C-terminal part after ALA374.

Not much is known of MKK4′s conformational dynamics. Recently, MKK4 was studied by small angle X-ray scattering (SAXS) with Ensemble Optimization Method (EOM) [37], revealing an ensemble of conformations in solution for all three structures. Only one study has been reported to date, where MD simulations were applied to investigate MKK4 [38]. In the study, a homology model based on 3aln structure was used and a single 400 ns simulation for wild-type MKK4 and G265D mutant was conducted. Overall, there is a lack of knowledge in the MKK4 dynamics, especially related to the specific autoinhibited state.

Here, we conducted microsecond timescale (a total of 40 μs) all-atom MD simulations to study the dynamics of autoinhibited MKK4. We studied the role and influence of all phosphorylation patterns in the activation segment. Finally, we investigated putative dimerization of MKK4 and the effect of phosphorylation to the stability of the homodimer, including the most common MKK4 mutant R134W to these simulations. Our results highlight the instability of the autoinhibited state as a monomer and suggest that it may exist as a stable dimer only when unphosphorylated.

2. Methods

2.1. MD simulations of monomer MKK4

For the simulations we used the autoinhibited MKK4 structure 3alo [35]. System preparation was done with Maestro 2017-2 (Schrödinger, LLC, New York, NY) with OPLS3 force field [39]. The disordered residues of the activation segment in the 3alo structure (residues SER277–GLY283) were added with Maestro's *cross link proteins* tool. First, the PRO277 residue was deleted, the rotamer of TYR284 was changed (to prevent the clash) and the sequence PRO277–GLY283 (PSASRQG) was added to link the chain. The terminal PRO389 was mutated to ALA as a terminal PRO residue distorts the dynamics (the C-terminal part ALA390–ASP399 that is missing from the structure was omitted). The systems were prepared with Protein Preparation Wizard with default settings (Cap-termini) [40]. We left out the disordered N-terminal part of MKK4 (residues 1–94), which is suggested to play a role in substrate recognition [41] and therefore was not considered critical for our simulations. For the phosphorylated systems, SER257 and THR261 were changed to the corresponding phosphoresidues pSER257 and pTHR261.

Desmond MD engine was used for the simulations [42]. Systems were solvated in a cubic box (minimum distance of 13 Å to the edges from the protein), and the total net charge was neutralized using Na^+-ions. The water was described with TIP3P model [43]. The final systems comprised ~59 k atoms. All simulations were run in NpT ensemble (T = 310 K, Nosé-Hoover method; p = 1.013 25 bar, Martyna-Tobias-Klein method) with default Desmond settings. RESPA integrator with 2 fs, 2 fs and 6 fs timesteps were used for bonded, near and far, respectively. The default value of 9 Å was used for Coulombic cut-off. All systems were relaxed using the default Desmond protocol prior to the production simulations. To obtain better sampling and to remove the potential initial bias in the systems, the ARG134 rotamer was changed to different ones for each replica (according to the rotamer library) as it was pointing towards the phosphoresidues in its initial configuration (in the unphosphorylated crystal structure there is no clear electron density for the side-chain of ARG134). Five replicas of production simulations were carried out for each system for 1000 ns (5 × 4 × 1000 ns = 20 μs). For each replica, a random seed was used. All replica simulations were run using OPLS3 force field, except three individual replicas were run with updated OPLS3e force field (using Maestro 2018-2) [44].

2.2. MD simulations of dimer MKK4

The dimer assembly (lowest energy assembly) was obtained with PDBePISA server (v.1.52) [45]. The dimer complexes were prepared as monomers, except the force field OPLS3e [44] was used (with Schrödinger Maestro 2018-2 and 2019-3). For the dimer simulations the cubic box was set to 15 Å from the protein. Final systems comprised ~120 k atoms. Five replicas of production simulations were carried out with same settings as mentioned above. Each system was simulated for 1000 ns (5 × 4 × 1000 ns = 20 μs). For each replica a random seed was used.

2.3. RMSD and RMSF

Root-mean-square fluctuations (RMSFs) of protein backbone and Root-mean-square deviations (RMSDs) of Cα-atoms were calculated using Maestro Simulation interaction analysis tool (Schrödinger, LLC, New York, NY). RMSDs of residue intervals used for angle calculations were conducted with MDAnalysis.analysis.rms module [46] of MDAnalysis library [47,48] for Python 3.7.

2.4. Principal component analysis (PCA)

The PCA was conducted with GROMACS (version 2019) covariance analysis tools (gmx covar, gmx anaeig) [49]. The PCA was conducted for all backbone atoms, excluding the residues SER257 and THR261 which differed among the systems (as a single PCA was conducted for all systems). For the further analysis we included the PCs that displayed >9% individual contribution: PC1 20.0%; PC2 11.8%; PC3 9.4% and PC4 9.1% (all combined 50.3%). The individual PC movements were illustrated with PyMOL-script Modevectors [50].

2.5. Secondary structure analysis

Secondary structure analysis was conducted with Maestro Simulation interaction analysis tool (Schrödinger, LLC, New York, NY). The percentage of secondary structure elements (%SSE) throughout the simulations was calculated with Python 3.7.

2.6. Angle calculations

Angle calculations between subunits of MKK4 were conducted using open-source MDAnalysis library for Python 3.7 [47,48]. For both monomer and dimer systems, first 250 ns were excluded from the analysis based on the system stabilization (SI Fig. S2–4). Sides of the angle are formed by: ILE250–ALA264 (ASH), GLU139–ARG154 (αC-Helix), the vertex is represented by GLU179–SER182 (HR). Calculated centre of geometry (cog) was used as an apex point for the angle calculations. The data from all frames for each system replica was combined and used for calculation of average (mean) and standard deviation for both mono- and dimer system of MKK4. In dimer system calculations were performed individually for both subunits A and B. Reference angles were calculated from frame 0 for dimer systems that are 14.9° and 17.9° for subunit A and B, respectively. For monomer systems 14.9° (subunit A) was chosen as reference value, that corresponds to the autoinhibited crystal conformation.

In order to confirm that the selection of particular ASH residues is not critical for angle value, we performed validation by testing different intervals. Residues ILE250–ALA264 (ASH), that are forming one side of the angle, were switched to ILE250–ALA259 and VAL255–ALA264 (SI Fig. S5). This way we shifted the selected residue interval of ASH by five residues back and forth. When compared to the switched ones, ILE250–ALA259 interval showed deviation of - ~1–2° from reference and VAL255–ALA264 of + ~1–2°. As a result, the selection of the residues itself do not play a critical role on the angle value as the two rays lie in a plane, but this plane does not have to be Euclidian one.

2.7. Distance calculations

The distances between subunits of MKK4 were calculated using GROMACS (version 2019) gmx distance tool [49]. Points for calculation were defined by cog of selected residue intervals. Following residues of each subunit were used for cog calculations: LEU102–GLN126 (N-lobe section), ILE250–ALA264 (ASH), VAL286–THR302 (αF-helix). The choice of this particular residues was based on that each of these intervals represent a specific region within MKK4 interface. Thus, N-terminal lobe, includes allosteric region, that predominantly bound the p38a peptide [35]; activation segment includes two phosphorylation sites of MKK4 (SER257, THR261), and αF-helix represents stable helix in lower part of C-lobe. Consequently, distance of the cog for chosen intervals represent three points within the dimer interface: upper, middle and lower part, where the middle part distance is perpendicular to others.

Distance calculations within the N-lobe MKK4 were done as above, using the following residue intervals of each subunit: VAL116–VAL120; VAL120–ILE127; PHE164–LEU168; ALA111–GLY114.

2.8. Interaction analysis

The interaction networks analysis related to Fig. 4 was conducted with Maestro (Schrödinger, LLC, New York, NY) scripts, analyse_trajectory_ppi.py for salt-bridges and trajectory_asl_monitor.py for hydrophobic interactions. Default salt-bridge interaction cut-off of 4.0 Å was used. For hydrophobic interaction definition, we used a sidechain atom distance below 2.5 Å. Residue 134 interaction analysis related to Fig. 7 was conducted with simulation interaction analysis tool of Maestro.

2.9. Data visualization

Results were plotted with Seaborn library for Python 3.7 [51]. Protein structures were visualized with PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.) Graphical representations of figures were arranged using Adobe Illustrator©. Supplementary movies were generated with PyMOL.

3. Results

3.1. Conformational dynamics of monomeric autoinhibited MKK4 with different phosphorylation states

3.1.1. MKK4's activation segment and C-lobe loop are highly dynamic

First, we investigated dynamics of autoinhibited MKK4 with all possible phosphorylation patterns in its activation segment (a total of 20 μs MD simulations). Four different phosphorylation states of MKK4 in this region are possible: unphosphorylated (Up), monophosphorylated at SER257 (pS257) or at THR261 (pT261) and double phosphorylated (pS257 + pT261) (Fig. 1A). Overall, these systems exhibit similar root-mean-square fluctuation (RMSF) values (Fig. 1B). All systems display the highest RMSF-values at two specific loop-regions, particularly among the residues CYS266–ARG281 of the activation segment and residues ASP315–LYS322 in a loop of the C-lobe. Generally, kinases have αG-helix located on this C-lobe loop location [5], but it appears as a loop in available MKK4 structures. We next compared how the observed dynamics based on RMSF is in agreement with observed B-factors and disorder of the MKK4 crystal structure.

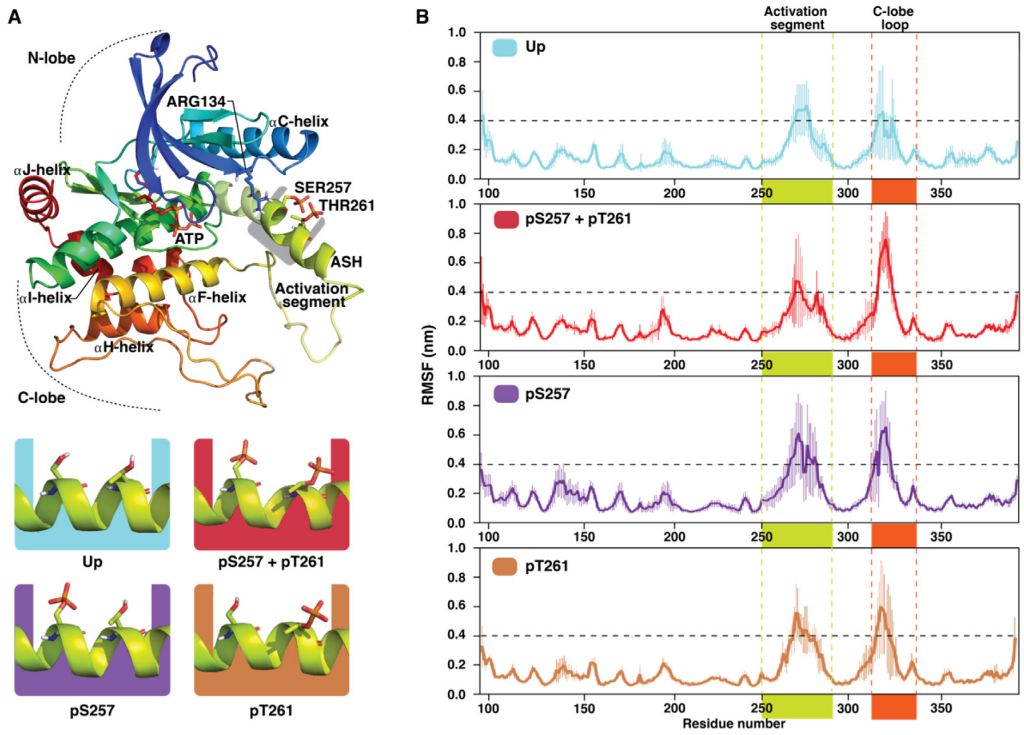


Fig. 1. Structure and dynamics of MKK4 with different phosphorylation states in its autoinhibited state. (A) Conformation of MKK4 in its autoinhibited state (PDB ID: 3alo). In its autoinhibited state, part of the activation (loop) segment is forming a helical structure; activation segment helix (ASH). Different phosphorylation patterns in its activation segment are colour coded throughout this article as: Up, teal; pS257 + pT261, red; pS257, violet; pT261, brown. (B) Root-mean-square fluctuation (RMSF) of protein backbone. Average of five replicas is shown with standard deviation (thin vertical lines). Highlighted regions, represented in same colours as in A, indicate activation segment (ILE250–SER292), green; C-lobe loop (PRO308–PHE340), dark orange. RMSF value of 0.4 nm is indicated with the horizontal dashed black line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Indeed, disorder (residues 278–283) or high B-factor values are observed in these regions (SI Fig. S6). Although generally a similar trend in RMSF values is observed among all systems, slightly different RMSF patterns are evident (see details in SI Table S1). For instance, higher RMSF (>0.1 nm compared to Up) of SER251–ALA259 is observed only in monophosphorylated pS257. This indicates an individual change in dynamics related to particular phosphorylation state of MKK4.

3.1.2. Phosphorylation defines MKK4 dynamics

To gain further insights into the protein dynamics and possible differences among the systems, we conducted principal component analysis (PCA). According to the PCA phosphorylation affects to the MKK4 dynamics, as each system displays individual profile in their PC (principal component) score plots (Fig. 2A). In PC1 the largest movements appear in activation segment and C-lobe loop, which also displayed the highest RMSF values (Fig. 2B, SI Movie M1). Moreover, considerable movement is observed in other regions of the protein: αC-helix and in C-lobe helices αF, αH, αI and αJ. Clearly lower values of PC1 are observed with the pS257 + pT261 in comparison to other systems. This indicates that the protruding movement of the activation segment (elongation of the ASH towards solvent *i.e.* original crystal structure conformation) is clearly disfavoured with the double phosphorylated system. In PC2, movements of activation segment and C-lobe loop are dominating (Fig. 2C, SI Movie M2). Both are folding towards the centre of the kinase. With this component the monophosphorylated pS257 system shows the highest values (Fig. 2A). With PC3 and PC4 (SI Fig. S7, SI Movies M3–4), pT261 is showing a unique subpopulation with high PC3 and low PC4 values. Overall, the areas displaying the highest contribution to the PCs occur in the regions that are responsible of the activation, substrate binding and regulatory actions in the kinase [52].

3.1.3. Activation segment helix movement is defined by its phosphorylation

In addition to the differences among the systems observed by RMSF and PCA, a system-specific movement is evident even by visual examination of the simulation trajectories. The activation segment appears to display a system specific shift that clearly deviates from the crystal conformation. This is also demonstrated by the resulting end conformations of the simulations (SI Fig. S8, SI Movie M5). Moreover, the intactness of the α-helical secondary structure in the activation segment deviates among the systems (SI Fig. S9).

To validate these visual observations, we further evaluated this movement by angle calculations (Fig. 3; SI Fig. S4). This allowed us to describe the activation segment movement in easily interpretable geometrical values. For these angle calculations, we applied centre of geometry (cog) of selected protein segments: activation segment helix (ASH; ILE250–ALA264), αC-helix (GLU139–ARG154) and hinge region (HR; GLU179–SER182). As hinge region and αC-Helix are relatively stable elements of the kinase, this angle-change provides information about the movement of ASH in respect to the protein and filters out their synchronized fluctuations. For instance, in monophosphorylated pS257 system ASH is constantly fluctuating towards the C-lobe, away from the αC-helix. This movement is directly related to the angle:

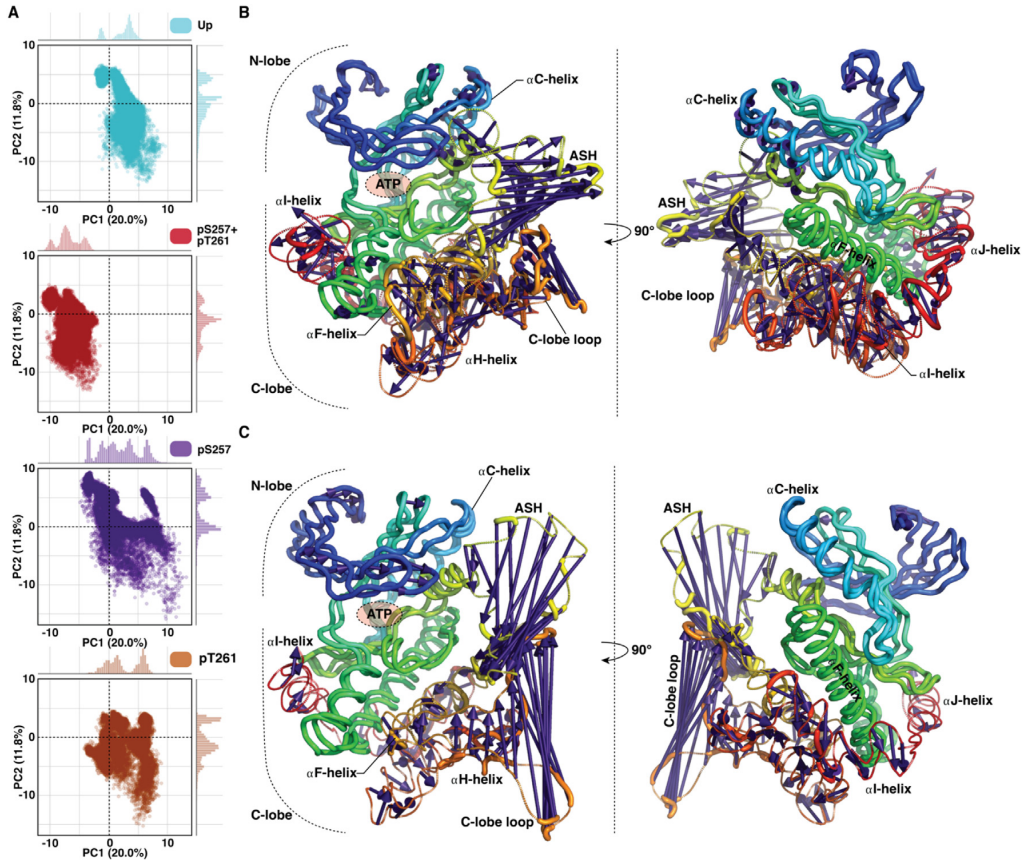


Fig. 2. MKK4 shows individual conformational dynamics with different phosphorylation states. (A) Principal component analysis (PCA) score plot of PC1 and PC2. A single PCA was conducted for all systems, ensuring the comparability of the score-plots. (B) The extreme movements of PC1. (C) The extreme movements of PC2. In B and C, protein is illustrated with rainbow colour (as in Fig. 1A) and the purple arrows indicate the extreme movements related to each principal component. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the more down along the Y-axis ASH moves, the wider the angle (Fig. 3A). Remarkably, none of the systems stay in the autoinhibited crystal conformation (reference angle 14.9°) (Fig. 3B). The highest variation of this angle exists with pS257 (~37–56°), which reflects to the visually observed high fluctuation of the ASH with this system. Interestingly, the double phosphorylation appears to fix MKK4 in a more specific configuration, as clearly less variation in the angle is observed with pS257 + pT261 system. To note, the beginning of ASH is relatively stable in all systems (Fig. 1B); therefore, even small angle variations indicate a considerable movement in the end part of ASH. We anticipate that the higher overall angle values of pS257 system are indeed related to the increased movement of the SER251–ALA259 residues, for which it shows higher RMSF values compared to other systems. These observations highlight clear influence of phosphorylation states on MKK4 dynamics, especially on the ASH region.

3.1.4. MKK4 interaction networks are phosphorylation state specific

As phosphorylation status clearly affects MKK4 dynamics, we next conducted interaction network analysis to determine the potential changes in the protein's interaction patterns. First, we analysed the salt-bridges and their frequencies in the close proximity of the ASH of each individual system (Fig. 4A). Especially in the monophosphorylated system pS257, LYS260 is fixed to the phosphorylated residue (90.4%). In the pS257 + pT261 system, LYS260 seems to be balancing between both of the phosphoresidues; still the major interaction occurring with pS257 (69.3% vs. 21.6%). The monophosphorylated pT261 displays the LYS260–pTHR261 interaction with 43.9% frequency. Interestingly, with pT261 system LYS260 displays 29.7% interaction to ASP138 that is located in the small loop connected to the αC-helix. This interaction is also present with Up (14.8%), but almost missing in pS257 (0.05%) and pS257 + pT261 (0.1%). Contact between LYS260 and ASP138 may explain the lower angle values observed in Up and pT261 systems (Fig. 3B), as it would fix and thus prevent the ASH bending, away from the αC-helix. ARG134, which is located in the same loop with ASP138, interacts with the pTHR261 (21.4% and 38.8% in pT261 and pS257 + pT261 systems, respectively). Interaction between ARG134 and pSER257 is almost nonexisting with double phosphorylated system (0.1%), whereas with monophosphorylated system (pS257) it occurs with 15.7%. ARG274 displays interaction with pSER257 only in the double phosphorylated system (18.9%). In addition, ARG228 interacts with the pSER257 in monophosphorylated pS257 system (11.9%), but not in double phosphorylated system (0.4%). Salt-bridge interaction frequencies are also altered among the systems with other residues than phosphoresidues in the ASH.

Systems also display differences in their hydrophobic interactions (Fig. 4B). A specific type of extensive bending of ASH towards G-loop occurs in pS257 + pT261 system, which can be traced to the

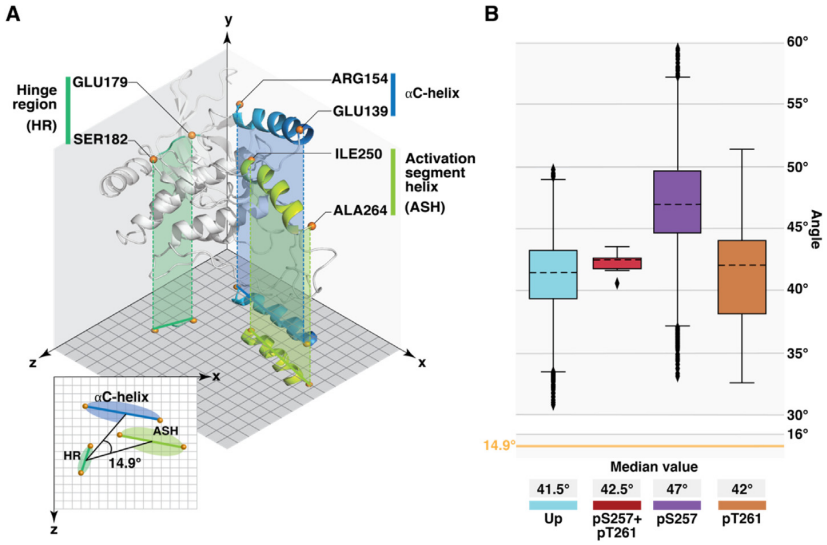


Fig. 3. Activation segment helix movement depends on its phosphorylation status. (A) Plane projection of residue intervals used for angle calculations. The change in the angle value describes a relative movement of these elements in respect to others. Apex points of the angle are represented by centre of geometry (cog) for each of the selected residue intervals throughout simulation time. *Cog* is visualised on the plane projection with semi-transparent ovals of the same colours as displayed residue intervals. Reference angle value (14.9°) corresponds to the autoinhibited crystal conformation. (B) Boxplot representation of observed angle values in MKK4 systems with different phosphorylation patterns (excluding the data of the first 250 ns of each simulation). The dashed black line in the box represents the median. Box displays the quartiles of the dataset (25–75%) and whiskers the rest of the data within 1.5 times of the interquartile range (IQR). Outliers are indicated with black diamonds. Reference angle value (14.9°) of the autoinhibited state is indicated with yellow line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

formation of hydrophobic interaction between ALA259–ALA112 (9.2%). The more stable conformations of the beginning of ASH appearing with Up and pT261 are manifested by the elevated interaction frequencies of ILE133–LEU254. Furthermore, pS257 system displays a clear shift in its interaction preferences in its αC-helix associated hydrophobic residues (orange spheres) in comparison to other systems.

3.2. Putative MKK4 homodimer and R134W mutation

3.2.1. MKK4 homodimer

As the autoinhibited MKK4 structure exhibits a long protruding activation segment, which appeared unstable in the monomer simulations, we evaluated the crystal assemblies of 3alo utilizing the PDBePISA server [45]. Indeed, the lowest energy assembly of the structure is identified as a dimer (ΔG^{int} = −53.5 kcal/mol; ΔG^{diss} = 10.3 kcal/mol) and not as a monomer (ΔG^{int} = −23.0 kcal/mol; ΔG^{diss} = 5.2 kcal/mol). In this homodimer assembly the apical activation segment region is stabilized and buried within the dimer interface (Fig. 5A). Therefore, we considered the possibility of MKK4 existing as a dimer in its autoinhibited state and decided to investigate this dimer and its stability by MD simulations. We conducted simulations of dimer MKK4 (a total of 20 μs) with different phosphorylation patterns: unphosphorylated (Up-DIM), double phosphorylated at SER257 and THR261 in one subunit (ppSA-DIM), double phosphorylated in both subunits (ppSA/ppSB-DIM) (Fig. 5A). In addition, as we noticed that the most frequent mutation of MKK4, R134W, is located in the dimer interface, we included unphosphorylated system accompanied with this mutation in both of the subunits (Up(R134W)-DIM).

First, we investigated the overall dynamics of the protein by RMSF. Compared to monomer systems, homodimer MKK4 displays significantly lower RMSF-values in its activation segment (Fig. 5B). This is perhaps not surprising, as the activation segment movements are hindered in the dimer by the other subunit. The C-lobe loop displays more comparable RMSF values with the monomer systems, although those are also generally lower.

3.2.2. MKK4 homodimer is stable only when unphosphorylated

Next, we evaluated the stability of the dimer complexes via a distance analysis. Three regions of MKK4 were selected for distance calculation with the following residue intervals: LEU102–GLN126 (N-lobe section), ILE250–ALA264 (ASH), VAL286–THR302 (αF-helix) (Fig. 6A). To note, distance between ASH is perpendicular compared to the other two selected intervals. Distances between αF-helices, which are buried deeply in the dimer interface in the C-lobe, remain close to the reference value of the crystal structure (Fig. 6B; SI Table S2). The ASH distance values indicate even tighter packing in the middle compared to the crystal. Here the shortest distance among wild-type dimer systems is displayed by Up-DIM, where with the unphosphorylated mutant it is even shorter. Overall, N-lobe from different subunits tends to come closer to each other with increase of simulation time among all wild-type systems. Remarkably, with R134W this is not the case; even increased distance compared to the crystal structure is observed. In the N-lobe the unphosphorylated mutant displays a striking difference compared to other systems (0.93 nm compared to Up-DIM), having the longest distance (2.90 nm) between N-lobes.

Next, we evaluated the relative movement of function-related structural elements in the dimer using angle calculations as described for monomer systems (Fig. 6C). Strikingly, here the difference between the angles of unphosphorylated and phosphorylated dimer is evident. Median value of Up-DIM is close to the reference value of the crystal ~14° and for ppSA/ppSB-DIM this is clearly higher ~26°. Thus, the relative position of these structural elements is well maintained only in the unphosphorylated dimer, suggesting that the phosphorylation leads to unstable dimer. Interestingly, both the highest value and variation of the angle occur

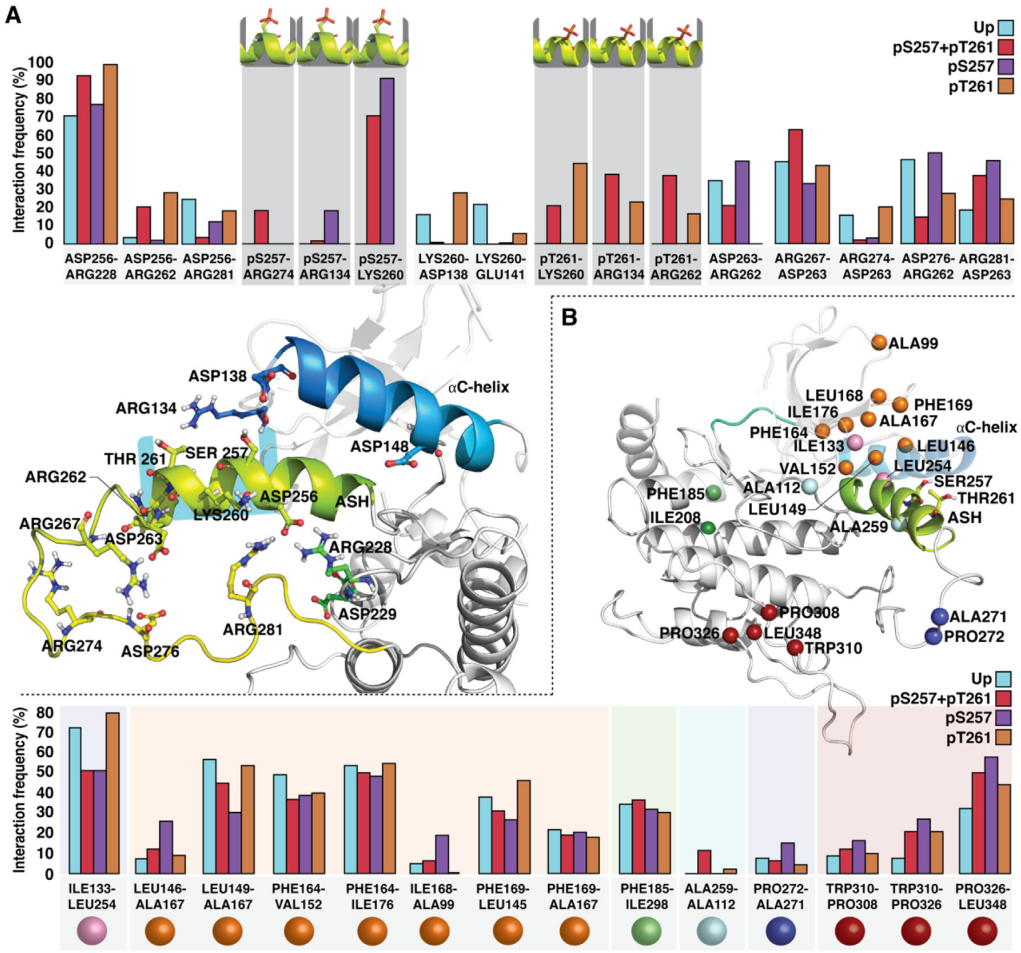


Fig. 4. Differences in protein interaction networks among different phosphorylation states. (A) Salt-bridges of the activation segment residues ILE250–ALA264 with their interaction frequencies (%). Only the salt-bridges with > 20% differences in their interaction frequencies among systems are shown. (B) Selected hydrophobic interactions of MKK4 and their interaction frequencies among different systems. The locations of the Cα-atoms of the hydrophobic residues are shown in spheres, which are coloured according to different hydrophobic clusters that exhibit linked interactions.

with Up(R134W)-DIM system, which clearly indicate that the mutation has influence on the unphosphorylated dimer stability and behaviour. Also, only the phosphorylation appears to distort the alpha-helical secondary structure of ASH, where in the R134W mutant this is stable as in the Up-DIM (SI Fig. S10). Similar root-mean-square deviation (RMSD) values are also observed for these structural elements in both subunits of the dimer, indicating synchronized movements between the subunits (SI Fig. S11).

The distance analysis did not reveal distortion in the dimer with the phosphorylated wild-type systems, whereas the angle calculations suggested that it clearly exists. Average end conformations of the systems reveal that in the systems where the ASH is unstable (based on angles), the dimer subunits begin to twist related to each other (SI Movies M6–8; SI Fig. S12). This explains the observation that even if the distances are comparable among the systems, the angle values reveal the distortion of the dimer. Therefore, simultaneous usage of both angle and distance calculations can provide more reliable information about protein movement. Each of these methods describe the motion along different geometric plane within the protein interface: distance calculations describe translational movements, whereas angle calculations consider rotations.

3.2.3. R134W mutant distort the dimer interactions in the N-lobe

Based on the distance and angle analysis, R134W clearly affects to dynamics of the unphosphorylated dimer. This occurs especially on the N-lobe interface where R134W is located in the loop between β3-sheet and αC-helix (Fig. 5A; SI Movie M8). Therefore, we decided to study in more detail this section of MKK4 and its interactions. In the dimer, the most frequent interaction of ARG134 in wild-type MKK4 occurs with TYR113 of the same subunit, via a water bridge (Fig. 7A, 7C). This water bridge occurs in all replicas of wild-type systems with ~56% on average. Interestingly, corresponding interaction between TRP134 and backbone amino group of TYR113 of the same subunit is direct, unlike in wild-type systems where the interaction is water-mediated (Fig. 7A). Overall, we see a clear difference in a number of interactions with the R134W mutant. In four out of five replicas, the mutated residue connects the subunits TRP134(SA)–TRP134(SB) via π-π interaction (~50% frequency) and/or hydrogen bonding (17%) (Fig. 7A). In the wild-type systems, ARG134s do not show interactions between each other and are located far apart (Fig. 7B). Moreover, interaction between TRP134 from subunit B and TYR113 from subunit A occurs in three mutant replicas

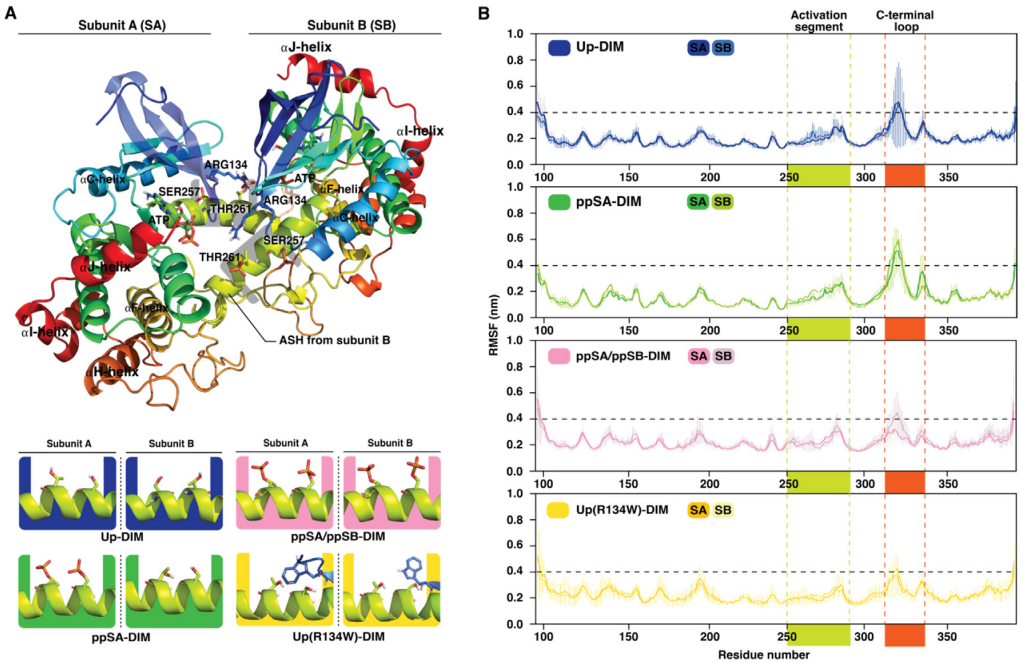


Fig. 5. Putative MKK4 homodimer and its dynamics. (A) Assembly of autoinhibited MKK4 homodimer. Phosphorylation patterns in simulated MKK4 dimer systems. Systems are coloured as: Up-DIM, blue; ppSA-DIM, green; ppSA/ppSB-DIM, pink; Up(R134W)-DIM, yellow. (B) Root-mean-square fluctuation (RMSF) of protein backbone. Average of five replicas is shown with standard deviation (thin vertical lines). Highlighted regions indicate activation segment (ILE250–SER292), green; C-lobe loop (PRO308– PHE340), dark orange. RMSF value of 0.4 nm is indicated with the horizontal dashed black line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(Fig. 7C), providing an additional link between the subunits. In double phosphorylated monomer MKK4, ARG134 forms a salt-bridge mainly with pTHR261 (Fig. 4A), and this interaction is also frequently observed with the phosphorylated dimers.

To get more comprehensive picture of the relative movements in N-lobe, we conducted additional distance calculations with four chosen intervals (Fig. 7D). The mutant system exhibits a notable difference especially in the upper part of N-lobe (VAL120–ILE127) compared to other systems (Fig. 7E; SI Table S3). Overall, the R134W mutation disrupts local contacts in the dimer interface and clearly modifies the N-lobe association between the unphosphorylated subunits.

4. Discussion

Here we investigated conformational dynamics of autoinhibited MKK4 with different phosphorylation states, putative dimer stability and the effect of R134W mutation by MD simulations. The two other publicly available crystal structures of MKK4 have a disordered activation segment, where only the 3alo exists in ordered autoinhibited conformation, forming a long α-helix with unphosphorylated SER257 and THR261 (SI Fig. S1). Based on our simulations, this autoinhibited conformation of unphosphorylated MKK4 appears unstable as a monomer.

Biggest movements in the MKK4 monomer occur in the activation segment helix and in the C-lobe loop between the αF- and αH-helices. This high flexibility observed in simulations perhaps reflect to the fact why high-quality MKK4 crystal structures are unavailable. Moreover, the results here are in agreement with the recent solution structure analysis of MKK4 by SAXS [37]. Interestingly, the movement pattern of these most dynamic regions of MKK4 appears phosphorylation dependent as different phosphorylation states exhibit unique effect on MKK4 dynamics. For instance, monophosphorylated pS257 displays high fluctuation of ASH, while double phosphorylated pS257 + pT261 appears to fix ASH in a relatively stable configuration. These dynamic differences of individual phosphorylation states may reflect to the observed MKK4 activity levels (*e.g.* double phosphorylation is required for full activation) [22]. To note, there is currently no data available of the relevance and biological activity of the monophosphorylated pT261 MKK4; therefore, it may represent an artificial system.

Although the autoinhibited MKK4 conformation is unstable as monomer even when unphosphorylated, our simulation results support the possibility that this unique inactive autoinhibited state is stable as a dimer. This observation is in agreement with the results obtained from the PDBePISA assembly evaluation [45]. Our simulation results indicate that this dimer configuration is stable only when these ASH residues are unphosphorylated. In these microsecond timescale simulations, the distortion of the dimer configuration with phosphorylated or mutated MKK4 is clearly demonstrated by the angle calculations. Here a distance analysis was unable to capture the distortion. This is probably due to the fact that a full dissociation of the dimer with a clear translational movement is not expected to occur within this timescale. Interestingly, a somewhat similar type of dimerization is observed with MEK1 and MEK2 in complex with inhibitors, where allosteric inhibitors occupy partially the corresponding autoinhibited ASH position in MKK4 (SI Fig. S13) [53].

The origin of the autoinhibited MKK4 configuration is from the co-crystallized complex with ANP and p38α peptide [37]. This fact leads us to speculate the potential existence of a N-lobe binding scaffold protein which stabilizes the inactive MKK4 dimer. This scaffold protein would provide an additional regulation mechanism for MKK4s activity. A putative candidate for this would be

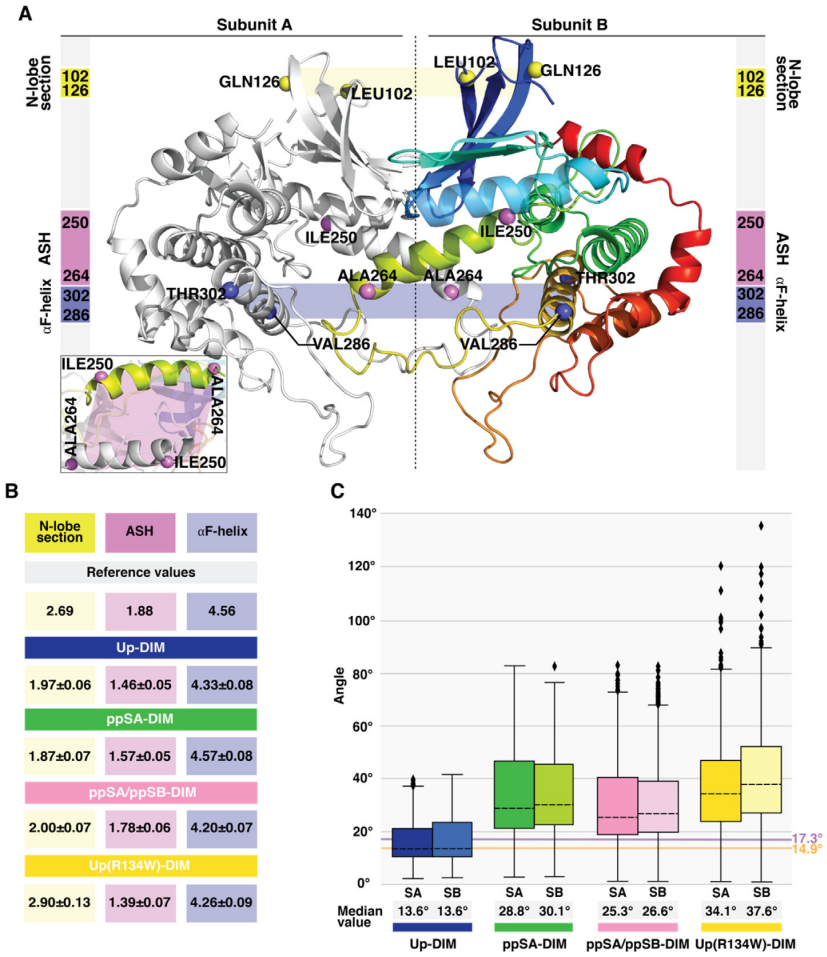


Fig. 6. Conformational stability of MKK4 homodimer. (A) Selected residue intervals and their locations for distance calculation between the subunits: LEU102–GLN126 (N-lobe section), ILE250–ALA264 (ASH) and VAL286–THR302 (αF-helix). (B) Average distance (nm) with standard deviation of the selected residue intervals between subunit A and B in MKK4 dimer systems. First 250 ns of the simulations were excluded from the analysis. See more details in SI Table S2. (C) Boxplot representation of αC-helix–HR–ASH angle values in dimer MKK4. Reference values of the corresponding angles are illustrated with orange line for subunit A (14.9°) and violet line for subunit B (17.34°). The black dashed horizontal line in the box represents the median. Box displays the quartiles of the dataset (25–75%) and whiskers the rest of the data within 1.5 times of the IQR. Outliers are indicated with black diamonds. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Scaffold protein C-Jun N-terminal kinase-interacting protein 4 (JIP4), which is known to interact with MKK4 and an increased association between these two proteins suppresses MKK4 phosphorylation [54]. This hypothesis of inactive MKK4 dimerization and the existence of a scaffold protein should be confirmed in further studies.

The most frequent MKK4 mutation in cancer, R134W, which role has not been disclosed to date, affects MKK4 dynamics on the putative inactive dimer interface. This mutation leads to dramatic alterations in the N-lobe interactions, demonstrated by the changes in frequency and nature of the interactions within the residue 134 in the N-lobe. Based on this, R134W may be an activating mutation via distorting the autoinhibited dimer state of MKK4. On the other hand, the timescale of the simulations is not sufficient to disclose a full disruption of the dimer complex. Therefore, this alteration with its shifted N-lobe interactions may even lead to enhanced stability of the inactive dimer regardless of the putative scaffold protein. This would mean that R134W is an inactivating MKK4 mutation. Overall, additional experimental evidence is required to disclose the role of this cancer associated mutation.

Our results demonstrate that the autoinhibited state of MKK4 is unstable as monomer and stable as dimer. Moreover, different phosphorylation patterns and the R134W mutation have all individual consequences for MKK4 dynamics. Better understanding of conformational changes and dimerization of protein kinases, occurring either due to phosphorylation (activation) processes or oncogenic mutations, is needed to provide comprehensive framework for disease causality. This will ensure and support a rational inhibitor design in a disease specific context related to aberrantly behaving protein kinases, which are currently the main target class in ongoing projects of the pharmaceutical industry [55].

Author contributions

T.P. performed simulations and designed the research. E.S, A.P. and T.P. analysed the data. Original draft was written by E.S. and T.P. All authors participated in writing the final manuscript. Figures were prepared by E.S and T.P.

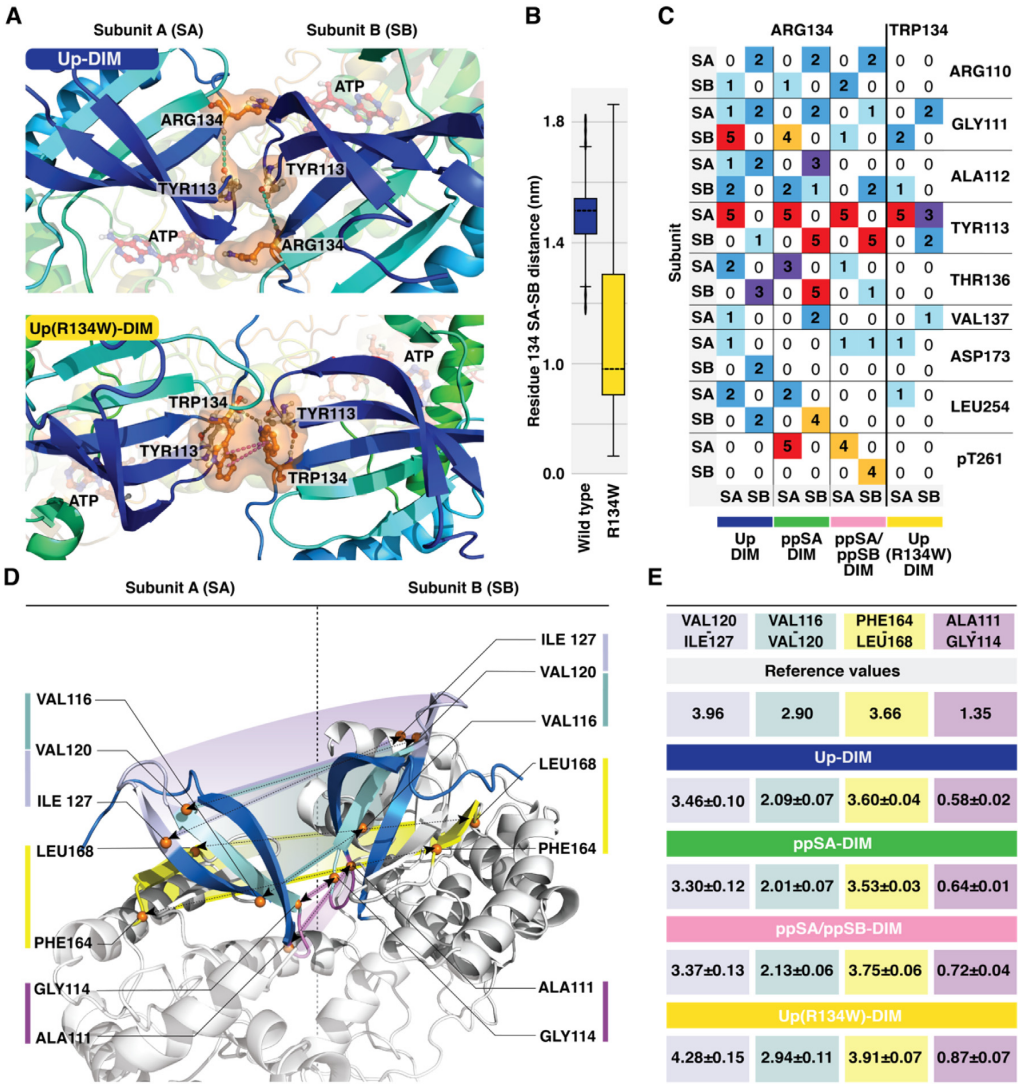


Fig. 7. Impact of phosphorylation and R134W mutation on N-lobe interactions. (A) A representative snapshot of the top view of N-lobe, displaying difference in contact pattern between Up-DIM and Up(R134W)-DIM at 500 ns. Water-bridged interactions are indicated with dashed cyan line; hydrogen bonds with dashed grey line; π-π interaction with dashed violet line. (B) Boxplot representation of distance between the residue 134 Cα-carbons of different subunits. Simulation time of 250–1000 ns from all replicas was included in the distance analysis. The black dashed horizontal line in the box represents the median values: 1.5 nm in wild-type and 1.08 nm in R134W. Box displays the quartiles of the dataset (25–75%) and whiskers the rest of the data within 1.5 times IQR. Outliers are indicated with black diamonds. (C) Heatmap of contacts with the residue R134/W134. Numbers (0–5) represent the number of individual replicas where the interaction occurs >30% of the simulation time. (D) Selected intervals for distance analysis within dimer N-lobe interface. (E) Average distances between selected N-lobe intervals (nm) with standard deviation (SD). Reference values are calculated for 3alo crystal structure. See more details in SI Table S3. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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

We would like to thank the CSC - IT Center for Science Ltd. (Espoo, Finland) for computational resources. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 839230 (T.P.). T.P. acknowledges the Orion Research Foundation sr for financial support. A.P. acknowledges the Academy of Finland (grant number 276509) for financial support. Mariana González Medina is acknowledged for help in data visualization for PCA.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2020.09.017. Full length raw trajectories for all simulations are freely available at https://doi.org/10.5281/zenodo.3933035.

References

[1] Cohen P. The origins of protein phosphorylation. Nat Cell Biol 2002;4(5):E127–30.
[2] Bhullar KS, Lagarón NO, McGowan EM, Parmar I, Jha A, Hubbard BP, Rupasinghe HPV. Kinase-targeted cancer therapies: progress, challenges and future directions. Mol Cancer 2018;17(1). https://doi.org/10.1186/s12943-018-0804-2.
[3] Sever R, Brugge JS. Signal Transduction in Cancer. Cold Spring Harb Perspect Med 2015;5:a006098. https://doi.org/10.1101/cshperspect.a006098
[4] Ferguson FM, Gray NS. Kinase inhibitors: the road ahead. Nat Rev Drug Discov 2018;17(5):353–77.
[5] Taylor SS, Keshwani MM, Steichen JM, Kornev AP. Evolution of the eukaryotic protein kinases as dynamic molecular switches. Phil Trans R Soc B 2012;367(1602):2517–28.
[6] Prowse CN, Lew J. Mechanism of activation of ERK2 by dual phosphorylation. J Biol Chem 2001;276(1):99–103.
[7] Guo X, Ward MD, Tiedebohl JB, Oden YM, Nyalwidhe JO, Semmes OJ. Interdependent phosphorylation within the kinase domain T-loop regulates CHK2 activity. J Biol Chem 2010;285(43):33348–57.
[8] Pike ACW, Rellos P, Niesen FH, Turnbull A, Oliver AW, Parker SA, Turk BE, Pearl LH, Knapp S. Activation segment dimerization: a mechanism for kinase autophosphorylation of non-consensus sites. EMBO J 2008;27(4):704–14.
[9] Luhovy AY, Jaberi A, Papillon J, Guillemette J, Cybulsky AV. Regulation of the Ste20-like kinase, SLK: involvement of activation segment phosphorylation. J Biol Chem 2012;287(8):5446–58.
[10] Lavoie H, Li JJ, Thevakumaran N, Therrien M, Sicheri F. Dimerization-induced allostery in protein kinase regulation. Trends Biochem Sci 2014;39(10):475–86.
[11] Oliver AW, Knapp S, Pearl LH. Activation segment exchange: a common mechanism of kinase autophosphorylation?. Trends Biochem Sci 2007;32(8):351–6.
[12] Meng H, Liu Z, Li X, Wang H, Jin T, Wu G, Shan B, Christofferson DE, Qi C, Yu Q, Li Y, Yuan J. Death-domain dimerization-mediated activation of RIPK1 controls necroptosis and RIPK1-dependent apoptosis. Proc Natl Acad Sci USA 2018;115(9):E2001–9.
[13] Rushworth LK, Hindley AD, O'Neill E, Kolch W. Regulation and role of Raf-1/B-Raf heterodimerization. MCB 2006;26(6):2262–72.
[14] Garnett MJ, Rana S, Paterson H, Barford D, Marais R. Wild-type and mutant B-RAF activate C-RAF through distinct mechanisms involving heterodimerization. Mol Cell 2005;20(6):963–9.
[15] Inouye K, Mizutani S, Koide H, Kaziro Y. Formation of the Ras dimer is essential for Raf-1 activation. J Biol Chem 2000;275(6):3737–40.
[16] Jura N, Endres NF, Engel K, Deindl S, Das R, Lamers MH, Wemmer DE, Zhang X, Kuriyan J. Mechanism for activation of the EGF receptor catalytic domain by the juxtamembrane segment. Cell 2009;137(7):1293–307.
[17] Chung I, Akita R, Vandlen R, Toomre D, Schlessinger J, Mellman I. Spatial control of EGF receptor activation by reversible dimerization on living cells. Nature 2010;464(7289):783–7.
[18] Arkhipov A, Shan Y, Das R, Endres N, Eastwood M, Wemmer D, Kuriyan J, Shaw D. Architecture and membrane interactions of the EGF receptor. Cell 2013;152(3):557–69.
[19] Dong C, Tang X, Xie Y, Zou Q, Yang X, Zhou H. The crystal structure of an inactive dimer of PDZ-binding kinase. Biochem Biophys Res Commun 2016;476(4):586–93.
[20] Chen Z, Gibson TB, Robinson F, Silvestro L, Pearson G, Xu B-e, Wright A, Vanderbilt C, Cobb MH. MAP kinases. Chem Rev 2001;101(8):2449–76.
[21] Deacon K, Blank JL. Characterization of the mitogen-activated protein kinase kinase 4 (MKK4)/c-Jun NH 2 -terminal kinase 1 and MKK3/p38 pathways regulated by MEK kinases 2 and 3: MEK kinase 3 activates MKK3 but does not cause activation of p38 kinase in vivo. J Biol Chem 1997;272(22):14489–96.
[22] Brancho D, Tanaka N, Jaeschke A, Ventura J-J, Kelkar N, Tanaka Y, et al. Mechanism of p38 MAP kinase activation in vivo. Genes Dev 2003;17:1969–78. https://doi.org/10.1101/gad.1107303.
[23] Cuenda A. Mitogen-activated protein kinase kinase 4 (MKK4). Int J Biochem Cell Biol 2000;32(6):581–7.
[24] Wuestefeld T, Pesic M, Rudalska R, Dauch D, Longerich T, Kang T-W, et al. A direct in vivo RNAi screen identifies MKK4 as a key regulator of liver regeneration. Cell 2013;153:389–401. https://doi.org/10.1016/j.cell.2013.03.026.
[25] An E, Brognard J. Orange is the new black: kinases are the new master regulators of tumor suppression. IUBMB Life 2019;71(6):738–48.
[26] Ahn Y-H, Yang Y, Gibbons DL, Creighton CJ, Yang F, Wistuba II, et al. Map2k4 functions as a tumor suppressor in lung adenocarcinoma and inhibits tumor cell invasion by decreasing peroxisome proliferator-activated receptor γ2 expression. Mol Cell Biol 2011;31:4270–85. https://doi.org/10.1128/MCB.05562-11.
[27] Wang P-N, Huang J, Duan Y-H, Zhou J-M, Huang P-Z, Fan X-J, Huang Y, Wang L, Liu H-L, Wang J-P, Huang M-J. Downregulation of phosphorylated MKK4 is associated with a poor prognosis in colorectal cancer patients. Oncotarget 2017;8(21):34352–61.
[28] Yamada SD, Hickson JA, Hrobowski Y, Vander Griend DJ, Benson D, Montag A, et al. Mitogen-activated protein kinase kinase 4 (MKK4) acts as a metastasis suppressor gene in human ovarian carcinoma. Cancer Res 2002;62:6717–23.
[29] Pavese JM, Ogden IM, Voll EA, Huang X, Xu L, Jovanovic B, et al. Mitogen-activated protein kinase kinase 4 (MAP2K4) promotes human prostate cancer metastasis. PLoS ONE 2014;9:e102289. https://doi.org/10.1371/journal.pone.0102289
[30] Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, Cole CG, Ward S, Dawson E, Ponting L, Stefancsik R, Harsha B, Kok CY, Jia M, Jubb H, Sondka Z, Thompson S, De T, Campbell PJ. COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res 2017;45(D1):D777–83.
[31] Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the catalogue of somatic mutations in cancer. Nucleic Acids Res 2019;47:D941–7. https://doi.org/10.1093/nar/gky1015
[32] Whitmarsh AJ, Davis RJ. Role of mitogen-activated protein kinase kinase 4 in cancer. Oncogene 2007;26(22):3172–84.
[33] Zaidi SH, Harrison TA, Phipps AI, Steinfelder R, Trinh QM, Qu C, Banbury BL, Georgeson P, Grasso CS, Giannakis M, Adams JB, Alwers E, Amitay EL, Barfield RT, Berndt SI, Borozan I, Brenner H, Brezina S, Buchanan DD, Cao Y, Chan AT, Chang-Claude J, Connolly CM, Drew DA, Farris III AB, Figueiredo JC, French AJ, Fuchs CS, Garraway LA, Gruber S, Guinter MA, Hamilton SR, Harlid S, Heisler LE, Hidaka A, Hopper JL, Huang W-Y, Huyghe JR, Jenkins MA, Krzyzanowski PM, Lemire M, Lin Yi, Luo X, Mardis ER, McPherson JD, Miller JK, Moreno V, Mu XJ, Nishihara R, Papadopoulos N, Pasternack D, Quist MJ, Rafikova A, Reid EEG, Shinbrot E, Shirts BH, Stein LD, Teney CD, Timms L, Um CY, Van Guelpen B, Van Tassel M, Wang X, Wheeler DA, Yung CK, Hsu Li, Ogino S, Gsur A, Newcomb PA, Gallinger S, Hoffmeister M, Campbell PT, Thibodeau SN, Sun W, Hudson TJ, Peters U. Landscape of somatic single nucleotide variants and indels in colorectal cancer and impact on survival. Nat Commun 2020;11(1). https://doi.org/10.1038/s41467-020-17386-z.
[34] Teng DH, Perry WL, Hogan JK, Baumgard M, Bell R, Berry S, et al. Human mitogen-activated protein kinase kinase 4 as a candidate tumor suppressor. Cancer Res 1997;57:4177–82.
[35] Matsumoto T, Kinoshita T, Kirii Y, Yokota K, Hamada K, Tada T. Crystal structures of MKK4 kinase domain reveal that substrate peptide binds to an allosteric site and induces an auto-inhibition state. Biochem Biophys Res Commun 2010;400(3):369–73.
[36] Matsumoto T, Kinoshita T, Kirii Y, Tada T, Yamano A. Crystal and solution structures disclose a putative transient state of mitogen-activated protein kinase kinase 4. Biochem Biophys Res Commun 2012;425(2):195–200.
[37] Matsumoto T, Yamano A, Murakawa Y, Fukada H, Sawa M, Kinoshita T. Ensemble structural analyses depict the regulatory mechanism of non-phosphorylated human MAP2K4. Biochem Biophys Res Commun 2020;521(1):106–12.
[38] Hudson AM, Stephenson NL, Li C, Trotter E, Fletcher AJ, Katona G, Bieniasz-Krzywiec P, Howell M, Wirth C, Furney S, Miller CJ, Brognard J. Truncation- and motif-based pan-cancer analysis reveals tumor-suppressing kinases. Sci Signal 2018;11(526):eaan6776. https://doi.org/10.1126/scisignal.aan6776.
[39] OPLS3: a force field providing broad coverage of drug-like small molecules and proteins | J Chem Theor Comput n.d. https://pubs.acs.org/doi/abs/10.1021/acs.jctc.5b00864 (accessed July 2, 2020).
[40] Madhavi Sastry G, 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(3):221–34.
[41] Delaforge E, Kragelj J, Tengo L, Palencia A, Milles S, Bouvignies G, Salvi N, Blackledge M, Jensen MR. Deciphering the dynamic interaction profile of an intrinsically disordered protein by NMR exchange spectroscopy. J Am Chem Soc 2018;140(3):1148–58.
[42] Bowers KJ, Chow E, Xu H, Dror RO, Eastwood MP, Gregersen BA, et al. Scalable algorithms for molecular dynamics simulations on commodity clusters. In: Proceedings of the 2006 ACM/IEEE conference on supercomputing, Tampa, Florida: association for computing machinery; 2006, p. 84–es. https://doi.org/10.1145/1188455.1188544
[43] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79(2):926–35.
[44] Roos K, Wu C, Damm W, Reboul M, Stevenson JM, Lu C, Dahlgren MK, Mondal S, Chen W, Wang L, Abel R, Friesner RA, Harder ED. OPLS3e: extending force field coverage for drug-like small molecules. J Chem Theory Comput 2019;15(3):1863–74.
[45] Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J Mol Biol 2007;372(3):774–97.
[46] Theobald DL. Rapid calculation of RMSDs using a quaternion-based characteristic polynomial. Acta Crystallogr A Found Crystallogr 2005;61(4):478–80.
[47] Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. J Comput Chem 2011;32(10):2319–27.
[48] Gowers RJ, Linke M, Barnoud J, Reddy TJE, Melo MN, Seyler SL, et al. MDAnalysis: a python package for the rapid analysis of molecular dynamics simulations. In: Proceedings of the 15th python in science conference 2016:98–105. https://doi.org/10.25080/Majora-629e541a-00e.
[49] Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 2015;1-2:19–25.
[50] Law SM. PyMol script: modevectors.py. Available from: https://raw.githubusercontent.com/Pymol-Scripts/Pymol-script-repo/master/modevectors.py

[51] Michael Waskom, Olga Botvinnik, Joel Ostblom, Saulius Lukauskas, Paul Hobson, MaozGelbart, et al. mwaskom/seaborn: v0.9.1 (January 2020). Zenodo; 2020. https://doi.org/10.5281/zenodo.3629445.
[52] McClendon CL, Kornev AP, Gilson MK, Taylor SS. Dynamic architecture of a protein kinase. Proc Natl Acad Sci 2014;111(43):E4623–31.
[53] Ohren JF, Chen H, Pavlovsky A, Whitehead C, Zhang E, Kuffa P, Yan C, McConnell P, Spessard C, Banotai C, Mueller WT, Delaney A, Omer C, Sebolt-Leopold J, Dudley DT, Leung IK, Flamme C, Warmus J, Kaufman M, Barrett S, Tecle H, Hasemann CA. Structures of human MAP kinase kinase 1 (MEK1) and MEK2 describe novel noncompetitive kinase inhibition. Nat Struct Mol Biol 2004;11(12):1192–7.
[54] Tanaka T, Iino M, Goto K. Knockdown of Sec8 enhances the binding affinity of c-Jun N-terminal kinase (JNK)-interacting protein 4 for mitogen-activated protein kinase kinase 4 (MKK4) and suppresses the phosphorylation of MKK4, p38, and JNK, thereby inhibiting apoptosis. FEBS J 2014;281(23):5237–50.
[55] Shih H-P, Zhang X, Aronov AM. Drug discovery effectiveness from the standpoint of therapeutic mechanisms and indications. Nat Rev Drug Discov 2018;17(1):19–33.

3.1.1 Supporting Information

SUPPLEMENTARY INFORMATION

The Autoinhibited State of MKK4: Phosphorylation, Putative Dimerization and R134W Mutant Studied by Molecular Dynamics Simulations.

Ekaterina Shevchenko[1], Antti Poso[1,2], Tatu Pantsar*[,2,3]

[1] Dept. of Internal Medicine VIII, University Hospital Tübingen, Otfried-Müller-Strasse 14, 72076 Tübingen, Germany

[2] School of Pharmacy, University of Eastern Finland, Yliopistonranta 1C, 70210 Kuopio, Finland

[3] Department of Pharmaceutical and Medicinal Chemistry, Institute of Pharmaceutical Sciences, Eberhard Karls Universität Tübingen, Auf der Morgenstelle 8, 72076 Tübingen, Germany

Content

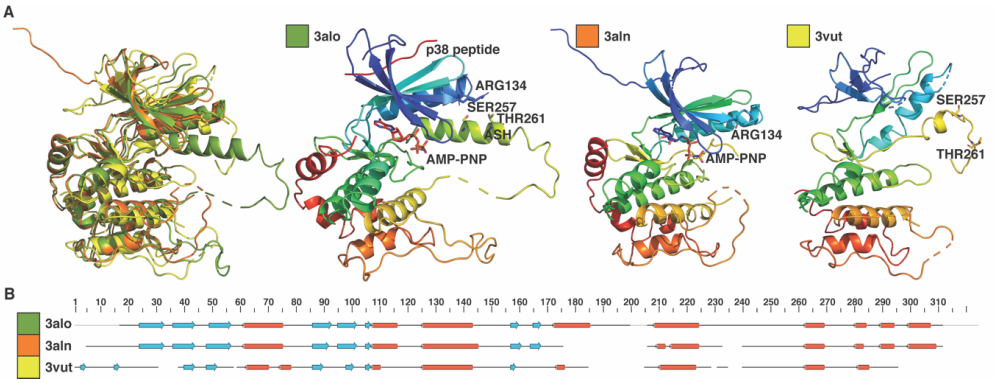


Figure S1. Superimposed MKK4 structures. (A) Three MKK4 structures are available in the RCSB Protein Data Bank (PDB IDs: 3alo, 3aln, 3vut). In the superimposed (left) structures 3alo is shown as green, 3aln as orange and 3vut as yellow. Sequence alignment **(B)** with secondary structure assignment: black lines indicate loop regions; blue arrows, β-strands; red barrels, α-helices. All structures are in inactive unphosphorylated form with length of 327 amino acids. 3alo (resolution 2.6Å) is crystallized with p38 peptide, bounded on top of N-terminal lobe of kinase. ASH (ILE250-ARG264) is in ordered autoinhibited conformation that is forming an α-helix. 3aln (resolution 2.3Å) has disordered regions: LEU254-GLY283 (alignment index 176-205), LYS309-GLY323 (alignment index 231-245). 3vut (resolution 3.5Å) is an apo structure has disordered regions: ASP263-ASP285 (alignment index 185-207), GLN316-VAL320 (alignment index 238-242), C-terminal part after ALA374 (alignment index 296) is missing.

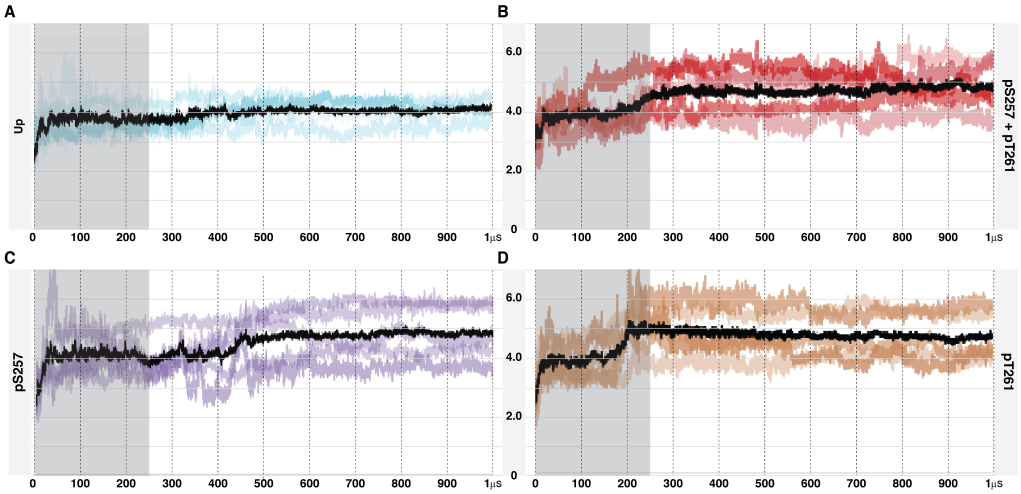


Figure S2. RMSD values of monomer MKK4 systems. (**A**) Up; (**B**) pS257+pT261; (**C**) pS257 and (**D**) pT261. RMSD was calculated for Cα-atoms. First 250 ns is highlighted with grey colour. The black line indicates the average and the other colours indicate individual replicas.

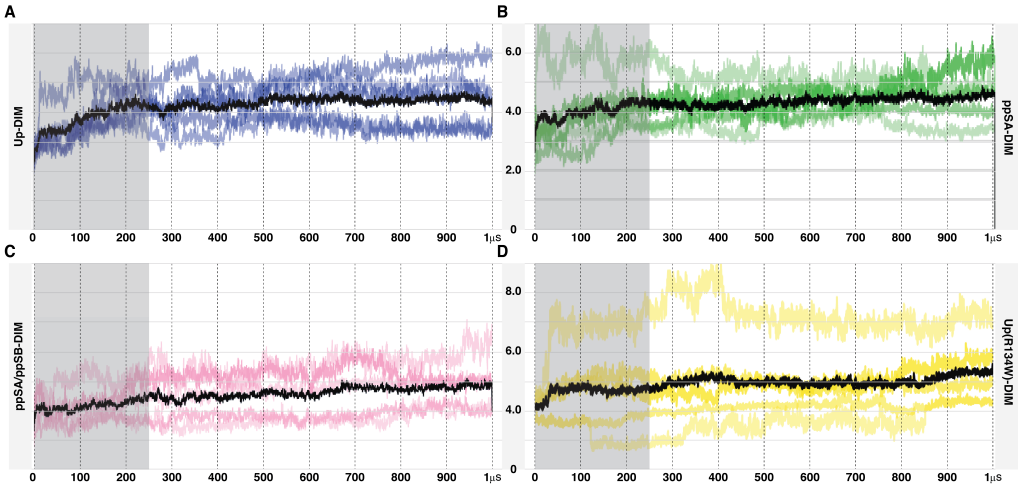


Figure S3. RMSD values of homodimer MKK4 systems. (**A**) Up-DIM; (**B**) ppSA-DIM; (**C**) ppSA/ppSB-DIM and (**D**) Up(R134W)-DIM. RMSD was calculated for Ca-atoms. First 250 ns is highlighted with grey colour. The black line indicates the average and the other colours indicate individual replicas.

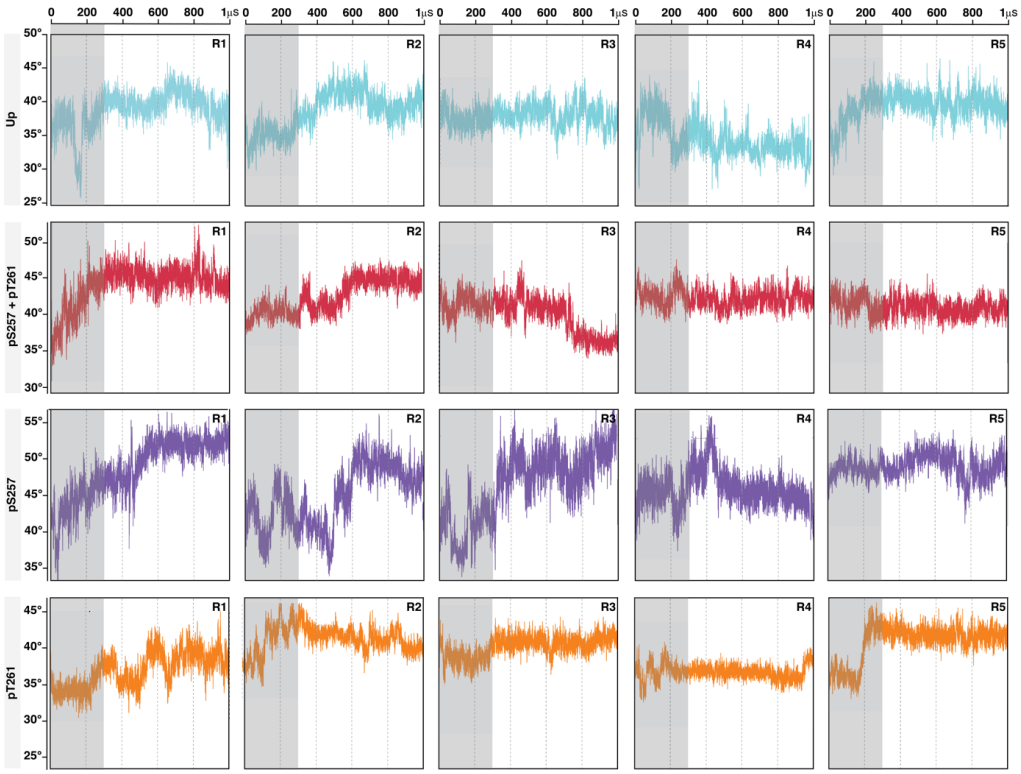


Figure S4. Observed angle values (aC-helix–HR–ASH) in monomer MKK4 systems. First 250 ns, which was excluded from the analysis related to Figure 3, is highlighted with grey colour. R1 = replica 1, R2 = replica 2, etc.

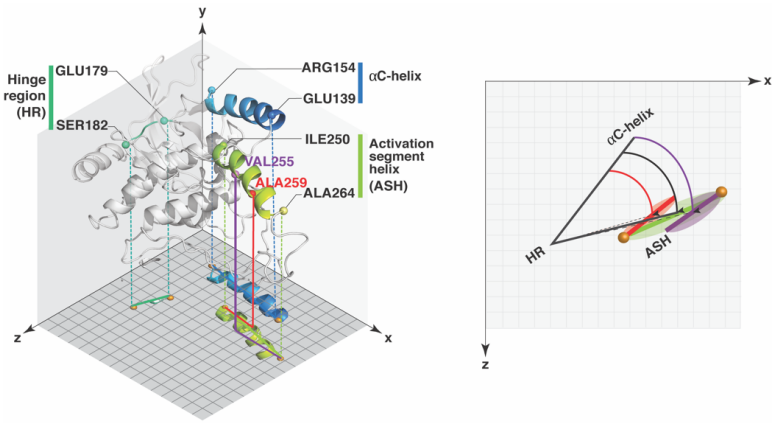


Figure S5. Plane projection of residue intervals used for angle calculation and validation. Monomer MKK4 system in three-dimensional space with flat representation of XZ-plane with different residue interval selections. Default selection of ASH defining residues ILE250–ALA264 (green) is switched to ILE250–ALA259 (red) and VAL255–ALA264 (purple). Black arrow heads on the XZ-plane indicate centre of geometry in each residue interval. Black curve indicates the reference angle (39.05°), from which red and purple display deviation of ~ -1–2° (ILE250–ALA259) and ~ +1–2° (VAL255–ALA264). Colours of selected residue intervals are consistent for both left and right images.

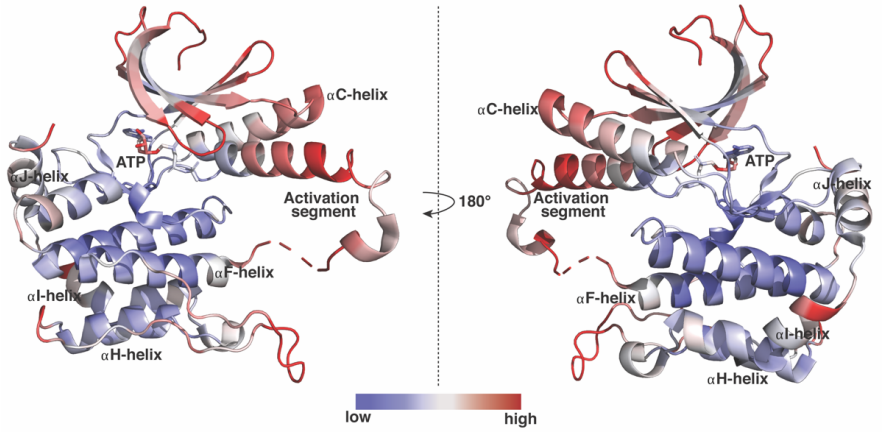


Figure S6. B-factor of 3alo crystal structure. Highest B factor values are indicated with red and lowest with blue colour. Disordered region is shown in dashed red line in the cartoon representation.

Table S1. Notable RMSF differences (> 0.1 nm) compared to the Up system. The most proximal residues of the C- and N-terminus (< ASP96 and > MET388) are excluded here.

	pS257 + pT261	pS257	pT261
Activation loop	K260, R262, R267–A271 (higher); I275–P277, R281, S288 (lower)	S251–K260, R262, P268–I275, S278–A279, Y284 (higher)	P268–M270, Q282–Y284 (higher)
C-terminal loop	P308–L317, (higher); V321, P325–L328 (lower)	W310–T318 (higher)	V313–G323 (higher)

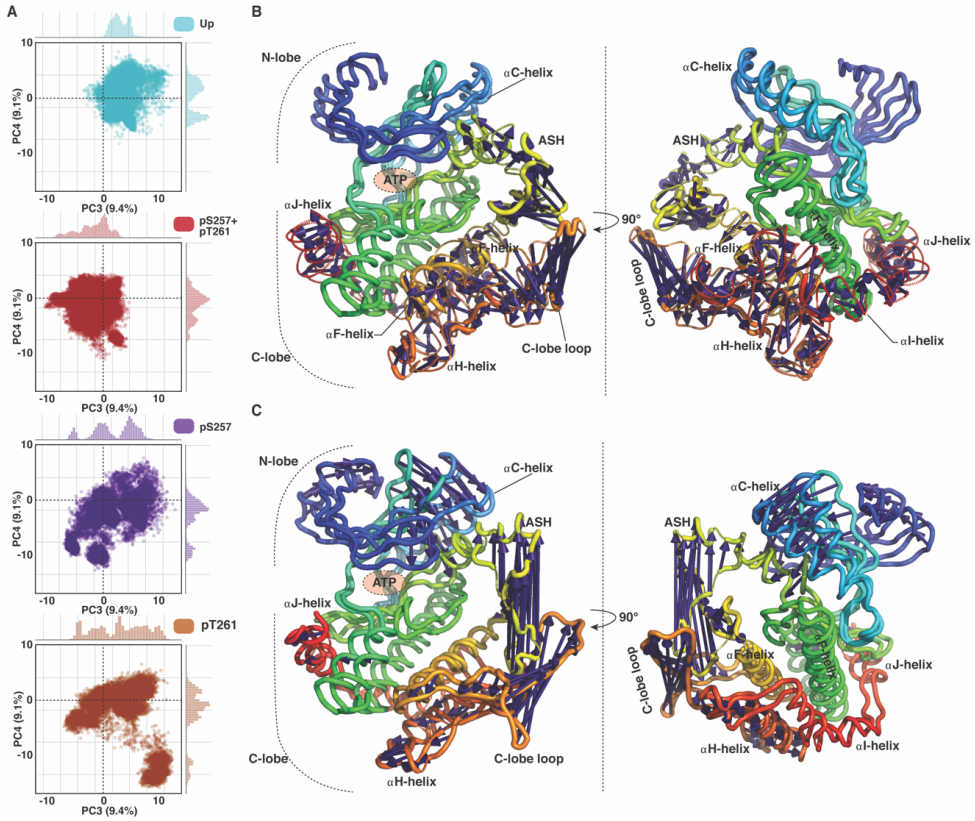
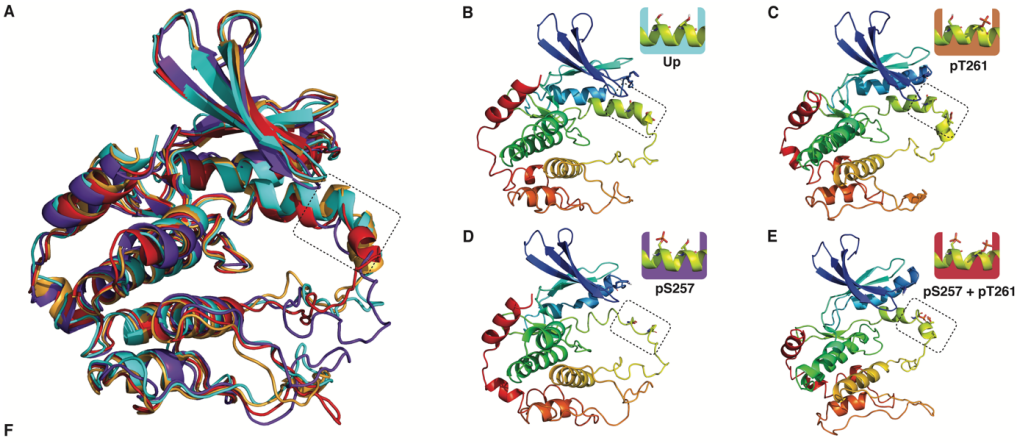


Figure S7. Principal components analysis: PC3 and PC4. (**A**) Principal component analysis (PCA) score plot of PC3 and PC4. (**B**) The extreme movements of PC3. (**C**) The extreme movements of PC4. In **B** and **C** protein is illustrated with rainbow colour and the purple arrows indicate the extreme movements.



F

	Replica 1	Replica 2	Replica 3	Replica 4	Replica 5
Up	2.8	3.6	2.9	2.9	2.6
pS257	2.2	2.6	2.2	4.0	3.1
pT261	2.9	2.7	3.4	3.1	4.6
pS257+pT261	4.5	3.2	3.5	3.8	2.7

Figure S8. Average monomer conformations at 1 μs of the simulations. All systems are superimposed in **A** and each average conformation is shown as individually in **B**–**E**. Dashed square indicates position of SER257 and THR261. See also **SI Movie M5** for the conformations. (**F**) RMSD of different systems compared to their average 1 μs conformations.

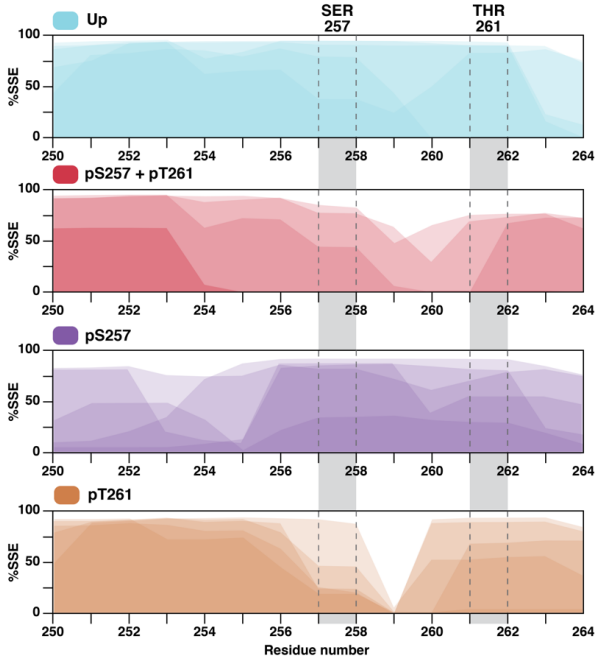


Figure S9. Secondary structure analysis of activation segment helix (ASH; I250–A264) in monomer MKK4. Plots display the percentage of time each residue contributes to each of the two secondary structure elements (%SSE): helix (coloured space inside the graphs), and loop (white space). Each replica (five replicas for each system in total) is represented with 20% colour opacity. The colour intensity is directly depended on the number of replicas, where particular secondary structure occurs. **Up**: helix 78%, loop 22%; **pS257+pT261**: helix 62%, loop 38%; **pS257**: helix 63%, loop 37%; **pT261**: helix 69%, loop 31%.

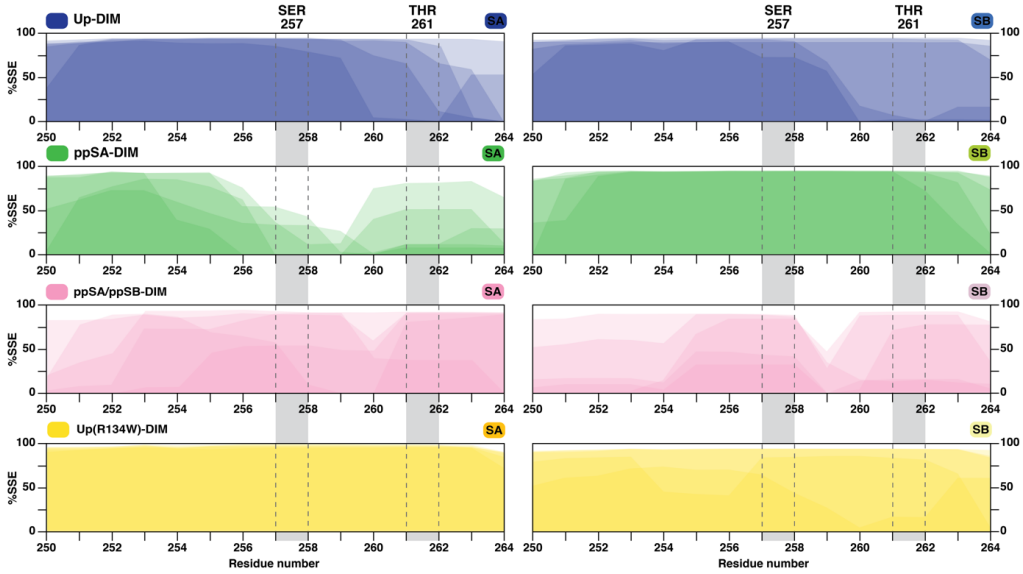


Figure S10. Secondary structure analysis of activation segment helix (ASH; I250–A264) in homodimer MKK4. Plots display the percentage of time each residue contributes to each of the two secondary structure elements (%SSE): helix (coloured space inside the graphs), and loop (white space). Analysis was performed separately for subunit A (SA) and subunit B (SB). Each replica (five replicas for each system in total) is represented on the graph with 20% colour opacity. The colour intensity is directly depended on the number of replicas, where particular secondary structure occurs. **Up-DIM** SA: helix 84%, loop 16%; SB: helix 83%, loop 17%; **ppSA-DIM** SA (phosphorylated): helix 50%, loop 50%; SB (unphosphorylated): helix 92%, loop 8%; **ppSA/ppSB-DIM** SA: helix 70%, loop 30%; SB: helix 52%, loop 48%. **Up(R134W)-DIM** SA: Helix 97%, loop 3%; SB: helix 86%, loop 14%.

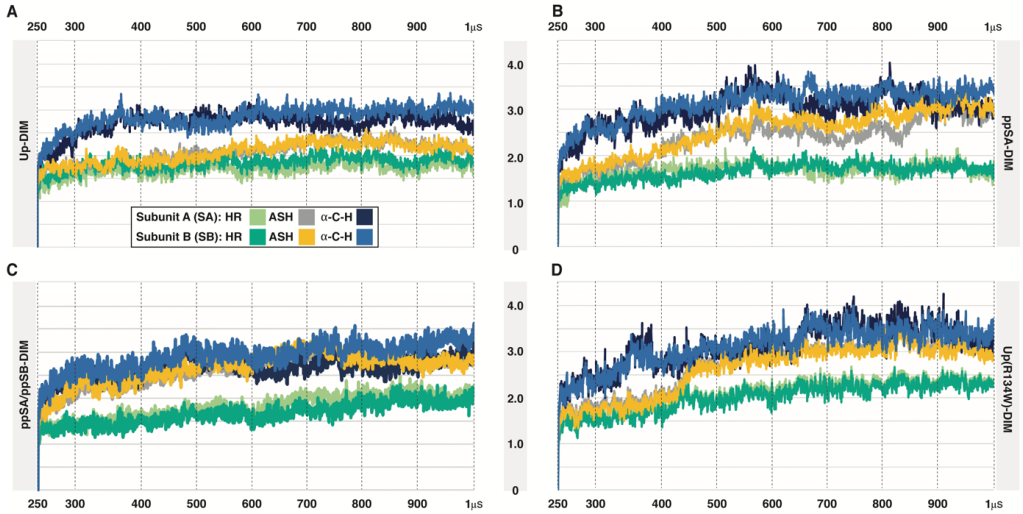
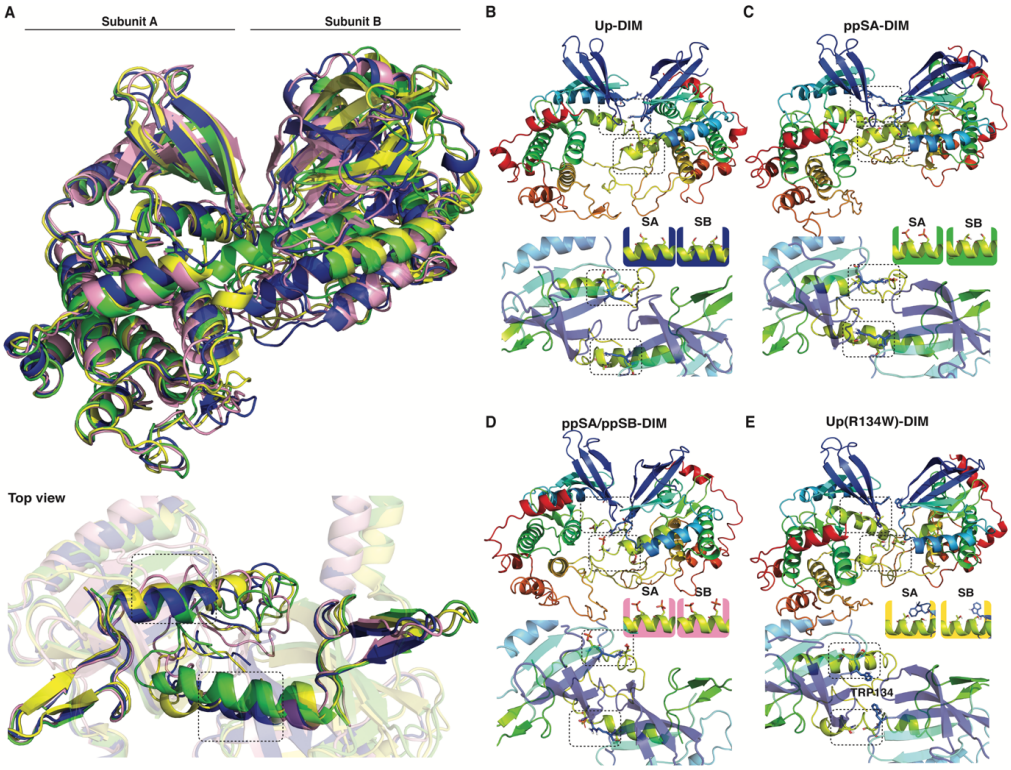


Figure S11. RMSD of residue intervals used for angle calculation in homodimer MKK4. RMSD is calculated for hinge region (HR; E179-S182; green), Activation segment helix (ASH; I250–A264; grey/yellow) and α-C-helix (α-C-H; E139–R154; blue) separately. (**A**) Up-DIM, (**B**) ppSA-DIM, (**C**) ppSA/ppSB-DIM and (**D**) Up(R134W)-DIM.



F

	Replica 1		Replica 2		Replica 3		Replica 4		Replica 5	
	SA	SB	SA	SB	SA	SB	SA	SB	SA	SB
Up-DIM	2.9	2.7	2.4	2.5	2.7	2.5	2.4	2.8	2.4	2.5
ppSA-DIM	2.5	2.4	2.9	2.5	2.6	3.5	2.4	2.6	2.2	2.5
ppSA/ppSB-DIM	2.5	3.0	3.0	3.0	3.0	2.7	3.4	2.3	3.3	2.7
Up(R134W)-DIM	2.2	2.9	2.1	2.8	2.3	2.2	2.6	3.1	2.5	2.3

Fig S12. Average homodimer conformations at 1 μs of the simulations. All systems are superimposed in **A** and each average conformation is shown as individually in **B**–**E**. Dashed square indicates position of SER257 and THR261. See also **SI Movies M6–M7** for the conformations. (**F**) RMSD of different systems compared to their average 1 μs conformations.

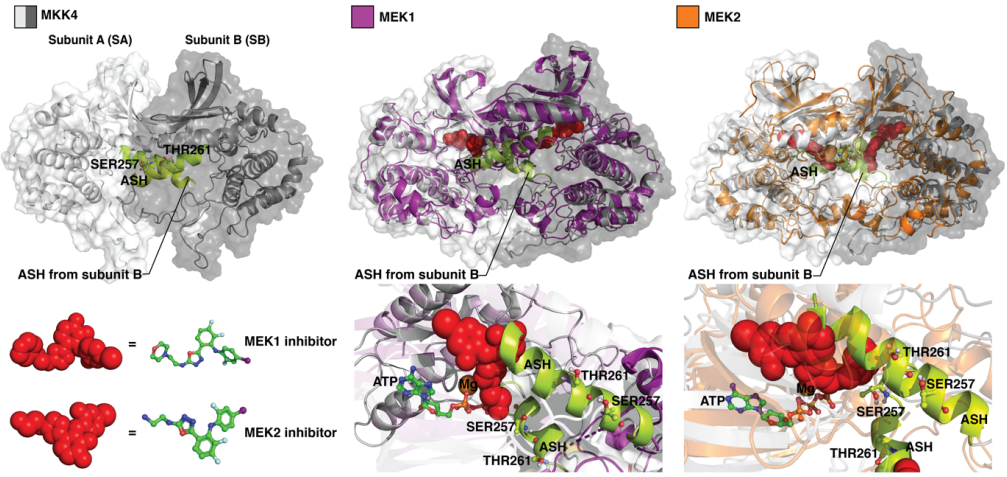


Fig S13. Superimposed structures of MKK4, MEK1 and MEK2. In the superimposed structures MKK4 is shown as grey (Subunit A) and dark grey (Subunit B), MEK1 is shown as purple, and MEK2 as orange. Activation segment helix (ASH) is shown in green. Allosteric MEK1/MEK2 inhibitors that partially occupy the corresponding ASH location in MKK4 are represented with red spheres. PDB IDs: 3alo (MKK4); 3eqb (MEK1),1s9i (MEK2).

Table S2. Distances between subunit A and subunit B of MKK4. R1 = replica 1; R2 = replica 2, etc.; SD = standard deviation

Distance (nm) Up-DIM						
Element Replica	N-lobe	SD	C-lobe	SD	Activation Segment	SD
R1	1.979	0.050	4.282	0.127	1.646	0.039
R2	2.040	0.068	3.998	0.083	1.218	0.041
R3	1.888	0.063	4.746	0.084	1.494	0.087
R4	1.901	0.047	4.035	0.078	1.496	0.043
R5	2.081	0.109	4.585	0.073	1.482	0.056
Average	**1.977**		**4.329**		**1.467**	
Distance (nm) ppSA-DIM						
Element Replica	N-lobe	SD	C-lobe	SD	Activation Segment	SD
R1	1.909	0.128	4.417	0.092	1.852	0.061
R2	1.900	0.082	4.523	0.089	1.720	0.073
R3	1.888	0.063	4.746	0.084	1.494	0.087
R4	1.890	0.058	4.631	0.080	1.423	0.036
R5	1.798	0.048	4.522	0.079	1.384	0.034
Average	**1.870**		**4.567**		**1.574**	
Distance (nm) ppSA/ppSB-DIM						
Element Replica	N-lobe	SD	C-lobe	SD	Activation Segment	SD
R1	1.877	0.061	4.205	0.087	1.514	0.066
R2	1.896	0.067	4.245	0.079	2.217	0.089
R3	1.954	0.065	4.233	0.072	1.759	0.089
R4	1.917	0.046	4.006	0.065	1.752	0.036
R5	2.382	0.125	4.322	0.057	1.706	0.047
Average	**2.005**		**4.202**		**1.789**	
Distance (nm) Up(R134W)-DIM						
Element Replica	N-lobe	SD	C-lobe	SD	Activation Segment	SD
R1	4.137	0.240	4.238	0.071	1.351	0.062
R2	2.850	0.187	4.145	0.099	1.170	0.106
R3	2.011	0.081	4.417	0.092	1.581	0.106
R4	2.635	0.068	4.376	0.135	1.510	0.036
R5	2.889	0.091	4.146	0.065	1.345	0.071
Average	**2.904**		**4.264**		**1.391**	

Table S3. Distances within the N-lobe of dimer MKK4. Overall table of distance statistics (nm) for each dimer system. R1 = replica 1, R2 = replica 2, etc.; SD = standard deviation

	R1		R2		R3		R4		R5		
Distance (nm) Up-DIM											
Distance	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD	**Total average**
V116-V120	2.102	0.055	2.177	0.075	1.980	0.050	2.004	0.053	2.227	0.132	**2.098**
V120-I127	3.525	0.086	3.607	0.124	3.188	0.097	3.444	0.084	3.540	0.148	**3.461**
F164-L168	3.653	0.038	3.582	0.048	3.620	0.035	3.477	0.035	3.704	0.074	**3.607**
A111-G11	0.567	0.027	0.614	0.026	0.617	0.031	0.550	0.019	0.586	0.030	**0.587**
Distance (nm) ppSA-DIM											
Distance	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD	**Total average**
V120-I127	3.070	0.221	3.462	0.110	3.430	0.126	3.367	0.094	3.194	0.134	**4.288**
F164-L168	3.592	0.078	3.497	0.048	3.615	0.060	3.465	0.038	3.494	0,071	**3.951**
A111-G11	0.809	0.059	0.579	0.024	0.679	0.030	0.585	0.021	0.582	0.069	**0.879**
Distance (nm) ppSA/ppSB-DIM											
Distance	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD	**Total average**
V120-I127	3.254	0.114	3.349	0.141	3.258	0.124	3.231	0.088	3.789	0.201	**3.376**
F164-L168	3.591	0.039	3.827	0.046	3.755	0.054	3.713	0.040	3.898	0.128	**3.757**
A111-G11	0.681	0.052	0.856	0.045	0.771	0.044	0.773	0.025	0.529	0.049	**0.722**
Distance (nm) Up(R134W)-DIM											
Distance	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD	**Total average**
V116-V120	4.390	0.244	2.177	0.075	2.148	0.091	2.878	0.067	3.121	0.093	**2.943**
V120-I127	5.895	0.268	3.607	0.124	3.439	0.114	4.100	0.120	4.398	0.134	**4.288**
F164-L168	5.032	0.155	3.582	0.048	3.554	0.073	3.786	0.049	3.804	0.070	**3.951**
A111-G11	1.650	0.1626	0.616	0.026	0.560	0.034	0.769	0.099	0.802	0.069	**0.879**

3.2 Publication II

Regulatory spine RS3 residue of protein kinases: a lipophilic bystander or a decisive element in the small-molecule kinase inhibitor binding?

Ekaterina Shevchenko [a], Tatu Pantsar [b]

[a] Department of Internal Medicine VIII, University Hospital Tübingen, Otfried-Müller-Strasse 14, 72076 Tübingen, Germany

[b] School of Pharmacy, University of Eastern Finland, Yliopistonranta 1C, 70210 Kuopio, Finland

Biochemical Society Transactions
Volume 50 (1), 2022, pages 633–648, DOI: 10.1042/BST20210837

Biochemical Society Transactions (2022) **50** 633–648
https://doi.org/10.1042/BST20210837

PORTLAND PRESS

Review Article

Regulatory spine RS3 residue of protein kinases: a lipophilic bystander or a decisive element in the small-molecule kinase inhibitor binding?

Ekaterina Shevchenko[1] and **Tatu Pantsar**[2]

[1]Department of Internal Medicine VIII, University Hospital Tübingen, Otfried-Müller-Strasse 14, Tübingen, DE 72076, Germany; [2]School of Pharmacy, Faculty of Health Sciences, University of Eastern Finland, Yliopistonranta 1, 70210 Kuopio, Finland

Correspondence: Tatu Pantsar (tatu.pantsar@uef.fi)

In recent years, protein kinases have been one of the most pursued drug targets. These determined efforts have resulted in ever increasing numbers of small-molecule kinase inhibitors reaching to the market, offering novel treatment options for patients with distinct diseases. One essential component related to the activation and normal functionality of a protein kinase is the regulatory spine (R-spine). The R-spine is formed of four conserved residues named as RS1–RS4. One of these residues, RS3, located in the C-terminal part of αC-helix, is usually accessible for the inhibitors from the ATP-binding cavity as its side chain is lining the hydrophobic back pocket in many protein kinases. Although the role of RS3 has been well acknowledged in protein kinase function, this residue has not been actively considered in inhibitor design, even though many small-molecule kinase inhibitors display interactions to this residue. In this minireview, we will cover the current knowledge of RS3, its relationship with the gatekeeper, and the role of RS3 in kinase inhibitor interactions. Finally, we comment on the future perspectives how this residue could be utilized in the kinase inhibitor design.

Introduction

Protein kinases are dynamic proteins that regulate a multitude of cellular signalling processes. They control the activity of their downstream targets mainly by phosphorylation, and their own activity is also usually controlled in the same manner. The human kinome comprises more than 500 protein kinases [1], and nearly 500 proteins contain a typical kinase domain [2]. Still, the biological role of many protein kinases is largely unknown, and there are ongoing efforts aiming to characterize these poorly understood protein kinases [3]. Although protein kinases display high similarity in their kinase domain, there is a higher level diversity in their structures; while some kinases consist (almost) solely of the kinase domain (e.g. MAPK14, GSK3B), other are larger multidomain proteins (e.g. LRRK2 [4]). The structure and function of the protein kinase domain is well-established (Figure 1A). For a comprehensive view of structural history of protein kinases, the reader is recommend a recent review by Taylor et al. [5].

In the protein kinase domain, one of the key dynamic elements in regulating protein kinase function is the hydrophobic regulatory-spine (R-spine), which was discovered already 15 years ago in 2006 [6]. The R-spine consists of four residues, named RS1–RS4, which connects the two lobes of the kinase domain (Figure 1A,B). Two of these residues, RS1 and RS2, belong the C-lobe. RS1 is His (sometimes Tyr) residue from the HRD (or YRD) motif [7]. RS2 is Phe (or Leu) from the DFG-motif, which is part of the activation loop of a protein kinase. The other two R-spine residues belong to the N-lobe of the protein kinase. RS4 is a residue from the β4-strand, which is less conserved but frequently Leu can be found in this position. Finally, RS3 is located four residues C-terminal from the αC-helix Glu that forms a salt-bridge to the Lys of β3-sheet. RS3 is usually (not always) accessible

Received: 22 December 2021
Revised: 15 February 2022
Accepted: 17 February 2022

Version of Record published:
28 February 2022

Biochemical Society Transactions (2022) **50** 633–648
https://doi.org/10.1042/BST20210837

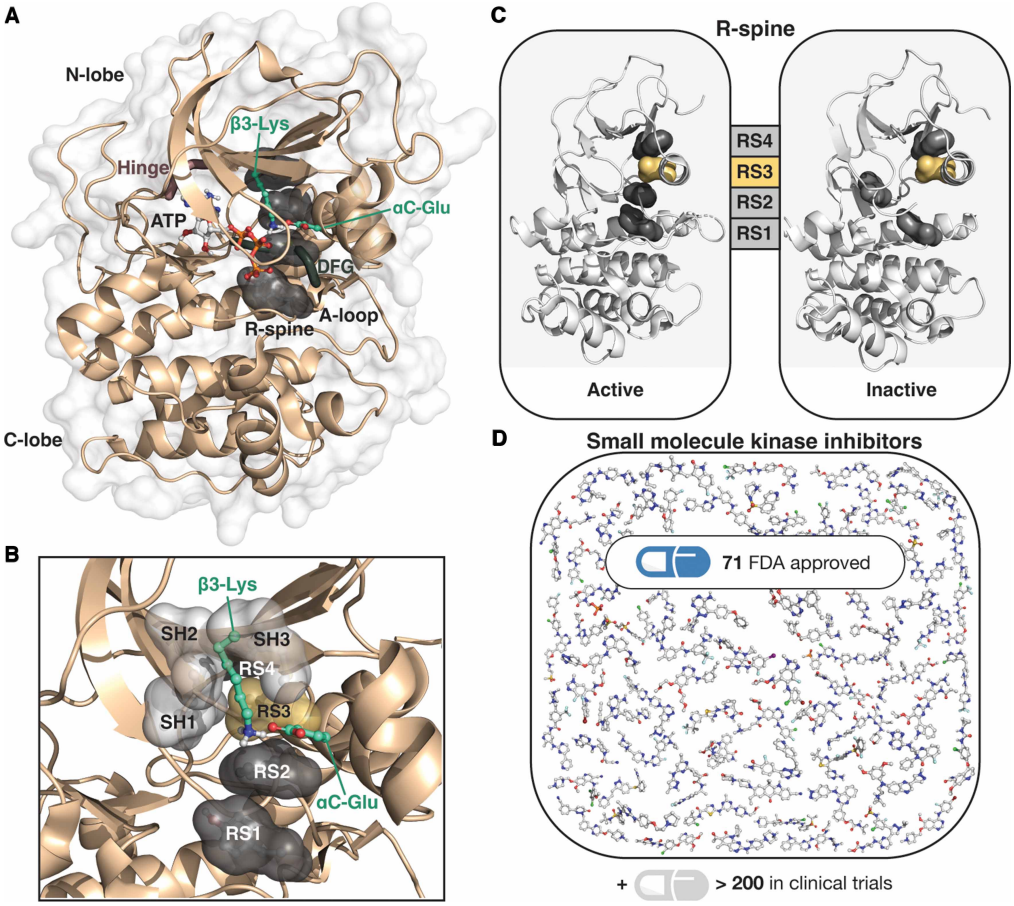


Figure 1. Protein kinase domains are important drug targets.

(**A**) A typical structure of a protein kinase domain. ATP-binding cleft is located between the N- and C-lobes of the kinase. In the figure, structure of cAMP-dependent protein kinase catalytic subunit alpha is depicted (PDB ID: 4wb5 [20]; inhibitory peptide is hidden). R-spine residues are illustrated with black surface. ATP, β3-Lys and αC-Glu are shown with stick model. (**B**) Shell residues SH1–SH3 (grey surface) are located next to the R-spine RS3 and RS4 residues. Shell residue SH2 (gatekeeper) is located close to RS3 (yellow surface). (**C**) The R-spine of a protein kinase is assembled in active conformation and disassembled in inactive conformation. In the figure, active and inactive configurations are illustrated with BRAF (PDB IDs: 4e26 [21] and 1uwh [22]). (**D**) Several small-molecule kinase inhibitors are already in clinical use, and dozens are in clinical trials.

from the ATP binding site as its side chain lines the active site cleft. Overall, αC-helix, where RS3 is located, has a central role in the kinase activation process [8]. In the catalytically active form of a protein kinase, R-spine is assembled and in the inactive state it is disassembled (Figure 1C). In the active state the location of RS2 as part of the assembled R-spine results in an open and extended conformation of the activation loop (A-loop), while a closed A-loop configuration is preferred in the inactive state. Notably, additional stabilization of the R-spine, such as via in-frame insertions or RS3 mutations [9,10], may result in increased catalytic activity of the protein kinase.

Next to the R-spine in the N-lobe are located three conserved residues, named as Shell (SH) residues (Figure 1C) [9]. These residues, which are usually hydrophobic, have a role in supporting R-spine and are therefore important for kinase activity. One of these residues, SH2, is found close to RS3. This SH2 residue is more commonly known as the gatekeeper residue, which is named due to its role in controlling access to the

hydrophobic pocket [11]. This shell residue participates in regulating R-spine dynamics, and gatekeeper mutations may stabilize the R-spine promoting the kinase activation [12].

In recent years, ever increasing efforts have been conducted by the pharmaceutical industry to target protein kinases [13]. These efforts have resulted in numerous small-molecule kinase inhibitors, totalling now over 70 FDA approved small-molecule inhibitors (Figure 1D). According to the Protein Kinase Inhibitors in Clinical Trials database (PKIDB) [14,15], approximately 300 small-molecule kinase inhibitors are either in clinical trials or already approved. Comprehensive reviews of the kinase inhibitor drug discovery and kinase inhibitor development are available [16–18]. Currently, oncology is dominating indication for the kinase inhibitors, but there is potential also in other therapeutic areas such as autoimmune and inflammatory diseases, and degenerative disorders [19].

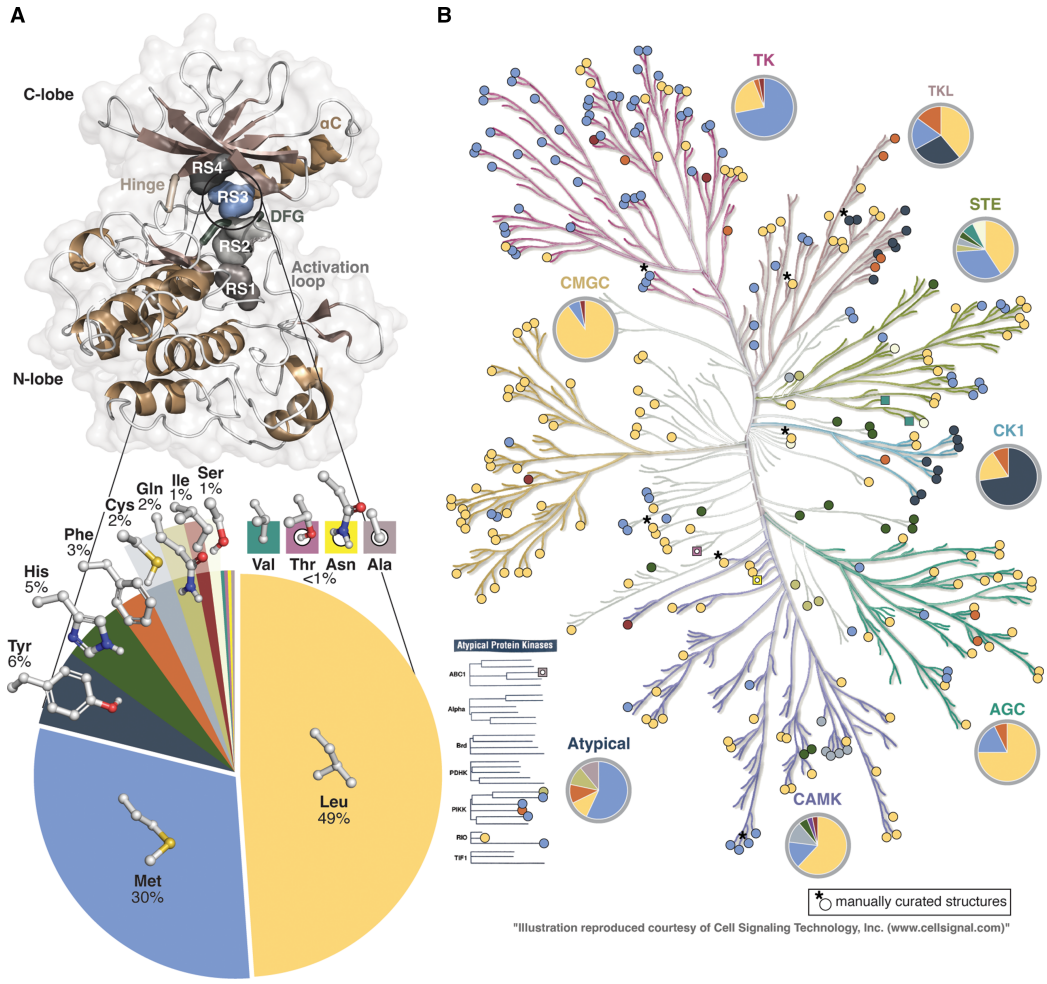


Figure 2. RS3 in human protein kinases with available structures.

(**A**) Occurrence of RS3 residues in human protein kinases with publicly available structural data (289 kinases). The location of RS3 is highlighted in IGFR1 kinase domain (PDB ID: 3qqu [28]). The shown frequencies are rounded up to the nearest %, for residues with <1% frequency, percentage is not shown. (**B**) RS3 residue distribution in human kinome. Colours of the residue types are as in **A**. Eight structures with KLIFS annotation errors that were manually curated (RS3 was properly assigned) are indicated with an asterisk. Data in **A** and **B** consist of human protein kinases with publicly available structures with (with lipid kinases excluded). Human kinome tree illustration was made with the help of KinMap [29].

PORTLAND PRESS

Biochemical Society Transactions (2022) **50** 633–648
https://doi.org/10.1042/BST20210837

Here, we review the characteristics of RS3 and its relationship with the neighbouring gatekeeper (SH2) based on the publicly available structural data. We also have a look at RS3 interactions to small-molecule kinase inhibitors, including the approved drugs. Finally, we end the review with available mutational data of RS3.

RS3 in the human kinome

A majority of the human protein kinases with publicly disclosed structures display a nonpolar aliphatic RS3 residue (Figure 2). Nearly half of these kinases exhibit Leu in RS3, and almost a third have Met (Figure 2A). In the overall human proteome Leu is also the most abundant residue (9.97%), while Met has the second lowest frequency (2.13%) of all amino acids [23]. Following the abundant Leu and Met in RS3, next preferred are aromatic residues. Tyr, His and Phe appear with the frequencies of 3–6%. Cys and Gln exist in 2%, Ile and Ser in 1%. Even more rare residues that are observed in this location are Val, Thr, Asn and Ala. The charged residues, Asp, Glu, Lys and Arg, as well as structurally more unique Trp, Gly or Pro are not present in the analysed set of human protein kinases in RS3. Based on the sequence alignment of a larger set of eucaryotic protein kinases these residues have been suggested to exist as RS3 residues, although rarely [7]. Regardless of the high existence of hydrophobic residues in RS3, there exists no clear trend related to the hydrophobicity ranking of the residues and their observed frequencies [24,25].

Protein kinases of different groups and families display distinct preferences for RS3 residues (Figure 2B). The majority of the kinases (72%) belonging to the TK group display Met RS3. More than a fifth (22%) of this group present Leu in this position, including protein kinases belonging to JakA family (JAK1–3; TYK2) and Trk family (NTRK1–3 also known as TRKA–C). Four protein kinases have either Ile (ALK; ERBB3 (ErbB3)) or a Phe (PTK7 (CCK4); LMTK3 (LMR3)). In EGFR family Met is preferred in RS3, except ErbB3 has Ile. Interestingly, ErbB3 has been identified to display considerably lower kinase activity [26,27]. However, ErbB3 displays also other unique characteristics that differ from other EGFRs (for instance, instead of αC-Glu ErbB3 has a His in this location). In the PDGFR family, KIT displays Leu instead of Met that is observed in other family members (FLT, PDGFRA, CSF1R (FMS)).

In contrast with the TKs, in the CMGC group Leu is clearly dominating RS3 (90%). Only three protein kinases have Met in RS3 (MAPK8 (JNK1); MAPK10(JNK3); GSK3B), and MAPK6 (Erk3) displays Ile in this position.

The TKLs prefer quite diversely Leu, Met, Tyr and Phe in their RS3. Members of the STKR family prefer an aromatic residue in this position: Tyr is observed in ACVR1 (ALK2), ACVR2A (ACTR2), ACVRL1 (ALK1), BMBR1B, BMBR2 and TGFBR1 (TGFbR1); Phe in ACVR2B (ACTR2B) and TGFBR2 (TGFbR2). Aromatic Phe is also present in two MLK family members: MAP3K9 (MLK1) and MAP3K21 (MLK4), while other kinases in this family display Leu (ILK; MAP3K12 (DLK); MAP3K20 (ZAK); MAP3K7 (TAK1); TNNI3K (HH498)).

Kinases of the STE group display mainly Leu and Met in RS3. However, also other RS3 residues are observed in this group, including His (STE11) and Ser (STRADA (STLK5); MAP2K6), Gln (MAP3K8 (COT)), Cys (MAP3K14 (NIK)) and Val (MAP2K4; MAP2K7). From these, MAP3K14 (NIK) and MAP3K8 (COT) belong to the STE-Unique family. Three members of the STE7 family display rare RS3 residues. These kinases are MAP2K6 (Ser), MAP2K4 also known as MKK4 (Val) and MAP2K7 also known as MKK7 (Val). Interestingly, unique autoinhibited conformational states have been reported for these kinases state [30–32]; with MKK4 this state may be related to its dimer form [33].

In the CK1 group, Tyr is dominating in RS3 (73%). VRK3 displays an aromatic Phe in this position. TTBK family kinases (TTBK1, TTBK2) are more diverse in this group with their aliphatic Leu in RS3.

The majority of the AGC and CAMK group kinases exhibit Leu in RS3, which is followed by Met with lower frequencies. Two kinases with an aromatic RS3 (Phe) are observed in AGC group, in PKN2 and PRKCI (PKCi). In CAMK group, kinases of the CAMK2 family (CAMK2A, CAMK2B, CAMK2D, CAMK2G) display Cys in RS3, as well as CASK. His is observed in MAPKAP family (MAPKAP2, MAPKAP3). TRIB1 (Trb1) represents Ile in RS3; however, in this pseudokinase the neighbouring Tyr may actually occupy the canonical RS3 position [34,35].

The protein kinases that are not belonging to any specific group display also family specific preferences. For instance, protein kinases of WEE family and PLK family exhibit His in RS3 (located above and below CK1 group in the kinome tree). Of the Atypical kinases, COQ8A that is also known as ADCK3 [36], is the only structure in the dataset that displays Ala in RS3.

Biochemical Society Transactions (2022) **50** 633–648
https://doi.org/10.1042/BST20210837

Polar RS3 are rare

Not only hydrophobic RS3 residues exist, but also polar residues are observed in this position. AURKA is an example of a widely studied kinase that has a polar RS3 (Gln) is. It was disclosed by Levinson et al. that this polar residue has a specific role in AURKA activation via a water-network [37]. Similarly, AURKB and AURKB have also Gln as their RS3. In addition, Gln is observed in MAP3K8 (COT, TPL2) and ATM. MAP3K8 controls inflammation [38] and ATR DNA damage responses [39].

In the available data, Ser is observed in three kinases. While in STRADA (STLK5, STE20) this residue appears unreachable from the binding cleft (PDB ID: 3gni [40], 2wtk [41]), in Haspin Ser is accessible (participates in water coordination next to αC-Glu (PDB ID: 4ouc [42]). RS3 Ser may also be accessible in MAP2K6 when it is not in its autoinhibited state (PDB ID: 3fme). In the autoinhibited state its neighbouring Met appears to take the regular RS3 position (PDB ID: 3vn9 [32]).

Two unique polar RS3 residues are present in the data. Asn is observed in the RS3 of CHK1, while all the other kinases belonging to the same CAMKL family have either a lipophilic Leu or Met in this position. CHK1 inhibition could be useful in the treatment of KRAS driven pancreatic ductal adenocarcinoma [43]. Thr is observed in ULK4, while other members of the ULK family (ULK1–ULK3) have Leu in the respective position.

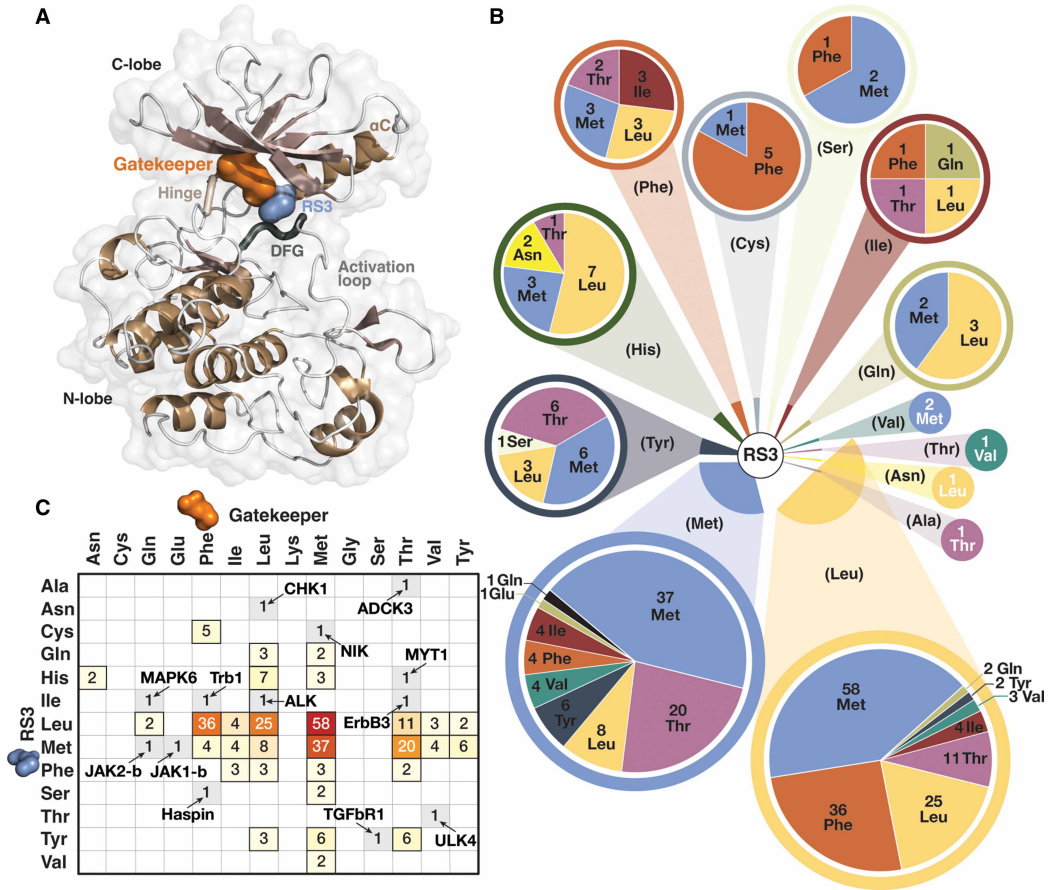


Figure 3. Gatekeeper and RS3.

(**A**) Gatekeeper is located near RS3, and in some cases these residues are in contact. In the figure, gatekeeper of the protein kinase IGF1R (PDB ID: 3qqu [28]) is shown with orange surface and RS3 with blue surface. (**B**) Distribution of gatekeeper residues according to their RS3 residue (in parenthesis) represented in pie charts. (**C**) Correlation matrix of RS3 and gatekeeper residues. The unique RS3-gatekeeper combination containing kinases are labelled.

Biochemical Society Transactions (2022) **50** 633–648
https://doi.org/10.1042/BST20210837

ULK4 is a pseudokinase and has an unusual structural characteristic in its αC-helix: it exhibits Trp residue in the location of αC-Glu. This Trp appears to participate in its R-spine formation [44].

RS3 relationship with gatekeeper

The access towards the hydrophobic pocket (and towards RS3) is controlled by gatekeeper, also known as SH2 residue. This residue may also influence R-spine dynamics and it can be found in close contact to RS3 (Figure 3A). Generally, protein kinases prefer Met, Leu, Phe and Thr gatekeepers (Figure 3B,C). In the available structures, Met is the most abundant gatekeeper (40%), followed by Leu (18%), Phe (16%) and Thr (15%). Less frequent gatekeepers — but presented in more than eight kinase domains — are Ile 4%, Tyr 3%, Val 3%.

Kinases with different RS3 residues display also distinct gatekeeper preferences (Figure 3B,C). With Met in RS3, less Phe gatekeepers are observed in comparison with when RS3 is Leu. Bulky aromatic RS3 residues (Tyr, His, Phe) do not exist in combination with aromatic gatekeepers. The kinase domains with Cys in RS3 appear to prefer Phe gatekeepers. Polar Gln and Asn display either Leu or Met as their gatekeepers. While Thr is rare RS3 residue, as a gatekeeper it is common with Leu, Met and Tyr in RS3. In addition to Thr, other polar gatekeepers do exist. Asn gatekeeper is observed only in combination with His RS3 in WEE1 (Wee1) (PDB ID: 3biz [45]) and WEE2 (Wee1B) (PDB ID: 5vdk [46]). Gln gatekeeper is observed in four kinases, including MAPK1 (Erk2) (PDB ID: 5hnv [47]), MAPK3 (Erk1) (PDB ID: 4qtb [48]), MAPK6 (Erk3) (PDB ID: 7aqb [49]), and pseudokinase domain JAK2-b (PDB ID: 4fvq [50]). Pseudokinase domain Jak1-b has a unique polar Glu as its gatekeeper (PDB ID: 4l00 [51]). Overall, thirteen protein kinases in the analysed dataset display unique RS3-gatekeeper combinations (Figure 3C).

RS3 and small-molecule kinase inhibitors

We searched the KLIFS database [52] and complemented our search using Protein–Ligand Database (PLDB) tool of Maestro (Schrödinger LLC) to map out all the existing protein kinase–ligand complexes that have contacts between RS3 and the ligand (Figure 4). In KLIFS, RS3 is named as residue #28 (αC-Glu is #24) [53]. Over 100 protein kinases have structures where RS3–ligand interactions are observed (Figure 4B). In total, more than 1000 structures with RS3–ligand interactions are available.

RS3–ligand interactions are observed with kinases containing Leu, Met, Tyr, Asn, Val, Gln, Ile or Ser in their RS3. No direct RS3–ligand contacts were observed for kinases with Ala, Cys, Phe, His and Thr. Polar interactions to RS3 are extremely rare. H-bond interaction occurs between inhibitor and RS3 only in 12 structures, including AURKA (Gln) (PDB ID: 4uzd [57]); CHK1 (Asn) (PDB IDs: 4fsq, 4fst, 4ftk, 4ftl, 4ftm, 4ftn); Haspin (Ser) (PDB ID: 6z5a); VRK1 (Tyr) (PDB IDs: 6btw, 6cfm, 6cnx, 6npn). Hydrophobic interactions are abundant, and hydrophobic contacts to Leu and Met appear in 551 and 410 structures, respectively. This is not surprising, based on their high frequency in RS3 (79%) among the available structures. Interactions to Tyr appear in 41 structures, while other RS3 residues with interactions are represented each with 10 or less structures.

Interactions to RS3 appear independent on the kinase conformation (Figure 4C,D) [55]. Based on the KinaMetrix [56], 'αC-helix in' conformations are dominating in the structures. CIDI is the most populated with 459 structures and CIDO appears in 323 structures. CODI and CODO structures with αC-helix out configuration exist in 148 and 42 structures, respectively. The ambiguous ωCD occurs in 47 structures. Leu and Met RS3–ligand interaction structures display all configurations, albeit less structures of CODI (16%), CODO (4%) and ωCD (4%) conformations exists. With less structural information containing RS3 residues the conformational representation does not cover all configurations. Nevertheless, all conformations are present in the structures with RS3–ligand interactions. As there exists distinct protein kinase conformation classifications, we also analysed the conformational distribution of RS3 contact structures with Kincore [58]. Table 1 shows the conformational distribution of these structures assigned with the Kincore. Overall, majority of the available compounds with the RS3 interactions exist in DFGin and DFGout spatial classes, covering different conformational classes (dihedrals).

Approved small-molecule kinase inhibitors and RS3

A publicly available structure exists for 49 out of the 71 FDA-approved small-molecule kinase inhibitors. From these, interaction to RS3 is displayed by 26 inhibitors (55%). These structures include targets with Met (Figure 5) and Leu RS3 residues (Figure 6). Inhibitors which exhibit RS3 interactions represent all types of kinase inhibitors that bind to the ATP-binding cleft. Of note, inhibitors of different type may engage RS3 site in a different manner. For instance, type II inhibitors, which bind the kinase in its inactive conformation, reach

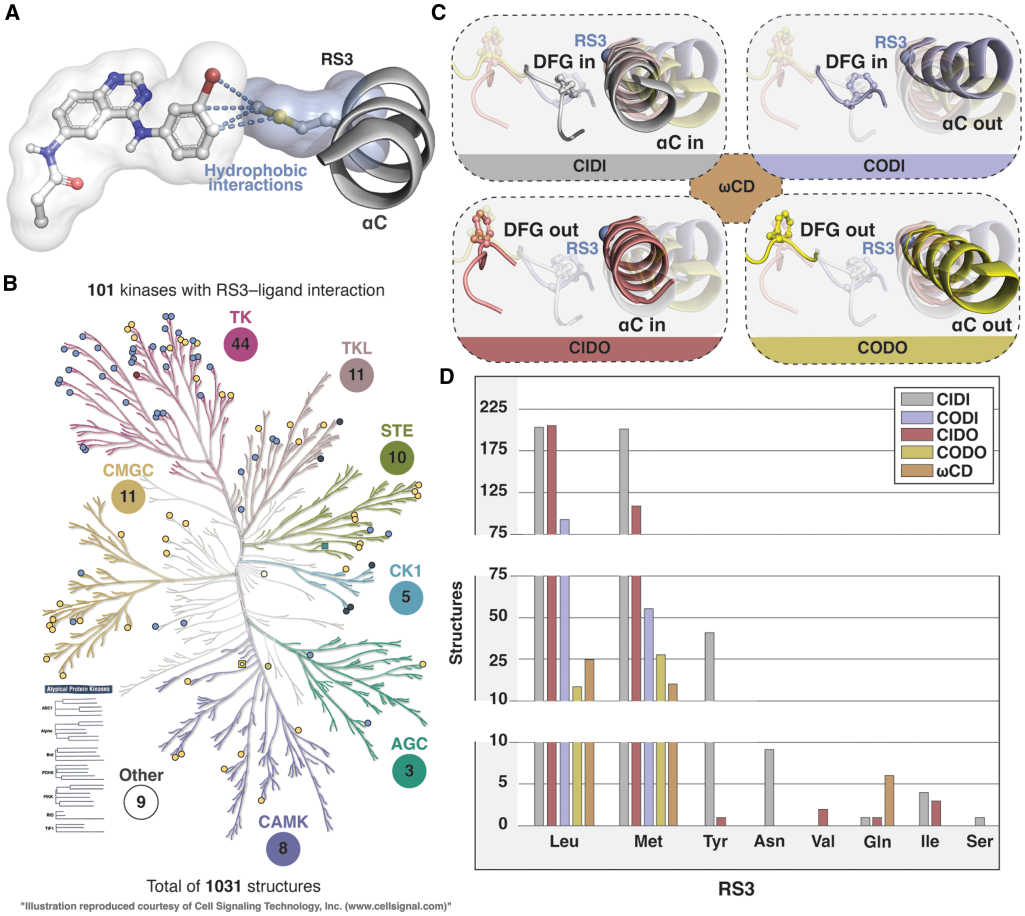


Figure 4. Publicly available structures that display RS3–ligand contacts.

(**A**) An example of a structure displaying hydrophobic contacts to RS3 (PDB ID: 4lqm [54]) (**B**) Structures with RS3–ligand interactions. Results of the KLIFS search, which was complemented with Protein–Ligand database (PLDB) (Schrödinger LLC) that identified 197 additional structures. (**C**) Conformation classification of protein kinases based on DFG and αC-helix conformations as disclosed by Ung et al.[55]. The ambiguous conformations are named as ωCD, which may represent the transition conformations between the four states. Location of RS3 Cα-atom is illustrated with a blue sphere. (**D**) Number of structures with RS3–ligand contacts and divided by RS3 residue type and KinaMetrix [56] defined kinase conformations. For 11 structures the conformation was undefined and those were excluded.

beyond the RS3 on the αC-helix side, and thereby can interact with the side chain to its 'side' or 'head' or both. Conversely, type I inhibitors, which bind to the active conformation of the kinase, interact mainly with the head of the RS3 side chain.

In occupying the region next to RS3, mainly six membered aromatic ring containing structural moieties are preferred (Figures 5, 6). Correspondingly, these aromatic rings participate in RS3 interactions in most cases. Tivozanib is the only inhibitor that presents a five-membered heteroaromatic ring in this location. Furthermore, aromatic ring attached halogens, Cl with RS3 Met (bosutinib; afatinib) and Br or I with RS3 Leu (vandetanib; trametinib, cobimetinib), display contacts to RS3. Vemurafenib and lenvatinib are the only exceptions that utilize nonaromatic moieties to interact with RS3. In addition to VEGFRs (Leu), lenvatinib also binds and inhibits FGFR-1 (Met). To both, Leu or Met, lenvatinib is in contact from its cyclopropyl urea moiety (PDB IDs: 3wzd [59], 5zv2 [60]). Vemurafenib displays contacts to RS3 Leu from aliphatic carbons of its propylsulfonamide group (PDB ID: 4rzv [61]).

Biochemical Society Transactions (2022) **50** 633–648
https://doi.org/10.1042/BST20210837

Table 1 RS3–ligand contact structures with Kincore defined conformations

Spatial Label	Dihedral label	Annotation	Leu[1]	Met[1]	Tyr	Asn	Val	Gln	Ile	Ser
DFGin	BLAminus	Active	71	96	39	9	-	-	-	1
	BLBplus	SRC-inactive	120	50	-	-	-	-	-	-
	ABAminus	Active-like	21	21	1	-	-	-	3	-
	BLBminus		36	33	-	-	-	-	-	-
	BLAplus	FGFR-inactive	23	39	-	-	-	-	-	-
	BLBtrans	CDK-inactive	-	7	-	-	-	-	-	-
	None		38	12	-	-	-	2	1	-
DFGinter	BABtrans		-	-	-	-	-	-	-	-
	None	AURKA-inactive	13	6	-	-	-	3	-	-
DFGout	BBAminus	Type-2 binding	151	101	1	-	1	1	2	-
	None	DFGout-like	59	32	-	-	1	2	1	-
None	None		13	4	-	-	-	2	-	-
Total			545	401	41	9	2	10	7	1

[1]Conformation definitions were unavailable in Kincore for six (Leu) and nine (Met) individual structures included in Figure 4D.

Some of these drugs display interactions with other kinases RS3 that are distinct from their main target. Bosutinib displays hydrophobic contacts to RS3 with MST3 (Leu) (PDB ID: 4qmn [62]), as well as dasatinib to Leu with MAPK14 (p38a) (PDB ID: 3lfa). Ponatinib displays hydrophobic interactions to Leu with RIPK2 (PDB ID: 4c8b [63]), KIT (PDB ID: 4u0i [64]) and BRAF (PDB ID: 6p3d [65]). Nilotinib to Leu with MAPK11 (p38b) (PDB ID: 3gp0). Imatinib displays contact to Leu when bound to KIT (PDB ID: 1t46 [66]) and MAPK14 (p38a) (PDB ID: 3hec [67]). Also, Pexidartinib exhibits contact to RS3 Leu with KIT (PDB ID: 7khg [68]). For inhibitors which main targets display Leu in RS3, apart from lenvatinib no Met RS3 interaction containing structures are available.

RS3 interaction is kinase dependent. Approved drugs with interactions to RS3 do not necessarily exhibit contacts to RS3 with other kinases that they bind to. For instance, gefitinib has been co-crystallized with GAK (Met), where it displays no contact to RS3 (PDB ID: 5y7z [69]). Ibrutinib has been also co-crystallized with MAP2K7 that has Val in RS3, but it does not display any contacts to this residue in this complex (PDB ID: 6yg2 [70]).

Mutations in RS3 exist rarely

According to the Catalogue of Somatic Mutations In Cancer (COSMIC; v.95) database [71], no clear tendency for mutations in RS3 exists. In total 82 kinases display at least one mutation (missense or silent) (Table 2). Only with ALK, several mutations at this location appear in the data. These mutations include, I1171N, I1171T and I1171S. For BRAF, L505H mutation is found in eight samples. Perhaps the low number of observed RS3 mutations is not surprising, due to the crucial role of this residue in the kinase function. In comparison, RS2 mutations are also rare, with BRAF F595L (13 samples in COSMIC v.95) being the most frequent in the analysed kinases. Meanwhile, RS2 flanking residues are common oncogenic drivers; for instance, BRAF V600E is found in 52 733 samples and EGFR L858R is present in 10 642 samples. Mutations at αC-helix may activate the kinase via destabilizing the kinase inactive conformation [72], but they are mainly found in other locations on the αC-helix than on RS3 [73].

In the literature, only few cases of studies including RS3 mutations have been reported. With BRAF, a secondary mutation in RS3 (L505H) induces resistance to vemurafenib [74]. Also, in another study resistance to dabrafenib or vemurafenib was demonstrated with BRAF L505H mutation [10]. Moreover, alectinib resistance was reported with I1171S and I1171N mutations in ALK [75]. The I1171N mutation was demonstrated to increase autophosphorylation level of ALK *in vitro* [76]. I1171T has been identified to induce crizotinib resistance [77]. Interestingly, these two inhibitors, crizotinib and alectinib do not display direct contacts to RS3. EGFR M766T mutation was reported to induce resistance to gefitinib and erlotinib [78].

Biochemical Society Transactions (2022) **50** 633–648
https://doi.org/10.1042/BST20210837

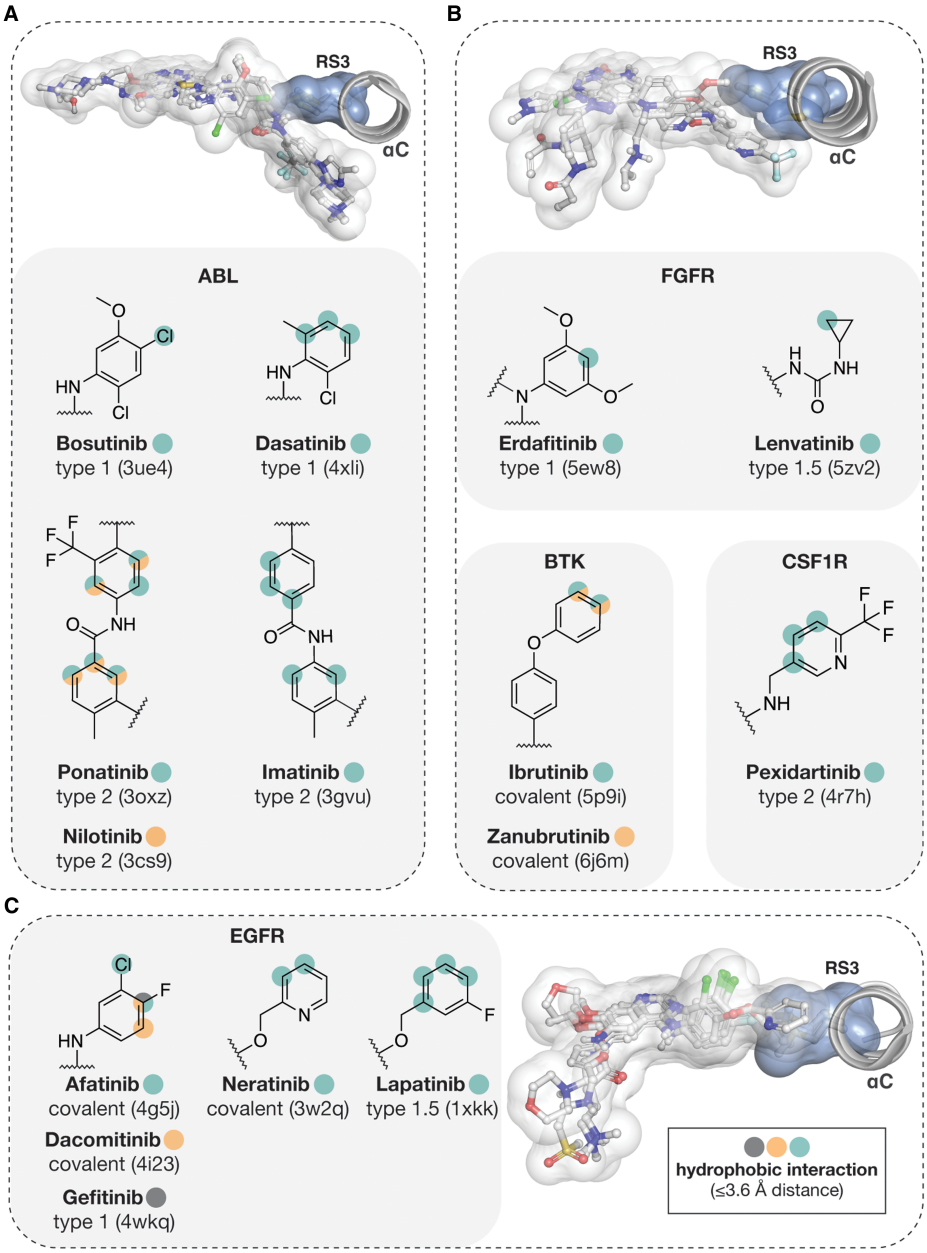


Figure 5. FDA approved small-molecule kinase inhibitors with Met RS3 contacts.

(**A**) ABL inhibitors. (**B**) FGFR, BTK and CSF1R inhibitors. (**C**) EGFR inhibitors. In 3D images, structures of the superimposed kinases are shown with their bound inhibitors. Full kinase inhibitor structures are represented with transparent white surface and stick models. RS3 Met is shown with transparent blue surface and sticks. The RS3 interacting atoms of the inhibitors are highlighted in 2D-structures of the structural moieties that are located near the RS3 residue. Hydrophobic interaction was defined based on a 3.6 Å (or shorter) distance between two hydrophobic atoms.

Biochemical Society Transactions (2022) **50** 633–648
https://doi.org/10.1042/BST20210837

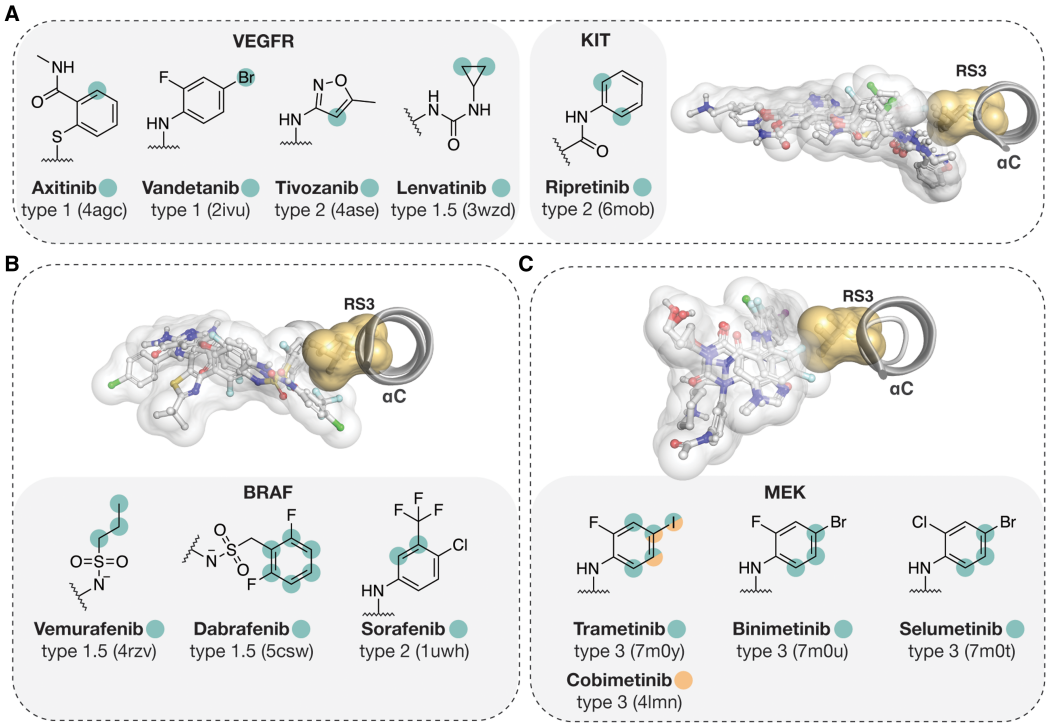


Figure 6. FDA approved small-molecule kinase inhibitors with Leu RS3 contacts.

(**A**) VEGFR and KIT inhibitors. (**B**) BRAF inhibitors. (**C**) MEK inhibitors. In 3D images, structures of the superimposed kinases are shown with their bound inhibitors. Full kinase inhibitor structures are represented with transparent white surface and stick models. RS3 Leu is shown with transparent yellow surface and sticks. The RS3 interacting atoms of the inhibitors are highlighted in 2D-structures of the structural moieties that are located near the RS3 residue. Hydrophobic interaction was defined based on a 3.6 Å (or shorter) distance between two hydrophobic atoms.

Conclusions

The function of the R-spine and the role of RS3 is quite conserved with typical protein kinases. Couple of residues are dominating RS3 in the available protein kinase domain structures. Nevertheless, also unique RS3 residues are observed, and in combination with gatekeeper even more kinase specific profiles for these residues are observed. Obviously, even with identical RS3–gatekeeper combinations the 3D-environment within this region can be quite different between two kinases. Kinase specific angles and absolute positions of these residues may provide important opportunities for selective targeting. Obviously, one must carefully consider this case-by-case, as RS3 is not targetable in all kinases. With pseudokinases [79–81], which compared with regular protein kinases can vary more in their structure in this region, the role of RS3 and its targeting would require further research.

The general understanding of RS3–ligand interactions are quite limited, even though numerous structures that contain these mainly hydrophobic interactions are available. Currently, no studies investigating specific effect of RS3 on ligand binding affinity exist that directly compare a set of ligands with selected mutations of this residue. Further research is needed to disclose the influence of RS3 residue for ligand binding and should be also extended to the cases where no direct contact between the residue and inhibitor exists. Of note, even with hydrophobic interactions (in the case of hydrophobic RS3) this should not be overlooked as these interactions may be crucial for the inhibitor binding [82]. For example, non-canonical interactions play a detrimental role in binding affinity of the ultra-potent small-molecule biotin [83]. There may be good possibilities available to optimize RS3-specific interactions, for instance, with enhanced interactions with the sulfur atom of Met [84].

Biochemical Society Transactions (2022) **50** 633–648
https://doi.org/10.1042/BST20210837

Table 2 Missense mutations in RS3 (COSMIC v.95) Part 1 of 3

				Mutation															
	Kinase	**RS3**	**RS3-contact[1]**	**N**	**Q**	**M**	**D**	**T**	**S**	**V**	**I**	**R**	**F**	**K**	**H**	**W**	**P**	**L**	**-[2]**	**Total**
	TK																			
1	ABL1	M290	+									1								1
2	ALK	I1171	+	18				11	5										2	36
3	DDR2	M629	+							2										2
4	EGFR	M766	+		1			1			1							1		4
5	FGFR2	M538	+								1	1						1		3
6	FLT1	L882	+																1	1
7	JAK2	L902	+														1			1
8	TYK2	L951	+																1	1
9	KDR	L889	+										1						1	2
10	KIT	L644	+																1	1
11	PDGFRa	M648	+								1									1
12	LCK	M292	+								4									4
13	SYK	M424	+					1												1
14	AXL	M589	-					2			1			2						5
15	EphA3	M674	-								2									2
16	EphA7	M686	-					1												1
17	EphB3	M686	-								1									1
18	ErbB3	I744	-					2												2
19	ITK	M410	-								1									1
20	JAK3	L875	-												2		1		5	8
21	ROR2	M526	-							1										1
22	ROS1	M2001	-															1		1
23	ZAP70	M390	-					1												1
	TKL																			
				N	Q	M	D	T	S	V	I	R	F	K	H	W	P	L	-	
24	MAP3K7	L81	+										1							1
25	BRAF	L505	+												8					8
26	ACVR2B	F234	-																4	4
27	MAP3K9	F195	-																3	3
28	RAF1	L397	-							1										1
29	HH498	L513	-								2		1				1			4
	STE																			
				N	Q	M	D	T	S	V	I	R	F	K	H	W	P	L	-	
30	PAK1	M319	+					1				1								2
31	PAK4	M370	+					1												1
32	STK10	L85	+														1			1
33	SLK	L83	+								1								1	2
34	MAP2K1	L118	+							1									2	3
35	MAP2K6	S103	-										2							2
36	MAP3K14	C444	-									1								1

Continued

PORTLAND PRESS

Biochemical Society Transactions (2022) **50** 633–648
https://doi.org/10.1042/BST20210837

Table 2 Missense mutations in RS3 (COSMIC v.95) Part 2 of 3

				Mutation															
	Kinase	**RS3**	**RS3-contact[1]**	**N**	**Q**	**M**	**D**	**T**	**S**	**V**	**I**	**R**	**F**	**K**	**H**	**W**	**P**	**L**	**-[2]**	**Total**
37	OXSR1	M67	-								1									1
38	PAK5	M498	-								1									1
39	TNIK	L73	-													1				1
	AGC																			
				N	Q	M	D	T	S	V	I	R	F	K	H	W	P	L	-	
40	AKT2	L204	+																1	1
41	ROCK1	M128	+								1									1
42	AKT1	L202	-																2	2
43	CDC42BPB	L128	-			1														1
44	DMPK1	L123	-														1		1	2
45	GPRK5	L238	-																1	1
46	MASTL	L85	-																1	1
47	PRKCh	L407	-																1	1
48	SGK1	L150	-						2	1			1						2	6
	CAMK																			
				N	Q	M	D	T	S	V	I	R	F	K	H	W	P	L	-	
49	CHEK1	N59	+				1													1
50	DRAK2	L84	+																1	1
51	PIM1	L93	+							1									3	4
52	RPS6KA3	L467	+								1		1							2
53	CAMK1d	L73	-																1	1
54	CAMK4	L93	-								3	1	3							7
55	MARK2	M104	-							1										1
56	MNK1	L98	-							2										2
57	MNK2	L133	-			1														1
58	SgK085	M155	-								2									2
59	PHKg2	L81	-										1		1					2
60	AMPKa2	L68	-																1	1
61	RPS6KA2	L113	-										2							2
62	SNRK	M67	-								1									1
63	STK17B	L84	-																1	1
	CDK																			
				N	Q	M	D	T	S	V	I	R	F	K	H	W	P	L	-	
64	GSK3B	M101	+															1		1
65	JNK3	M115	+							1										1
66	Erk5	L106	+																2	2
67	CDK4	L60	-		1												3			4
68	CDK6	L65	-																2	2
69	CK2a1	L85	-							1										1
70	DYRK1A	L207	-							1	1									2
71	HIPK2	L247	-			1														1

Continued

Table 2 Missense mutations in RS3 (COSMIC v.95) Part 3 of 3

				Mutation															
	Kinase	**RS3**	**RS3-contact[1]**	**N**	**Q**	**M**	**D**	**T**	**S**	**V**	**I**	**R**	**F**	**K**	**H**	**W**	**P**	**L**	**-[2]**	**Total**
72	JNK1	M77	-								1									1
	Atypical																			
				N	Q	M	D	T	S	V	I	R	F	K	H	W	P	L	-	
73	ATM	Q2729	-		1										2					3
74	ATR	M2339	-								1									1
75	ADCK3	A415	-				4													4
	Other																			
				N		M	D	T	S	V	I	R	F	K	H	W	P	L	-	
76	TTK	L575	+										1			1				2
77	Haspin	S539	-										1							1
78	NEK7	H86	-							1										1
79	TBK1	L59	-										2							2
80	TLK2	H518	-	2																2
81	ULK1	L67	-																1	1
82	WEE2	H263	-																3	3

[1]Hydrophobic interaction between ligand and RS3: +at least one structure with RS3–ligand contact available; - no contacts observed in available structures;
[2]Silent mutation.

Infrequency of mutations in RS3 may indicate a defiance against plausible point mutation in this position that could cause drug resistance [85]. Perhaps the somewhat buried location of RS3 in the quite rigid αC-helix position that offers a limited flexibility, renders the mutations in this location (at least in most cases) incompetent to drive kinase activation. This motivates further to optimize protein–ligand interactions for RS3. However, the data at hand may not necessarily cover potential drug therapy induced mutations in cancer patients. We believe that in near future, with accumulation of this data, this information will be more accessible, and a better estimate can be provided.

The full data presented in this review are freely available at https://doi.org/10.5281/zenodo.5796550

Perspectives

- The role of the conserved R-spine and RS3 residue in protein kinase function is well established.
- Protein kinases display diversity in their RS3 residue and in its surroundings. Many small-molecule protein kinase inhibitors, including approved drugs, display contacts to RS3.
- Considering the RS3 residue more carefully in the design of small-molecule kinase inhibitors may offer important advantage for the inhibitor binding and selectivity.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Biochemical Society Transactions (2022) **50** 633–648
https://doi.org/10.1042/BST20210837

Funding
This research is financially supported by Academy of Finland GeneCellNano Flagship funding 337120 and T.P. acknowledges funding by Orion Research Foundation sr.

Author Contributions
Both authors wrote the manuscript and made the figures. The original concept was conceived by T.P. T.P. directed the project.

Abbreviations
ATP, adenosine triphosphate; PLDB, Protein–Ligand Database.

References
1 Manning, G., Whyte, D.B., Martinez, R., Hunter, T. and Sudarsanam, S. (2002) The protein kinase complement of the human genome. *Science* **298**, 1912–1934 https://doi.org/10.1126/science.1075762
2 Modi, V. and Dunbrack, R.L. (2019) A structurally-validated multiple sequence alignment of 497 human protein kinase domains. *Sci. Rep.* **9**, 19790 https://doi.org/10.1038/s41598-019-56499-4
3 Berginski, M.E., Moret, N., Liu, C., Goldfarb, D., Sorger, P.K. and Gomez, S.M. (2021) The dark kinase knowledgebase: an online compendium of knowledge and experimental results of understudied kinases. *Nucleic Acids Res.* **49**, D529–D535 https://doi.org/10.1093/nar/gkaa853
4 Myasnikov, A., Zhu, H., Hixson, P., Xie, B., Yu, K., Pitre, A. et al. (2021) Structural analysis of the full-length human LRRK2. *Cell* **184**, 3519–3527.e10 https://doi.org/10.1016/j.cell.2021.05.004
5 Taylor, S.S., Wu, J., Bruystens, J.G.H., Rio, J.C.D., Lu, T.-W., Kornev, A.P. et al. (2021) From structure to the dynamic regulation of a molecular switch: a journey over 3 decades. *J. Biol. Chem.* **296**, 100746 https://doi.org/10.1016/j.jbc.2021.100746
6 Kornev, A.P., Haste, N.M., Taylor, S.S. and Eyck, L.F.T. (2006) Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proc. Natl Acad. Sci. U.S.A.* **103**, 17783–17788 https://doi.org/10.1073/pnas.0607656103
7 Meharena, H.S., Chang, P., Keshwani, M.M., Oruganty, K., Nene, A.K., Kannan, N. et al. (2013) Deciphering the structural basis of eukaryotic protein kinase regulation. *PLoS Biol.* **11**, e1001680 https://doi.org/10.1371/journal.pbio.1001680
8 Taylor, S.S., Shaw, A.S., Kannan, N. and Kornev, A.P. (2015) Integration of signaling in the kinome: architecture and regulation of the αC helix. *Biochim. Biophys. Acta Proteins Proteomics* **1854**, 1567–1574 https://doi.org/10.1016/j.bbapap.2015.04.007
9 Hu, J., Ahuja, L.G., Meharena, H.S., Kannan, N., Kornev, A.P., Taylor, S.S. et al. (2015) Kinase regulation by hydrophobic spine assembly in cancer. *Mol. Cell. Biol.* **35**, 264–276 https://doi.org/10.1128/MCB.00943-14
10 Yap, J., Deepak, R.N.V.K., Tian, Z., Ng, W.H., Goh, K.C., Foo, A. et al. (2021) The stability of R-spine defines RAF inhibitor resistance: a comprehensive analysis of oncogenic BRAF mutants with in-frame insertion of αC-β4 loop. *Sci. Adv.* **7**, eabg0390 https://doi.org/10.1126/sciadv.abg0390
11 Liu, Y., Shah, K., Yang, F., Witucki, L. and Shokat, K.M. (1998) A molecular gate which controls unnatural ATP analogue recognition by the tyrosine kinase v-Src. *Bioorg. Med. Chem.* **6**, 1219–1226 https://doi.org/10.1016/S0968-0896(98)00099-6
12 Nussinov, R., Zhang, M., Maloney, R., Tsai, C.-J., Yavuz, B.R., Tuncbag, N. et al. (2021) Mechanism of activation and the rewired network: new drug design concepts. *Med. Res. Rev.* **42**, 770–799 https://doi.org/10.1002/med.21863
13 Shih, H.-P., Zhang, X. and Aronov, A.M. (2018) Drug discovery effectiveness from the standpoint of therapeutic mechanisms and indications. *Nat. Rev. Drug Discov.* **17**, 19–33 https://doi.org/10.1038/nrd.2017.194
14 Bournez, C., Carles, F., Peyrat, G., Aci-Sèche, S., Bourg, S., Meyer, C. et al. (2020) Comparative assessment of protein kinase inhibitors in public databases and in PKIDB. *Molecules* **25**, 3226 https://doi.org/10.3390/molecules25143226
15 Carles, F., Bourg, S., Meyer, C. and Bonnet, P. (2018) PKIDB: a curated, annotated and updated database of protein kinase inhibitors in clinical trials. *Molecules* **23**, 908 https://doi.org/10.3390/molecules23040908
16 Ayala-Aguilera, C.C., Valero, T., Lorente-Macías, Á., Baillache, D.J., Croke, S. and Unciti-Broceta, A. (2021) Small molecule kinase inhibitor drugs (1995–2021): medical indication, pharmacology, and synthesis. *J. Med. Chem.* **65**, 1047–1131 https://doi.org/10.1021/acs.jmedchem.1c00963
17 Cohen, P., Cross, D. and Jänne, P.A. (2021) Kinase drug discovery 20 years after imatinib: progress and future directions. *Nat. Rev. Drug Discov.* **20**, 551–569 https://doi.org/10.1038/s41573-021-00195-4
18 Attwood, M.M., Fabbro, D., Sokolov, A.V., Knapp, S. and Schiöth, H.B. (2021) Trends in kinase drug discovery: targets, indications and inhibitor design. *Nat. Rev. Drug Discov.* **20**, 839–861 https://doi.org/10.1038/s41573-021-00252-y
19 Ferguson, F.M. and Gray, N.S. (2018) Kinase inhibitors: the road ahead. *Nat. Rev. Drug Discov.* **17**, 353–377 https://doi.org/10.1038/nrd.2018.21
20 Cheung, J., Ginter, C., Cassidy, M., Franklin, M.C., Rudolph, M.J., Robine, N. et al. (2015) Structural insights into mis-regulation of protein kinase A in human tumors. *Proc. Natl Acad. Sci. U.S.A.* **112**, 1374–1379 https://doi.org/10.1073/pnas.1424206112
21 Qin, J., Xie, P., Ventocilla, C., Zhou, G., Vultur, A., Chen, Q. et al. (2012) Identification of a novel family of BRAFV600E inhibitors. *J. Med. Chem.* **55**, 5220–5230 https://doi.org/10.1021/jm3004416
22 Wan, P.T.C., Garnett, M.J., Roe, S.M., Lee, S., Niculescu-Duvaz, D., Good, V.M. et al. (2004) Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. *Cell* **116**, 855–867 https://doi.org/10.1016/S0092-8674(04)00215-6
23 Kozlowski, L.P. (2017) Proteome-pI: proteome isoelectric point database. *Nucleic Acids Res.* **45**, D1112–D1116 https://doi.org/10.1093/nar/gkw978
24 Biswas, K.M., DeVido, D.R. and Dorsey, J.G. (2003) Evaluation of methods for measuring amino acid hydrophobicities and interactions. *J. Chromatogr. A* **1000**, 637–655 https://doi.org/10.1016/S0021-9673(03)00182-1
25 Simm, S., Einloft, J., Mirus, O. and Schleiff, E. (2016) 50 years of amino acid hydrophobicity scales: revisiting the capacity for peptide classification. *Biol. Res.* **49**, 31 https://doi.org/10.1186/s40659-016-0092-5
26 Roskoski, R. (2019) Small molecule inhibitors targeting the EGFR/ErbB family of protein-tyrosine kinases in human cancers. *Pharmacol. Res.* **139**, 395–411 https://doi.org/10.1016/j.phrs.2018.11.014

27 Shi, F., Telesco, S.E., Liu, Y., Radhakrishnan, R. and Lemmon, M.A. (2010) Erbb3/HER3 intracellular domain is competent to bind ATP and catalyze autophosphorylation. *Proc. Natl Acad. Sci. U.S.A.* **107**, 7692–7697 https://doi.org/10.1073/pnas.1002753107
28 Buchanan, J.L., Newcomb, J.R., Carney, D.P., Chaffee, S.C., Chai, L., Cupples, R. et al. (2011) Discovery of 2,4-bis-arylamino-1,3-pyrimidines as insulin-like growth factor-1 receptor (IGF-1R) inhibitors. *Bioorg. Med. Chem. Lett.* **21**, 2394–2399 https://doi.org/10.1016/j.bmcl.2011.02.075
29 Eid, S., Turk, S., Volkamer, A., Rippmann, F. and Fulle, S. (2017) Kinmap: a web-based tool for interactive navigation through human kinome data. *BMC Bioinformatics* **18**, 16 https://doi.org/10.1186/s12859-016-1433-7
30 Matsumoto, T., Kinoshita, T., Kirii, Y., Yokota, K., Hamada, K. and Tada, T. (2010) Crystal structures of MKK4 kinase domain reveal that substrate peptide binds to an allosteric site and induces an auto-inhibition state. *Biochem. Biophys. Res. Commun.* **400**, 369–373 https://doi.org/10.1016/j.bbrc.2010.08.071
31 Sogabe, Y., Hashimoto, T., Matsumoto, T., Kirii, Y., Sawa, M. and Kinoshita, T. (2016) A crucial role of Cys218 in configuring an unprecedented auto-inhibition form of MAP2K7. *Biochem. Biophys. Res. Commun.* **473**, 476–481 https://doi.org/10.1016/j.bbrc.2016.03.036
32 Matsumoto, T., Kinoshita, T., Matsuzaka, H., Nakai, R., Kirii, Y., Yokota, K. et al. (2012) Crystal structure of non-phosphorylated MAP2K6 in a putative auto-inhibition state. *J. Biochem.* **151**, 541–549 https://doi.org/10.1093/jb/mvs023
33 Shevchenko, E., Poso, A. and Pantsar, T. (2020) The autoinhibited state of MKK4: phosphorylation, putative dimerization and R134W mutant studied by molecular dynamics simulations. *Comput. Struct. Biotechnol. J.* **18**, 2687–2698 https://doi.org/10.1016/j.csbj.2020.09.017
34 Murphy, J.M., Nakatani, Y., Jamieson, S.A., Dai, W., Lucet, I.S. and Mace, P.D. (2015) Molecular mechanism of CCAAT-enhancer binding protein recruitment by the TRIB1 pseudokinase. *Structure* **23**, 2111–2121 https://doi.org/10.1016/j.str.2015.08.017
35 Jamieson, S.A., Ruan, Z., Burgess, A.E., Curry, J.R., McMillan, H.D., Brewster, J.L. et al. (2018) Substrate binding allosterically relieves autoinhibition of the pseudokinase TRIB1. *Sci Signal.* **11**, eaau0597 https://doi.org/10.1126/scisignal.aau0597
36 Asquith, C.R.M., Murray, N.H. and Pagliarini, D.J. (2019) ADCK3/COQ8A: the choice target of the UbiB protein kinase-like family. *Nat. Rev. Drug Discov.* **18**, 815–815 https://doi.org/10.1038/d41573-019-00158-w
37 Cyphers, S., Ruff, E.F., Behr, J.M., Chodera, J.D. and Levinson, N.M. (2017) A water-mediated allosteric network governs activation of aurora kinase A. *Nat. Chem. Biol.* **13**, 402–408 https://doi.org/10.1038/nchembio.2296
38 Xu, D., Matsumoto, M.L., McKenzie, B.S. and Zarrin, A.A. (2018) TPL2 kinase action and control of inflammation. *Pharmacol. Res.* **129**, 188–193 https://doi.org/10.1016/j.phrs.2017.11.031
39 Lee, J.-H. and Paull, T.T. (2021) Cellular functions of the protein kinase ATM and their relevance to human disease. *Nat. Rev. Mol. Cell Biol.* **22**, 796–814 https://doi.org/10.1038/s41580-021-00394-2
40 Zeqiraj, E., Filippi, B.M., Goldie, S., Navratilova, I., Boudeau, J., Deak, M. et al. (2009) ATP and MO25alpha regulate the conformational state of the STRADalpha pseudokinase and activation of the LKB1 tumour suppressor. *PLoS Biol.* **7**, e1000126 https://doi.org/10.1371/journal.pbio.1000126
41 Zeqiraj, E., Filippi, B.M., Deak, M., Alessi, D.R. and van Aalten, D.M.F. (2009) Structure of the LKB1-STRAD-MO25 complex reveals an allosteric mechanism of kinase activation. *Science* **326**, 1707–1711 https://doi.org/10.1126/science.1178377
42 Maiolica, A., de Medina-Redondo, M., Schoof, E.M., Chaikuad, A., Villa, F., Gatti, M. et al. (2014) Modulation of the chromatin phosphoproteome by the haspin protein kinase*. *Mol. Cell. Proteomics* **13**, 1724–1740 https://doi.org/10.1074/mcp.M113.034819
43 Klomp, J.E., Lee, Y.S., Goodwin, C.M., Papke, B., Klomp, J.A., Waters, A.M. et al. (2021) CHK1 protects oncogenic KRAS-expressing cells from DNA damage and is a target for pancreatic cancer treatment. *Cell Rep.* **37**, 110060 https://doi.org/10.1016/j.celrep.2021.110060
44 Preuss, F., Chatterjee, D., Mathea, S., Shrestha, S., St-Germain, J., Saha, M. et al. (2020) Nucleotide binding, evolutionary insights, and interaction partners of the pseudokinase Unc-51-like kinase 4. *Structure* **28**, 1184–1196.e6 https://doi.org/10.1016/j.str.2020.07.016
45 Smaill, J.B., Lee, H.H., Palmer, B.D., Thompson, A.M., Squire, C.J., Baker, E.N. et al. (2008) Synthesis and structure-activity relationships of soluble 8-substituted 4-(2-chlorophenyl)-9-hydroxypyrrolo[3,4-c]carbazole-1,3(2H,6H)-diones as inhibitors of the Wee1 and Chk1 checkpoint kinases. *Bioorg. Med. Chem. Lett.* **18**, 929–933 https://doi.org/10.1016/j.bmcl.2007.12.046
46 Zhu, J.-Y., Cuellar, R.A., Berndt, N., Lee, H.E., Olesen, S.H., Martin, M.P. et al. (2017) Structural basis of wee kinases functionality and inactivation by diverse small molecule inhibitors. *J. Med. Chem.* **60**, 7863–7875 https://doi.org/10.1021/acs.jmedchem.7b00996
47 Ward, R.A., Bethel, P., Cook, C., Davies, E., Debreczeni, J.E., Fairley, G. et al. (2017) Structure-guided discovery of potent and selective inhibitors of ERK1/2 from a modestly active and promiscuous chemical start point. *J. Med. Chem.* **60**, 3438–3450 https://doi.org/10.1021/acs.jmedchem.7b00267
48 Chaikuad, A., Tacconi, E.M.C., Zimmer, J., Liang, Y., Gray, N.S., Tarsounas, M. et al. (2014) A unique inhibitor binding site in ERK1/2 is associated with slow binding kinetics. *Nat. Chem. Biol.* **10**, 853–860 https://doi.org/10.1038/nchembio.1629
49 Schröder, M., Filippakopoulos, P., Schwalm, M.P., Ferrer, C.A., Drewry, D.H., Knapp, S. et al. (2020) Crystal structure and inhibitor identifications reveal targeting opportunity for the atypical MAPK kinase ERK3. *Int. J. Mol. Sci.* **21**, 7953 https://doi.org/10.3390/ijms21217953
50 Bandaranayake, R.M., Ungureanu, D., Shan, Y., Shaw, D.E., Silvennoinen, O. and Hubbard, S.R. (2012) Crystal structures of the JAK2 pseudokinase domain and the pathogenic mutant V617F. *Nat. Struct. Mol. Biol.* **19**, 754–759 https://doi.org/10.1038/nsmb.2348
51 Toms, A.V., Deshpande, A., McNally, R., Jeong, Y., Rogers, J.M., Kim, C.U. et al. (2013) Structure of a pseudokinase-domain switch that controls oncogenic activation of Jak kinases. *Nat. Struct. Mol. Biol.* **20**, 1221–1223 https://doi.org/10.1038/nsmb.2673
52 Kanev, G.K., de Graaf, C., Westerman, B.A., de Esch, I.J.P. and Kooistra, A.J. (2021) KLIFS: an overhaul after the first 5 years of supporting kinase research. *Nucleic Acids Res.* **49**, D562–D569 https://doi.org/10.1093/nar/gkaa895
53 van Linden, O.P.J., Kooistra, A.J., Leurs, R., de Esch, I.J.P. and de Graaf, C. (2014) KLIFS: a knowledge-based structural database To navigate kinase–ligand interaction space. *J. Med. Chem.* **57**, 249–277 https://doi.org/10.1021/jm400378w
54 Yasuda, H., Park, E., Yun, C.-H., Sng, N.J., Lucena-Araujo, A.R., Yeo, W.-L. et al. (2013) Structural, biochemical, and clinical characterization of epidermal growth factor receptor (EGFR) exon 20 insertion mutations in lung cancer. *Sci. Transl. Med.* **5**, 216ra177 https://doi.org/10.1126/scitranslmed.3007205
55 Ung, P.M.-U., Rahman, R. and Schlessinger, A. (2018) Redefining the protein kinase conformational space with machine learning. *Cell Chem. Biol.* **25**, 916–924.e2 https://doi.org/10.1016/j.chembiol.2018.05.002
56 Rahman, R., Ung, P.M.-U. and Schlessinger, A. (2019) Kinametrix: a web resource to investigate kinase conformations and inhibitor space. *Nucleic Acids Res.* **47**, D361–D366 https://doi.org/10.1093/nar/gky916

Biochemical Society Transactions (2022) **50** 633–648
https://doi.org/10.1042/BST20210837

57 Carry, J.-C., Clerc, F., Minoux, H., Schio, L., Mauger, J., Nair, A. et al. (2015) SAR156497, an exquisitely selective inhibitor of aurora kinases. *J. Med. Chem.* **58**, 362–375 https://doi.org/10.1021/jm501326k

58 Modi, V. and Dunbrack, R.L. (2022) Kincore: a web resource for structural classification of protein kinases and their inhibitors. *Nucleic Acids Res.* **50**, D654–D664 https://doi.org/10.1093/nar/gkab920

59 Okamoto, K., Ikemori-Kawada, M., Jestel, A., von König, K., Funahashi, Y., Matsushima, T. et al. (2015) Distinct binding mode of multikinase inhibitor lenvatinib revealed by biochemical characterization. *ACS Med. Chem. Lett.* **6**, 89–94 https://doi.org/10.1021/ml500394m

60 Matsuki, M., Hoshi, T., Yamamoto, Y., Ikemori-Kawada, M., Minoshima, Y., Funahashi, Y. et al. (2018) Lenvatinib inhibits angiogenesis and tumor fibroblast growth factor signaling pathways in human hepatocellular carcinoma models. *Cancer Med.* **7**, 2641–2653 https://doi.org/10.1002/cam4.1517

61 Karoulia, Z., Wu, Y., Ahmed, T.A., Xin, Q., Bollard, J., Krepler, C. et al. (2016) An integrated model of RAF inhibitor action predicts inhibitor activity against oncogenic BRAF signaling. *Cancer Cell* **30**, 485–498 https://doi.org/10.1016/j.ccell.2016.06.024

62 Olesen, S.H., Zhu, J.-Y., Martin, M.P. and Schönbrunn, E. (2016) Discovery of diverse small-molecule inhibitors of mammalian Sterile20-like kinase3 (MST3). *ChemMedChem* **11**, 1137–1144 https://doi.org/10.1002/cmdc.201600115

63 Canning, P., Ruan, Q., Schwerd, T., Hrdinka, M., Maki, J.L., Saleh, D. et al. (2015) Inflammatory signaling by NOD-RIPK2 Is inhibited by clinically relevant type II kinase inhibitors. *Chem. Biol.* **22**, 1174–1184 https://doi.org/10.1016/j.chembiol.2015.07.017

64 Garner, A.P., Gozgit, J.M., Anjum, R., Vodala, S., Schrock, A., Zhou, T. et al. (2014) Ponatinib inhibits polyclonal drug-resistant KIT oncoproteins and shows therapeutic potential in heavily pretreated gastrointestinal stromal tumor (GIST) patients. *Clin. Cancer Res.* **20**, 5745–5755 https://doi.org/10.1158/1078-0432.CCR-14-1397

65 Cotto-Rios, X.M., Agianian, B., Gitego, N., Zacharioudakis, E., Giricz, O., Wu, Y. et al. (2020) Inhibitors of BRAF dimers using an allosteric site. *Nat. Commun.* **11**, 4370 https://doi.org/10.1038/s41467-020-18123-2

66 Mol, C.D., Dougan, D.R., Schneider, T.R., Skene, R.J., Kraus, M.L., Scheibe, D.N. et al. (2004) Structural basis for the autoinhibition and STI-571 inhibition of c-Kit tyrosine kinase. *J. Biol. Chem.* **279**, 31655–31663 https://doi.org/10.1074/jbc.M403319200

67 Namboodiri, H.V., Bukhtiyarova, M., Ramcharan, J., Karpusas, M., Lee, Y. and Springman, E.B. (2010) Analysis of imatinib and sorafenib binding to p38alpha compared with c-Abl and b-Raf provides structural insights for understanding the selectivity of inhibitors targeting the DFG-out form of protein kinases. *Biochemistry* **49**, 3611–3618 https://doi.org/10.1021/bi100070r

68 Wagner, A.J., Severson, P.L., Shields, A.F., Patnaik, A., Chugh, R., Tinoco, G. et al. (2021) Association of combination of conformation-specific KIT inhibitors with clinical benefit in patients with refractory gastrointestinal stromal tumors: a phase 1b/2a nonrandomized clinical trial. *JAMA Oncol.* **7**, 1343–1350 https://doi.org/10.1001/jamaoncol.2021.2086

69 Ohbayashi, N., Murayama, K., Kato-Murayama, M., Kukimoto-Niino, M., Uejima, T., Matsuda, T. et al. (2018) Structural basis for the inhibition of cyclin G-associated kinase by gefitinib. *ChemistryOpen* **7**, 713–719 https://doi.org/10.1002/open.201800177

70 Schröder, M., Tan, L., Wang, J., Liang, Y., Gray, N.S., Knapp, S. et al. (2020) Catalytic domain plasticity of MKK7 reveals structural mechanisms of allosteric activation and diverse targeting opportunities. *Cell Chem. Biol.* **27**, 1285–1295.e4 https://doi.org/10.1016/j.chembiol.2020.07.014

71 Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N. et al. (2019) COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* **47**, D941–D947 https://doi.org/10.1093/nar/gky1015

72 Barouch-Bentov, R. and Sauer, K. (2011) Mechanisms of drug resistance in kinases. *Expert Opin. Investig. Drugs.* **20**, 153–208 https://doi.org/10.1517/13543784.2011.546344

73 Kim, P., Li, H., Wang, J. and Zhao, Z. (2021) Landscape of drug-resistance mutations in kinase regulatory hotspots. *Brief Bioinform.* **22**, bbaa108 https://doi.org/10.1093/bib/bbaa108

74 Hoogstraat, M., Hooijdonk, C.G.G., Ubink, I., Besselink, N.J.M., Pieterse, M., Veldhuis, W. et al. (2015) Detailed imaging and genetic analysis reveal a secondary BRAFL505H resistance mutation and extensive intrapatient heterogeneity in metastatic BRAF mutant melanoma patients treated with vemurafenib. *Pigment Cell Melanoma Res.* **28**, 318–323 https://doi.org/10.1111/pcmr.12347

75 Ou, S.-H.I., Klempner, S.J., Greenbowe, J.R., Azada, M., Schrock, A.B., Ali, S.M. et al. (2014) Identification of a novel HIP1-ALK fusion variant in non-small-cell lung cancer (NSCLC) and discovery of ALK I1171 (I1171N/S) mutations in two ALK-rearranged NSCLC patients with resistance to Alectinib. *J. Thorac. Oncol.* **9**, 1821–1825 https://doi.org/10.1097/JTO.0000000000000368

76 Bresler, S.C., Weiser, D.A., Huwe, P.J., Park, J.H., Krytska, K., Ryles, H. et al. (2014) ALK mutations confer differential oncogenic activation and sensitivity to ALK inhibition therapy in neuroblastoma. *Cancer Cell* **26**, 682–694 https://doi.org/10.1016/j.ccell.2014.09.019

77 Friboulet, L., Li, N., Katayama, R., Lee, C.C., Gainor, J.F., Crystal, A.S. et al. (2014) The ALK inhibitor ceritinib overcomes crizotinib resistance in non-small cell lung cancer. *Cancer Discov.* **4**, 662–673 https://doi.org/10.1158/2159-8290.CD-13-0846

78 Ruan, Z., Katiyar, S. and Kannan, N. (2017) Computational and experimental characterization of patient derived mutations reveal an unusual mode of regulatory spine assembly and drug sensitivity in EGFR kinase. *Biochemistry* **56**, 22–32 https://doi.org/10.1021/acs.biochem.6b00572

79 Mace, P.D. and Murphy, J.M. (2021) There's more to death than life: noncatalytic functions in kinase and pseudokinase signaling. *J. Biol. Chem.* **296**, 100705 https://doi.org/10.1016/j.jbc.2021.100705

80 Shrestha, S., Byrne, D.P., Harris, J.A., Kannan, N. and Eyers, P.A. (2020) Cataloguing the dead: breathing new life into pseudokinase research. *FEBS J.* **287**, 4150–4169 https://doi.org/10.1111/febs.15246

81 Kung, J.E. and Jura, N. (2019) Prospects for pharmacological targeting of pseudokinases. *Nat. Rev. Drug Discov.* **18**, 501–526 https://doi.org/10.1038/s41573-019-0018-3

82 Bissantz, C., Kuhn, B. and Stahl, M. (2010) A medicinal chemist's guide to molecular interactions. *J. Med. Chem.* **53**, 5061–5084 https://doi.org/10.1021/jm100112j

83 McConnell, D.B. (2021) Biotin's lessons in drug design. *J. Med. Chem.* **64**, 16319–16327 https://doi.org/10.1021/acs.jmedchem.1c00975

84 Beno, B.R., Yeung, K.-S., Bartberger, M.D., Pennington, L.D. and Meanwell, N.A. (2015) A survey of the role of noncovalent sulfur interactions in drug design. *J. Med. Chem.* **58**, 4383–4438 https://doi.org/10.1021/jm501853m

85 Ward, R.A., Fawell, S., Floc'h, N., Flemington, V., McKerrecher, D. and Smith, P.D. (2021) Challenges and opportunities in cancer drug resistance. *Chem. Rev.* **121**, 3297–3351 https://doi.org/10.1021/acs.chemrev.0c00383

3.3 Publication III

Do Go Chasing Waterfalls: Enoyl Reductase (FabI) in Complex with Inhibitors Stabilizes the Tetrameric Structure and Opens Water Channels

V. Gonçalves Maltarollo [a], E. Shevchenko [b,c,d], I. D. de Miranda Lima [e], E. A. Cino [e], G. M. Ferreira [f], A. Poso [b,c,d,g], T. Kronenberger [b,c,d]

[a] Departamento de Produtos Farmacêuticos, Faculdade de Farmácia, Universidade Federal de Minas Gerais (UFMG), 31270-901 Belo Horizonte, Minas Gerais, Brazil

[b] Institute of Pharmacy, Pharmaceutical/Medicinal Chemistry and Tübingen Center for Academic Drug Discovery, Eberhard Karls University Tübingen, Auf der Morgenstelle 8, 72076 Tübingen, Germany

[c] Department of Internal Medicine VIII, University Hospital Tübingen, Otfried-Müller-Strasse 14, 72076 Tübingen, Germany

[d] Tübingen Center for Academic Drug Discovery & Development (TüCAD2), 72076 Tübingen, Germany

[e] Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, Brazil

[f] Department of Clinical and Toxicological Analyses, School of Pharmaceutical Sciences, University of Sao Paulo, Av Prof Lineu Prestes 580, 05508-000 São Paulo, Brazil

[g] School of Pharmacy, University of Eastern Finland, Yliopistonranta 1C, 70210 Kuopio, Finland

Journal of Chemical Information and Modeling

2022, DOI:10.1021/acs.jcim.2c01178

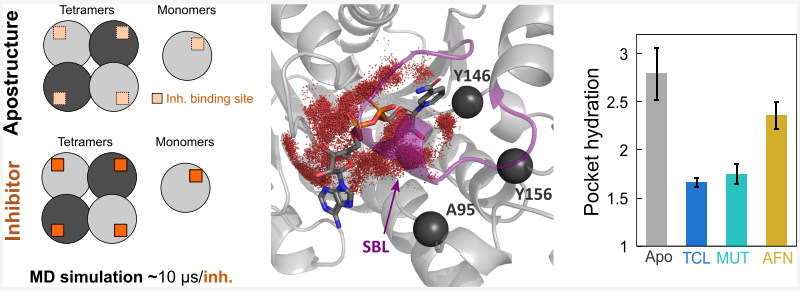
JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

Do Go Chasing Waterfalls: Enoyl Reductase (FabI) in Complex with Inhibitors Stabilizes the Tetrameric Structure and Opens Water Channels

Vinicius Gonçalves Maltarollo, Ekaterina Shevchenko, Igor Daniel de Miranda Lima, Elio A. Cino, Glaucio Monteiro Ferreira, Antti Poso,* and Thales Kronenberger*

Cite This: J. Chem. Inf. Model. 2022, 62, 5746−5761 Read Online



ABSTRACT: The enzyme enoyl-ACP reductase (FabI) is the limiting step of the membrane's fatty acid biosynthesis in bacteria and a druggable target for novel antibacterial agents. The FabI active form is a homotetramer, which displays the highest affinity to inhibitors. Herein, molecular dynamics studies were carried out using the structure of FabI in complex with known inhibitors to investigate their effects on tetramerization. Our results suggest that multimerization is essential for the integrity of the catalytic site and that inhibitor binding enables the multimerization by stabilizing the substrate binding loop (SBL, L:195-200) coupled with changes in the H4/5 (QR interface). We also observed that AFN-1252 (naphtpyridinone derivative) promotes unique conformational changes affecting monomer−monomer interfaces. These changes are induced by AFN-1252 interaction with key residues in the binding sites (Ala95, Tyr146, and Tyr156). In addition, the analysis of water trajectories indicated that AFN-1252 complexes allow more water molecules to enter the binding site than triclosan and MUT056399 complexes. FabI−AFN-1252 simulations show accumulation of water molecules near the Tyr146/147 pocket, which can become a hotspot to the design of novel FabI inhibitors.

1. INTRODUCTION

In the past decades, several new antibiotic classes have been approved by the Food and Drug Administration (FDA); however, most of them have limited usefulness against the most problematic "ESKAPE" pathogens.[1,2] The ESKAPE group, namely, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* species, contribute to ∼75% of infections and deaths from antibiotic-resistant bacteria.[3] Further, most of the recently discovered antibiotics target Gram-positive bacteria, as the Gram-negative pathogens have more impermeable cell membranes and promiscuous efflux pumps. In this sense, expanding the knowledge of relevant drugs for both groups and establishing comparative studies are relevant.

Enoyl-ACP reductase (FabI, Figure 1A) is a bacterial enzyme responsible for the limiting step of the cellular membrane's fatty acid biosynthesis. FabI catalyzes the final reductive step in the bacterial fatty acid synthesis cycle, with NAD(*P*)H as a hydride donor to an enoyl-ACP substrate and releasing the oxidized coenzyme ($NAD(P)^+$) with the product acyl-ACP. In that sense, FabI plays a major role in the elongation cycle of the bacterial fatty acid biosynthesis pathway (FAS-II) and the regulation of bacterial membrane homeostasis.[4] The FAS-II pathway is essential for Gram-negative pathogens' survival.[5] Its inhibition cannot be bypassed during infections through the uptake of exogenous nutrients.[5] Since mammalians rely on the FAS-I multicomplex, which has no similarity to FAS-II, the latter becomes the attractive target for antimicrobial development.

Received: September 20, 2022
Published: November 7, 2022

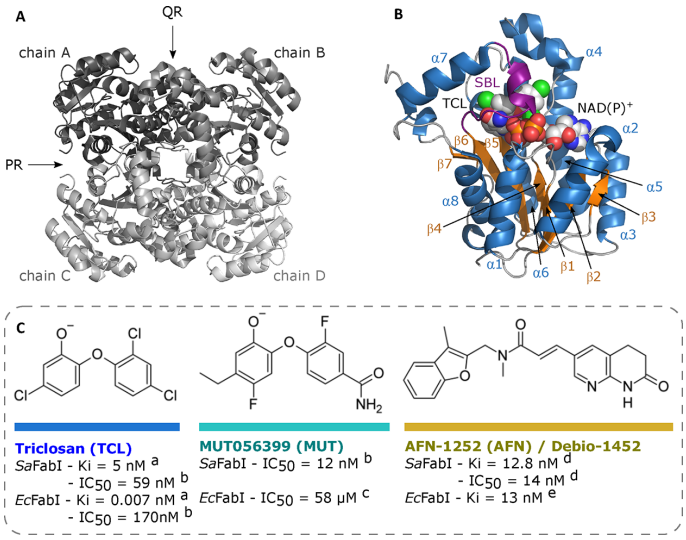


Figure 1. Tetrameric structure of FabI with the four chains represented in different shades of gray and the identification of the monomer−monomer (PR) and dimer−dimer (QR) interface positions (A). Structure of a FabI monomer identifying the eight helices (blue), seven strands (orange), inhibitor and NAD(P)$^+$ binding sites (represented as spheres models), and the substrate binding loop (SBL, in purple, L:195-200) (B). Structure and activities of studied FabI inhibitors (C). The reported activities were retrieved from (a) Lu and Tonge, 2008;[17] (b) Gerusz *et al.*, 2012,[18] (c) Bryskier, 2010;[19] (d) Kaplan *et al.*, 2012;[8] and (e) Parker *et al.*, 2020.[20]

FAS-II and, therefore, FabI relevance is showcased by the success of triclosan and isoniazid.[6,7]

The FabI structure is composed of a seven-stranded parallel β-sheet and eight α-helices forming a single domain (Figure 1B). Moreover, the Tyr-Xaa$_6$-Lys catalytic dyad is conserved in practically all FabI structures and placed in similar active site locations. The FabI catalytic site is covered by a flexible loop (substrate binding loop, or SBL, Figure 1B), which is very flexible and has been reported in various distinct conformations in different experiments.[8−10] Several FabI structures from *S. aureus* (*Sa*), *Escherichia coli* (*Ec*), and other organisms were reported in the past years comprising monomers,[8] dimers,[9,11,12] and tetramers.[10,13,14] FabI homodimerization relies on the PR interface composed of the interactions between helix 8 and β-sheet 7 of two monomers. Meanwhile, the QR interface is composed of the contact between helices 4/5 and 6 of two chains and is responsible for the final tetramerization (Figure 1A,B). PR and QR nomenclature was defined in previous work[15] and used this way since then to locate the mentioned structures.

To date, all available structures in active or inhibited states were reported as homotetramers, except the apo form of *Sa*FabI (PDB IDs: 3GNS and 3GNT[11]) and *Bacillus cereus* FabI (PDB ID: 3OJE[12]). This highlights the relationship between the ligand-bound state and tetramer formation. It is speculated that the cooperativity behavior of the overall tetramer depends on the interactions between the individual subunits. This is supported by mutations in the interface of tetramer subunits that promote resistance to FabI in clinical strains, such as Gly113Val of *S. aureus* enzyme.[16]

Triclosan (TCL, Figure 1C) is the most well-known FabI inhibitor with antibacterial properties. TCL is widely employed in healthcare products, and until the end of the 1990s, it was believed to act directly on the bacterial membrane without relying on specific protein targets. This changed with the work from McMurry *et al.* (1998) describing TCL as a FabI inhibitor.[6,13] Since then, many novel classes of FabI inhibitors with antibacterial properties and structure−activity studies of those compounds were reported.[17,21,22] We highlight two relevant inhibitors: MUT056399 (MUT) and Debio-1452, formerly known as AFN-1252 (AFN), which reached clinical trials as a treatment for bacterial infections.[23−26] TCL and MUT are diphenyl ether derivatives, with a phenolate that interacts with the catalytic tyrosine residue. AFN poses a lactam group (present in the naphtpyridinone ring) that forms an extra hydrogen bond with Ala97 from *Sa*FabI. MUT, despite being classified according to the same chemical class as TCL, was also designed to have an amide side chain to mimic this interaction with Ala97.[18]

The binding of these inhibitors to FabI leads to the SBL organization in a closed conformation, which until recently, was believed to be the only inhibition mode for this enzyme. However, recently, oligomeric transition was shown to be coupled with ligand binding in *Sa*FabI,[15] which supports a more complex inhibition mechanism. This cooperativity is supported by the increased ligand affinity depending on FabI's oligomerization state[27] and the observation of TCL-resistant mutants displaying substitutions in the dimer interface.[8] Until now, computational studies have focused only on the FabI monomeric state or employed subrelevant timescales to study large conformational changes.[28−30]

Our work aimed to investigate the binding mode of selected FabI inhibitors using simulations of a full tetramer in comparison with monomers. We considered that different ligands may affect the oligomerization of FabI by indirectly triggering protein−protein interactions ranging from binding site residues toward the interface.

2. MATERIALS AND METHODS

In this work, molecular dynamics simulations were carried out using *Sa*FabI and *Ec*FabI in the apo form and complexes with

https://doi.org/10.1021/acs.jcim.2c01178
J. Chem. Inf. Model. 2022, 62, 5746−5761

TCL, MUT, and AFN. Those systems were simulated using both a tetramer and monomer. Furthermore, apo forms of triclosan-resistant mutant *Sa*FabI-Gly112Val and *Ec*FabI-Gly113Val models were generated (tetramers/monomers) as control simulations for interface disruption.[16]

2.1. Modeling and Structure Preparation. *Sa*FabI and *Ec*FabI were simulated as tetramers in the apo form and complexes with TCL, MUT, and AFN. For that, *Sa*FabI and *Ec*FabI structures were modeled using SwissModel[31] and PDB ID 4ALI[15] and 1QG6[10] structures as templates, respectively; both tetramers bound to TCL. The initial validations of individual models were accessed using the Ramachandran plot, QMEANS score, and local similarity to the target protein (Supporting Information, Figure S1). Tetramer structures were assembled using PISA webserver[32] and compared with crystallographic structures.

The coordinates for TCL (PDB ID: 1QG6[10]), MUT (MD frame previously reported[28]), and AFN (PDB ID: 4FS3) were retrieved from respectives referenced structures.[8] The coordinates of $NADP^+$ (*S. aureus*) and NAD^+ (*E. coli*) were retrieved from the template structures. Further, steric bumps were minimized using Maestro (2021.4). Protein structures were prepared by adding hydrogen atoms and fixing missing side chains using Protein Preparation Wizard[33] (Maestro 2021.4, Schrödinger LLC, New York, NY, USA). Ionization states of amino acids were optimized with PROPKA (pH 7.4, Schrödinger), and as for the ligands, TCL and MUT were considered phenolates.[28]

2.2. Molecular Dynamics Simulations. Molecular dynamics (MD) simulations were carried out using the Desmond engine[34] with the OPLS4 force field.[35] Simulated systems encompassed the protein−ligand/cofactor complex, solvent (TIP3P[36]), and counterions (Na^+ or Cl^- adjusted to neutralize the overall system charge). The entire system was treated in a cubic box with periodic boundary conditions, specifying the size of the box as 13 Å distance from the box edges to any atom of the protein. Short-range coulombic interactions were calculated using 1 fs time steps and 9.0 Å cut-off value, whereas long-range coulombic interactions were estimated using the Smooth Particle Mesh Ewald method, which is a sufficiently good approximation to treat long-range interactions on large timescales.[37,38] Each system was subjected to at least 15 μs simulations (split among five replicas), which should be sufficient to observe the protein conformational changes of interest.[39,40]

2.3. Molecular Dynamics Trajectory Analyses.

2.3.1. Root Mean Square Deviation (RMSD). Root mean square deviation (RMSD) values of the protein backbone were used to monitor simulation equilibration and protein folding changes (Supporting Information, Figure S2). The fluctuation (RMSF) by residues was calculated using the initial MD frame as a reference and compared between ligand-bound and apo structure simulations (Supporting Information, Figure S3). The datasets generated and/or analyzed during the current study are available in the Zenodo repository (DOIs: 10.5281/zenodo.6917091, 10.5281/zenodo.6917244, and 10.5281/zenodo.7032391). Data available include the trajectory raw data. MD trajectories were visualized, and figures were generated using PyMOL v.2.5.2 (Schrödinger LCC, New York, NY, USA).

The obtained trajectories were initially analyzed by their protein backbone's RMSD variations, both for full complexes and individual chains (Figure S2), RMSF of single chains (Figure S3), and the radius of gyration, to investigate large changes. We also used the minimum distance between ligands and relevant residues for the binding (Ala95/97, Tyr146/147, and Tyr156/157, Figure S4) as a control for the ligand residence within the binding site. Those results show the stability of our simulated complexes and their suitability to further analyses. One single replica from all simulated systems (chain D of the MUT-*Sa*FabI complex) displayed high RMSF values due to instability at the end of the trajectory; therefore, this replica was removed for RMSF calculations and removed from further analyses. The same issue was noted for MUT binding to *Sa*FabI: in one single replica and one chain, the inhibitor left the binding site. However, these observations represent less than 5% of the total sampling.

2.3.2. Protein−Ligand Interactions. Atomic interactions and distances were determined using the Simulation Event Analysis pipeline as implemented in Maestro 2019v.4 (Schrödinger LCC). The criteria for the protein−ligand H-bond are 2.5 Å distance between the donor and acceptor atoms (D—H···A), ≥120° angle between the donor−hydrogen−acceptor atoms (D—H···A), and ≥90° angle between the hydrogen−acceptor-bonded atoms (H···A—X). Corresponding requirements for protein−water and water−ligand H-bonds are 2.8 Å (D—H···A), ≥110° (D—H···A), and ≥90° (H···A—X). Nonspecific hydrophobic interactions are defined by the presence of a hydrophobic side chain within 3.6 Å of the ligand's aromatic or aliphatic carbons. $\pi-\pi$ interactions are recorded when two aromatic groups are stacked face-to-face or face-to-edge and within 4.5 Å of distance.[41]

2.3.3. Principal Component Analysis (PCA). Principal component analysis (PCA) was used to study the main features of FabI backbone movements. PCA was conducted for all backbone atoms, which were selected and aligned using scripts (trj_selection_dl.py and trj_align.py) from Schrodinger package 2021v.4. Individual simulations from all runs were merged using the trj_merge.py script into a final trajectory and CMS file. Then, Desmond trajectories were aligned and transformed to the xtc format, keeping only backbone atoms, which were further used to generate the principal components. PCA was conducted for the backbone atoms using GROMACS tools (version 2019) (gmx covar and gmx anaeig).[42] For GROMACS analysis, figures representing the extreme motions were generated and visualized using the PyMOL script Modevectors (https://github.com/Pymol-Scripts/Pymol-script-repo/blob/master/modevectors.py), with motions with less than 5 Å distance between the initial point and the last point of the extreme motion frames not considered. For further analysis, we included the PCs that displayed >15% individual contribution. PCA graphics for the 2D projections were generated using a Python script. All commands were generated using JuPyter (Matplotlib, Seaborn, Numpy, and Pandas).

2.3.4. Domain Movement Analyses. The distance analysis scripts were used to study loop movements. The distance between selected regions was calculated by their center of mass distance using the script (trj_asl_distance.py) available on Schrodinger package 2021v.4. The amino acid ranges employed for calculations are mentioned in the respective sessions and figure legends.

2.3.5. Water Distribution and Water Occupancy Cluster. For each frame, the distance of all water oxygen atoms in relation to the COM of the tetramer's chains was calculated. Then, the distance distribution was plotted and normalized by the number of analyzed frames, generating a water distance distribution for each tetramer. Finally, the different chains of the same tetramer

https://doi.org/10.1021/acs.jcim.2c01178
J. Chem. Inf. Model. 2022, 62, 5746−5761

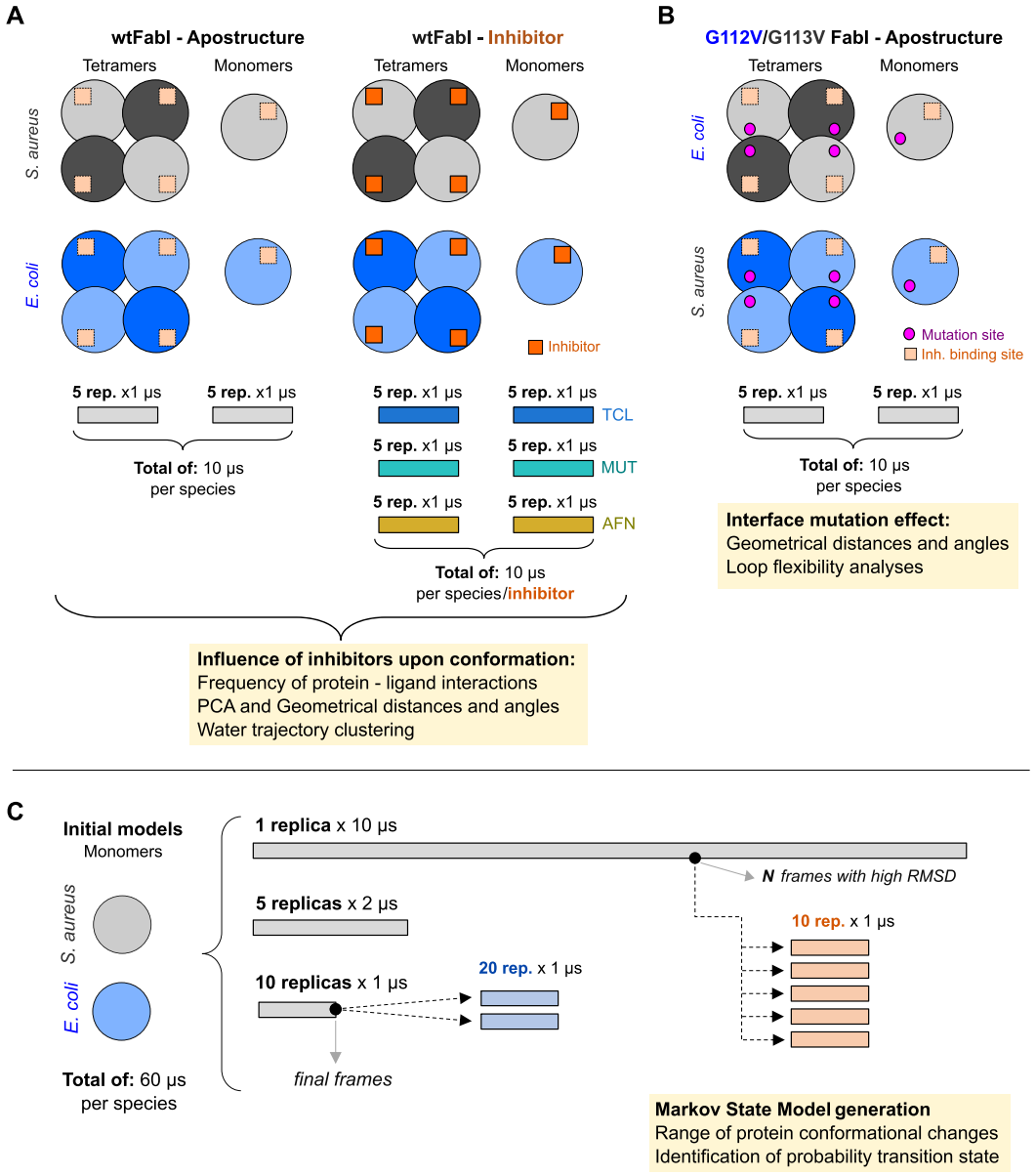


Figure 2. Experimental design of the simulations generated within this work. For each set of simulations, the hypothesis and developed analyses are highlighted in yellow boxes. (A) Study of the influence of different inhibitors (TCL, blue; MUT, turquoise; AFN, gold) on the tetramer stability for *S. aureus* (gray spheres) and *E. coli* (blue spheres) in comparison to their monomeric counterparts. Similar analyses were performed for the mutant Gly112/113Val (B), which is known to confer inhibitor resistance. The mutation spot in the dimerization interface is depicted as a purple sphere. (C) Monomeric apo structure simulations were performed in various lengths and number of replicas to generate a large and diverse set of conformations for Markov State Models.

were merged, that is, an averaged water distance distribution was generated. The area under the curve of the averaged water distribution up to a distance of 1 nm was estimated using the trapezoidal method. Waters that were found within a radius of 1 nm or less of the COM of the tetramer's chains in at least one frame of the trajectory had their coordinates extracted. Thus, a simple PDB was generated with only the coordinate of the oxygen atoms of the water located within this radius.

2.4. Markov State Model (MSM) Analysis. The extended sampling individual concatenated trajectories of monomeric states of apo *Sa*FabI and *Ec*FabI (see Figure 2C) were used as an input for MSM generation. Briefly, 16 replicas (1×10 μs, 5×2 μs, and 10×1 μs) shared the same starting frame, while 20 replicas used the final frames of 1 μs trajectories as an input, and the last 10 replicas used two high RMSD frames from 10 μs-long trajectories as a starting point.

MSM generation was conducted with PyEMMA2[43] according to ref 44. For featurization, we used the backbone torsions of whole protein. Initially, the obtained VampScores for each system (3.51 for *Sa*FabI and 3.62 for *Ec*FabI) suggest the reliability of calculated states, their probabilities, and the free energy conformational landscape.[45] The dimensional reduction was conducted with time-lagged independent component analysis (TICA).[46] Lag time τ = 800 step density and 273

https://doi.org/10.1021/acs.jcim.2c01178
J. Chem. Inf. Model. 2022, 62, 5746−5761

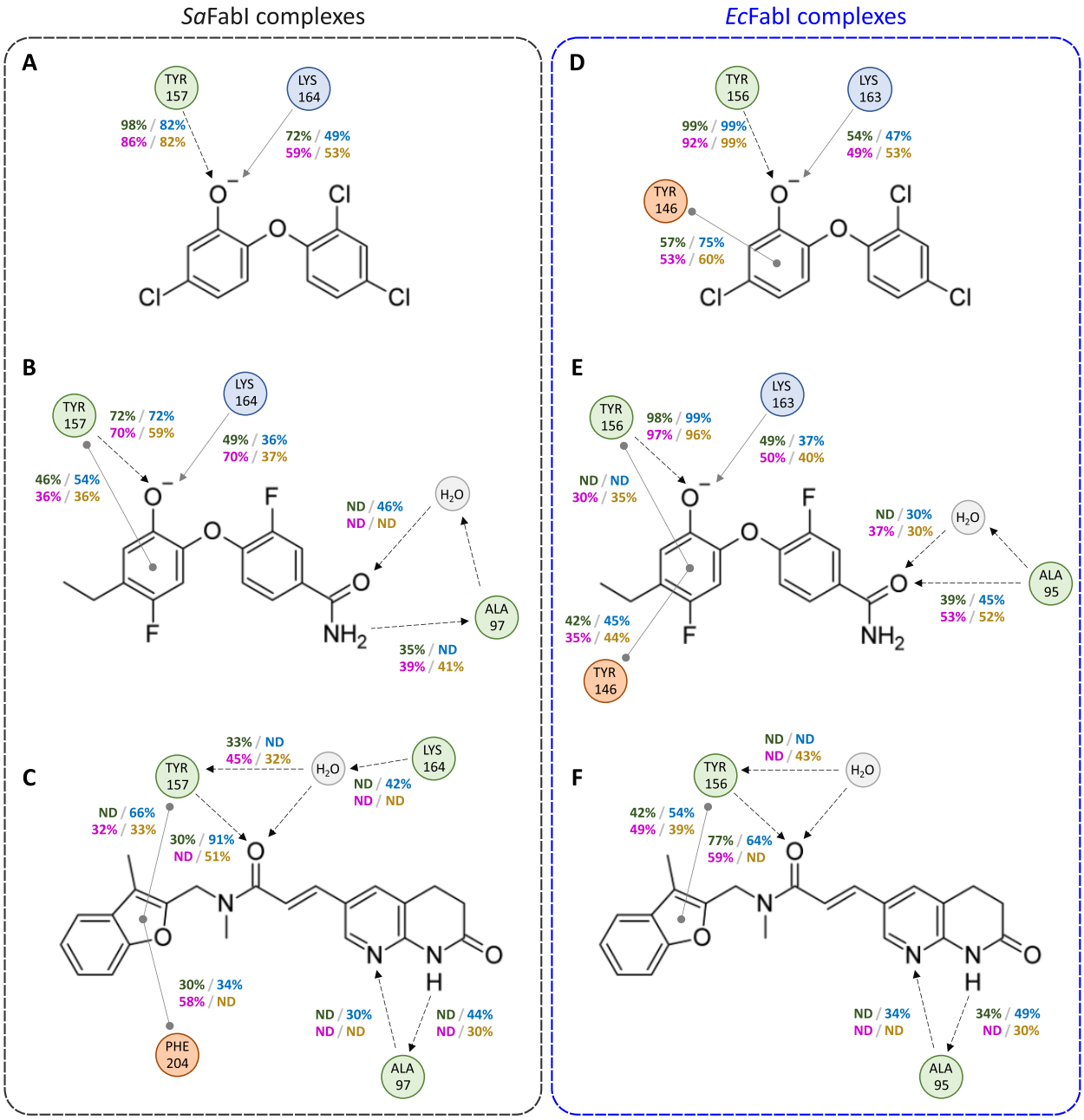


Figure 3. Simulation diagram interactions (SIDs) for TCL, MUT, and AFN complexes with *Sa*FabI (A−C, respectively) and *Ec*FabI (D−F, respectively). Green amino acid residues are involved in H-bonds, blue ones are involved in ionic interactions, and orange residues are involved in π-stacking interactions. Furthermore, percent values near arrows indicate the interaction frequency (with a cut-off of at least 30% of the trajectory) for chains A (green), B (cyan), C (magenta), and D (yellow). Dashed arrows represent the H-bond from the donor (tail) to the acceptor (arrow), solid arrows represent ionic interaction, and solid gray lines connecting gray circles represent π-stacking interactions.

dimensions were selected (Figure S4), 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^{++})[47] assigned the microstates to macrostates. Transition-path theory (TPT) was applied to investigate state transitions and the flux between metastable states.[48,49] The validation of MSM models and percentages of probability among the states are shown in Figures S5.

2.5. 3D Visualization of Sequence Similarity (SS3D).

The algorithm used was inspired by a previously published study.[50] Initially, a contact map of α-carbons was generated with a cut-off distance of 0.6 nm, in each frame, for all tetramers. Subsequently, contacts that remained present for less than 75% of the frames (3761 frames) were eliminated. The SS3D value for each of the selected contacts was calculated, having the apo protein tetramer as a reference among all the other structures compared. For that, we selected residues whose α-carbon was

https://doi.org/10.1021/acs.jcim.2c01178
J. Chem. Inf. Model. 2022, 62, 5746−5761

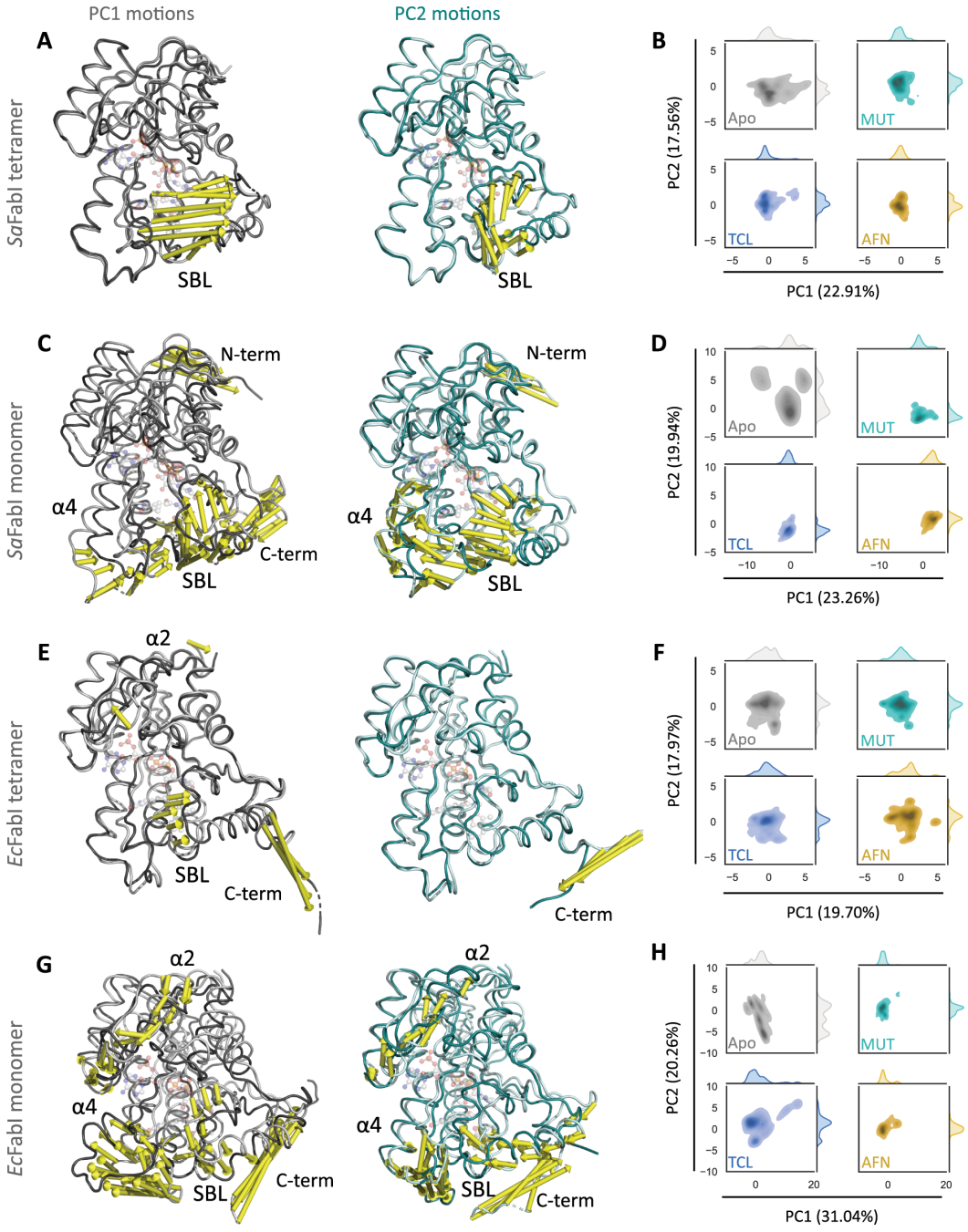


Figure 4. Protein movements (represented by yellow arrows) are described from the first and second PCs for *Sa*FabI tetramers (A) and monomers (C) and *Ec*FabI tetramers (E) and monomers (G). Distribution of frames' scatter plot histogram for the apo system (gray), TCL complexes (blue), MUT complexes (cyan), and AFN complexes (orange) along with the first PC for *Sa*FabI tetramers (B) and monomers (D) and *Ec*FabI tetramers (F) and monomers (H). NAD and AFN are presented in a transparent ball-and-sticks model as an indication of binding site position.

within a radius of 1 nm from any of the residues in contact, considering only residues that persisted for more than 75% of the frames within this radius. Originally, SS3D was intended to be used to highlight differences between homologous proteins by combining the sequence and structure information to do so. Since we are comparing the same tetramer interacting with different ligands, instead of calculating the SS3D using a substitution matrix, we are simply counting the number of common residues around the contact as a score, removing the substitution weight. Finally, all the different chains were merged

https://doi.org/10.1021/acs.jcim.2c01178
J. Chem. Inf. Model. 2022, 62, 5746−5761

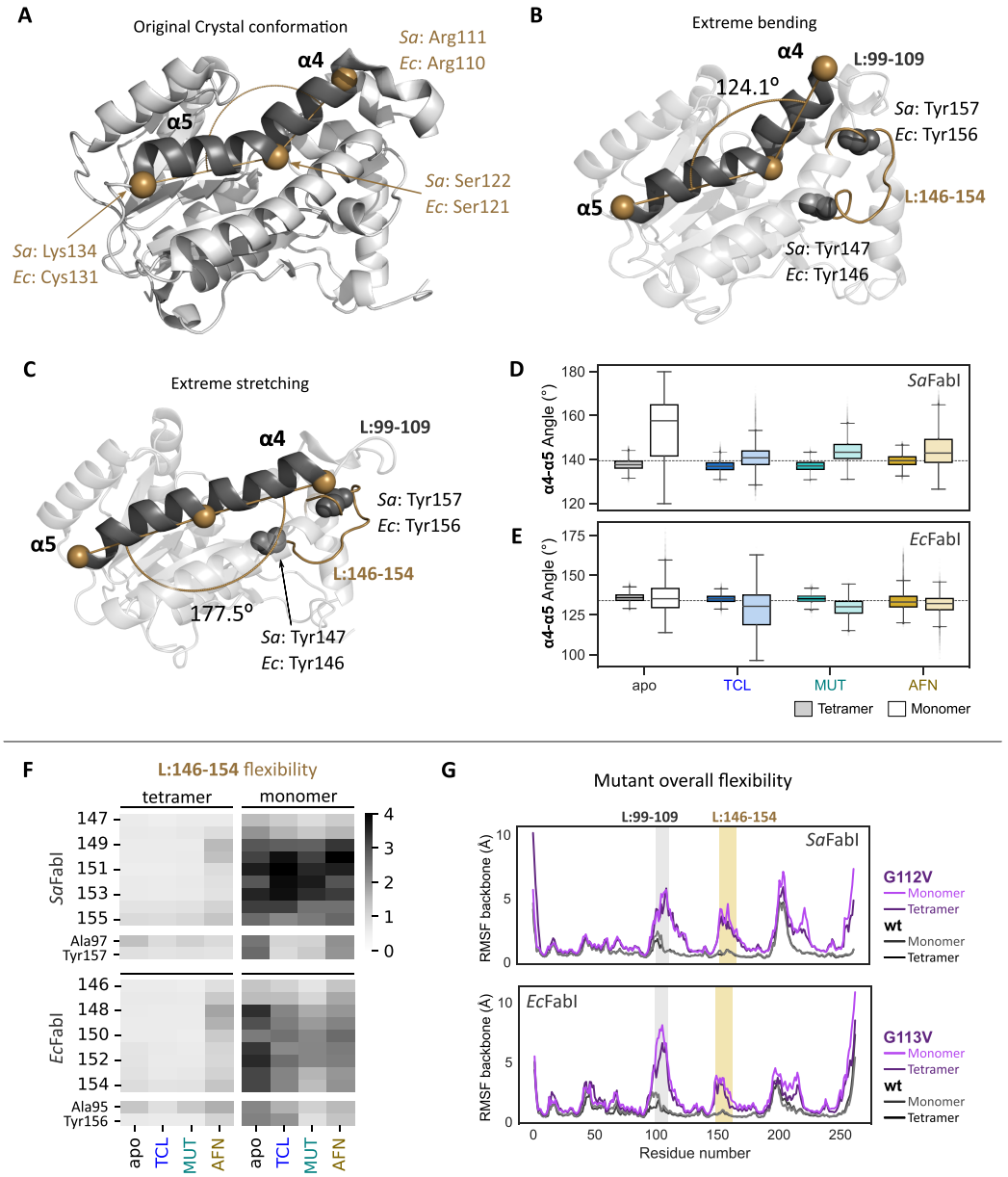


Figure 5. Illustration of the angle between α-helices 4 and 5 (A), with extreme bending (B) and stretching (C) conformations. The angle was calculated between the α-carbon of H4's first residue, the hinge residue (Ser121/122), and the α-carbon of the last H5's residue. Structures were selected from visually inspected relevant frames. Box plots of the angle measurements of H4−H5 bending for apo *Sa*FabI (D) and the complexes with TCL, MUT, and AFN and the same comparison for the *Ec*FabI systems (E). Box plots colored with lighter shades are from monomers and the darker ones are tetramers. (F) Heatmap with RMSF values for the loop 146−154 of *Sa*FabI and *Ec*FabI tetramers and monomers. (G) RMSF plots of the apo structure of wt FabI tetramer (black) and monomer (dark gray) and G112/113V interface mutant tetramer (dark purple) and monomer (light purple) for *Sa*FabI and *Ec*FabI. The gray shade represents the loop 99−109, while the yellow shade represents the loop 146−156. For both systems, the RMSF values of Ala97/95 and Tyr157/156 are also plotted as the control because they directly interact with ligands and show no differences between inhibited systems and apo structures for tetramers.

and the resulting SS3D value was mapped between 0 and 1. Contacts that were not common between the different chains were discarded.

3. RESULTS

3.1. Tetrameric FabI Inhibitor Simulations Reveal Asymmetry among Subunits. Simulated trajectories are displayed in Figure 2, encompassing FabI in complex with TCL, MUT, and AFN inhibitors either as tetramers or monomers, both from *S. aureus* and *E. coli* (Figure 2A). Also, two clinically relevant mutants, G113V *Sa*FabI and G112V *Ec*FabI, were simulated in monomeric and tetrameric states as the control of interfaces (5 × 1 μs, Figure 2B).

Our simulations reproduce the key protein−ligand interactions of experimentally determined binding modes (Figure 3).[8,10,18] All compounds maintain hydrogen bonds between

https://doi.org/10.1021/acs.jcim.2c01178
J. Chem. Inf. Model. 2022, 62, 5746−5761

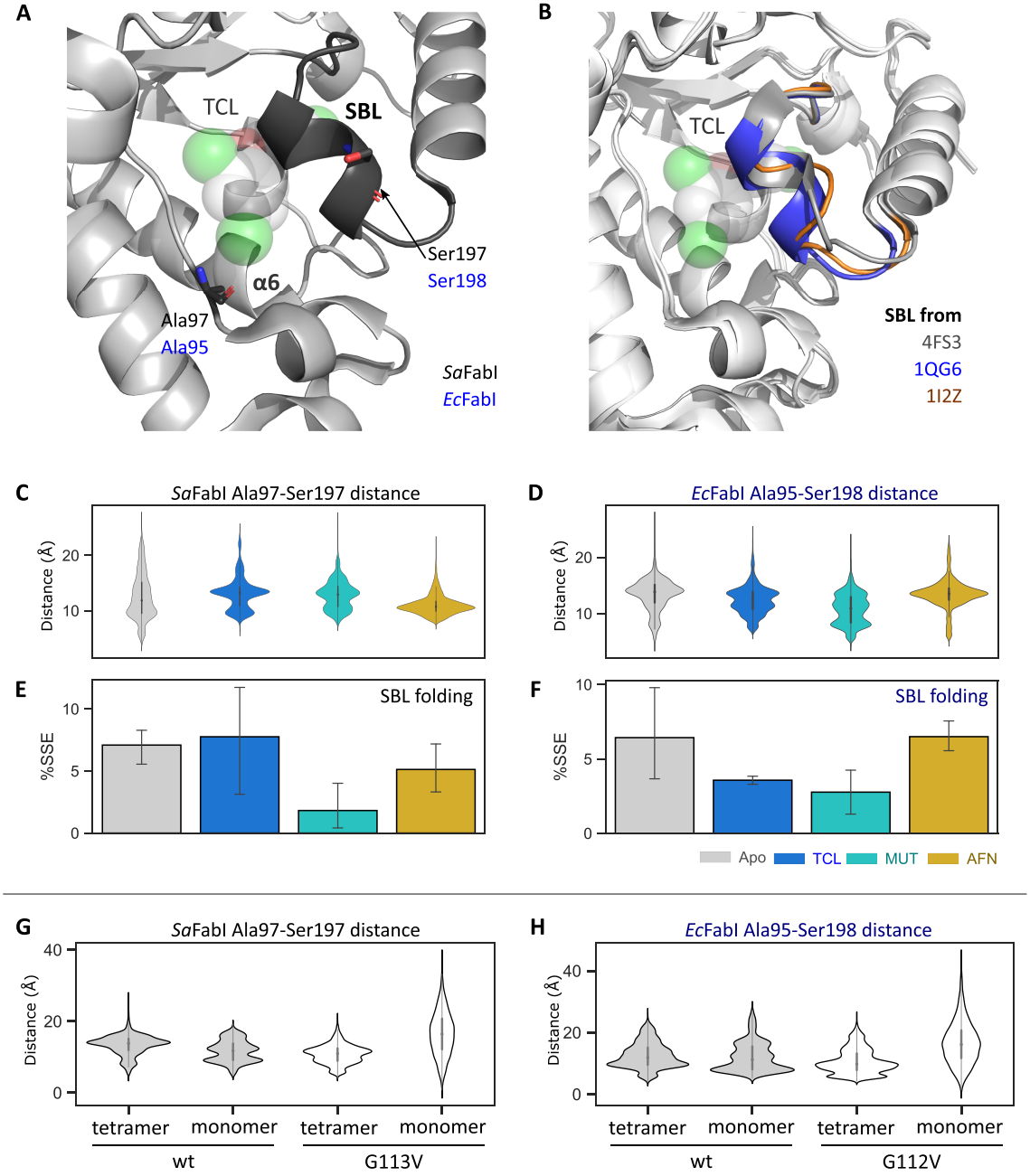


Figure 6. Cartoon representation of SBL and sticks representation of Ala97/95 and Ser197/198 (A). Comparison of SBL folding of experimental FabI structures deposited in PDB under IDs 4FS3[8] (gray), 1QG6[10] (blue), and 1I2Z[9] (orange) (B). Violin plots of the distance between SBL Ser197/198 and reference Ser97/95 for *Sa*FabI (C) and *Ec*FabI (D). Average percentage of SBL folding for the four monomers for *Sa*FabI (E) and *Ec*FabI (F). Distance between Ala97 and Ser197 (representing the COM of the SBL) of *Sa*FabI's centers of mass for wt and G113V mutant apo structures (G) and Ala95-Ser198 of *Ec*FabI for wt and G112V mutant apo structures' centers of mass (H).

their phenolate (TCL/MUT > 80%) or carbonyl (AFN, >30% in most cases) and Tyr157/156. Interestingly, MUT and AFN (but not TCL) display π-stacking interactions with Tyr157/156. Specific for AFN, π-stacking with Tyr157/156 is present even when the H-bond with the very same amino acid is less frequent than 30% of the trajectory. The TCL/MUT phenolate displays additional ionic interactions with Lys164/163. Additionally, MUT and AFN extended side-chain groups (an amide and a lactam, respectively) form a H-bond with Ala97/95, either directly or water-mediated. Finally, π-stacking interactions are observed between Tyr157/156, Tyr147/146, and/or Phe204 (*Sa*FabI) being complemented by hydrophobic pockets near the active site.

3.2. Tetramerization Stabilizes Substrate Binding Loop Movements. We individually analyzed the *Sa*FabI and *Ec*FabI trajectories using principal component analyses (PCAs), concatenating both monomer trajectories and each monomer in the tetramers. The first two components account for

https://doi.org/10.1021/acs.jcim.2c01178
J. Chem. Inf. Model. 2022, 62, 5746−5761

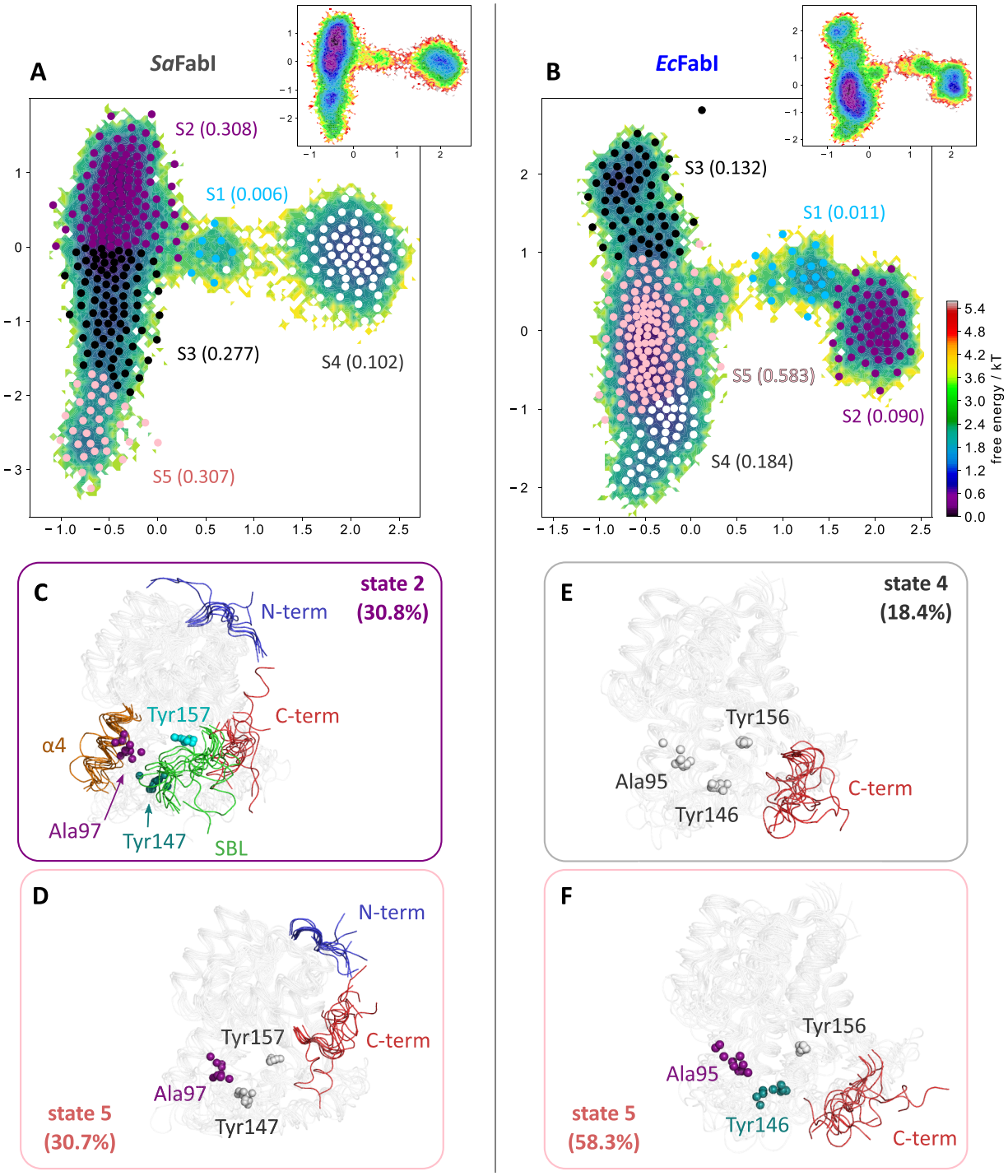


Figure 7. *Sa*FabI and *Ec*FabI structures separated by states and their probabilities along MSM components plus a graph in the top-right highlighting the relative free-energy landscape of structure transition (A and B, respectively). Two most probable states of *Sa*FabI (C and D) and *Ec*FabI (D and F) with disordered regions colored.

approximately between 38 and 52% in all simulations. Details of the PCA results are shown in Figure 4.

All tetrameric *Sa*FabI inhibitor simulations differ from the apo structure regarding SBL (L:195-200) movements (Figure 4A,B), which are explained by both PC1 and PC2 extreme motions. While the apo system transits between SBL-open and SBL-closed conformations, inhibited systems mainly stay in the closed state. For comparison, monomeric simulations presented additional motions of N- and C-termini and H4, besides the SBL, suggesting that the tetramer restricts their movement. The

https://doi.org/10.1021/acs.jcim.2c01178
J. Chem. Inf. Model. 2022, 62, 5746−5761

same extreme motions and relevant regions were observed for *Ec*FabI (Figure 4E,F). The distribution of PC1 scores shows that *Ec*FabI-MUT and *Ec*FabI-AFN complexes are more likely to have a closed SBL and C-terminus stabilized toward the protein core instead of the solvent. In contrast to *Sa*FabI complexes, the PC1 and PC2 score distributions of *Ec*FabI-TCL are distinct from the other systems. In addition, *Ec*FabI-TCL presents a conformation with an extended C-terminus, suggesting interaction with other subunits of the tetramer (Figure 4E,F).

Further, PCA indicated that monomeric simulations of *Sa*FabI with MUT and AFN display a distinct dynamic from TCL-bound and apo structures, where the SBL transits between open and closed conformations. In addition, PC1 motions display movements in the H4 and C-terminal substructures for both FabIs. Since these substructures compose the tetramer interfaces, their geometry was also investigated in the tetrameric trajectories by tracking the distances of the chain's centers of mass (COMs, Figure S6).

Only minor variations among apo, TCL, MUT complex, and AFN simulations were observed, suggesting inhibitors' influence on the dimer−dimer interfaces. However, distinct changes were observed in the PR and QR interfaces, which were previously considered symmetrical. The QR interface is composed of the helices 4/5 and 6 from two chains. Due to this, we decided to analyze in more detail ligand binding effects upon helices 4/5 as the main mechanism for the tetramer stabilization.

3.3. Indirect Effects of Interface Substructures on FabI's Binding Site. Interestingly, examples of extreme movements of H4 and H5 indicated that the loop directly attached to H4 (loop 99−109) moves as a "whip", inducing conformational changes to the loop 146−154, which forms an interface with the C-terminus. Therefore, we tracked the bending angle of H4/H5 (Figure 5A−C). All tetrameric systems have lower H4/H5 movement distribution compared to their monomeric counterparts (Figure 5D,E), as well as lower flexibility of the loop 146−154 (Figure 5F). The H4-connected loop 146−154 is directly attached to catalytic residues Tyr146/147 and Tyr156/157 (Figure 5B,C), suggesting that changes in these two substructures would propagate to the active site. Consistently, the RMSF of FabI-G113/112V mutant, a mutation at the PR interface, also showed higher fluctuation of loop 99−109 (Figure 5G) and this effect was more pronounced for monomer simulation.

AFN simulations displayed a distinct behavior, in comparison with all other simulated complexes, increasing the stretching of H4 and H5 movement for *Sa*FabI or bending in *Ec*FabI (Figure 5D,E). This AFN flexibility was confirmed by higher RMSF values of loop 146−154. Further, the distances between the twin tyrosine residues 147/146 and 157/156 to H4/5 show that AFN complexes promote larger movements in this region (Figure S7).

Further, PCA extreme motions suggest a larger amplitude of movement in the C-terminus, in comparison to other substructures. Interestingly, recent findings using truncated mutants at the C-terminus[51] suggest the loss of interactions with other subunits in the tetramer. In our simulations, the C-terminus interacts with the L:146−154 loop of another monomer (see the Supporting Information, Table S1 and associated discussion), displaying an additional indirect effect on the monomer's conformational changes (see Figure 4).

We also calculated movements and folding of the SBL (L:195−200) given that different crystals have shown its conformation changes upon ligand binding (Figure 6A,B). For both *Sa*FabI and *Ec*FabI in complex with AFN, the distance between the SBL and the core protein (Ala97/95) follows a unimodal distribution, while the others present bi- or even trimodal behavior (Figure 6C,D). *Sa*FabI's SBL has less helical-folding property for AFN and MUT, in comparison with other systems (Figure 6E). The SBL average folding of *Ec*FabI-AFN is very similar to the apo system but with a very narrow standard deviation (Figure 6F). Finally, the distance of Ala95/97 and Ser197/198 (Figure 6G,H) for the wild type and mutants suggests similar structural movements, except for the FabI mutant monomer. This indicates that the SBL movement of monomers is more susceptible to the interface mutation effects, further connecting these two regions. In this sense, FabI tetramerization is linked to the SBL movements, which depends on the presence of inhibitors to occur, a fact that explains the resistance generated by the PR interface mutation.

3.4. Markov State Models Identify Coordination between Ala97/95, Loop 146−154, and C-Terminus. At this point, our findings indicate that the FabI interface showed different movements for apo and inhibited simulations. However, there are several coordinated motions of the C-terminus, SBL, and loops 99−109 and 146−154, which imply an interaction network changing the conformational states of binding site elements. In other words, there are interactions in the PR and QR interfaces of FabI that impact the monomeric structure of the enzyme and, consecutively, the binding of inhibitors. PCA followed by calculations of specific distances and angles suggested that conformational states are able to distinguish apo structure systems from inhibited ones. This shows the mechanistic connection between the tetramer stabilization and inhibitor binding. However, at this point, we are not able to calculate the energy transition between those states nor their probabilities. Then, in parallel, several replicas with different timescales and starting frames of apo structures of *Sa*FabI and *Ec*FabI monomers were simulated, resulting in a concatenated trajectory of 60 μs for each protein (see Figure 2C), which were used as an input for Markov State Model analyses.

MSM suggested that the most probable state, S2 (30.8%), is in the region with the lowest relative free energy (Figure 7A) and presents disordered relevant structural features, Ala97, Tyr147, and Tyr157 as well as H4, N- and C-termini, SBL, and loop 146−154 (Figure 7C). Further, S5 (30.7%; the second-most probable state) kept H4 and SBL ordered and Ala97 and N- and C-termini remain disordered. Similar behavior was observed for *Ec*FabI (Figure 7B), where the most probable state (S5, 58.3%) is located at the lowest relative energy and presents a stable protein core, however with the C-terminus, Ala95, and Tyr146 being disordered. This state could transit to state S4 (18.4%), which displays a disordered C-terminus, or state S3 (13.2%) (Figure 7E,F).

Interestingly, a modified version of the Sequence Similarity 3D (SS3D)[50] analysis of full tetramer trajectories indicated that both complexes have conserved regions according to the protein movement and contacts, except the SBL, C-terminus, loops 99−109 and 146−154, Ala97/95, and helices H4/H5 (Figures S8 and S9). In other words, global analyses such as SS3D, MSM, and PCA corroborated the geometric and local distance analyses.

3.5. Water Molecule Occupancy Suggests Different Hydration Profiles and Potential Pockets. In parallel, the tetramer trajectories were analyzed, aiming at calculation of the frequency of water molecules around the monomer's COM

https://doi.org/10.1021/acs.jcim.2c01178
J. Chem. Inf. Model. 2022, 62, 5746−5761

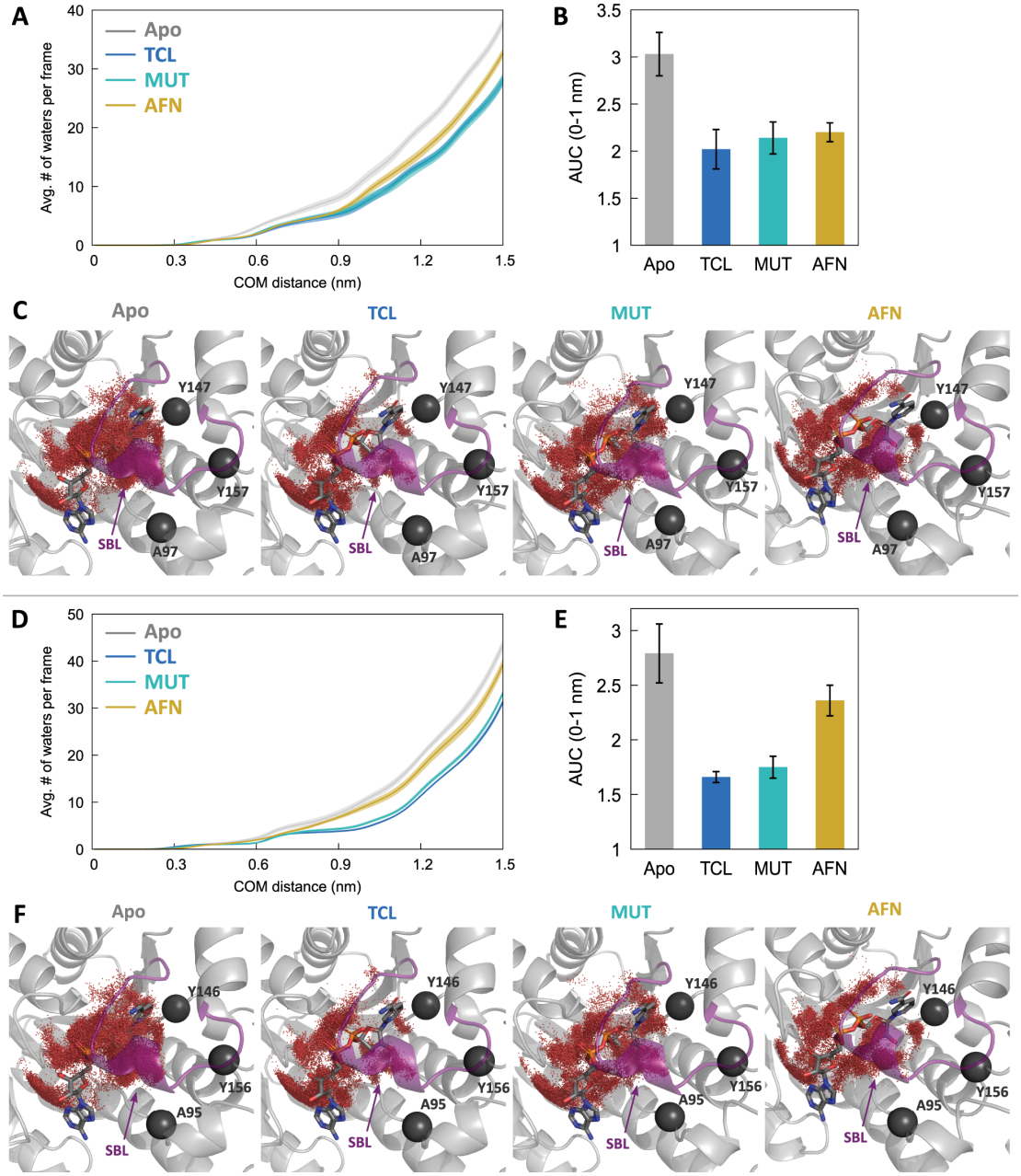


Figure 8. Average number of water molecules per frame versus the distance around the center of mass of each monomer during whole tetramer trajectories for all *Sa*FabI (A) and *Ec*FabI (D) simulated systems. Calculated area under the curve (AUC) values at 1.0 nm from COM (considered optimum value to distinguish the systems) for *Sa*FabI (B) and *Ec*FabI (E) simulated systems. Water clusters at 1.0 nm from COM for *Sa*FabI (B) and *Ec*FabI (E) apo structures and inhibitor complexes.

using a water clustering algorithm. The obtained results indicate that both *Sa*FabI and *Ec*FabI have similar behavior: the apo structure is more hydrated followed by FabI-AFN and, last, FabI-TCL and MUT (Figure 8A,D). In this sense, the average number of water molecules at 10 Å from the COM is optimal to represent those differences (Figure 8B,E). The observed water clusters at 10 Å from the COM (Figure 8C,F) indicate that in both the apo structure and FabI-AFN complexes, water molecules tend to more frequently occupy ligand binding pockets near SBL and Tyr147/146 in comparison to TCL and MUT complexes.

4. DISCUSSION

Previous FabI inhibitor studies employed shorter monomeric simulations focusing on protein−ligand interactions.[14,28,52−54] For instance, our group carried out short molecular dynamics of

https://doi.org/10.1021/acs.jcim.2c01178
J. Chem. Inf. Model. 2022, 62, 5746−5761

1 μs (5 × 200 ns), studying *Ec*FabI monomers with diphenyl ether derivatives,[28] where we established a relationship between protein−ligand interactions and compound potency. Previous *Sa*FabI full tetramer simulations (four individual chains with 25 ns each) identified structural determinants, such as the folding of H6/H7, that led to differences in the ligand residence time.[30]

Our current work discloses the longest published all-atom simulations of *S. aureus* and *E. coli*'s FabI to date, encompassing different inhibitor classes and oligomerization states. We decided to use longer simulations to investigate the extension of FabI conformation changes. Our current simulations allowed us to analyze relevant structural features, such as the dimerization interface (H2 and H4), changes in the substrate binding loop and C-terminal conformation diversity, and the network of interactions responsible for triggering conformational changes.[55−57]

Also, metastable states derived from monomeric simulations showed multiple conformations for the substrate binding loop, N- and C-termini, and helices H2 and H4 (QR interface). Some of these conformations were recapitulated by PCA and distance measurements in the tetramer simulations, while others were unique to the monomers. This prompts us to suggest that the stabilization of those substructures by inhibitor binding could occur before further dimerization/tetramerization. In the following sessions, we discuss the conformational changes in these individual regions.

4.1. Bending Movement of H4/H5 Influences the QR Interface. The formation of monomer−monomer and dimer−dimer interfaces of FabI was experimentally assessed, suggesting that multiple oligomeric states exist in equilibrium. For example, Rafi and colleagues[14] observed experimentally the FabI:ACP (2:1) stoichiometry in the FabI tetramer crystal and they were not able to find a 1:1 ratio by docking simulations due to steric bumps between monomers.

Later, it was shown that the oligomeric transition is coupled to ligand binding in *Sa*FabI,[15] which is supported by the increased ligand affinity depending on FabI's oligomerization state.[27] The link between interface stabilization and ligand binding was finally established for *Sa*FabI[15] and *Ba*FabI,[58] where inactive dimers displayed a large H5 inward bending (55°), blocking the NADH binding site. This is supported by the *Sa*FabI mutant Met99Thr, located at the loop between H4/5 and the loop with twin tyrosine residues, which induces a decrease in AFN activity by 2-fold but not TCL.[8]

It is worth noting that our measurements for H4/H5 bending angles (Figure 5D,E) in the *Sa*FabI-AFN complex repeat this phenomenon, while in the *Ec*FabI-AFN complex, these same helices undergo elongation in comparison to the apo form. This is the first report of those movements for *Ec*FabI complexes, which suggests that this could be a more general mechanism for FabI from different species. Furthermore, we propose that the AFN inhibition mechanism involves an indirect net of interactions from the active site toward the QR interface, which is different from the direct competitive behavior of inhibitors.

4.2. SBL Movements and Folding Changes upon Ligand Binding. The different inhibitory mechanisms between AFN/MUT and TCL were mainly attributed to the formation of H-bonds with Ala97/95.[8] However, these interactions were not stable in our simulations (Figure 3) and the bulkier naphtpyridinone ring seems to fit the hydrophobic pocket between Ala97/95 and SBL, rather than only interacting with them. These SBL interactions stabilize it in closed conformation simulations (Figure 6). This is consistent with the smaller distances between Ala97/95 and Ser198/197, as well as a unimodal distribution of those distances (Figure 6C,D,G,H), and differences in the SBL folding (Figure 6E,F). We observed that SBL movements are important to distinguish inhibited states (closed SBL) from apo structures (switching between open and closed states) in tetrameric forms (Figure 4). Meanwhile, the monomer simulations showed disordered movements of the SBL and active site (Figure 7). This conformation diversity was previously observed, to a lesser extent, in shorter simulations (50 ns), where the SBL assumed different folding profiles (α-helix or 3-10 helix).[52] The relevance of these states is also corroborated by structural information from other organisms' crystal structures, such as *Burkholderia pseudomallei* and *A. baumannii*.[59,60] In these structures, the SBL conformation shows great difference, especially when compared to *Sa*FabI and *Ec*FabI crystal structures. Strikingly, the FabI-kalimantacin A complex, which is a very large inhibitor compared to commonly reported small molecules, does not rely on H-bonds with Ala95 and Ser197 linking these two flexible loops.[29]

In summary, the SBL conformation displays a different movement/folding profile in the AFN simulations in comparison with TCL complexes and apo structures. However, since the selected ligands do not greatly differ in terms of potency (Figure 1), we hypothesized that those observations may be related to their residence time.[61−64]

4.3. The Water Occupancy Changes in the Active Site Changes upon Ligand Binding. Last, we tracked the water molecules around the COM of the FabI tetramer to investigate their hydration level and relevant water channels. Surprisingly, AFN forms more hydrated complexes than TCL/MUT and as hydrated as apo structures despite being larger than diphenylether inhibitors. We could interpret this result as a hydrophobic effect promoted by diphenylether derivatives, which contributed to expelling water molecules from the binding site. Since AFN has the lowest predicted log *P*, followed by MUT and TCL (consensus log *P* values predicted with SwissADME[65] equal to 2.86, 3.17, and 4.32, respectively), these water molecules around AFN are energetically allowed to stay in a pocket near Tyr146/147.

Interestingly, recent naphtpyridinone derivatives (including direct AFN analogues) help to illustrate the hypothesis that moderate lipophilicity contributes to allowing water molecules at FabI's binding site. The introduction of a $-NH_3^+$ group to AFN's naphtpyridinone ring resulted in a broad-spectrum compound against both Gram-positive and Gram-negative pathogens.[20] Molecular docking of this series suggests that the charged primary amine points toward the solvent.[20] Further development of these inhibitors[66] by changing the size and shape of the $-NH_3^+$ substituent resulted in a more potent derivative, namely, fabimycin. The *Ec*FabI-fabimycin cocrystal structure showed a favorable water network between the ligand, the NADH cofactor, and the enzyme. This novel water network justified the increase in antimicrobial activity, showcasing that this could be explored to improve the affinity of novel FabI inhibitors.

Finally, the complex dynamics herein reported will be valuable for future drug design efforts since identified hotspots and specific substructures could be targeted for the development of novel FabI inhibitors.[67] Our findings suggest different conserved binding modes for compounds from distinct chemical classes. Those binding modes rely not only upon direct interaction with

https://doi.org/10.1021/acs.jcim.2c01178
J. Chem. Inf. Model. 2022, 62, 5746−5761

binding site residues but also on promoting a network of intramolecular interactions that affects tetramer interfaces (clearly the QR interface but also others), suggesting cooperativity among FabI's subunits. Last but not least, the observed protein motions and their related mechanistic implications in addition to experimental data[51,58] could open a brand new possibility to target allosteric sites, aiming to increase the specificity and diversity of novel inhibitors.[68,69]

5. CONCLUSIONS

Our findings suggest that different inhibitors promote distinct motions in the oligomerization interfaces. FabI tetramerization into a catalytically competent homotetramer is a reversible process that relies on conserved structural elements. Our results suggest that multimerization is essential for the catalytic site's integrity and that inhibitor binding enables this multimerization by stabilizing not only the SBL but also the QR interface (H4/5). This proposed mechanism is consistent with inhibitor resistance caused by non-active site mutation, which we suggest is due to multimerization interference. Furthermore, there are significant changes in the hydration profiles between AFN complexes and diphenylether inhibitor complexes, especially near Tyr146/147, suggesting that water-mediated interactions could be explored (the chase for waterfall) to improve inhibitors' potency. For example, substitutions to promote and mediate those water-mediated hydrogen bonds (e.g., inclusion of polar groups such as hydroxyls, amines, etc.) could be done in the future design of novel FabI inhibitors.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.2c01178.

Figure S1: Ramachandran plots, QMEANS *Z*-score, and local similarity to the target protein generated by SWISS-Model; Figure S2: RMSD values of the entire tetramer and individual chains A, B, C, and D of *Sa*FabI and *Ec*FabI systems; Figure S3: RMSF values of individual chains A, B, C, and D of *Sa*FabI and *Ec*FabI systems; Figure S4: radius of gyration and distance of ligands' center of mass and Ala97/95, Tyr147/146, and Tyr157/156 for chains A, B, C, and D; Figure S5: validation of Markov State Models; Figure S6: FabI tetramer highlighting substructures of chain A's QR and PR interfaces as well as the other B, C, and D chains represented in transparent cartoons, representations of substructures present in the QR interface, PR interface, and loop 146−154/C-terminal interface, violin plot of the distances of the center of mass of each chain for *Sa*FabI and *Ec*FabI, and secondary structures present in the analyzed interfaces for *Sa*FabI and *Ec*FabI; Table S1: interactions of C-terminal amino acids from *Sa*FabI and *Ec*FabI and other parts of the tetramer, calculated by the simulation interaction diagram; Figure S7: distances between the α-carbon atoms for Tyr146 and Tyr156 toward the center of mass of H4/H5 (represented by Ser122/121's Cα) for *Sa*FabI and *Ec*FabI simulations; Figure S8: metastable states of *Ec*FabI and *Sa*FabI structures revealed by Markov state modeling; Figure S9: SS3D contact matrices for TCL-FabI, MUT-FabI, and AFN-FabI complexes in comparison with apo structures from both *S. aureus* and *E. coli* (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Antti Poso − *Institute of Pharmacy, Pharmaceutical/Medicinal Chemistry and Tübingen Center for Academic Drug Discovery, Eberhard Karls University Tübingen, 72076 Tübingen, Germany; Department of Oncology and Pneumonology, Internal Medicine VIII, University Hospital Tübingen, DE72076 Tübingen, Germany; Tübingen Center for Academic Drug Discovery & Development (TüCAD2), 72076 Tübingen, Germany; School of Pharmacy, Faculty of Health Sciences, University of Eastern Finland, 70211 Kuopio, Finland;* orcid.org/0000-0003-4196-4204; Email: antti.poso@uef.fi

Thales Kronenberger − *Institute of Pharmacy, Pharmaceutical/Medicinal Chemistry and Tübingen Center for Academic Drug Discovery, Eberhard Karls University Tübingen, 72076 Tübingen, Germany; Department of Oncology and Pneumonology, Internal Medicine VIII, University Hospital Tübingen, DE72076 Tübingen, Germany; Tübingen Center for Academic Drug Discovery & Development (TüCAD2), 72076 Tübingen, Germany; School of Pharmacy, Faculty of Health Sciences, University of Eastern Finland, 70211 Kuopio, Finland;* orcid.org/0000-0001-6933-7590; Phone: +49 (0)7071-29 74575; Email: thales.kronenberger@uni-tuebingen.de

Authors

Vinicius Gonçalves Maltarollo − *Departamento de Produtos Farmacêuticos, Faculdade de Farmácia, Universidade Federal de Minas Gerais (UFMG), 31270-901 Belo Horizonte, Minas Gerais, Brazil*

Ekaterina Shevchenko − *Institute of Pharmacy, Pharmaceutical/Medicinal Chemistry and Tübingen Center for Academic Drug Discovery, Eberhard Karls University Tübingen, 72076 Tübingen, Germany; Department of Oncology and Pneumonology, Internal Medicine VIII, University Hospital Tübingen, DE72076 Tübingen, Germany; Tübingen Center for Academic Drug Discovery & Development (TüCAD2), 72076 Tübingen, Germany*

Igor Daniel de Miranda Lima − *Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, Brazil*

Elio A. Cino − *Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, Brazil*

Glaucio Monteiro Ferreira − *Department of Clinical and Toxicological Analyses, School of Pharmaceutical Sciences, University of Sao Paulo, 05508-000 São Paulo, Brazil;* orcid.org/0000-0002-1952-9428

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.2c01178

Author Contributions

V.G.M. and T.K. designed the experiments and executed and analyzed the simulations. E.S. helped with the automation of analyses and scripting. G.M.F. generated and interpreted the Markov State Models. I.D.d.M.L. and E.A.C. developed and executed the SS3D and water analyses. T.K. and A.P. contributed to resources and overall study supervision. All authors helped with the discussion and writing. All authors prepared and reviewed the manuscript.

https://doi.org/10.1021/acs.jcim.2c01178
J. Chem. Inf. Model. 2022, 62, 5746−5761

Funding

This work was supported by the "Fundação de Amparo à Pesquisa do Estado de Minas Gerais" (FAPEMIG)−Brazil (project number APQ-01818-21) and the DAAD fellowship (reference number 91818683). G.M.F. is a recipient of a fellowship from FAPESP (grant #2021/11205-9, Brazil). T.K. is financed by the iFIT (EXC2180−390900677), Fortüne Initiative (NR.2613-0), and 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−Germany by the means of the program TüCAD2.

Notes

The authors declare no competing financial interest.

Further data discussion can be found in the Supporting Information. Additionally, all concatenated trajectories are provided as supporting information available at Zenodo (DOIs: 10.5281/zenodo.6917091 and 10.5281/zenodo.6917244). All trajectories are reported under the following IDs and DOIs: (i) *Sa*FabI and *Ec*FabI tetrameric systems (DOI: 10.5281/zenodo.6917091), (ii) *Sa*FabI and *Ec*FabI monomeric systems and G112/113V mutants (DOI: 10.5281/zenodo.6917244), and (iii) sampling trajectories for MSM analyses and water analyses (DOI: 10.5281/zenodo.7032391). Third-party software employed in the manuscript was as follows. ChemDraw version 19.0 (https://perkinelmerinformatics.com/), GraphPad PRISM version 9.2 (https://www.graphpad.com/), Schrödinger Suite 2019-2021.2 (https://www.schrodinger.com), and PyMOL version 2.5 (https://pymol.org/) are distributed under license.

■ ACKNOWLEDGMENTS

The authors thank the CSC-IT Center for Science, Finland, for the very generous computational resources.

■ ABBREVIATIONS

FabI, enoyl reductase; PDB, Protein Data Bank; NADH, nicotinamide adenine dinucleotide; NADPH, nicotinamide adenine dinucleotide phosphate

■ REFERENCES

(1) Silver, L. L. Challenges of Antibacterial Discovery. *Clin. Microbiol. Rev.* **2011**, *24*, 71−109.

(2) Rice, L. B. Federal Funding for the Study of Antimicrobial Resistance in Nosocomial Pathogens: No ESKAPE. *J. Infect. Dis.* **2008**, *197*, 1079−1081.

(3) Cassini, A.; Högberg, L. D.; Plachouras, D.; Quattrocchi, A.; Hoxha, A.; Simonsen, G. S.; Colomb-Cotinat, M.; Kretzschmar, M. E.; Devleesschauwer, B.; Cecchini, M.; Ouakrim, D. A.; Oliveira, T. C.; Struelens, M. J.; Suetens, C.; Monnet, D. L.; Strauss, R.; Mertens, K.; Struyf, T.; Catry, B.; Latour, K.; Ivanov, I. N.; Dobreva, E. G.; Andraševic, A. T.; Soprek, S.; Budimir, A.; Paphitou, N.; Žemlicková, H.; Olsen, S. S.; Sönksen, U. W.; Märtin, P.; Ivanova, M.; Lyytikäinen, O.; Jalava, J.; Coignard, B.; Eckmanns, T.; Sin, M. A.; Haller, S.; Daikos, G. L.; Gikas, A.; Tsiodras, S.; Kontopidou, F.; Tóth, Á.; Hajdu, Á.; Guólaugsson, Ó.; Kristinsson, K. G.; Murchan, S.; Burns, K.; Pezzotti, P.; Gagliotti, C.; Dumpis, U.; Liuimiene, A.; Perrin, M.; Borg, M. A.; Greeff, S. C. d.; Monen, J. C.; Koek, M. B.; Elstrøm, P.; Zabicka, D.; Deptula, A.; Hryniewicz, W.; Caniça, M.; Nogueira, P. J.; Fernandes, P. A.; Manageiro, V.; Popescu, G. A.; Serban, R. I.; Schréterová, E.; Litvová, S.; Štefkovicová, M.; Kolman, J.; Klavs, I.; Korošec, A.; Aracil, B.; Asensio, A.; Pérez-Vázquez, M.; Billström, H.; Larsson, S.; Reilly, J. S.; Johnson, A.; Hopkins, S. Attributable Deaths and Disability-Adjusted Life-Years Caused by Infections with Antibiotic-Resistant Bacteria in the EU and the European Economic Area in 2015: A Population-Level Modelling Analysis. *Lancet Infect. Dis.* **2019**, *19*, 56−66.

(4) Heath, R. J.; Rock, C. O. Enoyl-Acyl Carrier Protein Reductase (FabI) Plays a Determinant Role in Completing Cycles of Fatty Acid Elongation in Escherichia Coli(∗). *J. Biol. Chem.* **1995**, *270*, 26538−26542.

(5) Yao, J.; Rock, C. O. Exogenous Fatty Acid Metabolism in Bacteria. *Biochimie* **2017**, *141*, 30−39.

(6) McMurry, L. M.; Oethinger, M.; Levy, S. B. Triclosan Targets Lipid Synthesis. *Nature* **1998**, *394*, 531−532.

(7) Rawat, R.; Whitty, A.; Tonge, P. J. The Isoniazid-NAD Adduct Is a Slow, Tight-Binding Inhibitor of InhA, the Mycobacterium Tuberculosis Enoyl Reductase: Adduct Affinity and Drug Resistance. *Proc. Natl. Acad. Sci.* **2003**, *100*, 13881−13886.

(8) Kaplan, N.; Albert, M.; Awrey, D.; Bardouniotis, E.; Berman, J.; Clarke, T.; Dorsey, M.; Hafkin, B.; Ramnauth, J.; Romanov, V.; Schmid, M. B.; Thalakada, R.; Yethon, J.; Pauls, H. W. Mode of Action, In Vitro Activity, and In Vivo Efficacy of AFN-1252, a Selective Antistaphylococcal FabI Inhibitor. *Antimicrob. Agents Chemother.* **2012**, *56*, 5865−5874.

(9) Heerding, D. A.; Chan, G.; DeWolf, W. E.; Fosberry, A. P.; Janson, C. A.; Jaworski, D. D.; McManus, E.; Miller, W. H.; Moore, T. D.; Payne, D. J.; Qiu, X.; Rittenhouse, S. F.; Slater-Radosti, C.; Smith, W.; Takata, D. T.; Vaidya, K. S.; Yuan, C. C. K.; Huffman, W. F. 1,4-Disubstituted Imidazoles Are Potential Antibacterial Agents Functioning as Inhibitors of Enoyl Acyl Carrier Protein Reductase (FabI). *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2061−2065.

(10) Ward, W. H. J.; Holdgate, G. A.; Rowsell, S.; McLean, E. G.; Pauptit, R. A.; Clayton, E.; Nichols, W. W.; Colls, J. G.; Minshull, C. A.; Jude, D. A.; Mistry, A.; Timms, D.; Camble, R.; Hales, N. J.; Britton, C. J.; Taylor, I. W. F. Kinetic and Structural Characteristics of the Inhibition of Enoyl (Acyl Carrier Protein) Reductase by Triclosan. *Biochemistry* **1999**, *38*, 12514−12525.

(11) Priyadarshi, A.; Kim, E. E.; Hwang, K. Y. Structural Insights into Staphylococcus Aureus Enoyl-ACP Reductase (FabI), in Complex with NADP and Triclosan. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 480−486.

(12) Kim, S. J.; Ha, B. H.; Kim, K.-H.; Hong, S. K.; Shin, K.-J.; Suh, S. W.; Kim, E. E. Dimeric and Tetrameric Forms of Enoyl-Acyl Carrier Protein Reductase from Bacillus Cereus. *Biochem. Biophys. Res. Commun.* **2010**, *400*, 517−522.

(13) Stewart, M. J.; Parikh, S.; Xiao, G.; Tonge, P. J.; Kisker, C. Structural Basis and Mechanism of Enoyl Reductase Inhibition by Triclosan. *J. Mol. Biol.* **1999**, *290*, 859−865.

(14) Rafi, S.; Novichenok, P.; Kolappan, S.; Zhang, X.; Stratton, C. F.; Rawat, R.; Kisker, C.; Simmerling, C.; Tonge, P. J. Structure of Acyl Carrier Protein Bound to FabI, the FASII Enoyl Reductase from Escherichia Coli*. *J. Biol. Chem.* **2006**, *281*, 39285−39293.

(15) Schiebel, J.; Chang, A.; Lu, H.; Baxter, M. V.; Tonge, P. J.; Kisker, C. Staphylococcus Aureus FabI: Inhibition, Substrate Recognition, and Potential Implications for In Vivo Essentiality. *Structure* **2012**, *20*, 802−813.

(16) *Nonactive-Site Mutations in S. aureus FabI That Induce Triclosan Resistance | ACS Omega*. https://pubs.acs.org/doi/10.1021/acsomega.0c02942 (accessed 2022-05-27).

(17) Lu, H.; Tonge, P. J. Inhibitors of FabI, an Enzyme Drug Target in the Bacterial Fatty Acid Biosynthesis Pathway. *Acc. Chem. Res.* **2008**, *41*, 11−20.

(18) Gerusz, V.; Denis, A.; Faivre, F.; Bonvin, Y.; Oxoby, M.; Briet, S.; LeFralliec, G.; Oliveira, C.; Desroy, N.; Raymond, C.; Peltier, L.; Moreau, F.; Escaich, S.; Vongsouthi, V.; Floquet, S.; Drocourt, E.; Walton, A.; Prouvensier, L.; Saccomani, M.; Durant, L.; Genevard, J.-M.; Sam-Sambo, V.; Soulama-Mouze, C. From Triclosan toward the Clinic: Discovery of Nonbiocidal, Potent FabI Inhibitors for the Treatment of Resistant Bacteria. *J. Med. Chem.* **2012**, *55*, 9914−9928.

(19) Bryskier, A. *MUT056399: Mode of Action and Mechanisms of Resistance*, 2010.

from Molecular Docking and Molecular Dynamics. *J. Biomol. Struct. Dyn.* **2022**, *40*, 4021−4037.
(55) Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E. Long-Timescale Molecular Dynamics Simulations of Protein Structure and Function. *Curr. Opin. Struct. Biol.* **2009**, *19*, 120−127.
(56) Martinez-Rosell, G.; Giorgino, T.; Harvey, M. J.; de Fabritiis, G. Drug Discovery and Molecular Dynamics: Methods, Applications and Perspective Beyond the Second Timescale. *Curr. Top. Med. Chem.* **2017**, *17*, 2617−2625.
(57) Dror, R. O.; Jensen, M. Ø.; Borhani, D. W.; Shaw, D. E. Exploring Atomic Resolution Physiology on a Femtosecond to Millisecond Timescale Using Molecular Dynamics Simulations. *J. Gen. Physiol.* **2010**, *135*, 555−562.
(58) Kim, H. T.; Kim, S.; Na, B. K.; Chung, J.; Hwang, E.; Hwang, K. Y. Structural Insights into the Dimer-Tetramer Transition of FabI from Bacillus Anthracis. *Biochem. Biophys. Res. Commun.* **2017**, *493*, 28−33.
(59) Rao, K. N.; Lakshminarasimhan, A.; Joseph, S.; Lekshmi, S. U.; Lau, M.-S.; Takhi, M.; Sreenivas, K.; Nathan, S.; Yusof, R.; Abd Rahman, N.; Ramachandra, M.; Antony, T.; Subramanya, H. AFN-1252 Is a Potent Inhibitor of Enoyl-ACP Reductase from Burkholderia Pseudomallei−Crystal Structure, Mode of Action, and Biological Activity. *Protein Sci.* **2015**, *24*, 832−840.
(60) Rao, N. K.; Nataraj, V.; Ravi, M.; Panchariya, L.; Palai, K.; Talapati, S. R.; Lakshminarasimhan, A.; Ramachandra, M.; Antony, T. Ternary Complex Formation of AFN-1252 with Acinetobacter Baumannii FabI and NADH: Crystallographic and Biochemical Studies. *Chem. Biol. Drug Des.* **2020**, *96*, 704−713.
(61) Pantsar, T.; Kaiser, P. D.; Kudolo, M.; Forster, M.; Rothbauer, U.; Laufer, S. A. Decisive Role of Water and Protein Dynamics in Residence Time of P38α MAP Kinase Inhibitors. *Nat. Commun.* **2022**, *13*, 569.
(62) Amaral, M.; Kokh, D. B.; Bomke, J.; Wegener, A.; Buchstaller, H. P.; Eggenweiler, H. M.; Matias, P.; Sirrenberg, C.; Wade, R. C.; Frech, M. Protein Conformational Flexibility Modulates Kinetics and Thermodynamics of Drug Binding. *Nat. Commun.* **2017**, *8*, 2276.
(63) Berger, B.-T.; Amaral, M.; Kokh, D. B.; Nunes-Alves, A.; Musil, D.; Heinrich, T.; Schröder, M.; Neil, R.; Wang, J.; Navratilova, I.; Bomke, J.; Elkins, J. M.; Müller, S.; Frech, M.; Wade, R. C.; Knapp, S. Structure-Kinetic Relationship Reveals the Mechanism of Selectivity of FAK Inhibitors over PYK2. *Cell Chem. Biol.* **2021**, *28*, 686−698.e7.
(64) Glöckner, S.; Ngo, K.; Sager, C. P.; Hüfner-Wulsdorf, T.; Heine, A.; Klebe, G. Conformational Changes in Alkyl Chains Determine the Thermodynamic and Kinetic Binding Profiles of Carbonic Anhydrase Inhibitors. *ACS Chem. Biol.* **2020**, *15*, 675−685.
(65) Daina, A.; Michielin, O.; Zoete, V. SwissADME: A Free Web Tool to Evaluate Pharmacokinetics, Drug-Likeness and Medicinal Chemistry Friendliness of Small Molecules. *Sci. Rep.* **2017**, *7*, 42717.
(66) Parker, E. N.; Cain, B. N.; Hajian, B.; Ulrich, R. J.; Geddes, E. J.; Barkho, S.; Lee, H. Y.; Williams, J. D.; Raynor, M.; Caridha, D.; Zaino, A.; Shekhar, M.; Muñoz, K. A.; Rzasa, K. M.; Temple, E. R.; Hunt, D.; Jin, X.; Vuong, C.; Pannone, K.; Kelly, A. M.; Mulligan, M. P.; Lee, K. K.; Lau, G. W.; Hung, D. T.; Hergenrother, P. J. An Iterative Approach Guides Discovery of the FabI Inhibitor Fabimycin, a Late-Stage Antibiotic Candidate with In Vivo Efficacy against Drug-Resistant Gram-Negative Infections. *ACS Cent. Sci.* **2022**, *8*, 1145−1158.
(67) Robertson, J. C.; Hurley, N. C.; Tortorici, M.; Ciossani, G.; Borrello, M. T.; Vellore, N. A.; Ganesan, A.; Mattevi, A.; Baron, R. Expanding the Druggable Space of the LSD1/CoREST Epigenetic Target: New Potential Binding Regions for Drug-Like Molecules, Peptides, Protein Partners, and Chromatin. *PLoS Comput. Biol.* **2013**, *9*, No. e1003158.
(68) Hashimoto, K.; Nishi, H.; Bryant, S.; Panchenko, A. R. Caught in Self-Interaction: Evolutionary and Functional Mechanisms of Protein Homooligomerization. *Phys. Biol.* **2011**, *8*, No. 035007.
(69) Nishi, H.; Hashimoto, K.; Madej, T.; Panchenko, A. R. Evolutionary, Physicochemical, and Functional Mechanisms of Protein Homooligomerization. In *Progress in Molecular Biology and Translational Science*; Giraldo, J., Ciruela, F., Eds.; Oligomerization in Health and Disease; Academic Press, 2013; Vol. *117*, pp. 3−24, DOI: 10.1016/B978-0-12-386931-9.00001-5.

3.2.1 Supporting information

Supporting Information – Do go chasing waterfalls: enoyl reductase (FabI) in complex with inhibitors stabilize tetrameric structure and opens water channels

*Vinicius Gonçalves Maltarollo[1], Ekaterina Shevchenko[2], Igor Daniel de Miranda Lima[3], Elio A. Cino[3], Glaucio Monteiro Ferreira[4], Antti Poso[2,5], and Thales Kronenberger[2,5]**

[1] Departamento de Produtos Farmacêuticos, Faculdade de Farmácia, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Minas Gerais, Brazil

[2] Institute of Pharmacy, Pharmaceutical/Medicinal Chemistry and Tübingen Center for Academic Drug Discovery, Eberhard Karls University Tübingen, Auf der Morgenstelle 8, 72076 Tübingen, Germany. Department of Oncology and Pneumonology, Internal Medicine VIII, University Hospital Tübingen, Otfried-Müller-Straße 10, DE72076, Tübingen, Germany. Tübingen Center for Academic Drug Discovery & Development (TüCAD2), 72076, Tübingen, Germany

[3] Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

[4] Department of Clinical and Toxicological Analyses, School of Pharmaceutical Sciences, University of Sao Paulo, Av Prof Lineu Prestes 580, 05508-000, São Paulo, Brazil

[5] School of Pharmacy, Faculty of Health Sciences, University of Eastern Finland, 70211, Kuopio, Finland

Corresponding authors: thales.kronenberger@uni-tuebingen.de, (T.K.), +49 (0)7071-29 74575

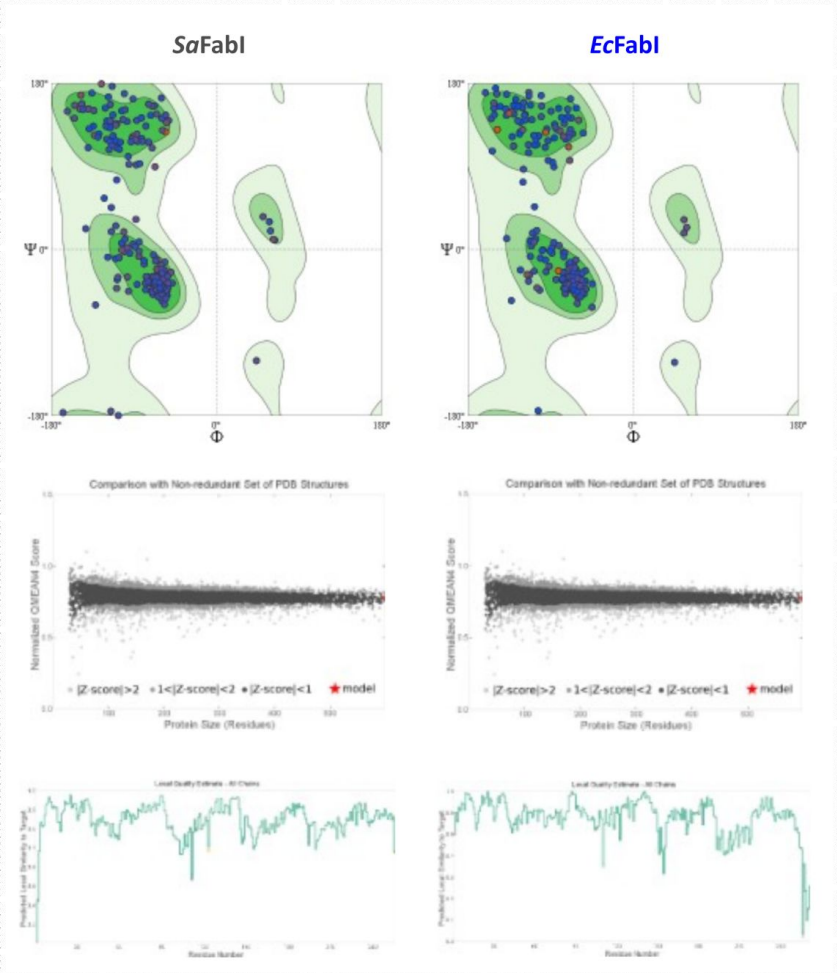


Figure S1. Ramachandran plots (upper graphs), QMEANS Z-score (medium panel), and local similarity to the target protein (lower graphs) generated by SWISS-Model.

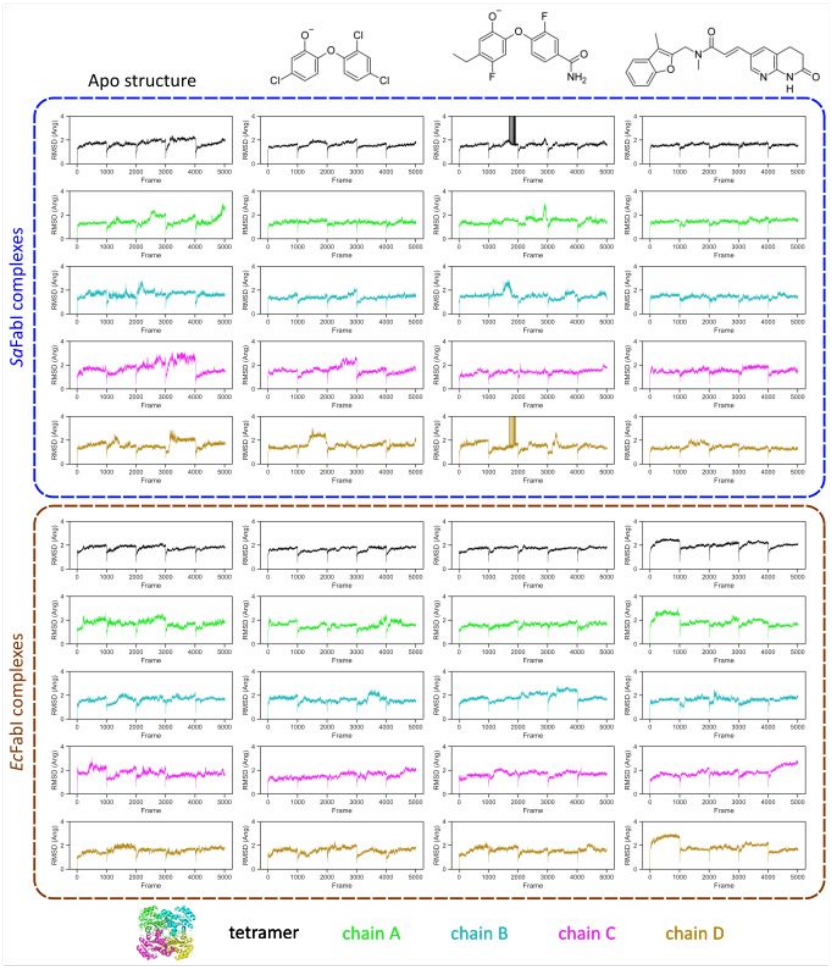


Figure S2. RMSD values of entire tetramer (black lines), and individual chains A (green lines), B (cyan lines), C (magenta lines), and D (yellow lines) of *Sa*FabI and *Ec*FabI systems.

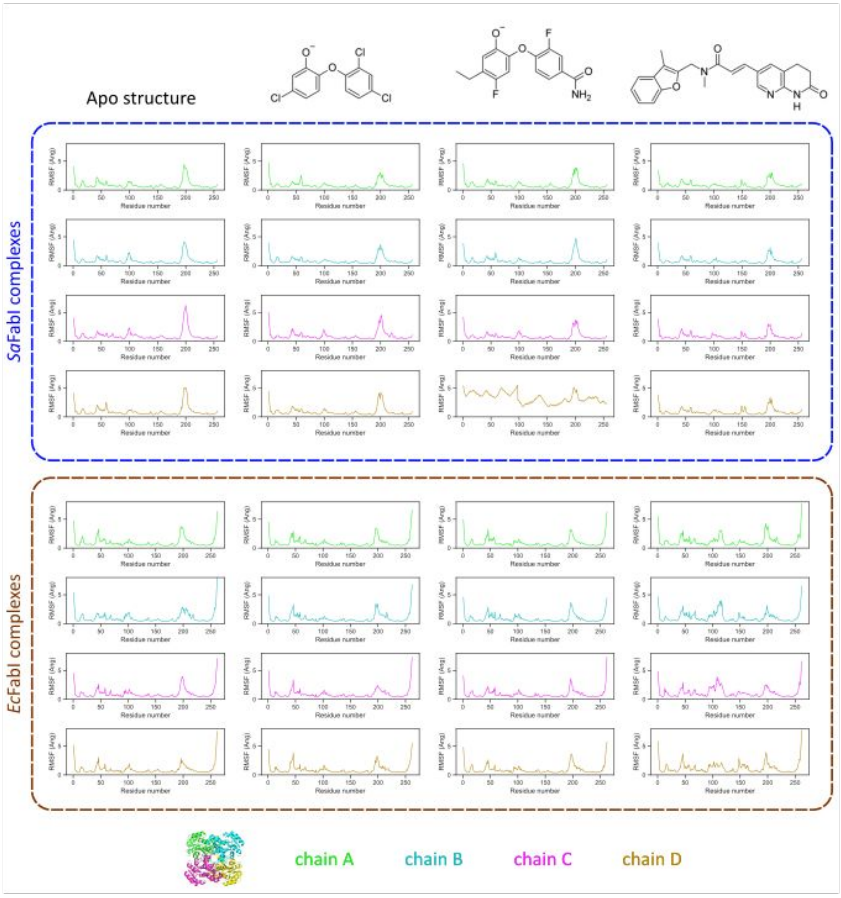


Figure S3. RMSF values of individual chains A (green lines), B (cyan lines), C (magenta lines), and D (yellow lines) of *Sa*FabI and *Ec*FabI systems.

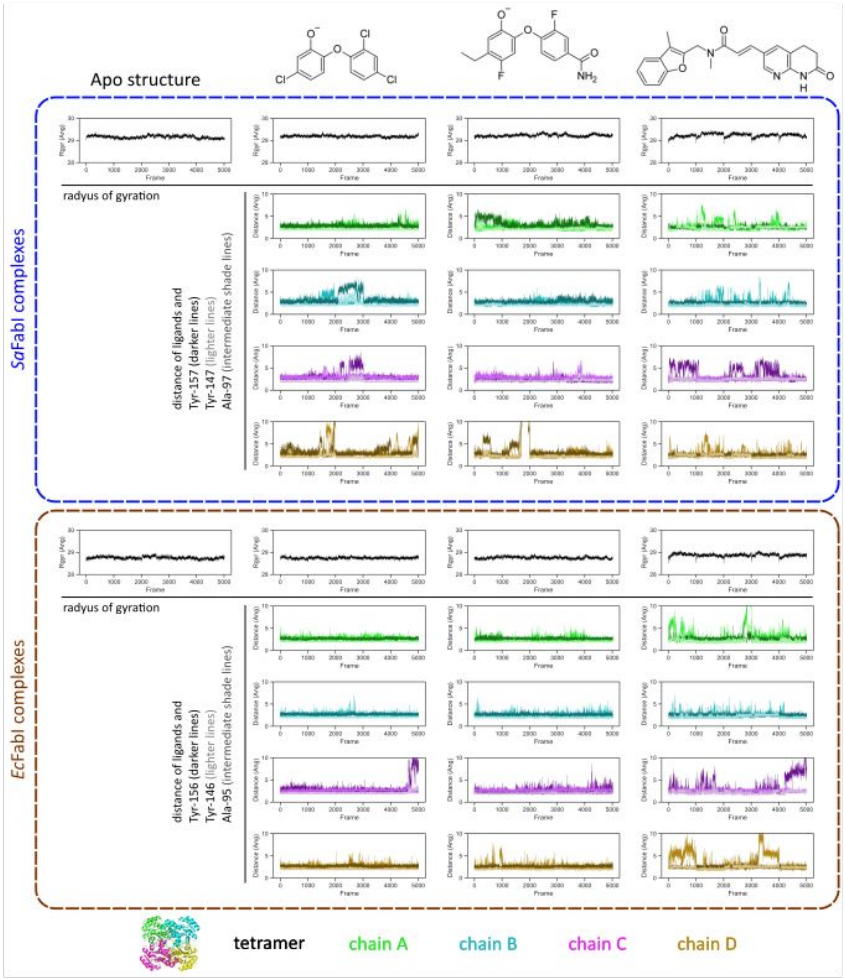


Figure S4. Radius of gyration (black lines) and distance of ligands' center of mass and Ala97/95, Tyr147/146, and Tyr157/156 for chains A (shades of green), B (shades of cyan), C (shades of magenta), and D (shades of yellow).

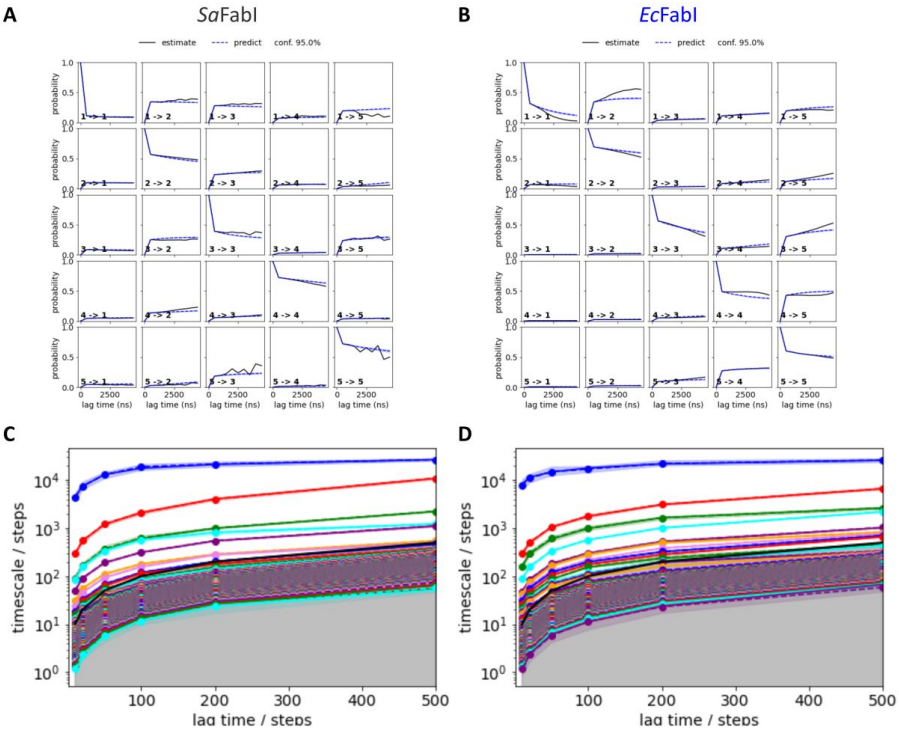


Figure S5. Validation of Markov State Models: Chapman-Kolmogorov tests demonstrate that the models follow expected estimates (A = *Sa*Fabl and B = *Ec*Fabl). In the selected lag times the implied timescales are converged (C = *Sa*Fabl and D = *Ec*Fabl).

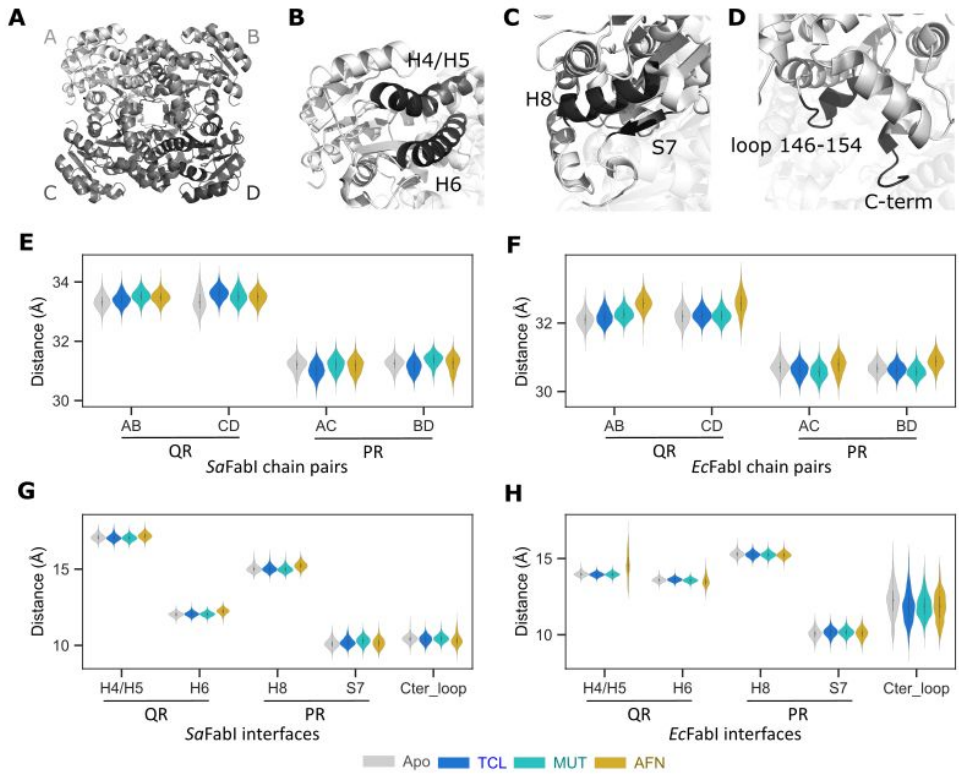


Figure S6. FabI tetramer highlighting substructures of chain A's QR and PR interfaces as well as the other B, C, and D chains represented in transparent cartoons (A). Representations of substructures present in QR interface (B), PR interface (C), and loop 146-154 / C-terminal interface (D). Violin plot of the distances between the centre of mass of each chain for *Sa*FabI (E) and *Ec*FabI (F), and the secondary structures present in the analysed interfaces for *Sa*FabI (G) and *Ec*FabI (H).

Comments: The folding of individual chains of tetramers was conserved without major changes during the trajectories. However, the distance between the centre of masses (COMs) of entire individual chains which compose the QR interface is on average smaller

for apo systems, indicating that the presence of inhibitors seems to change the overall conformation of the entire tetramer (Figures S4E,F). To illustrate it, the distance of chains A and B present in the QR interface of *Sa*FabI ranged from 33.3 Å in apo structure to 33.5 Å in AFN complex (Figure 4E). For *Ec*FabI, the average distance between chains A and B is equal to 32.1 Å (apo), 32.2 Å (TCL complex), 32.3 Å (MUT complex), and 32.6 Å (AFN complex). For chains C and D distances, the same finding was observed ranging from 32.2 Å (apo) to 32.6 Å (AFN complex). The same effect, but more slightly, was observed in the PR interface in the presence of inhibitors. This finding could be confirmed by the specific distances of secondary structures of QR interfaces (Figure S4G,H), specifically for the AFN complex. Interestingly, both *Sa*FabI and *Ec*FabI complexes with AFN stabilize more the C-terminal/loop (residues 146-154) interface than apo structures and other complexes.

77 **C-terminal conformational changes and interactions**

78 We studied the interaction frequency between the C-terminal's amino acids with other
79 parts of FabI. We observed fewer interactions in the inhibitor-bound complexes when
80 compared to the apo system. In particular, the C-terminal of AFN and *Sa*FabI complex
81 lost two out of four pi-cation interactions between Lys256 and Phe152 from distinct
82 chains (Table S1). Furthermore, one chain decreased by at least 20% the frequency of H-
83 bonds between the very same lysine and Ala153 from other chains. The same effect was
84 observed in the AFN and *Ec*FabI complex in comparison with other systems (Table S1).
85 In that case, the H-bond between Ala254 and Arg151 from distinct chains was lost for all
86 four chains.

87 As observed in our simulations, in both *Sa* and *Ec*FabI, C-terminal plays a role in
88 interactions (Table S1) with loop 146-154 of another monomer, that indirectly drive
89 active site conformational changes (see Figure 4). Until now, there are some indications
90 discussing the C-terminal effects over FabI activity. For example, the *Bacillus anthracis*
91 (*Ba*) FabI, when recombinantly expressed with a C-terminal tag, was reported as a dimer
92 while the untagged enzyme was reported as a tetramer. [1] In this work, Kim and
93 colleagues speculated that His_6-tag may block the dimer-tetramer *Ba*FabI transition. In
94 other words, they highlight the role of C-terminal interactions with other chains in
95 tetramer stabilization. Furthermore, the work of Radka and colleagues generated a series
96 of truncated mutants at C-terminal regions of *Af*FabI and found that mutants have
97 decreased catalytic efficiencies. Those mutations that prevented a disordered C-terminal
98 may decrease the interactions with other subunits in tetramer form and/or sealed water
99 channels related to enzymatic function. [2]

100

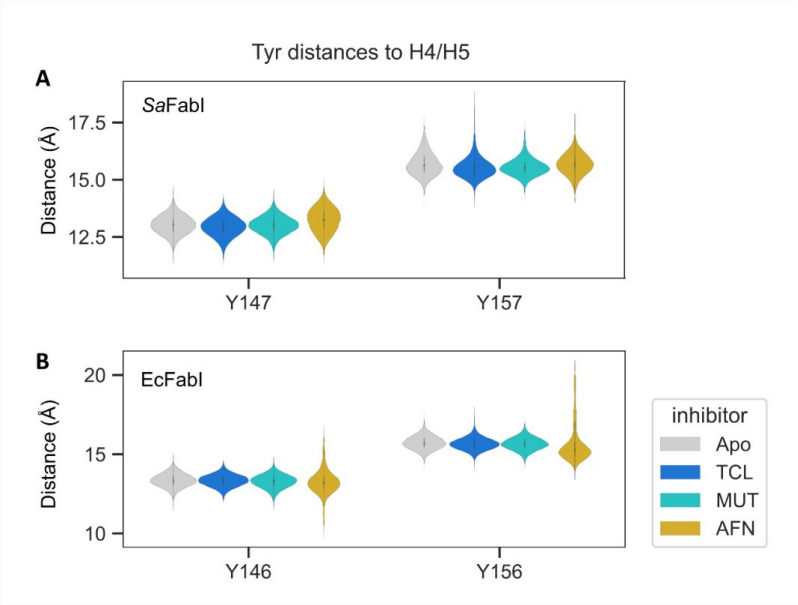
Table S1. Interactions of C-terminal amino acids from *Sa*FabI and *Ec*FabI and other parts of the tetramer are calculated by the simulation interaction diagram.

***Sa*FabI**		**Apo (%)**	**TCL (%)**	**MUT (%)**	**AFN (%)**
$Ala254_A$-$Ala153_D$	H-bond	99	99	99	99
$Lys256_A$-$Ala153_D$	H-bond	73	65	65	67
$Lys256_A$-$Phe152_D$	Pi-cation	38	35	32	-
$Ala254_B$-$Ala153_C$	H-bond	99	99	99	96
$Lys256_B$-$Ala153_C$	H-bond	58	62	52	69
$Lys256_B$-$Phe152_C$	Pi-cation	44	31	38	40
$Ala254_C$-$Ala153_B$	H-bond	99	99	99	99
$Lys256_C$-$Ala153_B$	H-bond	75	66	72	41
$Lys256_C$-$Phe152_B$	Pi-cation	80	33	-	31
$Ala254_D$-$Ala153_A$	H-bond	99	99	99	98
$Lys256_D$-$Ala153_A$	H-bond	61	74	59	60
$Lys256_D$-$Phe152_A$	Pi-cation	30	31	45	-

***Ec*FabI**		**Apo (%)**	**TCL (%)**	**MUT (%)**	**AFN (%)**
$Ser252_A$-$Arg151_D$	H-bond	45	55	52	31
$Ala254_A$-$Arg151_D$	H-bond	39	48^{H2O}	52	-

$Ser252_B$-$Arg151_C$	H-bond	46	59	-	30
$Ala254_B$-$Arg151_C$	H-bond	31	64^{H2O}	52^{H2O}	-
$Ser252_C$-$Arg151_B$	H-bond	45	36	45	66
$Ala254_C$-$Arg151_B$	H-bond	40	50	75	-
$Ser252_D$-$Arg151_A$	H-bond	58	60	44	44
$Ala254_D$-$Arg151_A$	H-bond	53	68	41^{H20}	-

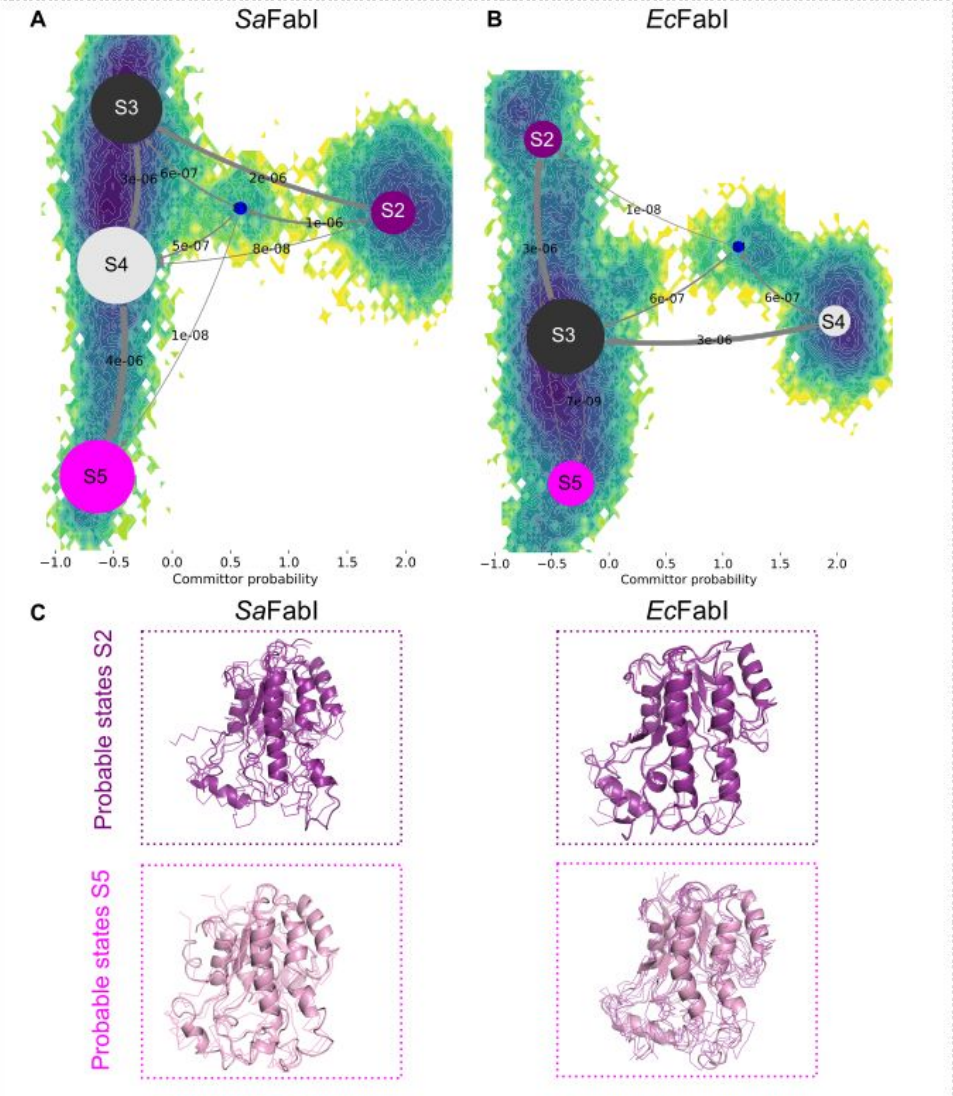
For *Ec*FabI Ser252 and Ala254 = mainchain / Gln258, Arg151 and Asn175 = sidechain. (-) means that interactions were present in less than 30% of trajectory. H_2O indicates water-mediated hydrogen bonds



107

108 **Figure S7.** Distances between the carbon alpha atoms' for Tyr146 and Tyr156 towards
109 the centre of mass of the H4/H5 (represented by Ser122/121's Cα), for *Sa*FabI (A) and
110 *Ec*FabI (B) simulations.

111



112

113 **Figure S8.** Metastable states of *Ec*Fbl and *Sa*Fbl structure revealed by Markov state

114 modelling. (A = *Sa*Fabl and B = *Ec*Fabl) Committor probability of the most representative

115 metastable states. (C) Each metastable state (S) is illustrated with ten representative

116 structures (coloured cartoons and ribbons), superimposed to a transparent cartoon with

117 the original crystal structure.

118

119

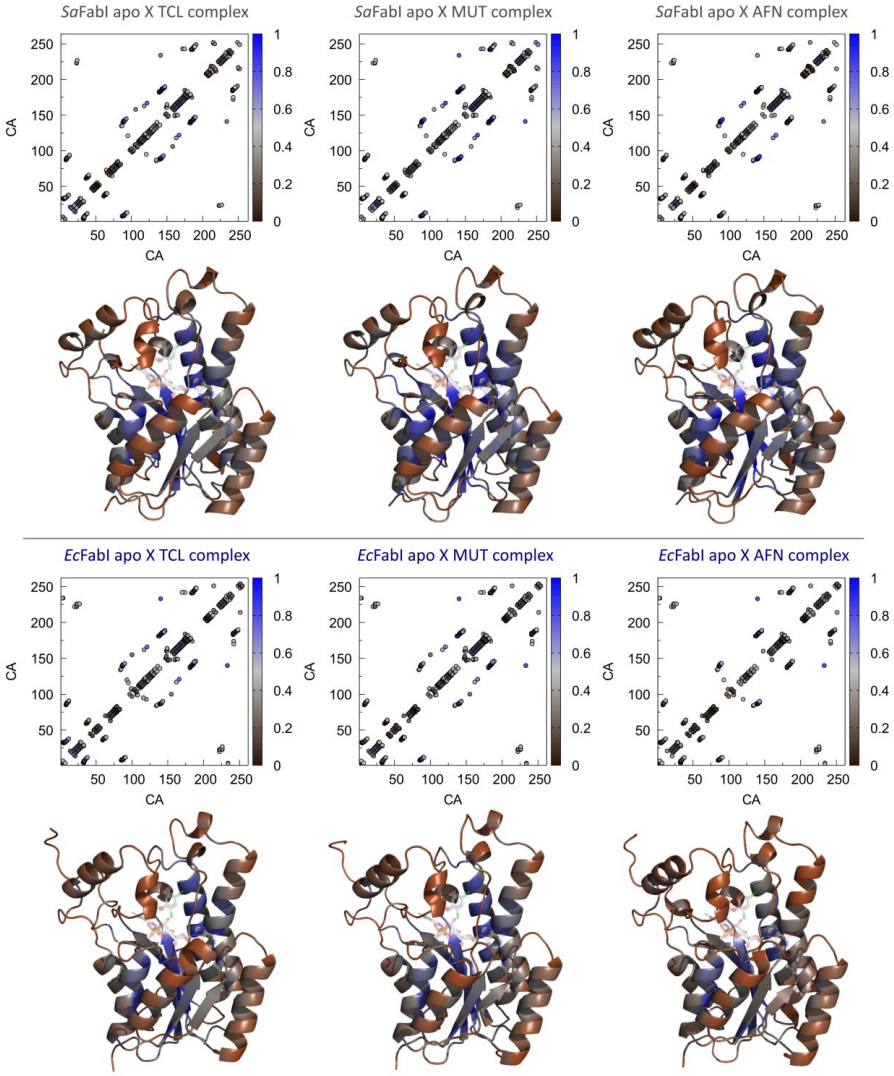


Figure S9. SS3D contact matrices for TCL-, MUT-, and AFN-FabI complexes in comparison with apo structures from both *S. aureus* and *E. coli*. Below contact plots, the structures of correspondent complexes were coloured according to 3D similarity ranging from brown (maximum dissimilarity), passing through grey and reaching blue regions (maximum similarity) are presented.

127 References

128 [1] Kim HT, Kim S, Na BK, Chung J, Hwang E, Hwang KY. Structural insights into
129 the dimer-tetramer transition of FabI from Bacillus anthracis. Biochemical and
130 Biophysical Research Communications 2017;493:28–33.
131 https://doi.org/10.1016/j.bbrc.2017.09.084.
132 [2] Radka CD, Frank MW, Yao J, Seetharaman J, Miller DJ, Rock CO. The genome of
133 a Bacteroidetes inhabitant of the human gut encodes a structurally distinct enoyl-
134 acyl carrier protein reductase (FabI). Journal of Biological Chemistry
135 2020;295:7635–52. https://doi.org/10.1074/jbc.RA120.013336.
136
137

VIII. Acknowledgements

The way this thesis lays through the years of the pandemic was indeed more complex than it was expected to be. Therefore, I would like to make the motto of this dissertation a quote from the Markov process, which endorses the existence of a random process in which the future is independent of the past, given the present.

I would like to thank my supervisors for lighting me the way into the world of science. Antti Poso, for believing in me, supporting my frequently not-trivial ideas, and always being there when I needed advice and help.

Stefan Laufer, for comprehensive pharmaceutical chemistry guidance, collaborations and making our lab in Morgenstelle come together. I am incredibly grateful to name these two outstanding professors and outstanding personalities as my supervisors.

Another person without whom this work will be fairly possible – is Thales Kronenberger. I would like to say that he is the warmest, kindest person that I have ever had a chance to meet. He is the person who welcomed me in Germany and educated me in every possible aspect of life – from establishing collaborations and achieving scientific thinking to overcoming highly stressful situations and keeping a clean mind. I hope the scientific journey that we have started to walk together is to this date in the very beginning.

I want to thank our Morgenstelle lab for lots of working hours and beyond that were spent together. These people definitely shaped my perception of Germany.

Before switching to the more personal part of acknowledgements, I would like to express my appreciation for CSC Finland, who allowed me to work with the HPC clusters and made long-scale Molecular Dynamics possible. Another gratitude is for the Tübingen cluster of excellence – iFIT and Baden-Württemberg Stiftung for providing the opportunities to make the science happen.

Switching from scientific life to private – I want to grate enormous gratitude for my Mother, who was, from my very childhood, the best example in my life and the main driving force behind every achievement I have made. I would like to thank her for every sacrifice she made to make my life go this way as it goes today. I am grateful that this powerful and outstanding woman was raising me and is every day behind my back despite the kilometres between us.

I would like to thank my friends (mainly three of them) from my home side, who were there for me during the sleepless nights in the last three years. They supported me during my transition to Germany and didn't give up on our friendship despite the distance and the lack of possibility of seeing each other in years.

I would love to say thousands of thanks to one more wonderful person whom I have the privilege to call my partner. He is the most supportive, kind-hearted and tolerant of my frequently unbearable personality. Without him, I can’t imagine I had the courage to come this far.