DOCTOR OF PHILOSOPHY

Molecular Modelling Of New Viral Targets From SARS-COV-2 To HIV

Pipito, Ludovico

Award date:
2024

Awarding institution:
Coventry University

Molecular Modelling of New Viral Targets: from SARS-CoV-2 to HIV

By

Ludovico Pipitò

PhD

January 2024

Molecular Modelling of New Viral Targets: from SARS-CoV-2 to HIV

By

Ludovico Pipitò

January 2024

A thesis submitted in partial fulfilment of the University's requirements for the Degree of Doctor of Philosophy.

Certificate of Ethical Approval

Applicant: Ludovico Pipito

Project Title: MOLECULAR MODELLING OF SARS-COV-2 ACE2 INTERACTIONS AND DRUG DISCOVERY OF POTENTIAL DISRUPTORS

This is to certify that the above named applicant has completed the Coventry University Ethical Approval process and their project has been confirmed and approved as Low Risk

Date of approval: 01 Feb 2021

Project Reference Number: P117209

Table of Contents:

6. SARS-CoV-2 Project Publications:

6.1. Literature review: Molecular dynamics studies reveal structural and functional features of the SARS-CoV-2 spike protein

6.2. Understanding The Role of Heparinoids on the SARS-CoV-2 Spike Protein through Molecular Dynamics Simulations

6.3. Is the Stalk of the SARS-CoV-2 Spike Protein Druggable?

6.4. A Pathway Model to Understand the Evolution of Spike Protein Binding to ACE2 in SARS-CoV-2 Variants

7. HIV Project Publications:

7.1. Computer-aided De Novo Design and Optimization of Novel Potential Inhibitors of HIV-1 Nef Protein

8. Extra-project Publications and Preprints:

8.1. Targeting hPKM2 in cancer: a bio isosteric approach for ligand design

8.2. Hidden GPCR structural transitions addressed by multiple walkers supervised molecular dynamics (mwSuMD)

9. Concluding Remarks and Future Research Directions

Thesis structure

Chapter 1: Introduction

This section provides a general overview of the project and background of the SARS-CoV-2 and HIV. In this section, I describe the methodologies used and the objectives pursued during my investigations .

Chapter 2: Molecular Binding's Driving Forces

This section describes general biophysical concepts at the base of molecular interaction. This chapter includes an overview of the fundamental atomic interactions, including the role of water in biological systems.

Chapter 3: Computer-Aided Drug Design: Modern Tools in Drug Discovery

This chapter introduces different approaches and concepts in modern drug discovery investigations with focus on field-based methods and their conceptual development.

Chapter 4: Fundamentals of Molecular Docking and Dynamics

The general principles behind molecular docking and dynamics are described. This section presents the fundamentals behind the design of the studies adopted throughout this project.

Chapter 5: References

Chapter 6: SARS-CoV-2 Project Publications

The results of the project were published in peer-reviewed journals. This section is an enriched paper collection of published manuscripts on SARS-CoV-2, which includes most of the methods deployed.

Chapter 7: HIV Project Publications

This section includes the published works targeting HIV's viral proteins. These publications address HIV's functional proteins through a structure-based drug design approach. These manuscripts describe additional techniques adopted for molecular modelling and drug design.

Chapter 8: Extra-project Publication and Preprints

This section is a collection of all the publications and manuscripts not related to SARS-CoV-2 or HIV including original publications and collaborations with other universities and collaborators.

Chapter 9: Concluding Remarks and Future Research Directions

Conclusions and reflections upon the achievements of this project and the future research perspectives.

Abstract

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus that caused the COVID-19 pandemic in 2020 and the human immunodeficiency virus (HIV) which leads to the acquired immunodeficiency syndrome (AIDS) have deeply affected the lives of more than 650 and 38 million people respectively in the last years. SARS-CoV-2 killed 6.5 million people in the last three years while HIV is accounted for more than 800 thousand deaths worldwide with a worrying yearly increase. While the COVID-19 outbreak has been tapered down by the advent of vaccines, the AIDS epidemic remains a global threat with 300 thousands new infections in 2023, as reported by the Joint United Nations Programme on HIV and AIDS reports. Despite the deployment of antiretroviral therapies, no effective vaccine is available on the market yet for HIV. The need to promptly identify effective drugs for HIV and COVID-19 treatment requires multiple scientific efforts and technologies to quickly reduce the high cost in terms of human lives and quality of life. The biopharmaceutical research sector has extensively used the computational technologies of the last twenty years in protein sequencing, drug design, cheminformatics, and artificial intelligence to meet unmet medical needs and anticipate emergencies. In the last decade, the increase in computational power improved pharmaceutical research by drastically reducing the time required for molecular modelling and computational chemistry calculations. With graphic processing units (GPU) becoming more accessible, computational chemistry and molecular modelling approaches are becoming more popular in drug discovery, providing atomic-level details and insights into the target-ligand molecular recognition mechanism.

This PhD project began when there were no drugs or vaccines available to efficiently treat SARS-CoV-2 infection and studies on the SARS-CoV-2 spike protein (S protein) were at their early stages. Later on, SARS-CoV-2's Alpha, Delta and Omicron variants were spreading and little was known about their infectivity or binding patterns or their antibody-escaping. Furthermore, In light of the coagulopathy effects triggered by SARS-CoV-2 infection, the role of heparinoids was being investigated, but their mechanism was unclear and experimental observations were conflicting.

During this PhD project, I identified and evaluated potential molecular candidates for the S protein, mapped conserved cryptic binding pockets on the S protein's stalk, characterised the S protein's binding patterns with the angiotensin-converting enzyme (ACE2) and defined the role of heparinoids as potential allosteric regulators. For the second part of this PhD project, I proposed a promising molecular candidate against the HIV's Negative factor protein (Nef). This manuscript will include a series of published works in

which we investigated the aforementioned molecular machinery structures and proposed working hypotheses based on our results.

Dedication

To the revered Masters, to whom my everlasting gratitude is bestowed. From Descartes to Sylvester, from Newton to Schrödinger, from Einstein to Dirac, from Hartree and Fock to Kohn and Hohenberg, from Fermi to Hawking. And to all the brilliant minds of these recent centuries, whose endeavours tirelessly strive to illuminate the complex laws of Nature guiding humanity to envision the boundlessly vast and infinitesimally small in the pursuit of Truth.

To the elevated and refined intellects of my mentors: to Doctor Deganutti, for his mastery and immense knowledge are paired with a kind spirit, an endless enthusiasm and an inextinguishable ardour for knowledge. My thanks and boundless appreciation extend to him, for his exemplary guidance has spurred me to consistently offer my utmost with assurance and optimism.

To Professor Reynolds: for his guidance and support throughout this PhD. For engaging in enlightening talks and constructive criticisms, stimulating my passion for molecular dynamics and the quantum world. Because he is an example of patience, generosity and trust.

To my esteemed colleagues who have collaborated and stood by me throughout these years. To Peter Griffin, whose adept ability to offer the right words of support and encouragement has been a source of solace during challenging moments. To Tal Weizmann, for your gracious and amiable conversations during periods of separation and constraints. To Roxana-Maria Rujan, whose invaluable support proved to be a beacon of light and kindness during darker moments. To Zamara Mariam, for her positivity and benevolence that inspires those around her. I wish you all, all the best.

To my life partner Antonella. Because without her immense love capable of enveloping the world, her patience, her help and her colours, nothing would have been possible. To her who at my side embraced my past, my pain and my present, filling a void that I thought was unbridgeable. I dedicate this work and my eternal love.

To my family for their love and support. In loving memory of Pina Scilla, Gino Cafeo and Maria Urbano.

Aim of The Project

This project was designed to expand our knowledge of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and propose molecular candidates that could prevent human immunodeficiency virus (HIV) progression to acquired immunodeficiency syndrome (AIDS). The proposed hypotheses were based on the experimental evidence available in the literature to investigate SARS-CoV-2 and HIV's crucial proteins through structural modelling and computational drug discovery methods. The rationale behind the structural approach is that a more profound knowledge of the proteins' anatomy and their mechanism could support the development of valid therapeutics through already existing and new methods. The new computational methods developed during this project expand and streamline the existing technologies to respond faster to unmet medical needs. SARS-CoV-2 studies were focused on the receptor binding domain (RBD) and the S2 domain (the stalk), while HIV's works were centred on the multifunctional negative factor (Nef).

The wide array of techniques implemented included molecular docking, classic molecular dynamics (cMD) - with the newly developed multiple-walker supervised molecular dynamics (mwSuMD) -, and mixed MD (mixMD). These techniques provided atomic-level details of the target-ligand interactions useful to characterise the molecular recognition mechanisms. Furthermore, we developed and deployed a set of computational tools which include a big-data scraping pipeline for large chemical databases, a high-throughput pipeline for molecular dynamics simulations bridged to virtual screening, an enhanced parallelised supervised molecular dynamic method and an explorative pipeline for antibody-cryptic epitope finding which explores possible antibody (Ab) and antigen (Ag) complex formation and evaluates their energy for a comparative screening.

1 Introduction

Human coronaviruses (HCoVs) have been associated with a wide range of respiratory illnesses with different severity degrees up to severe pneumonia (1). The viral infection aggravates the condition of adults with pre-existing diseases such as cardiovascular, hepatic, respiratory, gastrointestinal, and neurological diseases (2). The worsening symptoms are associated with the hyperbolic expression of proinflammatory signals and cytokines such as interleukins, interferon-gamma (IFN-γ), interferon-gamma induced protein 10 (IP-10), macrophage inflammatory protein 1A and 1B (MIP-1A, MIP1-B), platelet-derived growth factor (PDGF), tumour necrosis factor (TNF-α), and vascular endothelial growth factor (VEGF) (3).

There is an urgent need to understand the aetiological causes of the bleeding and thrombotic manifestations associated with COVID-19 coagulopathy, as well as a clinical testing protocol to predict bleeding, thrombosis, and severity of illness. The elevated level of D-dimer (4) with a prolonged prothrombin time has been associated with poor prognosis and the increased need for critical care (5). Disseminated intravascular coagulation (DIC) has been reported to develop in the majority of infection-related deaths (more than 70%) (5,6). Furthermore, infected patients had lower antithrombin values but higher D-dimer, fibrin/fibrinogen degradation products (FDP), and fibrinogen (FIB) in all severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) cases (6). D-dimer and FDP values were higher amongst those patients with aggressive SARS-CoV-2 infection than those with milder forms (6), however, the aetiology of the DIC is still unclear.

1.1 SARS-CoV-2 Overview.

The *Coronaviridae* family is characterised by an enveloped positive-stranded, non-segmented ribonucleic acid (RNA) with a genome of about 30 Kb (7). The severe acute respiratory syndrome (SARS-CoV) polymerase (Pol) and Nucleocapsid (N) genes were first used to reconstruct the phylogenetic tree for CoVs (8). Subsequentially, a whole genome-based phylogenetic analysis indicated SARS-CoV and Middle East respiratory syndrome coronavirus (MERS-CoV) as members of BetaCoV lineage B, characterized by high pathogenicity (9), suggesting bats as the initial viral reservoir. The RNA-dependent RNA polymerase (RdRP) and spike (S) protein gene sequence phylogenetic analysis confirmed SARS-CoV as a member of the BetaCoV subgroupEickmann et al., 2003.

Human coronaviruses (HCoVs) are composed of five major structural components: S protein, membrane protein (M), envelope protein (E), nucleocapsid proteins (N), and hemagglutinin (HA) as summarized in **Table 1.** Intriguingly, SARS-CoV-2 lacks HA (10) relying on the S protein for the attachment process (11). The S, M, and E proteins are

embedded in the viral envelope while the N protein protects the viral RNA genome located inside the virus (8) (Figure 1).

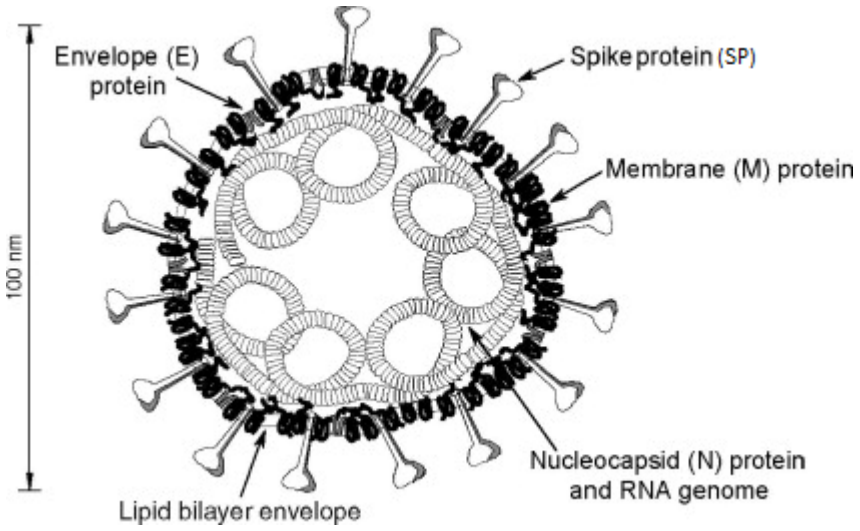


Figure 1 | Schematic representation of SARS-CoV-2 proteins. Minimal representation of a SARS-CoV-2 virion. SARS-CoV-2 lacks the HA protein with the "crown-like" S protein disposition. (adapted from (12)).

Table 1 - HCoVs Structural Proteins and Their Function

Protein	Role
Spike Protein (S protein)	ACE2 recognition and primary infection mechanism (13)
Membrane Protein (M)	Involved in SP glycosylation, recruits N and RNA (14)
Envelope Protein (E)	Favours membrane curvature, and mediates host immune responses through pore-forming and PDZ interaction (14)
Nucleocapsid Protein (N)	RNA-binding protein that favours viral compartment organization (15)
Hemagglutinin (HA)	Mediates attachment to O-acetylated sialic acids (10)

Furthermore, SARS-CoV-2 contains sixteen non-structural proteins (NSP1−16) which exert a range of necessary functions for viral replication as reported in Table 2. NSPSs play a crucial role in the activation of the post-infection viral replicative process by coordinating and favouring both the virion compartmentation and the assembly of the RNA (12).

Table 2 - Non-Structural-SARS-CoV-2 Proteins and their function (16)

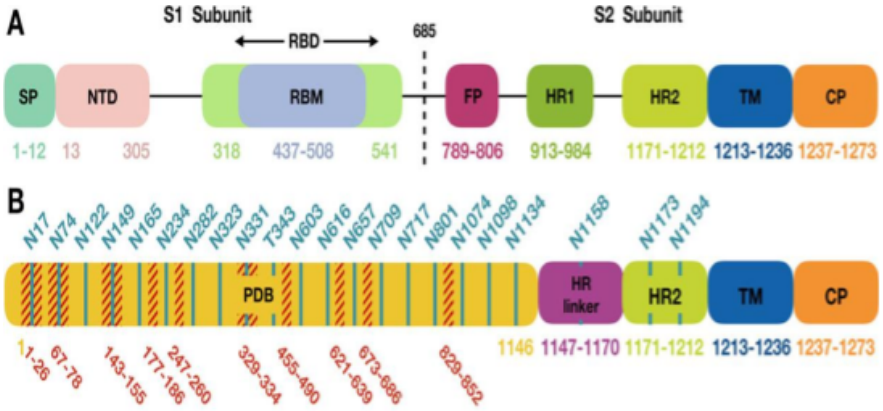
Name	Role
NSP1	RNA Replication
NSP2	Impairs Interferon production
NSP3	Separate post-translated proteins
NSP4	Contains transmembrane domain 2 (TM2) and modifies ER membranes
NSP5	Involved in protein polyprotein cleavage
NSP6	Transmembrane domain
NSP7	Favours Nsp12-Template-primer RNA interactions
NSP8	Favours Nsp12-Template-primer RNA interactions
NSP9	ssRNA-binding protein
NSP10	Responsible for cap methylation of viral mRNA
NSP11	Disordered protein with a supposed role in infection
NSP12	Contains RND-dependant RNA polymerase (RdRp)
NSP13	Adenosine triphosphate (ATP) and Zinc-binding domain
NSP14	Exoribonuclease domain
NSP15	Mn^{2+}-dependant endoribonuclease
NSP16	2'-O-ribose methyltransferase

1.1.1 SARS-CoV-2 Spike Protein

The SARS-CoV-2 infection mechanism depends on the transmembrane S protein (17,18), a highly conserved structure amongst the *coronaviridae* family responsible for extracellular binding and cell membrane fusion (19). The S protein characterises the shape of the *coronaviridae* family with the crown-like aspect they are named after (20) and shows a selective affinity for the angiotensin-converting enzyme-2 receptor (ACE2). ACE2 is a type 1

transmembrane protein with an external peptidase domain normally responsible for the conversion of angiotensin hormone into angiotensin II (21).

The S protein is a homotrimer class I fusion protein, with each protomer composed of domains S1 and S2 (13) (Figure 2). The S1 structure is responsible for binding with the ACE2 receptor (Figure 2A-C), before the conformational change in the stalk-like structure (Figure 2C) of the S2 subunit (22) and the subsequent membrane fusion after the cleavage of S1 from S2 by the host transmembrane protease serine 2 (TMPRSS2) (20).



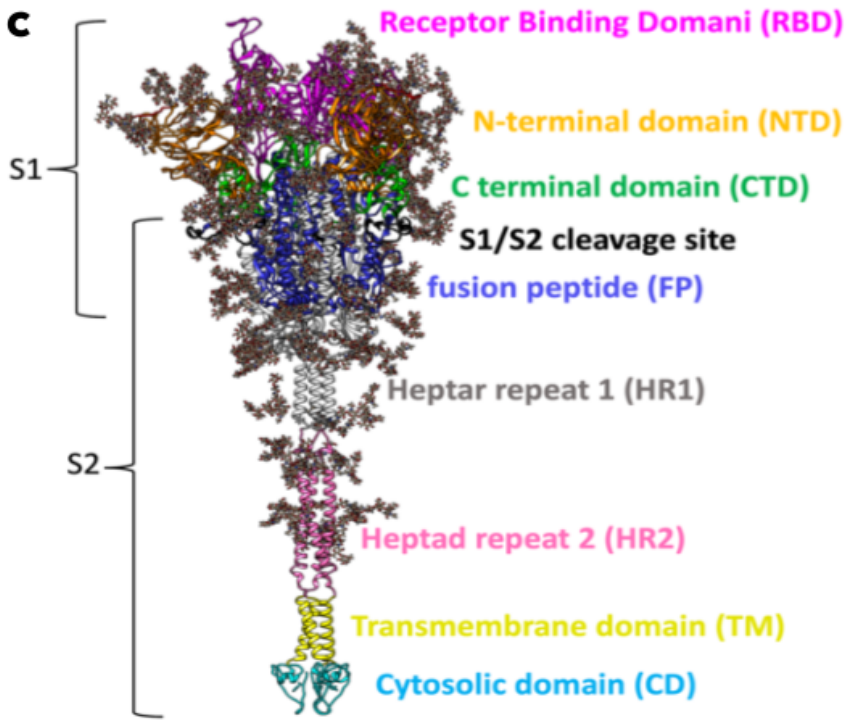


Figure 2 | Sequence and structure of the S protein domains in the pre-fusion state. A) SARS-CoV-2 S protein domains: signal peptide (SP), N-terminal domain (NTD), Receptor Binding Domain (RBD), receptor binding motif (RBM), fusion peptide (FP), heptad repeat 1 (HR1), heptad repeat 2 (HR2), transmembrane domain (TM), and cytosolic domain (CP) (adapted from (23)). **B)** N-Glycosylation sites or missing loops longer than 10 residues in PDB: 6VSB. (adapted from (23)). **C)** Fully glycosylated S protein (https://charmm-gui.org/?doc=archive&lib=covid19) with the S1 and S2 units highlighted and a colour-coded domain representation. Glycans are represented as sticks (24).

Cryogenic electron microscopy (Cryo-EM) studies revealed the metastable dynamic of the S protein where the arrangement of the N-terminal domain (NTD), receptor binding domain (RBD), and C-terminal domain (CTD) is folded over the trimer axis (Figure 2), covering the S2 domain residues at the S1/S2 furin cleavage site (13). In the S1 ectodomain (Figure 2), the apical portion of the S protein, composed of the RBD, NTD, and two CTD, folds in a hairpin motif that protects the prefusion conformation of S2 from the external environment (25). In its prefusion metastable state, it is speculated that a disulfide bond between C15-C136 could structure the apical portion of S1 (26). Downstream the fusion peptide (FP) another possible disulfide bond is between C840-C851, reinforced by a salt bridge between K835 and D848 and an extensive hydrogen bond network (27).

A distinguishing feature of the SARS-CoV-2 strain is an insertion of a polybasic PRRAR sequence in the protease S1/S2 cleavage site region, rich in arginine residues, which configures a furin recognition site, commonly found in highly virulent influenza viruses (28). The cleavage of the inter-region S1/S2 (Figure 2) allows for S2 structural conformation changes necessary for membrane fusion and post-fusion structure adaptation (19).

Upon RBD binding with ACE2, the S protein undergoes a conformational rearrangement with the HR1 forming an elongated three-stranded coiled axis while the S1/S2 form a three-stranded beta-sheet. The C-terminal segment of HR2 closes toward the viral membrane, shortening the three helical regions forming the main axis, which are reinforced by two disulfide bonds between C1064-C1108. The three helical regions are bundled against the groove of the axis formed by the central helices to form a short, rigid six-helix bundle structure (19) (Figure 3). The N- and C-termini regions of HR2 form another six-helix bundle structure, coated by the glycans on N1098, N1134, N1158, N1173, and N1194 (27).

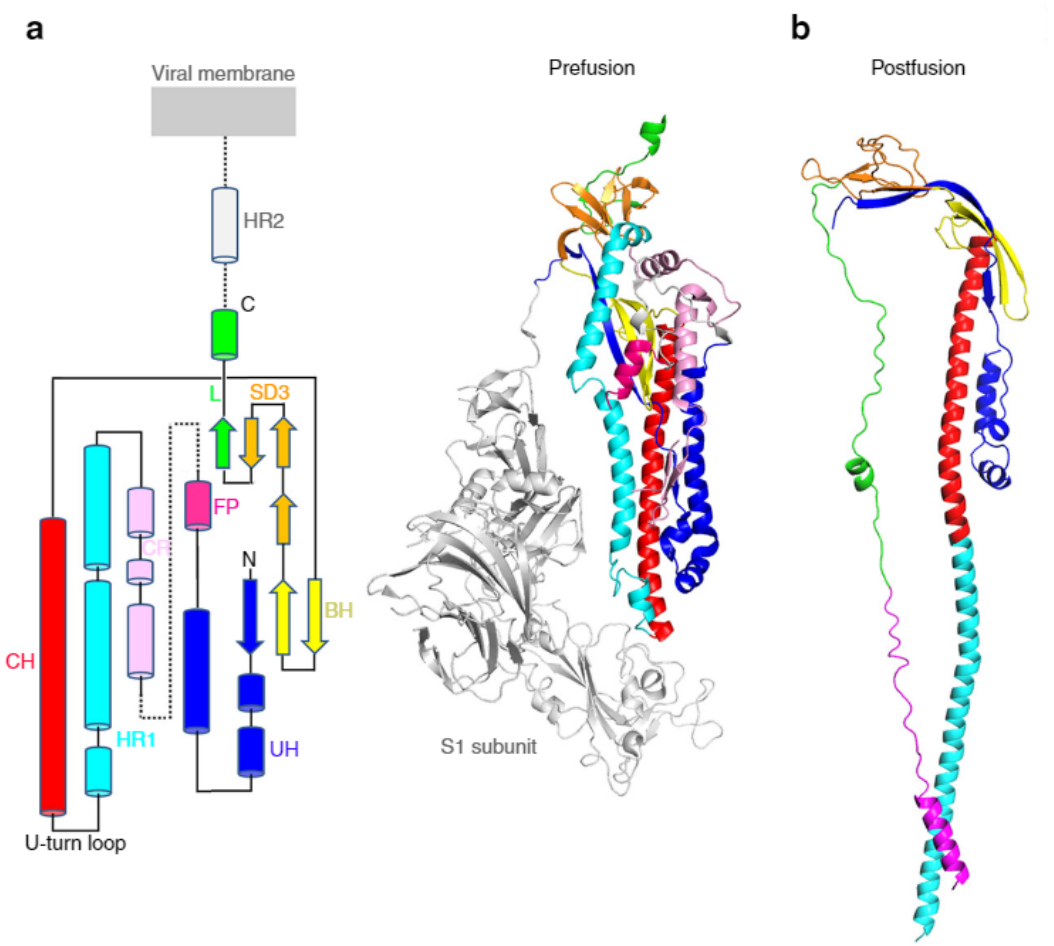


Figure 3 | The pre- and post-fusion conformations of the S protein. **A)** Schematic representation of SARS-CoV S2 subunits in the pre-fusion state with the C-terminal loop (L), the upstream helix (UH), β-hairpin (BH) motif, and subdomain 3 (SD3) adjacent to the HR1. **B)** Cartoon representation of SARS-CoV S2 subunits in both the pre and post-fusion state and coloured according to the schema (adapted from (19)).

For effective vaccine development, the SARS-CoV-2 evolution was studied to evaluate the possible effects of mutations on transmissibility, severity, and viral evasion of the immune mechanism (29–32). Strains that carried important mutations or deletions, especially on the RBD, were categorized as variants of concern (VOC) (29). Concerns among the scientific community have risen due to their potential to elude the immune system and overcome vaccine protection (33–35) despite showing an overall structural similarity between variants, which diverged only in terms of flexibility (36). A list of the mutations is reported in Table 3.

Table 3 – Important mutations found in VOCs (37).

		B.1.1.7 (Alpha)	B.1.351 (Beta)	P.1 (Gamma)	B.1.617.2 (Delta)	B.1.1.529 (Omicron)
Mutations on structural proteins	Spike	H69del, V70del, Y144del, N501Y, A570D, D614G, P681H, T716I, S982A, D1118H	D80A, D215G, L241del, L242del, A243del, K417N, E484K, N501Y, D614G, A701V	L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, T1027I, V1176F,	T19R, E156del, F157del, R158G, L452R, T478K, D614G, P681R, D950N	A67V, H69del, V70del, T95I, G142del, V143del, Y144del, Y145D, N211del, L212I, G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H, T547K, D614G, H655Y, N679K, P681H, N764K, D796Y, N856K, Q954H, N969K, L981F
	Nucleocapsid	D3L, R203K, G204R, S235F	T205I	P80R, R203K, G204R	D63G, R203M, D377Y	P13L, E31del, R32del, S33del, R203K, G204R
	Envelope		P71L			T9I
	Membrane				I82T	D3G, Q19E, A63T
Mutations on non-structural proteins	ORF1a	T1001I, A1708D, I2230T, S3675del, G3676del, F3677del	T265I, K1655N, K3353R, S3675del, G3676del, F3677del	S1188L, K1795Q, S3675del, G3676del, F3677del		K856R, S2083del, L2084I, A2710T, T3255I, P3395H, L3674del, S3675del, G3676del, I3758V
	ORF1b	P314L	P314L	P314L, E1264D	P314L, G662S, P1000L	P314L, I1566V
	ORF3a		Q57H	S253P	S26L	
	ORF7a				V82A, T120I	
	ORF8	Q27*, R52I, Y73C		E92K	D119del, F120del	
	ORF9b				T60A	P10S, E27del, N28del, A29del

*Stop codon.

At the time of this writing (November 2023), the new BA.5 Omicron VOC carrying N440K, G446S, L452R, S477N, 118 T478K, E484A, F486V, R493Q, G496S, Q498R, N501Y, and Y505H mutations became predominant over the Omicron BA.2 variant (38,39). Omicron VOC exhibits a faster and different entry mechanism (40–42), an enhanced ability to evade the immune system (32,43,44), and an increased affinity for ACE2 (45–47). However, BA.5's higher infectivity is mitigated by its milder pathogenic impact (48), similar to the new BA.2.75, XBB.1.5-like with the last update reported the 20 October 2023 by the European Centre for Disease Prevention and Control (49).

To prevent viral adhesion, a therapeutic approach might aim to intercept or prevent the S protein:ACE2 interaction required to initiate the infection mechanism (50). A pre-adhesion intervention might contain and reduce the symptoms, decreasing hospitalization rates while improving patients' prognosis. With its primary role in infection (51), the S protein has aroused the interest of medical and pharmaceutical research for antibodies' or disruptors' development as informed by the neutralizing antibodies response induced by the S protein (52).

Intriguingly, heparinoids such as heparin (HP) and heparan sulfate (HS), two relevant regulators of the coagulation cascade, seem to play a substantial role in the SARS-CoV-2 infection mechanism (53–55). The role of heparinoids is still unclear and, as a part of this project, I investigated the molecular interactions between EcHP, EcHS and the S protein.

Glycosaminoglycans (GAGs), such as extracellular heparin (ecHP) and extracellular heparan sulfate (ecHS), play a crucial role in regulating the immune response through cell adhesion, tuning cytokine, and chemokine function, and mediating inflammatory reactions (56,57), through HS-binding motifs (HSBM). EcHP, a natural glycosaminoglycan produced by basophils and mast cells (58), is constituted by highly sulfated repeating units of 1-4 pyranosyl uronic acid and 2-amino-2-deoxy glucopyranose (glucosamine) and it is known for its major role as an anticoagulant when formulated in low molecular weight (LMWH). EcHP has been proposed to play a role in promoting SARS-Cov-2 infection (51,59), probably by inducing conformational changes upon binding (55) through interactions on the S1/S2 cleavage site (60). The intermediary role of EcHS between the furin cleavage site and GAG has been also suggested by Schuurs and co-workers (61), suggesting the involvement of GAGs in favouring the membrane fusion mechanism. Clausen and collaborators (53) supported the allosteric role of EcHP to facilitate the interactions between S protein and ACE2 and provided a preliminary model of the RBD regions possibly implied in the recognition of both EcHS and EcHP. Other studies suggested the intriguing hypothesis that EcHP and low molecular weight heparins (LMWH) may function as antagonists (62,63) of the ACE2 binding, by competing for the EcHS binding site on the S protein (64). However, the role of heparinoids is still unclear and, as a part of this project, I investigated the molecular interactions between EcHP, EcHS and the S protein.

1.1.2 SARS-CoV-2 Main Protease

To overcome the mutability of the S protein, diverse strategies need to be explored, evaluating different targets of SARS-CoV-2 (65). From this perspective, targeting the NSPs (66,67) or interfering with the host's proteins that participate in the viral replication (68) can be a considerable choice. The interest in therapeutic research against coronaviruses in general arose during the SARS-CoV outbreak in July 2003. The main protease (Mpro, NSP5) and the papain-like protease (PLpro, NSP3) provided alternative targets for antiviral treatments (69,70). However, the deubiquitinase activity of PLpro, initiated by the molecular recognition between PLpro and the C-terminal sequence of ubiquitin, made the discovery of selective viral PLpro challenging due to the possibility of interfering with the host cell's deubiquitinase (71). On the other hand, the Mpro mechanism selectivity cleaves polypeptides after an LQ|SAG sequence (72). By exploiting the Mpro unique viral cleavage

system, it is, therefore, possible to extrapolate the catalytic site fingerprint to design molecular inhibitors, with, to the best of our knowledge, no side effects, due to the lack of specificity for human proteases. The SARS-CoV-2 Mpro is a 306 residue autolytic enzyme cleaving over 11 sites on the overlapping polyproteins pp1a and pp1ab to functional proteins. Its functional unit consists of two homodimers arranged almost perpendicular to each other (72) and each monomer comprises the catalytic dyad H41-C145. SARS-CoV-2 Mpro shares a 96% identity with SARS-CoV Mpro (72) and is divided into three domains (D1-3) (Figure 4).

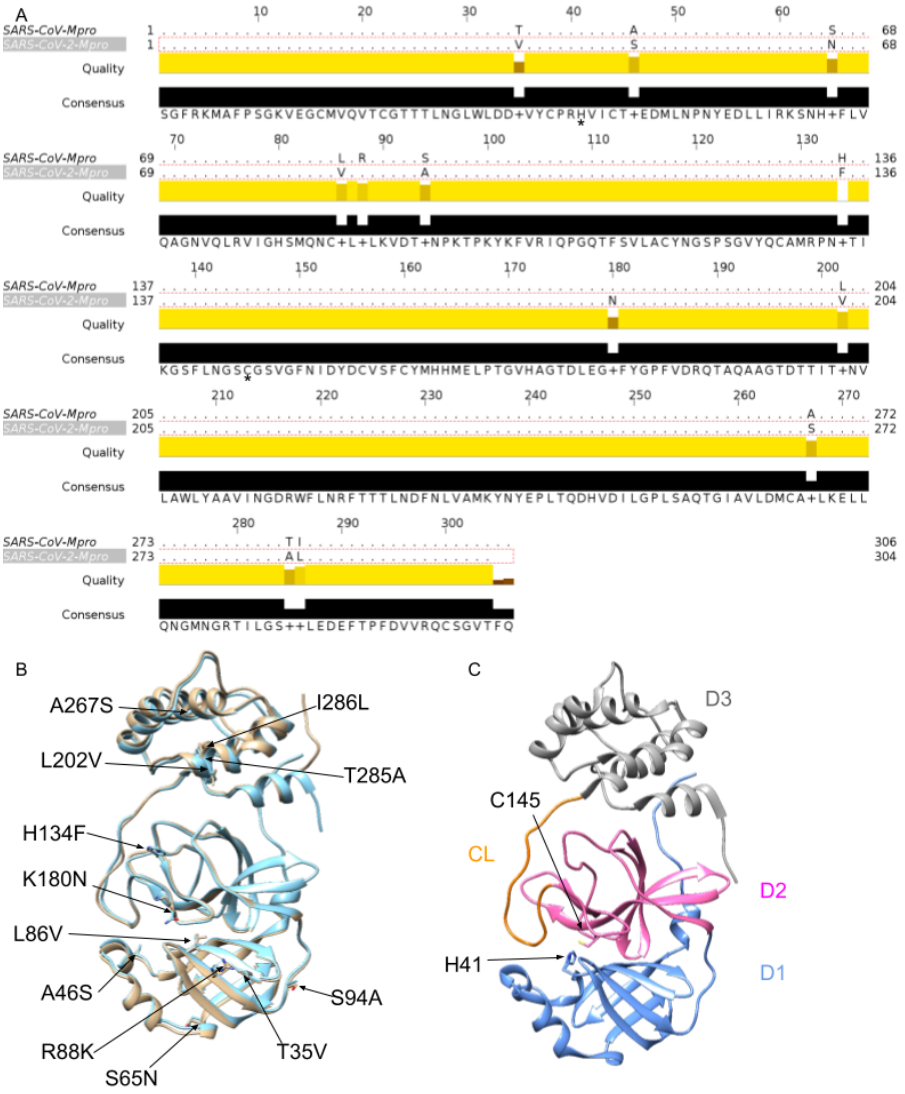


Figure 4 | The Mpro sequence and domains. A) Sequence comparison between SARS-CoV and SARS-CoV-2 Mpro. The catalytic dyad is marked with an asterisk. **B)** SARS-CoV and SARS-CoV-2 Mpro superposition with the mutations highlighted. **C)** Mpro domains: Domain 1 (D1), Domain 2 (D2), Connecting Loop (CL), and Domain 3 (D3). The catalytic dyad is situated between D1 and D2.

Domain 1 (D1, residues 1-99) and domain 2 (D2, residues 100-183) are arranged in an antiparallel beta-barrel conformation, while domain 3 (D3, residues 201-306) is placed in an antiparallel globular cluster, connected by a long connecting loop (CL, Figure 4C) (73). D3 contains five alpha-helices directly connected with the CL. The catalytic dyad, formed by H41 and C145, is positioned toward the solvent-exposed opening of Mpro and is characterized by an endopeptidase activity (73,74). Funk *et al (75)*, reported that although cysteine-based catalytic residues are prone to oxidative damage (76), the abundance of regulatory K-C redox switches in Mpro suggests a redox regulation of its activity exploiting the allosteric switch composed of sulfur-oxygen-nitrogen-oxygen-sulfur (SONOS) bridges formed between cysteines C22, C44 and K61 within D1 (75). Funk *et al (75)*, exploited the catalytic C145 oxidation as a drug design approach, additionally leading to the formation of oxidation products such as the mono-oxidized sulfenic acid and sulfinic acid (77) through maleimidoacetic acid N-hydroxysuccinimide ester (MAH) containing a nitrile warhead. The irreversible modification of C145 led to the disruption of the enzyme functionality, hampering viral replication. Following this direction, Pfizer began the production of PF-07304814, an Mpro inhibitor during the first SARS outbreak in 2003. However, this Mpro inhibitor could only be administered intravenously. The new SARS-CoV-2-Mpro inhibitor PF-07321332, instead, was developed from scratch during the SARS-CoV-2 pandemic and is now under the commercial name Paxlovid https://cen.acs.org/acs-news/acs-meeting-news/Pfizer-unveils-oral-SARS-CoV/99/i13.
Paxlovid is a reversible covalent inhibitor that targets Mpro catalytic cysteine C145. Recently, the binding pathway of Paxlovid has been proposed, using the supervised molecular dynamics (SuMD) protocol, indicating G143, H163, H164, E166, and Q192 electrostatic contribution for Paxlovid nitrile moiety to covalently bind to C145 (78).

1.1.3 Solved SARS-CoV-2 Mpro and Spike Protein Structures

Since the first cryo-EM structures of the S protein became available to the scientific community, it has been possible to investigate the conformational changes and the dynamic processes involving the S protein through MD simulations. One of the limits of the S protein structure experimentally determined is the scarcity of structural information about the post-translational glycosylation, due to the high dependency on the organism used for protein expression. To address this aspect, Woo and his group proposed a set of complete and fully glycosylated S protein models (23), corroborating Wrapp et al.'s structure of the spike protein glycosylated in 44 out of 66 possible sites in the ectodomain region (13) including also Watanabe et al. glycans' specifications (79). Several S protein structures have

been determined through cryo-EM and X-ray crystallography (Figure 5) (80). These include the inner S1 and the external S2 domains and indicate two different states in the RBD domain, named “up” and “down” (13,81–83), the former determining an active state (84) favourable to ACE2 binding.

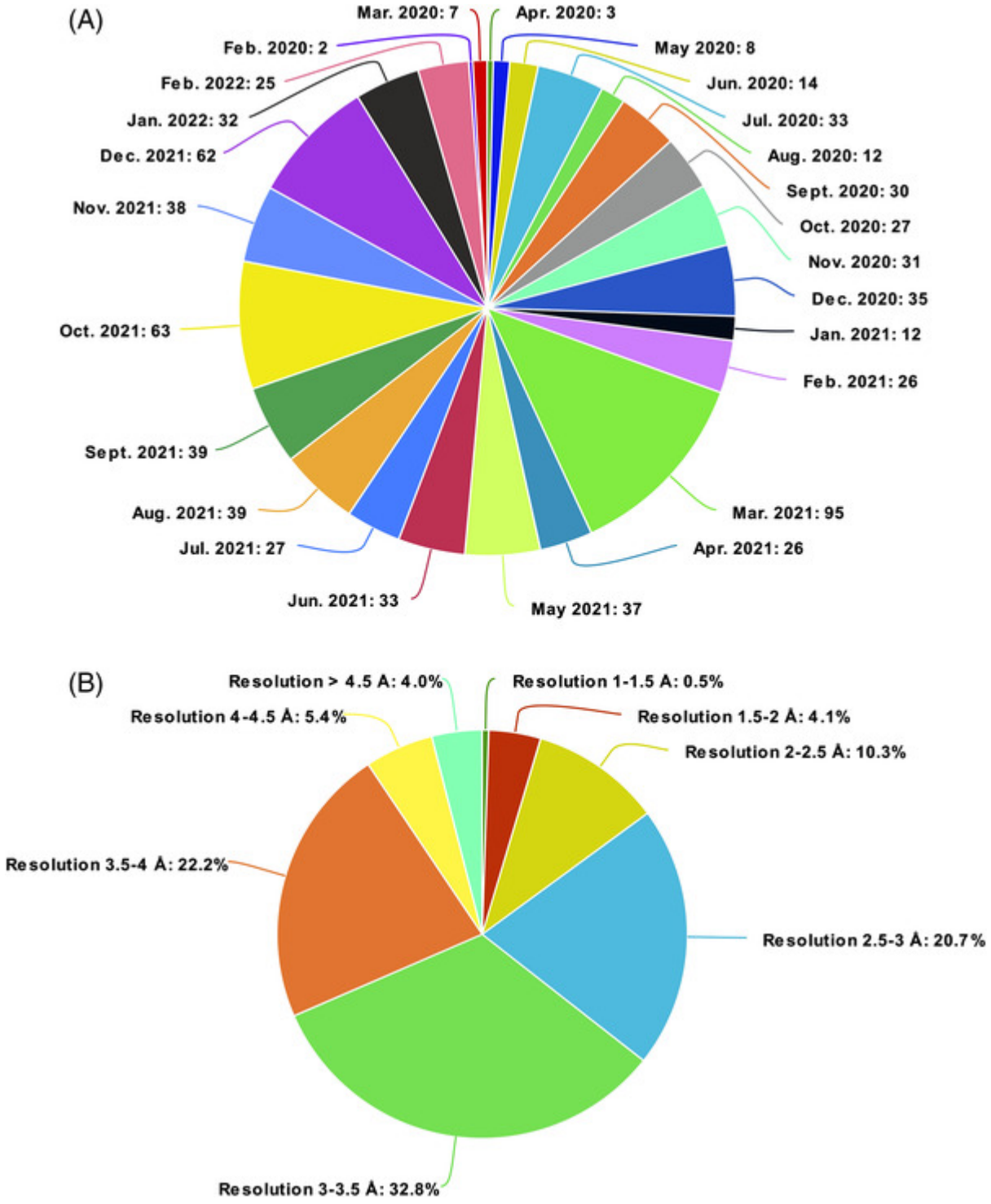


Figure 5 | S protein deposited structures in the protein data bank. A) SARS-CoV-2 protein structures released and ordered according to month and year of publication from the Protein Data Bank. **B)** S protein structures only, ordered according to resolution (85).

The structural description of the S protein, achieved through both crystallographic techniques (13) and *ab initio* modelling (86), has facilitated a more comprehensive investigation into the potential infection mechanisms and the structural modifications occurring in the S protein upon binding with ACE2 or in its transition from to the active state (87,88). Crystallographic data on the S protein and the NSPs also accelerated our understanding of the membrane fusion mechanism, increasing the set of therapeutic targets available (89,90), and broadening our options for the disruption of the infectious mechanism. Furthermore, the structural description of the S protein has proven to be pivotal in the development of antibodies (91), peptides (92), and vaccines (93–95), designed to target critical areas for molecular recognition with ACE2, such as the receptor-binding domain (RBD) or a set of epitopes (96–98) located in non-glycosylated regions of the S protein.

1.2 HIV Virus Overview

Human Immunodeficiency Viruses (HIV) are grouped into two morphologically similar types, HIV-type 1 (HIV-1) and HIV-type 2 (HIV-2) with a distinguished antigenic profile. HIV-1 is a genetically related member of the *Lentivirus* genus of the *Retroviridae* family while HIV-2 was closely related to a simian virus (SIV) that caused immunodeficiency in captive macaques (99). HIV-1 and HIV-2 are a result of the cross-species zoonotic transfer of two simian viruses infecting chimpanzees to humans (100). Lentivirus infection slowly and silently evolves into a chronic development of acquired immunodeficiency syndrome (AIDS), preceded by a long asymptomatic quiescent phase characterized by persistent viral replication and involvement of the central nervous system (101). HIV-1 is generally spread worldwide, while HIV-2 type is restricted to some Sud-Saharian, Western and Central Africa (99,102). Both HIV-1 and HIV-2 potentially cause AIDS with a lesser incidence of HIV-2 on the central nervous system (CNS) and generally taking longer to progress to the immunodeficiency pathological course. Both HIV strains share the same pathophysiological mechanism as well as the same replication cycle (Figure 6)

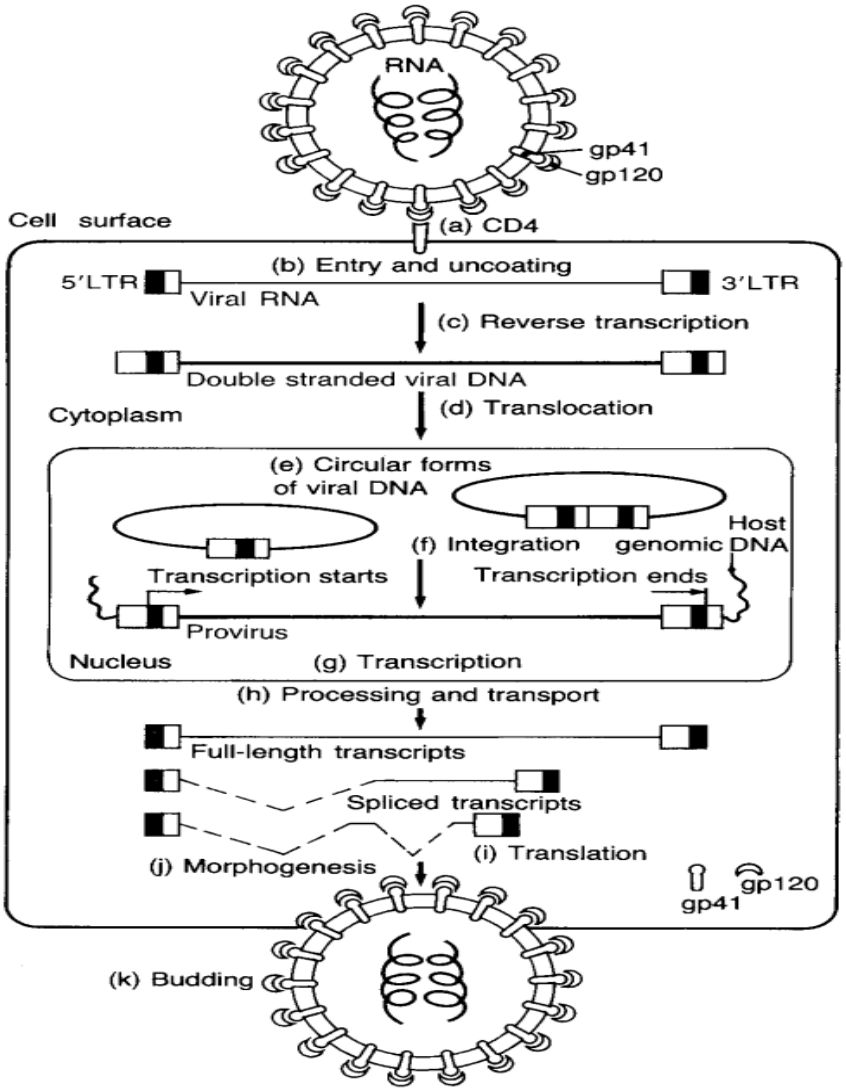


Figure 6 | HIV replication cycle. a) HIV attaches to CD4+ T-lymphocytes or macrophage through the heterotrimer gp120-gp41 surface protein **b)** HIV penetrates inside the host cell **c)** where it begins the reverse transcription of its viral-RNA into a double-stranded DNA through the RNA-associated reverse transcriptase. The viral genome **d)** is then translocated to the nucleus **e)** and converted to the supercoiled DNA sequence. Following the integration **f)** the DNA is truncated at the 5' and 3' and integrated by the integrase enzyme into the host cell DNA **g)** where the host's DNA polymerase II transcribes the newly integrated viral DNA **h)** which are then translocated and spliced back into the cytoplasm **i)** where *gag* and *pol* genes are fully translated as well as the other accessory proteins (gp41 and gp120 are shown only for clarity). **j)** The morphogenesis occurs with the formation of a ribonucleoprotein core dimer **k)** which will assemble with the rest of the gag and pol proteins ready to be shed by membrane budding which incorporates also some of the host cell's surface proteins in the process (adapted from (103)).

Peterlin reports that HIV's genome consists of a proviral hybrid DNA/RNA which is later translated into a double-strand DNA (103). HIV genome sequencing established that the proviral DNA genome is a 9.7 kilobase pairs sequence with retroviral features such as the presence of structural genes capped by long terminal repeats (LTR) (Figure 7). HIV genome encompasses structural proteins such as group-specific antigen (GAG), polymerase (POL), and envelope (ENV) and, in addition, four unique nonstructural genes, several of which appear to be essential in regulating virus replication, namely viral infectivity factor (Vif), viral protein R (Vpr), viral protein U (Vpu), and negative factor (Nef) with HIV-2 lacking for the most part Vpu (104). Table 4 briefly summarises the different proteins encoded and their role.

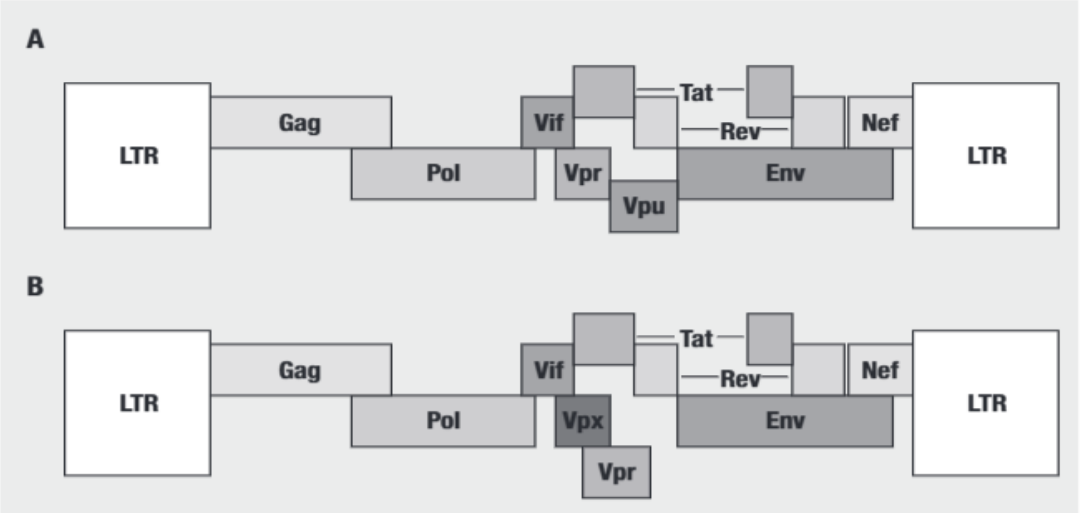


Figure 7 | HIV-1 and HIV-2 genome. A) HIV-1 genome organization with structural and accessory proteins. **B)** The HIV-2 genome differs from the presence of the viral protein X (Vpx) instead of Vpr (adapted from (101)).

HIV-1 and HIV-2 viruses display the presence of the same structural and non-structural proteins a common constant for all retroviruses. Additionally, two more genes, *tat* and *rev*, encode for the trans-activator of transcription (TAT) and Regulator of expression of virion proteins (REV) (105). TAT is expressed very early after infection and promotes the expression of HIV genes while REV ensures the export from the nucleus to the cytoplasm of the correctly processed messenger and genomic RNA.

Table 4 - HIV Proteins and their function

Name	Role
Matrix (p17)	Plasma membrane targeting of Gag for virion assembly, Env incorporation; and post-entry events;
Capsid (p24)	Virion core structure and assembly
Nucleocapsid (p7)	Virion packaging of genome RNA; RNA chaperone; virion assembly
p6	Promotes virion budding
Protease (PR)	Proteolytic processing of Gag and Gag-Pol polyproteins
Reverse Transcriptase (RT)	cDNA synthesis; RNaseH, domain degrades RNA
Integrase (IN)	Covalent insertion of virus cDNA into cellular DNA
Surface Glycoprotein (p120)	Binds cell-surface receptors and mediates virus attachment and entry
Transmembrane Glycoprotein (p41)	Contains fusion peptide, mediates membrane fusion and virus entry
Viral Infectivity Factor (Vif)	Suppresses APOBEC3G/APOBEC3F host factors that inhibit infection
Viral protein R/X (Vpr/Vpx)	Moderate enhancer of post-entry infectivity; G2/M cell cycle arrest
Trans-Activator of transcription (TAT)	Potent activator of viral transcription elongation
Regulator of expression of Virion protein (REV)	Induces nuclear export of intron-containing viral RNAs
Viral protein U (Vpu)	CD4/MHC downregulation and induces virion release from the host cell surface
Negative Factor (Nef)	CD4/MHC downregulation; T-cell activation; moderate enhancer of viral infectivity; blocks apoptosis; pathogenicity determinant.

HIV-1 and HIV-2 structure shows a strong morphological similarity characterised by the gag-encoded structural proteins (p24, p7, p6), matrix protein (p17), and the viral envelope glycoproteins gp120 and gp41, which recognize cell surface receptors (105,106). Additionally, the *pol* gene encodes also for the reverse transcriptase enzymes, necessary for converting the viral RNA into DNA. Integrase enzyme is coded from the *pol* gene, and it plays a crucial role in incorporating the viral DNA into the host chromosomal DNA by producing "sticky ends" on the vDNA to liberate 3′-hydroxyl groups attached to the invariant deoxycytidine–deoxyadenosine (dCdA) dinucleotides (107). Additionally, a viral protease cleaves large gag and pol protein precursors into their components until the active protease dimer is formed (108).

HIV viral particles have a diameter of ~100 nm and are surrounded by a lipoprotein-rich membrane (Figure 8). Each viral particle consists of two noncovalently linked heterotrimeric subunits - gp120 and gp41 - that form a 6-subunits structure bound to the membrane. Due to the noncovalent gp120 and gp41 binding (109), gp120 may be shed spontaneously within the local environment and be detected in the serum, as well as within the lymphatic tissue of HIV-infected patients.

When gp120 binds to the CD4 protein gp120 exposes a region capable of binding to chemokine receptors on the target's cell membrane. The natural ligands of these receptors are a set of chemokines that attract other immune system cells during inflammation. Chemokine receptors (CKR) are part of the superfamily of G protein-coupled receptors (GPCRs) that activate intracellular G protein- and β-arrestin-mediated pathways upon chemokine binding (110). Chemokines are described by the cysteine patent in the N-terminus and are classified by the number of residues between two cysteines and the splicing variant at the N-terminus.

CKR splicing variants CXCR4 and CCR5 are the most common chemokine receptors among HIV-1 with CXCR4 being expressed on many cells, including T lymphocytes, while CCR5 is more prevalent on monocytes, macrophages, dendritic cells and activated T lymphocytes (101). The presence of these specific receptors on target cells determines the HIV-specific tropism for either CXCR4 expressing cells, resulting in a T-lymphocyte-tropic (T-tropic) or a CCR5 receptor preference, resulting in a macrophage-tropic (M-tropic) tropism. Additionally, some strains can bind to both CCR5 and CXCR4 and are called dual tropic or X4R5 strains (101,111).

During the budding process where virion particles exit the infected cell, the virus can incorporate the host's proteins into its membrane, especially human leukocyte antigen (HLA) class I and II protein (103) or adhesion proteins such as Intercellular Adhesion Molecule 1 ICAM-1 and that may facilitate adhesion to other target cells (112).

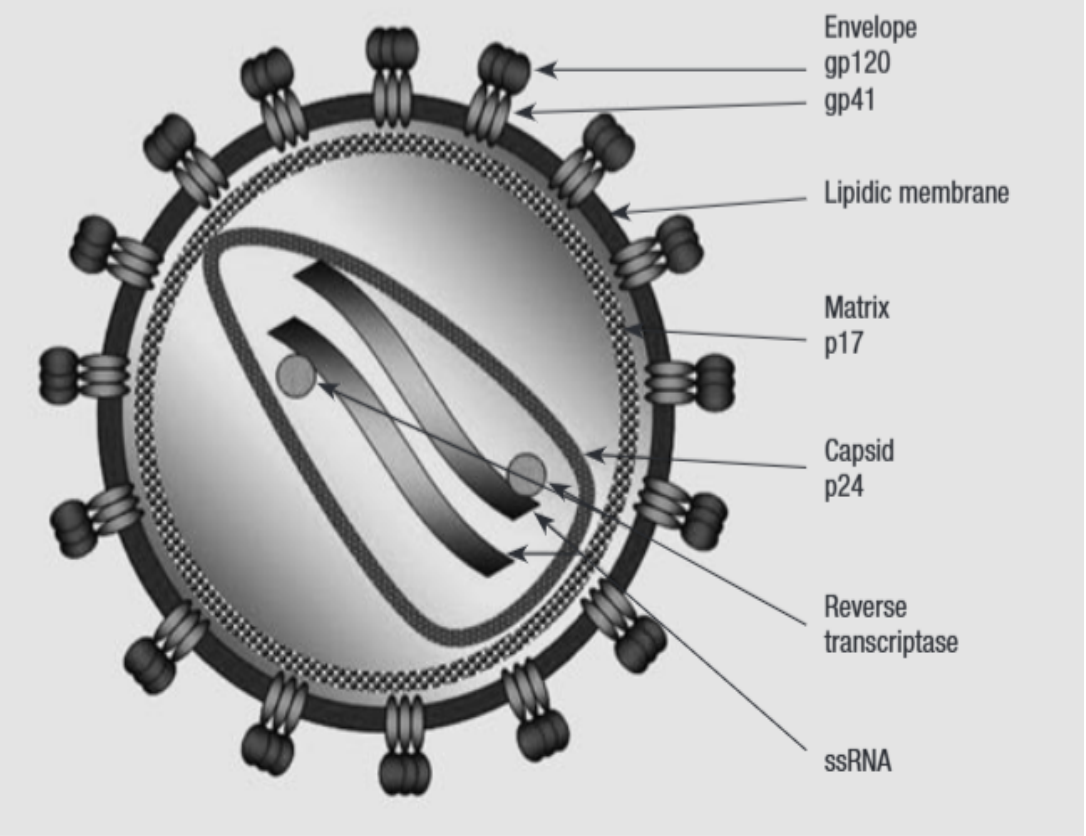


Figure 8 | HIV morphology. HIV-1 and HIV-2 morphology. HIV virion size is 100 nm in diameter, coated by a lipoproteic membrane. The membrane includes glycoprotein heterodimer complexes composed on the external surface of gp120 and gp41 which are not covalently bound. The gp120 unit may be shed and used as a diagnostic marker in the serum, as well as within the lymphatic tissue of HIV-infected patients (adapted from (101)).

The matrix protein (p17) is located inside the viral lipoprotein membrane. The virus membrane and the matrix protein include the capsid composed of polymers of the core antigen (p24). The capsid contains two copies of HIV RNA combined with a nucleoprotein and the enzymes reverse transcriptase, integrase and protease (106). Due to the high mutagenicity of HIV, different therapeutic targets might be approached for drug discovery. Nef is a nonenzymatic protein with a crucial role in viral replication and immune escape of HIV-infected which makes it an excellent target for drug design. Among its multiple functions, it has a role in kinase signalling pathways and endosomal trafficking, while being responsible for the downregulation of the CD4 receptor on the cell surface (114).

During this project, our investigation focussed on a *de-novo* discovery of a potential Nef dimerization mechanism disruptor (see Chapter 7). Its importance and mechanistic behaviour will be treated during this project to highlight its promising therapeutic potential in hampering viral replication and AIDS course progression (Section 1.2.1).

1.2.1 HIV Nef Regulatory Protein

Initially, Nef was identified as an open reading frame sequence (B, E', F orf) encoding for a 25-27 kDa protein whose amino-terminal was bound to myristyl alcohol (103). Nef amino-terminal region resembles the proto-oncogene tyrosine-protein kinase (src) protein family and, similarly to src protein, can bind and hydrolyse guanosine triphosphate (GTP), while acting as a "negative factor" for HIV-1 replication, which led to its initial naming (115). Nef is an important player in viral pathogenesis by promoting viral replication and enabling immune escape in infected hostsStaudt et al., 2020. Early pieces of evidence showed that expression of Nef in CD4+ T-cells and macrophages causes AIDS-like disease in transgenic mice and that patients infected with nef defective HIV-1 fail to progress to AIDS suggesting a direct role for Nef in HIV-1 pathogenesis (116,117).

Activated Nef consists of a globular domain dimer and disordered N and C-terminal structures with an SH3 kinase binding domain (Figure 9). Myristilation of the ammino-terminal group anchors Nef's flexible structure in the lipid membrane. Both HIV and SIV Nef globular domains are relatively well conserved between the two viruses with more variable terminals (118). Nef differences between HIV and SIV viruses translate into differential T-cell receptor (TCR-CD3) down-regulation by Nef. Transgenic mouse models highlight progressive AIDS due to the progressive loss of CD4+ T-cells, T-cell activation, lymphadenopathy, and immunodeficiency where Nef also showed significant importance for viral replication (119).

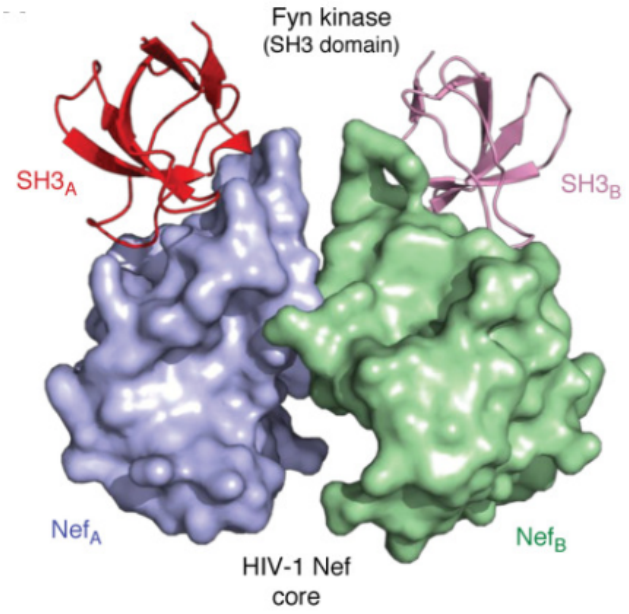


Figure 9 | X-ray crystal structures of HIV-1 Nef in complex with the Fyn SH3 domain. Nef monomers are coloured in blue (NefA) and green (NefB), respectively, with the SH3 domains in red (SH3A) and pink (SH3B) (adapted from (114)).

Nef contributes to HIV pathology by modulating protein trafficking, interfering with signalling pathways and apoptosis and stimulating viral replication/virulence. More than 70 Nef-interacting proteins in the human proteome have been identified and they include clathrin-coated vesicle machinery, coatomer, an endosomal sorting complex required for transport (ESCRT) machinery, an endosomal sorting complex required for transport (ESCRT) machinery and multiple clusters of differentiation (CD) receptors (120).

Nef possesses multiple protein-binding domains on its surface such as the Src homology (SH3) binding domain, involved in interactions between Nef and hematopoietic cell kinase (Hck), lymphocyte-specific protein tyrosine kinase (Lck) and p21-activated kinase (PAK2) and multiple tyrosine-based and dileucine motifs that mediate the interactions with the adaptor proteins (AP) from vesicular coats and induce the down-regulation of CD4, MHC-I, CD8, CD28 (118). Among its functions, Nef induces two T-lymphocyte-killing mechanisms: promoting the programmed cell death of HIV-specific cytotoxic T-lymphocytes (CTL) by upregulating Fas ligand (FAS) expression on the cell surface and down-regulating the expression of the major histocompatibility complex class I (MHC-I), specifically HLA-A and –B antigens used for CTL recognition (120).

2. Molecular Binding's Driving Forces

Biochemical phenomena are a consequence of the collisions and energy exchanges between the particles in a combination of attractive and repulsive forces between the atoms.

Biological systems often consist of a thermodynamic solute–solvent ensemble where the solute (e.g. protein with or without a membrane or a multi-component assembly) and the solvent (i.e. water and ions) are in contact with each other (129). The interactions and heat exchange among these substances are influenced by the intrinsic properties of each atom, their structural arrangement, and their distance. Conformational changes depend on the way heat and energy are transferred according to the laws of thermodynamics (130).

The attraction or repulsion forces between molecules or atoms depend on their positions and the results of the summative effect of both their non-bonded long and short-range forces (131). Intramolecular stability, on the other hand, relies on internal short-range forces (covalent bonds, ionic bridges, London dispersion forces, dipole-dipole forces). Forces and the energy distribution between atoms and molecules determine the overall capacity of a system to produce work in the form of bond formation or breaking, heat exchange and conformational changes (132,133). The energy exchange that follows such dynamics can be quantified as free energy variation.

The Gibbs free energy is a thermodynamic potential that measures the capacity of a thermodynamic system to do maximum or reversible work at a constant temperature and pressure (isothermal, isobaric conditions). The chemical potential energy was elegantly formulated by J.W Gibbs (134) and it was originally formulated as the capacity of a system to produce work in the form of energy or heat exchange (non-pressure work) when two elements come into contact is measured by the variation of its total energy once the two elements get into contact.

A chemical system is defined by the number of moles (N), volume (V), and temperature (T). Chemical potential also provides a measure of the tendency of a substance to participate in chemical reactions as described by Gibbs:

$$U = Ur + Um + Uc = TS - pV + \mu N \qquad \textbf{(1)}$$

Equation 1 is the complete form which connects the seven basic thermodynamic variables, Ur (thermal potential energy), Um (mechanical potential energy), Uc (chemical potential energy), T (temperature), S (entropy), p (pressure), V (Volume), μ (chemical potential), and N (number of moles) (135) and can be rewritten as

$$U - TS + pV = \mu N$$

$$= \Delta (H - TS) = \Delta H - \Delta (TS) = \Delta H - (T\Delta S + S\Delta T) \quad \textbf{(2)}$$

The free energy variation (∆G) at any moment in time during molecular association is therefore given by the formula:

$$\Delta G = \Delta H - T\Delta S \quad \textbf{(3)}$$

where ∆H and ∆S refer to enthalpy and entropy changes of the system upon ligand binding and T represents the absolute temperature in Kelvin (Figure 11).

Enthalpy is a state function that represents the total heat energies of a thermodynamic system (136). ∆H is negative during exothermic processes, for example, during combustion, and positive during endothermic reactions such as photosynthesis. Entropy is also a state function of a system which is associated with the state of randomness or chaos of a system, and it measures how evenly the heat energy is distributed over the overall thermodynamic system (137). The second law of thermodynamics dictates that thermal energy moves from high-temperature regions to those characterised by lower temperatures diminishing the state of organisation inherent to the initial system. Consequently, entropy could be used to quantify the level of “disorder” or “chaos” stochasticity represented by atoms and molecules at a given thermal state.

Receptor-ligand binding results in a reduction of the rotational, translational and torsional degrees of freedom (Figure 11) defined as a loss in entropy which can be compensated by the enthalpy increase provided by the ligand-protein interaction (138).

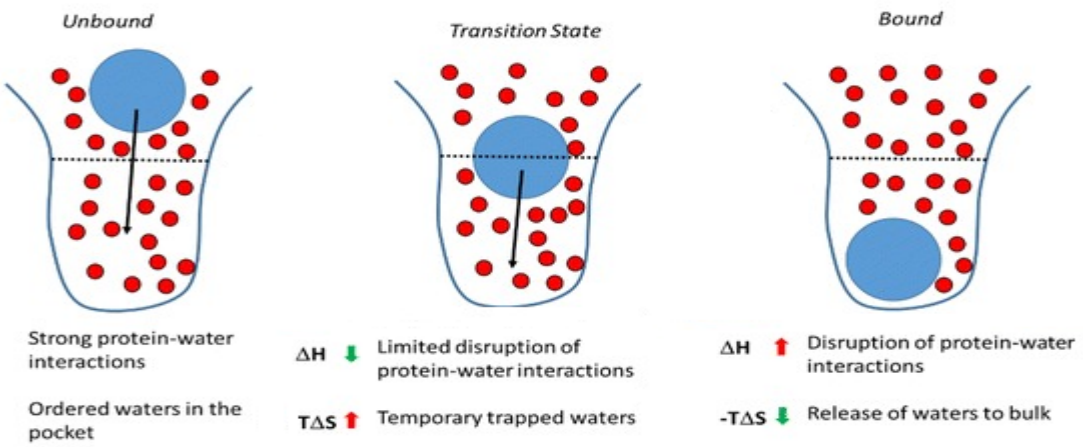


Figure 11 | Entropy and enthalpy changes upon binding. Schematic representation of the enthalpy and entropy changes as a result of both solvent displacement and ligand stabilization against the binding surface of a receptor/protein. The transition state is associated with a temporary increase in entropy and a proportional decrease in enthalpy due to the formation and disruption of non-covalent bonds between the protein-ligand interface and the protein-water interactions. Upon binding, the entropy loss is paired with an increase in enthalpy increase due to the disruption of protein-water interactions (adapted from (139)).

Spontaneous processes are characterized by a negative change in Gibbs free energy (ΔG) when the energy of the system is negative at the equilibrium state with constant pressure and temperature. However, an equilibrium system is not static even at 0 K as the vibrational frequency of atoms never reaches zero. More realistically, the forces acting on a system at equilibrium are determined by continuous local thermal fluctuations (140) that could determine changes in the local entropy (141) affecting, for example, the binding interface of a protein, therefore altering the protein-ligand complexation.

2.1 Conformational Changes and Molecular Recognition

Proteins are complex biochemical machines that regulate cell functions, structure, and cell's life cycle. The conceptualization of the quaternary structure (QS) was first discovered by Svedberg in the 1920' while investigating the molecular weight of haemoglobin by sedimentation in the ultracentrifuge (147). The formation of multimer assembly is what determines the QS of a protein (148), with a specific folding of its sequence forming its super secondary structures (SSS), correlated with biological functions (149). Since the elucidation of myoglobin structure in 1958 (142) protein structures' description was enriched by an additional layer of complexity to represent the natural molecular folding (143). The folding mechanism depends on the protein's amino acid sequence (144), the complex inter and

intra-residue interactions (145), and the contribution from the solvation energy (146). The interplay of those elements determines the possible shapes of a protein and its relative biological function.

In the last century, the importance of the QS gave birth to the intriguing “protein folding problem” (150,151) which is necessary to describe and predict the conformation (and possibly the associated biological activity) of a targeted structure. The characterisable structural features include exposed epitopes, catalytic residues’ disposition, conformational changes available, or molecular recognition binding motifs, to name just a few.

The ability of proteins to fold reversibly in a precise and efficient manner and to access thermodynamically stable states requires microseconds to seconds (151,152). This suggests that proteins naturally explore multiple energetical states before folding in favourable conformation. The combination of conformations and associated energy describes a set of finite states described by a free energy surface (FES) or potential energy surface (PES), in which every protein conformation, possibly associated with a specific biological activity, has a specific energy and probability to exist (153). The PES provides a better understanding of the kinetics and thermodynamic properties of proteins by quantifying energy barriers between stable conformations.

The statistical mechanical description of proteins defines a number of possible states with specific configurational entropy (S) (154) and free energy (E), in relation to their degrees of freedom (155). More structured proteins have lower E and S and are characterised by an efficient folding pattern that undergoes transitional metastable states (156). Generally, a protein’s efficient folding occurs by minimising the number of explorative pathways required in a step-wise folding mechanism (157). The FES represents the ensemble of states in which a protein folds and the respective potential energy (Figure 12A).

The FES of a polypeptide or protein is characterised by multiple local minima (i.e. favourable metastable conformations), hence the energy minimisation of the protein reaches only a local minimum instead of the global minimum (Figure 12B). Several methods are being deployed to investigate the “surroundings” of the saddle points of FESs which represent the many-object modelisation of the structure (158).

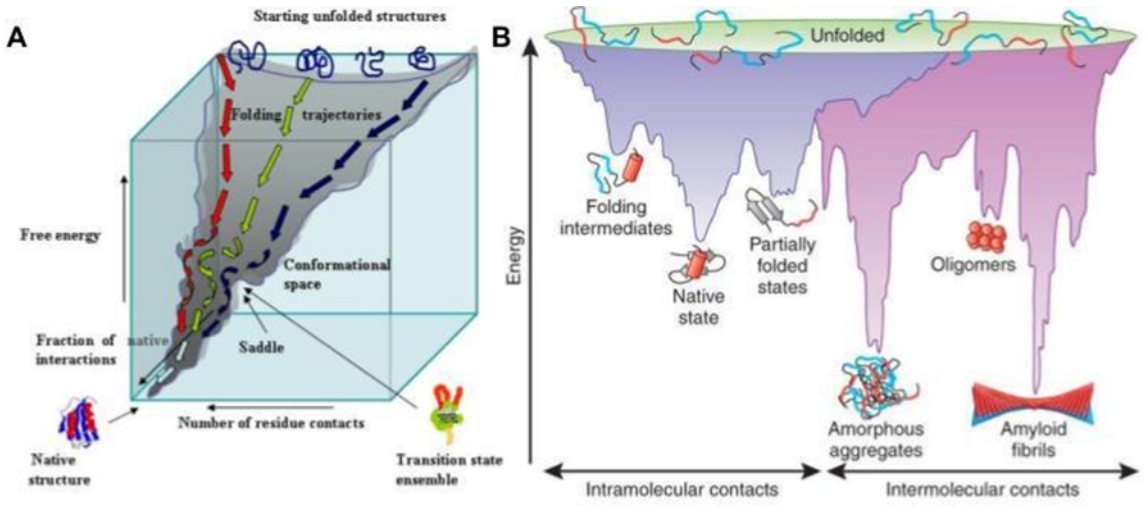


Figure 12 | Free energy surface and multiple minima. A) 3D representation of the microstate transition in the folding process of a protein to its native state at the lowest energy available. The pathway to the native state N encounters a series of saddles and peaks, indicating that the folding pathway is not a conjugated gradient descent but more likely a complex sequence of events and shifts **B)** Multiple minima in the FES are associated with folding intermediates or metastable states (partially folded or intermediates). Protein states fall into a plethora of low-energy intermediates and the investigation of their transition might unveil new saddle points, disclosing unexplored structural features. (adapted from (159,160)).

Thirty years later, Perutz's crystallographic studies of haemoglobin subunits' structures confirmed Svedberg's description of its QS and indicated QS changes upon oxygen-haemoglobin binding (161). Further experimental evidence demonstrated the correlation between structural changes and biological function (162,163), with each conformation associated with an energy state dependent on the intramolecular forces acting inside the system (160).

Conformational changes are necessary to exert a biological activity, usually exerted by exposing chemically reactive sites for modifications such as (de) phosphorylation, (de) sulfonation, or glycosylation only to name a few. Stimuli such as photochemical reactions (166), drug binding, changes in the surrounding medium (167), and receptor engagement (168) are just a few examples of cascade triggers meant to modulate the protein activity and shape. Specifically, protein-target binding may be responsible for exposing cryptic binding pockets (169) or altering the surface (or even the electrostatic landscape (170)) of proteins. The importance of QS becomes more evident considering the biochemistry of pathology. For

example, protein misfolding in cystic fibrosis (164) leads to pathogenesis and the dysregulation of physiological homeostasis in lung cells. Mutations in the Cystic fibrosis Transmembrane conductance Regulator (CFTR) gene dysregulate the activation of the Cl^- ion transport mechanism, leading to ionic imbalance and pathology (165).

Three binding mechanism models were proposed to describe the protein-ligand binding: the "lock-and-key", the "induced fit", and the "conformational selection" (171–173). In the lock-and-key model, the binding process occurs when the protein and the ligand match a precise molecular fingerprint in what was called the “one enzyme, one substrate” type of interaction. In this scenario, their binding interfaces must match to trigger the intended structural modifications, both chemically and geometrically speaking. However, the lock-and-key model cannot explain “imperfect” match bindings and does not explain conformations changes (174). The induced fit model fills those gaps assuming that the binding site is a flexible space that interacts with a ligand in an adaptive manner, triggering dynamic conformational changes in the binding site of the protein. A binding event happens via a mix of conformational selection and induced fit, giving the protein dynamic that exhibits different degrees of flexibility throughout their QS. Finally, the conformational selection model takes flexibility into account from the potential energy surface (PES) (175–177) theory of protein structure and dynamics. In PES the native state of a protein (intended as the relation between the low-state energy associated with specific conformations of a protein) is represented as a large (but finite) set of conformational states/sub-states. Ligands selectively bind to one of the most populated protein conformations. Therefore, the unbound protein (UP) has a finite probability of adopting the necessary conformation to favour the ligand-bound state (137).

2.2 The Binding and Unbinding Mechanisms

In physiological conditions of temperature and pressure, fundamental biochemical processes occur upon the binding or unbinding of an exogenous or endogenous ligand. The laws of thermodynamics regulate the driving forces and the kinetics that lead to the formation of intermolecular complexes.

In 2006, the kinetic concept of residence time, which describes the period for which the receptor is occupied by a ligand (178) was first introduced as the reciprocal of the dissociation rate constant k_{off}. The pharmacodynamic relevance of the residence time (*i.e.* expressed in seconds) provides a measure of the drug’s potency associated with its stability inside the receptor (179).

When a protein [P] and a ligand [L] reach equilibrium, the dissociation constant k_d is expressed by:

$$k_d = \frac{[P][L]}{[PL]} = \frac{koff}{kon} = k_a^{-1} \quad \textbf{(4)}$$

where [PL] indicates the protein-ligand complex concentration, and k_{on} on and k_{off} represent the kinetic rate constants for binding and reverse unbinding (or dissociation) reactions. The units of k_{on} and k_{off} are $M^{-1}s^{-1}$and s^{-1}, respectively.

At equilibrium, the binding and unbinding rates are balanced, so Equation 4 can be written as:

$$k_{on} = [P][L] = k_{off}[PL] \quad \textbf{(5)}$$

The k_{off} depends on the difference of free energy between the bound state and the transition state (TS) along the dissociation pathway and usually is related to the stability of interactions occurring in the binding pocket and entropic factors (140). A ligand with a fast-binding rate (i.e. low energy binding TS) accompanied by a slow dissociation rate (i.e. high energy unbinding TS) will have a high affinity for the target. The association process includes a diffusion phase, where the ligand's chance of a productive collision with the receptor relies on its own concentration, and an interaction phase, where recognition is based on binding mechanism and energy progress. Once molecular recognition occurs, the binding and unbinding mechanism will rely on the interatomic forces and energy changes that will determine its overall kinetics. The recognition event is a second-order reaction (expressed in $M^{-1}s^{-1}$) and is dependent on the concentration of both the ligand and the receptor (180), while the dissociation is a function of the intermolecular complex concentration and is expressed in s^{-1} (181).

The relationship between Gibbs free energy and the equilibrium constant k_e includes both the entropic and the enthalpic contribution described by the formula:

$$k_{eq} = e^{(-\frac{\Delta G}{RT})} = \Delta G° = - RT\ ln\ k_{eq} \quad \textbf{(6)}$$

as the exponential of the energy variation ΔG over RT, with R as the gas constant and T as the temperature of the system.

In Eyring's equation derivation (182), the kinetics constants are proportional to the exponential of the energies of activation of the respective transition states (ΔG‡ on/off),

through a pre-exponential factor that combines Boltzmann's constant K_b, Planck's constant h and the absolute temperature T:

$$K_{on/off} = \frac{K_b T}{h} e^{-\frac{-\Delta G \pm on/off}{RT}} \quad \textbf{(7)}$$

Adapting equation 4, the relationship between the enthalpy and entropy changes and K_d can be rewritten as

$$k_d = e^{-\frac{\Delta G_{binding}}{RT}} = k_a^{-1} \quad \textbf{(8)}$$

with a form much similar to the temperature-dependent Arrhenius equation:

$$k = Ae^{\frac{-E}{RT}} \quad \textbf{(9)}$$

where k is the rate constant at which fruitful collisions occur, A is the Arrhenius pre-exponential factor (expressed as s^{-1}) and E is the activation energy.

A broader statistical mechanic approach to Eyring's transition for more complex systems implies that the reaction's direction and kinetics are defined by different transition states and pathways each contributing to the kinetic. These concepts are more easily evaluated through Marcelin's (183) contribution to the representation of the chemical reaction direction by a 2D motion of a point in phase and in Rice's treatise. In Rice's extended implications (182), the state of a molecular system can be expressed as a set of Lagrange generalized coordinates and their derivatives with respect to time to describe the potential energy of a system. By restricting the potential energy dependence on one coordinate, the course of the reaction can therefore be regarded as the motion of a point in 2n-dimensional phase space (Figure 13).

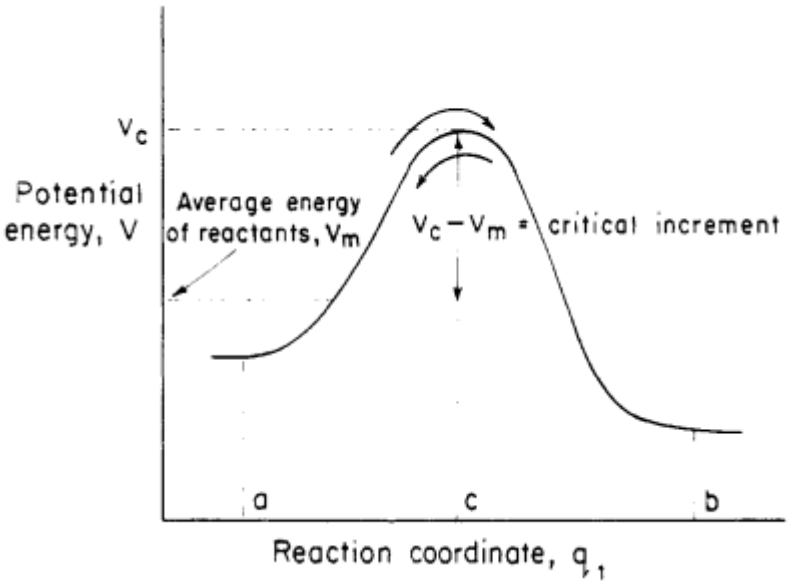


Figure 13 | 2D reaction coordinate profile. Following the simplified 2D coordinates, Rice's interpretation of Marcelin's transition pathway depends on the "critical increment" Vc - Vm which corresponds to the activation energy between the unbound [P] [L] to the bound state [PL] (Adopted from (182)).

The stochastic nature of intermolecular recognition events implies a first molecular contact that is dependent on the probability of the ligand diffusing in the medium to form productive collisions (184). Statistical mechanics can be used as an effective conceptualization to describe and simulate protein-ligand interactions. In statistical mechanics, the evolution of a system of particles can be described by the probability of a system occupying a position with momentum in time. The dynamic trajectory of the set of points in phase space that the system visits over time carries information about the state of the system. Together, positions and momenta define a point in the phase space of the system (a portion of the PES) which correlates the system's coordinates with the calculated potential energy. Such information also expresses all experimentally measurable quantities (such as temperature, pressure, diffusion coefficients, spectra, etc. (185)) as functions of the points in phase space visited. States with low energy are thermodynamically favoured and more populated in experimental conditions, representing the starting point for structure-based drug design investigations of molecular methods (see Chapter 3). Free energy and thermodynamic estimation are a common aspect tackled by the statistical mechanical interpretation of the PES, which can support drug design decision-making

2.3 Intermolecular Interactions

Intermolecular interactions are of great importance in biology. Different pH conditions can alter the ionization (Figure 14) of amino acids whose chemical structure includes basic or acidic functional groups (186). Amino acids with ionizable side chains such as aspartate, glutamate, histidine, lysine and arginine are responsible for numerous crucial pH-regulated protein functions (denaturation, structural stabilisation, catalytic site role) and intramolecular connections (187). Charged residues in the proximity of solvent-exposed areas of the protein might be subject to dielectric constant changes from those residues buried in the protein core, with values that can diverge from ≈20-80 (at the bulk solvent interface) to ≈ 1-4 (inside buried protein cavities) (188).

Atomic charges may result in stabilizing or repulsive interactions depending on the charge density of the atoms involved, their distance and the dielectric constant (189). Electrostatic changes on the binding surface of proteins appear evident in the case of RBD-ACE2 molecular recognition between SARS-CoV-2 variants (190) where mutations on

key polar residues responsible for molecular recognition are proposed to change the electrostatic interactions involved in the binding pathway (191). It is important to notice that charge density, electric moment dipole and interatomic distances should be treated as dynamic values that change with respect to time rather than static properties attached to a point.

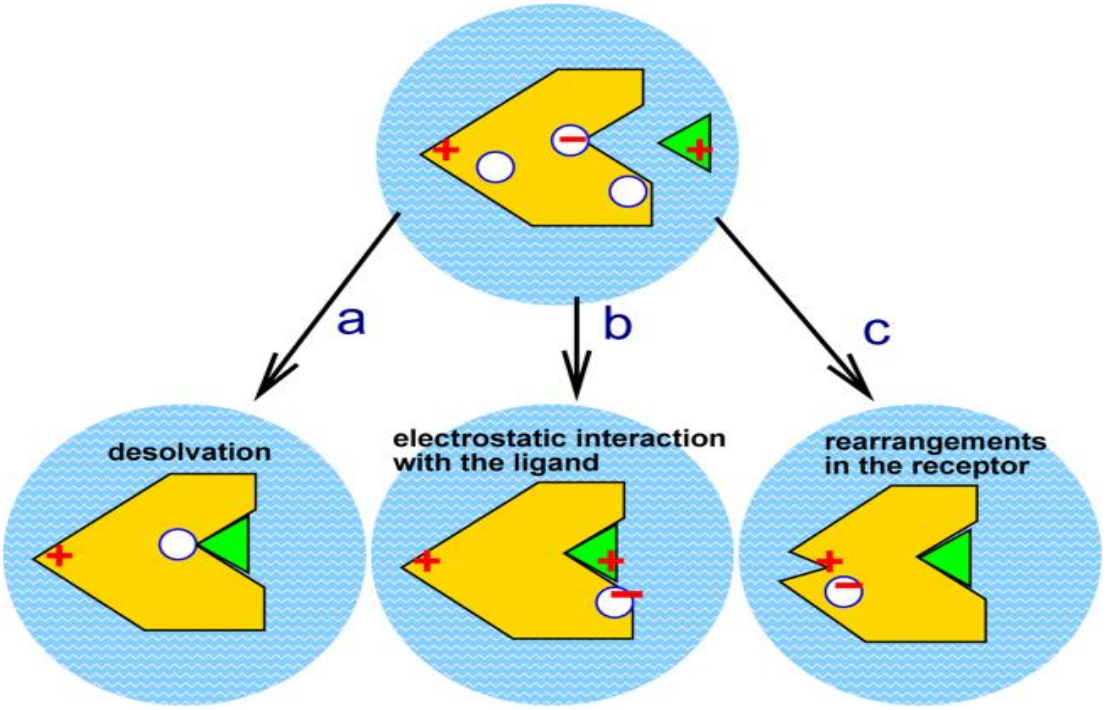


Figure 14 | Protonation and ionization mechanisms. Three main mechanisms are behind changes in the protonation state of an ionizable group in receptor protein upon ligand binding. a): desolvation of the group upon binding of the ligand, b): direct electrostatic interaction with the ligand, c): structural re-arrangements in the receptor protein caused by the binding (adapted from (186)).

Hydrogen bonds play a major role in many biological processes determining the infrastructural cohesion of proteins (192,193) or active site catalysis (194). As such, they are a key element in drug design (195). Hydrogen bonds usually involve atoms with electronegativity higher than hydrogen, similar to water (196). Hydrogen bonds, however, can occur in heterogeneous protein-protein interface systems, protein-ligand biophase or in correspondence with water-exposed sites of the protein (197). The contribution of the hydrogen bonds to energy stabilisation varies depending on the nature of the hydrogen bond donor or acceptor the distance and angle at which the atoms are positioned, and the atomic species involved (198).

It appears to be a universal relationship between hydrogen bond length and the difference in proton affinity of the hydrogen bond donor and acceptor and the energy contribution seems to depend on differential interactions of the solvent or surroundings between the free and hydrogen-bound states (199).

These considerations highlight the crucial role of hydrogen bonding in complex molecular systems, especially regarding the abundance, and distribution of this type of interaction and the importance of water molecules in biological systems.

2.4 The role of water in protein-ligand interaction

Water has a crucial role in the binding mechanism, as it acts as a ligand-protein bridge. We distinguish between "bulk water" and "biological water" according to its proximity to either other waters or biological molecules (197). Water molecules in biomolecular complexes are enthalpically more favourable than bulk solvent interactions with buried water molecules having longer residence times in protein structures (200) due to the contribution of nearby hydrophilic residues. It has been suggested that the presence of water molecules close to an active site interface positively contributes to the free energy of interaction with an estimated free energy change of -7.0 kJ/ mol (201,202). It follows that the presence of “biological water” could be used for a water-driven drug design, by using water as a positional marker for hydrogen bond donor or acceptor ligand extension (203). A variety of MD-based techniques and tools are available to identify the position of coordinating water in proteins, as well as tools able to quantify the energetic contribution produced by water (204–206).

3 Computer-Aided Drug Design: Modern Tools in Drug Discovery

The drug discovery process is hampered by a variety of obstacles: from the cost of synthesis to *in vivo* poor efficacy, stability, and safety of the compounds. As such, scrupulous attention must be given to the pharmacokinetics properties: absorption, distribution, metabolism, excretion, and toxicity (ADMET) to minimise complications at a later stage. Due to all these challenges, the number of drugs approved per financial investment has drastically diminished with a cost/benefit ratio that has increased in the last two decades. Generally, less than 10% of drugs that pass Phase I clinical trials obtain the Food and Drug Administration (FDA) approval for the intended therapeutic indication, including antibodies and natural products (207,207–209).

To overcome the serendipity of drug discovery (e.g. Sildenafil (210)) and save consistent amounts of time and resources cheminformatics methods can be deployed (209). Despite being far from perfect, many computational approaches were pivotal in the development of new drugs (211). The wide set of computational pharmacology techniques and software implemented falls under the general classification of computer-aided drug design (CADD) (212). An overview of CADD approaches can be seen in Table 4.

Strategies focused on ligands fall under the ligand-based drug design (LBDD), while the use of receptor's structural insights is categorized as structure-based drug design (SBDD). Specifically, SBDD is suited for target-based ligand optimisation strategies, which include the evaluation of binding pockets (213). Additionally, recent advancements in genomics, including the human genome project, (218) proteomics and structural information elevated the importance of SBDD in drug discovery (219). Since the early advancements in X-ray crystallography (161,220,221), nuclear magnetic resonance (NMR) (222), cryo-electron microscopy (cryo-EM) (223), and artificial intelligence (224) an increasing number of biological structures have been detailed to the atomic level. Knowledge of ternary and quaternary structures aided the pharmacological investigation of the protein-target interactions, contributing to solving specific selectivity or potency issues (225).

An important aspect of CADD is the definition of scoring functions and metrics. Such descriptors are based on mathematical models, such as molecular force fields, or quantum chemical parameters (214) used to extrapolate qualitative and quantitative data. Dar *et al* (215) defined scoring functions as "the most approximate mathematical methods used in computational drug designing for the prediction of the strength of the non-covalent interaction/binding affinity between two molecules. [...] most frequently, one molecule is a ligand and the other being the biological target such as a protein receptor " (215). CADD tools are not mutually exclusive, and it is generally a good practice to consider each strategy as a part of a broader validating process.

Table 4. A brief overview of molecular approaches and techniques in CADD

CADD Method	Description	Criticalities
Structure-Based Drug Design (SBDD)	Relies on knowledge of the cryo-EM, X-ray crystallography or NMR-derived 3D structure of a target	● Crystallographic artefacts ● Poor structure resolution
Ligand-Based Drug Design (LBDD)	Focuses on the geometric space or the atomic properties of a series of ligands, applying combinatorial techniques or 3D structure approaches	● May not consider multiple target interactions ● Low interest in target structure ● High risk of taking into consideration unsafe structures
Pharmacophore Based Drug Design (PBDD)	A specific aspect of LBDD where the attention is focussed on the biologically active moiety of a given ligand, exploring through combinatorial techniques or bioisosteres	● Limited chemical space explored ● Risk of inert or toxic compound ● Molecules might be difficult to synthesise
Fragment-Based Drug Design (FBDD)	An SBDD approach that starts from low-weight (< 300 kDa) molecules and builds the ligand structure with an additive approach to identify possible leads	● A delicate balance between efficacy/complexity and synthesis issues ● Built models might behave differently when combined

Generally, a holistic approach that joins SBDD and LBDD might be better suited for a comprehensive perspective on target structural interaction, binding energy evaluation, and the prediction of structures' toxicity and potency when investigating libraries of compounds (216,217). With the increased computational power, the integration of informatics, combinatorial chemistry and SBDD became a consolidated method to screen libraries of compounds and accelerate the drug discovery process (226).

3.1 Atomic Forces and Properties: From the Quantum Realm to Force Fields

"*How do molecules form?"* is still one of the most intriguing and complex questions in chemistry and physics. The quantum mechanical (QM) postulations (227) of molecular geometries and the properties associated with their energy level were established in the early twentieth century (181,228–230). The formulations that constitute the core of this project's methods derive from the approximation of the QM representation of atoms and molecules (231,232). Although the treatise of the QM complexity is beyond the scope of this project, I will present a brief highlight of the QM principles from which the force fields originate.

From a QM perspective, a molecule can be described as a multi-particle system whose properties are the result of electronic and nuclear interactions. The resulting electron-electron and electron-nuclei interaction energies define the geometry and energy of a particle system and can be represented as the linear combination of atomic orbitals (LCAO). Atomic orbitals are conceptualised as *wavefunctions*, which are a mathematical description of the quantum state of a system that corresponds to the probability amplitude of a particle to be found in a region of space (233).

In 1924, Louis de Broglie proposed the *particle-wave* duality, and this idea was later experimentally confirmed through the double-slit experiment. While Max Born and Werner Heisenberg developed matrix mechanics, Erwin Schrödinger postulated the equivalent mathematical model which used wave functions to describe the behaviour of particles (234). The intuition of Schrödinger's equation established the fundamental postulate of quantum mechanics where the solutions of the equation represent the state of a particle in an isolated system in a time-dependent manner. As indicated by the variational principle postulation, the Hamiltonian operator acting on a wavefunction describes the upper limit of the ground state energy (the *eigenvalue*) *(235)* of the kinetic and potential energy (equations 10 and 11) which represents an overestimation of the energy of the system.

$$H\,|\phi\rangle \;=\; \varepsilon|\phi \qquad \textbf{(10)}$$

In the above equation, H is the Hamiltonian acting on a trial function ϕ, returning the eigenvalue ε that represents the state of that function. Specifically, the Hamiltonian of a system

$$H \;=\; -\frac{\hbar^2}{2m}\nabla^2 \;+\; V\,(r) \qquad \textbf{(11)}$$

considers the kinetic energy (first term) as the second derivative with respect to the coordinates (the ∇ Laplacian operator) plus the potential energy V as a function of the position *r*. The definition of the energy of a system includes the nucleus-nucleus, nucleus-electron, and electron-electron interactions (Figure 15). Although the contribution of

the nuclei is not null, it is common practice in quantum mechanics to take into account only the electronic contribution, due to the minimal impact of nuclei on the state of a system (according to the Born-Oppenheimer approximation) (230).

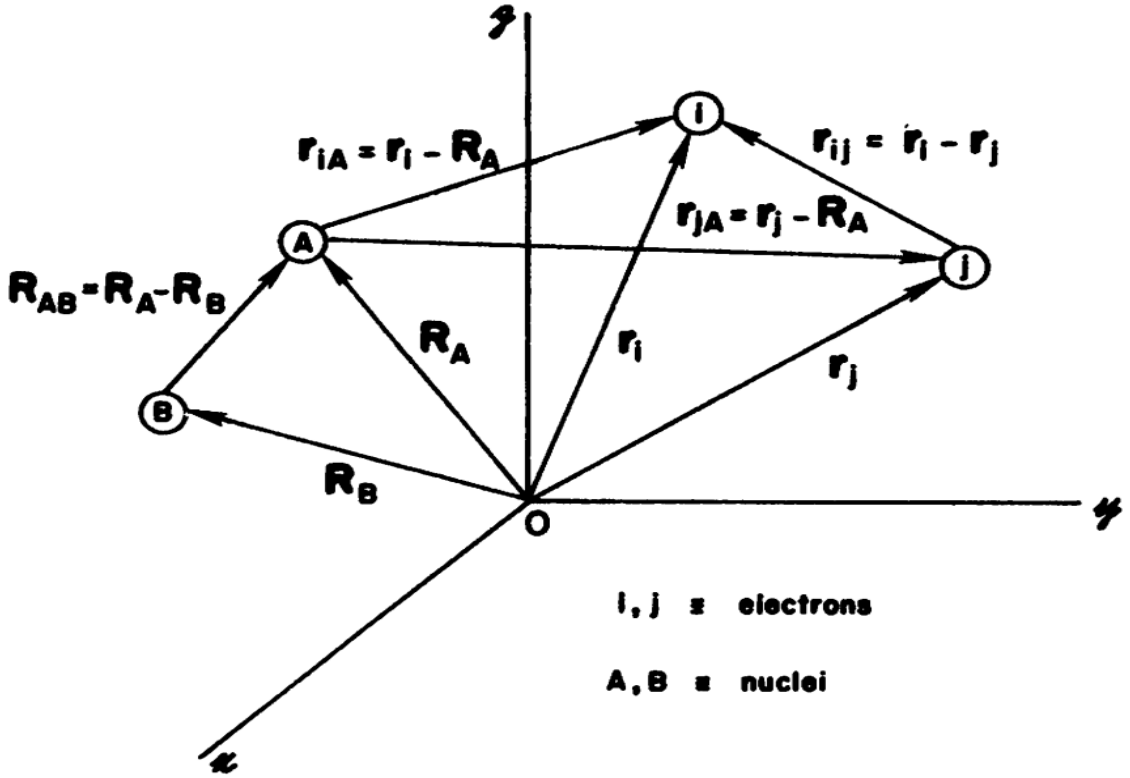


Figure 15 | Molecular coordinate system. The particle coordinates' system takes into account the internuclear distance |RA – RB| as well as the nuclear-electron distance |ri-RA|, the electron-electron distance |ri-rj| and the electron-nucleus distance |rj-RA| here represented on cartesian coordinates. Distances determine both the interaction level and the degree of interference between each component. The scheme is representative of an essential model of interaction of a multi-component system. Due to the negligible contribution of the nuclei due to their mass, their contribution is neglected (adapted from (230)).

The time-dependent Schrödinger differential equation explicit the Hamiltonian and represents the stationary state of a particle in a potential field V. The energy ε is explicitly represented on the right side of the equation:

$$-\frac{\hbar^2}{2m}\frac{d^2\Psi}{d\nabla^2} + V(r,t)\Psi = i\hbar\frac{d\Psi}{dt} \qquad \textbf{(12)}$$

where ħ is ≈ Planck's constant h over 2 π, m is the mass of the particle (i.e. generally the electron), V (r,t) is the potential energy as a function of position and time, and $i\hbar\frac{d\Psi}{dt}$ is the energy associated with that state.

As previously mentioned, the wave function Ψ is a mathematical function that describes the probability density of finding a particle at a specific position in time (236). Its exponential complexity is related to the difficulty of treating the interelectronic repulsion as a many-body system, which makes it impossible to find mathematically *exact solutions,* requiring approximate methods such as the variational principle or the perturbation (230). Ψ can be expressed as a linear combination of basis functions that can be determined by the self-consistent field method (SCF) to represent atomic orbitals (237). The combination of Gaussian-type orbital functions, namely a basis set, is used to conveniently describe the wave function. Ideally, the use of larger basis sets increases the accuracy of the energy estimation in exchange for additional computational costs. The plethora of basis sets and theoretical approaches available (230) require a rational decision between accuracy, speed and suitability for the considered chemical system (238–240).

3.2 Molecular Mechanics' Force Field

Molecular force fields (FFs) are a simplified representation that approximates the QM-derived geometries and energies of a molecule to a lower level of theory. The additive linear terms (239) of the FF formulation represent an approximation of the QM potential energy surfaces (241) to the molecular mechanics (MM) potential energy equation. In MM, the contribution of the electronic energy is simplified by the attribution of a formal charge defined by Coulomb's law, while the nuclei are represented as points in space with no correlating effect on the surrounding electron cloud. The electronic cloud represented by the wave function is replaced by a uniform distribution and a distance-dependent potential (242). The rovibrational contribution that characterises the vibrational spectra is substituted by the angle bending and the torsional excursion ranges (243). Although this method is convenient, it leaves out important information on the electronic structure (electron shell contributions, polarisation phenomena, induced dipole moment, as well as the ability to represent bond breaking and formation etc) representing a compromise between accuracy and calculation speed.

In this regard, force field implementation strikes a balance between the rigour of QM calculation and the increased, deemed necessary when tackling large systems with a conspicuous number of atoms. In such scenarios, unfortunately, the full QM calculation would be computationally demanding. Furthermore, MM force fields can be easily adjusted to experimental observations, becoming an asset for modern computational drug discovery.

Different force fields such as the Assisted Model Building with Energy Refinement (AMBER), Optimized Potentials for Liquid Simulations (OPLS), and the Chemistry at Harvard Macromolecular Mechanics (CHARMM) adopt different strategies for QM-experimental

parameter fitting overall showing a similar response for protein, nucleic adics, membrane or ligand (245). The translation of QM data to MM force field terms generates transferable parameters across different molecules (Figure 16).

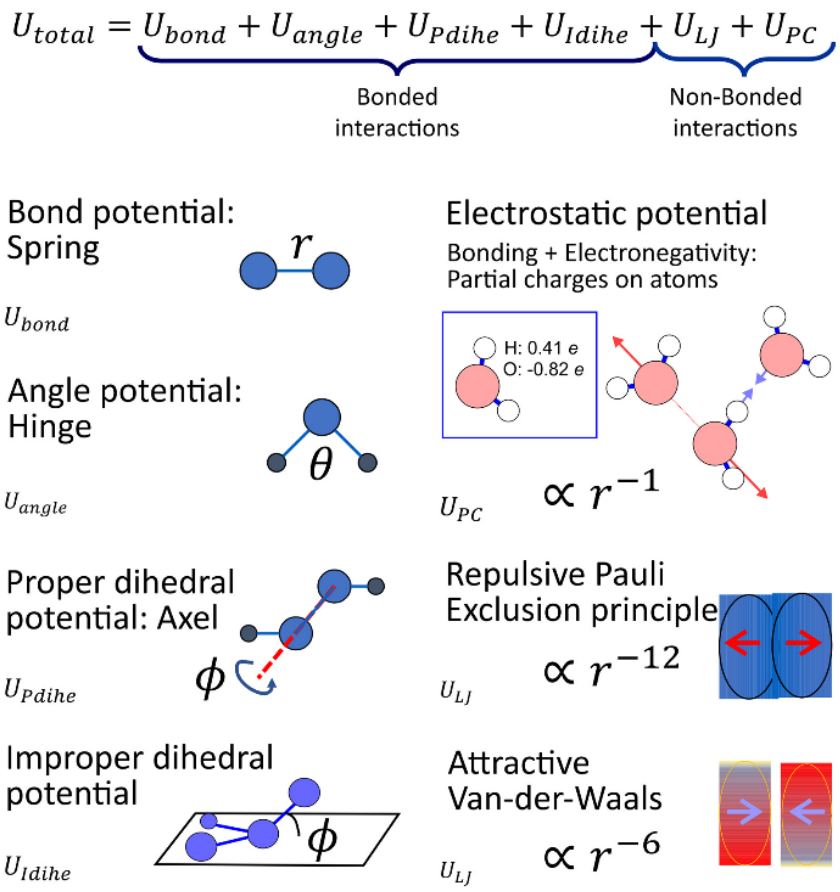


Figure 16 | Overview of the force field parameters. Visual representation of the additive members that constitute the forces driving the motion of atoms in MD. The non-bonded short-range attractive and repulsive forces are described conveniently by LJ formalism (adapted from (244)).

For the CHARMM force field (245), largely used in this project, numerous studies (246–248) experimentally corrected the FF parameters for nucleic acids, proteins, and membranes with the addition of additional correction terms. The following equations describe CHARMM's general force field which includes the aforementioned correction terms (249). The total potential energy of a molecule V_r can be divided into internal V_i and external V_e additive contributions such that

$$V_r = V_i + V_e$$

where

$$V_i = \sum_{bonds} K_b (b - b_0)^2 + \sum_{angles} K_\Theta (\Theta - \Theta_0)^2 + \sum_{dihedrals} K_\chi [(1 + cos(n\chi - \sigma)]$$

$$+ \sum_{impr\ dihe} K_\varphi (\varphi - \varphi_0)^2 + \sum_{Urey-Bradley} K_{UB} (r_{1,3} - r_{1,3,0})^2 + CMAP \quad \textbf{(13)}$$

defines the pairwise internal energy described by bonds, angles and dihedrals between connected atoms and

$$V_e = \sum_{nb\ atom\ pairs} (\varepsilon_{ij} [(\frac{R_{min,ij}}{r_{ij}})^{12} - (\frac{R_{min,ij}}{r_{ij}})^6] + \frac{q_i q_j}{\varepsilon_D r_{ij}}) \quad \textbf{(14)}$$

describes the nonbonded potential energy between atom pairs.

Equation 13 describes the types of bonded interactions that determine the geometry of a molecule in which each term is paired with an equilibrium value that minimises its potential energy. In the first term that models the vibration between two covalently bound atoms, the stretching of the bonds is approximated by a harmonic spring function. This harmonic spring describes the oscillation around an equilibrium bond length b_0 with bond constant K_b. The second term approximates the angle bending contributions to the potential energy and is defined for each triplet of atoms. It is also approximated by a harmonic function describing oscillation about an equilibrium angle θ_0 with force constant K_θ. The third term models the torsional energy between four consecutive atoms. The torsion angle χ is the angle of rotation with respect to two middle covalently bound atoms. It is calculated as the sum of cosine functions with n multiplicities where n is the number of multiple minima generated in a 360° rotation. The fourth term represents the out-of-plane contribution to the potential energy, usually in the form of an improper dihedral, where the potential energy is harmonic as a function of the out-of-plane angle φ.

The empirical validation of force field potential is a non-trivial effort that struggles with compatibility issues between experimental and QM data. In CHARMM, such adjustments required additional correction parameters such as a 2D dihedral energy correction map (255) (CMAP) to match high-level QM data fitting and the Urey-Bradley-Shimanouchi term(256), to match the spectroscopic anharmonic vibrational effect in proteins.

The first term of the sum in Equation 14 describes the van der Waals (vdW) interactions (250) using a 6-12 Lennard-Jones potential (LJ) term (251) (Figure 17) while the second term expresses the pairwise Coulomb interaction (252). Specifically, the first term inside the LJ member represents the exchange repulsion between atoms associated with the overlap of the electron clouds of the individual atoms (i.e., the Pauli exclusion principle) considering into account the difference between attractive forces (dipole-dipole, dipole-induced dipole, and London interactions) and repulsive forces. Although the LJ

potential was designed to describe the behaviour of noble gases (253) it is still widely used in biological systems.

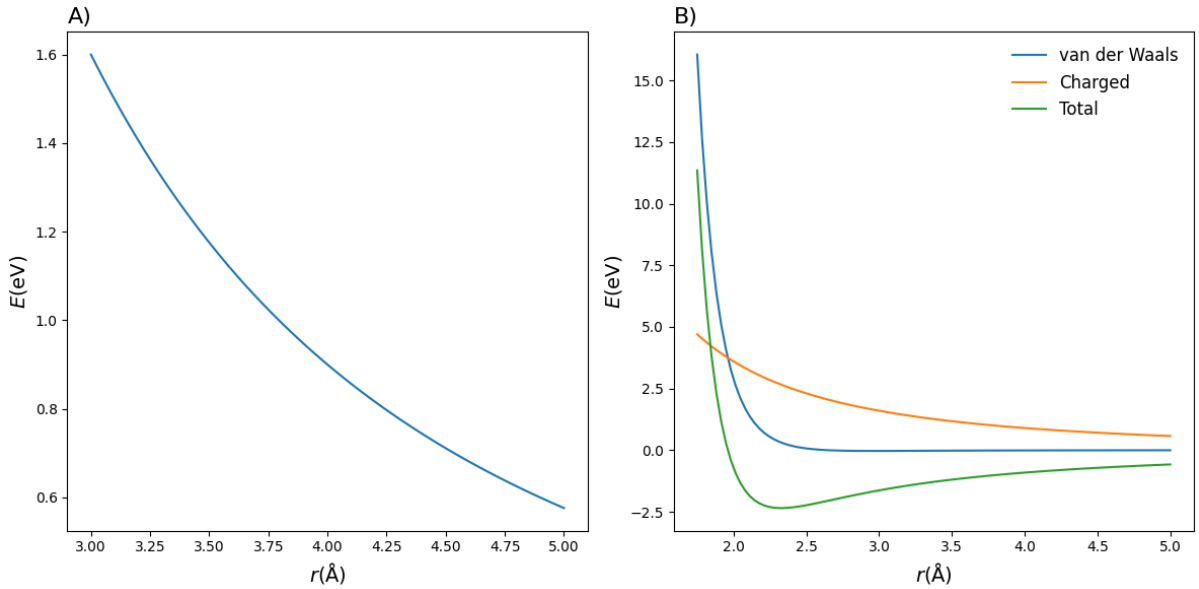


Figure 17 | Coulomb repulsion and vdW interactions. A) Coulomb interaction expressed with respect to the distance. **B)** Implementation of the LJ potential to calculate the energy of interaction expressed in electron volt (eV). The summative effect of Coulomb and LJ forces is shown in green and identifies the minimum potential well at ≃ 2.3 Å

The strong distance dependence of the repulsion is indicated by the 12th power of the first term of Equation 14, representing London dispersion interactions. The dipole–induced dipole interactions, instead, are the second 6th power term, which is negative, indicating its attractive nature as per convention. The epsilon ε_{ij} indicates the degree of magnitude of London's dispersion between atoms i, j and influences the depth of the curve; $R_{min,ij}$ is the minimum threshold distance at which the atoms i and j "feel" the minimum LJ interaction between them. R_{min} is related to the vdW radius of an atom intended as half the minimum distance between two nuclei. εi and $R_{min,ij}$ parameters are specific for individual atom types. For multiple atomic species (where, for instance, sp2 or sp3 carbons coexist in the same molecule) the combining rules are either the arithmetic or the geometric mean [i.e., εij (εiεj)1/2] for a simpler means of estimating the parameters (241). The simplification in atomic charges (q_i q_j) in the second term of Equation 14 is a computationally convenient Coulombic formalism (254) where r_i r_j is the distance between the atom i and j and ε_D is the dielectric constant.

4 Fundamentals of Molecular Docking and Dynamics

Molecular docking and all-atoms molecular dynamics (MD) simulations are widely adopted SBDD techniques (257,258). In molecular docking which predicts the binding modes and the affinity between a receptor and a ligand, the conformations of the ligand are explored by a search algorithm (259). The complementarity between the receptor and the generated docking poses is determined by a scoring function and ranked. Water molecules (260) or ions (261) can be considered in the calculations should they play a role in the receptor-ligand interaction. Interactions between the putative ligand and the target are calculated statically: receptor-ligand interactions do not induce any dynamic change in the system and the receptor is a rigid surface against which the ligand is tested. This limit inspired the exploration of flexible molecular docking protocols which treated a portion of the receptor as a flexible structure (262), allowing for ligand-target-induced adjustments. However, the solvation effect on the protein and the dynamic ligand-receptor interplay are neglected.

MD overcome the limitations of the docking approach by simulating the trajectory of the particles of a system over time (Section 4.2). While molecular docking portrays a snapshot of the ligand-receptor conformation, MD samples both receptor and ligand conformational changes as the result of the reciprocal interaction in continuum with the explicit solvent (192,263). The interplay between the solute and solvent reproduces solvation and hydrophobic effects and provides structural insights into the driving mechanism of binding selectivity (264,265) or kinetics (244). Furthermore, the versatility of MD allows the development of new approaches that can either speed up the sampling of rare kinetic events or provide additional information on a system's properties. The following sections outline docking and MD methods deployed in this project.

4.1 Molecular Docking

Molecular docking predicts the noncovalent binding between a receptor and a ligand and measures its binding affinity. As for all SBDD approaches, molecular docking requires the structure of the target which can be obtained through X-ray, cryo-electron microscopy (cryo-EM), and nuclear magnetic resonance (NMR). More recently, the neural-network-based model AlphaFold (224) and its advancement AlphaFold 2 (266,267) have predicted the structure of proteins in remarkable agreement with experimental structures. In molecular docking, a search algorithm evaluates ligands' possible conformations and rototranslational positionings allowed by a semiempirical force field's parameters (268). For each conformer, a score is calculated (269) by a function which represents the stability of a ligand within the target site.

As a good compromise between accuracy and speed, I adopted AutoDock Vina (270,271) and its enhanced version, QuickVina (Qvina) (272) as the main docking programs. The setup and workflow for a general molecular docking analysis are shown in Figure 18.

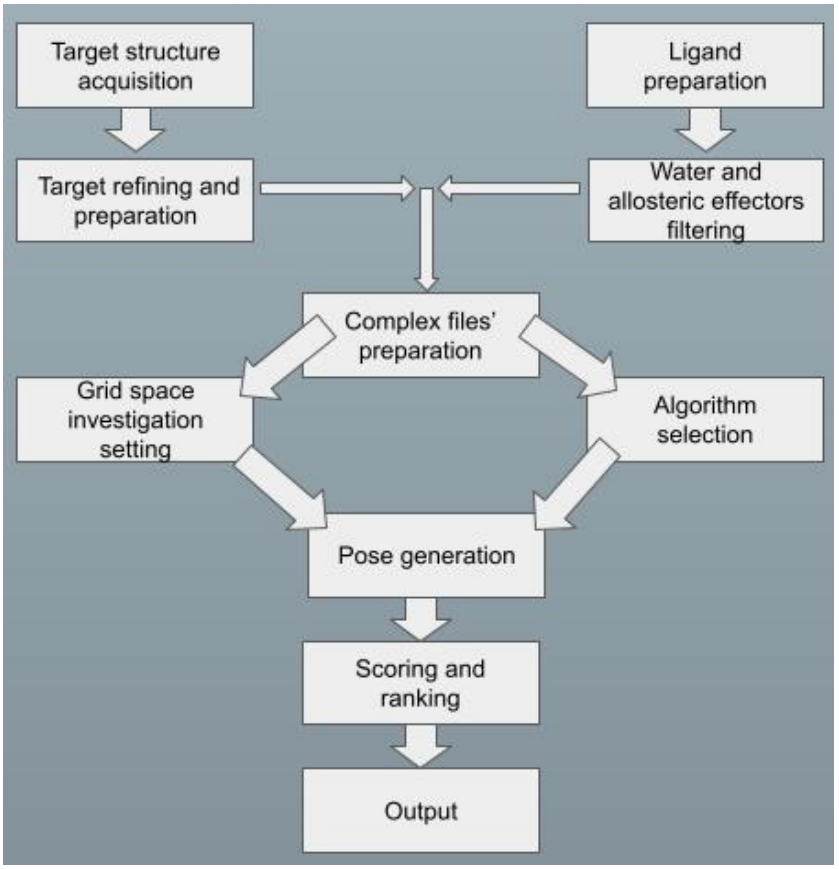


Figure 18 | Molecular Docking flow chart. Molecular docking requires a receptor and a ligand. The receptor's structure can be retrieved from repositories and might require additional steps before molecular docking like modelling missing residues or atoms and assigning protonation states to tritatable side chains Ligands need to be prepared as well and protonated accordingly. Water and cofactors should be carefully considered and kept should they participate in molecular recognition. Molecular docking requires that both receptor and ligand files are prepared and formatted to be used for the conformational search algorithm. Ultimately, a set of poses is generated, ranked and outputted.

In both Vina and QVina, the binding energy is calculated as the difference between the energies of the ligand and the protein in their unbound and bound states:

$$\Delta G = (V^{Lb-Lb} - V^{Lu-Lu}) + (V^{Rb-Rb} - V^{Ru-Ru}) + (V^{Rb-Lb} - V^{Ru-Lu} + \Delta S_{conf}) \quad \textbf{(15)}$$

where V is the potential energy, L refers to the ligand and R to the receptor. Lb and Lu refer to the ligand in the bound and unbound state respectively and Rb and Ru to the receptor in

its bound and unbound states. The computation of the potential energy differences between the bound and unbound states determines the ligand's binding disposition toward the selected binding site, where the estimation of the conformational entropic loss ΔS is derived from the summation of all the torsional degrees of freedom of the molecule. The potential energy terms are evaluated with a pair-wise forcefield:

$$V = W_{vdw} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) + W_{hbound} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}\right) +$$

$$+ W_{el} \sum_{i,j} \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} + W_{sol} \sum_{i,j} (S_i V_j + S_j V_i) e^{\left(\frac{-r_{ij}^2}{2\sigma^2}\right)} \quad \textbf{(16)}$$

where weights W are implemented to adjust the van der Waals, H-bond, electrostatic, and solvation terms to empirical values. A and B are the Lennard-Jones parameters extrapolated from the AMBER force field while C and D are atom-specific parameters adjusted to give a maximal well depth of -5 kcal/mol at 1.9 Å for O-H and N-H. E (t) is a smoothing directionality coefficient dependent on the deviation of the angle coefficient t from the H-bonds theoretical optimal geometry (273). The third term is a Coulomb electrostatic potential. The final term is a desolvation potential based on the volume (*V*) occupied by all the atoms surrounding a given atom, weighted by a solvation parameter (*S*) and an exponential term based on the square of the distance (274).

Despite being a widely accepted method, molecular docking's results, are not exhaustive: i) they are bound to the semiempirical forcefield that relies on LJ formulas and weights ii) do not directly consider the solvation effects of explicit water molecules iii) rely on target rigid structures that do not represent well the flexible nature of biological macromolecules. They are, however, a consolidated and relevant starting point for MD investigations (275).

4.2 A Classical Approach to Particles

In 1828 Robert Brown's observations of the movement of dust and pollen particles on a water droplet brought to attention the dynamic behaviour of particles in motion. Later in 1905, Albert Einstein explained Brownian's motion principles as the random thermal fluctuations that cause the dynamic movement of the particles in a system. The stochastic effect of these motions is due to the continuous collisions of the solvent's atoms with the solute, which follows the Gaussian distribution law (276,277). The summative contribution of these collisions varies the velocities and directions of the particles, causing their deflection or

redirection which determines a trajectory over time. Brownian motion's principles led to the development of mathematical techniques for the investigation of probabilistic effects (278,279) and led to the application of mathematical models and concepts deployed in MD to simulate multi-particle systems.

MD methods (Section 4.3) are based on FFs to evaluate the potential energy of molecules. Specifically, MD aims to replicate the thermodynamic fluctuation observed in nature in multiatomic systems and to determine the trajectories and velocities of those particles over time course (280).

Knowing the atomic forces, initial coordinates, and the masses of every atom of a system, it is possible to integrate the position changes with respect to time and determine the atoms' trajectory. Hence, it is possible to investigate biological systems using the classical mechanic's Newtonian equation of motion. According to Newton's second law and its time derivative:

$$F_i = m_i a_i = m_i \frac{\delta^2 r}{\delta t^2} \quad \textbf{(17)}$$

defines F (i) as the force acting on the "i" th particle, m (i) the mass of particle "i", and a (i) its acceleration represented as the second derivative of the particle position r with respect to time. The force is determined by the gradient of the potential energy function, which is also a function of all the atomic coordinates r:

$$F_i = \nabla_i U(r) \quad \textbf{(18)}$$

The force formulation of the second law can be rewritten in terms of the Hamiltonian form where:

$$H = \frac{\delta H(r,p)}{\delta p} \quad \textbf{(19)}$$

which indicates the variation of the total energy depending on the momentum and the position of the individual components of the system. Knowing the positions *r* and the momenta *p* of each particle, the calculation of the force and displacement can be derived at any time interval using a finite-difference approach to solve the differential equation. By approaching a discrete Taylor series expansion, it is possible to estimate the positions at each given timestep variation δt.

$$r(t + \Delta t) = r(t) + r(t)\Delta t \frac{1}{2!} a(t)\Delta t^2 +$$

$$+ r(t)\Delta t \frac{1}{3!} a(t)\Delta^3 + r(t)\Delta t \frac{1}{n!} a(t)\Delta t^n \ldots \quad \textbf{(20)}$$

The integration is calculated step by step and in its discrete form can be rewritten indicating the position occupied at step n (at time t) and r (n+1) and the next step indicated with “n+1” (at time t Δt) as:

$$r_{n+1} = r_n + v_n \Delta t + \frac{1}{2!} \left(\frac{F_n}{m}\right) \Delta t^2 + O\left(\Delta t^3\right) \quad \textbf{(21)}$$

However, due to the truncation of the series expansion (usually at the second term), this algorithm carries an error, that is largely dependent on the last truncation term O (Δt^3).

There are multiple mathematical approaches to minimise the truncation error in MD (281) with different advantages and disadvantages. One of the most used is Verlet’s integrator algorithm based on the summation of the back and forward series expansion such that:

$$r_{n-1} = r_n + v_n \Delta t + \frac{1}{2!} \left(\frac{F_n}{m}\right) \Delta t^2 - O\left(\Delta t^3\right) \quad \textbf{(22)}$$

The summation of the two series (eq 22 and eq 23) leads to the cancellation of the O (Δt^3) (as all the odd terms) and the square of the even terms of the series (282).

The advantages of the velocity Verlet algorithm lie in its numerical stability, reversibility, and accuracy with a minimal cost in overhead computation. Once the evolution of all the positions of the atoms present in the system is determined, it is possible to calculate the interatomic forces between the components of the system and determine whether attractive or repulsive interactions occur. In MD simulations, the initial velocities are assigned according to a low-temperature Boltzmann distribution (283) necessary to start the numerical integration of the equation of motion.

4.3 Molecular Dynamics

MD's popularity increased over the years since its first deployment (284,285) and it is now a powerful and widely used tool in chemistry, biology, physics, and materials science (286). MD is capable of sampling transition pathways or different local minima of the phase space (287), providing insights into ligand-receptor stabilisation, where the bound complex corresponds to the global minimum of the energy landscape (212,214).

Recently, thanks to the advancements in the graphic processing unit (GPU) and computational power, longer timescales are becoming more common in MD investigations (132). While the protein-ligand association can be simulated using classic MD approaches, the dissociation pathway requires specific enhanced sampling methods (288,289) to overcome the energy barrier associated with the intermolecular complex unbinding (290).

As previously stated, an MD simulation samples the phase space of a system by numerically integrating the equation of motion of the atoms with respect to time through the forces that are acting upon them (291). There are different treatises to define the equation of motion such as Newton's classical expression (292,293), the Brownian equation of motion (294), Langevin's stochastic equation (295,296), or their derivation extrapolated from a combination of quantum and classical mechanics terms (QM/MM) (297) up to a complete *ab-initio* formulation (298,299). In this project, we refer to Newton's treatise for its fast numerical integration and the extrapolation of qualitative and quantitative thermodynamic information.

In a classical system, the Hamiltonian (H) is a function dependent on the coordinates *r* and momenta *p* which are used to describe the system's energy. The atoms' positions are used to calculate the forces acting between them according to the force field parameters. The Hamiltonian is equal to the total energy:

$$H = H(r,p) = K(p) + U(r) = \sum_i \frac{p_i}{2m_i} + U(r) \quad \textbf{(23)}$$

where *K (p)* indicates the kinetic energy and *U (r)* is the potential energy. In this scenario, the phase space can be described by the relationship between its energy intended as the sum of the angular momenta of each particle p_i over $2m_i$, the potential energy U dependent on the position of each particle *r (282)*. An MD system where the number of particles (N) volume (V) and energy (E) are constant (NVE) is called a microcanonical ensemble. The microcanonical ensemble undergoes adiabatic processes where no thermal energy is exchanged outside the system. It follows that in case of structural changes or prolonged collisions between the atoms, the instantaneous kinetic energy is redistributed throughout the system, causing a temperature drift (300). Energy drifts can derive also from the accumulation of numerical errors (301), leading to an excessive increment in temperature (T) which might result in non-Newtonian dynamics or unphysical results. However, experimental systems are not generally held at constant V, but they are performed under constant pressure (P) or T. While NVE might be appropriate to simulate gas-phase systems (302), variations in T and P are more representative of the thermodynamic exchanges occurring in nature or biological experiments. It is, therefore, preferable to perform such simulations in other ensembles, such as the canonical (NVT) or the isothermal-isobaric (NPT). To correct the temperature or pressure drifts, the system is coupled with a thermostat or a barostat which act on the kinetic energy of the system or by scaling the location of the particles with respect to the simulation cell size respectively (303). Many barostats and thermostat methods were developed (304) with the Lanvegin thermostat and Monte Carlo barostat (305) being the ones used in this project by the ACEMD engine (306).

Controlling T might be desirable to determine conformational changes at a given temperature or to simulate an annealing protocol where the variation of T allows for a broader phase-space exploration as T is related to the total kinetic energy (307). For T control, Berendsen (308) proposed coupling the system to an external bath at a fixed T to adjust the momenta of the particles by adding or subtracting heat from the system. This scaling is obtained by applying dissipative Langevin forces in the equations of motion as

$$a_i(t) = \frac{F_i(T)}{m_i} + \frac{p_i(t)}{m_i \tau}\left[\frac{T_0}{T(t)} - 1\right] \quad \textbf{(24)}$$

where T (t) is the temperature at the timestep t and τ is a scaling factor that controls the impact on the momentum of the *i*-th particle.

In an isobaric ensemble, at constant pressure P, the volume V needs to accommodate the pressure by coupling our MD system to a barostat (309) where the volume V is adjusted to maintain the desired P. From the ideal gas statistical mechanics, the pressure P is described as

$$P(t) = \frac{1}{V(t)}\left[Nk_bT(t) + \frac{1}{3}\sum_i^N \sum_{j>1}^N F_{ij} r_{ij}\right] \quad \textbf{(25)}$$

where V is the volume at the timestep t, N is the number of particles, K_b is the Boltzmann's constant and Fij rij are the forces and the distances between particles. In ACEMD's Monte Carlo barostat, a change in volume is estimated after a time interval when forces are calculated. Iterative volume changes are generated and are kept should they fall within a statistical Monte Carlo acceptance threshold of 40-50% (310). The system's box length and the coordinates of each molecule are then rescaled according to:

$$l'_i = l_i \sqrt[3]{\frac{V'}{V}} \text{ and } r'_i = (r_i - c_i)\sqrt[3]{\frac{V'}{V}} + C_i \quad \textbf{(26)}$$

where i= 1, 2, 3, V'= V + ΔV and Ci are the coordinates of the centre of the periodic box (305). Careful consideration of both thermostat and barostat parameters needs to be dedicated to ensuring the numerical stability of the system, as well as its consistency with biochemical experiments.

To extrapolate statistically representative averages of observable variables of an ensemble during an MD simulation, one should know the probability (ρ) of finding that system configuration at each point of the phase (311) space by calculating the Boltzmann distribution as follows:

$$\rho(r, p) = -\frac{exp[-H(r,p)/KbT]}{Z}$$

(27)

where H is the Hamiltonian and Z is the partition function that represents the number of thermally accessible states of a system.
Specifically, the “canonical” NVT partition function Z elegantly describes all the possible states of a system as:

$$Z\,(N,V,T) \;=\; \frac{1}{N!h^{3N}}\int e^{-\beta H\,(r,p)}\,d^{3N}p\,d^{3N}q \qquad \textbf{(28)}$$

where *h* is Plank’s constant and the function is integrated over the whole phase space for each particle of the system. The ergodic principle assumes that in a long *enough* time, a system will explore all its possible microstates (307), and that is it therefore, possible, once the probability distribution is known, to calculate the phase space averages of any dynamic variable at any coordinates and momenta. In reality, very long computational simulations are still uncommon, and this strategy would be computationally expensive, requiring also a very large number of MD replicas.

4.4 Enhanced and Adaptive MD Sampling

MD is a powerful tool for studying qualitative and quantitative changes in biological systems. However, as already stated in Chapter 2, biological systems are characterized by numerous minima that sit at different positions in the PES. These many local minima are often separated by high-energy barriers (312), which could result in a simulation being stuck inside a saddle point of the PES. In the case of large conformational changes or the dissociation of intermolecular complexes, classic MD approaches can be time-consuming and carry the risk of falling into one of the PES’s saddle points (158,313). Overcoming the energetic barriers or avoiding the oversampling of metastable states requires, therefore, new approaches to improve the exploration of the PES. Such improvement can be achieved by acting on the simulated properties of the system to favour the sampling of statistically rare events. These techniques are generally distinguished into two groups: enhanced sampling and adaptive sampling.

Adaptive sampling techniques are characterized by the preservation of the thermodynamic ensemble where the sampling enhancement is achieved by either restarting MD trajectories at different conditions or particularly chosen seeds (314). Enhanced sampling techniques accelerate the exploration of the system by modifying the PES through the addition of a force or a bias potential to the Hamiltonian of the systems (315). This bias acts on the valleys of the PES, decreasing the energy barrier needed to sample transitions. This allows for a gradual energy increase that pushes the system into adjacent accessible states over the PES. A brief non-exhaustive list of methods that alter the potential energy function includes: i) selectively scaled MD (316), where specific energy terms of the potential

energy function are gradually altered to promote dissociative events during simulations, ii) accelerated MD (317) where the potential energy function is varied according to reaching a threshold, iii) temperature replica exchange (T-REMD) (318) where the states of parallel simulations performed at different temperatures are exchanged according to a statistical criterion iv) Hamiltonian replica exchange MD (H-REMD) where the various replicas are simulated at constant or variable temperatures, but with different parameter sets for the equations of motion (319).

Enhanced methods like metadynamics (320) (321) (322) require order parameters upon which the energetic bias is applied. The force field as well as atomic coordinates can be combined to define the order parameters. Hence, the order parameter is the mathematical combination of multiple degrees of freedom such as the angles formed by non-bonded atoms or distances between groups of atoms or other experimentally derived variables (323) that describe a specific state of a system. The behaviour of an MD simulation can be altered by influencing one or more order parameters named collective variables (CVs) through the addition of a scalar value (324) to the potential energy of that portion of the system. With the term CVs, we refer to the projection of the numerous degrees of freedom of a simulated system over usually just one to three metrics to identify and separate different macrostates (i.e. deep local energy minima). Throughout the simulation, an energy bias can be added to these CVs to increase the probability of observing the desired transition (Figure 19).

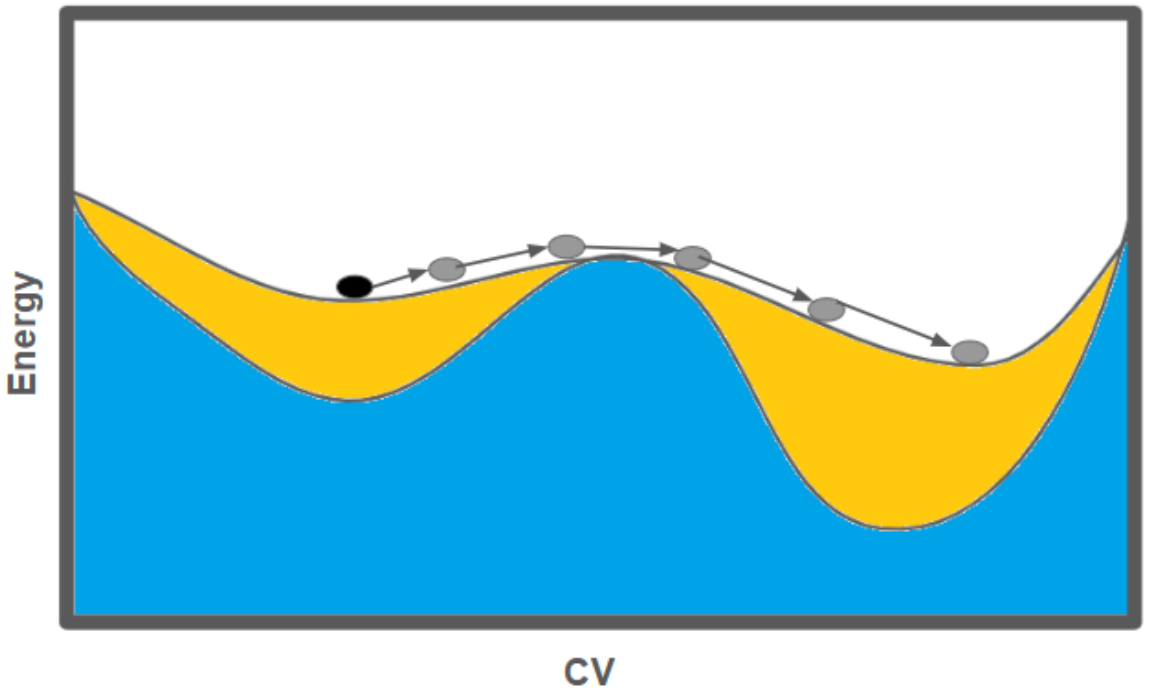


Figure 19 | Energy bias applied to a collective variable. A schematic representation of an incremental energy bias (yellow) is used to overcome the transitional energy barrier in the direction indicated by the arrow. This bias is necessary to explore adjacent minima in the PES and can also be gradually tuned for a natural system relaxation to the new saddle point.

Generally, the definition of the order parameters to describe binding or unbinding events or structural changes is not trivial. Conformational changes, for instance, might require specific sequences of structural rearrangement or simultaneous events that might require a trial-and-error approach. Additionally, the height, the width, and the time interval at which the Gaussian bias is introduced influence the behaviour of the system. Ideally, the height of the Gaussian should be considerably smaller than the highest energy barrier to prevent overfitting or unwanted unfolding. The bell shape, as well, should be small enough to maintain a sufficient level of resolution inside multiple adjacent saddles: a broad Gaussian potential might cover one or more adjacent minima, hiding meaningful transitory states that might be connected through close smaller paths. The time interval plays also a crucial role in our sampling since i) the error introduced during metadynamics is inversely proportional to the square root of the time deposition (325) ii) the system will need adequate time to explore the new set point iii) a frequent energy bias might lead to the instability of the system. It follows that an adequately large time interval should be set.

The issues related to the overfilling effect of metadynamics led to the development of well-tempered metadynamics (WTmetaD) which sets a threshold parameter to automatically smooth the energy addition based on the previous history, limiting the PES exploration and minimising the bias added (322).

4.5 Supervised and Multiple Walker Supervised Molecular Dynamics (mwSuMD)

Although enhanced methods are a viable aid to observe conformational shifts or to simulate binding or unbinding pathways, they might force a system to the desired state, rather than observing a spontaneous behaviour. In this scenario, unbiased adaptive methods could still offer an alternative to the addition of external forces to the system. Supervised molecular dynamics (326,327) (SuMD) is a powerful adaptive MD technique for studying ligand-receptor binding and unbinding pathways (328). SuMD drastically reduces the timescale necessary to sample intermolecular complex formations and dissociations to less than hundreds of nanoseconds timescale. A tabu–like algorithm (Figure 20) is designed to monitor the distance between the centre of masses (COM) of the ligand and a selected binding site during short classic MD simulations. Each simulation represents a step from which the distances of the two COMs are collected. The distances are used to define a linear

function with angular coefficient *m for two* consecutive steps. If the distance between two steps decreases and the angular coefficient *m* of the regression line is negative (in case of an unbinding) then a subsequent simulation step is run. Otherwise, another unbiased simulation is restarted from the previous step.

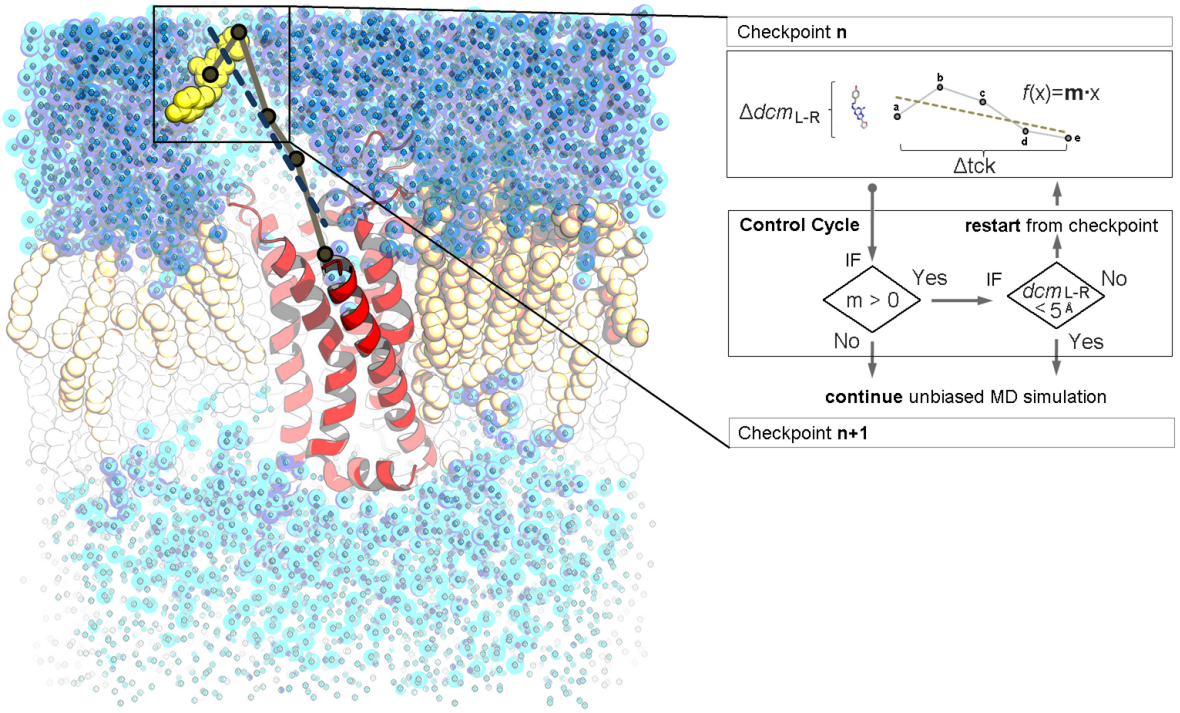


Figure 20 | SuMD tabu-like algorithm. The distance vector between two selections ((dcmL-R) is collected at regular intervals. The result is then interpolated to describe a linear regression curve where the angular coefficient m is used as a discriminant. In the case of binding, if the coefficient m is negative, the unbiased simulation continues with a new short unbiased MD, otherwise, the simulation will be restarted from the previous step (figure courtesy of the Molecular Modelling Section - University of Padova).

We developed an enhanced version of SuMD named multiple walker supervised molecular dynamics (mwSuMD). At its core, mwSuMD operates similarly to SuMD by supervising one or two metrics of the system. Additionally, we expanded the metrics' choices to the positional Root Mean Square Deviation (RMSD) of a selection to measure the geometric difference from reference coordinates, intermolecular distances for binding/unbinding monitoring, and the occurrence of molecular contacts between two entities to investigate molecular interaction patterns. The data are used to observe the evolution of one or two selections defined by the user in a SuMD-like scheme of check-continue-else-restart.

By monitoring these metrics, mwSuMD provides invaluable insights into the binding and unbinding pathways of interacting molecules. Moreover, paired with free energy estimation methods such as MMPBSA and MMGBSA (329) it delves into the quantitative and qualitative aspects of molecular contacts, shedding light on the intricacies of molecular recognition and binding affinities. The method's ability to discern binding pathways and conformational dynamics offers a powerful tool for investigating the interactions between molecules of interest.

With mwSuMD, we unveiled hidden G protein-coupled receptors (GPCRs) structural transitions (330), provided crucial ligand-receptor binding insights for small molecule design against HIV’s Nef protein (331) and investigated the unbinding pathway for SARS-CoV-2 Mpro-S-217622, the first oral noncovalent inhibitor (332). Additionally, mwSuMD's strength lies in its parallel architecture and compatibility with ACEMD, NAMD, GROMACS and OPENMM engines as well as its user-friendly design.

5 Chapter 1-4 References:

1. Pene F, Merlat A, Vabret A, Rozenberg F, Buzyn A, Dreyfus F, et al. Coronavirus 229E-related pneumonia in immunocompromised patients. Clin Infect Dis. 2003 Oct 1;37 (7):929–32.
2. Pawlowski C, Venkatakrishnan AJ, Ramudu E, Kirkup C, Puranik A, Kayal N, et al. Pre-existing conditions are associated with COVID-19 patients' hospitalization, despite confirmed clearance of SARS-CoV-2 virus. EClinicalMedicine. 2021 Apr;34:100793.
3. Ragab D, Salah Eldin H, Taeimah M, Khattab R, Salem R. The COVID-19 Cytokine Storm; What We Know So Far. Front Immunol. 2020 Jun 16;11:1446.
4. Guan W-J, Ni Z-Y, Hu Y, Liang W-H, Ou C-Q, He J-X, et al. Clinical characteristics of coronavirus disease 2019 in china. N Engl J Med. 2020 Apr 30;382 (18):1708–20.
5. Al-Samkari H, Karp Leaf RS, Dzik WH, Carlson JCT, Fogerty AE, Waheed A, et al. COVID-19 and coagulation: bleeding and thrombotic manifestations of SARS-CoV-2 infection. Blood. 2020 Jul 23;136 (4):489–500.
6. Tang N, Li D, Wang X, Sun Z. Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. J Thromb Haemost. 2020 Apr;18 (4):844–7.
7. Siddell SG. The coronaviridae: an introduction. In: Siddell SG, editor. The Coronaviridae. Boston, MA: Springer US; 1995. p. 1–10.
8. Kirtipal N, Bharadwaj S, Kang SG. From SARS to SARS-CoV-2, insights on structure, pathogenicity and immunity aspects of pandemic human coronaviruses. Infect Genet Evol. 2020 Nov;85:104502.
9. Memish ZA, Mishra N, Olival KJ, Fagbo SF, Kapoor V, Epstein JH, et al. Middle East respiratory syndrome coronavirus in bats, Saudi Arabia. Emerging Infect Dis. 2013 Nov;19 (11).
10. Zeng Q, Langereis MA, van Vliet ALW, Huizinga EG, de Groot RJ. Structure of coronavirus hemagglutinin-esterase offers insight into corona and influenza virus evolution. Proc Natl Acad Sci USA. 2008 Jul 1;105 (26):9065–9.
11. Xu C, Wang Y, Liu C, Zhang C, Han W, Hong X, et al. Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM. Sci Adv. 2021 Jan 1;7 (1).
12. Masters PS. The Molecular Biology of Coronaviruses. Elsevier; 2006. p. 193–292.

13. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science. 2020 Mar 13;367 (6483):1260–3.

14. Zhang Z, Nomura N, Muramoto Y, Ekimoto T, Uemura T, Liu K, et al. Structure of SARS-CoV-2 membrane protein essential for virus assembly. Nat Commun. 2022 Aug 5;13 (1):4399.

15. Bai Z, Cao Y, Liu W, Li J. The SARS-CoV-2 Nucleocapsid Protein and Its Role in Viral Structure, Biological Functions, and a Potential Target for Drug or Vaccine Mitigation. Viruses. 2021 Jun 10;13 (6).

16. Yadav R, Chaudhary JK, Jain N, Chaudhary PK, Khanra S, Dhamija P, et al. Role of Structural and Non-Structural Proteins and Therapeutic Targets of SARS-CoV-2 for COVID-19. Cells. 2021 Apr 6;10 (4).

17. Bosch BJ, van der Zee R, de Haan CAM, Rottier PJM. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. J Virol. 2003 Aug;77 (16):8801–11.

18. Xia X. Domains and Functions of Spike Protein in Sars-Cov-2 in the Context of Vaccine Design. Viruses. 2021 Jan 14;13 (1).

19. Fan X, Cao D, Kong L, Zhang X. Cryo-EM analysis of the post-fusion structure of the SARS-CoV spike glycoprotein. Nat Commun. 2020 Jul 17;11 (1):3618.

20. Al Adem K, Shanti A, Stefanini C, Lee S. Inhibition of SARS-CoV-2 Entry into Host Cells Using Small Molecules. Pharmaceuticals (Basel). 2020 Dec 8;13 (12).

21. Tikellis C, Thomas MC. Angiotensin-Converting Enzyme 2 (ACE2) Is a Key Modulator of the Renin Angiotensin System in Health and Disease. Int J Pept. 2012 Mar 20;2012:256294.

22. Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature. 2020 May;581 (7807):215–20.

23. Woo H, Park S-J, Choi YK, Park T, Tanveer M, Cao Y, et al. Developing a Fully Glycosylated Full-Length SARS-CoV-2 Spike Protein Model in a Viral Membrane. J Phys Chem B. 2020 Aug 20;124 (33):7128–37.

24. Pipitò L, Rujan R-M, Reynolds CA, Deganutti G. Molecular dynamics studies reveal structural and functional features of the SARS-CoV-2 spike protein. Bioessays. 2022 Sep;44 (9):e2200060.

25. Benton DJ, Wrobel AG, Xu P, Roustan C, Martin SR, Rosenthal PB, et al. Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. Nature. 2020 Dec;588 (7837):327–30.

26. McCallum M, De Marco A, Lempp FA, Tortorici MA, Pinto D, Walls AC, et al. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. Cell. 2021 Apr 29;184 (9):2332-2347.e16.

27. Cai Y, Zhang J, Xiao T, Peng H, Sterling SM, Walsh RM, et al. Distinct conformational states of SARS-CoV-2 spike protein. Science. 2020 Sep 25;369 (6511):1586–92.

28. Shang J, Wan Y, Luo C, Ye G, Geng Q, Auerbach A, et al. Cell entry mechanisms of SARS-CoV-2. Proc Natl Acad Sci USA. 2020 May 26;117 (21):11727–34.

29. Choi JY, Smith DM. SARS-CoV-2 Variants of Concern. Yonsei Med J. 2021 Nov;62 (11):961–8.

30. Thakur V, Ratho RK. OMICRON (B.1.1.529): A new SARS-CoV-2 variant of concern mounting worldwide fear. J Med Virol. 2022 May;94 (5):1821–4.

31. Gupta D, Sharma P, Singh M, Kumar M, Ethayathulla AS, Kaur P. Structural and functional insights into the spike protein mutations of emerging SARS-CoV-2 variants. Cell Mol Life Sci. 2021 Dec;78 (24):7967–89.

32. Iketani S, Liu L, Guo Y, Liu L, Chan JF-W, Huang Y, et al. Antibody evasion properties of SARS-CoV-2 Omicron sublineages. Nature. 2022 Apr;604 (7906):553–6.

33. Al-Tawfiq JA, Koritala T, Alhumaid S, Barry M, Alshukairi AN, Temsah M-H, et al. Implication of the emergence of the delta (B.1.617.2) variants on vaccine effectiveness. Infection. 2022 Jun;50 (3):583–96.

34. Mistry P, Barmania F, Mellet J, Peta K, Strydom A, Viljoen IM, et al. SARS-CoV-2 Variants, Vaccines, and Host Immunity. Front Immunol. 2021;12:809244.

35. McCarthy KR, Rennick LJ, Nambulli S, Robinson-McCarthy LR, Bain WG, Haidar G, et al. Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. Science. 2021 Mar 12;371 (6534):1139–42.

36. Mandal N, Padhi AK, Rath SL. Molecular insights into the differential dynamics of SARS-CoV-2 variants of concern. J Mol Graph Model. 2022 Apr 14;114:108194.

37. Thakur S, Sasi S, Pillai SG, Nag A, Shukla D, Singhal R, et al. SARS-CoV-2 Mutations and Their Impact on Diagnostics, Therapeutics and Vaccines. Front Med (Lausanne). 2022 Feb 22;9:815389.

38. Yamasoba D, Kosugi Y, Kimura I, Fujita S, Uriu K, Ito J, et al. Sensitivity of novel SARS-CoV-2 Omicron subvariants, BA.2.11, BA.2.12.1, BA.4 and BA.5 to therapeutic monoclonal antibodies. BioRxiv. 2022 May 3;

39. Mahase E. Covid-19: What we know about the BA.4 and BA.5 omicron variants. BMJ. 2022 Aug 9;378:o1969.

40. Peacock TP, Brown JC, Zhou J, Thakur N, Newman J, Kugathasan R, et al. The SARS-CoV-2 variant, Omicron, shows rapid replication in human primary nasal epithelial cultures and efficiently uses the endosomal route of entry. BioRxiv. 2022 Jan 3;

41. Callaway E. Why does the Omicron sub-variant spread faster than the original? Nature. 2022 Feb;602 (7898):556–7.

42. Arora P, Zhang L, Krüger N, Rocha C, Sidarovich A, Schulz S, et al. SARS-CoV-2 Omicron sublineages show comparable cell entry but differential neutralization by therapeutic antibodies. Cell Host Microbe. 2022 May 6;

43. Unable to find information for 12902880.

44. Meng B, Ferreira I, Abdullahi A, Kemp SA, Goonawardane N, Papa G, et al. SARS-CoV-2 Omicron spike mediated immune escape, infectivity and cell-cell fusion. BioRxiv. 2021 Dec 21;

45. Kumar R, Murugan NA, Srivastava V. Improved binding affinity of the Omicron's spike protein with hACE2 receptor is the key factor behind its increased virulence. BioRxiv. 2021 Dec 28;

46. Yin W, Xu Y, Xu P, Cao X, Wu C, Gu C, et al. Structures of the Omicron spike trimer with ACE2 and an anti-Omicron antibody. Science. 2022 Mar 4;375 (6584):1048–53.

47. Lupala CS, Ye Y, Chen H, Su X-D, Liu H. Mutations on RBD of SARS-CoV-2 Omicron variant result in stronger binding to human ACE2 receptor. Biochem Biophys Res Commun. 2022 Jan 29;590:34–41.

48. Shuai H, Chan JF-W, Hu B, Chai Y, Yuen TT-T, Yin F, et al. Attenuated replication and pathogenicity of SARS-CoV-2 B.1.1.529 Omicron. Nature. 2022 Mar;603 (7902):693–9.

49. SARS-CoV-2 variants of concern as of 20 October 2023 [Internet]. [cited 2023 Nov 8]. Available from: https://www.ecdc.europa.eu/en/covid-19/variants-concern

50. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a

Clinically Proven Protease Inhibitor. Cell. 2020 Apr 16;181 (2):271-280.e8.

51. Zamorano Cuervo N, Grandvaux N. ACE2: Evidence of role as entry receptor for SARS-CoV-2 and implications in comorbidities. eLife. 2020 Nov 9;9.

52. Du L, He Y, Zhou Y, Liu S, Zheng B-J, Jiang S. The spike protein of SARS-CoV--a target for vaccine and therapeutic development. Nat Rev Microbiol. 2009 Mar;7 (3):226–36.

53. Clausen TM, Sandoval DR, Spliid CB, Pihl J, Perrett HR, Painter CD, et al. SARS-CoV-2 Infection Depends on Cellular Heparan Sulfate and ACE2. Cell. 2020 Nov 12;183 (4):1043-1057.e15.

54. Kim SY, Jin W, Sood A, Montgomery DW, Grant OC, Fuster MM, et al. Characterization of heparin and severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) spike glycoprotein binding interactions. Antiviral Res. 2020 Sep;181:104873.

55. Mycroft-West C, Su D, Elli S, Guimond S, Miller G, Turnbull J, et al. The 2019 coronavirus (SARS-CoV-2) surface protein (Spike) S1 Receptor Binding Domain undergoes conformational change upon heparin binding. BioRxiv. 2020 Mar 2;

56. Simon Davis DA, Parish CR. Heparan sulfate: a ubiquitous glycosaminoglycan with multiple roles in immunity. Front Immunol. 2013 Dec 18;4:470.

57. Meneghetti MCZ, Hughes AJ, Rudd TR, Nader HB, Powell AK, Yates EA, et al. Heparan sulfate and heparin interactions with proteins. J R Soc Interface. 2015 Sep 6;12 (110):0589.

58. Lim GB. Milestone 1: Discovery and purification of heparin. Nat Rev Cardiol. 2017 Dec 14;

59. Yue J, Jin W, Yang H, Faulkner J, Song X, Qiu H, et al. Heparan Sulfate Facilitates Spike Protein-Mediated SARS-CoV-2 Host Cell Invasion and Contributes to Increased Infection of SARS-CoV-2 G614 Mutant and in Lung Cancer. Front Mol Biosci. 2021 Jun 11;8:649575.

60. Kim SY, Jin W, Sood A, Montgomery DW, Grant OC, Fuster MM, et al. Glycosaminoglycan binding motif at S1/S2 proteolytic cleavage site on spike glycoprotein may facilitate novel coronavirus (SARS-CoV-2) host cell entry. BioRxiv. 2020 Apr 15;

61. Schuurs ZP, Hammond E, Elli S, Rudd TR, Mycroft-West CJ, Lima MA, et al. Evidence of a putative glycosaminoglycan binding site on the glycosylated

SARS-CoV-2 spike protein N-terminal domain. Comput Struct Biotechnol J. 2021 May 4;19:2806–18.

62. Unable to find information for 9239468.

63. Kwon PS, Oh H, Kwon S-J, Jin W, Zhang F, Fraser K, et al. Sulfated polysaccharides effectively inhibit SARS-CoV-2 in vitro. Cell Discov. 2020 Jul 24;6 (1):50.

64. Mycroft-West C, Su D, Pagani I, Rudd T, Elli S, Guimond S, et al. Heparin inhibits cellular invasion by SARS-CoV-2: structural dependence of the interaction of the surface protein (spike) S1 receptor binding domain with heparin. BioRxiv. 2020 Apr 28;

65. Goyal B, Goyal D. Targeting the Dimerization of the Main Protease of Coronaviruses: A Potential Broad-Spectrum Therapeutic Strategy. ACS Comb Sci. 2020 Jun 8;22 (6):297–305.

66. Vardhan S, Sahoo SK. Exploring the therapeutic nature of limonoids and triterpenoids against SARS-CoV-2 by targeting nsp13, nsp14, and nsp15 through molecular docking and dynamics simulations. J Tradit Complement Med. 2022 Jan;12 (1):44–54.

67. Vivek-Ananth RP, Krishnaswamy S, Samal A. Potential phytochemical inhibitors of SARS-CoV-2 helicase Nsp13: a molecular docking and dynamic simulation study. Mol Divers. 2022 Feb;26 (1):429–42.

68. Vicenti I, Zazzi M, Saladini F. SARS-CoV-2 RNA-dependent RNA polymerase as a therapeutic target for COVID-19. Expert Opin Ther Pat. 2021 Apr;31 (4):325–37.

69. Anand K, Ziebuhr J, Wadhwani P, Mesters JR, Hilgenfeld R. Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. Science. 2003 Jun 13;300 (5626):1763–7.

70. Ratia K, Pegan S, Takayama J, Sleeman K, Coughlin M, Baliji S, et al. A noncovalent class of papain-like protease/deubiquitinase inhibitors blocks SARS virus replication. Proc Natl Acad Sci USA. 2008 Oct 21;105 (42):16119–24.

71. Barretto N, Jukneliene D, Ratia K, Chen Z, Mesecar AD, Baker SC. The papain-like protease of severe acute respiratory syndrome coronavirus has deubiquitinating activity. J Virol. 2005 Dec;79 (24):15189–98.

72. Zhang L, Lin D, Sun X, Curth U, Drosten C, Sauerhering L, et al. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. Science. 2020 Apr 24;368 (6489):409–12.

73. Jin Z, Du X, Xu Y, Deng Y, Liu M, Zhao Y, et al. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. Nature. 2020 Jun;582 (7811):289–93.

74. Xue X, Yang H, Shen W, Zhao Q, Li J, Yang K, et al. Production of authentic SARS-CoV M (pro) with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction. J Mol Biol. 2007 Feb 23;366 (3):965–75.

75. Funk L-M, Poschmann G, Chari A, von Pappenheim FR, Stegmann K-M, Dickmanns A, et al. Redox regulation of the SARS-CoV-2 main protease provides new opportunities for drug design. BioRxiv. 2022 Apr 19;

76. Schwarz KB. Oxidative stress during viral infection: a review. Free Radic Biol Med. 1996;21 (5):641–9.

77. Paulsen CE, Carroll KS. Cysteine-mediated redox signaling: chemistry, biology, and tools for discovery. Chem Rev. 2013 Jul 10;113 (7):4633–79.

78. Pavan M, Bolcato G, Bassani D, Sturlese M, Moro S. Supervised Molecular Dynamics (SuMD) Insights into the mechanism of action of SARS-CoV-2 main protease inhibitor PF-07321332. J Enzyme Inhib Med Chem. 2021 Dec;36 (1):1646–50.

79. Allen JD, Chawla H, Samsudin F, Zuzic L, Shivgan AT, Watanabe Y, et al. Site-Specific Steric Control of SARS-CoV-2 Spike Glycosylation. Biochemistry. 2021 Jul 13;60 (27):2153–69.

80. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. Nucleic Acids Res. 2000 Jan 1;28 (1):235–42.

81. Ke Z, Oton J, Qu K, Cortese M, Zila V, McKeane L, et al. Structures and distributions of SARS-CoV-2 spike proteins on intact virions. Nature. 2020 Dec;588 (7838):498–502.

82. Henderson R, Edwards RJ, Mansouri K, Janowska K, Stalls V, Gobeil SMC, et al. Controlling the SARS-CoV-2 spike glycoprotein conformation. Nat Struct Mol Biol. 2020 Oct;27 (10):925–33.

83. Turoňová B, Sikora M, Schürmann C, Hagen WJH, Welsch S, Blanc FEC, et al. In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. Science. 2020 Oct 9;370 (6513):203–8.

84. Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell. 2020 Apr 16;181 (2):281-292.e6.

85. Pipitò L, Rujan RM, Reynolds CA, Deganutti G. Molecular dynamics simulations reveal the importance of conformational changes, glycosylation, mutations, and drug repurposing for the Sars-Cov-2 spike protein. 2022 Mar 3;

86. Ching W-Y, Adhikari P, Jawad B, Podgornik R. Ultra-large-scale ab initio quantum chemical computation of bio-molecular systems: The case of spike protein of SARS-CoV-2 virus. Comput Struct Biotechnol J. 2021 Feb 19;19:1288–301.

87. Gur M, Taka E, Yilmaz SZ, Kilinc C, Aktas U, Golcuk M. Conformational transition of SARS-CoV-2 spike glycoprotein between its closed and open states. J Chem Phys. 2020 Aug 21;153 (7):075101.

88. Sztain T, Ahn S-H, Bogetti AT, Casalino L, Goldsmith JA, Seitz E, et al. A glycan gate controls opening of the SARS-CoV-2 spike protein. Nat Chem. 2021 Oct;13 (10):963–8.

89. Faheem, Kumar BK, Sekhar KVGC, Kunjiappan S, Jamalis J, Balaña-Fouce R, et al. Druggable targets of SARS-CoV-2 and treatment opportunities for COVID-19. Bioorg Chem. 2020 Nov;104:104269.

90. Wu C, Liu Y, Yang Y, Zhang P, Zhong W, Wang Y, et al. Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. Acta Pharm Sin B. 2020 May;10 (5):766–88.

91. Baum A, Fulton BO, Wloga E, Copin R, Pascal KE, Russo V, et al. Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. Science. 2020 Aug 21;369 (6506):1014–8.

92. Roth A, Lütke S, Meinberger D, Hermes G, Sengle G, Koch M, et al. LL-37 fights SARS-CoV-2: The Vitamin D-Inducible Peptide LL-37 Inhibits Binding of SARS-CoV-2 Spike Protein to its Cellular Receptor Angiotensin Converting Enzyme 2 *In Vitro*. BioRxiv. 2020 Dec 2;

93. Ramasamy MN, Minassian AM, Ewer KJ, Flaxman AL, Folegatti PM, Owens DR, et al. Safety and immunogenicity of ChAdOx1 nCoV-19 vaccine administered in a prime-boost regimen in young and old adults (COV002): a single-blind, randomised, controlled, phase 2/3 trial. Lancet. 2021 Dec 19;396 (10267):1979–93.

94. Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. N Engl J Med. 2020 Dec 31;383 (27):2603–15.

95. Chung JY, Thone MN, Kwon YJ. COVID-19 vaccines: The status and perspectives in delivery points of view. Adv Drug Deliv Rev. 2021 Mar;170:1–25.

96. Sikora M, von Bülow S, Blanc FEC, Gecht M, Covino R, Hummer G. Computational epitope map of SARS-CoV-2 spike protein. PLoS Comput Biol. 2021 Apr 1;17 (4):e1008790.

97. Lai S-C, Chong PC-S, Yeh C-T, Liu LS-J, Jan J-T, Chi H-Y, et al. Characterization of neutralizing monoclonal antibodies recognizing a 15-residues epitope on the spike protein HR2 region of severe acute respiratory syndrome coronavirus (SARS-CoV). J Biomed Sci. 2005 Oct;12 (5):711–27.

98. Walls AC, Tortorici MA, Frenz B, Snijder J, Li W, Rey FA, et al. Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. Nat Struct Mol Biol. 2016 Oct;23 (10):899–905.

99. Sharp PM, Hahn BH. Origins of HIV and the AIDS pandemic. Cold Spring Harb Perspect Med. 2011 Sep;1 (1):a006841.

100. Hahn BH, Shaw GM, De Cock KM, Sharp PM. AIDS as a zoonosis: scientific and public health implications. Science. 2000 Jan 28;287 (5453):607–14.

101. Fanales-Belasio E, Raimondo M, Suligoi B, Buttò S. HIV virology and pathogenetic mechanisms of infection: a brief overview. Ann Ist Super Sanita. 2010;46 (1):5–14.

102. Dwyer-Lindgren L, Cork MA, Sligar A, Steuben KM, Wilson KF, Provost NR, et al. Mapping HIV prevalence in sub-Saharan Africa between 2000 and 2017. Nature. 2019 Jun;570 (7760):189–93.

103. Peterlin BM, Luciw PA. Molecular biology of HIV. AIDS. 1988;2 Suppl 1:S29-40.

104. Strebel K. HIV accessory proteins versus host restriction factors. Curr Opin Virol. 2013 Dec;3 (6):692–9.

105. Swanson CM, Malim MH. SnapShot: HIV-1 proteins. Cell. 2008 May 16;133 (4):742, 742.e1.

106. Gonda MA. Molecular genetics and structure of the human immunodeficiency virus. J Electron Microsc Tech. 1988 Jan;8 (1):17–40.

107. Maertens GN, Engelman AN, Cherepanov P. Structure and function of retroviral integrase. Nat Rev Microbiol. 2022 Jan;20 (1):20–34.

108. Tabler CO, Wegman SJ, Chen J, Shroff H, Alhusaini N, Tilton JC. The HIV-1 Viral Protease Is Activated during Assembly and Budding Prior to Particle Release. J Virol. 2022 May 11;96 (9):e0219821.

109. Torrents de la Peña A, Del Moral Sánchez I, Burger JA, Bontjer I, Isik G, Eggink D, et al. Convergent HIV-1 Evolution upon Targeted Destabilization of the gp120-gp41

Interface. J Virol. 2021 Nov 23;95 (24):e0053221.

110. Kufareva I. Chemokines and their receptors: insights from molecular modeling and crystallography. Curr Opin Pharmacol. 2016 Oct;30:27–37.

111. Melikyan GB. HIV entry: a game of hide-and-fuse? Curr Opin Virol. 2014 Feb;4:1–7.

112. Kondo N, Melikyan GB. Intercellular adhesion molecule 1 promotes HIV-1 attachment but not fusion to target cells. PLoS ONE. 2012 Sep 6;7 (9):e44827.

113. Piller SC, Ewart GD, Premkumar A, Cox GB, Gage PW. Vpr protein of human immunodeficiency virus type 1 forms cation-selective channels in planar lipid bilayers. Proc Natl Acad Sci USA. 1996 Jan 9;93 (1):111–5.

114. Staudt RP, Alvarado JJ, Emert-Sedlak LA, Shi H, Shu ST, Wales TE, et al. Structure, function, and inhibitor targeting of HIV-1 Nef-effector kinase complexes. J Biol Chem. 2020 Oct 30;295 (44):15158–71.

115. Ahmad N, Venkatesan S. Nef protein of HIV-1 is a transcriptional repressor of HIV-1 LTR. Science. 1988 Sep 16;241 (4872):1481–5.

116. Jolicoeur P. The CD4C/HIV (Nef)transgenic model of AIDS. Curr HIV Res. 2011 Oct;9 (7):524–30.

117. Rhodes DI, Ashton L, Solomon A, Carr A, Cooper D, Kaldor J, et al. Characterization of three nef-defective human immunodeficiency virus type 1 strains associated with long-term nonprogression. Australian Long-Term Nonprogressor Study Group. J Virol. 2000 Nov;74 (22):10581–8.

118. Ariën KK, Verhasselt B. HIV Nef: role in pathogenesis and viral fitness. Curr HIV Res. 2008 May;6 (3):200–8.

119. Jamieson BD, Aldrovandi GM, Planelles V, Jowett JB, Gao L, Bloch LM, et al. Requirement of human immunodeficiency virus type 1 nef for in vivo replication and pathogenicity. J Virol. 1994 Jun;68 (6):3478–85.

120. Buffalo CZ, Iwamoto Y, Hurley JH, Ren X. How HIV nef proteins hijack membrane traffic to promote infection. J Virol. 2019 Dec 15;93 (24).

121. Morellet N, Bouaziz S, Petitjean P, Roques BP. NMR structure of the HIV-1 regulatory protein VPR. J Mol Biol. 2003 Mar 14;327 (1):215–27.

122. Sapp N, Burge N, Cox K, Prakash P, Balasubramaniam M, Thapa S, et al. HIV-1 Preintegration Complex Preferentially Integrates the Viral DNA into Nucleosomes Containing Trimethylated Histone 3-Lysine 36 Modification and Flanking Linker DNA. J Virol. 2022 Sep 28;96 (18):e0101122.

123. Kogan M, Rappaport J. HIV-1 accessory protein Vpr: relevance in the pathogenesis of HIV and potential for therapeutic intervention. Retrovirology. 2011 Apr 13;8:25.

124. Varin A, Decrion A-Z, Sabbah E, Quivy V, Sire J, Van Lint C, et al. Synthetic Vpr protein activates activator protein-1, c-Jun N-terminal kinase, and NF-kappaB and stimulates HIV-1 transcription in promonocytic cells and primary macrophages. J Biol Chem. 2005 Dec 30;280 (52):42557–67.

125. Levy DN, Refaeli Y, MacGregor RR, Weiner DB. Serum Vpr regulates productive infection and latency of human immunodeficiency virus type 1. Proc Natl Acad Sci USA. 1994 Nov 8;91 (23):10873–7.

126. Wheeler EDA, Achim CL, Ayyavoo V. Immunodetection of human immunodeficiency virus type 1 (HIV-1) Vpr in brain tissue of HIV-1 encephalitic patients. J Neurovirol. 2006 Jun;12 (3):200–10.

127. Arunagiri C, Macreadie I, Hewish D, Azad A. A C-terminal domain of HIV-1 accessory protein Vpr is involved in penetration, mitochondrial dysfunction and apoptosis of human CD4+ lymphocytes. Apoptosis. 1997;2 (1):69–76.

128. Yao S, Torres AM, Azad AA, Macreadie IG, Norton RS. Solution structure of peptides from HIV-1 Vpr protein that cause membrane permeabilization and growth arrest. Journal of Peptide Science. 1998 Nov 1;

129. Tompa P. The principle of conformational signaling. Chem Soc Rev. 2016 Jul 25;45 (15):4252–84.

130. Limongelli V. Ligand binding free energy and kinetics calculation in 2020. WIREs Comput Mol Sci. 2020 Jan 9;

131. Ninham BW. Long-range vs. short-range forces. The present state of play. J Phys Chem. 1980 Jun;84 (12):1423–30.

132. Chu W-T, Zheng Q-C. Conformational changes of enzymes and DNA in molecular dynamics: influenced by pH, temperature, and ligand. Adv Protein Chem Struct Biol. 2013;92:179–217.

133. Zhao Q. Molecular and thermodynamic mechanisms for protein adaptation. Eur Biophys J. 2022 Dec;51 (7–8):519–34.

134. Gibbs JW. On the equilibrium of heterogeneous substances. Am J Sci. 1878 Dec 1;s3-16 (96):441–58.

135. Chen L-Q. Chemical potential and Gibbs free energy. MRS Bull. 2019 Jul;44 (7):520–3.

136. Cooper A, Johnson CM. Introduction to microcalorimetry and biomolecular energetics. Methods Mol Biol. 1994;22:109–24.

137. Du X, Li Y, Xia Y-L, Ai S-M, Liang J, Sang P, et al. Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods. Int J Mol Sci. 2016 Jan 26;17 (2).

138. Chang C-E, Gilson MK. Free energy, entropy, and induced fit in host-guest recognition: calculations with the second-generation mining minima algorithm. J Am Chem Soc. 2004 Oct 13;126 (40):13156–64.

139. Deganutti G, Zhukov A, Deflorian F, Federico S, Spalluto G, Cooke RM, et al. Impact of protein-ligand solvation and desolvation on transition state thermodynamic properties of adenosine A2A ligand binding kinetics. In Silico Pharmacol. 2017 Nov 20;5 (1):16.

140. Schwesinger F, Ros R, Strunz T, Anselmetti D, Güntherodt HJ, Honegger A, et al. Unbinding forces of single antibody-antigen complexes correlate with their thermal dissociation rates. Proc Natl Acad Sci USA. 2000 Aug 29;97 (18):9972–7.

141. Wand AJ, Sharp KA. Measuring entropy in molecular recognition by proteins. Annu Rev Biophys. 2018 May 20;47:41–61.

142. Kendrew J, Bodo G, Dintzis H, Parrish R, Wyckoff H, Phillips D. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. Nature. 1958 Mar 8;181 (4610):662–6.

143. Eisenberg D. The discovery of the alpha-helix and beta-sheet, the principal structural features of proteins. Proc Natl Acad Sci USA. 2003 Sep 30;100 (20):11207–10.

144. Bork P, Koonin EV. Protein sequence motifs. Curr Opin Struct Biol. 1996 Jun;6 (3):366–76.

145. Bryngelson JD, Wolynes PG. Intermediates and barrier crossing in a random energy model (with applications to protein folding). J Phys Chem. 1989 Sep;93 (19):6902–15.

146. Ben-Naim A. Solvent effects on protein association and protein folding. Biopolymers. 1990 Feb 15;29 (3):567–96.

147. Pedersen KO. The development of svedberg's ultracentrifuge. Biophys Chem. 1976 Jul;5 (1–2):3–18.

148. Janin J, Bahadur RP, Chakrabarti P. Protein-protein interaction and quaternary structure. Q Rev Biophys. 2008 May;41 (2):133–80.

149. Koch I, Schäfer T. Protein super-secondary structure and quaternary structure

topology: theoretical description and application. Curr Opin Struct Biol. 2018 Jun;50:134–43.

150. Nassar R, Dignon GL, Razban RM, Dill KA. The protein folding problem: the role of theory. J Mol Biol. 2021 Oct 1;433 (20):167126.

151. Service RF. Problem solved* (*sort of). Science. 2008 Aug 8;321 (5890):784–6.

152. Zwanzig R, Szabo A, Bagchi B. Levinthal's paradox. Proc Natl Acad Sci USA. 1992 Jan 1;89 (1):20–2.

153. Lee J, Liwo A, Ripoll DR, Pillardy J, Scheraga HA. Calculation of protein conformation by global optimization of a potential energy function. Proteins: Structure, Function, and Bioinformatics. 1999 Jan 1;

154. Kobayashi Y. Statistical mechanics of protein structural transitions: Insights from the island model. Biophys Physicobiol. 2016 Nov 18;13:251–62.

155. Yakubovich AV. Degrees of freedom in polypeptides and proteins. Theory of phase transitions in polypeptides and proteins. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. p. 27–54.

156. Ghosh DK, Ranjan A. The metastable states of proteins. Protein Sci. 2020 Jul;29 (7):1559–68.

157. Gruebele M. Protein folding: the free energy surface. Curr Opin Struct Biol. 2002 Apr;12 (2):161–8.

158. Scheraga HA. The Multiple-Minima Problem in Protein Folding. AIP Conference Proceedings. AIP; 1991. p. 97–107.

159. Hartl FU, Hayer-Hartl M. Converging concepts of protein folding in vitro and in vivo. Nat Struct Mol Biol. 2009 Jun;16 (6):574–81.

160. Stefani M. Protein folding and misfolding on surfaces. Int J Mol Sci. 2008 Dec 9;9 (12):2515–42.

161. Perutz MF, Rossmann MG, Cullis AF, Muirhead H, Will G, North AC. Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis. Nature. 1960 Feb 13;185 (4711):416–22.

162. Lumry R, Eyring H. Conformation changes of proteins. J Phys Chem. 1954 Feb;58 (2):110–20.

163. Wess J, Han S-J, Kim S-K, Jacobson KA, Li JH. Conformational changes involved in G-protein-coupled-receptor activation. Trends Pharmacol Sci. 2008 Dec;29 (12):616–25.

164. Fraser-Pitt D, O'Neil D. Cystic fibrosis - a multiorgan protein misfolding disease. Future Science OA. 2015 Sep 1;1 (2):FSO57.

165. Vankeerberghen A, Cuppens H, Cassiman J-J. The cystic fibrosis transmembrane conductance regulator: an intriguing protein with pleiotropic functions. J Cyst Fibros. 2002 Mar;1 (1):13–29.

166. Kandori H, Shichida Y, Yoshizawa T. Photoisomerization in rhodopsin. Biochemistry Mosc. 2001 Nov;66 (11):1197–209.

167. Hu Z, Wang Y, Cheng C, He Y. Structural basis of the pH-dependent conformational change of the N-terminal region of human mannose receptor/CD206. J Struct Biol. 2019 Dec 1;208 (3):107384.

168. Nwamba CO, Ibrahim K. The role of protein conformational switches in pharmacology: its implications in metabolic reprogramming and protein evolution. Cell Biochem Biophys. 2014 Apr;68 (3):455–62.

169. Vajda S, Beglov D, Wakefield AE, Egbert M, Whitty A. Cryptic binding sites on proteins: definition, detection, and druggability. Curr Opin Chem Biol. 2018 Jun;44:1–8.

170. Callis PR. Exploring the Electrostatic Landscape of Proteins with Tryptophan Fluorescence. Reviews in fluorescence 2007. New York, NY: Springer New York; 2009. p. 199–248.

171. Fischer E. Einfluss der Configuration auf die Wirkung der Enzyme. Ber Dtsch Chem Ges. 1894 Oct;27 (3):2985–93.

172. Koshland DE. Application of a Theory of Enzyme Specificity to Protein Synthesis. Proc Natl Acad Sci USA. 1958 Feb;44 (2):98–104.

173. Csermely P, Palotai R, Nussinov R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. Trends Biochem Sci. 2010 Oct;35 (10):539–46.

174. Parker MJ, Spencer J, Clarke AR. An integrated kinetic analysis of intermediates and transition states in protein folding reactions. J Mol Biol. 1995 Nov 10;253 (5):771–86.

175. Dill KA, Chan HS. From Levinthal to pathways to funnels. Nat Struct Biol. 1997 Jan;4 (1):10–9.

176. Frauenfelder H, Sligar SG, Wolynes PG. The energy landscapes and motions of proteins. Science. 1991 Dec 13;254 (5038):1598–603.

177. Henzler-Wildman K, Kern D. Dynamic personalities of proteins. Nature. 2007 Dec

13;450 (7172):964–72.

178. Copeland RA, Pompliano DL, Meek TD. Drug-target residence time and its implications for lead optimization. Nat Rev Drug Discov. 2006 Sep 1;5 (9):730–9.

179. Zhang R, Monsma F. The importance of drug-target residence time. Curr Opin Drug Discov Devel. 2009 Jul 1;12 (4):488–96.

180. Pollard TD, De La Cruz EM. Take advantage of time in your experiments: a guide to simple, informative kinetics assays. Mol Biol Cell. 2013 Apr;24 (8):1103–10.

181. Held M, Noé F. Calculating kinetics and pathways of protein-ligand association. Eur J Cell Biol. 2012 Apr;91 (4):357–64.

182. Laidler KJ, King MC. Development of transition-state theory. J Phys Chem. 1983 Jul;87 (15):2657–64.

183. Teramoto H, Toda M, Komatsuzaki T. Understandings of chemical reaction dynamics in terms of dynamical systems theory. AIP Publishing LLC; 2015. p. 090042.

184. Gabdoulline RR, Wade RC. Simulation of the diffusional association of barnase and barstar. Biophys J. 1997 May;72 (5):1917–29.

185. Wereszczynski J, McCammon JA. Statistical mechanics and molecular dynamics in evaluating thermodynamic properties of biomolecular recognition. Q Rev Biophys. 2012 Feb;45 (1):1–25.

186. Onufriev AV, Alexov E. Protonation and pK changes in protein-ligand binding. Q Rev Biophys. 2013 May;46 (2):181–209.

187. Pang X, Zhou H-X. Rate Constants and Mechanisms of Protein-Ligand Binding. Annu Rev Biophys. 2017 May 22;46:105–30.

188. Raabe G, Sadus RJ. Molecular dynamics simulation of the dielectric constant of water: the effect of bond flexibility. J Chem Phys. 2011 Jun 21;134 (23):234501.

189. Sheinerman FB, Norel R, Honig B. Electrostatic aspects of protein-protein interactions. Curr Opin Struct Biol. 2000 Apr;10 (2):153–9.

190. Nie C, Sahoo AK, Netz RR, Herrmann A, Ballauff M, Haag R. Charge matters: mutations in omicron variant favor binding to cells. Chembiochem. 2022 Mar 18;23 (6):e202100681.

191. Pipitò L, Reynolds CA, Mobarec JC, Vickery O, Deganutti G. A Pathway Model to Understand the Evolution of Spike Protein Binding to ACE2 in SARS-CoV-2 Variants. Biomolecules. 2022 Oct 31;12 (11).

192. Pattanayak SK, Chowdhuri S. Effect of water on solvation structure and dynamics of ions in the peptide bond environment: importance of hydrogen bonding and dynamics of the solvents. J Phys Chem B. 2011 Nov 17;115 (45):13241–52.

193. Bondar A-N. Graphs of Hydrogen-Bond Networks to Dissect Protein Conformational Dynamics. J Phys Chem B. 2022 Jun 9;126 (22):3973–84.

194. Gerlt JA, Kreevoy MM, Cleland W, Frey PA. Understanding enzymic catalysis: the importance of short, strong hydrogen bonds. Chem Biol. 1997 Apr;4 (4):259–67.

195. Kenny PW. Hydrogen-Bond Donors in Drug Design. J Med Chem. 2022 Nov 10;65 (21):14261–75.

196. Kollman PA, Allen LC. Theory of the hydrogen bond. Chem Rev. 1972 Jun;72 (3):283–303.

197. Adhikari A, Park W-W, Kwon O-H. Hydrogen-Bond Dynamics and Energetics of Biological Water. ChemPlusChem. 2020 Dec;85 (12):2657–65.

198. Cornell WD, Cieplak P, Bayly CI, Kollmann PA. Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. J Am Chem Soc. 1993 Oct;115 (21):9620–31.

199. Herschlag D, Pinney MM. Hydrogen Bonds: Simple after All? Biochemistry. 2018 Jun 19;57 (24):3338–52.

200. Ladbury JE. Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. Chem Biol. 1996 Dec;3 (12):973–80.

201. Dunitz JD. The entropic cost of bound water in crystals and biomolecules. Science. 1994 Apr 29;264 (5159):670.

202. Dunitz JD. Win some, lose some: enthalpy-entropy compensation in weak intermolecular interactions. Chem Biol. 1995 Nov;2 (11):709–12.

203. Congreve M, Andrews SP, Doré AS, Hollenstein K, Hurrell E, Langmead CJ, et al. Discovery of 1,2,4-triazine derivatives as adenosine A (2A) antagonists using structure based drug design. J Med Chem. 2012 Mar 8;55 (5):1898–903.

204. Ross GA, Morris GM, Biggin PC. Rapid and accurate prediction and scoring of water molecules in protein binding sites. PLoS ONE. 2012 Mar 1;7 (3):e32036.

205. Cappel D, Sherman W, Beuming T. Calculating Water Thermodynamics in the Binding Site of Proteins - Applications of WaterMap to Drug Discovery. Curr Top Med Chem. 2017;17 (23):2586–98.

206. Abel R, Young T, Farid R, Berne BJ, Friesner RA. Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. J Am Chem Soc. 2008 Mar 5;130 (9):2817–31.

207. Boyd NK, Teng C, Frei CR. Brief overview of approaches and challenges in new antibiotic development: A focus on drug repurposing. Front Cell Infect Microbiol. 2021 May 17;11:684515.

208. Hodos RA, Kidd BA, Shameer K, Readhead BP, Dudley JT. In silico methods for drug repurposing and pharmacology. Wiley Interdiscip Rev Syst Biol Med. 2016 May;8 (3):186–210.

209. Hilal-Dandan R editor, Brunton LL editor, Goodman LS. Goodman and Gilman's manual of pharmacology and therapeutics. Second edition / editors, Randa Hilal-Dandan, Laurence L. Brunton. 2019.

210. Galiè N, Ghofrani HA, Torbicki A, Barst RJ, Rubin LJ, Badesch D, et al. Sildenafil citrate therapy for pulmonary arterial hypertension. N Engl J Med. 2005 Nov 17;353 (20):2148–57.

211. Mortier J, Rakers C, Bermudez M, Murgueitio MS, Riniker S, Wolber G. The impact of molecular dynamics on drug design: applications for the characterization of ligand-macromolecule complexes. Drug Discov Today. 2015 Jun;20 (6):686–702.

212. Yu W, MacKerell AD. Computer-Aided Drug Design Methods. Methods Mol Biol. 2017;1520:85–106.

213. Leelananda SP, Lindert S. Computational methods in drug discovery. Beilstein J Org Chem. 2016 Dec 12;12:2694–718.

214. Baig MH, Ahmad K, Roy S, Ashraf JM, Adil M, Siddiqui MH, et al. Computer aided drug design: success and limitations. Curr Pharm Des. 2016;22 (5):572–81.

215. Dar KB, Bhat AH, Amin S, Hamid R, Anees S, Anjum S, et al. Modern computational strategies for designing drugs to curb human diseases: A prospect. Curr Top Med Chem. 2018;18 (31):2702–19.

216. Wang J. Comprehensive assessment of ADMET risks in drug discovery. Curr Pharm Des. 2009;15 (19):2195–219.

217. Chandrasekaran B, Abed SN, Al-Attraqchi O, Kuche K, Tekade RK. Computer-Aided Prediction of Pharmacokinetic (ADMET) Properties. Dosage form design parameters. Elsevier; 2018. p. 731–55.

218. Green ED, Watson JD, Collins FS. Human Genome Project: Twenty-five years of big

biology. Nature. 2015 Oct 1;526 (7571):29–31.

219. Anderson AC. Structure-based functional design of drugs: from target to lead compound. Methods Mol Biol. 2012;823:359–66.

220. Shi Y. A glimpse of structural biology through X-ray crystallography. Cell. 2014 Nov 20;159 (5):995–1014.

221. Wang H-W, Wang J-W. How cryo-electron microscopy and X-ray crystallography complement each other. Protein Sci. 2017 Jan;26 (1):32–9.

222. Pellecchia M, Sem DS, Wüthrich K. NMR in drug discovery. Nat Rev Drug Discov. 2002 Mar;1 (3):211–9.

223. Danev R, Yanagisawa H, Kikkawa M. Cryo-Electron Microscopy Methodology: Current Aspects and Future Directions. Trends Biochem Sci. 2019 Oct;44 (10):837–48.

224. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021 Aug;596 (7873):583–9.

225. Anderson AC. The process of structure-based drug design. Chem Biol. 2003 Sep;10 (9):787–97.

226. Sliwoski G, Kothiwale S, Meiler J, Lowe EW. Computational methods in drug discovery. Pharmacol Rev. 2014;66 (1):334–95.

227. Mayne CG, Saam J, Schulten K, Tajkhorshid E, Gumbart JC. Rapid parameterization of small molecules using the Force Field Toolkit. J Comput Chem. 2013 Dec 15;34 (32):2757–70.

228. Gendenshteĭn, L. É. Derivation of exact spectra of the Schrödinger equation by means of supersymmetry. Soviet Journal of Experimental and Theoretical Physics Letters. 1983 Sep;

229. Walton JR, Rivera-Rivera LA, Lucchese RR, Bevan JW. Canonical Force Distributions in Pairwise Interatomic Interactions from the Perspective of the Hellmann-Feynman Theorem. J Phys Chem A. 2016 May 26;120 (20):3718–25.

230. Szabo A, Ostlund NS. Modern quantum chemistry: introduction to advanced electronic structure theory. Courier Corporation; 2012.

231. Hagler AT. Force field development phase II: Relaxation of physics-based criteria… or inclusion of more rigorous physics into the representation of molecular energetics. J Comput Aided Mol Des. 2019 Feb;33 (2):205–64.

232. Cavasotto CN, Adler NS, Aucar MG. Quantum Chemical Approaches in Structure-Based Virtual Screening and Lead Optimization. Front Chem. 2018 May 29;6:188.

233. McQuarrie DA. Quantum Chemistry. 2nd ed. Sausalito, Calif: University Science Books; 2007.

234. Arvind, Dorai K, Chaturvedi S, Mukunda N. The development of quantum mechanics. Reson. 2018 Oct;23 (10):1077–100.

235. Rosenbrock HH. A variational principle for quantum mechanics. Physics Letters A. 1985 Aug;110 (7–8):343–6.

236. Faye J. Copenhagen interpretation of quantum mechanics. 2002;

237. Weimann S, Perez-Leija A, Lebugle M, Keil R, Tichy M, Gräfe M, et al. Implementation of quantum and classical discrete fractional Fourier transforms. Nat Commun. 2016 Mar 23;7:11027.

238. Hill JG. Gaussian basis sets for molecular applications. Int J Quantum Chem. 2013 Jan 5;113 (1):21–34.

239. Merz KM. Using quantum mechanical approaches to study biological systems. Acc Chem Res. 2014 Sep 16;47 (9):2804–11.

240. Liu J, He X. QM implementation in drug design: does it really help? Methods Mol Biol. 2020;2114:19–35.

241. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J Comput Chem. 2010 Mar;31 (4):671–90.

242. Piquemal J-P, Jordan KD. Preface: special topic: from quantum mechanics to force fields. J Chem Phys. 2017 Oct 28;147 (16):161401.

243. Huang J, MacKerell AD. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. J Comput Chem. 2013 Sep 30;34 (25):2135–45.

244. Swope WC, Pitera JW, Suits F. Describing protein folding kinetics by molecular dynamics simulations. 1. theory†. J Phys Chem B. 2004 May;108 (21):6571–81.

245. Guvench O, MacKerell AD. Comparison of protein force fields for molecular dynamics simulations. Methods Mol Biol. 2008;443:63–88.

246. Siu SWI, Pluhackova K, Böckmann RA. Optimization of the OPLS-AA Force Field for Long Hydrocarbons. J Chem Theory Comput. 2012 Apr 10;8 (4):1459–70.

247. Aduri R, Psciuk BT, Saro P, Taniga H, Schlegel HB, SantaLucia J. AMBER Force Field Parameters for the Naturally Occurring Modified Nucleosides in RNA. J Chem Theory Comput. 2007 Jul;3 (4):1464–75.

248. Patel S, Brooks CL. CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. J Comput Chem. 2004 Jan 15;25 (1):1–15.

249. Mackerell AD. Empirical force fields for biological macromolecules: overview and issues. J Comput Chem. 2004 Oct;25 (13):1584–604.

250. Holstein BR. The van der Waals interaction. Am J Phys. 2001 Apr 1;69 (4):441–9.

251. Tee LS, Gotoh S, Stewart WE. Molecular Parameters for Normal Fluids. Lennard-Jones 12-6 Potential. Ind Eng Chem Fund. 1966 Aug;5 (3):356–63.

252. Bohm D, Pines D. A Collective Description of Electron Interactions: III. Coulomb Interactions in a Degenerate Electron Gas. Phys Rev. 1953 Nov 1;92 (3):609–25.

253. Wang X, Ramírez-Hinestrosa S, Dobnikar J, Frenkel D. The Lennard-Jones potential: when (not) to use it. Phys Chem Chem Phys. 2020 May 21;22 (19):10624–33.

254. Tannor DJ, Marten B, Murphy R, Friesner RA, Sitkoff D, Nicholls A, et al. Accurate First Principles Calculation of Molecular Charge Distributions and Solvation Energies from Ab Initio Quantum Mechanics and Continuum Dielectric Theory. J Am Chem Soc. 1994 Dec;116 (26):11875–82.

255. Buck M, Bouguet-Bonnet S, Pastor RW, MacKerell AD. Importance of the CMAP correction to the CHARMM22 protein force field: dynamics of hen lysozyme. Biophys J. 2006 Feb 15;90 (4):L36-8.

256. Nakamoto K. The urey — bradley force field: its significance and application. In: Forrette JE, Lanterman E, editors. Developments in applied spectroscopy. Boston, MA: Springer US; 1964. p. 158–68.

257. Liu X, Shi D, Zhou S, Liu H, Liu H, Yao X. Molecular dynamics simulations and novel drug discovery. Expert Opin Drug Discov. 2018 Jan;13 (1):23–37.

258. Doerr S, Harvey MJ, Noé F, De Fabritiis G. HTMD: High-Throughput Molecular Dynamics for Molecular Discovery. J Chem Theory Comput. 2016 Apr 12;12 (4):1845–52.

259. Dias R, de Azevedo WF. Molecular docking algorithms. Curr Drug Targets. 2008 Dec;9 (12):1040–7.

260. Lie MA, Thomsen R, Pedersen CNS, Schiøtt B, Christensen MH. Molecular docking

with ligand attached water molecules. J Chem Inf Model. 2011 Apr 25;51 (4):909–17.

261. Salha D, Andaç M, Denizli A. Molecular docking of metal ion immobilized ligands to proteins in affinity chromatography. J Mol Recognit. 2021 Feb;34 (2):e2875.

262. Totrov M, Abagyan R. Flexible ligand docking to multiple receptor conformations: a practical alternative. Curr Opin Struct Biol. 2008 Apr;18 (2):178–84.

263. Jokinen EM, Gopinath K, Kurkinen ST, Pentikainen OT. Detection of Binding Sites on SARS-CoV-2 Spike Protein Receptor-Binding Domain by Molecular Dynamics Simulations in Mixed Solvents. IEEE/ACM Trans Comput Biol Bioinform. 2021 Aug 6;18 (4):1281–9.

264. Hu G, Wang J. Ligand selectivity of estrogen receptors by a molecular dynamics study. Eur J Med Chem. 2014 Mar 3;74:726–35.

265. Arfeen M, Patel R, Khan T, Bharatam PV. Molecular dynamics simulation studies of GSK-3β ATP competitive inhibitors: understanding the factors contributing to selectivity. J Biomol Struct Dyn. 2015 Nov 27;33 (12):2578–93.

266. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Tunyasuvunakool K, et al. AlphaFold 2. Fourteenth Critical Assessment of Techniques for Protein Structure Prediction; DeepMind: London, UK. 2020;

267. Skolnick J, Gao M, Zhou H, Singh S. Alphafold 2: why it works and its implications for understanding the relationships of protein sequence, structure, and function. J Chem Inf Model. 2021 Oct 25;61 (10):4827–31.

268. Huey R, Morris GM, Olson AJ, Goodsell DS. A semiempirical free energy force field with charge-based desolvation. J Comput Chem. 2007 Apr 30;28 (6):1145–52.

269. Agu PC, Afiukwa CA, Orji OU, Ezeh EM, Ofoke IH, Ogbu CO, et al. Molecular docking as a tool for the discovery of molecular targets of nutraceuticals in diseases management. Sci Rep. 2023 Aug 17;13 (1):13398.

270. Eberhardt J, Santos-Martins D, Tillack AF, Forli S. Autodock vina 1.2.0: new docking methods, expanded force field, and python bindings. J Chem Inf Model. 2021 Aug 23;61 (8):3891–8.

271. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010 Jan 30;31 (2):455–61.

272. Hassan NM, Alhossary AA, Mu Y, Kwoh C-K. Protein-Ligand Blind Docking Using QuickVina-W With Inter-Process Spatio-Temporal Integration. Sci Rep. 2017 Nov

13;7 (1):15451.

273. Boobbyer DN, Goodford PJ, McWhinnie PM, Wade RC. New hydrogen-bond potentials for use in determining energetically favorable binding sites on molecules of known structure. J Med Chem. 1989 May;32 (5):1083–94.

274. Stouten PFW, Frömmel C, Nakamura H, Sander C. An effective solvation term based on atomic occupancies for use in protein simulations. Mol Simul. 1993 Aug;10 (2–6):97–120.

275. Santos LHS, Ferreira RS, Caffarena ER. Integrating molecular docking and molecular dynamics simulations. Methods Mol Biol. 2019;2053:13–34.

276. Uhlenbeck GE, Ornstein LS. On the Theory of the Brownian Motion. Phys Rev. 1930 Sep 1;36 (5):823–41.

277. Lavenda BH. Brownian Motion. Sci Am. 1985 Feb;252 (2):70–85.

278. Klafter J, Shlesinger M, Zumofen G. Fractal generalizations of Brownian motion have proven to be a rich field in probability theory, statistical physics and chaotic dynamics.

279. Karatzas I, Shreve SE. Brownian motion and stochastic calculus. New York, NY: Springer New York; 1998.

280. Greiner W. Classical mechanics: systems of particles and hamiltonian dynamics. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010.

281. Leimkuhler BJ, Reich S, Skeel RD. Integration methods for molecular dynamics. In: Mesirov JP, Schulten K, Sumners DW, editors. Mathematical approaches to biomolecular structure and dynamics. New York, NY: Springer New York; 1996. p. 161–85.

282. Computational Biochemistry and Biophysics - 1st Edition - Oren M. Bec [Internet]. [cited 2023 May 31]. Available from: https://www.routledge.com/Computational-Biochemistry-and-Biophysics/Becker-Jr-Roux-Watanabe/p/book/9780367397579

283. Rowlinson * JS. The Maxwell–Boltzmann distribution. Mol Phys. 2005 Nov 10;103 (21–23):2821–8.

284. Alder BJ, Wainwright TE. Phase transition for a hard sphere system. J Chem Phys. 1957 Nov;27 (5):1208–9.

285. McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. Nature. 1977 Jun 16;267 (5612):585–90.

286. Hollingsworth SA, Dror RO. Molecular dynamics simulation for all. Neuron. 2018 Sep

19;99 (6):1129–43.

287. Bussi G, Tribello GA. Analyzing and Biasing Simulations with PLUMED. Methods Mol Biol. 2019;2022:529–78.

288. Rodriguez-Bussey IG, Doshi U, Hamelberg D. Enhanced molecular dynamics sampling of drug target conformations. Biopolymers. 2016 Jan;105 (1):35–42.

289. Bernardi RC, Melo MCR, Schulten K. Enhanced sampling techniques in molecular dynamics simulations of biological systems. Biochim Biophys Acta. 2015 May;1850 (5):872–7.

290. Huang D, Caflisch A. The free energy landscape of small molecule unbinding. PLoS Comput Biol. 2011 Feb 3;7 (2):e1002002.

291. Hug S. Classical molecular dynamics in a nutshell. Methods Mol Biol. 2013;924:127–52.

292. Meller J. Molecular Dynamics. In: John Wiley & Sons, Ltd, editor. eLS. Wiley; 2001.

293. Binder K, Horbach J, Kob W, Paul W, Varnik F. Molecular dynamics simulations. J Phys: Condens Matter. 2004 Feb 11;16 (5):S429–53.

294. Chen JC, Kim AS. Brownian Dynamics, Molecular Dynamics, and Monte Carlo modeling of colloidal systems. Adv Colloid Interface Sci. 2004 Dec 31;112 (1–3):159–73.

295. Lü J-T, Hu B-Z, Hedegård P, Brandbyge M. Semi-classical generalized Langevin equation for equilibrium and nonequilibrium molecular dynamics simulation. Prog Surf Sci. 2018 Sep;94 (1):21–40.

296. Barth E, Schlick T. Overcoming stability limitations in biomolecular dynamics. I. Combining force splitting via extrapolation with Langevin dynamics inLN. J Chem Phys. 1998 Aug;109 (5):1617–32.

297. Groenhof G. Introduction to QM/MM simulations. Methods Mol Biol. 2013;924:43–66.

298. Laasonen K. Ab initio molecular dynamics. Methods Mol Biol. 2013;924:29–42.

299. Kirchner B, di Dio PJ, Hutter J. Real-world predictions from ab initio molecular dynamics simulations. Top Curr Chem. 2012;307:109–53.

300. Tobias DJ, Martyna GJ, Klein ML. Molecular dynamics simulations of a protein in the canonical ensemble. J Phys Chem. 1993 Dec;97 (49):12959–66.

301. Cottrell D, Tupper PF. Energy drift in molecular dynamics simulations. Bit Numer Math. 2007 Sep 13;47 (3):507–23.

302. Hammonds KD, Heyes DM. Shadow Hamiltonian in classical NVE molecular dynamics simulations involving Coulomb interactions. J Chem Phys. 2021 May 7;154 (17):174102.

303. Cramer CJ, Bickelhaupt FM. Essentials of computational chemistry. ANGEWANDTE CHEMIE-INTERNATIONAL EDITION IN ENGLISH-. 2003;42 (4):381–381.

304. Andersen HC. Molecular dynamics simulations at constant pressure and/or temperature. J Chem Phys. 1980;72 (4):2384.

305. Åqvist J, Wennerström P, Nervall M, Bjelic S, Brandsdal BO. Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. Chem Phys Lett. 2004 Jan;384 (4–6):288–94.

306. Harvey MJ, Giupponi G, Fabritiis GD. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. J Chem Theory Comput. 2009 Jun 9;5 (6):1632–9.

307. Cho K, Joannopoulos JD. Ergodicity and dynamical properties of constant-temperature molecular dynamics. Phys Rev, A. 1992 May 15;45 (10):7089–97.

308. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys. 1984;81 (8):3684.

309. Martyna GJ, Tobias DJ, Klein ML. Constant pressure molecular dynamics algorithms. J Chem Phys. 1994;101 (5):4177.

310. Jayaram B. On the acceptance ratio in Monte Carlo computer simulations. J Math Chem. 1996;20 (2):395–7.

311. Becker O, MacKerell Jr A, Roux B, editors. Computational biochemistry and biophysics. CRC Press; 2001.

312. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. Annu Rev Phys Chem. 1997;48:545–600.

313. Shrivastav G, Vanden-Eijnden E, Abrams CF. Mapping saddles and minima on free energy surfaces using multiple climbing strings. J Chem Phys. 2019 Sep 28;151 (12):124112.

314. DE Kleiman, Nadeem H, Shukla D. Adaptive Sampling Methods for Molecular Dynamics in the Era of Machine Learning. arXiv preprint arXiv:230709664. 2023;

315. Yang YI, Shao Q, Zhang J, Yang L, Gao YQ. Enhanced sampling in molecular dynamics. J Chem Phys. 2019 Aug 21;151 (7):070902.

316. Deb I, Frank AT. Accelerating rare dissociative processes in biomolecules using

selectively scaled MD simulations. J Chem Theory Comput. 2019 Nov 12;15 (11):5817–28.

317. Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. J Chem Phys. 2004 Jun 22;120 (24):11919–29.

318. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett. 1999 Nov;314 (1–2):141–51.

319. Fukunishi H, Watanabe O, Takada S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. J Chem Phys. 2002;116 (20):9058.

320. Ensing B, De Vivo M, Liu Z, Moore P, Klein ML. Metadynamics as a tool for exploring free energy landscapes of chemical reactions. Acc Chem Res. 2006 Feb;39 (2):73–81.

321. Barducci A, Bonomi M, Parrinello M. Linking well-tempered metadynamics simulations with experiments. Biophys J. 2010 May 19;98 (9):L44-6.

322. Barducci A, Bussi G, Parrinello M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. Phys Rev Lett. 2008 Jan 18;100 (2):020603.

323. Perilla JR, Woolf TB. Towards the prediction of order parameters from molecular dynamics simulations in proteins. J Chem Phys. 2012 Apr 28;136 (16):164101.

324. Fiorin G, Klein ML, Hénin J. Using collective variables to drive molecular dynamics simulations. Mol Phys. 2013 Dec;111 (22–23):3345–62.

325. Laio A, Gervasio FL. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. Rep Prog Phys. 2008 Dec 1;71 (12):126601.

326. Sabbadin D, Salmaso V, Sturlese M, Moro S. Supervised molecular dynamics (sumd) approaches in drug design. Methods Mol Biol. 2018;1824:287–98.

327. Sabbadin D, Moro S. Supervised molecular dynamics (SuMD) as a helpful tool to depict GPCR-ligand recognition pathway in a nanosecond time scale. J Chem Inf Model. 2014 Feb 24;54 (2):372–6.

328. Deganutti G, Moro S, Reynolds CA. A Supervised Molecular Dynamics Approach to Unbiased Ligand-Protein Unbinding. J Chem Inf Model. 2020 Mar 23;60 (3):1804–17.

329. Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate

ligand-binding affinities. Expert Opin Drug Discov. 2015 May;10 (5):449–61.

330. Deganutti G, Pipito L, Rujan RM, Weizmann T, Griffin P, Ciancetta A, et al. Hidden GPCR structural transitions addressed by multiple walker supervised molecular dynamics (mwSuMD). BioRxiv. 2022 Oct 27;

331. Majumder S, Deganutti G, Pipitò L, Chaudhuri D, Datta J, Giri K. Computer-aided de novo design and optimization of novel potential inhibitors of HIV-1 Nef protein. Comput Biol Chem. 2023 Apr 15;104:107871.

332. Unoh Y, Uehara S, Nakahara K, Nobori H, Yamatsu Y, Yamamoto S, et al. Discovery of S-217622, a Noncovalent Oral SARS-CoV-2 3CL Protease Inhibitor Clinical Candidate for Treating COVID-19. J Med Chem. 2022 May 12;65 (9):6499–512.

6. SARS-CoV-2 Project Publications

The first period of the PhD project was focused on the collection of the available structural and biological data on SARS-CoV-2 proteins. These data included a set of available SARS-CoV-2 protein structures alone or in complex with antibodies, either isolated from convalescent patients and preliminary data on new or repurposed drugs. These data guided the initial steps of work, with particular attention on the role of heparin and heparan sulfate as infection modulators and cofactors.

6.1 Molecular Dynamics studies on the SARS-CoV-2 spike protein

I reviewed the available literature to summarise the information on the SARS-CoV-2 S protein for a broader understanding of the state of research concerning drug repurposing, and computational methods deployed for drug discovery. Since the first stage of the project revolved around MD, I focused the review on this method applied to SARS-CoV-2 drug discovery. The published results, as well as the deployed methods, structures, ligands, antibodies, and proteins analysed, were gathered from publicly available datasets and publications, unveiling structural details that inspired the subsequent works and resulted in the publication: “Molecular dynamics studies reveal structural and functional features of the SARS-CoV-2 spike protein” published on BioEssays, Volume 44, Issue 9, https://doi.org/10.1002/bies.202200060.

Authors: Ludovico Pipitò, Roxana-Maria Rujan, Christopher A. Reynolds, and Giuseppe Deganutti

Abstract

The SARS-CoV-2 virus is responsible for the COVID-19 pandemic the world experienced since 2019. The protein responsible for the first steps of cell invasion, the spike protein, has probably received the most attention in light of its central role during infection. Computational approaches are among the tools employed by the scientific community in the enormous effort to study this new burden. One of these methods, namely molecular dynamics (MD), has been used to characterize the function of the spike protein at the atomic level and unveil its structural features from a dynamic perspective. In this review, we focus on these main findings.

Introduction

The year 2019 signalled the start of the worldwide outbreak of Coronavirus Disease (COVID-19) (1–3) from the coronaviridae SARS-CoV-2 virus, which counts around 360 million cases around the world with more than 5.6 million certified deaths (WHO dashboard, 27 January 2022). Coronaviridae is an enveloped positive-stranded, non-segmented RNA virus with a genome of about 30 Kb (4). Coronaviridae viruses are responsible for cardiovascular, hepatic, respiratory, gastrointestinal, and neurological diseases, with major symptoms associated with a hyperbolic expression of proinflammatory signals and cytokines such as interleukins, interferon-gamma (IFN-γ), interferon-gamma induced protein 10 (IP-10), macrophage inflammatory protein 1A and 1B (MIP-1A, MIP1-B), platelet-derived growth factor (PDGF), tumour necrosis factor (TNF-α), and vascular endothelial growth factor (VEGF) (5).

The SARS-CoV-2 infection mechanism depends on the transmembrane spike protein (S protein, Figure 1a,d) (6,7), a highly conserved structure amongst the coronaviridae family responsible for extracellular binding and cell membrane fusion (8). It characterizes the shape of this family of viruses, giving it the “solar” crown aspect (9) they are named after. The SARS-CoV-2 strain shows a selective affinity for the angiotensin-converting enzyme 2 (ACE2, Figure 1a) receptor, a type 1 transmembrane protein with an external peptidase domain normally responsible for the conversion of angiotensin hormone into angiotensin II (10).

The S protein has aroused the interest of medical and pharmaceutical research, to prevent infection and reduce the burden of clinical intervention. It is a homotrimer class I fusion protein, with each protomer composed of domain S1 and S2 (in prefusion conformation, Figure 1a) (11). The S1 structure is responsible for binding ACE2 (Figure 1a-c), before the conformational change in the stalk-like structure (Figure 1a,d) of the S2 subunit (12) and the subsequent membrane fusion after the cleavage of S1 from S2 by the host transmembrane protease serine 2 (TMPRSS2) (9). The cleavage of the inter-region S1/S2 (Figure 1a,d) allows for S2 structural conformation changes necessary for membrane fusion and post-fusion structure adaptation (8). In the S1 ectodomain (Figure 1d), the apical portion of the S protein, composed of the receptor-binding domain (RBD), the N-terminal domains (NTDs), and two C-terminal domains (CTDs), folds in a hairpin motif that protects the prefusion conformation of S2 from the external environment (13). A distinguishing feature of the SARS-CoV-2 strain is an insertion in the protease S1/S2 cleavage site region, rich in arginine, which configures a furin recognition site, commonly found in highly virulent influenza viruses (14).

Several S protein structures have been determined through cryo-electron microscopy (cryo-EM) and X-ray crystallography (Figure 2) (15). These include the inner S1 and the

external S2 domains and indicate two different states in the RBD domain, named “up” and “down” (11,16–18), the former determining an active state (19) favourable to ACE2 binding (Figure 1a-c). For S1 to bind ACE2 (20), the RBD must undergo a conformational hinge movement, exposing the hydrophobic region between A570 -T572, F855-N856 at the interface between RBD and S2 in an “up” conformation state (17,21). The coronaviridae family has a distinctive morphology characterized by a spherical virion with a diameter of 91 ± 11 nm measured at the membrane, on whose surface there are 24 ± 9 S trimers unevenly distributed with a prevalence of 97% of trimers in "down" conformation (16) at room temperature. Although cryo-EM studied by Benton et al showed that only 11% of the total trimeric structures were fully closed, 20% were open state either with one RBD (16%) or two RBD up (4%) (13). The RBD is responsible for ACE2-specific binding through an ensemble of 16 well-conserved residues directly interacting with the receptor (22) (Figure 1a-c). Three different sites (Figure 1c), named according to which part of ACE2 they bind, can be distinguished. Site 1 (identified by residue Q498, T500, N501, Y505) and Site 3 (N487 and F486) bind to the α1 helix C (Q24 and T27), while Site 2 (R403, Y453, L455, F456, and Q493) binds to the centre of the helix (D30, K31, H34, D38) which is slightly bent outwards, exposing polar amino acids for interaction (23). The RBD is an important target for preventing or treating the SARS-CoV-2 infection (9,24). A common trait shared among the coronavirus family is the post-translational N- and O-glycosylation used to mask the S protein epitopes and escape from immune system recognition (25,26), covering approximately 40% of the surface protein, especially N343 which seems to hinder antibody binding. A recent cryo-EM-derived S protein model revealed that 44 out of 66 potential sites are heavily N-glycosylated in the ectodomain region (Figure 1a) (11).

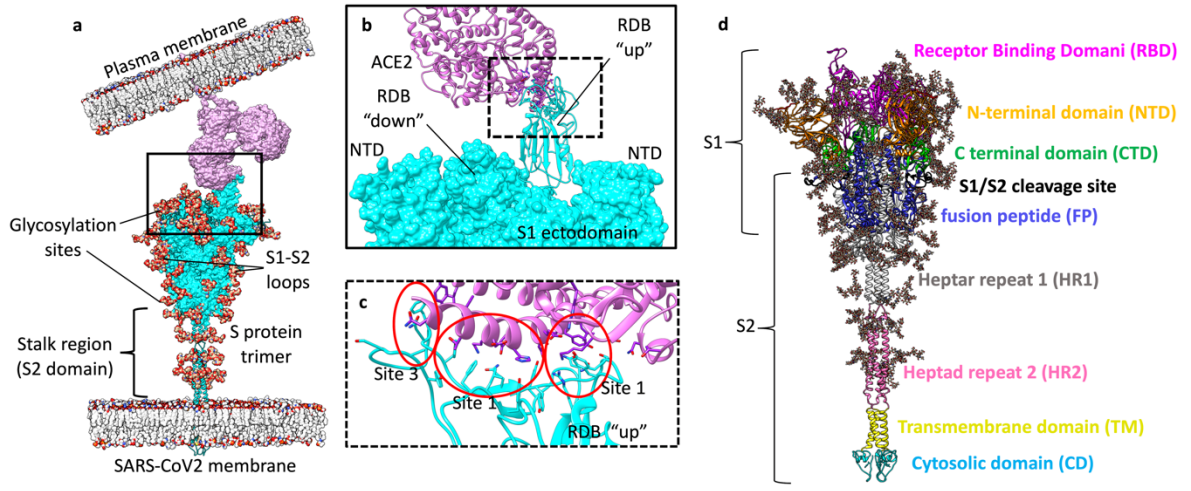


Figure 1. The S protein is the first promoter of SARS-CoV-2 internalization. a) Overall architecture of the complex between S protein (prefusion conformation, cyan) and ACE2 (violet); glycans on S protein are in van der Waals spheres; the relative positions of the plasma and viral membranes are reported. b) Magnification of the S1 ectodomain (glycans removed for clarity); the RBD in the up conformation is responsible for binding ACE2. c) Magnification of the interface between the RBD and one of the ACE2 monomers; the interactions can be divided according to the relative position into Sites 1 to 3 (red circles). d) Fully glycosylated S protein (https://charmm-gui.org/?doc=archive&lib=covid19) with the S1 and S2 units highlighted; B) The N-terminal domain (NTD), receptor-binding domain (RBD), C-terminal domain (CTD), S1/S2 cleavage site, fusion peptide (FP), heptad repeat 1 (HR1), heptad repeat 2 (HR2), transmembrane domain (TM), cytosolic domain (CD) are reported. Glycans are shown in licorice.

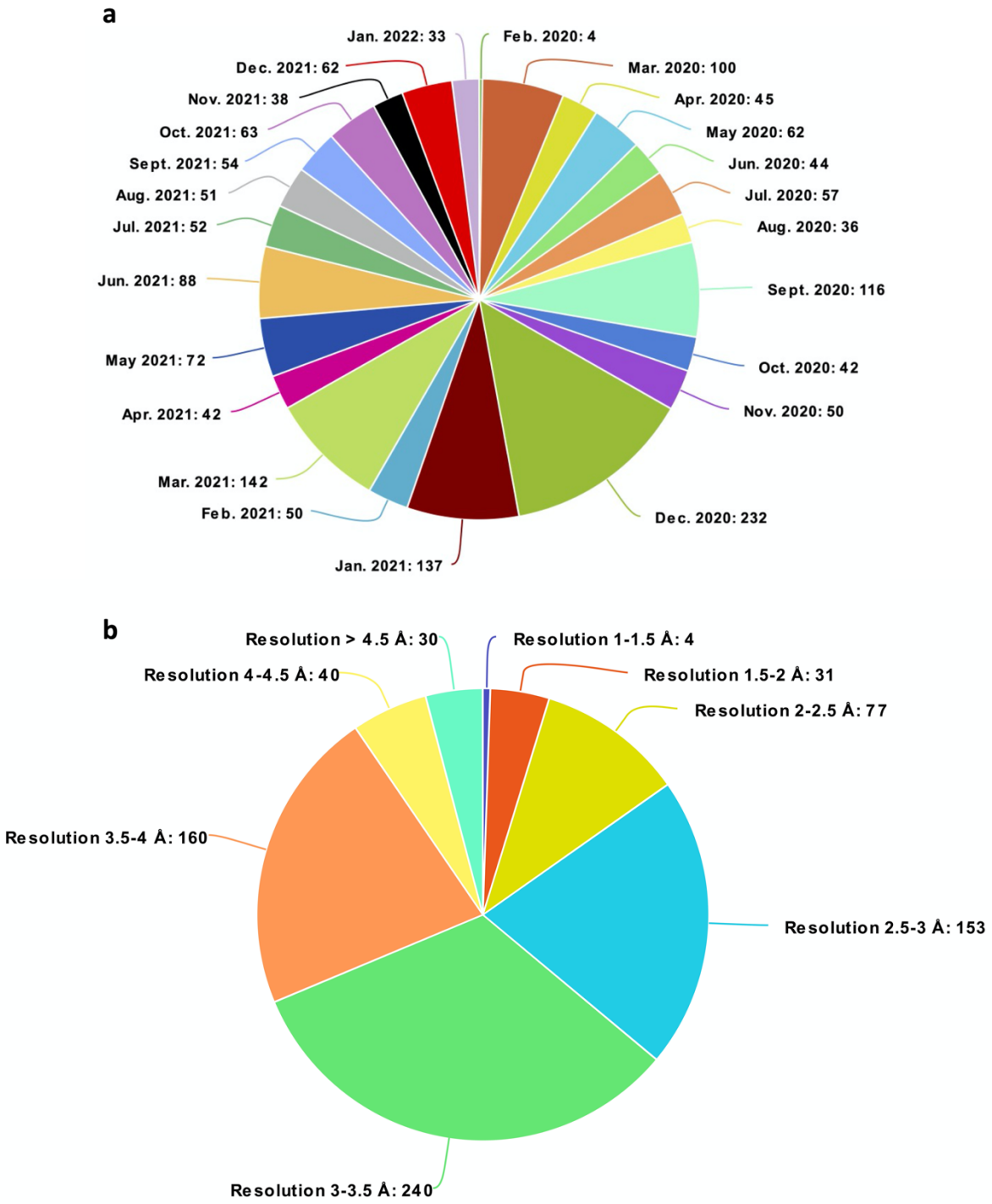


Figure 2. S protein deposited structures in the protein data bank. a) SARS-CoV-2 protein structures released and ordered according to month and year of publication from the Protein Data Bank (15) b) S protein structures only, ordered according to resolution.

Molecular dynamics (MD) is a computational technique that simulates the evolution over time of (bio)molecular structures. It represents a state-of-the-art tool for biophysical studies and structure-based drug design (23) as it describes the positional changes of the

atoms of a chemical system that explicitly includes water, ions, and other biological components such as protein, membrane, nucleic acids (27), allowing the conformational exploration of biological structures. The evolution over time is simulated by integrating Newton's classical equation of motion for each atom of the system. The result of this many-particle motion is a trajectory, from which it is possible to extrapolate thermodynamic, kinetics, and physical properties through statistical mechanics models (28). One of the advantages of MD is to overcome the unnatural rigidity that characterizes X-ray crystallography and cryo-EM structures, allowing the investigation of possible cryptic binding pockets, allosteric effects, and structural changes in response to the binding.

In this review, we recapitulate the MD studies that have expanded our knowledge of the SARS-CoV-2 S protein flexibility and antibody (Ab) recognition and summarise their contribution to drug repurposing campaigns.

Molecular dynamics simulations uncover the S protein flexibility

Since the first cryo-EM structures of the S protein became available to the scientific community, it has been possible to investigate the conformational changes and the dynamic processes involving the S protein through MD simulations. One of the limits of the S protein structure experimentally determined is the scarcity of structural information about the post-translational glycosylation, due to the high dependency on the organism used for protein expression. To address this aspect, Woo and his group proposed a set of complete and fully glycosylated (Figure 1) S protein models (29), corroborating Wrapp et al.'s structure of the spike protein glycosylated in 44 out of 66 possible sites in the ectodomain region (11) including also Watanabe et al. glycans' specifications (30).

A study by Turonova (18) as cited by Choi and co-workers (31), showed that the S1 domain displays structural compactness, while the stalk is characterized by two flexible portions, at the heptad-repeat 2 (HR2) linker and the heptad repeat transmembrane portion (HR2-TM) respectively. Such findings are in line with the experimental observation that the S protein can tilt up to 90° toward the membrane, with an inclination of 48° (to the membrane normal) being most likely to occur. Such movements might favour ACE2 binding (31) by scanning the surrounding space for a possible alignment with the receptor, while indirectly exposing cryptic epitopes (16). These findings, in context with the cryo-EM results (18,32), highlighted the importance of S2 flexibility, which plays a crucial role in conformational changes, (33), alignment, and membrane fusion process (7,16). However, the limitations of the timescale in MD simulations restrict the exploration of long-lasting contacts between the protein and the receptor, thus limiting our ability to characterize the interaction network that contributes to the binding process.

The sequence of conformational changes on the S2 domain, necessary for membrane fusion, is triggered at the S1/S2 cleavage site on residues P681-R684 (34) and facilitates the exposure of the FP (35). However, the experimental determination of these conformational changes is difficult due to the rapid timescale involved. A computational attempt was made by Remington et al (36), through the use of nontargeted parallel cascade selection MD (nt-PaCS-MD). The variational approach to Markov processes (VAMP) analysis indicated distinct conformational changes in cleaved SARS-CoV-2-spike models at the level of residues V705–D1146 and residues S816 –D1146 (36). These changes seemed necessary to expose the fusion peptide (FP) and rearrange the region between residues I818 – V826 of the FP into an outward-facing helical structure which might mechanically initiate membrane fusion, indicating the crucial role of the S1/S2 cleavage site in facilitating the fusion mechanism. These findings suggest that drug discovery could target the S1/S2 cleavage site to hamper an efficient exposition of the FP, therefore interfering with the membrane fusion mechanism.

The differences between SARS-CoV and SARS-CoV-2 were addressed to understand the reasons behind SARS-CoV-2's high infectivity and the molecular mechanisms required for effective therapy development. Furthermore, due to the presence of multiple mutations that differentiated SARS-CoV-2 and its variants from the original SARS-CoV, the molecular investigation of residues and conformational differences became necessary for a prompt pharmaceutical response. MD studies of SARS-CoV-2 have indicated accentuated flexibility compared to its predecessor SARS-CoV in segments of the RBD implicated in the molecular recognition of ACE2, more precisely in the region comprising residues Q474–G485, C488–F490, and S494–Y505 of the RBD, which enhances binding to the ACE2 receptor (37) in B.1.351 and B.1.1.7 variants as well. It was possible, through MD simulations, to appraise the effect of mutations such as N501Tyr and E484Lys, which improved the binding to ACE2 by −4.5 and −1.3 kcal/mol respectively, as determined by free energy perturbation (FEP). The flexibility of RBD in the "up" conformation has been proposed as a determinant for the high propensity of SARS-CoV-2 to reach ACE2, giving rise to the high infectivity associated with SARS-CoV-2 (38) compared to SARS-CoV (37). MD investigations allowed a broader analysis of the interaction network between ACE2 and RBD, which was not observed in the static cryo-EM or X-ray crystal structures. MD simulations showed a large interaction network between residues I21, Q24, T27, F28, D30, E35, D38, A80, M82, and Y83 of ACE2 and RBD (39).

A study by Barros et al. (40) indicated that ACE2 presents great motility when in contact with S protein, suggesting that the rotation of the catalytic zinc-binding peptidase domain (PD) along the transmembrane domain axis could sterically accommodate multiple ACE2 bindings. This large shift appeared to be enhanced by five glycan residues bound to

N53, N90, N103, N322, and N546 of ACE2, with N53 involved in both intramolecular homodimer and heterodimer contacts (40,41). The results by Williams and co-workers (42), in conjunction with those by Barros, indicated that in the RBD-ACE2 interaction pattern, residues F486, N487, and Y489 are responsible for the adaptive flexibility of RBD in establishing strong interactions with ACE2. Taken together, these results describe the synergy between a strong ACE2-binding RBD which, once locked, is carried by the rotation of the ACE2 axis, allowing for multiple receptor engagement and a subsequential binding mechanism. At the same time, this study demonstrated how mutations in that RBD sub-region are crucial in the selective pressure of the virus, altering the flexibility of RBD and interfering in intra-monomer interactions within RBD (42). From a geometric perspective, effective interaction between SARS-CoV-2-spike and ACE2 would occur at an angle of inclination between the apical portion of RBD "up" and ACE2 of at least 52° (43). Such MD results indicated that RBD “up” conformations have a large degree of manoeuvre to achieve sufficient residue exposure for ACE2 binding.

Although MD is able to describe the dynamic events that lead to conformational changes and new interactions, predictions are still limited by the computational cost and represent a simplified scenario, where the complexity of a cellular microenvironment cannot be adequately represented.

Molecular dynamics simulations to explore rare S protein conformational changes

The activation of a protein occurs through a series of conformational changes driven by molecular interactions with the intended target. The exploration of metastable states is necessary to understand the intermediate steps occurring during molecular events, and, therefore, identify possible therapeutic targets to interfere with the functional pathway.

However, large protein conformational rearrangements usually take place in the millisecond or second timescale, far beyond the time simulated in MD, which is usually within tens of microseconds. In such a context, the implication of this is that rare conformational changes can be missed. It is, therefore, necessary to apply enhanced or adaptive sampling algorithms to overcome this intrinsic limitation of the sampling that can be achieved to explore drastic structural changes in proteins.

From this perspective, weighted ensemble (WE) MD allows the sampling of rare events (44), drastically increasing the computational efficiency. With WE, multiple simulations are run in parallel and the trajectories that explored new values of a metric decided a priori (a distance between atoms in the simplest case) are retained and replicated, thus minimizing the randomness of conformational exploration. By using the WE path-sampling strategy, Sztain et al. (45) were able to simulate the transition state of RBD from "down" to "up", uncovering the crucial role of several glycan residues in allosterically stabilizing the “up”

state. While N165 and N264 shield the RBM acting as an “up” state stabilizer (46,47), N343 pushes the RBD to the final “up” state interacting with residues F490, Y489, F456, and R457 on the interaction portion of the ACE2 binding motif (45). More recently, it has been suggested that glycans attached to N165 and N343 contribute to the overall stability of the RBD open conformation (46).

An approach combining WE and artificial intelligence (AI) was adopted by Casalino et al. (48) to evaluate transition conformations during the binding between fully glycosylated S protein and ACE2. This confirmed the role that the two N-glycan residues linked to N165 and N234 have in modulating the dynamics of the S protein's RBD, contributing to the axial mobility of ACE2 while triggering the opening of RBD in a “hand jive” motion. Yao and co-workers (49), analyzed the molecular architecture of SARS-CoV-2, from cryo-electron tomography (cryo-ET) and subtomogram averaging (STA) highlighting the complex composition of N-glycans, which is the result of unions between branched oligomannose and hybrids units. Such complex glycan ramification also appears to be present on N234, whose allosteric role in the conformational change of RBD from “down” to “up” has been demonstrated by Amaro et al. (48).

In a separate study, all-atom steered MD (SMD) forced the RBD from “down” to “up” and highlighted the conformational changes that occur during the breaking of the salt bridges between RBD and the neighbouring protomers with the hydrogen bonds that keep RBD in an inactive “down” state (50). These intramolecular salt bridges, K378-E988 and K386-D985 within the S2 domain, and E516-K202 within the NTD are mainly responsible for the inactive "down" state of the monomers and prevent the interactions with ACE2. Data obtained through targeted molecular dynamics (TMD) have shown how glycans on RBD residues N165, N234, and N343, can act as position locker for the active “up” conformation (51), stabilizing a set of interdomain salt bridges involving K417, R408, and K378. Furthermore, glycans on N165 and N234 were proposed as shielding the epitopes (Figure 3), while locking RBD in the “up” state (47).

The description of the RBD transition from the down to the up states is a nontrivial task, a determinant for understanding the protein activation and providing valuable information about cryptic binding pockets. A recent study by Dokainish et al, (52) described the opening of RBD by adopting the new generalized replica exchange with solute tempering of selected surface charged residues method (gREST_SSCR), an enhanced method derived from generalized replica exchange with solute tempering (gREST) (53). In this study a selection of charged residues at the RBD interface was regarded as the solute region for gREST, exploring a range of temperatures while the solvent was kept at a constant temperature. The results highlighted the important intra-chain interaction between residues R408 (chain A) and the proximal D406 (chain C) and the stabilizing role of glycans on N165

for the “down” state, while glycan on N343 and N234 supported the opening of the chain and the stabilization of the “up” state, respectively. Glycan on N343 interacts with Y489 and Q493 in the “up” state contributing to the structural stabilization, with the interdomain contribution of residues S477-T385, Q493-C379, Y489-T385, and Q493-K378 pairs (51).

A remarkable effort was made by Zimmerman et al (54), to explore drastic conformational changes through Markov-state models (MSMs) combined with the computational power provided by “citizen-scientists” from the “Folding@home” project (http://foldingathome.org), Very long time scale simulations, in conjunction with the FAST algorithm, described drastic conformational changes on the S protein which opened the RBD from its “down” to the “up” state, while the RBD domain twisted outward, exposing new cryptic epitopes.

Hide and seek: the hunt for epitopes through MD

Access to the S protein epitope(s) is necessary for an antibody’s binding (Figure 3) (55). From this standpoint, long MD simulations might unveil cryptic epitopes. Sikora et al. (55) performed an extensive simulation of four S proteins embedded in a membrane, for a total of 2.5 μs. The resulting trajectories were analyzed through simulated illumination analysis and rigid docking of the antibody CR3022. In the illumination analysis, randomly oriented rays emanated from a half-sphere at the centre of mass of the S protein. Rays are then absorbed by the first heavy atom they meet within 1.5 Å. Single S protein structures are collected every 10 ns. To evaluate the shielding effect of glycans, the analysis was replicated without their presence. The results indicated that glycans reduce the S protein accessibility by up to 80%, with the most marked effect occurring in the stalk region close to the viral membrane. An interesting ab initio epitope mapping method was used by Serapian et al (56). Possible epitopes were classified according to the coupling energy with the rest of the structure which identifies sites on the S protein surface that are at lower binding energy levels and could possibly energetically prefer bound states. This method suggested that residues forming an epitope prefer to form molecular interactions with external elements (56). The data agrees with the experimentally detected epitope recognized by antibodies CR3022, 4A8, S309, and EY6A.

The use of small molecules or cosolvents as probes in MD simulations (mixMD) increases the chance of discovering cryptic niches or epitopes on the surface of a protein. Through MixMD is possible to map interesting interaction sites by considering the frequency of contacts between protein and probe, resulting in a volumetric map. Using a concentration of around 1-5% cosolvent usually improves the sampling of hot spots for interactions without denaturing the protein. Pyrimidine (Py), acetonitrile, and isopropanol were used to discover possible docking niches on the interface between RBD and ACE2 and to inspire the drug

design of antagonists or antibodies (57). Py showed the most relevant volumetric maps within the RBM that span from residue Q498 to residue Y505. Knowledge of these types of interactions, in conjunction with the molecular mechanics with generalized Born surface area solvation (MM-GBSA) analysis, leads drug design processes and virtual screening in concordance with experimental data from antibodies, with the data suggesting a set of new molecules (DB02651, DB03714, DB08248, and DB14826) as possible RBD interaction modulators.

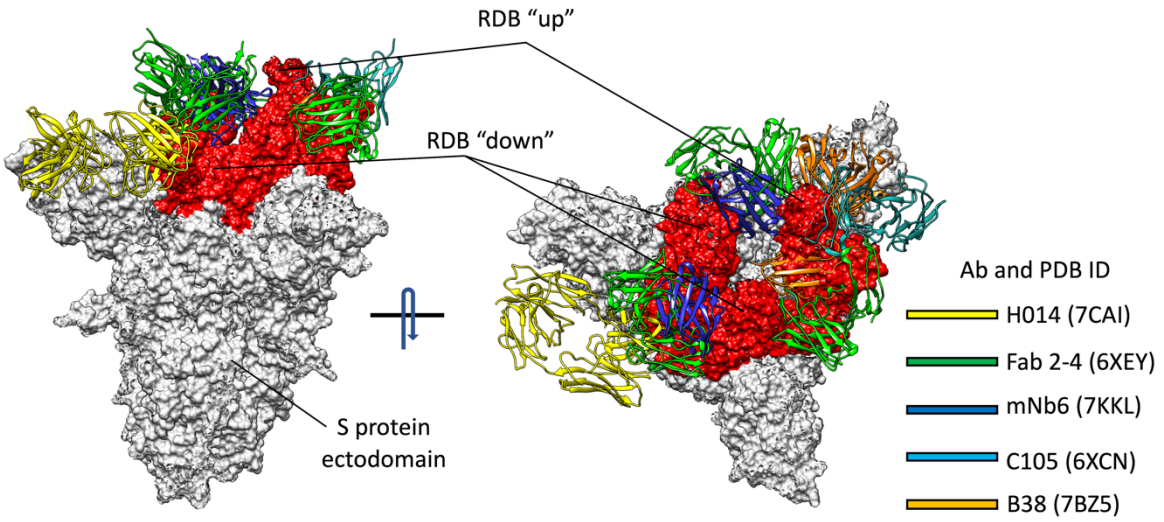


Figure 3. Human antibodies can bind to different S protein epitopes. The binding position of five human antibodies (Abs) on the surface of the S protein, is coloured according to the legend. Abs names and protein data bank database IDs are reported in the legend. The S protein is represented as a white surface, with the RBD in red.

Computer-aided drug repurposing to tackle COVID-19: the role of molecular dynamics simulations

Box 1. A possible strategy to overcome the barriers in the research and development of new active compounds is through drug repurposing of existing formulations for a different therapeutic indication (drug repurposing is usually characterized by a substantial variation from its original use). Since the COVID-19 outbreak, the alarming spread of the virus and the gravity of the infection led the scientific and medical community to seek rapid responses. The general approach to drug repurposing ideally starts with three steps: identifying the candidate molecule to generate the hypothesis,

preclinical studies of the candidate molecule, and evaluation of efficacy in phase II clinical trials (58). A wide set of computational techniques and software, which falls under the category of computer-aided drug design (CADD), is routinely used during the first step of drug repurposing to pinpoint potential drug candidates. CADD includes a plethora of ligand-based and structure-based approaches, involving target structural validation, binding energy evaluation in both static and dynamic models, and pharmacokinetics prediction.

A huge drug repurposing effort (Box 1) was put in place (59), worldwide, to shorten as much as possible the approval of therapeutics against SARS-CoV-2 validated targets. Necessarily, the S protein has been one of the most investigated COVID-19 therapeutic targets due to its unique function and central role in the early stage of infection.

The general idea behind targeting the spike protein is to act as a preventive defence against infection, with the intent of minimizing the risk of triggering a potentially dangerous over-reaction of the immune system, reducing de facto the burden on the public health sector. In May 2020, news about the efficacy against COVID-19 of hydroxychloroquine (HCQ) originated in China, and the use of HCQ and azithromycin (ATM) was indicated as a possible front-line treatment. Simulations indicated that HCQ and ATM would have a synergistic effect in the treatment of the infection, where HCQ acts as a competitive binder against gangliosides, another proposed receptor for S protein, and ATM interacts with the tip of SARS-CoV-2-spike (60). Although these results seemed promising, the outcomes of clinical trials appeared highly controversial and the hypothesis of adopting the combined HCQ and ATM therapy has been abandoned.

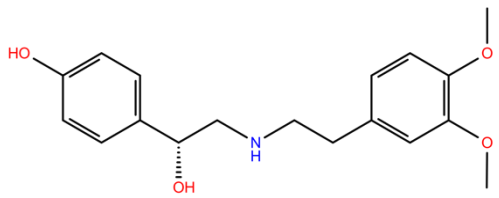
To face the threat of SARS-CoV-2 and its mutations, including the British (alpha) variant (61,62), large companies such as Pfizer BioNTech and AstraZeneca (63,64) have developed vaccines capable of activating an immunogenic response against the S protein. A global vaccination campaign has started, with more vaccines currently under development all around the world (65). However, as low-income countries struggle to have access to vaccines and immunosuppressed and allergic subjects cannot take advantage of the protection offered, alternative therapeutic approaches are still needed. Also, despite the high efficacy of vaccines, the full compliance of the population of high-income countries is yet to be reached, due to the limited knowledge of the long-term effects of new mRNA technologies and their implementation (64). In this scenario, drug repositioning could bring many advantages in terms of risk control and unwanted side effect management – because repurposed drugs have already passed safety assessments. Understandably, antiviral

agents were among the first agents to be tested against COVID-19. This approach led to the approval of Remdesivir as the first treatment for hospitalized patients (66–68), but not without controversies, due to uncertain outcomes of many clinical trials (67,69,70).

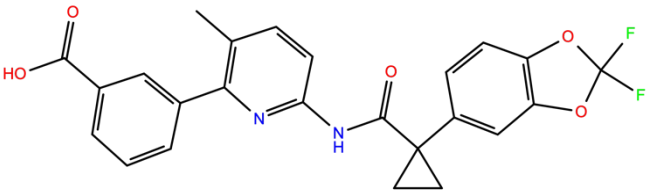
Long MD simulations have become a state-of-the-art computational tool in CADD (23) as they represent the best tool to validate in silico results of molecular docking and virtual screening campaigns. Here we report insights from MD simulations applied to the discovery of potential drugs able to interfere with the binding between RBD and ACE2. Only molecules tested both in vitro and in silico are reported.

One of the first computational works on SARS-CoV-2 proposed denopamine (Table 1A), bometolol, and Rotigaptide as possible inhibitors of S protein-ACE2 binding (71). The authors tested denopamine in vitro, observing a diminishing of RBD binding at denopamine concentrations > 100 μM (71). An in silico study highlighted simeprevir and lumacaftor as putative RBD binders (72). Lumacaftor (Table 1B) was subsequently proved to weakly bind to S protein with an IC_{50} of 84 ± 4 μM, although showing a good inhibition profile in Vero-E6 assays (73). Simeprevir (Table 1C) reduces the cellular viral load, synergizing with Remdesivir, but this effect was attributed to a direct action on the main protease and the RNA-dependent RNA polymerase (RdRp) (74). Post-docking MD simulations identified KT203, BMS195614, KT185, RS504393, and GSK1838705A (Table 1G-H), five compounds from the Sigma-Aldrich library of pharmacologically active compounds (LOPAC), as potential binders of the S protein (75). A retrospective MD investigation on arbidol (Table 1I), a therapeutic agent approved in China and Russia for influenza, showed an inhibitor effect on the original SARS spike protein (76); they proposed arbidol intercalated between different spike protein subunits, and so affecting the trimerization of the S protein (77). Docking and MD simulations performed by ourselves (78) and others (79) proposed Nilotinib (Table 1J) as a potential binder of the RBD or disruptor of the RBD-ACE2 complex. The anti-SARS-CoV potential of nilotinib was first reported in 2016 in the early stages of infection by inhibiting viral fusion at the endosomal level (80). A couple of years later further results pointed out an action of nilotinib and other Abl kinase inhibitors, on the virus-cell membrane fusion (81). In a recent study, the EC_{50} of imatinib was quantified as 1.44 μM and 3.06 μM in Vero-E6 cells and human respiratory cells respectively (82). Therefore, no experimental evidence for imatinib binding to RBD has been reported. The same goes for nafamostat, which we suggested as a putative RBD binder (78), but it is proposed to act as a TMPRSS2 inhibitor in the low nanomolar range (83,84).

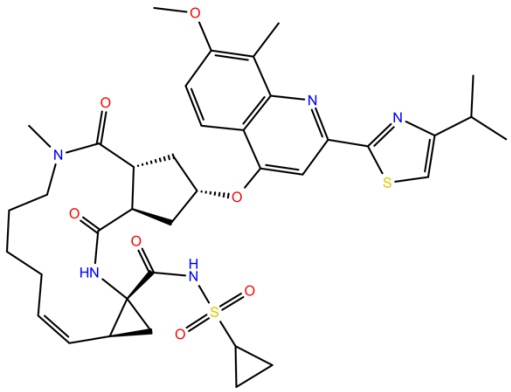
Table 1. Summary of the drugs, recently identified as protective against SARS-CoV-2 in vitro, proposed as RBD binders by MD simulations.



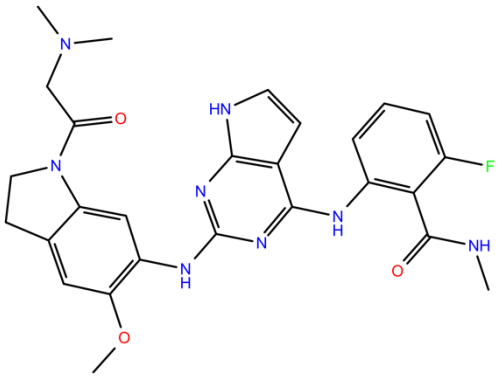
Denopamine: cardiotonic drug acting as an agonist at β1 adrenergic receptor; used in the treatment of angina (85)



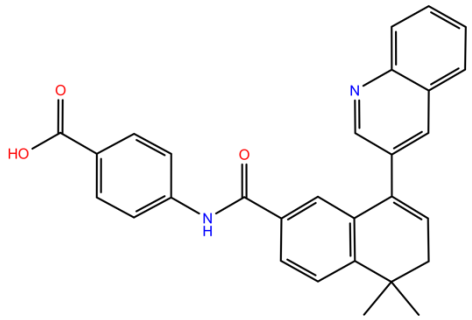
Lumacaftor: used for the treatment of cystic fibrosis in patients that present the F508del in the CFTR (cystic fibrosis transmembrane conductance regulator) proteins (86); IC_{50} of 84 ± 4 μM towards the S protein



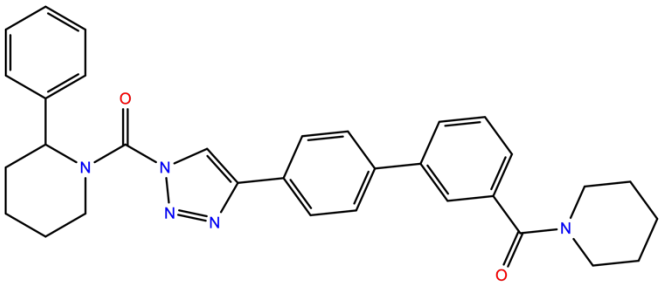
Simeprevir: inhibitor of the hepatitis C virus (HCV) NS3/NS4A protease (87) [71] IC_{50} of 9.6 ± 2.3 μM towards the M^{pro} and an IC_{50} value of 5.5 ± 0.2 μM towards the RdRp (RNA-dependent RNA polymerase) (74)



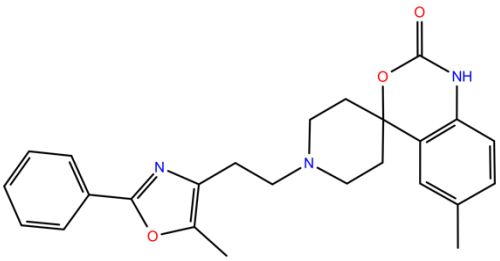
GSK1838705A: inhibitor of the insulin-like growth factor-1 receptor (IGF-IR), insulin receptor and anaplastic lymphoma kinase (ALK) (88)



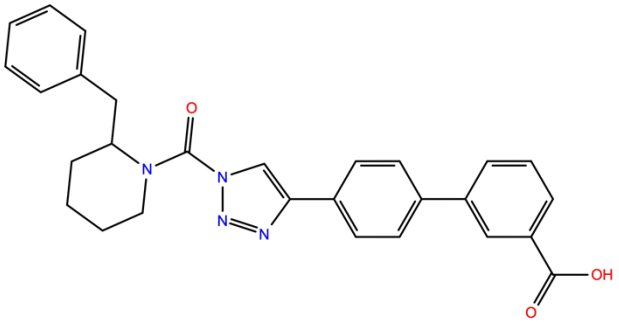
BMS195614: antagonist of the retinoic acid receptor (RAR)



KT185: inhibitor of α/β-hydrolase domain-containing 6 (ABHD6) in the brain and liver of mice

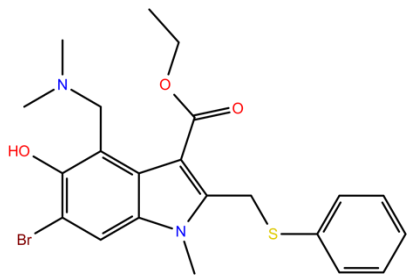


RS504393: antagonist of the CC2 chemokine receptor

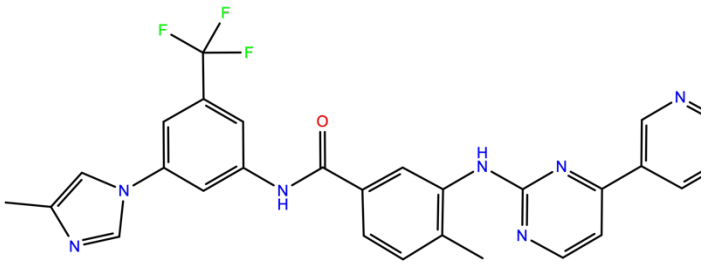


KT203: inhibitor of ABHD6 activity in the liver of mice

I.



J.



Arbidol: used as a treatment for influenza and other respiratory infections in Russia and China (77)

Nilotinib: a BCR-ABL tyrosine kinase inhibitor used for the treatment of chronic myelogenous leukaemia (CML) (89)

Perspectives for Targeting the Spike Protein

Although the worldwide effort to discover approved drugs to repurpose against the SARS-CoV-2 S protein, to date no MD-based study has delivered working hypotheses resulted in clinical trials. Open access COVID-19 drug repurposing databases (90,91) are a precious source of information but do not consider potential therapeutic agents proposed in silico, thus, there is a coordination gap between theoretical and experimental scientific communities (92).

Box 2. From a technical perspective, the amount of MD sampling to confirm molecular docking predictions has been generally limited to the time scale of a few tens of nanoseconds and this has probably produced numerous in silico false positives, undermining the credibility of computation studies. Simulations over tens or a few hundreds of nanoseconds showing a docking complex as stable should not be faithfully trusted. For example, a molecule with a residence time of a few microseconds (way longer than usual MD post docking simulations) and an optimistic binding k_{on} of $\approx 10^{7}$ $M^{-1}s^{-1}$ would have a kinetic affinity of about 10^{-2} M and therefore would not be a binder despite the indication provided by MD.

The discrepancy between the time scale of the real world and the simulated models (Box 2) can be partially overcome with end-state methods such as the MM-PBSA or MM-GBSA (93), which can quantify the binding free energy using short MD simulations. However, the accuracy of these methods is system-dependent and usually best suited for comparisons between congeneric ligands (94) rather than very different chemotypes, as is usually required by repurposing strategies. Enhanced MD sampling techniques such as metadynamics (95) speed up the time required to dissociate docking complexes and allow estimation of the stability of the bound ligand, therefore can aid in recognizing docking false positives.

We screened in silico more than 2000 approved small molecules (78). After the docking and post-docking MD simulations of the best-ranked compounds docked in situ to the RBD, we performed a further step consisting of simulating the encounter of the RBD and ACE2 in the presence of the best compounds Despite the stability displayed during the cefsulodin / RBD simulations, ACE2 easily displaced the proposed ligand in half of the supervised MD (SuMD (96,97) replicas. Increasing the complexity of the modelled system highlighted the limit of common computational protocols for correctly selecting small molecules able to hinder the recognition between S protein and ACE2. Therefore, dynamic approaches that consider the formation of the ternary systems between the S protein, ACE2, and the potential binding inhibitor should be routinely considered. Besides this, structure-based drug repurposing strategies should take into account the fully glycosylated S protein and the discovery of allosteric sites on the RBD to overcome the targeting obstacles. The extreme flexibility of the glycans on the surface of the S protein and the effective steric hindrance they offer affect the ability of antibodies or potential therapeutic molecules to bind to a sufficiently exposed epitope (98). To combat this, Haji-Ghassemi et al. suggested searching for drugs to target this shield (99) this is a different approach from the traditional protein-oriented one.

Molecular Dynamics insights on new SARS-CoV-2 variants:

Since the beginning of the pandemic, the evolutionary impact of SARS-CoV-2 has been kept under observation by the scientific community to evaluate the possible effects of mutations on transmissibility, severity, and viral evasion of the immune mechanism (100–103). Among the SARS-CoV-2 variants, the major preoccupations regarded those strains that carried important mutations and deletions, especially on the RBD. (categorized as variants of concern (VOC)) (100). Such VOC have important RBD mutations: B.1.1.7 (Alpha), carries E484K, N501Y, D614G, P681H; B.1.351 (Beta) carries K417N, E484K, N501Y, D614G, A701V; P1 (Gamma) carries K417T, E484K, N501Y, D614G, H655Y; B.1.617.2 (Delta) carries L452R, T478K, D614G, P681R (102). Concerns among the scientific community have risen due to their potential to elude the immune system and overcome vaccine protection (104–106) despite showing an overall similarity between variants, which diverged only in terms of flexibility (107).

More recently, a new B.1.1.529 (Omicron) VOC (108–110) carrying N440K, G446S, S477N, 118 T478K, E484A, Q493R, G496S, Q498R, N501Y, and Y505H mutations, and its lineages became predominant over the Delta variant, possibly due to a more rapid entry or different mechanism (111–113), an enhanced ability to evade the immune system (103,114,115), and its increased affinity for ACE2 (116–118) although showing a milder pathogenic impact (119). New VOCs are expected to pose a new threat should they become

widespread (120,121) and further studies should follow to evaluate the potential risk of new mutations.

MD-based computational efforts evaluated the effect of omicron's mutations on ACE2 binding strength (118,122,123), suggesting that YG339D, N440K, S477N, T478K, Q493K, N501Y increase the binding affinity, as also reported by Socher et al, (124). S371L, S373P, S375F, K417N, G446S, E484A, G496S, Q498R, and Y505H, on the other hand, decreased the binding affinity for ACE2, in agreement with a compensatory effect that moderates the binding strength of the enhancing mutations (125). However, the reinforced network of hydrogen bonds involving T500-D355, G502-K353, N487-Y83, as well as R493-D38, and A475-S19, paired with the electrostatic matching between R493-D38 and the loop shift caused by E484A and T478K mutation as suggested by Zhao et al, (126), suggesting an overall increase in the binding energy. These shifts seem to increase the complementarity between ACE2 and Omicron's RBD and could be the reason for the increased binding affinity, as also highlighted by Nie et al, (127).

The advent of the new VOC highlighted the necessity to follow multiple paths, for a broad-spectrum therapeutic approach, which should not only consider RBD as the target of main interest but should also consider more conserved viral proteins among the variants. MD studies were carried out to investigate non-structural proteins (NSP) as potential druggable targets (128,129) whereas Vivek et al, suggested NSP13 helicase ATP-binding sites as a druggable target, while Vardhan et al, included the NSP14 and NSP15 exonuclease and endonuclease respectively. Both groups used phytochemicals small molecules as target binders, which, however, still require experimental data to confirm their efficacy.

Alternatively, PF-07321332 a promising oral antiviral candidate against the main protease (MPro) catalytic dyad on residues H41-C145 was investigated using MD simulation by Macchiagodena et al, using preliminary data structures available (130). According to Macchiagodena's work, the formation of the thiolate-imidazolium, paired with the exposition of the nitrile warhead in the proximity of the C145 would allow for the electrophilic attack on the MPro, for effective enzyme inhibition.

However, MD is not the most adequate method to represent bond formation or breaking, but the insights provided by contact frequency and interatomic distances could support the description of the experimental data, once the PF-07321332-MPro complex structure becomes publicly available.

Other attempts against MPro were made (130–133) but require further experimental data to validate the hypothesis. The RNA polymerase, (134) as well as the nucleocapsid (135), and envelope protein (136) were proposed as a druggable target, but these studies

will require further confirmation by experimental data to verify whether NSPs could be considered viable targets.

Conclusion

As a state-of-the-art computational technique, MD has been broadly employed to interrogate the structure and function of the S protein at the atomic level to understand how its inherent flexibility modulates the binding to ACE2 receptors and, therefore, SARS-CoV-2 virulency. MD suggested unexpected flexibility of the stalk region S2, the role of glycans on the S protein surface, and the contribution of single residues on the RBD to the interactions with ACE2. MD contributed to describing important dynamics and structural elements such as the minimum angulation required for molecular recognition between ACE2 and RBD, the effects of mutations on the binding capacity of the S protein, as well as the structural and protective role of glycans. Through MD it was possible to understand the spontaneous motions that open RBD from the "down" to "up" conformation, revealing numerous cryptic pockets, possible targets of new drugs. The “down” to “up” transition that the RBD undergoes before ACE2 recognition was another important phenomenon MD MD-delivered structural insights.

From a future perspective, we believe there is scope for an increasingly important contribution of MD in the study of Ab and their rational development as therapeutic agents. Also, MD contributed to rationalizing in vitro data on potential S protein binding antagonists, but with limited utility in drug repurposing. Approaches to address COVID-19 start to fade away from drug repurposing and the S protein to more classic rational strategies to target functional viral proteins, as demonstrated by the main protease (Mpro) inhibitor nirmatrelvir, the first oral anti-COVID-19 drug approved by the FDA. In this scenario, it is plausible that MD will regain a central role in aiding the development of future new classes of therapeutics against SARS-CoV-2.

Videos available at https://www.biorxiv.org/content/10.1101/2022.07.05.498807v1.supplementary-material.

Bibliography

1. Chan JF-W, Yuan S, Kok K-H, To KK-W, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet. 2020 Feb 15;395(10223):514–23.
2. Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020

Mar;579(7798):265–9.

3. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020 Mar;579(7798):270–3.

4. Siddell SG. The coronaviridae: an introduction. In: Siddell SG, editor. The Coronaviridae. Boston, MA: Springer US; 1995. p. 1–10.

5. Ragab D, Salah Eldin H, Taeimah M, Khattab R, Salem R. The COVID-19 Cytokine Storm; What We Know So Far. Front Immunol. 2020 Jun 16;11:1446.

6. Bosch BJ, van der Zee R, de Haan CAM, Rottier PJM. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. J Virol. 2003 Aug;77(16):8801–11.

7. Xia X. Domains and Functions of Spike Protein in Sars-Cov-2 in the Context of Vaccine Design. Viruses. 2021 Jan 14;13(1).

8. Fan X, Cao D, Kong L, Zhang X. Cryo-EM analysis of the post-fusion structure of the SARS-CoV spike glycoprotein. Nat Commun. 2020 Jul 17;11(1):3618.

9. Al Adem K, Shanti A, Stefanini C, Lee S. Inhibition of SARS-CoV-2 Entry into Host Cells Using Small Molecules. Pharmaceuticals (Basel). 2020 Dec 8;13(12).

10. Tikellis C, Thomas MC. Angiotensin-Converting Enzyme 2 (ACE2) Is a Key Modulator of the Renin Angiotensin System in Health and Disease. Int J Pept. 2012 Mar 20;2012:256294.

11. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science. 2020 Mar 13;367(6483):1260–3.

12. Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature. 2020 May;581(7807):215–20.

13. Benton DJ, Wrobel AG, Xu P, Roustan C, Martin SR, Rosenthal PB, et al. Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. Nature. 2020 Dec;588(7837):327–30.

14. Shang J, Wan Y, Luo C, Ye G, Geng Q, Auerbach A, et al. Cell entry mechanisms of SARS-CoV-2. Proc Natl Acad Sci USA. 2020 May 26;117(21):11727–34.

15. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. Nucleic Acids Res. 2000 Jan 1;28(1):235–42.

16. Ke Z, Oton J, Qu K, Cortese M, Zila V, McKeane L, et al. Structures and distributions of SARS-CoV-2 spike proteins on intact virions. Nature. 2020 Dec;588(7838):498–502.

17. Henderson R, Edwards RJ, Mansouri K, Janowska K, Stalls V, Gobeil SMC, et al. Controlling the SARS-CoV-2 spike glycoprotein conformation. Nat Struct Mol Biol. 2020 Oct;27(10):925–33.

18. Turoňová B, Sikora M, Schürmann C, Hagen WJH, Welsch S, Blanc FEC, et al. In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. Science. 2020 Oct 9;370(6513):203–8.

19. Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell. 2020 Apr 16;181(2):281-292.e6.

20. Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. Science. 2020 Mar 27;367(6485):1444–8.

21. Cai Y, Zhang J, Xiao T, Peng H, Sterling SM, Walsh RM, et al. Distinct conformational states of SARS-CoV-2 spike protein. Science. 2020 Sep 25;369(6511):1586–92.

22. Chen Y, Guo Y, Pan Y, Zhao ZJ. Structure analysis of the receptor binding of 2019-nCoV. Biochem Biophys Res Commun. 2020 Feb 17;

23. Deganutti G, Moro S, Reynolds CA. A Supervised Molecular Dynamics Approach to Unbiased Ligand-Protein Unbinding. J Chem Inf Model. 2020 Mar 23;60(3):1804–17.

24. Chi X, Yan R, Zhang J, Zhang G, Zhang Y, Hao M, et al. A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. Science. 2020 Aug 7;369(6504):650–5.

25. Walls AC, Tortorici MA, Frenz B, Snijder J, Li W, Rey FA, et al. Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. Nat Struct Mol Biol. 2016 Oct;23(10):899–905.

26. Grant OC, Montgomery D, Ito K, Woods RJ. Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition. Sci Rep. 2020 Sep 14;10(1):14991.

27. Hollingsworth SA, Dror RO. Molecular dynamics simulation for all. Neuron. 2018 Sep 19;99(6):1129–43.

28. Padhi AK, Rath SL, Tripathi T. Accelerating COVID-19 Research Using Molecular Dynamics Simulation. J Phys Chem B. 2021 Jul 28;

29. Woo H, Park S-J, Choi YK, Park T, Tanveer M, Cao Y, et al. Developing a Fully Glycosylated Full-Length SARS-CoV-2 Spike Protein Model in a Viral Membrane. J Phys Chem B. 2020 Aug 20;124(33):7128–37.

30. Allen JD, Chawla H, Samsudin F, Zuzic L, Shivgan AT, Watanabe Y, et al. Site-Specific Steric Control of SARS-CoV-2 Spike Glycosylation. Biochemistry. 2021 Jul 13;60(27):2153–69.

31. Choi YK, Cao Y, Frank M, Woo H, Park S-J, Yeom MS, et al. Structure, Dynamics, Receptor Binding, and Antibody Binding of the Fully Glycosylated Full-Length SARS-CoV-2 Spike Protein in a Viral Membrane. J Chem Theory Comput. 2021 Apr 13;17(4):2479–87.

32. Pramanick I, Sengupta N, Mishra S, Pandey S, Girish N, Das A, et al. Conformational flexibility and structural variability of SARS-CoV2 S protein. Structure. 2021 Aug 5;29(8):834-845.e5.

33. Verkhivker G. Conformational Flexibility and Local Frustration in the Functional States of the SARS-CoV-2 Spike B.1.1.7 and B.1.351 Variants: Mutation-Induced Allosteric Modulation Mechanism of Functional Dynamics and Protein Stability. Int J Mol Sci. 2022 Jan 31;23(3).

34. Takeda M. Proteolytic activation of SARS-CoV-2 spike protein. Microbiol Immunol. 2022 Jan;66(1):15–23.

35. Schaefer SL, Jung H, Hummer G. Binding of SARS-CoV-2 Fusion Peptide to Host Endosome and Plasma Membrane. J Phys Chem B. 2021 Jul 22;125(28):7732–41.

36. Remington JM, McKay KT, Ferrell JB, Schneebeli ST, Li J. Enhanced sampling protocol to elucidate fusion peptide opening of SARS-CoV-2 spike protein. Biophys J. 2021 Jul 20;120(14):2848–58.

37. Pavlova A, Zhang Z, Acharya A, Lynch DL, Pang YT, Mou Z, et al. Machine Learning Reveals the Critical Interactions for SARS-CoV-2 Spike Protein Binding to ACE2. J Phys Chem Lett. 2021 Jun 17;12(23):5494–502.

38. Giron CC, Laaksonen A, Barroso da Silva FL. Up State of the SARS-COV-2 Spike Homotrimer Favors an Increased Virulence for New Variants. Front Med Technol. 2021 Jun 30;3:694347.

39. Ali A, Vijayan R. Dynamics of the ACE2-SARS-CoV-2/SARS-CoV spike protein interface reveal unique mechanisms. Sci Rep. 2020 Aug 26;10(1):14214.

40. Barros EP, Casalino L, Gaieb Z, Dommer AC, Wang Y, Fallon L, et al. The flexibility of

ACE2 in the context of SARS-CoV-2 infection. Biophys J. 2021 Mar 16;120(6):1072–84.

41. Mehdipour AR, Hummer G. Dual nature of human ACE2 glycosylation in binding to SARS-CoV-2 spike. Proc Natl Acad Sci USA. 2021 May 11;118(19).

42. Williams JK, Wang B, Sam A, Hoop CL, Case DA, Baum J. Molecular dynamics analysis of a flexible loop at the binding interface of the SARS-CoV-2 spike protein receptor-binding domain. Proteins. 2022 May;90(5):1044–53.

43. Peng C, Zhu Z, Shi Y, Wang X, Mu K, Yang Y, et al. Exploring the Binding Mechanism and Accessible Angle of SARS-CoV-2 Spike and ACE2 by Molecular Dynamics Simulation and Free Energy Calculation. 2020 Feb 21;

44. Zuckerman DM, Chong LT. Weighted ensemble simulation: review of methodology, applications, and software. Annu Rev Biophys. 2017 May 22;46:43–57.

45. Sztain T, Ahn S-H, Bogetti AT, Casalino L, Goldsmith JA, Seitz E, et al. A glycan gate controls opening of the SARS-CoV-2 spike protein. Nat Chem. 2021 Oct;13(10):963–8.

46. Harbison AM, Fogarty CA, Phung TK, Satheesan A, Schulz BL, Fadda E. Fine-tuning the spike: role of the nature and topology of the glycan shield in the structure and dynamics of the SARS-CoV-2 S. Chem Sci. 2022 Jan 5;13(2):386–95.

47. Casalino L, Gaieb Z, Goldsmith JA, Hjorth CK, Dommer AC, Harbison AM, et al. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. ACS Cent Sci. 2020 Oct 28;6(10):1722–34.

48. Casalino L, Dommer A, Gaieb Z, Barros EP, Sztain T, Ahn S-H, et al. AI-Driven Multiscale Simulations Illuminate Mechanisms of SARS-CoV-2 Spike Dynamics. BioRxiv. 2020 Nov 20;

49. Yao H, Song Y, Chen Y, Wu N, Xu J, Sun C, et al. Molecular Architecture of the SARS-CoV-2 Virus. Cell. 2020 Oct 29;183(3):730-738.e13.

50. Gur M, Taka E, Yilmaz SZ, Kilinc C, Aktas U, Golcuk M. Conformational transition of SARS-CoV-2 spike glycoprotein between its closed and open states. J Chem Phys. 2020 Aug 21;153(7):075101.

51. Mori T, Jung J, Kobayashi C, Dokainish HM, Re S, Sugita Y. Elucidation of interactions regulating conformational stability and dynamics of SARS-CoV-2 S-protein. Biophys J. 2021 Mar 16;120(6):1060–71.

52. Dokainish HM, Re S, Mori T, Kobayashi C, Jung J, Sugita Y. The inherent flexibility of

receptor binding domains in SARS-CoV-2 spike protein. eLife. 2022 Mar 24;11.

53. Kamiya M, Sugita Y. Flexible selection of the solute region in replica exchange with solute tempering: Application to protein-folding simulations. J Chem Phys. 2018 Aug 21;149(7):072304.

54. Zimmerman MI, Porter JR, Ward MD, Singh S, Vithani N, Meller A, et al. SARS-CoV-2 simulations go exascale to predict dramatic spike opening and cryptic pockets across the proteome. Nat Chem. 2021 Jul;13(7):651–9.

55. Sikora M, von Bülow S, Blanc FEC, Gecht M, Covino R, Hummer G. Computational epitope map of SARS-CoV-2 spike protein. PLoS Comput Biol. 2021 Apr 1;17(4):e1008790.

56. Serapian SA, Marchetti F, Triveri A, Morra G, Meli M, Moroni E, et al. The Answer Lies in the Energy: How Simple Atomistic Molecular Dynamics Simulations May Hold the Key to Epitope Prediction on the Fully Glycosylated SARS-CoV-2 Spike Protein. J Phys Chem Lett. 2020 Oct 1;11(19):8084–93.

57. Jokinen EM, Gopinath K, Kurkinen ST, Pentikainen OT. Detection of Binding Sites on SARS-CoV-2 Spike Protein Receptor-Binding Domain by Molecular Dynamics Simulations in Mixed Solvents. IEEE/ACM Trans Comput Biol Bioinform. 2021 Aug 6;18(4):1281–9.

58. Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, et al. Drug repurposing: progress, challenges and recommendations. Nat Rev Drug Discov. 2019 Jan;18(1):41–58.

59. Janik E, Niemcewicz M, Podogrocki M, Saluk-Bijak J, Bijak M. Existing Drugs Considered as Promising in COVID-19 Therapy. Int J Mol Sci. 2021 May 21;22(11).

60. Fantini J, Chahinian H, Yahi N. Synergistic antiviral effect of hydroxychloroquine and azithromycin in combination against SARS-CoV-2: What molecular dynamics studies of virus-host interactions reveal. Int J Antimicrob Agents. 2020 Aug;56(2):106020.

61. Gu H, Krishnan P, Ng DYM, Chang LDJ, Liu GYZ, Cheng SSM, et al. Probable Transmission of SARS-CoV-2 Omicron Variant in Quarantine Hotel, Hong Kong, China, November 2021. Emerging Infect Dis. 2022 Feb;28(2):460–2.

62. McCallum M, Walls AC, Sprouse KR, Bowen JE, Rosen LE, Dang HV, et al. Molecular basis of immune evasion by the Delta and Kappa SARS-CoV-2 variants. Science. 2021 Dec 24;374(6575):1621–6.

63. Ramasamy MN, Minassian AM, Ewer KJ, Flaxman AL, Folegatti PM, Owens DR, et

al. Safety and immunogenicity of ChAdOx1 nCoV-19 vaccine administered in a prime-boost regimen in young and old adults (COV002): a single-blind, randomised, controlled, phase 2/3 trial. Lancet. 2021 Dec 19;396(10267):1979–93.

64. Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. N Engl J Med. 2020 Dec 31;383(27):2603–15.

65. Chung JY, Thone MN, Kwon YJ. COVID-19 vaccines: The status and perspectives in delivery points of view. Adv Drug Deliv Rev. 2021 Mar;170:1–25.

66. Beigel JH, Tomashek KM, Dodd LE, Mehta AK, Zingman BS, Kalil AC, et al. Remdesivir for the Treatment of Covid-19 - Final Report. N Engl J Med. 2020 Nov 5;383(19):1813–26.

67. Wang Y, Zhang D, Du G, Du R, Zhao J, Jin Y, et al. Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. Lancet. 2020 May 16;395(10236):1569–78.

68. Gottlieb RL, Vaca CE, Paredes R, Mera J, Webb BJ, Perez G, et al. Early Remdesivir to Prevent Progression to Severe Covid-19 in Outpatients. N Engl J Med. 2022 Jan 27;386(4):305–15.

69. Cohen J. The ‘very, very bad look’ of remdesivir, the first FDA-approved COVID-19 drug. Science. 2020 Oct 27;

70. Sreekanth Reddy O, Lai W-F. Tackling COVID-19 Using Remdesivir and Favipiravir as Therapeutic Options. Chembiochem. 2021 Mar 16;22(6):939–48.

71. Durdagi S, Orhan MD, Aksoydan B, Calis S, Dogan B, Sahin K, et al. Screening of Clinically Approved and Investigation Drugs as Potential Inhibitors of SARS-CoV-2: A Combined in silico and in vitro Study. Mol Inform. 2021 Sep 16;e2100062.

72. Trezza A, Iovinelli D, Prischi F, Santucci A, Spiga O. An integrated drug repurposing strategy for the rapid identification of potential SARS-CoV-2 viral inhibitors. 2020 Apr 30;

73. Day CJ, Bailly B, Guillon P, Dirr L, Jen FE-C, Spillings BL, et al. Multidisciplinary Approaches Identify Compounds that Bind to Human ACE2 or SARS-CoV-2 Spike Protein as Candidates to Block SARS-CoV-2-ACE2 Receptor Interactions. MBio. 2021 Mar 30;12(2).

74. Lo HS, Hui KPY, Lai H-M, He X, Khan KS, Kaur S, et al. Simeprevir Potently Suppresses SARS-CoV-2 Replication and Synergizes with Remdesivir. ACS Cent

Sci. 2021 May 26;7(5):792–802.

75. Choudhary S, Malik YS, Tomar S. Identification of SARS-CoV-2 Cell Entry Inhibitors by Drug Repurposing Using in silico Structure-Based Virtual Screening Approach. Front Immunol. 2020 Jul 10;11:1664.

76. Zhang J, Zhou L, Yang Y, Peng W, Wang W, Chen X. Therapeutic and triage strategies for 2019 novel coronavirus disease in fever clinics. Lancet Respir Med. 2020 Mar;8(3):e11–2.

77. Vankadari N. Arbidol: A potential antiviral drug for the treatment of SARS-CoV-2 by blocking trimerization of the spike glycoprotein. Int J Antimicrob Agents. 2020 Aug;56(2):105998.

78. Deganutti G, Prischi F, Reynolds CA. Supervised molecular dynamics for exploring the druggability of the SARS-CoV-2 spike protein. J Comput Aided Mol Des. 2021 Feb;35(2):195–207.

79. Razizadeh M, Nikfar M, Liu Y. Small molecule therapeutics to destabilize the ACE2-RBD complex: A molecular dynamics study. Biophys J. 2021 Jul 20;120(14):2793–804.

80. Coleman CM, Sisk JM, Mingo RM, Nelson EA, White JM, Frieman MB. Abelson kinase inhibitors are potent inhibitors of severe acute respiratory syndrome coronavirus and middle east respiratory syndrome coronavirus fusion. J Virol. 2016 Oct 1;90(19):8924–33.

81. Sisk JM, Frieman MB, Machamer CE. Coronavirus S protein-induced fusion is blocked prior to hemifusion by Abl kinase inhibitors. J Gen Virol. 2018 May;99(5):619–30.

82. Cagno V, Magliocco G, Tapparel C, Daali Y. The tyrosine kinase inhibitor nilotinib inhibits SARS-CoV-2 in vitro. Basic Clin Pharmacol Toxicol. 2021 Apr;128(4):621–4.

83. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell. 2020 Apr 16;181(2):271-280.e8.

84. Zhu H, Du W, Song M, Liu Q, Herrmann A, Huang Q. Spontaneous binding of potential COVID-19 drugs (Camostat and Nafamostat) to human serine protease TMPRSS2. Comput Struct Biotechnol J. 2021;19:467–76.

85. Nakajima D, Negoro N, Nakaboh A, Nakakoji T, Hoshiga M, Nariyama J, et al. Effectiveness of low dose denopamine, a beta1-adrenoceptor agonist, in a patient

with vasospastic angina refractory to intensive medical treatment. Int J Cardiol. 2006 Apr 4;108(2):281–3.

86. Saint-Criq V, Gray MA. Role of CFTR in epithelial physiology. Cell Mol Life Sci. 2017 Jan;74(1):93–115.

87. Raboisson P, de Kock H, Rosenquist A, Nilsson M, Salvador-Oden L, Lin T-I, et al. Structure-activity relationship study on a novel series of cyclopentane-containing macrocyclic inhibitors of the hepatitis C virus NS3/4A protease leading to the discovery of TMC435350. Bioorg Med Chem Lett. 2008 Sep 1;18(17):4853–8.

88. Sabbatini P, Korenchuk S, Rowand JL, Groy A, Liu Q, Leperi D, et al. GSK1838705A inhibits the insulin-like growth factor-1 receptor and anaplastic lymphoma kinase and shows antitumor activity in experimental models of human cancers. Mol Cancer Ther. 2009 Oct;8(10):2811–20.

89. Kantarjian H, Giles F, Wunderle L, Bhalla K, O'Brien S, Wassmann B, et al. Nilotinib in imatinib-resistant CML and Philadelphia chromosome-positive ALL. N Engl J Med. 2006 Jun 15;354(24):2542–51.

90. Excelra | Covid-19-Drug-Repurposing-Database [Internet]. [cited 2021 Oct 11]. Available from: https://www.excelra.com/covid-19-drug-repurposing-database/

91. Brimacombe KR, Zhao T, Eastman RT, Hu X, Wang K, Backus M, et al. An OpenData portal to share COVID-19 drug repurposing data in real time. BioRxiv. 2020 Jun 5;

92. Galindez G, Matschinske J, Rose TD, Sadegh S, Salgado-Albarrán M, Späth J, et al. Lessons from the COVID-19 pandemic for advancing computational drug repurposing strategies. Nat Comput Sci. 2021 Jan;1(1):33–41.

93. Wang E, Sun H, Wang J, Wang Z, Liu H, Zhang JZH, et al. End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. Chem Rev. 2019 Aug 28;119(16):9478–508.

94. Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discov. 2015 May;10(5):449–61.

95. Barducci A, Bonomi M, Parrinello M. Linking well-tempered metadynamics simulations with experiments. Biophys J. 2010 May 19;98(9):L44-6.

96. Sabbadin D, Moro S. Supervised molecular dynamics (SuMD) as a helpful tool to depict GPCR-ligand recognition pathway in a nanosecond time scale. J Chem Inf Model. 2014 Feb 24;54(2):372–6.

97. Cuzzolin A, Sturlese M, Deganutti G, Salmaso V, Sabbadin D, Ciancetta A, et al.

Deciphering the Complexity of Ligand-Protein Recognition Pathways Using Supervised Molecular Dynamics (SuMD) Simulations. J Chem Inf Model. 2016 Apr 25;56(4):687–705.

98. Homans SW, Dwek RA, Boyd J, Mahmoudian M, Richards WG, Rademacher TW. Conformational transitions in N-linked oligosaccharides. Biochemistry. 1986 Oct 7;25(20):6342–50.

99. Haji-Ghassemi O, Blackler RJ, Martin Young N, Evans SV. Antibody recognition of carbohydrate epitopes†. Glycobiology. 2015 Sep;25(9):920–52.

100. Choi JY, Smith DM. SARS-CoV-2 Variants of Concern. Yonsei Med J. 2021 Nov;62(11):961–8.

101. Thakur V, Ratho RK. OMICRON (B.1.1.529): A new SARS-CoV-2 variant of concern mounting worldwide fear. J Med Virol. 2022 May;94(5):1821–4.

102. Gupta D, Sharma P, Singh M, Kumar M, Ethayathulla AS, Kaur P. Structural and functional insights into the spike protein mutations of emerging SARS-CoV-2 variants. Cell Mol Life Sci. 2021 Dec;78(24):7967–89.

103. Iketani S, Liu L, Guo Y, Liu L, Chan JF-W, Huang Y, et al. Antibody evasion properties of SARS-CoV-2 Omicron sublineages. Nature. 2022 Apr;604(7906):553–6.

104. Al-Tawfiq JA, Koritala T, Alhumaid S, Barry M, Alshukairi AN, Temsah M-H, et al. Implication of the emergence of the delta (B.1.617.2) variants on vaccine effectiveness. Infection. 2022 Jun;50(3):583–96.

105. Mistry P, Barmania F, Mellet J, Peta K, Strydom A, Viljoen IM, et al. SARS-CoV-2 Variants, Vaccines, and Host Immunity. Front Immunol. 2021;12:809244.

106. McCarthy KR, Rennick LJ, Nambulli S, Robinson-McCarthy LR, Bain WG, Haidar G, et al. Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. Science. 2021 Mar 12;371(6534):1139–42.

107. Mandal N, Padhi AK, Rath SL. Molecular insights into the differential dynamics of SARS-CoV-2 variants of concern. J Mol Graph Model. 2022 Apr 14;114:108194.

108. Classification of Omicron (B.1.1.529): SARS-CoV-2 Variant of Concern [Internet]. [cited 2022 May 14]. Available from: https://www.who.int/news/item/26-11-2021-classification-of-omicron-(b.1.1.529)-sars-cov-2-variant-of-concern

109. Kannan S, Shaik Syed Ali P, Sheeza A. Omicron (B.1.1.529) - variant of concern - molecular profile and epidemiology: a mini review. Eur Rev Med Pharmacol Sci. 2021

Dec;25(24):8019–22.

110. Gao S-J, Guo H, Luo G. Omicron variant (B.1.1.529) of SARS-CoV-2, a global urgent public health alert! J Med Virol. 2022 Apr;94(4):1255–6.

111. Peacock TP, Brown JC, Zhou J, Thakur N, Newman J, Kugathasan R, et al. The SARS-CoV-2 variant, Omicron, shows rapid replication in human primary nasal epithelial cultures and efficiently uses the endosomal route of entry. BioRxiv. 2022 Jan 3;

112. Callaway E. Why does the Omicron sub-variant spread faster than the original? Nature. 2022 Feb;602(7898):556–7.

113. Arora P, Zhang L, Krüger N, Rocha C, Sidarovich A, Schulz S, et al. SARS-CoV-2 Omicron sublineages show comparable cell entry but differential neutralization by therapeutic antibodies. Cell Host Microbe. 2022 May 6;

114. Unable to find information for 12902880.

115. Meng B, Ferreira I, Abdullahi A, Kemp SA, Goonawardane N, Papa G, et al. SARS-CoV-2 Omicron spike mediated immune escape, infectivity and cell-cell fusion. BioRxiv. 2021 Dec 21;

116. Kumar R, Murugan NA, Srivastava V. Improved binding affinity of the Omicron's spike protein with hACE2 receptor is the key factor behind its increased virulence. BioRxiv. 2021 Dec 28;

117. Yin W, Xu Y, Xu P, Cao X, Wu C, Gu C, et al. Structures of the Omicron spike trimer with ACE2 and an anti-Omicron antibody. Science. 2022 Mar 4;375(6584):1048–53.

118. Lupala CS, Ye Y, Chen H, Su X-D, Liu H. Mutations on RBD of SARS-CoV-2 Omicron variant result in stronger binding to human ACE2 receptor. Biochem Biophys Res Commun. 2022 Jan 29;590:34–41.

119. Shuai H, Chan JF-W, Hu B, Chai Y, Yuen TT-T, Yin F, et al. Attenuated replication and pathogenicity of SARS-CoV-2 B.1.1.529 Omicron. Nature. 2022 Mar;603(7902):693–9.

120. Yamasoba D, Kosugi Y, Kimura I, Fujita S, Uriu K, Ito J, et al. Sensitivity of novel SARS-CoV-2 Omicron subvariants, BA.2.11, BA.2.12.1, BA.4 and BA.5 to therapeutic monoclonal antibodies. BioRxiv. 2022 May 3;

121. Khan K, Karim F, Ganga Y, Bernstein M, Jule Z, Reedoy K, et al. Omicron sub-lineages BA.4/BA.5 escape BA.1 infection elicited neutralizing immunity. medRxiv. 2022 May 1;

122. Wu L, Zhou L, Mo M, Li Y, Han J, Li J, et al. The effect of the multiple mutations in Omicron RBD on its binding to human ACE2 receptor and immune evasion: an investigation of molecular dynamics simulations. 2021 Dec 7;

123. Jawaid MZ, Baidya A, Mahboubi-Ardakani R, Davis RL, Cox DL. Simulation of the omicron variant of SARS-CoV-2 shows broad antibody escape, weakened ACE2 binding, and modest increase in furin binding. BioRxiv. 2021 Dec 15;

124. Socher E, Heger L, Paulsen F, Zunke F, Arnold P. Molecular dynamics simulations of the delta and omicron SARS-CoV-2 spike - ACE2 complexes reveal distinct changes between both variants. Comput Struct Biotechnol J. 2022 Feb 26;20:1168–76.

125. Wu L, Zhou L, Mo M, Liu T, Wu C, Gong C, et al. SARS-CoV-2 Omicron RBD shows weaker binding affinity than the currently dominant Delta variant to human ACE2. Signal Transduct Target Ther. 2022 Jan 5;7(1):8.

126. Zhao X, Xiong D, Luo S, Duan L. Origin of the tight binding mode to ACE2 triggered by multi-point mutations in the omicron variant: a dynamic insight. Phys Chem Chem Phys. 2022 Apr 13;24(15):8724–37.

127. Nie C, Sahoo AK, Netz RR, Herrmann A, Ballauff M, Haag R. Charge matters: mutations in omicron variant favor binding to cells. Chembiochem. 2022 Mar 18;23(6):e202100681.

128. Vivek-Ananth RP, Krishnaswamy S, Samal A. Potential phytochemical inhibitors of SARS-CoV-2 helicase Nsp13: a molecular docking and dynamic simulation study. Mol Divers. 2022 Feb;26(1):429–42.

129. Vardhan S, Sahoo SK. Exploring the therapeutic nature of limonoids and triterpenoids against SARS-CoV-2 by targeting nsp13, nsp14, and nsp15 through molecular docking and dynamics simulations. J Tradit Complement Med. 2022 Jan;12(1):44–54.

130. Macchiagodena M, Pagliai M, Procacci P. Characterization of the non-covalent interaction between the PF-07321332 inhibitor and the SARS-CoV-2 main protease. J Mol Graph Model. 2022 Jan;110:108042.

131. Shree P, Mishra P, Selvaraj C, Singh SK, Chaube R, Garg N, et al. Targeting COVID-19 (SARS-CoV-2) main protease through active phytochemicals of ayurvedic medicinal plants - Withania somnifera (Ashwagandha), Tinospora cordifolia (Giloy) and Ocimum sanctum (Tulsi) - a molecular docking study. J Biomol Struct Dyn. 2022 Jan;40(1):190–203.

132. El Khatabi K, Aanouz I, Alaqarbeh M, Ajana MA, Lakhlifi T, Bouachrine M. Molecular docking, molecular dynamics simulation, and ADMET analysis of levamisole

derivatives against the SARS-CoV-2 main protease (MPro). Bioimpacts. 2022;12(2):107–13.

133. Raman APS, Singh MB, Vishvakarma VK, Jain P, Kumar A, Sachdeva S, et al. An investigation for the interaction of gamma oryzanol with the Mpro of SARS-CoV-2 to combat COVID-19: DFT, molecular docking, ADME and molecular dynamics simulations. J Biomol Struct Dyn. 2022 Jan 22;1–11.

134. Elfiky AA, Mahran HA, Ibrahim IM, Ibrahim MN, Elshemey WM. Molecular dynamics simulations and MM-GBSA reveal novel guanosine derivatives against SARS-CoV-2 RNA dependent RNA polymerase. RSC Adv. 2022 Jan 18;12(5):2741–50.

135. Ahamad S, Gupta D, Kumar V. Targeting SARS-CoV-2 nucleocapsid oligomerization: Insights from molecular docking and molecular dynamics simulations. J Biomol Struct Dyn. 2022 Apr;40(6):2430–43.

136. Kuzmin A, Orekhov P, Astashkin R, Gordeliy V, Gushchin I. Structure and dynamics of the SARS-CoV-2 envelope protein monomer. Proteins. 2022 May;90(5):1102–14.

6.2 Understanding The Role of Heparinoids on the SARS-CoV-2 Spike Protein through Molecular Dynamics Simulations

The review of the literature on the S protein brought to my attention the challenges in drug discovery for SARS-CoV-2 due to the epitope-masking effect of the glycocalyx, the high mutagenicity of the spike protein (S protein), the structural mobility of the SARS-CoV-2 stalk (the S2 domain) and the unclear role of heparinoids in the molecular recognition with the ACE2 receptor. Since the start of the pandemic, up to the time of this writing (November 2023), the treatments for SARS-CoV-2 infection and symptoms include a combination of antiviral protease inhibitors Nirmatrelvir and Ritonavir (under the name Paxlovid), Remdesevir and Molnupiravir (nucleoside analogues), and an injectable neutralising monoclonal antibody (Sotrovimab). No S protein:ACE2 inhibitor has been approved for clinical use. Intriguingly, the role of “heparinoids” (i.e. heparin - and its various degrees of sulfonation - as well as heparan sulfate) in the S protein:ACE molecular recognition mechanism was unclear with different findings being reported up until November 2020. To formulate functional hypotheses I investigated the role of different types of heparinoid disaccharides, from fully sulfated heparin to heparan sulfate, against the S protein, highlighting the crucial role of sulphate groups in the binding mechanism. My findings analyses are published in the preprint: “Understanding The Role of Heparinoids on the SARS-CoV-2 Spike Protein through Molecular Dynamics Simulations”. These findings inspired the screening of a large database of compounds for heparin-like matching molecules.

Authors: Ludovico Pipitò, Christopher A. Reynolds, and Giuseppe Deganutti*

Abstract

The pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) resulted in an estimated number of more than 6.8 million deaths. SARS-CoV-2 entry into the cell is mediated by its transmembrane spike glycoprotein (S protein) interacting with the angiotensin-converting enzyme 2 (ACE2) receptor on the human cells’ surface through the receptor binding domain (RBD). The extracellular heparan sulphate (EcHS) and heparin (HP) in the proximity of the ACE2 receptor enhance S protein binding through a poorly understood mechanism. Surprisingly, low molecular weight heparin (LMWH) and heparan sulphate disaccharides (dHS) hinder the S protein binding to ACE2, despite their structural similarity to EcHS and HP. In our study, we highlighted the region-selectivity of SARS-CoV-2 RBD for the highly sulphated glycosaminoglycans (GAGS) while identifying binding sites intended for low or non-sulphated extracellular GAGS.

By using both desulphated heparan disaccharides (dH) and heparin disaccharides (dHP) probes, we identified key-specific S Protein sites intended specifically for high and low-sulphated heparinoids using both mixed molecular dynamics (MixMD) and classic molecular dynamic (cMD) simulations. By mapping the unique interaction areas for each disaccharide on the S protein and the ACE2 receptor we propose the dHP role in hampering the opening of RBD. Furthermore, we identified dH-specific binding areas that suggest their role in the ACE2:RBD alignment.
Our study advances the understanding of the different mechanisms behind the SARS-CoV-2 infection dHS inhibitory effect and the EcHS enhancing mechanism with respect to their sulphonation.

Keywords: SARS-CoV-2, Spike Protein, ACE2, Heparan Sulphate, Heparin; Molecular Dynamics, MixMD

1. Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike protein (S protein) has a strong affinity for the human angiotensin-converting enzyme 2 (ACE2) receptor, a type 1 transmembrane protein responsible for the extracellular conversion of the angiotensin hormone into angiotensin II (1). The S protein (Figure 1A) is a highly glycosylated trimeric structure that is conserved amongst the *coronaviridae* family. Each S protein monomer can be divided into two main domains, S1 and S2 (2). While S1 recognizes ACE2, S2 perforates the membrane of the host cell to transfer the genetic material into the cytoplasm. The receptor-binding domain (RBD) of S1 (residues R319-F541) is responsible for the molecular recognition of ACE2 (Figure 1B) and for triggering conformational changes that initiate the infectious mechanism.

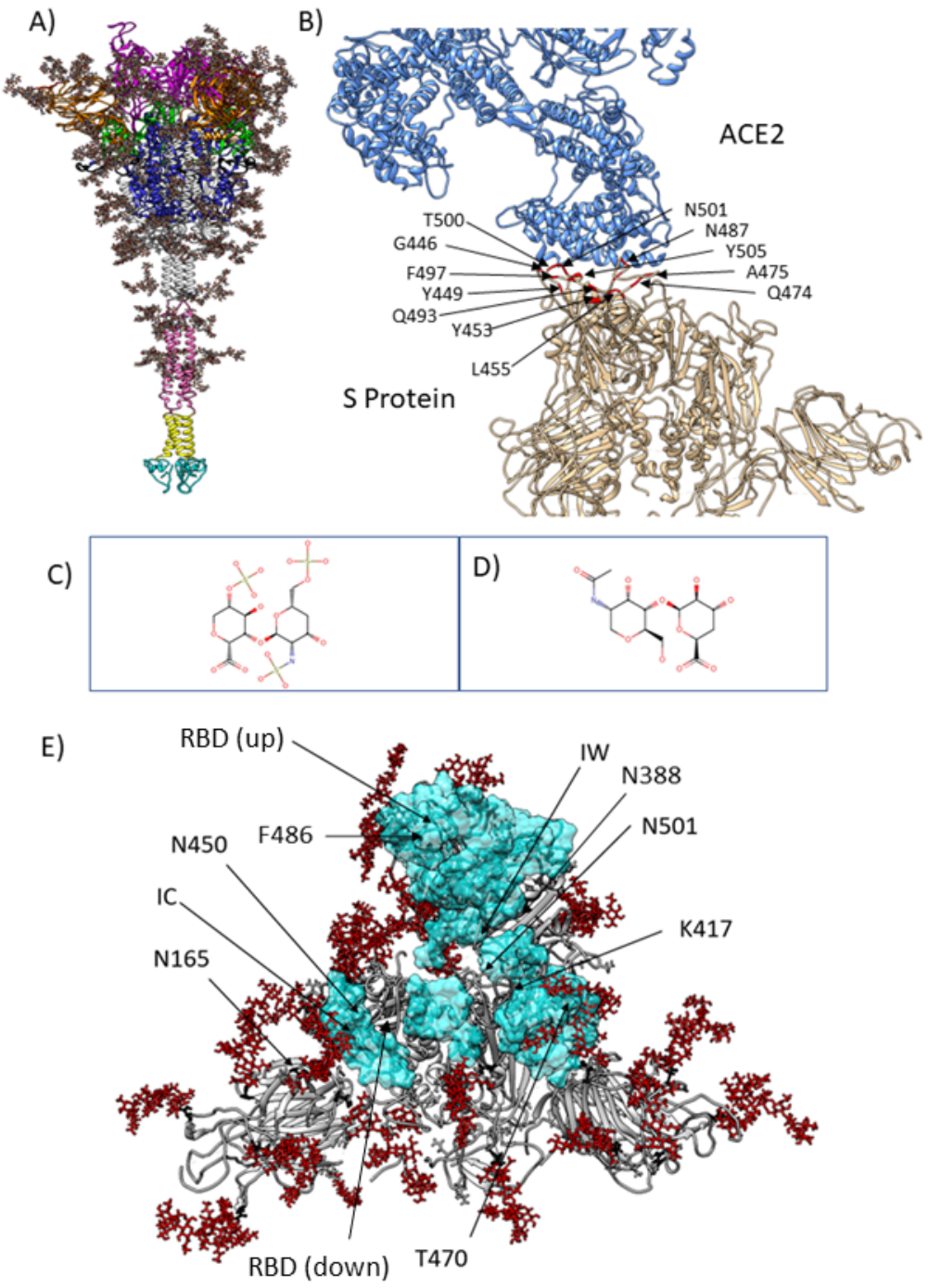


Figure 1. The S protein with its domains and docking results. A) The fully glycosylated S protein with each domain coloured as follows: signal peptide (SP) (red), N-terminal domain (NTD) (orange), receptor binding domain (RBD) (magenta), C-terminal domain (CTD) (green), S1/S2 cleavage site (black), the fusion peptide (FP) (blue), heptad repeat 1 (HR1) (light grey), heptad repeat 2 (HR2) (pink), transmembrane domain (TM) (yellow), and cytosolic domain (CD) (cyan). Glycans are depicted in sticks. (3). **B**) S1 (grey) - ACE2 (blue) complex; RBD "up" residues (red) close to ACE2 are responsible for molecular recognition. **C**) Heparin's most representative disaccharide unit (dHP), consisting of alpha-L-iduronate

2-O-sulphate and glucosamine 2,6-disulphate. **D)** The dH model used for simulations (glucuronic acid and N-acetylglucosamine). dH is also a common unit in HS (4). **E**) Results for Molecular docking dH and dH to S protein. The RBD in the “up” state (RBD up) is covered by overlapping dH and dHP (cyan surface). dH and dHP units were predicted to bind a large area of the S protein (grey ribbon) around residues N388, K417-F486, and N501 of RBD “up” (IW). Both disaccharides were also predicted on residues N165, N450-T470, and N501 (IC) on RBD “down”. The spaces around glycan residues (red stick representation) on N165 and the RBD indicated gaps between the RBD and the NTD where disaccharides were also predicted to bind.

Several S protein structures in the pre-fusion state (Figure 1 A) have been solved through cryo-electron microscopy (cryo-EM) (5, 6), allowing the distinction of two different RBD configurations, named “up” and “down” (7), the former responsible for binding to ACE2. A set of 13 RBD residues (T402, R439, Y436, N440, Y455, N473, Y475, F486, G488, Y491, Q493, Q498, N501) is involved in interactions with ACE2 (8) and constitutes the receptor binding motif (RBM) on the apical portion of RBD (Figure 1B**)** (9).

1.1. Heparin and Heparan Sulfate

Heparin (HP) and heparan sulphate (HS) are similar molecules but can be distinguished by their chemical structure and their degree of sulfonation. HP is a high molecular weight glycosaminoglycan that occurs naturally primarily in the connective tissue of mammals, such as intestinal and lung mucosa, liver, and skin and it is used as an anticoagulant for its antithrombotic effects (10), thus preventing blood from clotting.

HS, on the other hand, is another type of glycosaminoglycan found in animal tissues, such as the basement membrane, connective tissues, and blood plasma, with a high molecular weight and is composed of sugar chains. However, unlike HP, HS has no anticoagulant properties but plays a regulatory role in inflammation (11).

HP consists of a linear chain of repeating disaccharides, each of which is composed of a residue of uronic acid (D-glucuronate or L-iduronate) and N-acetylglucosamine (GlcNAc), with an average of 2.7 units of disaccharide containing at least one sulphate group (12). In contrast, HS has a similar structure but contains a higher percentage of iduronic acid than HP, and around one sulphate group (13).

The degree of sulfonation influences the chemical and biological properties of the molecules with HP being more hydrophilic, negatively charged, and with a greater ability to interact with proteins and clotting factors, inhibiting blood clotting. HS, on the other hand with a lower degree of sulfonation, has a lower anticoagulant capacity.

In summary, HP and HS are two different molecules, although they share some similar structural characteristics. HP has anticoagulant properties, while HS does not. Our aim is to create two precise characterizations in chemical terms: a very electronegative HP with a more important steric encumbrance due to the presence of sulphate groups and a sulfate-free heparan to reduce the "noise" caused by the presence of single functional groups such as heparan sulfate. By generating two different chemical probes, in which one lacks sulphate groups, we will determine the importance of the sulphate constituting the ECM and their role in the Spike and ACE2 alignment. At the same time, highly sulphated units such as HP will be used to detect possible sulphate-dependent overlapping zones across the S protein.

1.2. Glycosaminoglycans play a role in infection

Glycosaminoglycans (GAGs), such as HP and HS (EcHS), play a crucial role in regulating the immune response by regulating cell adhesion, tuning cytokine and chemokine function, and mediating inflammatory reactions (10,14) through HS-binding motifs (HSBM). HP, also produced by basophils and mast cells (15), is constituted by highly sulphated repeating units of 1-4 pyranosyl uronic acid and 2-amino-2-deoxy glucopyranose (Figure 1C) (glucosamine) and it is known for its major role as an anticoagulant when formulated in low molecular weight (LMWH). HP is characterized by a wide structural heterogeneity and high negative charge due to the presence of numerous sulphate groups per disaccharide unit of O-sulfonated, D-glucosamine, and O-sulfonated hexuronic acid (Figure 1D) (16). The high level of sulfonation makes the HP formal charge higher than EcHS (10) and is responsible for LMWH pharmacodynamics properties on the coagulation cascade by binding to antithrombin (13). EcHS is diffusely expressed on the cellular surface in the form of high molecular weight HS proteoglycans (17). EcHS has various roles as an immune regulator (14), and anticoagulant (18), and reportedly assists microbial and viral infections by acting as molecular adhesion receptors in human immunodeficiency virus (HIV).

The role of EcHS in assisting viral infections in general (19) is carried out both by promoting virus-host cell adhesion and by acting as a "glue" between viral proteins and host cell receptors. Schuurs and co-workers suggested an intermediary role of EcHS between the S protein furin cleavage site (20), favouring the membrane fusion mechanism. HP has been proposed to play a role in promoting SARS-Cov-2 infection (21,22), probably by inducing conformational changes upon binding (23) through interactions on the S1/S2 cleavage site (24). Clausen and collaborators (25) experimentally demonstrated that HP allosterically facilitates interactions between S protein and ACE2 and provided a preliminary model of the RBD regions possibly implied in the recognition of both EcHS and HP. Other studies

suggested the intriguing hypothesis that HP and low molecular weight heparins (LMWH) may function as antagonists (26,27) of ACE2 binding by competing for the EcHS binding site on the S protein (28). While existing studies have proposed the role of HP as an RBD:ACE2 intermediator (29,30) proposing HP-like properties to compete with HP (31), to date, much remains unexplored about the totality of the S Protein S1 domain and its interactions between differently sulfonated HP and HS molecules.

To pinpoint possible divergent binding sites on the S protein and ACE2 between HP and EcHS, and inform on why sulphated groups, abundant on HP, play a role in the SARS-CoV-2 binding to ACE2, we present an investigation of the binding profile of these two heparinoids on both S protein and ACE2 receptor by using a mixed molecular dynamics (MixMD) approach (32).

2. Methods

2.1. Structure Preparation and Force Field Settings

All systems were prepared using the CHARMM36 (33,34)/CGenFF 3.0.1 (35,36) force field combination. To speed up the simulations, only the apical portion of the fully glycosylated S protein was kept from the spike protein model 6VSB_1_2_1 available at (https://charmm-gui.org/?doc=archive&lib=covid19) (37). Such regions (residues 1-1003) included the S1 and S2 domains SP (1 - 13), NTD (14-305), RBD (306-541), CTD (542 - 652), S1 / S2 (653 - 686), and FP (687-911). The dHP and dH (Figure 1C, 1D) were designed with the VMD Molefacture plugin (https://www.ks.uiuc.edu/Research/vmd/plugins/molefacture/).

The ACE2 peptidase domain dimer (ACE2-PD) was taken from the PDB entry 6M17 (residues I21-L731) (38). Hydrogen atoms were added to the S protein and ACE2 by Propka (39) at a simulated pH of 7.0, while disulfide bonds were identified by HTMD (40), visually inspected, and patched manually through VMD (41) as previously reported (42). The resulting structure was minimised using the conjugate gradient algorithm by ACEMD (40).

2.2. Molecular Docking

Multiple docking experiments were performed using Autodock Vina (43) on the apical portion of the fully glycosylated S protein S1 domain prepared as reported above. As a reference for docking, the centre of the grid box was located on residue C432 of each monomer in the “up” and “down” state (Figure 1E). The docking was repeated six times by moving the box +5 Å on the z-axis on the RBD “up” to map the whole surface of S1, including residues Q493-Y505. Finally, blind docking was performed on the whole S protein to extend the investigation for possible binding sites to the whole system. We extended the docking area

to compare the set of affinities of all possible binding pockets and, at the same time, to add regions of S1 that had not been sufficiently covered with the previous dockings, including, therefore, the NTD subdomains.

2.3. MixMD system preparation

To map possible S protein and ACE2 sites interacting with HS and HP, each probe was simulated individually using a MixMD approach (44). In total, four systems were prepared using dimeric HP (dHP) or dimeric HS (dHS) (100 molecules in each system) through Packmol (45). A minimum distance of 4 Å was set to avoid clashes and secure an ordered placement of the molecules within a 150 Å x150 Å x150 Å box (46). TIP3P water molecules (47) were added to the simulation box considering a 15 Å water padding by the Solvate plugin 1.5 (http://www.ks.uiuc.edu/Research/vmd/plugins/solvate/). The charge neutrality of the system was achieved by adding Na^+ /Cl^- to the concentration of 0.150 M using the Autoionize plugin 1.3 (http: //www.ks.uiuc. edu/Research/vmd /plugins/autoionize/). ACEMD was used for both equilibration and MD production. The energy of the systems was reduced through 2000 conjugate-gradient minimization (CG) steps to eliminate possible clashes and optimize atomic distances. Equilibration was reached in isothermal-isobaric conditions (NPT) using the Berendsen barostat (48) (target pressure 1 atm) and the Langevin thermostat (49) (target temperature 310 K) during a 4 ns simulation (integration time step 2 fs). During the equilibration, a positional restraint of 1 kcal/ mol $Å^2$ was applied only on the alpha carbons at the base of the S protein, where the structure was cut from the rest of the protein, leaving the glycans and the three RBDs free to move. Two 500 ns productive trajectories were produced with an integration time step of 4 fs, using the hydrogen mass repartition (HMR) (50), in the canonical ensemble (NVT), with the same positional restraints used in the equilibration. The cut-off distance for electrostatic interactions was set at 9 Å, with a switching function applied beyond 7.5 Å. Long-range Coulomb interactions were handled using the particle mesh Ewald summation method (PME) (51) with default ACEMD settings. In summary, two trajectories were produced for ACE2-HP, two for ACE2-HS, two S protein-HP, and two S protein-HS, reassigning the atomic velocities on each replica, for a total of 1000 ns for each system.

2.4. MD trajectories analysis

MD Trajectories were merged using MDtraj (52). The root-mean-square deviation (RMSD) and Root-Mean-Square Fluctuation (RMSF) analysis were computed using VMD (41). Ligand-protein contacts, including hydrogen bonds, were detected using GetContacts (https://getcontacts.github.io), with a threshold distance and angle of 3.5 Å and 120°, respectively. Contacts and hydrogen bonds were expressed as occupancy (% of total MD frames) over the two merged replicas for each system. Molecular graphics images were

produced using UCSF Chimera (53). Volumetric maps were computed using the VMD VolMap plugin with a space grid of 0.25 A, averaging all frames (https://www.ks.uiuc.edu/Research/vmd/vmd-1.9.1/ug/node153.html).

3 Results

Low molecular weight HS (LMWHS) have been proposed as possible antagonists of the ACE2 binding to the S protein (26) thanks to the high degree of sulfonation (22). We tried to clarify the heparinoids' S protein binding profile to better understand this mechanism, focusing on the crucial S protein regions involved in ACE2 molecular recognition (9). Our hypothesis was that the electrostatic interactions driving the binding of EcHS and LMWH to the S protein can be sampled also by shorter disaccharide units (dH and dHP).

3.1. Molecular Docking Suggests a Broad Distribution of HP and HS Interaction Sites

Molecular docking of disaccharide HP (dHP) and HS (dHS) was performed to reproduce the data obtained by Clausen *et al*. (25) using a different protocol. In Clausen's study, tetra-saccharide HP (dp4) units were docked with ClusPro against RBD in both "up" and "down" states, suggesting interactions with residues R346, R355, K444, and R466. Residues F347, S349, N354, G447, Y449, and Y451 were suggested favoring the S protein - ACE2 alignment. Our docking results agree and show a homogeneous distribution of poses on the S protein. Furthermore, our results predicted a broader distribution of both dHP and dH on the RBD "up" conformation, in the vicinity of residues N388 - F486 (Figure 1E) with scores ranging from -7.2 to -8.1 kcal/mol. Docking on RBDs "down" indicated a lesser engagement, with dH and dHP, predicted to bind to a small cleft between NTD and RBD close to residues K150 and T470, i.e. the region between K150 (chain B, "down" state) and T470 (chain A, "up" state (Figure 1E) suggesting a stronger engagement of dHP in respect to dH. Additionally, dH was also predicted to bind the RBD, responsible for the molecular recognition with ACE2, at the level of residues K417- T500 but with the lowest docking score among all poses. These initial results suggest that the disaccharides are likely to lodge into niches at the base of RBD, rather than engaging with the ACE2 binding site directly.

3.2. MD simulation elucidates the flexibility of the S protein and its glycoshield

Although molecular docking predicted a large area of the S protein potentially involved in interactions with dH and dHP, it did not consider the protein flexibility and explicit solvent. To overcome these intrinsic limits, we performed MixMD simulations to sample S protein sites potentially interacting with dH and dHP, in a flexible and fully hydrated environment (Video S1, Video S2). During the MixMD replicas, the RBDs "down" remained stable due to the presence of salt bridges between K378-E988 and K386-D985 in the S2 domain, and E516-K202 in the N-terminal domain (NTD), in accordance with Gur et al (54). The RBD “up” showed overall higher mobility but moderate flexibility in the loop between F485 and Y505 and high flexibility in the I468-Q493 region. The glycan residues bound to N165 intercalated between N427 and E465 at the base of the RBD, preventing it from folding back to the “down” conformation (55). Overall, glycan chains displayed the highest mobility among all the residues in the system. Polysaccharides bound to N165, N234, and N343 restricted the access to the apical portion of S1, while the glycans linked to N165 locked RBD in the “up” state, in agreement with Casalino and co-workers (56). These results indicate that the dynamics of the S protein, combined with the sweeping motion of glycans, impede access to the inner wall of the RBD (residues R346, N354-R357), provide protection to the cavity formed by the junction between the NTD of one chain and the RBD of the adjacent chain (between N165 and N450), and reduce accessibility to residues R403-R408 at the centre of the S1 domain. Areas across the NTD and close to the RBD “up” were more likely to engage dH and dHP in productive contacts due to the lower shielding of glycans.

3.3. Simulations identified specific zones for dHP and dH on the S protein

MixMD simulations against the glycosylated S protein indicated areas involved in interactions with dH or dHP, including the NTD and both RBDs “up” and “down”. dHP and dH engaged the S protein in persistent contacts in areas not covered by the glycan residues' sweeping motion (Video S1).

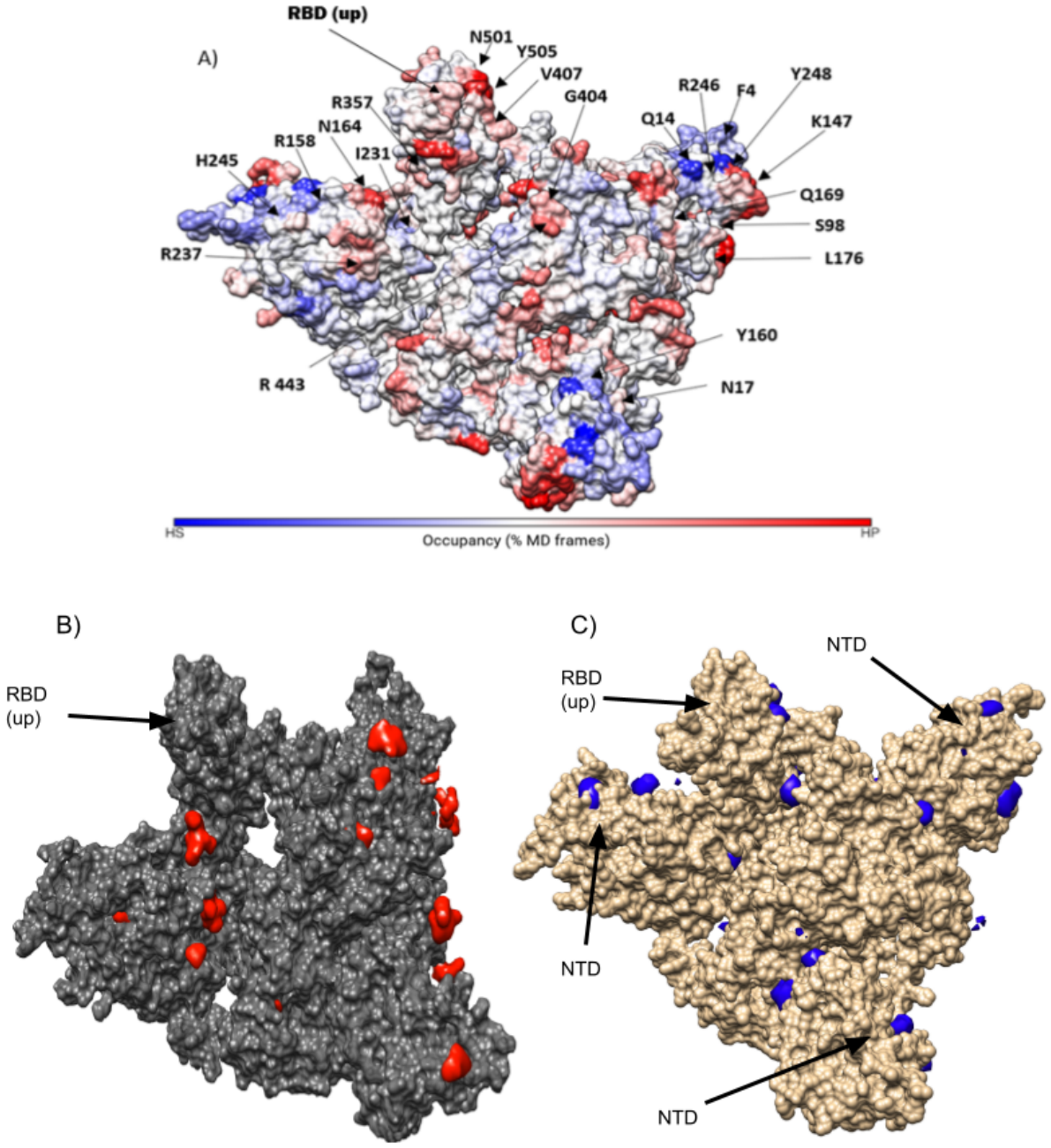


Figure 2. Interactions between disaccharide HS or HP and the S1 domain. A) Differences in S protein contacts with dH or dHP, plotted on the S protein (surface representation) and coloured according to the occupancy prevalence (% MD frames). The polar residues R158, N164, R357, and V407 made more contacts with dHP, while F4, Q14, N17, Q169, and Y248 engaged more dHS, at the "corners" of the S protein, in agreement with Clausen's results. **B**) Volumetric density map of dHP (red, isovalue 20%) showing the averaged atomic positions for the ligand across the simulation box. **C**) Volumetric density map of dH (blue, isovalue 20%); The circles in both images indicated the spots with maximum persistence, around NTD and SP domains, located at the base of S1.

Many persistent contacts occurred between dHP and most of the S protein surface, specifically on residues N164, R355-R357, V407, N501, and Y505, and the inner RBD “down” site on residues I231, G404, N501, and Y505 (Figure 2A). Residues R355-R357 and residues N501-T505 were engaged by dHP in both RBD “up” or “down” conformations suggesting the importance of sulphate groups in the disaccharides’ stabilization. Notably, none of these residues were part of the binding site for ACE2, except for N501-Y505. dHP engaged shortly with R357, G404, V407, N501, and Y505 on the monomer in the open conformation, and more prominently on R237 and N164 on the NTD. dHP formed more persistent interactions than dH with R354-K356, due to the negative charge of its sulphate groups, which favoured the contacts between the arginine-rich chains of the S protein constituting the trimeric system. The volumetric density maps highlight the differences between dHP and dH's most persistent interaction sites (Figure 2B, 2C). dH stationed by the NTD apical region close to the S protein base on residues S98-K147, R237, R246-Y248, R357 and by the RBD “up” residues R355-R357, K378, G404.

Notably, other S protein positions, like residues V407, S477, Q474, and the surrounding areas at the centre of the trimer did not show selective interactions between the S protein and one heparinoid over the other. Our results highlight a selective affinity for HP with respect to HP on residues R346, N354, R355, K356, R357, R466, and K444 in agreement with experimental nuclear magnetic resonance (NMR) results (57), indicating a sulfonation-dependent selectivity on the S protein. Furthermore, this selectivity enforces the idea that negatively charged chains of HP find a multi-modal lodge around R355 which triggers the opening of the RBD, without interfering in the ACE2:S protein molecular recognition.

In summary, dH and dHP probes clearly identified three domain-specific sites of interaction along the S protein surface. dHP resided specifically along the RBD in the “up” state, while dH mapped the corners of the S protein on the NTD domain. We speculate that the S protein might exhibit a domain-specific interaction preference, where LMWH is stabilized mostly on the RBD, while EcHS are more specific for the NTD. The protruding portion of the NTD may represent a possible engagement point for EcHS, providing support in the alignment of the S protein and ACE2, while HP could support the RBD “up” conformation without interfering with the ACE2 interaction. The contacts between G502 and Y505 would provide interaction sites for HP to support the opening of RBD in the “up” conformation, in agreement with Clausen’s experimental results. Finally, we identified specific interaction hotspots intended for dH on the NTD. dH probes, used in this study because considered representative of EcHS interactions, would aid the alignment between the S1 domain and ACE2. We speculate that SARS-2 might distinguish between EcHS and HP thanks to their different sulfonation state

and a specific molecular recognition fingerprint for the RBD, therefore using HP to lock the RBD in its “up” state and HS for a favourable alignment on top of ACE2.

3.4. Simulations identified specific zones for dHP and dH on ACE2

We focused on ACE2 to pinpoint possible interaction sites for HP and EcHS, involved in the enhanced binding between ACE2 and S protein (58,59). During MixMD simulations, the ACE2 pincer-shaped domains underwent an outward opening movement (Video S3, Video S4) and exhibited flexibility at the PD domain level on chain D.

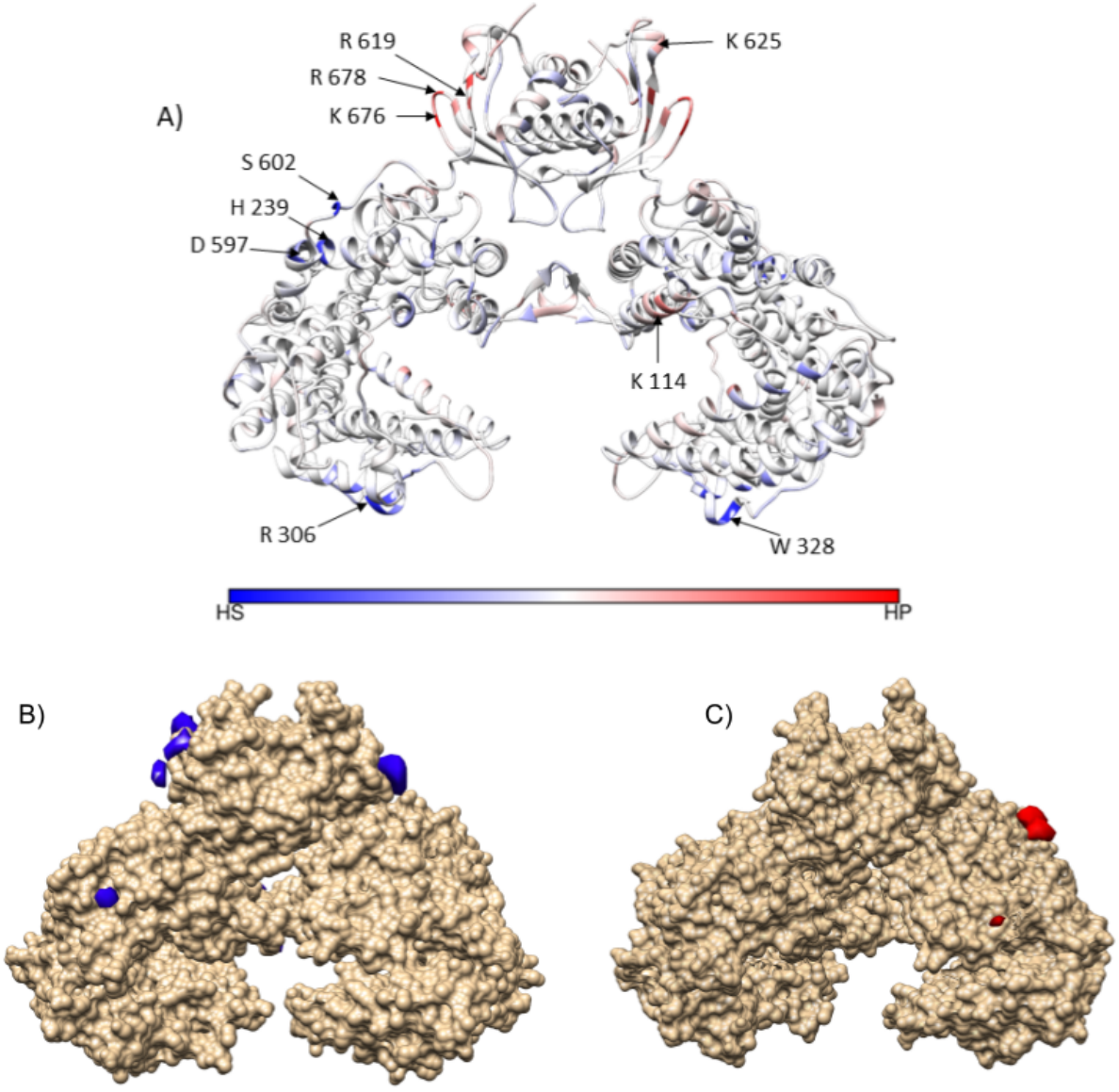


Figure 3. Interactions between dHP or dH and ACE2. A) Contact frequency difference between dH and dHP according to the frequency of interactions during the MixMD

simulations. **B)** ACE2-dH density map (blue, isovalue 20%). **C**) dH density map (red, isovalue 20%) on ACE2.

Overall, the higher sulfonation of dHP increased electrostatic interactions with lysine and arginine residues indicating that dHP is more likely to engage with ACE, especially at the base of the extracellular domain. Interactions were predominant around residues K625, K676-R678 (Figure 3A), and G726. dH MixMD simulations indicated lower distribution around ACE2 (Figure 3B) suggesting a lower engagement in the S protein-dHS-ACE2.
From a mechanistic perspective, we propose that HP could bind the ACE2 while engaging the S protein without hindering the S protein: ACE2 molecular recognition mechanism. dH engaged ACE2 on residues H239, R306, W328, D597, and S602 (Figure 3C).

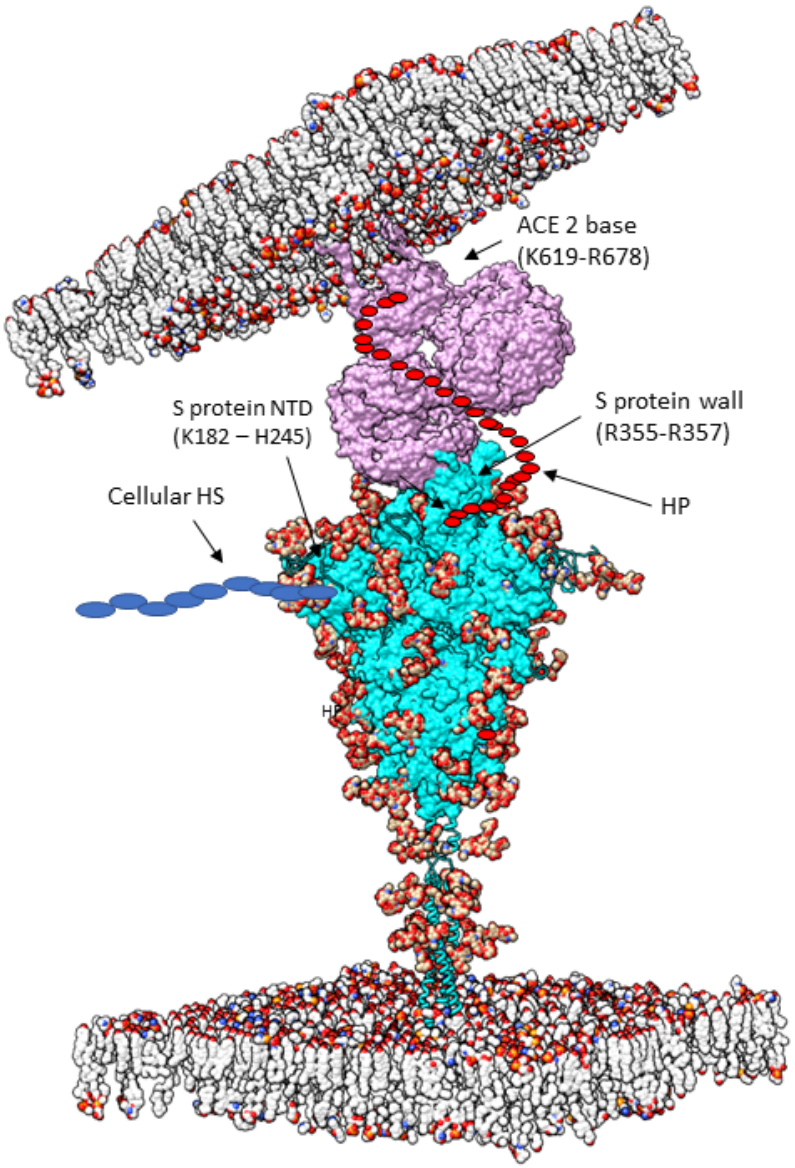


Figure 4. The proposed interaction model for EcHS and HP on the S protein-ACE2 complex. The presence of EcHS in the glycocalyx on the host's extracellular matrix could be exploited by SARS-2 to better anchor and align the virion to the host cell, without interfering with the receptor binding motif (RBM). Long EcHS chains could also act as three-point adhesives, enhancing the SARS-2 infective mechanism. HP could lock the RBD in the "up" state without affecting the molecular recognition with ACE2. (Figure modified with permission from (60)).

Our results indicated that the most persistent contacts between dHP and ACE2 occurred at the base of the receptor. According to the distinct interaction specificity of both dHP and dHS, we speculate that cellular EcHS could bind to the corners of the S protein to facilitate a favourable orientation of the trimer toward ACE2 around residues K182-H245. At the same time, long EcHS chains could favour the molecular recognition between ACE2 and the S protein by supporting the "up" state of the RBD. We hypothesize the synergic interaction of long HP chains with the extracellular ACE2 domain around residues K619-R678.
In summary, simulations suggested that long-chained heparinoids bound to the S protein can extend toward the base of the ACE2, providing an additional anchoring point to facilitate viral binding without occupying the site intended for RBD molecular recognition.

4. Discussion

We have identified potential sites on the S protein and ACE2 receptor where EcHS and HP could bind and compete. Our results indicate that both heparinoids can linger between the RBD external wall, around R355, and the adjacent NTD close to N165. dH formed more contacts at the corners of each S protein monomer NTD domain, including a portion of the SP. Strong similarities between the interaction heatmaps of dH and dHP indicate that SARS-CoV-2 could exploit both HP and EcHS, regardless of their sulfonation state, to initially approach the host's cellular membrane by using the proteoglycans-bound EcHS, while using the highly sulphated heparin to initiate the conformational changes. The presence of restricted S protein areas selective for either of the two heparinoids suggests that the long HP chain could unwind from K148, pass close to N165, and stabilize the RBD "up" in correspondence with N165 and between R355-R357 (Figure 4) using their high sulphonation degree as a region-selective means.
We propose the hypothesis where the S protein initially exploits both HP and EcHS on residues R346, R355-R377, G404, R408, K444, and G502-G504, to favor the initial RBD transition to the "up" state with HP specifically supporting the transition between the states in

a more efficient manner. Hence, the role of heparin would be to favour, together with the glycan on N165, the opening and stabilization of RBD towards the "up" state.

On the other hand, SARS-CoV-2 could use low-sulphated EcHS to approach the membrane through the residues F4, Q14, N17, H245, R246, and Y248 to favour a proper S1:ACE2 alignment. Our hypothesis sees EcHS as a landing “hook” for the S protein NTD orientation of the RBD toward ACE2, as indicated by the numerous contacts on multiple residues on NTD. In this scenario, the HP present on the ECM could additionally stabilize and anchor the S protein to ACE2, as indicated by the contacts on both the S protein and the extracellular portion of ACE2, suggesting a synergic effect of EcHS and HP, in agreement with Clausen’s experimental results.

We speculate that long EcHS chains projected towards the extracellular lumen could intercept the S protein through electrostatic interaction with the less sulphated disaccharide chains of EcHS on residues V407, S477 and Q474 where the sulphonation profile was not decisive. This approaching mechanism, in conjunction with HP RBD stabilisation, might regulate alignment and affinity with the subsequent conformational changes, in agreement with Tandon’s group binding experimental data (61).

It is reasonable to assume that the inhibition experimentally displayed by dHP and dH is due to the ability of short chains to saturate all the binding sites for EcHS and HP onto the S protein and ACE2 surface, preventing longer HP chains from binding and facilitate the conformational shift of RBD from “down” to “up”, or hindering the alignment with ACE2.

5. Conclusion:

Using dH and dHP as molecular probes, we suggest both common zones and specific sites of the glycosylated S protein and ACE2 receptor where HP and EcHS can potentially engage. dH is proposed to bind areas around the NTD responsible for the molecular recognition of the extracellular matrix, hampering the correct S protein-ACE2 alignment and altering S protein stabilization in the proximity of the membrane. dHP interacts on the RBD, preventing the opening and locking of the ACE-2 binding domain, reducing SARS-CoV-2 infectivity. Our results support the scenario in which short-chain HP and HS could saturate S protein binding spots RBD for glycocalyx EcHS on the NTD and RBD, in agreement with the experimental observations available in the literature.

In summary, we distinguished the different behaviour of the dH and dHP, suggesting that HP plays an opening and stabilizing role on the RBD, while low-sulphated heparan is responsible for the RBD:ACE2 alignment, addressing the reasons behind the different behaviour of similar disaccharides to their sulphonation profile.

6. Videos:

https://www.biorxiv.org/content/10.1101/2022.07.05.498807v1.supplementary-material

References

1. Tikellis C, Thomas MC. Angiotensin-Converting Enzyme 2 (ACE2) Is a Key Modulator of the Renin Angiotensin System in Health and Disease. Int J Pept. 2012 Mar 20;2012:256294.
2. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science. 2020 Mar 13;367 (6483):1260–3.
3. Zhou T, Tsybovsky Y, Olia AS, Gorman J, Rapp MA, Cerutti G, et al. A pH-dependent switch mediates conformational masking of SARS-CoV-2 spike. BioRxiv. 2020 Jul 4;
4. Rabenstein DL. Heparin and heparan sulfate: structure and function. Nat Prod Rep. 2002 May 23;19 (3):312–31.
5. Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature. 2020 May;581 (7807):215–20.
6. Pipitò L, Rujan RM, Reynolds CA, Deganutti G. Molecular dynamics simulations reveal the importance of conformational changes, glycosylation, mutations, and drug repurposing for the Sars-Cov-2 spike protein. 2022 Mar 3;
7. Cai Y, Zhang J, Xiao T, Peng H, Sterling SM, Walsh RM, et al. Distinct conformational states of SARS-CoV-2 spike protein. Science. 2020 Sep 25;369 (6511):1586–92.
8. Walls AC, Tortorici MA, Frenz B, Snijder J, Li W, Rey FA, et al. Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. Nat Struct Mol Biol. 2016 Oct;23 (10):899–905.
9. Deganutti G, Prischi F, Reynolds CA. Supervised molecular dynamics for exploring the druggability of the SARS-CoV-2 spike protein. J Comput Aided Mol Des. 2021 Feb;35 (2):195–207.
10. Meneghetti MCZ, Hughes AJ, Rudd TR, Nader HB, Powell AK, Yates EA, et al. Heparan sulfate and heparin interactions with proteins. J R Soc Interface. 2015 Sep 6;12 (110):0589.
11. Parish CR. The role of heparan sulphate in inflammation. Nat Rev Immunol. 2006 Sep;6 (9):633–43.
12. Toida T, Yoshida H, Toyoda H, Koshiishi I, Imanari T, Hileman RE, et al. Structural differences and the presence of unsubstituted amino groups in heparan sulphates from different tissues and species. Biochem J. 1997 Mar 1;322 (Pt 2) (Pt 2):499–506.

13. Shriver Z, Capila I, Venkataraman G, Sasisekharan R. Heparin and heparan sulfate: analyzing structure and microheterogeneity. Handb Exp Pharmacol. 2012; (207):159–76.

14. Simon Davis DA, Parish CR. Heparan sulfate: a ubiquitous glycosaminoglycan with multiple roles in immunity. Front Immunol. 2013 Dec 18;4:470.

15. Lim GB. Milestone 1: Discovery and purification of heparin. Nat Rev Cardiol. 2017 Dec 14;

16. Capila I, Linhardt RJ. Heparin-protein interactions. Angew Chem Int Ed. 2002 Feb 1;41 (3):391–412.

17. Kreuger J, Kjellén L. Heparan sulfate biosynthesis: regulation and variability. J Histochem Cytochem. 2012 Dec;60 (12):898–907.

18. Liu J, Pedersen LC. Anticoagulant heparan sulfate: structural specificity and biosynthesis. Appl Microbiol Biotechnol. 2007 Feb;74 (2):263–72.

19. Liu J, Thorp SC. Cell surface heparan sulfate and its roles in assisting viral infections. Med Res Rev. 2002 Jan;22 (1):1–25.

20. Schuurs ZP, Hammond E, Elli S, Rudd TR, Mycroft-West CJ, Lima MA, et al. Evidence of a putative glycosaminoglycan binding site on the glycosylated SARS-CoV-2 spike protein N-terminal domain. Comput Struct Biotechnol J. 2021 May 4;19:2806–18.

21. Zamorano Cuervo N, Grandvaux N. ACE2: Evidence of role as entry receptor for SARS-CoV-2 and implications in comorbidities. eLife. 2020 Nov 9;9.

22. Yue J, Jin W, Yang H, Faulkner J, Song X, Qiu H, et al. Heparan Sulfate Facilitates Spike Protein-Mediated SARS-CoV-2 Host Cell Invasion and Contributes to Increased Infection of SARS-CoV-2 G614 Mutant and in Lung Cancer. Front Mol Biosci. 2021 Jun 11;8:649575.

23. Mycroft-West C, Su D, Elli S, Guimond S, Miller G, Turnbull J, et al. The 2019 coronavirus (SARS-CoV-2) surface protein (Spike) S1 Receptor Binding Domain undergoes conformational change upon heparin binding. BioRxiv. 2020 Mar 2;

24. Kim SY, Jin W, Sood A, Montgomery DW, Grant OC, Fuster MM, et al. Glycosaminoglycan binding motif at S1/S2 proteolytic cleavage site on spike glycoprotein may facilitate novel coronavirus (SARS-CoV-2) host cell entry. BioRxiv. 2020 Apr 15;

25. Clausen TM, Sandoval DR, Spliid CB, Pihl J, Perrett HR, Painter CD, et al.

SARS-CoV-2 Infection Depends on Cellular Heparan Sulfate and ACE2. Cell. 2020 Nov 12;183 (4):1043-1057.e15.

26. Unable to find information for 9239468.

27. Kwon PS, Oh H, Kwon S-J, Jin W, Zhang F, Fraser K, et al. Sulfated polysaccharides effectively inhibit SARS-CoV-2 in vitro. Cell Discov. 2020 Jul 24;6 (1):50.

28. Mycroft-West C, Su D, Pagani I, Rudd T, Elli S, Guimond S, et al. Heparin inhibits cellular invasion by SARS-CoV-2: structural dependence of the interaction of the surface protein (spike) S1 receptor binding domain with heparin. BioRxiv. 2020 Apr 28;

29. Wang X, Bie L, Gao J. Structural Insights into the Cofactor Role of Heparin/Heparan Sulfate in Binding between the SARS-CoV-2 Spike Protein and Host Angiotensin-Converting Enzyme II. J Chem Inf Model. 2022 Feb 14;62 (3):656–67.

30. Gupta Y, Maciorowski D, Zak SE, Kulkarni CV, Herbert AS, Durvasula R, et al. Heparin: A simplistic repurposing to prevent SARS-CoV-2 transmission in light of its in-vitro nanomolar efficacy. Int J Biol Macromol. 2021 Jul 31;183:203–12.

31. Sreekumar S, M Kuthe A, Chandra Tripathi S, C Patil G, Ravikumar C. Integrated computational approach towards identification of HSPG and ACE2 mimicking moieties for SARS-CoV-2 inhibition. J Mol Liq. 2022 Dec 1;367:120566.

32. Graham SE, Leja N, Carlson HA. MixMD Probeview: Robust Binding Site Prediction from Cosolvent Simulations. J Chem Inf Model. 2018 Jul 23;58 (7):1426–33.

33. Huang J, MacKerell AD. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. J Comput Chem. 2013 Sep 30;34 (25):2135–45.

34. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods. 2017 Jan;14 (1):71–3.

35. Vanommeslaeghe K, MacKerell AD. Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing. J Chem Inf Model. 2012 Dec 21;52 (12):3144–54.

36. Vanommeslaeghe K, Raman EP, MacKerell AD. Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. J Chem Inf Model. 2012 Dec 21;52 (12):3155–68.

37. Woo H, Park S-J, Choi YK, Park T, Tanveer M, Cao Y, et al. Developing a Fully Glycosylated Full-Length SARS-CoV-2 Spike Protein Model in a Viral Membrane. J

Phys Chem B. 2020 Aug 20;124 (33):7128–37.

38. Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. Science. 2020 Mar 27;367 (6485):1444–8.

39. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK Predictions. J Chem Theory Comput. 2011 Feb 8;7 (2):525–37.

40. Doerr S, Harvey MJ, Noé F, De Fabritiis G. HTMD: High-Throughput Molecular Dynamics for Molecular Discovery. J Chem Theory Comput. 2016 Apr 12;12 (4):1845–52.

41. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996 Feb;14 (1):33–8, 27.

42. Fabri L, Maruta H, Muramatsu H, Muramatsu T, Simpson RJ, Burgess AW, et al. Structural characterisation of native and recombinant forms of the neurotrophic cytokine MK. J Chromatogr. 1993 Aug 27;646 (1):213–25.

43. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010 Jan 30;31 (2):455–61.

44. Ghanakota P, Carlson HA. Moving Beyond Active-Site Detection: MixMD Applied to Allosteric Systems. J Phys Chem B. 2016 Aug 25;120 (33):8685–95.

45. Martínez L, Andrade R, Birgin EG, Martínez JM. PACKMOL: a package for building initial configurations for molecular dynamics simulations. J Comput Chem. 2009 Oct;30 (13):2157–64.

46. Martínez JM, Martínez L. Packing optimization for automated generation of complex system's initial configurations for molecular dynamics and docking. J Comput Chem. 2003 May;24 (7):819–25.

47. 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.

48. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys. 1984;81 (8):3684.

49. Loncharich RJ, Brooks BR, Pastor RW. Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. Biopolymers. 1992 May;32 (5):523–35.

50. Hopkins CW, Le Grand S, Walker RC, Roitberg AE. Long-Time-Step Molecular

Dynamics through Hydrogen Mass Repartitioning. J Chem Theory Comput. 2015 Apr 14;11 (4):1864–74.

51. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. J Chem Phys. 1995;103 (19):8577.

52. McGibbon RT, Beauchamp KA, Harrigan MP, Klein C, Swails JM, Hernández CX, et al. Mdtraj: A modern open library for the analysis of molecular dynamics trajectories. Biophys J. 2015 Oct 20;109 (8):1528–32.

53. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem. 2004 Oct;25 (13):1605–12.

54. Gur M, Taka E, Yilmaz SZ, Kilinc C, Aktas U, Golcuk M. Conformational transition of SARS-CoV-2 spike glycoprotein between its closed and open states. J Chem Phys. 2020 Aug 21;153 (7):075101.

55. Sztain T, Ahn S-H, Bogetti AT, Casalino L, Goldsmith JA, Seitz E, et al. A glycan gate controls opening of the SARS-CoV-2 spike protein. Nat Chem. 2021 Oct;13 (10):963–8.

56. Casalino L, Gaieb Z, Goldsmith JA, Hjorth CK, Dommer AC, Harbison AM, et al. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. ACS Cent Sci. 2020 Oct 28;6 (10):1722–34.

57. Parafioriti M, Ni M, Petitou M, Mycroft-West CJ, Rudd TR, Gandhi NS, et al. Evidence for Multiple Binding Modes in the Initial Contact Between SARS-CoV-2 Spike S1 Protein and Cell Surface Glycans. Chem Eur J. 2023 Jan 2;29 (1):e202202599.

58. Ali A, Vijayan R. Dynamics of the ACE2-SARS-CoV-2/SARS-CoV spike protein interface reveal unique mechanisms. Sci Rep. 2020 Aug 26;10 (1):14214.

59. Xu C, Wang Y, Liu C, Zhang C, Han W, Hong X, et al. Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM. Sci Adv. 2021 Jan 1;7 (1).

60. Pipitò L, Rujan R-M, Reynolds CA, Deganutti G. Molecular dynamics studies reveal structural and functional features of the SARS-CoV-2 spike protein. Bioessays. 2022 Sep;44 (9):e2200060.

61. Tandon R, Sharp JS, Zhang F, Pomin VH, Ashpole NM, Mitra D, et al. Effective Inhibition of SARS-CoV-2 Entry by Heparin and Enoxaparin Derivatives. J Virol. 2021 Jan 13;95 (3).

6.3 Is the SARS-CoV-2 Spike Protein Stalk Druggable?

The mutations of the RBD of the S protein led to the diffusion of different SARS-CoV-2 variants. These variants, named after the Greek alphabet letters, highlighted the criticalities of the antibody discovery approach, due to epitope changes caused by the mutations. My efforts primarily targeted the most diffused and concerning variants, namely the Alpha (B1.1.1.7), Delta (B.1.617.2) and Omicron (B.1.1.529). Epitope mutation is a prominent escape mechanism and represents an urgent challenge for the development of infective treatments (see the HIV section, Chapter 7). Identifying conserved structures between the different variants to target represented an intriguing approach for drug discovery. The structural comparison of the different SARS-CoV-2 variants inspired a new study for targeting conserved regions of SARS-CoV-2. Sequence and structural alignment of Alfa, Delta, and Omicron revealed conserved regions at the base of the S2/S1 domain and along the stalk (S2 domain) shared across the “variants of concern”. The S2 domain is responsible for the S protein’s flexibility mediated by three hinges. This flexibility explains why multiple spikes simultaneously engage onto the surface of a host cell, providing an effective S protein:ACE2 alignment. Identifying S2 binders would possibly lead to the development of a pan-coronavirus treatment to reduce the stalk’s fexibility and it’s ability to align to ACE2. For this purpose, I performed the virtual screening of a small but optimised library of fragments. These results defined the opportunities and limits of targeting SARS-CoV-2 flexible and dynamic structures. The published manuscript that collects the results and observations is titled “Is the Stalk of the SARS-CoV-2 Spike Protein Druggable?” published on Viruses 2022, 14 (12), 2789; https://doi.org/10.3390/v14122789.

Authors: Ludovico Pipitò, Christopher A Reynolds, Giuseppe Deganutti

Abstract

The SARS-CoV-2 virus spike protein (SP) is the vector of the virus infectivity. The high propensity to mutate in key regions responsible for the recognition of the human angiotensin-converting enzyme 2 (hACE2) or the antibodies produced by the immune system following infection or vaccination makes subunit 1 of the SP a difficult to target and, to date, efforts have not delivered any ACE2 binding inhibitor yet. The inherent flexibility of the stalk region within subunit S2 is key to SARS-CoV-2 high infectivity because it facilitates the receptor binding domain encounter with ACE2. Thus, it could be a valuable therapeutic target. By employing a fragment-based strategy, we computationally studied the druggability of the conserved part of the SP stalk by means of an integrated approach that combines molecular docking with high-throughput molecular dynamics simulations. Our results suggest

that the druggability of the stalk is challenging and provide the structural basis for such difficulty.

Introduction

The severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2) pandemic that was first identified in 2019 in the city of Wuhan (1) continues to raise concerns amongst governments and the scientific community with almost 530 million cases around the world with more than 6.2 million certified deaths (WHO dashboard, 08 June 2022, https://covid19.who.int). SARS-CoV-2 shows a strong affinity for the human angiotensin-converting-enzyme 2 (ACE2) receptor, a type 1 transmembrane protein responsible for the extracellular conversion of the angiotensin hormone into angiotensin II (2) through the SARS-CoV-2 spike protein (SP). The SP is a highly glycosylated trimeric structure, common amongst the coronaviridae family (3), which is constituted by two main units named S1 and S2. While S1 is responsible for molecular recognition of ACE2, S2 is paramount to structural stability and orientation of the whole SP and membrane fusion to deliver the viral genome (4). The efforts of the scientific community were dedicated to promptly developing vaccines or a variety of small molecules(5–8), specifically designed to bind and neutralize the area on S1 responsible for ACE2 binding, namely the receptor-binding domain (RBD).

In 2020, the term variants of concern (VOC) was introduced to describe new SARS-CoV-2 strains which differentiated from the original SARS-CoV-2 wild type (WT) through a series of mutations, mainly on the RBD, which cause drastic changes in transmissibility and pathogenicity (9–11). The B.1.617.2 strain (Delta variant) was identified in India by January 2021 and spread rapidly across the globe (12), overcoming the WT in a short amount of time. n November 2021 the B.1.1.529 (Omicron variant) became the dominant VOC over the Delta (13).

Among the SARS-CoV-2 VOCs, major preoccupations regarded those strains that carried important mutations and deletions, especially on the RBD (14). VOC has important RBD mutations: B.1.1.7 (Alpha), carries E484K, N501Y, D614G, P681H; B.1.351 (Beta) carries K417N, E484K, N501Y, D614G, A701V; P1 (Gamma) carries K417T, E484K, N501Y, D614G, H655Y; B.1.617.2 (Delta) carries L452R, T478K, D614G, P681R (15). Concerns among the scientific community have risen due to their potential to elude the immune system and overcome vaccine protection (16,17) despite showing an overall similarity between variants, which diverged only in terms of flexibility of SP. More recently, a new B.1.1.529 (Omicron) VOC (18) carrying N440K, G446S, S477N, 118 T478K, E484A, Q493R, G496S, Q498R, N501Y, and Y505H mutations, and its lineages became predominant over the Delta variant, possibly due to a more rapid entry or different mechanism (19–21), an enhanced ability to

evade the immune system(22–24), and its increased affinity for ACE2 (25–27) although showing a milder pathogenic impact (28). New VOCs are expected to pose a new threat should they become widespread (22,29) and further studies should follow to evaluate the potential risk of new mutations.

The vaccine technology developed so far is designed to specifically target RBD, where the majority of the mutation occurred, increasing the risk for antibody inefficacy (30–32). The potential loss of efficacy against the Omicron variant was attenuated by a loss of replication and lethality power, probably due to Omicron's inefficiency to exploit the cellular transmembrane protease serine 2 (TMPRSS2) (28). However, with the continuous viral diffusion, the likelihood of new mutations remains critical and new variant-specific vaccines need to be developed regularly to keep up with the rate of mutation (33). While S1 and RBD are the SP domains most prone to mutation, S2 has a higher level of conservation among the coronaviridae family (34). The only S2 mutations identified so far are N764K, D796Y, N856K, Q954H, N969K, and L981F. Residues I921, S980, V1187, F1218, and I1219 (Figure 1), conserved in both the Delta and Omicron strains, are pivotal residues that confer increased flexibility to the stalk(35,36). Interestingly, the region between residues L1145-L1186 (Conserved Region 1, Figure1), and between E1188-W1217 (Conserved Region 2, Figure1) contain Loop 1 and Loop2 that contribute to the S1 domain flexibility (37), are conserved amongst all the VOC, and exhibit specific highly-conserved sequences (38). Molecular dynamics (MD) simulations highlighted unexpected flexibility of the SP stalk (36), which has been proposed as paramount for ACE2 binding and infectivity (39). In principle, a small molecule able to target the stalk region would be effective on all the VOC by impairing the flexibility of the SP, thus the infectivity of SARS-CoV-2. For this reason, we integrated molecular docking and molecular dynamics (cMD) simulations to study the druggability of the conserved S2 stalk region E1144-R1185 (Figure 1B) and its potential as a drug target. Possible cryptic binding sites were sampled using a mixed MD (mixMD) approach(31,40,41) to evaluate the accessibility of the stalk in the presence of the branched glycans on the SP surface. We docked a small library of optimized fragments (32) to the SP stalk and performed hundreds of high-throughput post-docking cMD simulations combined with binding free-energy calculations to determine if small molecules can target this important SARS-CoV-2 protein domain.

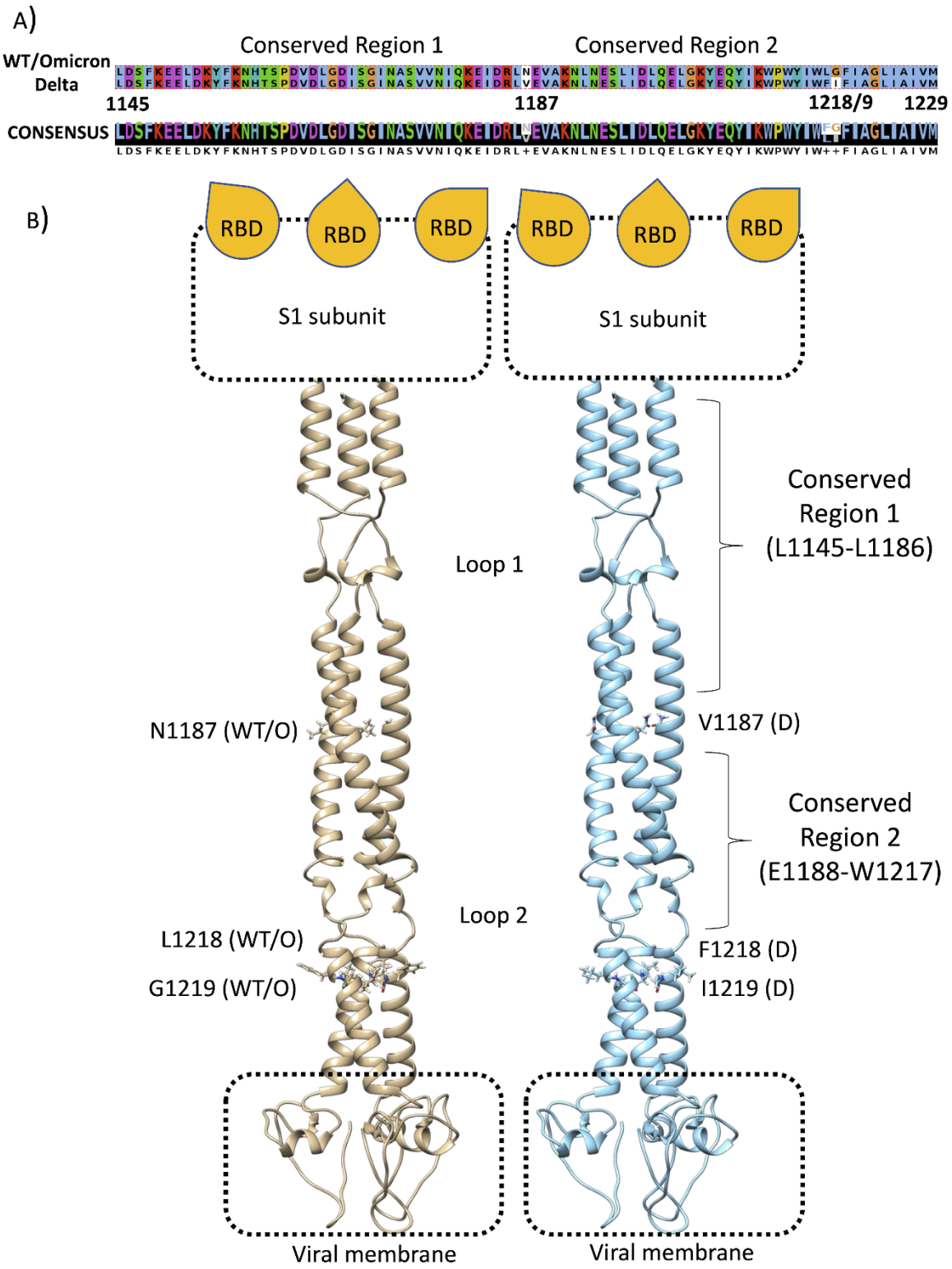

A)
Conserved Region 1
Conserved Region 2
WT/Omicron
Delta
1145
1187
1218/9
1229
CONSENSUS
B)
RBD
RBD
RBD
RBD
RBD
RBD
S1 subunit
S1 subunit
Loop 1
Conserved
Region 1
(L1145-L1186)
N1187 (WT/O)
V1187 (D)
Conserved
Region 2
(E1188-W1217)
Loop 2
L1218 (WT/O)
G1219 (WT/O)
F1218 (D)
I1219 (D)
Viral membrane
Viral membrane

Figure 1T. A) Sequence alignment between the conserved SP stalk region of the wild type (WT) or Omicron and Delta variants. B) Structural comparison between WT or Omicron (tan ribbon) and Delta S2 region (cyan). The area between residues L1145-W1217 is almost identical between the strains and could represent a therapeutic target. Glycans were removed for clarity; the S1 subunit with the three receptor binding domains (RBDs)and the viral, membrane are schematically represented.

Methods

General Workflow

MD simulations and molecular docking were combined in a computational pipeline (Figure 2) aimed to discern potential fragments able to overcome the steric barrier provided by glycosylation and engage the SP stalk in specific interactions. Initial MD simulations of the stalk prepared the structure for molecular docking, while mixMD sampled the accessibility of potential pockets. Molecular docking followed by high throughput post docking MD simulations discerned the stability of the predicted poses, narrowing the number of putative binders to five, which were further evaluated in further, extended MD and mixMD simulations.

Preliminary classic MD of the spike protein stalk

The SP stalk was prepared using CHARMM36 (42,43). The fully glycosylated SP model was retrieved from the CHARMM-GUI repository (https://charmm-gui.org/?doc=archive&lib=covid19), (44), and subsequentially trimmed from residue E1144-W1214, keeping the glycans in their original position at their original length. Hydrogen atoms in the S2 domain were added by Propka (45) at a simulated pH of 7.0, while structural integrity was checked through HTMD (46), visually inspected, and patched manually through VMD (47) according to previous structural knowledge (4). Each system was solvated with TIP3P water molecules (48) added to the simulation box considering a 15 Å padding in every direction by Solvate plugin 1.5 (http://www.ks.uiuc.edu/Research/vmd/plugins/solvate/). The charge neutrality was achieved by adding Na^+/Cl^- to the concentration of 0.150 M using Autoionize plugin 1.3 (http: //www.ks.uiuc. edu/Research/vmd /plugins/autoionize/). The initial geometry and internal energy were optimized using the conjugate gradient algorithm by ACEMD (49)to eliminate possible clashes and optimize atomic distances. The equilibration was achieved in isothermal-isobaric conditions (NPT) using the Berendsen barostat (50) (target pressure 1 atm) and the Langevin thermostat (51) (target temperature 300 K) with low damping of 1 ps^{-1}. During the 4 ns equilibration, a positional restraint of 1 kcal/ mol $Å^2$ was applied on the alpha carbons for the first 3 ns, and on protein side chains for the first 2 ns.

Fragments preparation and molecular docking

A set of 240 optimized molecular fragments from X-ray complexes (32), the SpotXplorer database, was converted to 3D conformers through the RDkit module (52), protonated at pH 7.4 with Chimera (53), and energy minimized with RDkit. Each fragment was docked using Autodock Vina(54,55), to the Conserved Region 1 of the SP stalk using the structure from the last frame of the cMD equilibration and the residue V1164 as the center of a grid with a 46 Å side length, for a broad exploration of the stalk surface. For each fragment, ten poses were ranked according to the docking score. Poses in contact with glycans or outside the conserved region were discarded. The rationale behind our selection was to narrow our list of fragments to those able to specifically target protein residues of the stalk. Poses away from the density maps produced by benzene (BENZ), formic acid (FAC), and methylamine (MTA) (see below) or not engaging simultaneously with at least two monomers, were also excluded.

Post-docking cMD

The best 559 docking complexes were subjected to 10 ns of post docking cMD. Initial CGenFF force field(56,57) topology and parameter files of molecular fragments were obtained from the CGenFF software. Restrained electrostatic potential (RESP) charges were calculated with AmberTools20 (58) after geometry optimization through Gaussian09 at the HF/6-31G* level of theory (59). Each complex was prepared for cMD, equilibrated, and simulated as reported below. For each simulation, similarly to Sabbadin *et al (60)*, we calculated the dynamic energy score (DES, Equation 1), which is the sum, over all the MD frames, of the ratio between the generalized born surface area GBSA binding energy and fragment root mean square deviation (RMSD) to the initial docking pose, using AmberTools20 and VMD.

$$DES = \sum_{i=1}^{n} \frac{GBSA}{RMSD}$$ **Equation 1**

We excluded all the complexes with an average RMSD < 10 Å, and RMSD standard deviation < 5 Å, and ranked them according to the best final score. We then chose the five best fragments, which were visually inspected to avoid important interactions with glycans. These candidates were then further evaluated with 500 ns of cMD or mixMD (see below).

Mixed Molecular Dynamics (MixMD)

Fragments 1-3 and three common molecular probes such as benzene (BENZ), formic acid (FAC), and methylamine (MTA) were used for mixMD(31,40) to explore the accessibility of the stalk region to both very low (BENZ, FAC, MTA) and intermediate (fragments 1-3) molecular weight molecules, as well as possible cryptic binding sites. MixMD systems were prepared using PACKMOL (61), setting a minimum distance of 4 Å between each component to avoid clashes and secure a broad placement of the cosolvent molecules. An adequate number of cosolvent molecules were introduced to reach a virtual concentration of 5% w/w. The systems were then solvated, neutralized, equilibrated, and simulated as reported above. Density maps were computed using the Volmap VMD plugin (https://www.ks.uiuc.edu/Research/vmd/plugins/volmapgui/) setting a grid of 0.5 Å while Solvent accessible surface area (SASA) were estimated using vmdICE (62) and Chimera.

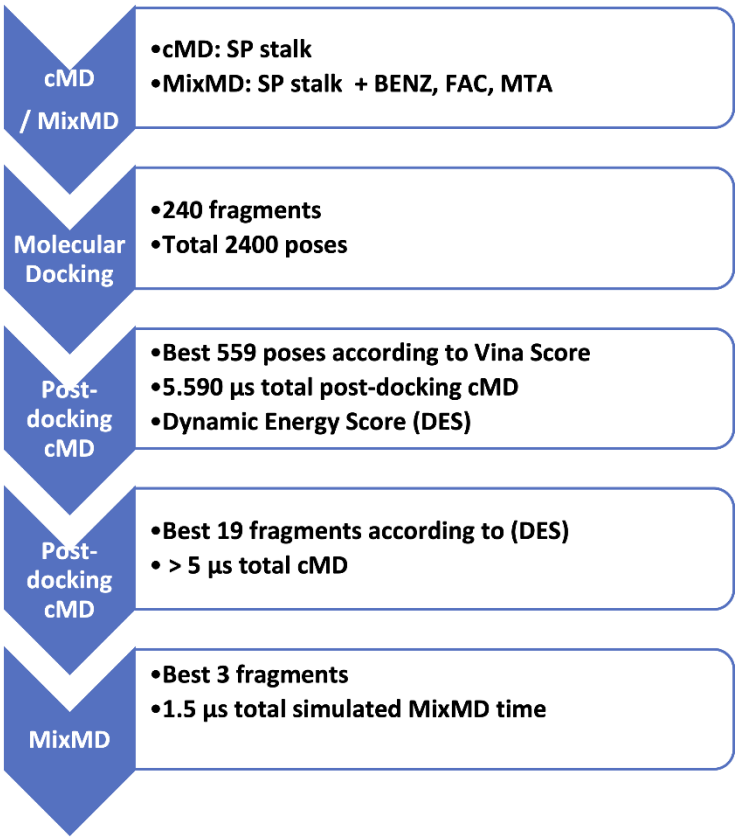


Figure 2. Computational workflow of the study. Preliminary classic molecular dynamics (cMD) and mixed MD (mixMD) were performed on the spike protein stalk (SP); an MD-extracted conformation of the SP stalk was used to dock 240 molecular fragments. The best 560 poses according to Vina score were subjected to 10 ns of cMD ed evaluated according to the dynamic energy score (DES, Equation 1). The best 5 fragments according to DES were finally further simulated through cMD and mixMD.

Results

The flexible loops promote stalk flexibility.

Our investigation focused on the SP stalk Conserved Region 1, between residues L1145 and L1186, which is conserved and less glycosylated than other SP domains. During preliminary cMD simulations of the stalk (Figure 3), the trimer maintained a stable quaternary structure in the region L1154-L1166, while displaying the highest flexibility at the level of the N-terminus (residues 1144-1156) and C-terminus (residues 1202-1214). The flexible Loop 1 (residues 1160-1170) was characterized by intermediate flexibility. The high RMSF (Figure 3A) of the N- and C- *termini* is ascribable to the artificial cut necessary to isolate the stalk region from the rest of the SP, which created unnatural protein ends. For this reason, we excluded the terminal four N-terminal amino acids for the successive docking studies. Computational studies have suggested high flexibility of the SP stalk due to the presence of two unstructured knees (63). Our simulations displayed similar behavior, where the bending of the stalk was aided by the opening of transitory pockets within the flexible loops, which temporarily moved away more than 20 Å from each other (Figure 3B) with an angle of about 133° between the alfa carbons of P1143-V1164-I1172.

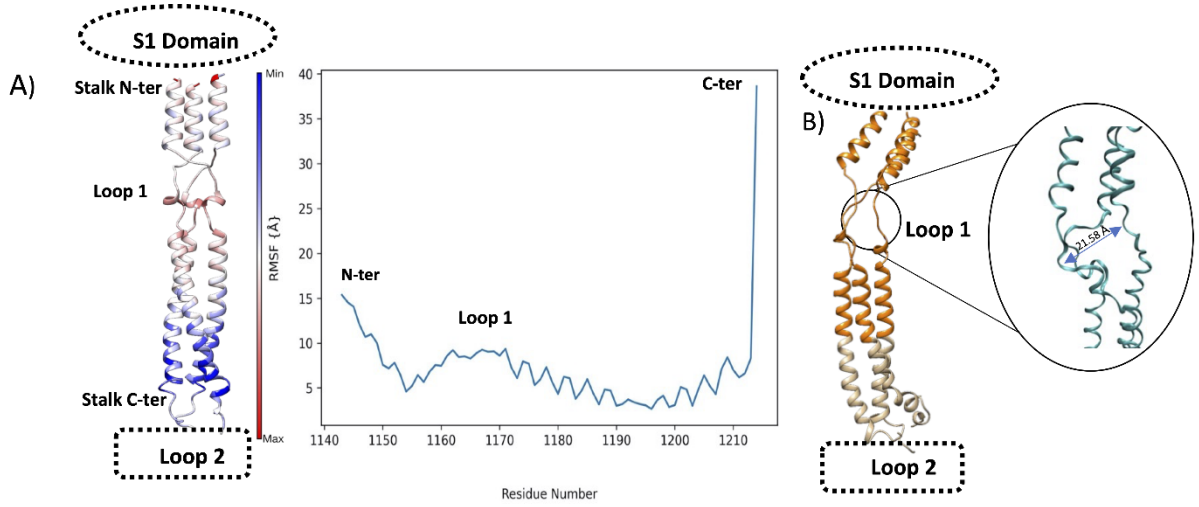


Figure 3. A) Root mean square fluctuation (RMSF) of the SP stalk during cMD simulations. The RMSF of each residue is mapped on the structure (left panel) and color coded according to the value (flexible residues are red) plotted; RMSF are also plotted on the sequence (right panel); **B)** the SP stalk Conserved region 1 (orange) is subjected to high flexibility during cMD simulations at the level of Loop 1; the position of the S1 domain and the Loop 2 is reported for reference.

The conserved region of the stalk is accessible to solvent and small fragments

Preliminary MixMD simulations were employed to assess the accessibility of the Conserved Region 1. The probes benzene (BENZ), formic acid (FAC), and methylamine (MTA) indicated accessible sections across the stalk. BENZ sampled possible hydrophobic pockets on both the N- (close to K1149) and C-terminal (close to E1195) of the stalk. The latter was explored also by FAC, alongside a further interhelical volume in the proximity of E1182. MTA, and to a lesser extent BENZ, weakly interacted with the stalk at the level of the flexible loops' residues V1164-S1170. These results indicate that the glycosylation of the stalk efficiently protects the SP, although some small areas are accessible for potential binding. Interestingly, molecular probes were able to intercalate within the three stalk helices, indicating possible pockets. SASA analysis indicated more solvent-exposed sites at the N-terminal of the stalk, below the connection with S1, and below the flexible Loop 1 in agreement with mixMD results.

Molecular docking and high throughput post-docking cMD predict interactions with Loop 1

Molecular docking of the optimized small library of fragments, performed on the fully glycosylated stalk, produced binding poses with scores ranging from -6.7 kcal/mol to -2.3 kcal/mol with limited convergences with density maps obtained from MixMD simulations. We discarded the last half of the poses (poses 1201-2400) as their docking score was lower than the arbitrary value of -4.5 kcal/mol and all the poses that did not interact with at least two monomers of the stalk or that made contacts with the glycans. The remaining 559 poses were then simulated during 10 ns of post-docking MD simulations, for a total MD sampling of almost 5.6 μs, in a fully hydrated and flexible environment. We computed the DES (Equation 1), which considers both the generalized born surface area (GBSA) binding energy and the root mean square deviation (RMSD) to the initial docking pose, for each fragment. After visually inspecting the resulting best poses according to the DES, we kept only those closest to the Conserved Region 1 and not in contact with any glycan residue, retaining 19 fragments, that were subjected to a further 500 ns cMD simulation.

Longer post-docking MD simulations rebut molecular docking predictions

Disappointedly, 14 out of the 19 fragments were completely displaced in the first 100 ns of the extended simulations. Rotations, openings, and closures of the flexible Loop 1 rapidly disentangled the fragments, causing the rapid displacement of most of the fragments. The remaining five compounds (fragments 1-5, Figure 4, Table 1, Video S1) initially interacted with Loop 1, in correspondence with residues N1158-S1170, before moving away from the initial position. Overall, all the fragments changed orientations towards and formed

alternative interactions before being displaced after less than 300 ns (Figure 4E-I). Planar compounds, predicted by docking in correspondence with the volumes identified by the initial mixMD as possible pockets, displayed better interactions with Loop 1 residues N1158-S1170. 1-3 resided at the center of the trimer for more than 100 ns, before moving upward towards the N-terminal through the three chains and dissociating; 4 and 5 resided mainly on the C-terminal end of Loop 1 before unbinding through a temporary tunnel formed between the stalk chains.

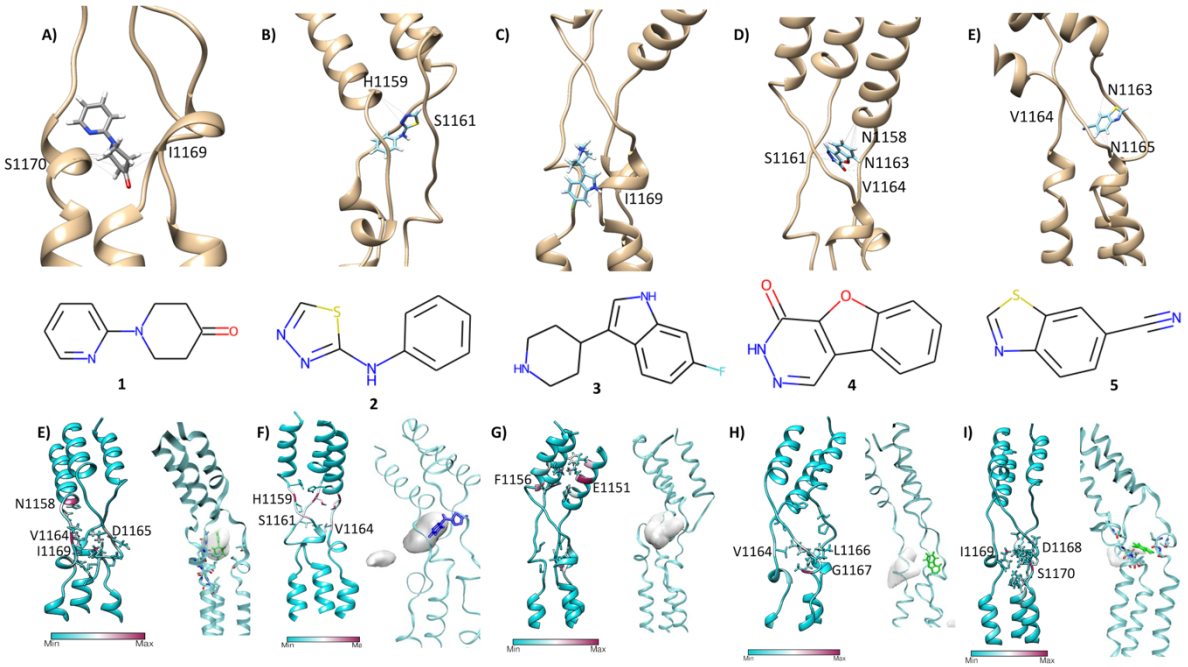


Figure 4. A-D) The best five fragments (1-5) according to molecular docking followed by 10 ns of post-docking cMD simulations; the final conformation (stick representation) after 10 ns is reported. **E-I**) Stalk residues (ribbon) that most interacted with fragments 1-5 during 500 ns of cMD; residues with the highest occupancy are in maroon. The density maps (iso surface 20%, grey surfaces) of the fragments are also reported.

Table 1: Summary of the best 5 fragments after 500 ns of cMD.

Fragment	IUPAC NAME	Ligand Position on the stalk before unbinding	Displacement Time
1	1-pyridin-2-ylpiperidin-4-one	N-terminal	~160 ns
2	N-phenyl-1,3,4-thiadiazol-2-amine	N-terminal	~200 ns
3	6-fluoro-3-piperidin-4-yl-1H-indole	N-terminal	~300 ns

4	3H-[1]benzofuro[2,3-d]pyridazin-4-one	C-terminal	~200 ns
5	1,3-benzothiazole-6-carbonitrile	C-terminal	~110 ns

Mix MD to check the fragments' accessibility to Conserved Region 1

Since mixMD simulations of the molecular probes BENZ, FAC, and MTA suggested some degree of accessibility to the stalk despite the high glycosylation of the SP, we run further mixMD simulations using 1-3 (Table 1, Figure 4) to investigate the accessibility of larger compounds, and any convergence with the metastable configurations sampled during the 500 ns post-docking cMD simulations. During mixMD simulations, the fragments were able to reach the stalk protein surface on isolated spots, overcoming the shield provided by the glycans (Figure 5). Fragment 1 made fewer interactions with the stalk among the three compounds , mainly engaging residues T1155, K1149, E1151, and L1152 and transiently intercalating between the stalk chains (Figure 5A, Video S2), within a cryptic niche formed by L1152, F1148, Y1155, and F1156. Fragment 2 formed interatomic contacts with residues K1149, Y1155, E1151, L1152, V1176, H1159, Q1180, N1173, F1156, S1175, and F2248 (, Figure 5B). Compound 3 formed the most persistent interactions with the stalk Conserved Region1, engaging Y1155, E1182, E1151, N1178, Q1180, K1149, H1159, L1152, R1185, ASP1153, F1156, V1177, and D1153 side chains along both the N- and C-termini (Figure 5C). Interactions highlighted the presence of a possible cryptic pseudo-pocket around residues Y1155-F1156 and H1159. In contrast with the previous docking predictions and cMD simulations, MixMD identified putative interactions along the N-*terminal* section of the Conserved Regon1 of the stalk, right above Loop 1.

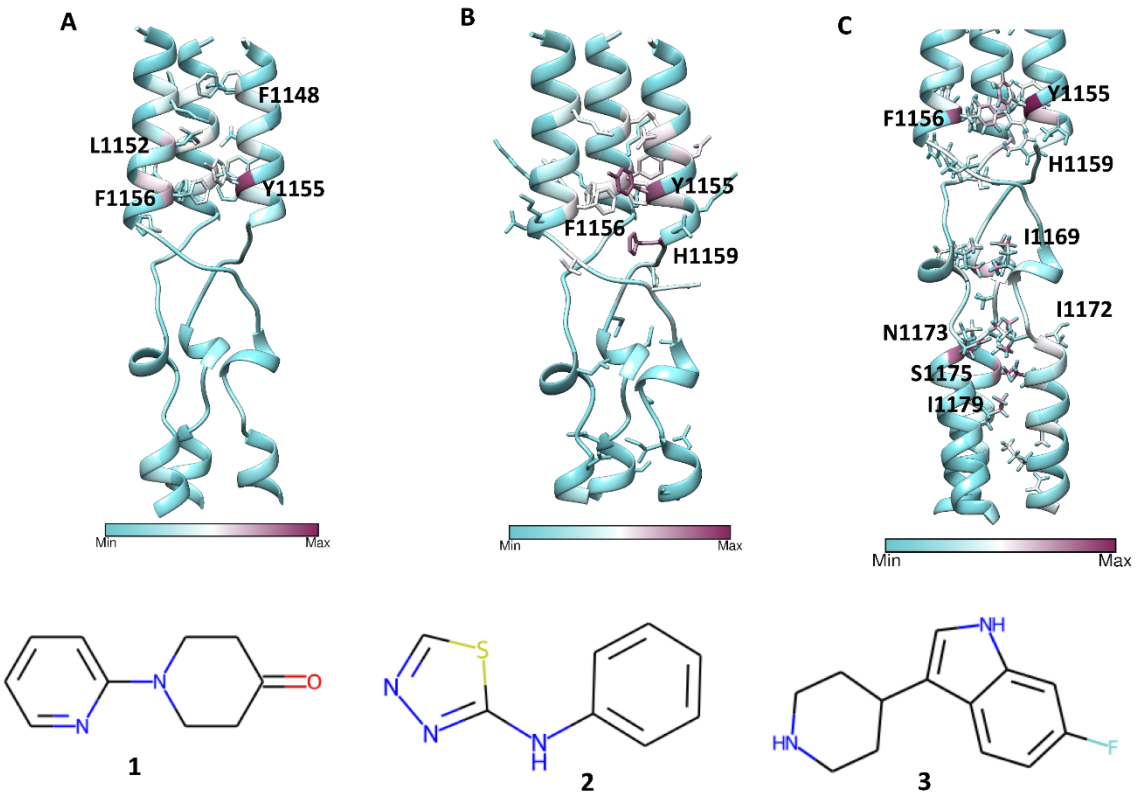


Figure 5. Contacts formed during mixMD simulations of fragments 1-3. The stalk is reported as a ribbon, while the most involved residues (maroon) are in stick representation.

Conclusion

The SP stalk region is conserved amongst the SARS-CoV-2 VOCs. Given its role in orienting the RBD for binding to ACE2, impairing the flexibility of the loops formed by residues 1160-1170 through the binding of a drug, could represent a therapeutic approach to explore(38,64). We investigated the druggability of the SP stalk using 240 molecular fragments and considered the shielding effect of the glycans on the protein surface. Our computational workflow combined molecular docking, high-throughput MD simulations, and mixMD as orthogonal methods to evaluate putative interactions on the stalk region. Molecular docking predicted putative interaction sites around residues T1160 - S1170. High throughput cMD simulations of 559 docking poses suggested the instability of docked fragments, except for a few ligands that were further evaluated in longer simulations. Possible metastable interactions on the stalk region were confirmed in the proximity of residues H1159 – I1169. Finally, mixMD simulations of the three most promising fragments sampled a few narrow binding sites within the three helices of the stalk, which allowed for the transient binding of small, planar ligands.

Overall, our results suggest that the binding at the level of the most rigid part of the stalk is possible for planar fragments, however, the flexibility of Loop 1 does not provide

structural stability for binding, reducing the overall druggability of Conserved Region 1. Although it would be possible to design chemicals bearing a planar group that intercalates between the stalk helices and a linear portion that impairs Loop1 dynamics and likely, the whole SP flexibility, this strategy appears complicated. Future investigations could explore alternative approaches to exploit the symmetry of the trimeric stalk with high molecular weight molecules, rather than a fragment-based one we reported here.

Video:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9786045/bin/viruses-14-02789-s001.zip

References

1. Kumar A, Singh R, Kaur J, Pandey S, Sharma V, Thakur L, et al. Wuhan to World: The COVID-19 Pandemic. Front Cell Infect Microbiol. 2021 Mar 30;11:596201.
2. Tikellis C, Thomas MC. Angiotensin-Converting Enzyme 2 (ACE2) Is a Key Modulator of the Renin Angiotensin System in Health and Disease. Int J Pept. 2012 Mar 20;2012:256294.
3. Bosch BJ, van der Zee R, de Haan CAM, Rottier PJM. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. J Virol. 2003 Aug;77(16):8801–11.
4. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science. 2020 Mar 13;367(6483):1260–3.
5. Al Adem K, Shanti A, Stefanini C, Lee S. Inhibition of SARS-CoV-2 Entry into Host Cells Using Small Molecules. Pharmaceuticals (Basel). 2020 Dec 8;13(12).
6. Sun C, Zhang J, Wei J, Zheng X, Zhao X, Fang Z, et al. Screening, simulation, and optimization design of small molecule inhibitors of the SARS-CoV-2 spike glycoprotein. PLoS ONE. 2021 Jan 25;16(1):e0245975.
7. Razizadeh M, Nikfar M, Liu Y. Small molecule therapeutics to destabilize the ACE2-RBD complex: A molecular dynamics study. Biophys J. 2021 Jul 20;120(14):2793–804.
8. Pandey P, Rane JS, Chatterjee A, Kumar A, Khan R, Prakash A, et al. Targeting SARS-CoV-2 spike protein of COVID-19 with naturally occurring phytochemicals: an in silico study for drug development. J Biomol Struct Dyn. 2021 Oct;39(16):6306–16.

9. Han P, Li L, Liu S, Wang Q, Zhang D, Xu Z, et al. Receptor binding and complex structures of human ACE2 to spike RBD from omicron and delta SARS-CoV-2. Cell. 2022 Feb 17;185(4):630-640.e10.

10. Wu L, Zhou L, Mo M, Liu T, Wu C, Gong C, et al. SARS-CoV-2 Omicron RBD shows weaker binding affinity than the currently dominant Delta variant to human ACE2. Signal Transduct Target Ther. 2022 Jan 5;7(1):8.

11. Zhang X, Wu S, Wu B, Yang Q, Chen A, Li Y, et al. SARS-CoV-2 Omicron strain exhibits potent capabilities for immune evasion and viral entrance. Signal Transduct Target Ther. 2021 Dec 17;6(1):430.

12. Kumar S, Thambiraja TS, Karuppanan K, Subramaniam G. Omicron and Delta variant of SARS-CoV-2: A comparative computational study of spike protein. J Med Virol. 2022 Apr;94(4):1641–9.

13. Kannan S, Shaik Syed Ali P, Sheeza A. Omicron (B.1.1.529) - variant of concern - molecular profile and epidemiology: a mini review. Eur Rev Med Pharmacol Sci. 2021 Dec;25(24):8019–22.

14. Choi JY, Smith DM. SARS-CoV-2 Variants of Concern. Yonsei Med J. 2021 Nov;62(11):961–8.

15. Mandal N, Padhi AK, Rath SL. Molecular insights into the differential dynamics of SARS-CoV-2 variants of concern. J Mol Graph Model. 2022 Apr 14;114:108194.

16. Gao S-J, Guo H, Luo G. Omicron variant (B.1.1.529) of SARS-CoV-2, a global urgent public health alert! J Med Virol. 2022 Apr;94(4):1255–6.

17. Peacock TP, Brown JC, Zhou J, Thakur N, Newman J, Kugathasan R, et al. The SARS-CoV-2 variant, Omicron, shows rapid replication in human primary nasal epithelial cultures and efficiently uses the endosomal route of entry. BioRxiv. 2022 Jan 3;

18. Classification of Omicron (B.1.1.529): SARS-CoV-2 Variant of Concern [Internet]. [cited 2022 May 14]. Available from: https://www.who.int/news/item/26-11-2021-classification-of-omicron-(b.1.1.529)-sars-cov-2-variant-of-concern

19. Pia L, Rowland-Jones S. Omicron entry route. Nat Rev Immunol. 2022 Mar;22(3):144.

20. Callaway E. Why does the Omicron sub-variant spread faster than the original? Nature. 2022 Feb;602(7898):556–7.

21. Arora P, Zhang L, Krüger N, Rocha C, Sidarovich A, Schulz S, et al. SARS-CoV-2 Omicron sublineages show comparable cell entry but differential neutralization by therapeutic antibodies. Cell Host Microbe. 2022 May 6;

22. Khan K, Karim F, Ganga Y, Bernstein M, Jule Z, Reedoy K, et al. Omicron sub-lineages BA.4/BA.5 escape BA.1 infection elicited neutralizing immunity. medRxiv. 2022 May 1;

23. Iketani S, Liu L, Guo Y, Liu L, Chan JF-W, Huang Y, et al. Antibody evasion properties of SARS-CoV-2 Omicron sublineages. Nature. 2022 Apr;604(7906):553–6.

24. Meng B, Ferreira I, Abdullahi A, Kemp SA, Goonawardane N, Papa G, et al. SARS-CoV-2 Omicron spike mediated immune escape, infectivity and cell-cell fusion. BioRxiv. 2021 Dec 21;

25. Kumar R, Murugan NA, Srivastava V. Improved binding affinity of the Omicron's spike protein with hACE2 receptor is the key factor behind its increased virulence. BioRxiv. 2021 Dec 28;

26. Yin W, Xu Y, Xu P, Cao X, Wu C, Gu C, et al. Structures of the Omicron spike trimer with ACE2 and an anti-Omicron antibody. Science. 2022 Mar 4;375(6584):1048–53.

27. Lupala CS, Ye Y, Chen H, Su X-D, Liu H. Mutations on RBD of SARS-CoV-2 Omicron variant result in stronger binding to human ACE2 receptor. Biochem Biophys Res Commun. 2022 Jan 29;590:34–41.

28. Shuai H, Chan JF-W, Hu B, Chai Y, Yuen TT-T, Yin F, et al. Attenuated replication and pathogenicity of SARS-CoV-2 B.1.1.529 Omicron. Nature. 2022 Mar;603(7902):693–9.

29. Yamasoba D, Kosugi Y, Kimura I, Fujita S, Uriu K, Ito J, et al. Sensitivity of novel SARS-CoV-2 Omicron subvariants, BA.2.11, BA.2.12.1, BA.4 and BA.5 to therapeutic monoclonal antibodies. BioRxiv. 2022 May 3;

30. Pipitò L, Rujan RM, Reynolds CA, Deganutti G. Molecular dynamics simulations reveal the importance of conformational changes, glycosylation, mutations, and drug repurposing for the Sars-Cov-2 spike protein. 2022 Mar 3;

31. Smith RD, Carlson HA. Identification of Cryptic Binding Sites Using MixMD with Standard and Accelerated Molecular Dynamics. J Chem Inf Model. 2021 Mar 22;61(3):1287–99.

32. Bajusz D, Wade WS, Satała G, Bojarski AJ, Ilaš J, Ebner J, et al. Exploring protein hotspots by optimized fragment pharmacophores. Nat Commun. 2021 May

27;12(1):3201.

33. Duan L, Zheng Q, Zhang H, Niu Y, Lou Y, Wang H. The SARS-CoV-2 Spike Glycoprotein Biosynthesis, Structure, Function, and Antigenicity: Implications for the Design of Spike-Based Vaccine Immunogens. Front Immunol. 2020 Oct 7;11:576622.

34. Shah P, Canziani GA, Carter EP, Chaiken I. The Case for S2: The Potential Benefits of the S2 Subunit of the SARS-CoV-2 Spike Protein as an Immunogen in Fighting the COVID-19 Pandemic. Front Immunol. 2021 Mar 9;12:637651.

35. Turoňová B, Sikora M, Schürmann C, Hagen WJH, Welsch S, Blanc FEC, et al. In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. Science. 2020 Oct 9;370(6513):203–8.

36. Williams JK, Wang B, Sam A, Hoop CL, Case DA, Baum J. Molecular dynamics analysis of a flexible loop at the binding interface of the SARS-CoV-2 spike protein receptor-binding domain. Proteins. 2022 May;90(5):1044–53.

37. Barros EP, Casalino L, Gaieb Z, Dommer AC, Wang Y, Fallon L, et al. The flexibility of ACE2 in the context of SARS-CoV-2 infection. Biophys J. 2021 Mar 16;120(6):1072–84.

38. Elshabrawy HA, Coughlin MM, Baker SC, Prabhakar BS. Human monoclonal antibodies against highly conserved HR1 and HR2 domains of the SARS-CoV spike protein are more broadly neutralizing. PLoS ONE. 2012 Nov 21;7(11):e50366.

39. Dokainish HM, Re S, Mori T, Kobayashi C, Jung J, Sugita Y. The inherent flexibility of receptor binding domains in SARS-CoV-2 spike protein. eLife. 2022 Mar 24;11.

40. Graham SE, Leja N, Carlson HA. MixMD Probeview: Robust Binding Site Prediction from Cosolvent Simulations. J Chem Inf Model. 2018 Jul 23;58(7):1426–33.

41. Ghanakota P, Carlson HA. Moving Beyond Active-Site Detection: MixMD Applied to Allosteric Systems. J Phys Chem B. 2016 Aug 25;120(33):8685–95.

42. Huang J, MacKerell AD. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. J Comput Chem. 2013 Sep 30;34(25):2135–45.

43. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods. 2017 Jan;14(1):71–3.

44. Woo H, Park S-J, Choi YK, Park T, Tanveer M, Cao Y, et al. Developing a Fully Glycosylated Full-Length SARS-CoV-2 Spike Protein Model in a Viral Membrane. J Phys Chem B. 2020 Aug 20;124(33):7128–37.

45. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK Predictions. J Chem Theory Comput. 2011 Feb 8;7(2):525–37.

46. Doerr S, Harvey MJ, Noé F, De Fabritiis G. HTMD: High-Throughput Molecular Dynamics for Molecular Discovery. J Chem Theory Comput. 2016 Apr 12;12(4):1845–52.

47. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996 Feb;14(1):33–8, 27.

48. 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.

49. Harvey MJ, Giupponi G, Fabritiis GD. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. J Chem Theory Comput. 2009 Jun 9;5(6):1632–9.

50. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys. 1984;81(8):3684.

51. Loncharich RJ, Brooks BR, Pastor RW. Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. Biopolymers. 1992 May;32(5):523–35.

52. Landrum G, Tosco P, Kelley B, Sriniker, Gedeck, NadineSchneider, et al. rdkit/rdkit: 2020_03_1 (Q1 2020) Release. Zenodo. 2020;

53. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem. 2004 Oct;25(13):1605–12.

54. Eberhardt J, Santos-Martins D, Tillack AF, Forli S. Autodock vina 1.2.0: new docking methods, expanded force field, and python bindings. J Chem Inf Model. 2021 Aug 23;61(8):3891–8.

55. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010 Jan 30;31(2):455–61.

56. Vanommeslaeghe K, MacKerell AD. Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing. J Chem Inf Model. 2012 Dec 21;52(12):3144–54.

57. Vanommeslaeghe K, Raman EP, MacKerell AD. Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic

charges. J Chem Inf Model. 2012 Dec 21;52(12):3155–68.

58. Salomon-Ferrer R, Case DA, Walker RC. An overview of the Amber biomolecular simulation package. WIREs Comput Mol Sci. 2013 Mar;3(2):198–210.

59. Frisch A. Gaussian 09w reference. Wallingford, USA, 25p. 2009;

60. Sabbadin D, Ciancetta A, Deganutti G, Cuzzolin A, Moro S. Exploring the recognition pathway at the human A2A adenosine receptor of the endogenous agonist adenosine using supervised molecular dynamics simulations. Med Chem Commun. 2015;6(6):1081–5.

61. Martínez L, Andrade R, Birgin EG, Martínez JM. PACKMOL: a package for building initial configurations for molecular dynamics simulations. J Comput Chem. 2009 Oct;30(13):2157–64.

62. Knapp B, Lederer N, Omasits U, Schreiner W. vmdICE: a plug-in for rapid evaluation of molecular dynamics simulations using VMD. J Comput Chem. 2010 Dec;31(16):2868–73.

63. Verkhivker G. Conformational Flexibility and Local Frustration in the Functional States of the SARS-CoV-2 Spike B.1.1.7 and B.1.351 Variants: Mutation-Induced Allosteric Modulation Mechanism of Functional Dynamics and Protein Stability. Int J Mol Sci. 2022 Jan 31;23(3).

64. Xia X. Domains and Functions of Spike Protein in Sars-Cov-2 in the Context of Vaccine Design. Viruses. 2021 Jan 14;13(1).

6.4 A Pathway Model to Understand the Evolution of Spike Protein Binding to ACE2 in SARS-CoV-2 Variants

With Omicron cases rising worldwide, the concerns of the scientific community were focused on predicting the effects of the mutation on both antibodies' affinity and the pathogenic effects. Surface plasmon resonance (SPR) experiments in the literature reported a stronger binding affinity to ACE2 of the Omicron variant but a slower binding kinetics, compared to the Delta and Alfa. Supervised molecular dynamics (SuMD) revealed the RBD:ACE2 binding pathway for the Alpha, Delta, and Omicron variants, and highlighted the effect of the mutations on the RBD in the stability of the complex. My work compared the binding properties and pathways of different SARS-CoV-2 RBD variants (RBD^{Δ}, RBD^{wt} and RBD°) to understand how key mutations impact infectivity. The mutation's impact on binding dynamics affects the association rate by destabilizing unbinding states (USs), enhancing RBD° complex stability, and potentially prolonging receptor residence time. The insights and observations led to the publication of the work titled: "A Pathway Model to Understand the Evolution of Spike Protein Binding to ACE2 in SARS-CoV-2 Variants" published on Biomolecules. 2022 Nov; 12 (11): 1607; doi: 10.3390/biom12111607.

Authors: Ludovico Pipitò, Christopher A. Reynolds, Juan Carlos Mobarec, Owen Vickery, and Giuseppe Deganutti

Abstract

After the SARS-CoV-2 Wuhan variant that gave rise to the pandemic, other variants named Delta, Omicron, and Omicron-2 sequentially became prevalent. Mutations spread around the viral genome, including on the spike (S) protein resulting in gains in infectivity, potentially through enhanced binding of the virus to the angiotensin-converting enzyme 2 (ACE2) receptor on the cell's surface. We interrogated *in silico* the molecular recognition between the receptor-binding domain (RBD) of different variants and ACE2 through supervised molecular dynamics (SuMD) and classic molecular dynamics (MD) simulations to address the effect of mutations on the possible S protein binding pathways. Our results indicate that compensation between binding path efficiency and stability of the complex exists for the Omicron receptor binding domain, while Omicron-2's mutations putatively improved the dynamic recognition of the ACE2 receptor, suggesting an evolutionary advantage over the previous strains.

Introduction

The new strain of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), B.1.1.529 (Omicron), originated in South Africa (https://www.gisaid.org/phylodynamics/west-africa/) was identified by the World Health Organization (WHO) on the 24th of November 2021 and a declared variant of concern (VOC) two days later (1). Various Omicron cases were reported from travellers coming from South Africa and people all around the world (2,3), raising concerns amongst the scientific community and governments alike. Omicron cases in South Africa, America, and India drastically reached their peak in January right after the Delta variant started to be under control (4), with a similar scenario in Europe where Delta and Omicron are still competing in the infection's ratios, while new Omicron variants B.A.2 and B.A.3 are kept under observation (4).

The presence of more than 50 mutations (Figure 1), including deletions, raised concerns and speculations about Omicron's ability to evade the innate immune response. K417N, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y and Y505H are part of the immunodominant antigenic site I (5). Despite the concerns about Omicron and its evasion mechanism, experimental data indicated how Omicron's mutation heavily impact the viral replication and pathogenicity due to inefficiency to exploit the cellular serine protease TMPRSS2 (6) compared to Alfa, Beta and Delta variants. In Shuai's experiments, mice infected with the Omicron showed a drastic reduction in viral replication and a strongly reduced pro-inflammatory response as indicated by the modest gene expression of interferon-gamma induced protein 10 (IP-10) and the reduced interferon-gamma production (IFNγ). The mutations on the S1/S2 domain and the N-terminal Domain (NTD) suggested the intriguing hypothesis that Omicron could compromise the cell's ability to degrade its viral components while also reducing the efficacy of the majority of the vaccines (7) due to 15 mutations, some of which are conserved between Beta and Delta strains (8), present on the receptor-binding domain (9) (RBD). According to deep mutational scanning experiments, almost all the mutations involving the receptor-binding motif (RBM) did not increase ACE2 binding affinity when present individually (10), while N501Y mutation enhances RBD binding to ACE2 by 6-fold or more(11–13) relative to other strains, due to increased shape complementarity with ACE2 Y41^{ACE2} and K353^{ACE2} side chains. Mutations Q493K (or Q493R) and Q498R introduce new ionic interactions with E35^{ACE2} and E38^{ACE2} but displayed slightly reduced avidity when tested individually in yeast-displayed SARS-CoV-2 RBD (10). The K417N mutation, on the other hand, worsens ACE2 recognition by about 3-fold(11,14) through loss of a salt bridge with D30^{ACE2}, although the effect on the binding when combined with other mutations is reportedly smaller (15). Notably, these numerous omicron mutations

seem to compensate each other when it comes to the binding affinity for ACE2, as the RBD of Omicron (RBD^O) was similar to other strains(15–17).

In the present work, we first interrogated the ACE2 ectodomain in complex with the RBDs from the wild type S protein, Delta, and omicron variants by means of molecular dynamics (MD) simulations, proposing a unique network of hydrogen bonds characterizing omicron. We then studied the out-of-equilibrium binding process of Delta TBD (RBD^D) or RBD^O to ACE2 employing supervised MD(18,19) (SuMD) simulations. We propose that the same mutations stabilizing the omicron complex with ACE2 hamper the kinetics of binding, accounting for the overall compensation on the measured affinity for the receptor.

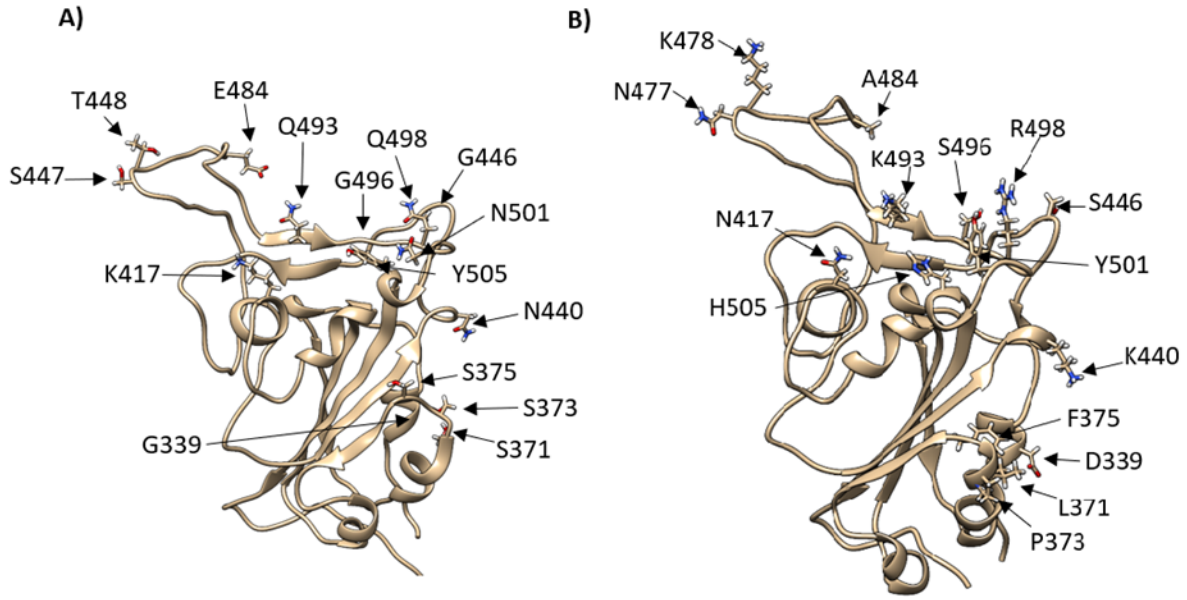


Figure 1. Comparison between SARS-CoV-2 RBD wild type and SARS-CoV-2 RBD Omicron. A) SARS-2 WT RBD model showing residues T333-P527, with the original WT amino acids represented as licorice. **B**) SARS-2 WT RBD model showing residues T333-P527, with the Omicron mutations highlighted and represented as licorice. With a total of 50 mutations, 15 of which on RBD, the Omicron variant possesses a different configuration of polar residues in the region between N477-H505 and an intriguing K417N mutation.

Methods

Structure Preparation and Force Field Settings

All systems were prepared using the CHARMM36(20,21)/CGenFF 3.0.1(22,23) force field combination. The model of the RBD^{O} of the spike protein RBD model was modeled through alphafold2 (24) and comprised residues T333-P527. RBD^{D} was retrieved from PDB ID 7V8B. ACE2 residues S19 to A614 were obtained from PDB ID 6M17.

The protonation state of residues side chains was calculated by Propka (25) at a simulated pH of 7.45, to match the crystallography experimental data, and added by pdb2pqr (26) while disulfide bonds were identified by HTMD (27), visually inspected, and patched manually through VMD (28). The initial geometry and internal energy were optimized using ACEMD (29).

System preparation for classic molecular dynamics (MD)

The RBD^{WT}:ACE2 complex from PDB 6M17, as per our previous work (19), was used as a reference for the preparation of both RBD^{D}:ACE2 and RBD^{O}:ACE2 complexes. RBD^{D}:ACE2 was obtained by superimposing RBD^{D} from PDB 7V8B onto RBD^{WT}:ACE2, while the RBD^{O}:ACE2 complex was obtained superimposing an RBD^{O} model obtained by AlphaFold2 (24) on the RBD^{WT}:ACE2 complex. Glycan residues were removed and topology files were prepared using VMD's Psfgen plugin (https://www.ks.uiuc.edu/Research/vmd/plugins/psfgen/), and the resulting structures were visually inspected after their creation. The systems were simulated for a total time of 500 ns in triplicate with TIP3P water molecules (30) added to the simulation box considering a 15 Å padding in every direction by Solvate plugin 1.5 (http://www.ks.uiuc.edu/Research/vmd/plugins/solvate/). The charge neutrality was achieved by adding Na^{+}/Cl^{-} to the concentration of 0.150 M using Autoionize plugin 1.3 (http: //www.ks.uiuc. edu/Research/vmd /plugins/autoionize/). ACEMD was used for both the equilibration and the productive MD trajectories . The energy of the systems was reduced through 1000 conjugate-gradient minimization steps to eliminate possible clashes and optimize atomic distances. Equilibration was reached in isothermal-isobaric conditions (NPT) using the Berendsen barostat (31) (target pressure 1 atm) and the Langevin thermostat (32) (target temperature 310 K) during a 4 ns long MD simulation (integration time step 2 fs). During the equilibration, a positional restraint of 1 kcal/ mol $Å^2$ was applied on the alpha carbons of both ACE2 and RBD for the first 3 ns, and on protein side chains for the first 2 ns. Productive trajectories were produced in triplicate with an integration time step of 4 fs, using

the hydrogen mass repartition (33) in the canonical ensemble (NVT), with no positional restraints. The cut-off distance for electrostatic interactions was set at 9 Å, with a switching function applied beyond 7.5 Å. Long-range Coulomb interactions were handled using the particle mesh Ewald summation method (PME)(34) with default ACEMD settings. Atomic velocity was reassigned in each replicate to increase the sampling and explore possible alternate conformations.

Supervised molecular dynamics (SuMD)

SuMD is an adaptive sampling method (35) for speeding up the simulation of the binding(18,36) and unbinding processes (37). During SuMD, sampling is gained without the input of any energetic bias, by applying a "tabu–like" algorithm to monitor the distance between centers of mass (or geometrical centers) chosen on ligand and receptor. Consecutive unbiased short MD simulations are performed, and, after each simulation, the distances (collected at regular time intervals) are fitted to a linear function. If the slope of the linear fitting function is negative, then the next short MD will start from the last coordinates and velocities, otherwise, the simulation will be restarted by randomly assigning the atomic velocities according to the Boltzmann distribution (38).

An initial distance between RBD^{WT} and ACE2 was set at 25 Å, allowing conformational exploration during the binding path as per our previous work (19). The initial position of RBD^{D} or RBD^{O} was obtained by superimposing them on RBD^{WT} through Chimera's aligning feature, producing identical starting conditions. Ultimately, the topology files were built through VMD's Psfgen and visually inspected. TIP3P water molecules were added to the simulation box considering a 15 Å water padding, using the minimum and maximum coordinates of the structures as a reference. The charge neutrality of the system was achieved by adding Na^{+} /Cl^{-} to the concentration of 0.150 M using Autoionize plugin 1.3. Eight independent replicas of SuMD were produced for each DV and OV system. The simulations were produced by supervising the distance between RBD residue $K31^{ACE2}$ and $Q493^{D}$ or $K493^{O}$ on the RBD "up" binding motif (RBM). A series of 600 ps-long time windows were simulated until the distance reached a value lower than 10 Å without further improvements in the distance. Frames were saved every 200 ps and used to interpolate the linear function of the distance during the simulated 600 ps. To simulate the engagement between RBD and ACE2, a 200 ns long MD trajectory was produced starting from the last coordinate produced by SuMD.

MD trajectories analysis

Out of eight SuMD binding replicas, the best four and three replicas in terms of reproducing the experimental complex geometry (Figure 3A, C) were analyzed for RBD^{D} and RBD^{O}, respectively. The root mean square deviations (RMSD) analysis were computed using VMD and MDTraj (39). Ligand-protein contacts, including hydrogen bonds, were detected using the GetContacts scripts tool (https://getcontacts.github.io), with a threshold distance and angle of 3.5 Å 120°, respectively. Contacts and HB were expressed as occupancy (% of total MD frames). The molecular mechanics energy combined with the generalized Born surface area (MM-GBSA) was computed with the MMPBSA.py (40) script (AmberTools20 suite at http://ambermd.org/), converting the CHARMM psf topology files to Amber prmtop format through ParmEd (http://parmed.github.io/ParmEd/html/index.html).

RESULTS

Omicron mutations strengthen the interaction with ACE2 compared to WT and Delta complexes

During the preparation of this manuscript, the cryo-EM structure of omicron S protein in complex with ACE2 was released (41). We assessed the quality of our model by measuring the RMSD of RBD^{O} to the experimental structure during MD simulations of the complex with ACE2, which resulted in alpha carbon atoms displacement of 0.8 Å and side-chain atoms displacement of 3.1 Å, respectively. The latter value is closed to the nominal resolution (2.45 Å to 3.40 Å) of the available RBD^{O} structures (7T9L, 7T9K, and 7WBL), indicating the validity of the AlphaFold2 model we used. We assessed the structural binding properties of RBD^{WT}, RBD^{D}, or RBD^{O} in complex with ACE2 through MD simulations of each system, performed in triplicate (Table 1). The comparison of the three RBD strands in terms of thermic fluctuations show similar RMSD values for DV and OV, while the WT resulted more flexible when bound to ACE2 (Figure 2A). This is in line with previous work suggesting higher dynamicity of the RBD^{WT} complex compared to the omicron strand (42).

Mutations characterizing the RBD^{D} did not modify the interaction pattern observed for RBD^{WT} with ACE2 (Figure 2B), although residues common to both strands formed a different number of contacts in the two complexes. More precisely, $Q498^{D}$, $T500^{D}$, and $N501^{D}$ on RBD^{D} engaged ACE2 more than RBD^{WT} did, while the latter tent to interact more through $N487^{WT}$ and $Y505^{WT}$. This slightly asymmetric interaction pattern is not mirrored by MM-GBSA energy analysis (Figure 2C). Indeed, per residue energy contribution to the stabilization of the ACE2 complexes show a high degree of similarity between RBD^{WT} and

RBD^{D} and overall very similar computed binding energies of -23.38 kcal/mol for RBD^{WT}:ACE2 and -22.95 kcal/mol for RBD^{D}:ACE2, respectively. Intriguingly, simulations of RBD^{O} suggested a substantial change of the interaction pattern with ACE2, compared to RBD^{WT} (Figure 2C). RBD^{O} residues $R498^{O}$, $K493^{O}$, $S496^{O}$, $Y501^{O}$ ($Q493^{WT}$, $G496^{WT}$, $N501^{WT}$ in RBD^{WT}), and $T500^{O}$ formed more contacts with the receptor than RBD^{WT}, which was instead more engaged at the level of $Y505^{WT}$ ($H505^{O}$ in RBD^{O}) and $K417^{WT}$ ($N417^{O}$ in RBD^{O}). $N477^{O}$ had a very limited impact on the hydrogen bonds network with ACE2 by transitory interacting with $Q24^{ACE2}$, $S19^{ACE2}$, and $T20^{ACE2}$. All the other mutated residues characterizing RBD^{O} did not participate in hydrogen bonds with the receptor. N501Y appears particularly important for the RBD affinity towards ACE2(12,13).

From an energetic perspective, this interactions frame is suggested to be particularly important for $K493^{O}$, which was able to form stabilizing salt bridges with the ACE2 residues $E35^{ACE2}$ and $D38^{ACE2}$, compensating unfavorable interactions with $D30^{ACE2}$ and $K31^{ACE2}$. Both RBD^{WT} and RBD^{D} present, instead, a glutamine residue, $Q493^{WT/D}$, forming a simple hydrogen bond with $E35^{ACE2}$, while $Y449^{WT/D}$ engaged $D38^{ACE2}$ in a further, weak, hydrogen bond. The computed binding energy of the RBD^{O}:ACE2 complex was -28.49 kcal/mol , about 5 kcal/mol more stable than RBD^{WT}:ACE2 and RBD^{D}:ACE2, likely thanks to these specific electrostatic interactions involving $K493^{O}$. Our MM-GBSA binding energy results are consistent with Rajender K, *et al* (43) and Lupala C.S *et al* (44), but in disagree with findings from other groups(45–47) suggesting RBD^{D} as a stronger ACE2 binder than RBD^{WT} and RBD^{O}. Reasons for these discrepancies could lie in the different lengths of the simulations or the divergent number of replicas considered for the binding energy computation. Other technical aspects such as the force field in the simulations and the GBSA parameters should influence the output limitedly (48). Surface plasmon resonance (SPR) binding assays quantified the RBD^{O} binding affinity for ACE2 being either 2.4-fold higher (16) than RBD^{WT} or unchanged (41), with relative differences between $RBD^{D}$and RBD^{O} in the range of 1 to 3 fold(17,41) in favor of the latter. Surprisingly, such affinities indicate very similar binding properties between RBD variants despite the high number of mutated residues present on the RBD^{O} receptor binding motif (RBM).

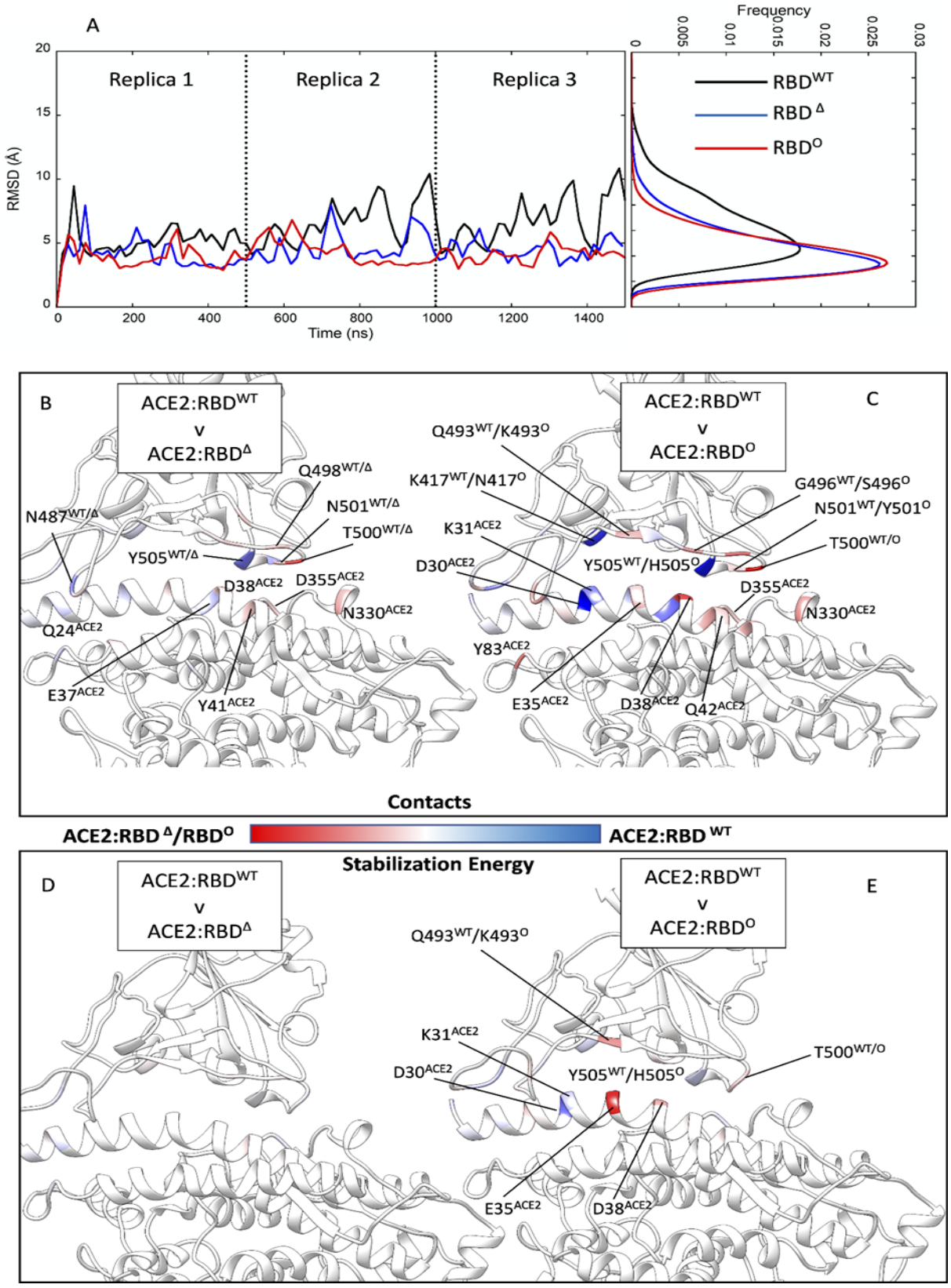


Figure 2. MD of ACE2 in complex with RBD^{WT}, RBD^{D}, and RBD^{O}. A) RMSD of RBD^{WT}, RBD^{D}, and RBD^{O} over the time course of three replicas (left panel, the curves were

smoothed to interpolate the RMSD values) and the relative frequency distribution. **B)** Comparison between the intermolecular contacts formed in ACE2:RBD^{WT} and ACE2:RBD^{D} complexes; red residues interacted more in ACE2:RBD^{D}, while blue residues were more engaged in ACE2:RBD^{D}. **C)** Comparison between the intermolecular contacts formed in ACE2:RBD^{WT} and ACE2:RBD^{O} complexes; red residues interacted more in ACE2:RBD^{O}, while blue residues were more engaged in ACE2:RBD^{D}. **D)** Comparison between the per residue interaction energy in ACE2:RBD^{WT} and ACE2:RBD^{D} complexes; red residues stabilized ACE2:RBD^{D}, while blue residues stabilized more ACE2:RBD^{D}. **E)** Comparison between the per residue interaction energy in ACE2:RBD^{WT} and ACE2:RBD^{O} complexes; red residues stabilized ACE2:RBD^{D}, while blue residues stabilized more ACE2:RBD^{O}.

Omicron RBM mutations impede the binding to ACE2

Since equilibrium MD simulations of the ACE2:RBDs complexes indicated gain in binding stability for ACE2:RBD^{O}, not supported by experimental binding data in the recent literature, we further investigated the RBD binding properties by means of SuMD, an energetically unbiased out-of-equilibrium MD technique. The goal was to study the first step of the molecular recognition between ACE2 and RBD^{D} or RBD^{O}, starting from the completely dissociated heterodimer and allowing a direct comparison between the two dominant SARS-CoV-2 variants (Video S1 and Video S2). We first performed height SuMD binding replicas for both RBD^{D} and RBD^{O}, followed by 200 ns of unsupervised, classic MD to allow the metastable complexes during the SuMD stage to relax. The replicas better reproducing the experimental complex geometry (four for ACE2:RBD^{O} and replicas for ACE2:RBD^{D}, Figure 3A,C) were further analyzed. In three SuMD replicas out of four (Figure 3B) the ACE2:RBD^{D} complex reached stabilization energy close to the one computed for the complex (e.g. -23.38 kcal/mol, Table S3), while all the three ACE2:RBD^{O} SuMD binding simulations (Figure 3D) showed a transitory stabilization of about -20 to -30 kcal/mol before experiencing an energy increase at the end of the simulations. The lower stabilization of ACE2:RBD^{O} during the simulated binding events compared to ACE2:RBD^{D} is supported by the higher RMSD values to the bound complex (Figure 3A,C) and preliminarily suggests a less efficient propensity to engage ACE2 for some reasons.

To address the differences between ACE2:RBD^{D} and ACE2:RBD^{O} that emerged during SuMD binding simulations, we extracted and analyzed all the frames with an MM-GBSA energy > 5 kcal/mol. The rationale for this is that the kinetics of binding is governed by the energy of the transition states (TS) along the path and therefore the RBD propensity to bind ACE2 can be understood by determining the driving forces of potential transition states. Importantly, we did not consider the less stable configurations from SuMD simulations as the

actual TS of binding, for two reasons. The first is that TS inherently suffer from poor MD sampling and extensive simulations are required to capture high energy states of the system. The second reason is that the MM-GBSA analysis we performed, which uses an implicit solvent, neglected the entropic contribution to the energy of binding. It follows that the conformational entropy of the proteins, the roto-translational entropy of water molecules and the contribution of desolvation to the free energy of binding were overlooked. However, we assumed the frames with an MM-GBSA energy > 5 kcal/mol to be closed enough to give insight into the enthalpic nature of the TS during RBD^{D} and RBD^{O} to ACE2. In these unstable states (USs) RBD^{O} made atomic contacts and hydrogen bonds with ACE2 mainly through the apical portion of the domain, especially involving $Y449^{O}$ and $S446^{O}$ (Figure 3E,F). RBD^{D}, instead, made spreader interactions with the receptor, through $N501^{D}$, $Y453^{D}$, and $N417^{D}$ (Figure 3E,F). The MM-GBSA per residue energy decomposition (Figure 3 G,H) highlights the RBD^{D} and RBD^{O} residues that stabilized or destabilized the USs during binding. $D405^{D}$, $E406^{D}$, $D420^{D}$ contributed to the high energy of the complex during the binding, while $K417^{D}$ and $V503^{D}$ putatively stabilized these states thanks to hydrogen bonds with ACE2 residues D30 in the case of the former residue . Moving to RBD^{O}, the most stabilizing residues during binding were $Y449^{O}$, which formed a hydrogen bond with E37 and $F486^{O}$, while $K493^{O}$ increased the energy of the USs intermediate states (Figure 3 G,H) despite weak hydrogen bonds with E35 and D38.

Altogether, these results suggest a different binding path for RBD^{D} and RBD^{O} driven by some of the mutations occurring between the two strains of the virus. Residue $K417^{D}$ appears pivotal in orienting the binding to ACE2 by forming the strong hydrogen bond with D30 since the first step of RBD^{D} recognition. From this standpoint, the smaller and neutral $N493^{D}$ side chain is predicted not to affect the binding transition states, compared to K493, which instead destabilized the binding path.

Kinetics experiments ruled any influence of mutation N501Y on the RBD binding on-rate (13), however, $N501^{D}$ formed a stabilizing hydrogen bond with $K353^{ACE2}$ in the first steps of RBD^{O} binding to ACE2, whilst $Y501^{O}$ was not involved during USs. This inconsistency could be due to the fact that the kinetics data refers to the single RBD mutant or to the inherent limits of the MM-GBSA model.

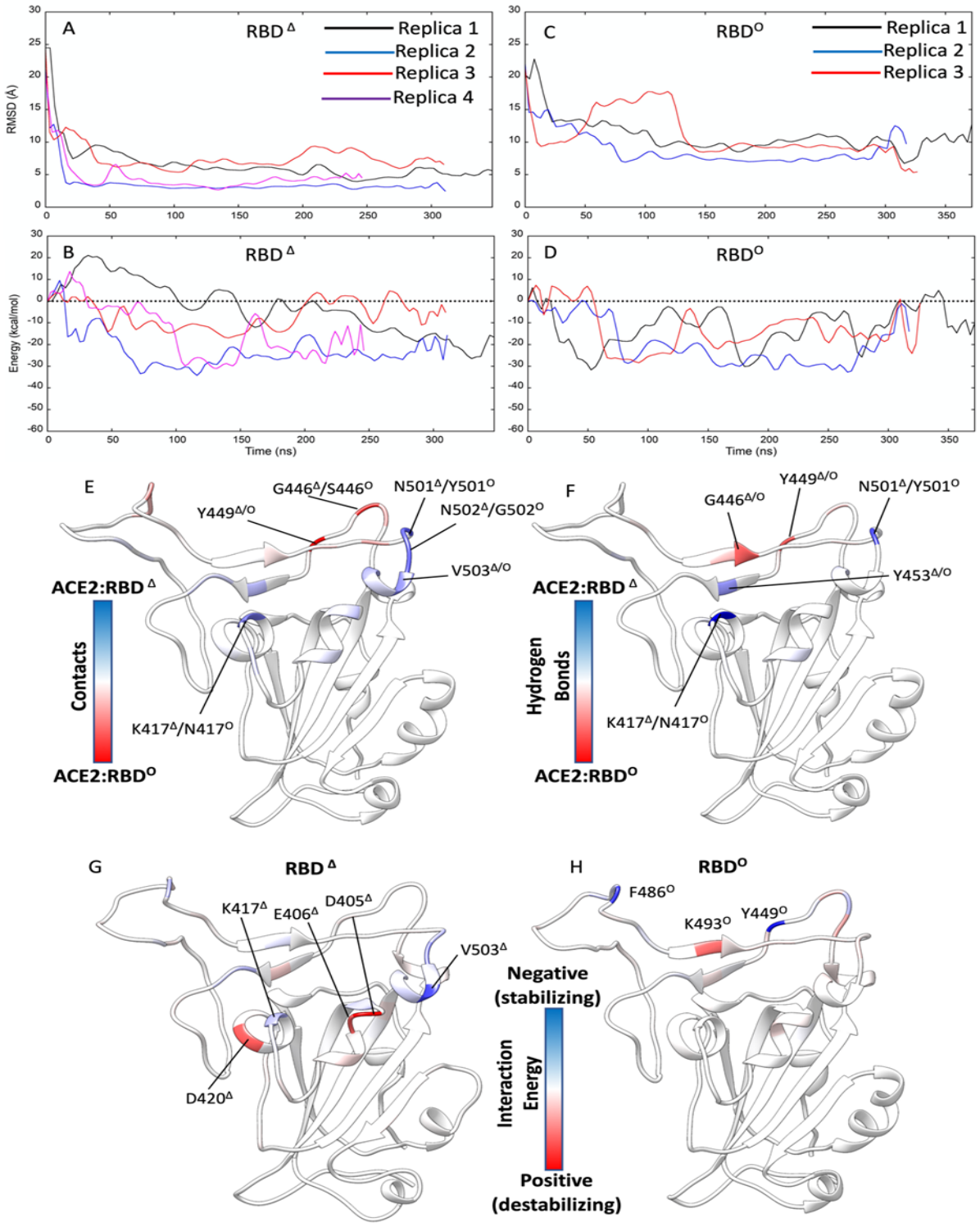


Figure 3. SuMD binding of RBD^{D} and RBD^{O} to ACE2. Only frames with binding energy > 5 kcal/mol were analyzed (unstable states). **A)** RMSD of RBD^{D} to the bound final complex over the time course of the best four SuMD replicas; **B)** MM-GBSA energy over the time course of the best four RBD^{D} SuMD replicas; **C)** RMSD of RBD^{O} to the bound final complex over the time course of the best three SuMD replicas; **D)** MM-GBSA energy over the time course of

the best three RBD^O SuMD replicas. Curves were smoothed to interpolate the RMSD values. **E)** Comparison between the intermolecular contacts formed by RBD^D or RBD^O in the unstable states of SuMD binding to ACE2; red residues interacted more in RBD^O while blue residues were more engaged in RBD^D. **F)** Comparison between the hydrogen bonds formed by RBD^D or RBD^O in the unstable states of SuMD binding to ACE2; red residues interacted more in RBD^O while blue residues were more engaged in RBD^D. **G)** Per residue energy decomposition of RBD^D in the unstable states of SuMD binding to ACE2; **F)** Per residue energy decomposition of RBD^O in the unstable states of SuMD binding to ACE2.

Conclusion

The late months of 2021 and early 2022 have witnessed the spread of the omicron SARS-CoV-2 strain as the dominant variant responsible for COVID-19, replacing the Delta variant. Important efforts have been addressed to understand the role of new Omicron mutations in the etiology of its rapid diffusion. Mutations of RBD residues in contact with the N-terminal domain (NTD) of the S protein can favor a shift of the up and down conformational equilibrium towards the up form, increasing the probability of ACE2 recognition and, therefore, infectivity. However, this has not yet been reported for Omicron, which shows instead a high number of mutations grouped on the apical part of the RBD, responsible for direct interactions with ACE2 receptors. In this work, we computationally assessed and compared the binding properties of RBD^D and RBD^O to understand the putative role of key mutations in enhancing the infectivity, despite the binding affinity for ACE2 being almost unchanged, compared to the Delta variant. Our simulations of RBD^{WT}, RBD^D and RBD^O suggest that RBD^O:ACE2 is more stable than the former complexes, thanks to the contribution of new salt bridges formed by $K493^O$. The same mutation disfavored the dynamic binding to ACE2 by destabilizing the transition states during the recognition. In this scenario, a compensation between kinetics constants of binding, kon and koff, would explain the similar affinity for ACE2 displayed by RBD^D and RBD^O. The higher stability of the RBD^O complex could produce a higher residence time on the receptor, increasing the chances for the TMPRSS2 to cleavage the S protein and start the membrane fusion, with the final effect of enhancing the infectivity. Bearing in mind the inherent limits of MM-GBSA computations (49), we believe this is a new angle to understand infectivity from the dynamic perspective of RBD binding.

Videos:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9687612/bin/biomolecules-12-01607-s001.zip

References

1. Gao S-J, Guo H, Luo G. Omicron variant (B.1.1.529) of SARS-CoV-2, a global urgent public health alert! J Med Virol. 2022 Apr;94(4):1255–6.

2. Gu H, Krishnan P, Ng DYM, Chang LDJ, Liu GYZ, Cheng SSM, et al. Probable Transmission of SARS-CoV-2 Omicron Variant in Quarantine Hotel, Hong Kong, China, November 2021. Emerging Infect Dis. 2022 Feb;28(2):460–2.

3. Callaway E. Heavily mutated Omicron variant puts scientists on alert. Nature. 2021 Dec;600(7887):21.

4. Desingu PA, Nagarajan K, Dhama K. Emergence of Omicron third lineage BA.3 and its importance. J Med Virol. 2022 Jan 18;

5. McCallum M, Czudnochowski N, Rosen LE, Zepeda SK, Bowen JE, Walls AC, et al. Structural basis of SARS-CoV-2 Omicron immune evasion and receptor engagement. Science. 2022 Feb 25;375(6583):864–8.

6. Shuai H, Chan JF-W, Hu B, Chai Y, Yuen TT-T, Yin F, et al. Attenuated replication and pathogenicity of SARS-CoV-2 B.1.1.529 Omicron. Nature. 2022 Mar;603(7902):693–9.

7. VanBlargan LA, Errico JM, Halfmann PJ, Zost SJ, Crowe JE, Purcell LA, et al. An infectious SARS-CoV-2 B.1.1.529 Omicron virus escapes neutralization by therapeutic monoclonal antibodies. Nat Med. 2022 Mar;28(3):490–5.

8. Kannan S, Shaik Syed Ali P, Sheeza A. Omicron (B.1.1.529) - variant of concern - molecular profile and epidemiology: a mini review. Eur Rev Med Pharmacol Sci. 2021 Dec;25(24):8019–22.

9. Thakur V, Ratho RK. OMICRON (B.1.1.529): A new SARS-CoV-2 variant of concern mounting worldwide fear. J Med Virol. 2022 May;94(5):1821–4.

10. Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, et al. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. Cell. 2020 Sep 3;182(5):1295-1310.e20.

11. Collier DA, De Marco A, Ferreira IATM, Meng B, Datir RP, Walls AC, et al. Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. Nature. 2021 May;593(7857):136–41.

12. Tian F, Tong B, Sun L, Shi S, Zheng B, Wang Z, et al. N501Y mutation of spike protein in SARS-CoV-2 strengthens its binding to receptor ACE2. eLife. 2021 Aug 20;10.

13. Liu Y, Liu J, Plante KS, Plante JA, Xie X, Zhang X, et al. The N501Y spike substitution enhances SARS-CoV-2 infection and transmission. Nature. 2022 Feb;602(7896):294–9.

14. McCallum M, Walls AC, Sprouse KR, Bowen JE, Rosen LE, Dang HV, et al. Molecular basis of immune evasion by the Delta and Kappa SARS-CoV-2 variants. Science. 2021 Dec 24;374(6575):1621–6.

15. Dejnirattisai W, Huo J, Zhou D, Zahradník J, Supasa P, Liu C, et al. SARS-CoV-2 Omicron-B.1.1.529 leads to widespread escape from neutralizing antibody responses. Cell. 2022 Feb 3;185(3):467-484.e15.

16. Cameroni E, Bowen JE, Rosen LE, Saliba C, Zepeda SK, Culap K, et al. Broadly neutralizing antibodies overcome SARS-CoV-2 Omicron antigenic shift. Nature. 2022 Feb;602(7898):664–70.

17. Zhang X, Wu S, Wu B, Yang Q, Chen A, Li Y, et al. SARS-CoV-2 Omicron strain exhibits potent capabilities for immune evasion and viral entrance. Signal Transduct Target Ther. 2021 Dec 17;6(1):430.

18. Cuzzolin A, Sturlese M, Deganutti G, Salmaso V, Sabbadin D, Ciancetta A, et al. Deciphering the Complexity of Ligand-Protein Recognition Pathways Using Supervised Molecular Dynamics (SuMD) Simulations. J Chem Inf Model. 2016 Apr 25;56(4):687–705.

19. Deganutti G, Prischi F, Reynolds CA. Supervised molecular dynamics for exploring the druggability of the SARS-CoV-2 spike protein. J Comput Aided Mol Des. 2021 Feb;35(2):195–207.

20. Huang J, MacKerell AD. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. J Comput Chem. 2013 Sep 30;34(25):2135–45.

21. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods. 2017 Jan;14(1):71–3.

22. Vanommeslaeghe K, MacKerell AD. Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing. J Chem Inf Model. 2012 Dec 21;52(12):3144–54.

23. Vanommeslaeghe K, Raman EP, MacKerell AD. Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. J Chem Inf Model. 2012 Dec 21;52(12):3155–68.

24. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021 Aug;596(7873):583–9.

25. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK Predictions. J Chem Theory Comput. 2011 Feb 8;7(2):525–37.

26. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res. 2004 Jul 1;32(Web Server issue):W665-7.

27. Doerr S, Harvey MJ, Noé F, De Fabritiis G. HTMD: High-Throughput Molecular Dynamics for Molecular Discovery. J Chem Theory Comput. 2016 Apr 12;12(4):1845–52.

28. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996 Feb;14(1):33–8, 27.

29. Harvey MJ, Giupponi G, Fabritiis GD. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. J Chem Theory Comput. 2009 Jun 9;5(6):1632–9.

30. 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.

31. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys. 1984;81(8):3684.

32. Loncharich RJ, Brooks BR, Pastor RW. Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N’-methylamide. Biopolymers. 1992 May;32(5):523–35.

33. Hopkins CW, Le Grand S, Walker RC, Roitberg AE. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. J Chem Theory Comput. 2015 Apr 14;11(4):1864–74.

34. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. J Chem Phys. 1995;103(19):8577.

35. Deganutti G, Moro S. Estimation of kinetic and thermodynamic ligand-binding parameters using computational strategies. Future Med Chem. 2017 Apr;9(5):507–23.

36. Sabbadin D, Moro S. Supervised molecular dynamics (SuMD) as a helpful tool to depict GPCR-ligand recognition pathway in a nanosecond time scale. J Chem Inf Model. 2014 Feb 24;54(2):372–6.

37. Deganutti G, Moro S, Reynolds CA. A Supervised Molecular Dynamics Approach to Unbiased Ligand-Protein Unbinding. J Chem Inf Model. 2020 Mar 23;60(3):1804–17.

38. Sabbadin D, Salmaso V, Sturlese M, Moro S. Supervised molecular dynamics (sumd) approaches in drug design. Methods Mol Biol. 2018;1824:287–98.

39. McGibbon RT, Beauchamp KA, Harrigan MP, Klein C, Swails JM, Hernández CX, et al. Mdtraj: A modern open library for the analysis of molecular dynamics trajectories. Biophys J. 2015 Oct 20;109(8):1528–32.

40. Miller BR, McGee TD, Swails JM, Homeyer N, Gohlke H, Roitberg AE. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. J Chem Theory Comput. 2012 Sep 11;8(9):3314–21.

41. Han P, Li L, Liu S, Wang Q, Zhang D, Xu Z, et al. Receptor binding and complex structures of human ACE2 to spike RBD from omicron and delta SARS-CoV-2. Cell. 2022 Feb 17;185(4):630-640.e10.

42. Rath SL, Padhi AK, Mandal N. Scanning the RBD-ACE2 molecular interactions in Omicron variant. Biochem Biophys Res Commun. 2022 Jan 6;592:18–23.

43. Kumar R, Murugan NA, Srivastava V. Improved binding affinity of the Omicron's spike protein with hACE2 receptor is the key factor behind its increased virulence. BioRxiv. 2021 Dec 28;

44. Lupala CS, Ye Y, Chen H, Su X-D, Liu H. Mutations on RBD of SARS-CoV-2 Omicron variant result in stronger binding to human ACE2 receptor. Biochem Biophys Res Commun. 2022 Jan 29;590:34–41.

45. Jawaid MZ, Baidya A, Mahboubi-Ardakani R, Davis RL, Cox DL. Simulation of the omicron variant of SARS-CoV-2 shows broad antibody escape, weakened ACE2 binding, and modest increase in furin binding. BioRxiv. 2021 Dec 15;

46. Wu L, Zhou L, Mo M, Liu T, Wu C, Gong C, et al. SARS-CoV-2 Omicron RBD shows weaker binding affinity than the currently dominant Delta variant to human ACE2. Signal Transduct Target Ther. 2022 Jan 5;7(1):8.

47. Wu L, Zhou L, Mo M, Li Y, Han J, Li J, et al. The effect of the multiple mutations in Omicron RBD on its binding to human ACE2 receptor and immune evasion: an investigation of molecular dynamics simulations. 2021 Dec 7;

48. Wang E, Sun H, Wang J, Wang Z, Liu H, Zhang JZH, et al. End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. Chem Rev. 2019 Aug 28;119(16):9478–508.

49. Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discov. 2015 May;10(5):449–61.

7. HIV Project Publication

Viral escape mechanisms pose numerous challenges to the development of potent antiretroviral therapies (ART). Such mechanisms include epitope mutation (1) and shielding (2,3), the alteration of the host's immune response and biological response to infection (4), as well as the manumission of the host cell's replicative cycle (5). Specifically, HIV-1 and HIV-2 of the Lentivirus genus of the Retroviridae family, adopt a combination of immune escape mechanisms that pose a difficult challenge to pharmacological research. While most of HIV's proteins are well known, the structural characterization of accessory proteins (Vif, Vpu, Vpr, Vpx, and Nef) along with structural /enzymatic (Gag, Pol, and Env) and gene-expression regulatory proteins (Tat and Rev) essential for viral replication is not fully understood. As the role of multifunctional proteins becomes clearer, the interest in these non-structural proteins increases among the scientific community, offering new opportunities for the development of new treatments or structural characterizations.

During the project, I collaborated with Kolkata University (India) the *de-novo* drug discovery of molecular disruptors against HIV-1's Negative Factor protein (Nef). In this phase of the project, I expanded the array of methods at my disposal, including generative combinatorial techniques (6) for new molecules and the use of water-based insights for ligand optimization. Furthermore, I used the new multiple-walker supervised molecular dynamics (mwSuMD, see Chapters 4.5 and 8.2) to describe Nef's binding pathway. The scientific publication related to these exertions is titled "Computer-aided De Novo Design and Optimization of Novel Potential Inhibitors of HIV-1 Nef Protein".

References

1. Caetano DG, Côrtes FH, Bello G, Teixeira SLM, Hoagland B, Grinsztejn B, et al. Next-generation sequencing analyses of the emergence and maintenance of mutations in CTL epitopes in HIV controllers with differential viremia control. Retrovirology. 2018 Sep 10;15 (1):62.
2. Peterlin BM, Trono D. Hide, shield and strike back: how HIV-infected cells avoid immune eradication. Nat Rev Immunol. 2003 Feb;3 (2):97–107.
3. McMichael AJ. HIV. The immune response. Curr Opin Immunol. 1996 Aug;8 (4):537–9.
4. Kazer SW, Walker BD, Shalek AK. Evolution and Diversity of Immune Responses during Acute HIV Infection. Immunity. 2020 Nov 17;53 (5):908–24.

5. Ferguson MR, Rojo DR, von Lindern JJ, O'Brien WA. HIV-1 replication cycle. Clin Lab Med. 2002 Sep;22 (3):611–35.

6. Spiegel JO, Durrant JD. AutoGrow4: an open-source genetic algorithm for de novo drug design and lead optimization. J Cheminform. 2020 Apr 17;12 (1):25.

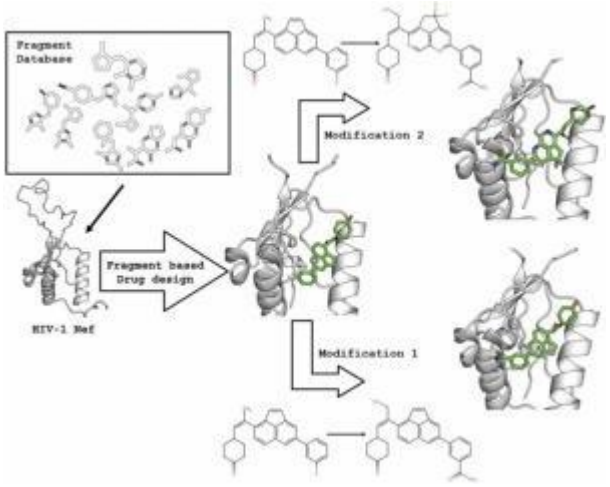
7.1 Computer-aided de novo design and optimization of novel potential inhibitors of HIV-1 Nef protein

Authors: Satyabrata Majumder, Giuseppe Deganutti, Ludovico Pipitò, Dwaipayan Chaudhuri, Joyeeta Datta, Kalyan Giri

Abstract

Nef is a small accessory protein pivotal in the HIV-1 viral replication cycle. It is a multifunctional protein and its interactions with kinases in host cells have been well characterized through many in vitro and structural studies. Nef forms a homodimer to activate the kinases and subsequently the phosphorylation pathways. The disruption of its homodimerization represents a valuable approach in the search for novel classes of antiretroviral. However, this research avenue is still underdeveloped as just a few Nef inhibitors have been reported so far, with very limited structural information about their mechanism of action. To address this issue, we have employed an *in silico* structure-based drug design strategy that combines de novo ligand design with molecular docking and extensive molecular dynamics simulations. Since the Nef pocket involved in homodimerization has high lipophilicity, the initial de novo-designed structures displayed poor drug-likeness and solubility. Taking information from the hydration sites within the homodimerization pocket, structural modifications in the initial lead compound have been introduced to improve the solubility and drug-likeness, without affecting the binding profile. We propose lead compounds that can be the starting point for further optimizations to deliver long-awaited, rationally designed Nef inhibitors.

Graphical Abstract



Keywords: HIV-1 Nef protein, De Novo ligand design, Supervised molecular dynamics

1. Introduction

Human immunodeficiency virus 1 (HIV-1), a member of the Lentiviral family, is the causative agent of acquired immune deficiency syndrome (AIDS). It encodes six accessory proteins (Tat, Rev, Vpu, Vif, Nef, and Vpr) along with other regulatory proteins (Sterbel et al., 2013). Nef is a relatively small (27–34 kDa) nonenzymatic auxiliary protein that is expressed early in the viral life cycle and performs a plethora of functions inside the host cells. Early research indicated that Nef was a "negative factor" for HIV-1 replication, which led to its initial naming (Ahmad and Venkatesan, 1988). Nef is an important player in viral pathogenesis by promoting viral replication and enabling immune escape in infected hosts (Staudt et al., 2020). Early pieces of evidence showed that expression of Nef in $CD4^+$ T-cells and macrophages causes AIDS-like disease in transgenic mice and that patients infected with *nef* defective HIV-1 fail to progress to AIDS suggesting a direct role for Nef in HIV-1 pathogenesis (Jolicoeur, 2011; Hanna et al., 1998, Rhodes et al., 2000; Kirchhoff et al., 1995).

Nef is composed of two main parts: the N-terminal anchor region and a folded core region in which a flexible internal loop (~25 residues) is present. The N-terminal 60 amino acids that make up the anchor region are mainly unstructured and contain a myristoylation site at the starting glycine residue. This myristoylation helps the Nef anchor to the flexible core of the membrane leaflet. It interacts with a plethora of membrane proteins and factors such as CD4, CXCR4, CCR5, MHC-I, T-cell receptor, SERINC5, trafficking proteins, guanine nucleotide exchange factors, protein kinases, etc. (Usami et al., 2015; Saksela, 2011; Gerlach et al., 2010, Geyer et al., 1999 Akgun et al., 2013; Jäger et al., 2011, Kent et al., 2010). Numerous structural and in vitro studies on the recruitment and activation of

Src-family kinases indicated that Nef preferentially binds to the SH3 domain of the Hck and Lyn kinases and displaces the SH3 domain from the kinase regulatory position resulting in constitutive kinase activation in vitro systems (Moarefi et al., 1997; Briggs et al., 1997). Nef-dependent activation of the Hck is a conserved function of all HIV-1 Nef M-group subtypes (Narute et al., 2012) Previous studies showed that both the expression of a dominant-negative Hck mutant and suppression of Hck expression impairs HIV-1 transcription and viral replication in macrophages (Komuro et al., 2003; Biggs et al., 1999). Structural studies show that Nef preferentially forms dimer when it is complexed with Fyn kinase SH3 domain and Hck kinase SH3-SH2 domain. The additional contacts resulting from the dimer formation contribute to the kinase activation (Alvarado et al., 2014; Staudt et al., 2020). It was shown that when L112 and Y115, two important residues for homodimerization, are replaced with aspartate residues (Nef-DD), the Nef-DD mutant interacts with Itk kinase but is unable to stimulate the kinase activity, indicating the essential role for Nef homodimer formation in the Itk activation mechanism (Poe et al., 2014). Another study showed that when Nef-ER (estrogen receptor) and Hck are expressed together in rodent fibroblasts, 4-hydroxytamoxifen treatment induces Nef-ER dimer formation, Hck activation, and oncogenic transformation (Ye et al., 2004). Nef residues that play a crucial role in homodimer formation (e.g. L112, Y115, F121) are highly conserved across HIV-1 M-group subtypes with many conserved or homologous amino acids in the corresponding positions of Nef from HIV-2 and SIV Nef (Narute et al., 2012; Arold et al., 2000). Nef harbours a conserved PxxPxR motif which serves as the docking site for the recruitment of Src-family kinases and other proteins that contain SH3 domains (Herna et al., 2000). To date, there are no reported Nef structures in complex with antiviral, small molecules, or binder, with only Nef dimer X-ray structure being characterized (PDB: 6B72). The homodimerization is mainly induced by the Nef αB helices and there exists a hydrophobic pocket that participates in many essential hydrophobic contacts to stabilize the dimer (Fig. 1) (Wu et al., 2018). Homodimerization is essential for Nef to activate the Kinase function. So disrupting the dimerization of Nef could prevent the dimer formation and the activation of the kinase.

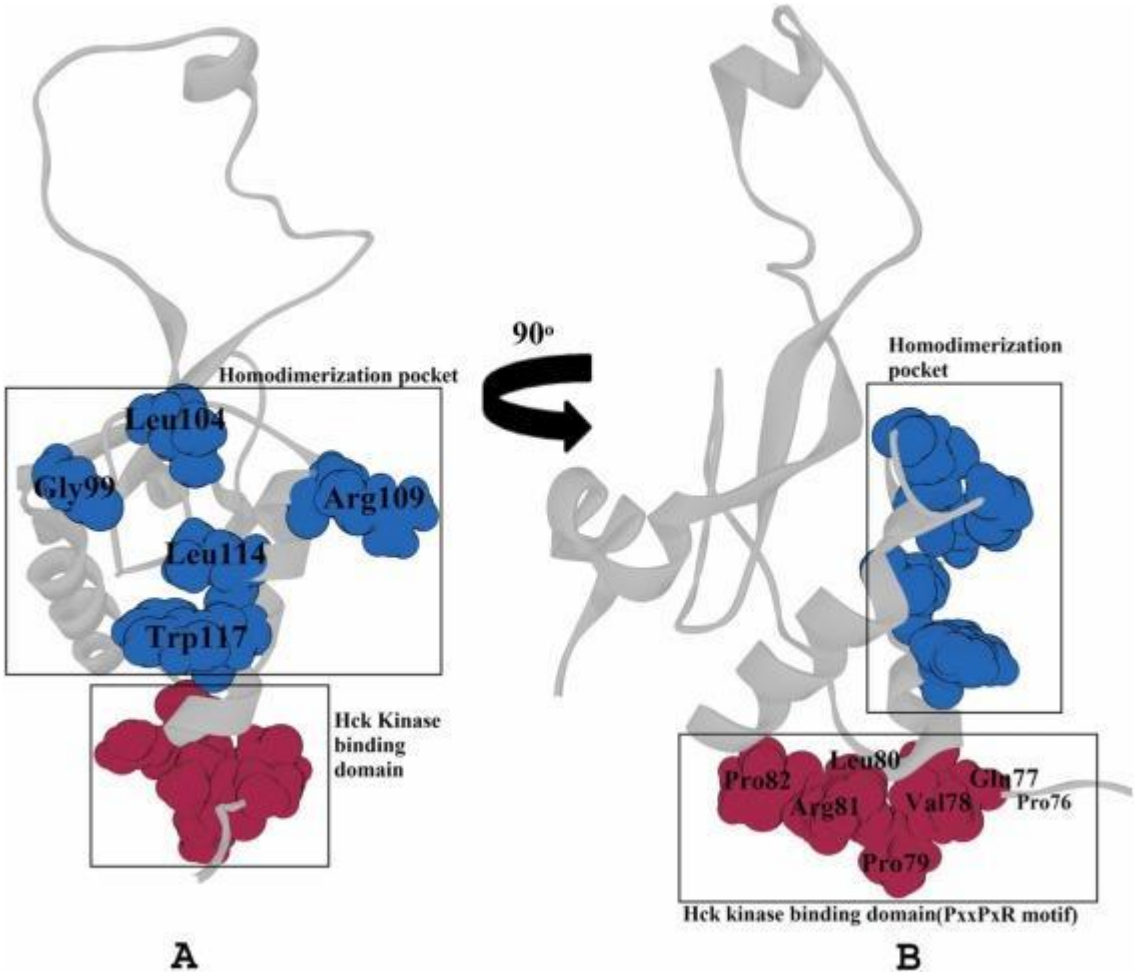


Fig. 1. Nef homodimerization pocket (blue residues) and Hck kinase binding domain (red residues). Residues are represented as van der Waals spheres.

To date, several compounds have been identified through high-throughput screening (HTS) assays mainly harnessing kinase activation by Nef for anti-retroviral drug discovery. A diphenyl hydroxypyrazolodiazene compound (also known as B9) was first identified through HTS and originally showed inhibitory activity against multiple Nef functions, including the enhancement of viral replication via activation of Src-family kinase and MHC-I downregulation (Emert-Sedlak et al., 2013; Mujib et al., 2017). Later, using the hydroxypyrazole core of B9, several new compounds were synthesized. They not only retained the MHC-I downregulation activity but also completely suppressed Itk kinase activity (Shi et al., 2020). However, whether these compounds bind to Nef directly or to the Nef-kinase complex and the mechanism of binding at the atomic level are still unknown.

In a previous study, using docking and molecular dynamics (MD) simulation, it was proposed that B9 binds to the Nef homodimerization surface, located far from the Hck-kinase binding site, and that the kinase becomes active after the homodimer formation (Moonsamy et al., 2017). However, B9 was docked onto the already-formed dimer and the possibility that this molecule binds to the dimer-forming pocket of the monomeric Nef was not explored. In another study, a class of 4-amino-diphenylfurano-pyrimidine (DFP) compounds were discovered in a small library of kinase inhibitor–based compounds. These compounds function through the Hck active site, but in the presence of Nef, they demonstrated increased potency for Hck inhibition, indicating that Nef binding may have an allosteric effect on the Hck active site to improve the inhibitor binding (Emert-Sedlak et al., 2009). This finding is further supported by subsequent hydrogen exchange mass spectroscopy (HX MS) studies of the Nef-Hck complex in the presence of a DFP-based compound with antiretroviral activity (Wales et al., 2015). A phenoxyacetamido benzoic acid analogue called D1 was identified using combined *in silico* screening with a cell-based protein-protein interaction assay that blocked Nef-SH3 complex formation (Betzi et al., 2007). Several other compounds, such as the isothiazolopyridinone analog SRI-37264, the synthetic analog known as '2 C' which was derived from the *Streptomyces* metabolite UCS15A, lovastatin, and concanamycin A were shown to inhibit Nef mediated MHC-I downregulation. It was found that compounds that block the MHC-I downregulation pathways, also block the Nef-kinase activity, but whether these compounds could prevent the homodimerization event is not understood (Emert-Sedlak et al., 2022).

In this study, we first designed *in silico* novel compounds to specifically target the Nef pocket fundamental for homodimer formation, through a fragment *de novo* approach, combined with extensive MD simulation and molecular mechanics with generalized Born and surface area solvation (MM/GBSA) binding free energy calculation. The best compound was further analyzed using the ColDock method (Takemura et al., 2018). Structural modifications were informed by protein hydration site analysis and led to the evaluation of further modified compounds with lower lipophilicity and increased drug-likeness. Using multiple walker supervised molecular dynamics (mwSuMD) simulations, the mechanism of binding of the most promising compound inside the pocket was assessed (Deganutti et al., 2022). This study represents the first rational design of Nef homodimerization inhibitors and paves the way for a new antiretroviral therapy.

2. Materials and methods

2.1. Protein structure preparation

We retrieved the Nef protein structure from the Protein Data Bank (PDB) entry 4U5W which is a dimer in complex with Hck SH3-SH2 kinase (Alvarado et al., 2014). We excluded the unstructured region 1–71, which harbors only the myristoylation sites for membrane fusion. Nef from 4U5W was compared with all other available Nef structures (containing the region of interest) in the PDB and found to be very similar in terms of 3D structure (RMSD of the backbone atoms < 1.5 Å). This step ensures that Nef does not undergo major conformational change upon binding with various cellular partners. Modelling of the missing segments was done in CHARMM-GUI (Lee et al., 2016). Hydrogen atoms were added considering the pH of 7.4; the protonation of the titratable side chains was checked by visual inspection.

2.2. Identification of druggable pockets

To identify viable druggable binding sites on Nef, we used DeepSite and FTMap (Jiménez et al., 2017, Kozakov et al., 2015). DeepSite is a knowledge-based approach where 7622 proteins from the scPDB database of binding sites were used to construct the convolutional neural networks (CNN) (evaluated using both a distance and a volumetric overlap approach) (Desaphy et al., 2014). Pockets with a high DeepSite score (in the range of 0–1) were taken into consideration. FTMap method distributes 16 small organic probe molecules of varying shape, size, and polarity on the protein surface; then it finds the most favorable positions for each probe type, clustering and ranking the results on the basis of their average energy. Clusters with the highest density of probes are considered possible hotspots.

2.3. Fragment-based small molecule design

AutoGrow4 was used for fragment-based drug design (FBDD) to generate novel drug-like molecules (Spiegel and Durrant, 2020). It uses a genetic algorithm to create new predicted ligands. First, an initial population of seed molecules is used to create a new population (i.e. a generation) of potential ligands. These compounds are then docked using Autodock VINA (Trott and Olson, 2010) to a user-specified protein pocket and the poses are ranked by their calculated fitness. New generations are seeded from the top-scoring (VINA score expressed in kcal/mol) molecules of the previous generation. In this study, we have used all its default fragment library molecules, which are divided into four groups according to molecular weight ($MW \leq 100$ Da, 100 Da $< MW \leq 150$ Da, 150 Da $< MW \leq 200$ Da, and 200 Da $< MW \leq 250$ Da) and are obtained from ZINC15 database (Sterling and Irwi, 2015). OpenBabel 2.3.1 was used to convert the SMILES fragment libraries into 3D pdbqt format, adding hydrogen atoms and optimizing the intramolecular hydrogen-bond network (pH 7.4) (O'Boyle et al., 2011). We set a docking grid of 22 Å × 20 Å× 20 Å at the specified pocket by taking the center of mass (COM) of the FTMap probe cluster (probes that bound with pocket 3) as the grid

center (grid center coordinate: −1.57, −0.653, 19.762 in x, y and z direction). RobustRxn set, a library of 58 reactions, was used for the in silico reactions. Evolving molecules were subjected to the lenient form of Lipinski's filter to increase the drug-like properties. The number of generations was set to 50 for all the MW groups but 150 Da < MW ≤ 200 Da fragments, as the docking affinity (VINA score), reached a plateau from generation 24 (see result and discussion). Docking exhaustiveness was set to 15. The docking of modified compounds into the Nef pocket utilized the same grid coordinates, grid box, and other parameters as in Autogrow4. Prior to docking, we calculated the Kollhman charge of Nef, which was determined to be −5. Additionally, the Gasteiger charges of the modified compound were calculated to be 0. The top 10 poses were generated for each docking run.

2.4. General MD simulation setup and MMGBSA binding free energy calculation

The MD engine GROMACS 2020.5 was used for both equilibration and productive simulations (Abraham et al., 2015). The AMBER99SB-IILDN protein force field was used for the simulations and ligand molecules were parameterized using the Acpype tool by introducing bcc charge types and GAFF atom types (Lindorff-Larsen et al., 2010, Sousa da Silva and Vranken, 2012. Nef was solvated using TIP3P water molecules in a dodecahedron box (box edge of 12 Å) and the box edge distance from protein atoms was set to 10 Å. Overall charge neutrality was reached by adding Na^+/Cl^- counter ions up to the final concentration of 0.150 M. Energy minimization of the neutralized system was done using 5000 steps of the steepest-descent algorithm. 500 ps long NVT and NPT equilibrations were run with a positional constraint of 1 kcal/mol $Å^{-2}$ on protein and ligand heavy atoms at 300 K temperature. The NPT equilibrated system was then subjected to 100 ns long production runs with 2 fs timestep. Twenty independent simulation runs were conducted for Nef in pseudo-apo form (total of 2 μs), while for Nef-ligand complexes, it was triplicated. MMGBSA binding free energy was calculated using the gmx_MMPBSA tool (Valdés-Tresanco et al., 2021). Each trajectory was saved at every 100 ps resulting in a total of 1000 snapshots and using this trajectory the free energy is calculated for every run. General MD simulation analyses such as radius of gyration (Rg), root mean square fluctuation (RMSF), and solvent accessible surface area (SASA) were calculated with GROMACS's internal tool, using default settings (Gupta et al., 2022).

2.5. Principal component analysis (PCA)

PCA is used in molecular dynamics data to study essential motions of the protein backbone and sample out the most probable conformation of the protein. It is also used to construct the free energy landscape of the proteins where the free energy is mainly plotted along the first

two principal components (PCs). In summary, to obtain PC modes first correlation matrix is computed using the following formula:

$$C_{ij} = \frac{1}{N} \sum_{k=1}^{N} \left(x_i^k - \bar{x}_i\right)\left(x_j^k - \bar{x}_j\right)$$

Here x (k, i) and x (k, j) are a pair of elements of vector (where is a vector containing the cartesian coordinates of the atoms of the protein), which describes the configuration of the system at time step, while, are their average values calculated from the structures sampled in the MD simulation. Then this correlation matrix is diagonalized using the formula $\Lambda = T^T CT$ where T is the transformation matrix whose columns are the eigenvectors of the motions and the diagonal elements of Λ are the associated eigenvalues (Majumder et al., 2021). PCA was carried out using PyEMMA 2.5.7 and the plotting was done using matplotlib (Scherer et al., 2015).

2.6. ColDock

In the ColDock method, multiple ligands were randomly placed within 20 Å of the protein atoms inside a cubic box with dimensions 30 Å × 30 Å × 30 Å (Takemura et al., 2018). We prepared separate systems with the following numbers of ligand molecules: 6, 8, and 10. Systems were solvated with TIP3 water molecules and salt concentration was set to 0.15 M. AMBER99SB-IILDN force field was used for the simulations and five independent runs of 50 ns each were conducted: 2 × 6 ligand system, 2 × 8 ligand system, 1 × 10 ligand system. This high concentration of ligands increases the likelihood that the ligand will explore the protein surface, leading to spontaneous binding to the proper binding sites. Ligand aggregation is prevented during MD simulation by imposing repulsive force E_{LJ} in the form of a Lennard-Jones potential between ligand molecules:

$$V_{\mathrm{LJ}}(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

where $\varepsilon = 10^{-3}$ kJ/mol, $\sigma = 20$ Å, and r is the distance between the centres of mass of the ligands. When ε is extremely low, the attractive force is negligibly small in comparison to thermal noise, and if two ligands approach one another within 10 Å, the repulsive force becomes significantly large. This prevents aggregation of the ligands. Preliminary MD simulations of 50 ns have been conducted using 33 ligand molecules inside a 12 Å × 12 Å × 12 Å cubic box with TIP3 water molecules and 0.15 M Na^+ and Cl^- ion concentration. Then

radial distribution function (RDF) is computed using the central C18 atom to see if the parameters prevent aggregation.

Without adding a virtual site, the ligand molecules aggregated (peak around 6 Å) due to the hydrophobic interactions. When a repulsive force was added, the ligand molecules did not aggregate. The flat value of radial distribution function (g (r)) around unity at r > 10 Å indicates a uniform distribution of ligand molecules during the MD simulation at this distance range. When the ligands came closer to each other (<10 Å), the repulsive force pushed the ligands from each other and prevents aggregation. The distance between protein atoms and ligands was kept at least 20 Å.

After conducting Nef-multiple ligands system simulations, ligand poses were clustered. During clustering, the ligands in contact with N_{thr} (threshold number, 6 by default) or more protein residues were considered to be bound to Nef. A contact is defined when the shortest heavy atom distance between the protein and the ligand is less than 5 Å. This procedure permits the selection of multiple ligand poses from one MD snapshot. Each ligand pose is translated and rotated so that the corresponding protein pose is superimposed onto a reference; then the ligand poses are clustered based on a defined RMSD cutoff (2 Å is used in this study).

2.7. AquaMMapS water analysis

The Nef protein structure was retrieved from the PDB entry 4U5W and prepared for the CHARMM36 force field as follows: the Hck SH3-SH2 kinase monomer was removed and hydrogen atoms were added employing the pdb2pqr and propka software (considering a simulated pH of 7.0); the protonation of titratable side chains was checked by visual inspection and TIP3P water molecules were added to the simulation box utilizing the VMD Solvate plugin 1.5 (Solvate Plugin, Version 1.5. at <http://www.ks.uiuc.edu/Research/vmd/plugins/solvate/) (Huang et al., 2017, Huang and MacKerell, 2013; Dolinsky et al., 2004; Jorgensen et al., 1983, Olsson et al., 2011). Finally, overall charge neutrality was reached by adding Na^+/Cl^- counter ions up to the final concentration of 0.15 M, using the VMD Autoionize plugin 1.3 (Autoionize Plugin, Version 1.3. at <http://www.ks.uiuc.edu/Research/vmd/plugins/autoionize/). The MD engine ACEMD3 was employed for both the equilibration and the productive simulation (Harvey et al., 2009). The equilibration was achieved in isothermal-isobaric conditions (NPT) using the Berendsen barostat (target pressure 1 atm) and the Langevin thermostat (target temperature 300 K) with

low damping of 1 ps^{-1} (Berendsen et al., 1984; Loncharich et al., 1992). A multi-stage procedure was performed (integration time step of 2 fs): first, clashes between atoms were reduced through 500 conjugate-gradient minimization steps, then 1 kcal/mol $Å^{-2}$ positional restraints on protein atoms other than Cα was gradually removed over 4 ns, keeping the same restraints on protein Cα atoms. The productive trajectory was computed with an integration time step of 2 fs in the canonical ensemble (NVT). The target temperature was set at 300 K, using a thermostat damping of 0.1 ps^{-1}; the M-SHAKE algorithm was employed to constrain the bond lengths involving hydrogen atoms (Forester and Smith, 1998; Krautler et al., 2001). The cut-off distance for electrostatic interactions was set at 9 Å, with a switching function applied beyond 7.5 Å. Long-range Coulomb interactions were handled using the particle mesh Ewald summation method (PME) by setting the mesh spacing to 1.0 Å (Essmann et al., 1995). Structural water molecules were detected in the simulation cell using AquaMMapS (Cuzzolin et al., 2018). A short 10 ns simulation was performed with time step 2 fs, restraining the Cα atoms and saving a frame every 10 ps of simulation.

2.8. Multiple walkers supervised molecular dynamics (mwSuMD) simulations

The Nef protein structure was retrieved from the PDB entry 4U5W and prepared with the Amber14SB force field. A single molecule of ligand was placed about 35 Å from pocket 3 and parameterized for the GAFF force field using Antechamber (Maier et al., 2015; Wang et al., 2004). Six independent mwSuMD replicas were performed (Deganutti et al., 2022). mwSuMD needs as input the initial coordinates of the system as a PDB file, the coordinates, and the atomic velocities of the system from the equilibration stage, the topology file of the system, and all the necessary force field parameters. Then, a series of batches of short unbiased MD simulations (walkers) are performed (integration time step if 4 fs in this work). When all the walkers of a batch are terminated, the best walker is selected and extended by seeding the same number of walkers, with the same duration as the step before. The decision to continue one walker after any batch was based on the distance between the centroid of the ligand and the centroid of pocket 3 residues' L95, L101, and L114. The single metric score (SMscore) score was used: (1)

$$SMscore = \sqrt{X_{last\ \ \ frame} * \overline{X}}$$

The SMscore is computed as the square root of the product between the distance in the last frame ($X_{last\ frame}$) and the average distance value over the short simulation ($\overline{X}$). After any batch of walkers, the walker with the lowest SMscore was continued by seeding the new batch of

walkers. The current implementation of mwSuMD is for python3 and exploits MDAnalysis and MDTRaj modules (Michaud-Agrawal et al., 2011; McGibbon R.T. et al., 2015). Atomic contacts were computed using the GetContacts analysis tool (at https://getcontacts.github.io/), with the donor-acceptor threshold distance set to 3.5 Å and the angle set to 120°. The MMPBSA.py script, from the AmberTools20 suite (The Amber Molecular Dynamics Package, at http://ambermd.org/), was used to compute molecular mechanics energies combined with the generalized Born and surface area continuum solvation (MM/GBSA) method (Miller et al., 2012).

3. Results

3.1. Selection of the most suitable pocket for FBDD

To identify putative druggable sites on Nef, we employed both the convolutional neural network (CNN) based predictor DeepSite and the probe scanning-based method FTMap. We considered the three pockets with the highest DeepSite score, predicted in the Nef core region (79–153 and 184–201, amino acid numbering based on the crystal structure 4U5W). Then we compared the DeepSite predicted pockets with the FTMap result and found that probe densities were highest in the three predicted DeepSite pockets than in other sub-pockets (Fig. 2). Residues that constitute each pocket were identified as the residues within 3.5 Å from the FTMap cluster COMs.

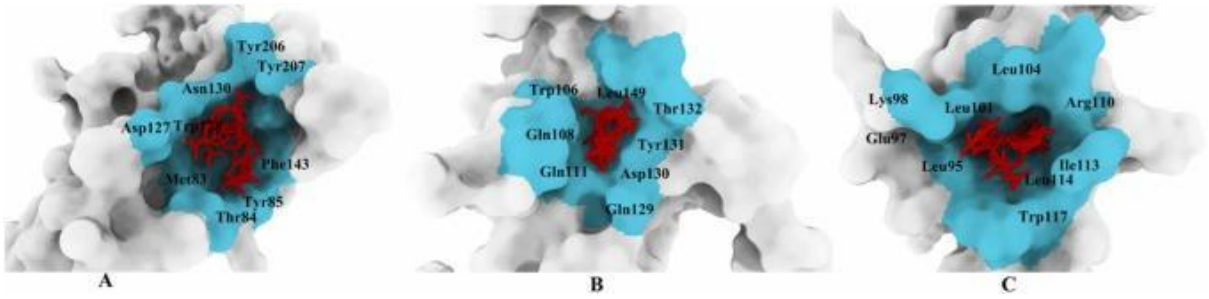


Fig 2. Top 3 pockets predicted through DeepSite and FTmap. A) Pocket 1, B) pocket 2, C) pocket 3. The pocket grooves are colored cyan and important pocket residues are labeled. FTMap probes are colored in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

From 20 independent MD simulations of monomeric apo Nef, we computed the root mean square deviation (RMSD) probability histogram of each pocket and found that pocket 3 was the most promising for structure-based drug design (SBDD) because it displayed very low overall RMSD (≤ 1.5 Å) indicating inherent stability. Pocket 1 was considered less promising

than pocket 3 because of a secondary RMSD peak around 3 Å indicating slight instability in some of the trajectories. Based on PDB entry 6B72, the residues lining pocket 3 are very important for dimerization where the dimer is without any kinase. The important dimer forming contacts are (numbering based on PDB entry 6B72): L100 and R106 form two Pi-alkyl interactions with the phenyl ring of F68 from the other monomer; P69 and L110 form Pi-alkyl interactions with W113, while the backbone oxygen atom of the W113 forms a hydrogen bond with G96; R105 and R106 form three important salt bridges with E64. We analyzed the structural integrity of the core region (79–153 and 184–201), containing the three pockets, through PCA analysis considering only the core region backbone atoms, as the loop region would have added noises because of its inherent high flexibility. The first four principal components (PCs) were considered as they correspond to 85 % of the total variance. Representative structures were taken from the high-density region of each combination of PC plots and an extra structure was taken from the low-density region of the PC1 vs PC2 plot, randomly. The heatmap shows that the randomly chosen structure from the low-density region (named PC1_PC2_2) has slightly higher RMSD values in comparison to all other extracted structures (< 2.4 Å in most cases) and the PDB 4U5W (2.8 Å). The PCA-extracted structures are also very similar to all other available Nef structures in the PDB (mean RMSD ≤ 2.0 Å). This similarity confirms that the pockets formed in the core region are very stable and suitable for FBDD. Some high flexibility was present in the residues lining pocket 3, although this did not affect the overall pocket RMSD. Pocket 3 groove is highly hydrophobic but presents some H-bond donors on the pocket boundary (R110, L114, G99, and G100). In PDB 4UW5, the backbone of R110 forms a hydrogen bond with (4 S)− 2-methylpentane-2,4-diol cosolvent molecule, suggesting a possible role for this residue's backbone in ligand design. Furthermore, G100, E97, and F94 form backbone H-bonds with a water molecule.

3.2. *De-novo* generation of a lead compound

We employed Autogrow4 to construct a lead compound against pocket 3, halting the algorithm at generation 30 because docking scores reached a plateau after generation 24. Fig. 3 reports the docking scores per generation and the top-ranked compounds. One common structural feature is the presence of lipophilic ring moieties in the scaffold, as a consequence of the hydrophobic nature of pocket 3.

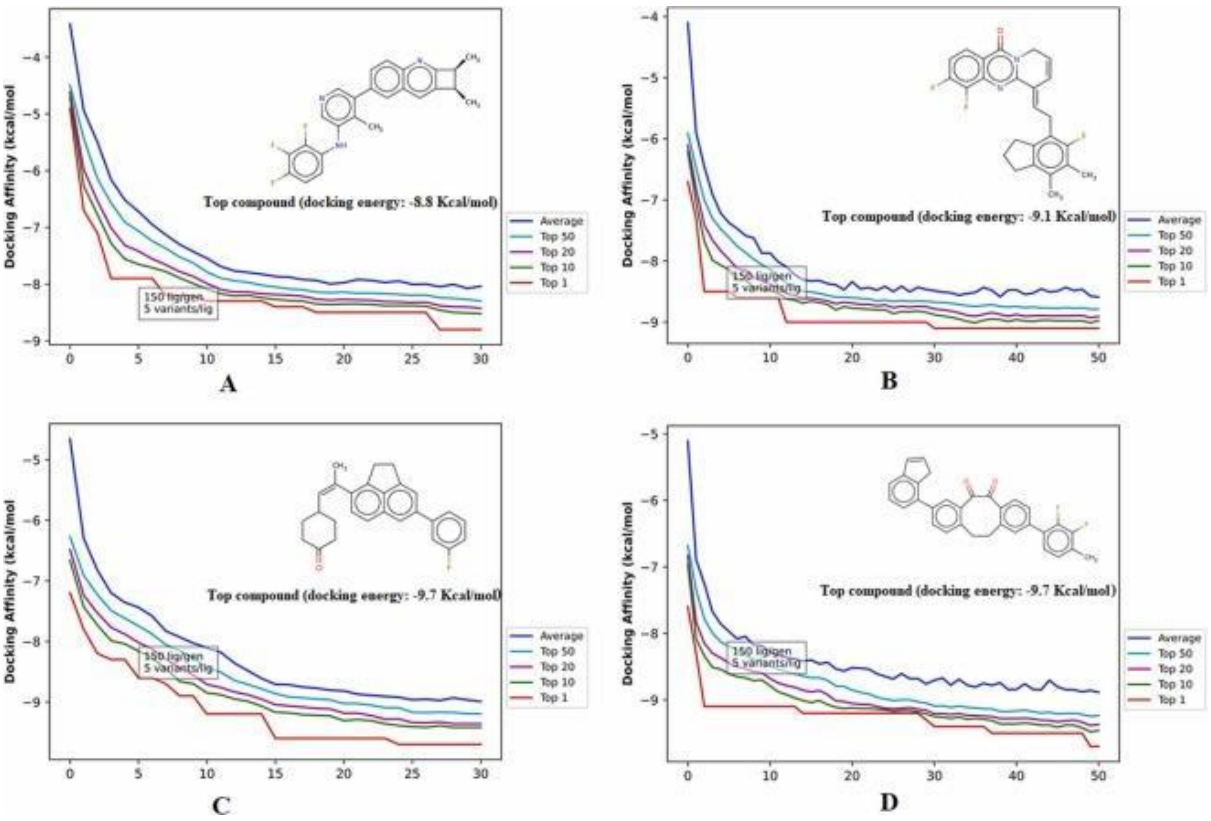


Fig. 3. Autogrow4 generated compounds using ZINC fragment library of molecular weight (MW) A) ≤100; B) 100≤MW ≤150; C)150≤MW ≤200; D) 200≤MW ≤250. The docking energy of the top compound from each category is reported.

Various physicochemical properties of the compounds were calculated using SwissADME[14]**.** The four top-ranked compounds are poorly soluble in water, and in the docked poses formed multiple hydrophobic contacts with pocket 3. No H-bond was formed due to the absence of any donor or acceptor group in the aromatic scaffolds. We considered the compounds with the highest VINA score (−9.7 kcal/mol) for further MD simulations, named compound 1 and compound 2.

3.3. MD simulation of selected compounds

We conducted five independent replicas of 100 ns for compound 1 or compound 2 Nef docked complex, starting from the same energy-minimized system but assigning different initial velocities. RMSD values distribution (Fig. 4) suggested that compound 2 was more stable than compound 1, as indicated by the high probability of RMSDs around 2.0 Å (Fig. 4B). On the other hand, compound 1 displayed overall instability and it did not retain the initial docking pose during all the simulations (lowest RMSD peak at 3 Å and another prominent peak around 7 Å). To investigate the reason behind compound 2 stabilization, we

performed the PCA analysis of the ligand (using the concatenated trajectories) followed by K-means clustering using the first 20 PCs and the top four densest clusters. We also extracted the water molecules within 3.5 Å from the ligand's heavy atoms to analyze any kind of water involvement. The oxygen atom of the cyclohexanone moiety was solvent-exposed in every cluster but did not engage Nef in direct electrostatic interactions.

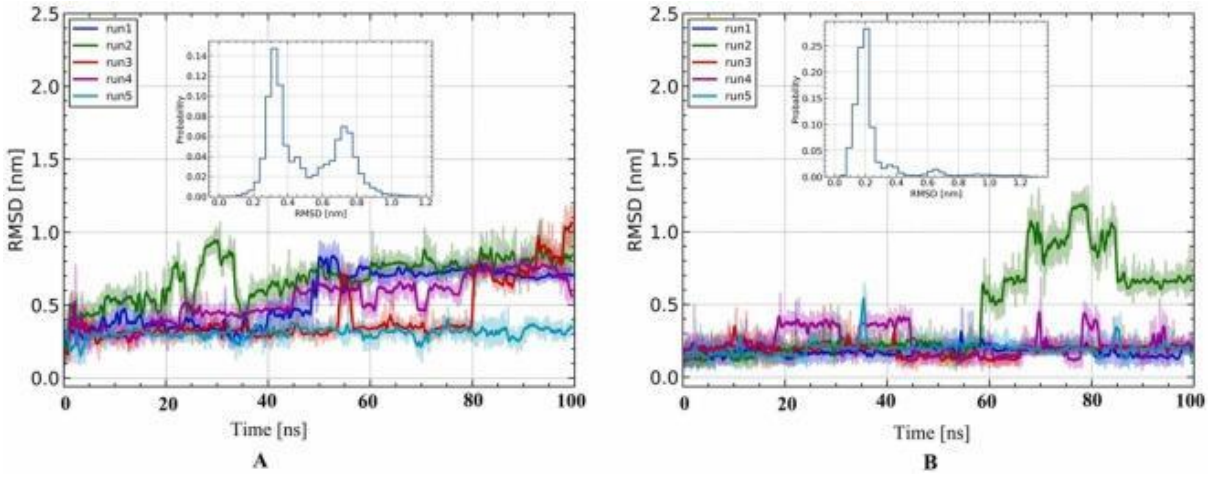


Fig. 4. RMSD graph (with a running average of 100 ps interval) and probability histogram (inset) of A) compound 1 and B) compound 2.

The fluorine (F) atom also formed H-bonds with water molecules in three out of four top cluster representative structures. In cluster 2 and cluster 4, one of the water molecules that formed H-bonds with the oxygen atom of the ligand also formed two potential H-bonds with R110. On the other hand, a water molecule that contacted the F atom in cluster 2 also formed H-bonds with E97 and F94. These water bridge interactions likely compensated for the lack of direct H-bonds between compound 2 and Nef. This kind of water molecule-mediated stabilization is absent in compound 1. In the crystal structure (4U5W) there are water molecules close to F94, E97, P99, G100, and an MPD molecule that forms an H-bond with R110 backbone, which suggests that compound 2 mimics a similar H-bond network. The MMGBSA analysis indicated similar binding energy for compound 1 (−27.92 ± 2.3 kcal/mol) and compound 2 (−24.53 ± 2.3 kcal/mol).

3.4. Assessing the accessibility of compound 2 toward pocket 3

To determine if compound 2 can spontaneously bind to the Nef pocket 3, we performed ColDock simulations. We run a total of five simulations (two replicas considering six ligand molecules, two replicas considering eight ligand molecules, and one replica considering 10 ligand molecules). Ligand binding to several distinct positions was observed during the

simulation. After 50 ns, compound 2 poses were clustered and representative structures were extracted from the largest cluster. Two replicas (one from the six-ligands system and the other from the ten-ligands system) out of five produced poses very similar to the VINA docked pose (Fig. 5**),** indicating that compound 2 is a potential binder for pocket 3. Convergent binding poses were also obtained with lower simulated ligand concentration. In another two replicas (one six-ligand system and one eight-ligand system), we found that the cyclohexanone moiety of compound 2 in the highest cluster tends to reach toward pocket 3 but did not produce the native VINA pose. This data also suggests the specificity of compound 2.

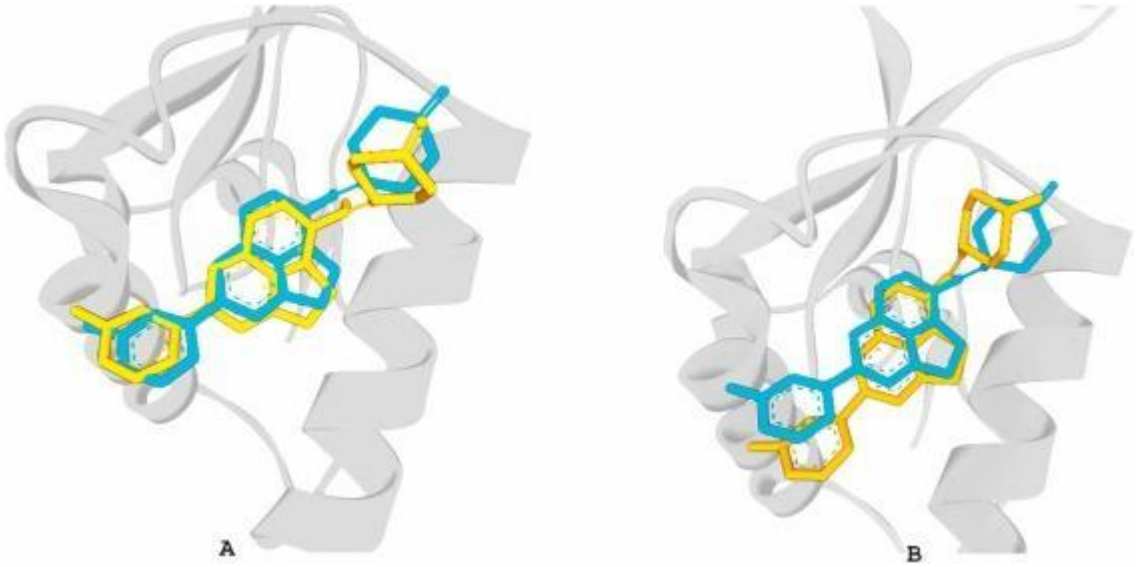


Fig. 5. Alignment of VINA docked pose (Cyan) and ColDock generated pose (Orange). A) Central representative snapshot from the largest cluster of the six-ligands systems. RMSD between two poses is 1.37 Å. B) Central representative snapshots from the largest cluster of the eight-ligand system. RMSD between two poses is 1.49 Å. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.5. Ligand modification to increase the solubility and druggability

Compound 2 (Chemical name: 4-[(1Z)− 2-[7-(3-fluorophenyl)− 1,2-dihydroacenaphthylen-3-yl] prop-1-en-1-yl] cyclohexan-1-one) is very lipophilic as it bears only an H-bond acceptor, and no H-bond donor atoms. It is predicted to be poorly soluble in water (as evident from SwissADME prediction) in the following solubility prediction methods: log S (ESOL), log S (Ali), log S (SILICOS-IT). Its pharmacokinetics is likely to be also very poor. To fine-tune the drug-likeness and the solubility of compound 2,

we introduced some structural modifications. We reasoned that it would be possible to mimic the hydrogen bonds between structural water molecules and Nef. In 4U5W, a water molecule is coordinated by the backbone within the loop formed by F94, E97, and G100, while an oxygen atom of the co-solvent MPD forms a hydrogen bond with the backbone of R110. We also inspected the structural water molecules in the apo Nef through the AquaMMapS analysis. Interestingly, a stable water molecule was suggested in correspondence with the MPD oxygen atom, confirming the presence of a hydrophilic spot within pocket 3. Therefore, we designed a new ligand, compound 2_mod-1 (chemical name: 3-{8-[(1E)− 3-hydroxy-1-(4-oxocyclohexyl) prop-1-en-2-yl]acenaphthylen-4-yl}benzamide) by adding a hydroxyl group (-OH) to the methyl group of compound 2 to introduce an H-bond with the R110 backbone. Also, a carboxamide ($CONH_2$) group was added in place of the fluorine atom because this H-bond donor/acceptor would be able to interact with the backbone of residues F94 and G100 as suggested by the water molecule present in the X-ray structure. In a further ligand, Compound 2_mod-2 (chemical name: 3-{2,2-difluoro-5-[(1E)− 3-hydroxy-1-(4-oxocyclohexyl)prop-1-en-2-yl]− 6,9-diazatricyclo[6.3.1.04,12]dodeca-1 (12),4,6,8,10-pentaen-10-ylbenzamide), we replaced two carbon atoms of the planar scaffold of modification 1 with nitrogen atoms, retaining aromaticity, and two hydrogens with fluorine atoms at one of the methylene groups of the scaffold, to increase the polarity of the molecule (Fig. 6). These modifications increased the computed solubility and drug-likeness of the molecules. Compound 2_mod-1 and compound 2_mod-2 were docked in pocket 3 using VINA taking the same Autogrow grid center. As expected, in compound 2_mod-1 the OH group was predicted to make an H-bond with the backbone of R110, while the $CONH_2$ group formed an H-bond with the backbone of F94. For compound 2_mod-2, instead, the $CONH_2$ group was predicted to form an H-bond with E97 and the OH to H-bond with L114. Several new hydrophobic contacts between Nef and modified compounds were seen (Fig. 7).

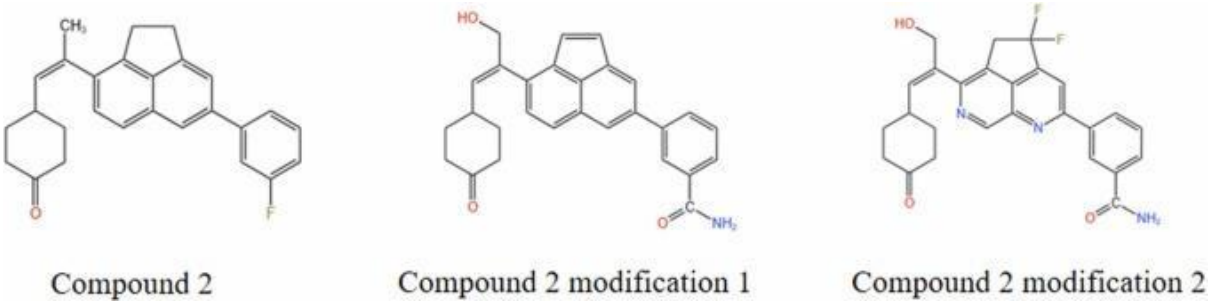


Fig. 6. Modifications of the original lead compound (compound 2) to increase the solubility and drug-likeness.

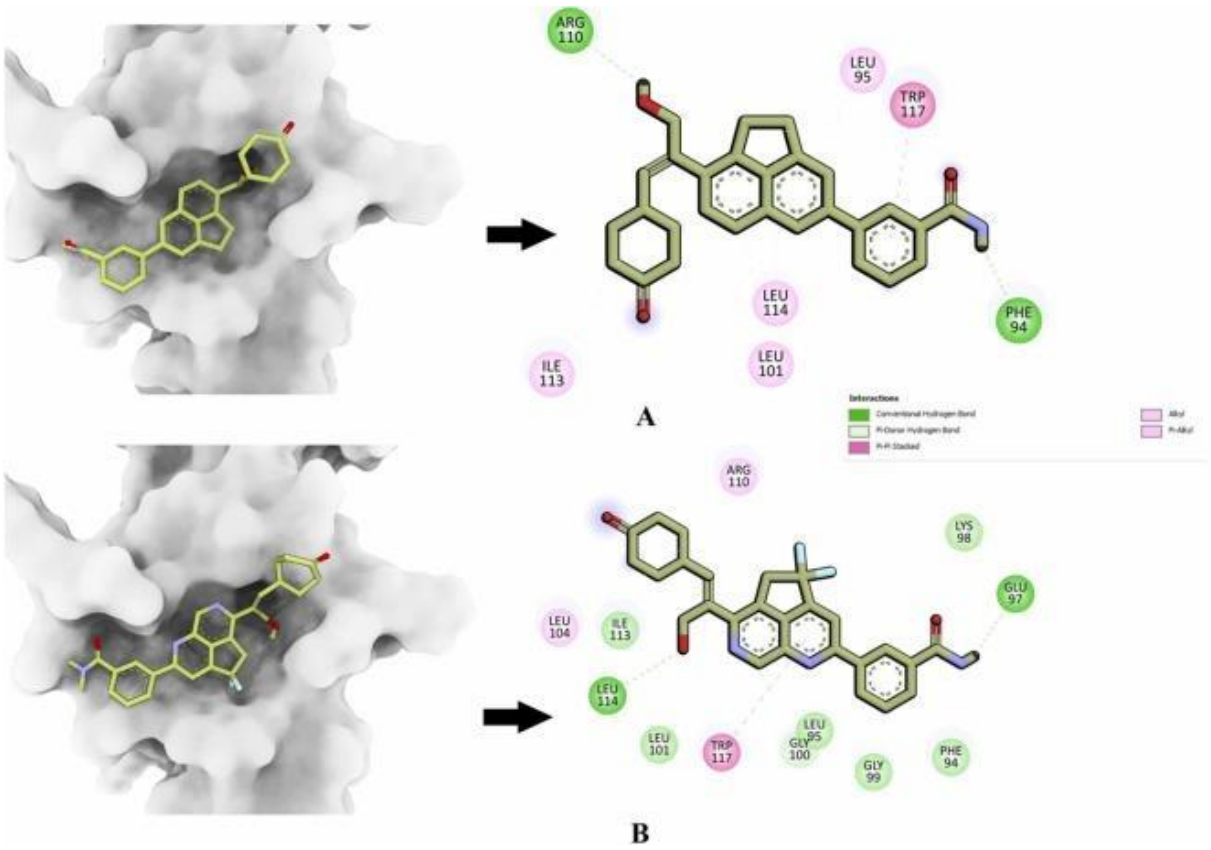


Fig. 7. Docked poses of the modified compounds with Nef. A) Compound2_mod-1 and B) compound2_mod-2. Dashed colored lines are the various interaction types. 2D interaction diagrams are generated in BIOVIA discovery studio.

Three independent 100 ns post-docking MD simulations were carried out for each Nef-modified compound, and the MMGBSA binding free energy was calculated. Although the modifications increased the number of H-bonds as expected (Fig. 8**C)**, compound 2_mod-1 was less stable than the original compound 2, as evident from the RMSD values throughout the trajectory (Fig. 8A), showing a peak around 4 Å and another large peak at about 5.5 Å. Compound 2_mod-2, on the other hand, was more stable inside the pocket, with a larger peak at 1.4 Å and a smaller peak with a relatively low probability at 3.5 Å. From the PCA-based clustering analysis, we extracted the representative structures from the three largest clusters and analyzed the poses and interaction patterns. Compound 2_mod-1 disrupted the starting binding pose and formed solely an H-bond with H170, which is part of the loop region. Also, simulations displayed the involvement of the N-terminal random coil region (residues 72–78) of Nef in the ligand binding. P73 made a hydrophobic contact with the ligand molecule. In this pose, two H-bonds were formed with G100 and E162, which are in the loop region. Only hydrophobic contacts restrained compound 2_mod-1 within pocket 3. The N-terminal random coil region made several hydrophobic contacts also with compound

2_mod-2 to stabilize it inside the pocket groove (Fig. 8B) in all three cluster representatives. Interestingly, compound 2_mod-2 flipped by 180° along its major axes relative to the initial VINA docked pose. The OH group, which made an H-bond with L114 in the initial pose, reoriented out of the groove and does not involve in any H-bond with Nef. Another stabilizing interaction of compound 2_mod-2 was the H-bond between carboxamide ($CONH_2$) and F94 backbone during the full course of the simulation. The MMGBSA free energy calculation (Fig. 8D) shows an opposite scenario to the RMSD analysis; compound 2_mod-1 has a relatively stronger binding free energy (−52.22 ± 6.5 kcal/mol) than compound 2_mod- 2 (−35.6 ± 2.91 kcal/mol). We also computed other parameters such as RMSF of the residues responsible for binding the compounds, the Rg of Nef backbone atoms, and SASA for the comparison with apo-Nef. Rg and SASA show a similar pattern for apo-Nef and complexed Nef indicating no drastic effect of the compounds on the whole protein structural level. From RMSF data we found that the compounds significantly reduce the fluctuations of the residues in pocket 3 in comparison with the apo-Nef, consistently with a reduction of the local kinetic energy in favor of the potential energy necessary to form noncovalent interactions with the ligands.

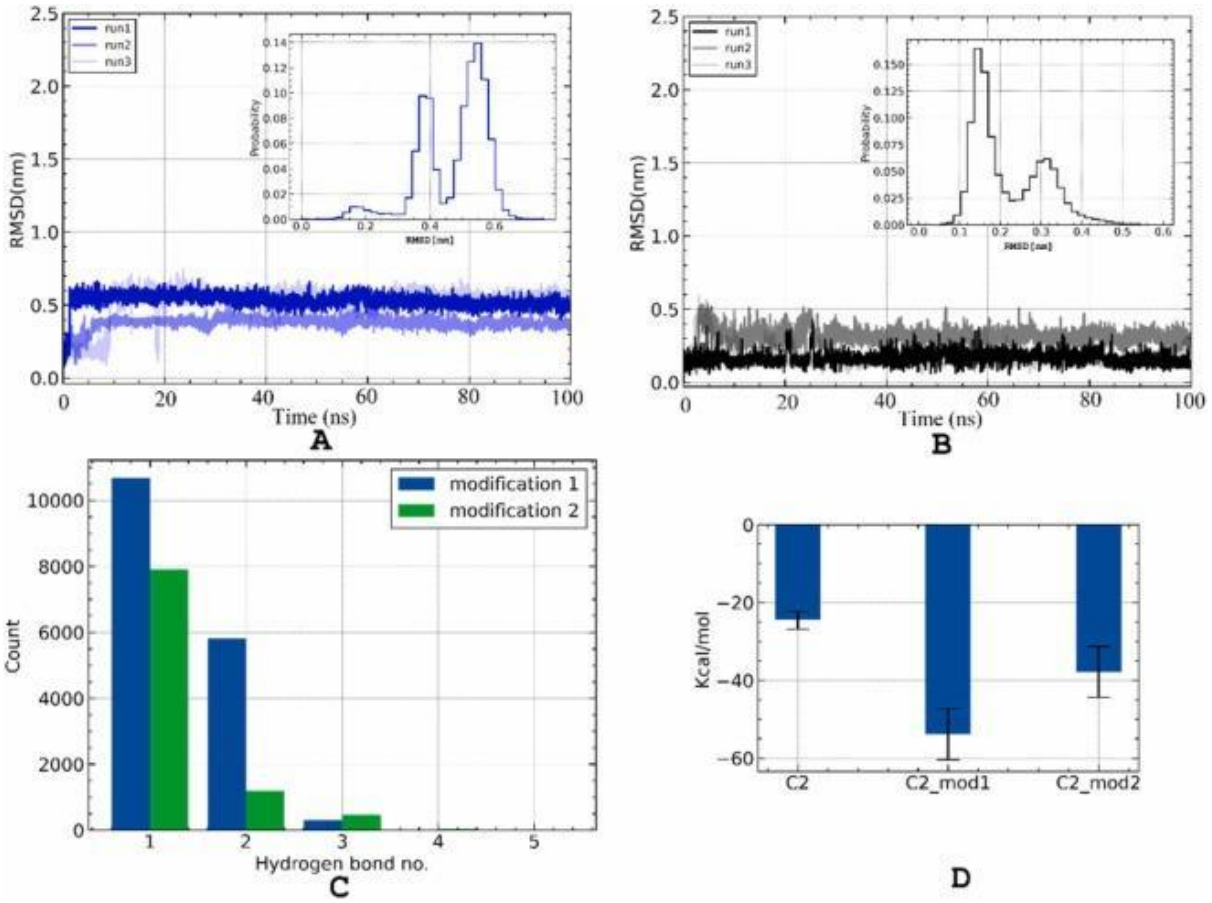


Fig. 8. Post-docking MD simulation analysis of the Nef and modified ligand systems. A) RMSD graph and probability histogram (inset) of compound2_mod-1; B) RMSD graph and probability histogram (inset) of compound2_mod-2; C) Hydrogen bond number counts calculated from concatenated trajectories. D) MMGBSA binding free energy calculation at 300 K (C2 =compound 2, C2_mod1 = compound2_mod-1, C2_mod2 = compound2_mod-2).

3.6. Binding simulations of compound 2_mod-2

Previous high-concentration ColDock simulations of compound 2 showed a spontaneous propensity to bind pocket 3 in a fashion similar to the docking prediction (Fig. 5). Similarly, to assess the binding of compound 2_mod-2 to Nef we performed multiple walkers supervised MD (mwSuMD) simulations. mwSuMD is an adaptive method to simulate the (un)binding of ligands without the introduction of any energy bias.

A total of six replicas were collected and the best one in terms of binding stability was extended for a further 200 ns of classic MD to evaluate the binding characteristic of the formed complex. Along the binding path to Nef, the ligand first formed metastable interactions (< −20 kcal/mol) with pocket 3 boundary residues at about 5–7 Å from the binding site (Fig. 9A, Video S1), then reached Pocket 3 orienting the cyclohexanone ring toward R100 side chain and the benzamide ring towards F94. During the post-mwSuMD classic MD simulation, compound 2_mod-2 flipped along its major axes to align the fluorine atoms in the direction of G100, in accordance with post-docking MD simulations. While the scaffold remained stably bound, the benzamide ring experienced different orientations and interacted with F94, E97, and W117 . Overall, the stabilization gained ranged between − 20 and − 45 kcal/mol (Fig. 9B), with an average of − 31.4 ± 5.8 kcal/mol. The side chains most involved in the binding were F94, W117, L95, and L115 (Fig. 9C).

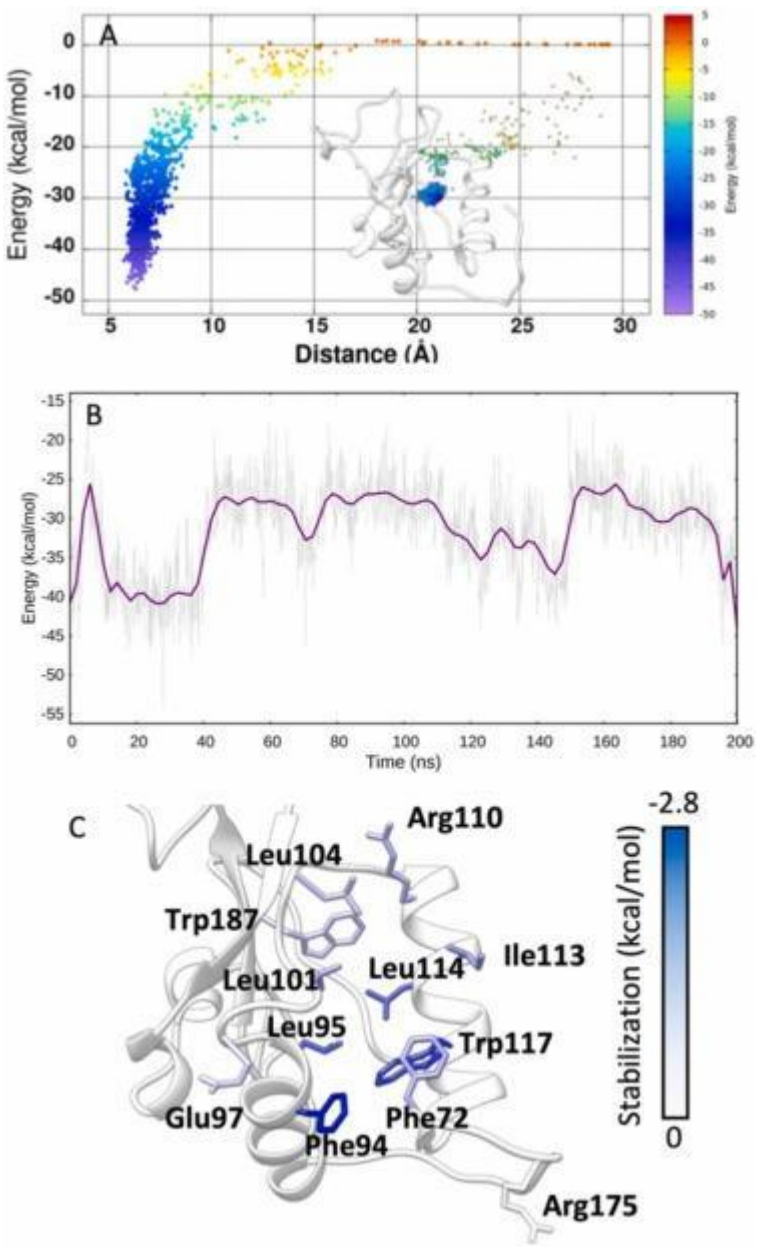


Fig 9. MwSuMD binding simulations pf Compound 2_mod-2. A) Binding energy landscape; B) MMGBSA analysis of the post-mwSuMD classic MD simulation; C) Nef per-residue interaction energy during the post-mwSuMD classic MD simulation.

4. Discussion

Nef is a crucial protein for HIV-1 pathogenesis and several crystal and NMR structures are available in the PDB database. Apart from the NMR structure of truncated Nef and the anchor domain, in all other structures, Nef is in complex with different proteins (Geyer et al., 1999, Grzesiek et al., 1997). Recruitment and activation of Src-family kinases by Nef have been the focus of many structural, cellular, and in vitro studies, and several crystal structures of Nef were solved in complex with Kinase SH3 or SH3-SH2 domain (Staudt et al., 2020). As

Nef homodimers form complexes with kinases, Nef mutants that are unable to form homodimers are also unable to stimulate kinase (e.g. Itk) activity (Poe et al., 2014). For this reason, inhibiting the formation of the Nef homodimer is a potential therapeutic strategy, understudied so far.

We have identified three druggable pockets using a neural network-based method and a probe scanning-based method. MD simulations suggest the highest stability for pocket 3, which is directly involved in homodimerization, and mainly hydrophobic but presents H-bond donor residues at the backbone level of G99, G100, R110, and L114 (PDB ID 4U5W). Indeed, FTMap showed that multiple polar probes (e.g. urea, benzaldehyde, ethanol, N,N-dimethylformamide, phenol, isopropanol, and dimethyl ether) are able to form H-bond with these residues, while the core of the pocket is encompassed by hydrophobic probes only (i.e. cyclohexane, benzene). The hydrophobic nature of pocket 3 drove the *de novo* generation of compounds by AutoGrow4 biased towards lipophilicity. This was because hydrophobic complementarity gave the best VINA docking scores during the compound generation process. Amongst the structures predicted by AutoGrow4, compound 2 was the most stable inside pocket 3 thanks to hydrophobic contacts.

Protein-ligand binding MD simulations are attractive methods for assessing ligand specificity and affinity, but they suffer from the stochastic nature of rare events, requiring computationally expensive simulations to capture spontaneous association events. Adaptive sampling methods have been applied to capture spontaneous ligand binding events (Betz, Dror, 2019). Another way to address this issue is to increase the chances of binding by placing a high concentration of ligands around protein. In this way, compound 2 achieved docking-like poses in two ColDock replicas out of five. This corroborated the specificity of compound 2 towards pocket 3.

Several studies have proved the importance of water molecules in increasing ligand binding affinity and promoting specificity towards proteins (Rudling et al., 2018; Balázs Zoltán, Csaba, 2021). In the case of pocket 3, most of the water molecules are unstable and can be expelled, increasing binding affinity. Pocket 3 is primarily hydrophobic, however, the solvent-exposed backbone of F94, E97, G99, G100, R110, and L114 represents a potential site for H-bonding, as suggested by hydration analysis. Therefore, we introduced the hydroxyl and carboxamide groups to compound 2 (compound2_mod-1). These two modifications increased the predicted water solubility and aided the formation of H-bonds with the backbone of R110 and F94. We further replaced two carbon atoms of the planar scaffold with nitrogen, retaining aromaticity, and added two fluorine atoms at one of the methylene groups of the scaffold (compound2_mod-2) to further increase the predicted

solubility. Molecular docking suggested an H-bond between compound2_mod-2 and the L114 backbone rather than R110.

Since compound2_mod-2 was the best-proposed structure in terms of predicted pharmacokinetics, drug-likeness, and stability inside pocket 3, we further characterized its binding mechanism using mwSuMD simulations, which suggested the flipping of the scaffold compared to the initial docking pose, in accordance with the three PCA clusters extracted from post-docking MD simulations. Average MMGBSA binding energy from the mwSuMD simulation was comparable to post-docking MD simulations, supporting compound2_mod-2 as a Nef binder for pocket 3 and indicating the possibility of using this compound for in vitro testing.

5. Conclusion

This study represents the first rational attempt to design inhibitors against Nef homodimerization. Using a fragment-based *de novo* approach and extensive molecular dynamics simulation we have designed compounds specifically toward the dimerization pocket 3. Computationally, the initial lead compound 2 was very specific towards the homodimerization pocket and the modified compounds also showed the same characteristics. Although these compounds have high *in silico* synthetic accessibility scores and good ADMET properties, further in vitro studies are needed to completely assess their predicted properties, binding specificity, and whether they could inhibit the kinase activity. Although our findings are computational and experimental validation will be needed, we believe that this predictive computational study will pave the way for the rational development of new Nef binders and possibly anti-HIV drugs specifically targeting Nef.

Videos:

https://www.sciencedirect.com/science/article/pii/S1476927123000622?via%3Dihub#ec0005

References

1. Abraham MJ, Murtola T, Schulz R, Szilárd Páll, Smith JC, Hess B, Lindahl E, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX,Volumes 1–2,2015,Pages 19-25,ISSN 2352-7110
2. Ahmad, N., and Venkatesan, S. (1988) Nef protein of HIV-1 is a transcriptional repressor of HIV-1 LTR. Science 241, 1481–1485
3. Akgun, B., Satija, S., Nanda, H., Pirrone, G. F., Shi, X., Engen, J. R., and Kent, M. S. (2013) Conformational transition of membrane-associated terminally acylated HIV-1 Nef. Structure 21, 1822–1833 CrossRef Medline
4. Alvarado JJ, Tarafdar S, Yeh JI, Smithgall TE. Interaction with the Src homology (SH3-SH2) region of the Src-family kinase Hck structures the HIV-1 Nef dimer for kinase activation and effector recruitment. J Biol Chem. 2014 Oct 10;289(41):28539-53. doi: 10.1074/jbc.M114.600031. Epub 2014 Aug 13. PMID: 25122770; PMCID: PMC4192505.
5. Arold, S., Hoh, F., Domergue, S., Birck, C., Delsuc, M. A., Jullien, M., and Dumas, C. (2000) Characterization and molecular basis of the oligomeric structure of HIV-1 Nef protein. Protein Sci. 9, 1137–1148 CrossRef Medline
6. Balázs Zoltán Zsidó, Csaba Hetényi, The role of water in ligand binding, Current Opinion in Structural Biology, Volume 67, 2021, Pages 1-8, ISSN 0959-440X, https://doi.org/10.1016/j.sbi.2020.08.002
7. Betz RM, Dror RO. How Effectively Can Adaptive Sampling Methods Capture Spontaneous Ligand Binding? J Chem Theory Comput. 2019 Mar 12;15(3):2053-2063. doi: 10.1021/acs.jctc.8b00913. Epub 2019 Feb 4. PMID: 30645108; PMCID: PMC6795214
8. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81, 3684 (1984).
9. Betzi, S.; Restouin, A.; Opi, S.; Arold, S.T.; Parrot, I.; Guerlesquin, F.; Morelli, X.; Collette, Y. Protein protein interaction inhibition (2P2I) combining high throughput and virtual screening: Application to the HIV-1 Nef protein. Proc. Natl. Acad. Sci. USA 2007, 104, 19256–19261.
10. Biggs, T. E., Cooke, S. J., Barton, C. H., Harris, M. P., Saksela, K., and Mann, D. A. (1999) Induction of activator protein 1 (AP-1) in macrophages by human immunodeficiency virus type-1 NEF is a cell-type-specific response that requires both hck and MAPK signaling events. J. Mol. Biol. 290, 21–35 CrossRef Medline

11. Briggs, S. D., Sharkey, M., Stevenson, M., and Smithgall, T. E. (1997) SH3- mediated Hck tyrosine kinase activation and fibroblast transformation by the Nef protein of HIV-1. J. Biol. Chem. 272, 17899–17902
12. Cuzzolin, A., Deganutti, G., Salmaso, V., Sturlese, M. & Moro, S. AquaMMapS: An Alternative Tool to Monitor the Role of Water Molecules During Protein-Ligand Association. ChemMedChem 13, 522–531 (2018).
13. Deganutti G, Pipitò L, Rujan R.M., Weizmann T, Griffin P, Ciancetta A, Moro S, Reynolds C.A. Hidden GPCR structural transitions addressed by multiple walker supervised molecular dynamics (mwSuMD). bioRxiv 2022.10.26.513870; doi: https://doi.org/10.1101/2022.10.26.513870
14. Desaphy J, Bret G, Rognan D, Kellenberger E. sc-PDB: a 3D-database of ligandable binding sites--10 years on. Nucleic Acids Res. 2015 Jan;43(Database issue):D399-404. doi: 10.1093/nar/gku928. Epub 2014 Oct 9. PMID: 25300483; PMCID: PMC4384012.
15. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res. 32, W665-7 (2004).
16. Emert-Sedlak, L., Kodama, T., Lerner, E. C., Dai, W., Foster, C., Day, B. W., Lazo, J. S., and Smithgall, T. E. (2009) Chemical library screens targeting an HIV-1 accessory factor/host cell kinase complex identify novel antiretroviral compounds. ACS Chem. Biol. 4, 939–947 CrossRef Medline
17. Emert-Sedlak, L. A., Narute, P., Shu, S. T., Poe, J. A., Shi, H., Yanamala, N., Alvarado, J. J., Lazo, J. S., Yeh, J. I., Johnston, P. A., and Smithgall, T. E. (2013) Effector Kinase Coupling Enables High- Throughput Screens for Direct HIV-1 Nef Antagonists with Antiretroviral Activity. Chem. Biol. 20 (1), 82−91.
18. Emert-Sedlak LA, Shi H, Tice CM, Chen L, Alvarado JJ, Shu ST, Du S, Thomas CE, Wrobel JE, Reitz AB, Smithgall TE. Antiretroviral Drug Discovery Targeting the HIV-1 Nef Virulence Factor. Viruses. 2022 Sep 13;14(9):2025. doi: 10.3390/v14092025. PMID: 36146831; PMCID: PMC9503669.
19. Essmann, U. et al. A smooth particle mesh Ewald method. J. Chem. Phys. 103, 8577 (1995).
20. Forester, T. R. & Smith, W. SHAKE, rattle, and roll: Efficient constraint algorithms for linked rigid bodies. J. Comput. Chem. (1998).
21. Gerlach, H., Laumann, V., Martens, S., Becker, C. F., Goody, R. S., and Geyer, M. (2010) HIV-1 Nef membrane association depends on charge, curvature, composition and sequence. Nat. Chem. Biol. 6, 46–53 CrossRefMedline
22. Geyer M, Munte CE, Schorr J, Kellner R, Kalbitzer HR. Structure of the

anchor-domain of myristoylated and non-myristoylated HIV-1 Nef protein. J Mol Biol. 1999 May 28;289(1):123-38. doi: 10.1006/jmbi.1999.2740. PMID: 10339411.

23. Grzesiek S, Bax A, Hu JS, Kaufman J, Palmer I, Stahl SJ, Tjandra N, Wingfield PT. Refined solution structure and backbone dynamics of HIV-1 Nef. Protein Sci. 1997 Jun;6(6):1248-63. doi: 10.1002/pro.5560060613. PMID: 9194185; PMCID: PMC2143732.
24. Gupta DN, Dalal V, Savita BK, Alam MS, Singh A, Gubyad M, Ghosh DK, Kumar P, Sharma AK. Biochemical characterization and structure-based in silico screening of potent inhibitor molecules against the 1 cys peroxiredoxin of bacterioferritin comigratory protein family from Candidatus Liberibacter asiaticus. J Biomol Struct Dyn. 2022 Jul 9:1-13. doi: 10.1080/07391102.2022.2096118. Epub ahead of print. PMID: 35815556.
25. Hanna, Z., Kay, D. G., Rebai, N., Guimond, A., Jothy, S., and Jolicoeur, P. (1998) Nef harbors a major determinant of pathogenicity for an AIDS-like disease induced by HIV-1 in transgenic mice. Cell 95, 163–175
26. Harvey, M. J., Giupponi, G. & Fabritiis, G. D. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. J. Chem. Theory Comput. 5, 1632–1639 (2009).
27. Herna, R. G., and Saksela, K. (2000) Interactions of HIV-1 NEF with cellular signal transducing proteins. Front. Biosci. 5, D268–D283
28. Huang, J. & MacKerell, A. D. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. J. Comput. Chem. 34, 2135–2145 (2013).
29. Huang, J. et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat. Methods 14, 71–73 (2017).
30. Jäger, S., Cimermancic, P., Gulbahce, N., Johnson, J. R., McGovern, K. E., Clarke, S. C., Shales, M., Mercenne, G., Pache, L., Li, K., Hernandez, H., Jang, G. M., Roth, S. L., Akiva, E., Marlett, J., et al. (2011) Global landscape of HIV-human protein complexes. Nature 481, 365–370 CrossRef Medline
31. J Jiménez, S Doerr, G Martínez-Rosell, A S Rose, G De Fabritiis, DeepSite: protein-binding site predictor using 3D-convolutional neural networks, Bioinformatics, Volume 33, Issue 19, 01 October 2017, Pages 3036–3042
32. Jolicoeur, P. (2011) The CD4C/HIV(Nef)transgenic model of AIDS. Curr. HIV. Res. 9, 524–530
33. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79, 926 (1983).

34. Kent, M. S., Murton, J. K., Sasaki, D. Y., Satija, S., Akgun, B., Nanda, H., Curtis, J. E., Majewski, J., Morgan, C. R., and Engen, J. R. (2010) Neutron reflectometry study of the conformation of HIV Nef bound to lipid membranes. Biophys. J. 99, 1940–1948 CrossRef Medline
35. Kirchhoff, F., Greenough, T. C., Brettler, D. B., Sullivan, J. L., and Desrosiers, R. C. (1995) Brief report: absence of intact nef sequences in a longterm survivor with nonprogressive HIV-1 infection. N. Engl. J. Med. 332, 228–232 CrossRef Medline
36. Komuro, I., Yokota, Y., Yasuda, S., Iwamoto, A., and Kagawa, K. S. (2003) CSF-induced and HIV-1-mediated distinct regulation of Hck and C/ EBPb represent a heterogeneous susceptibility of monocyte-derived macrophages to M-tropic HIV-1 infection. J. Exp. Med. 198, 443–453 CrossRef Medline
37. Kozakov, D., Grove, L., Hall, D. et al. The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. Nat Protoc 10, 733–755 (2015). https://doi.org/10.1038/nprot.2015.043
38. Krautler, V., van Gunsteren, W. F. & Hanenberger, P. H. A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. J. Comput. Chem. 22, 501–508 (2001).
39. Lee J, Cheng X, Swails JM, Yeom MS, Eastman PK, Lemkul JA, Wei S, Buckner J, Jeong JC, Qi Y, Jo S, Pande VS, Case DA, Brooks CL 3rd, MacKerell AD Jr, Klauda JB, Im W. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. J Chem Theory Comput. 2016 Jan 12; 12(1):405-13. doi: 10.1021/acs.jctc.5b00935. Epub 2015 Dec 3. PMID: 26631602; PMCID: PMC4712441.
40. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins. 2010 Jun;78(8):1950-8. doi: 10.1002/prot.22711. PMID: 20408171; PMCID: PMC2970904.
41. Loncharich, R. J., Brooks, B. R. & Pastor, R. W. Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. Biopolymers 32, 523–535 (1992).
42. Maier, J. A. et al. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J. Chem. Theory Comput. 11, 3696–3713 (2015).
43. Majumder S, Giri K. An insight into the binding mechanism of Viprinin and its morpholine and piperidine derivatives with HIV-1 Vpr: molecular dynamics simulation, principal component analysis and binding free energy calculation study. J Biomol Struct Dyn. 2021 Jul 23:1-13. doi: 10.1080/07391102.2021.1954553. Epub ahead of

print. PMID: 34296659.

44. McGibbon, R. T. et al. Mdtraj: A modern open library for the analysis of molecular dynamics trajectories. Biophys. J. 109, 1528–1532 (2015).
45. 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).
46. Miller, B. R. et al. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. J. Chem. Theory Comput. 8, 3314–3321 (2012).
47. Moarefi, I., LaFevre-Bernt, M., Sicheri, F., Huse, M., Lee, C. H., Kuriyan, J., and Miller, W. T. (1997) Activation of the Src-family tyrosine kinase Hck by SH3 domain displacement. Nature 385, 650–653
48. Moonsamy S, Bhakat S, Ramesh M, Soliman ME. Identification of Binding Mode and Prospective Structural Features of Novel Nef Protein Inhibitors as Potential Anti-HIV Drugs. Cell Biochem Biophys. 2017 Mar;75(1):49-64. doi: 10.1007/s12013-016-0774-1. Epub 2016 Dec 16. PMID: 27981421.
49. Mujib, S., Saiyed, A., Fadel, S., Bozorgzad, A., Aidarus, N., Yue, F. Y., Benko, E., Kovacs, C., Emert-Sedlak, L., Smithgall, T. E., and Ostrowski, M. A. (2017) Pharmacologic HIV-1 Nef Blockade Enhances the Recognition and Elimination of Latently HIV-1 Infected CD4 T cells by Autologous CD8 T cells. J. Clin. Invest. Insight 2 (17), No. e93684.
50. Narute, P. S., and Smithgall, T. E. (2012) Nef alleles from all major HIV-1 clades activate Src-family kinases and enhance HIV-1 replication in an inhibitor-sensitive manner. PLoS One 7, e32561 CrossRef Medline
51. O'Boyle, N.M., Banck, M., James, C.A. et al. Open Babel: An open chemical toolbox. J Cheminform 3, 33 (2011). https://doi.org/10.1186/1758-2946-3-33
52. Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK Predictions. J. Chem. Theory Comput. 7, 525–537 (2011).
53. Poe, J. A., Vollmer, L., Vogt, A., and Smithgall, T. E. (2014) Development and Validation of a High-Content Bimolecular Fluorescence Complementation Assay for Small-Molecule Inhibitors of HIV-1 Nef Dimerization. J. Biomol. Screening 19, 556−565.
54. Rhodes, D. I., Ashton, L., Solomon, A., Carr, A., Cooper, D., Kaldor, J., and Deacon,N. (2000) Characterization of three nef-defective human immunodeficiency virus type 1 strains associated with long-term nonprogression. Australian Long-Term Nonprogressor Study Group. J. Virol. 74, 10581–10588 CrossRef Medline
55. Rudling A, Orro A, Carlsson J. Prediction of Ordered Water Molecules in Protein

Binding Sites from Molecular Dynamics Simulations: The Impact of Ligand Binding on Hydration Networks. J Chem Inf Model. 2018 Feb 26;58(2):350-361. doi: 10.1021/acs.jcim.7b00520. Epub 2018 Feb 5. PMID: 29308882; PMCID: PMC6716772.

56. Saksela, K. (2011) Interactions of the HIV/SIV pathogenicity factor Nef with SH3 domain-containing host cell proteins. Curr. HIV. Res. 9,531–542
57. Scherer MK, Trendelkamp-Schroer B, Paul F, Pérez-Hernández G, Hoffmann M, Plattner N, Wehmeyer C, Prinz JH, Noé F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. J Chem Theory Comput. 2015 Nov 10;11(11):5525-42. doi: 10.1021/acs.jctc.5b00743. Epub 2015 Oct 14. PMID: 26574340.
58. Shi H, Tice CM, Emert-Sedlak L, Chen L, Li WF, Carlsen M, Wrobel JE, Reitz AB, Smithgall TE. Tight-Binding Hydroxypyrazole HIV-1 Nef Inhibitors Suppress Viral Replication in Donor Mononuclear Cells and Reverse Nef-Mediated MHC-I Downregulation. ACS Infect Dis. 2020 Feb 14;6(2):302-312. doi: 10.1021/acsinfecdis.9b00382. Epub 2019 Dec 16. PMID: 31775511; PMCID: PMC8142392.
59. Sousa da Silva, A.W., Vranken, W.F. ACPYPE - AnteChamber PYthon Parser interfacE. BMC Res Notes 5, 367 (2012). https://doi.org/10.1186/1756-0500-5-367
60. Spiegel, J.O., Durrant, J.D. AutoGrow4: an open-source genetic algorithm for de novo drug design and lead optimization. J Cheminform 12, 25 (2020). https://doi.org/10.1186/s13321-020-00429-4
61. Staudt RP, Alvarado JJ, Emert-Sedlak LA, Shi H, Shu ST, Wales TE, Engen JR, Smithgall TE. Structure, function, and inhibitor targeting of HIV-1 Nef-effector kinase complexes. J Biol Chem. 2020 Oct 30;295(44):15158-15171. doi: 10.1074/jbc.REV120.012317. Epub 2020 Aug 29. PMID: 32862141; PMCID: PMC7606690.
62. Sterling T, Irwin JJ. ZINC 15--Ligand Discovery for Everyone. J Chem Inf Model. 2015 Nov 23;55(11):2324-37. doi: 10.1021/acs.jcim.5b00559. Epub 2015 Nov 9. PMID: 26479676; PMCID: PMC4658288.
63. Strebel K. HIV accessory proteins versus host restriction factors. Curr Opin Virol. 2013 Dec;3(6):692-9. doi: 10.1016/j.coviro.2013.08.004. Epub 2013 Nov 15. PMID: 24246762; PMCID: PMC3855913.
64. Takemura K, Sato C, Kitao A. ColDock: Concentrated Ligand Docking with All-Atom Molecular Dynamics Simulation. J Phys Chem B. 2018 Jul 26;122(29):7191-7200. doi: 10.1021/acs.jpcb.8b02756. Epub 2018 Jul 11. PMID: 29993242.
65. Trott O., Olson A.J., AutoDock Vina: improving the speed and accuracy of docking

with a new scoring function, efficient optimization and multithreading, Journal of Computational Chemistry 31 (2010) 455-461

66. Usami, Y., Wu, Y., and Göttlinger, H. G. (2015) SERINC3 and SERINC5 restrict HIV-1 infectivity and are counteracted by Nef. Nature 526, 218–223 CrossRef Medline
67. Valdés-Tresanco, M.S., Valdés-Tresanco, M.E., Valiente, P.A. and Moreno E. gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS. Journal of Chemical Theory and Computation, 2021 17 (10), 6281-6291
68. Wales, T. E., Hochrein, J. M., Morgan, C. R., Emert-Sedlak, L. A., Smithgall, T. E., and Engen, J. R. (2015) Subtle dynamic changes accompany Hck activation by HIV-1 Nef and are reversed by an antiretroviral kinase inhibitor. Biochemistry 54, 6382–6391 CrossRef Medline
69. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. J. Comput. Chem. 25, 1157–1174 (2004).
70. Wu M, Alvarado JJ, Augelli-Szafran CE, Ptak RG, Smithgall TE. A single β-octyl glucoside molecule induces HIV-1 Nef dimer formation in the absence of partner protein binding. PLoS One. 2018 Feb 7;13(2):e0192512. doi: 10.1371/journal.pone.0192512. Erratum in: PLoS One. 2018 Apr 17;13(4):e0196229. PMID: 29415006; PMCID: PMC5802939.
71. Ye, H. H., Choi, H. J., Poe, J., and Smithgall, T. E. (2004) Oligomerization is required for HIV-1 Nef-induced activation of the Src family protein-tyrosine kinase, Hck. Biochemistry 43, 15775–15784 CrossRef Medline

8 Extra-project Publications and Preprints

While most of the PhD studies were dedicated to SARS-CoV-2 and HIV, I also worked on two further projects: a new approach to elucidate and inhibit the mechanism of the human pyruvate kinase 2 (hPKM2) enzyme expressed in many cancer phenotypes and the enhancement of the supervised molecular dynamics (SuMD) protocol.

Anaerobic glycolysis is a key feature in cancer cells, sustaining their unregulated proliferation through hPKM2-mediated ATP production, particularly in hypoxic tumour microenvironments. Targeting overexpressed hPKM2 could disrupt glycolytic ATP production, potentially reducing cancer cell replication and enhancing chemotherapy effectiveness. I characterised the molecular recognition mechanism for hPKM2 and phosphoenol pyruvate (PEP), providing insights into its binding mode while screening and designing PEP isosteres, with optimized interactions and an overall safer toxicological profile. It is the first publication that describes the hPKM2-PEP binding mode through MD and highlights the importance of the presence of the anionic centres for a bioisosteric drug design.

8.1 Targeting hPKM2 in cancer: a bio isosteric approach for ligand design

The metabolic reprogramming of cancer cells under a hypoxic tumour microenvironment was first observed by Otto Warburg in 1920 who reported an abnormal glucose uptake. This metabolic reprogramming is shared amongst many cancer phenotypes as well as the overexpression of the human pyruvate kinase 2 (hPKM2), a rate-limiting kinase enzyme responsible for the phosphorylation of adenosine diphosphate (ADP) to adenosine triphosphate (ATP) in cancer, necessary for cancer pro-survival mechanisms. In addition to the phosphate transfer, hPKM2 plays a crucial role as a transcriptor factor which leads to the overexpression of glucose transporters, multidrug resistance proteins, as well as a wide array of oncogenes that promote and sustain cancer development. This work aimed to describe the molecular recognition mechanism between hPKM2 and phosphoenol pyruvate (PEP) and propose a set of viable bioisosteric replacements based on previous experimental approaches. This work included extensive parameterization work, as well as a structure-based optimization strategy with different generative approaches. The results of such exertion were published with the title: “Targeting hPKM2 in cancer: a bio isosteric approach for ligand design” published in Computers in Biology and Medicine, Volume 158; https://doi.org/10.1016/j.compbiomed.2023.106852

Authors: Ludovico Pipitò, Thomas Arron Illingworth, Giuseppe Deganutti

Abstract:

The term cancer refers to a plethora of diseases characterized by the development of abnormal cells that divide uncontrollably and can infiltrate further proximal or distal body tissues. Each type of cancer can be defined by aggressiveness, localization, metabolism, and response to available treatments. Among the most common hallmarks of cancer is a more acidic intracellular microenvironment. Offset pH values are due to an excess of lactate and an increased hypoxia-inducible factor (HIF) expression, which leads to a hypoxic state and a metabolic shift towards glycolysis to produce adenosine-5'-triphosphate (ATP) necessary for cellular metabolism. Warburg's hypothesis underpins this concept, making glycolysis and its central enzyme pyruvate kinase (hPKM2), an ideal target for drug development. Using molecular docking and extensive molecular dynamics (MD) simulations we investigated the binding mode of phosphoenolpyruvate (PEP) inside the hPKM2 active site, and then evaluated a set of known bio-isosteric inhibitors to understand the differences caused by their substitutions on their binding mode. Ultimately, we propose a new molecular entity to hamper hPKM2, unbalance cellular energy, and possibly trigger autophagic mechanisms.

Introduction:

Cancer is characterised by an uncontrolled proliferative capacity (1), driven by genetic and epigenetic alterations of cell functions like metabolism (2), and proliferative and survival pathways. The result is cells that are capable of growing exponentially, impairing tissue and organ function in the process (3). Cancer cells have higher metabolic outputs compared to healthy cells, which is achieved through over-expression and hyperactivity of various enzymes (2) involved in key metabolic pathways, including glucose uptake and glycolysis (4) to promote a pro–survival environment for cancer. The "Warburg hypothesis"(5) describes the glycolytic metabolic shifts in both aerobic and anaerobic conditions observed within solid-state tumours (6,7), gained through overexpression of glucose transporters, increased ATP production, and lactate accumulation in the tumour microenvironment (TME). Metabolic remodelling of the tumour favours disease progression and impedes treatment opportunities, due in large part to chemotherapy resistance, shifting preference for hypoxic states through lactate-dependent mechanisms to produce ATP in absence of oxygen (8). The result of this is a survival advantage with respect to energy production, leading to tumour progression, angiogenesis, and metastasis (9). Oxygen deficiency also promotes hypoxia-inducible factors (HIF), which in turn over expresses glucose transporter (GLUT1) to facilitate glucose uptake for tumour progression (10). HIF-1α triggers several tumour-promoting pathways. HIF1α alters lactate dehydrogenase A isoform expression (LDH-A), impacting cellular pH, ion

balance, as well as hypoxia response elements linked to cancer progression (11). Additionally, HIF-1 is involved in the transcription of glycolytic enzymes ultimately linked with metabolic regulators such as the mammalian target of rapamycin (mTOR), through the AMP-activated protein kinase pathway (AMPK 5). This inhibits apoptotic response elements in response to increased concentrations of ATP (10). Cancers exhibiting a HIF-glycolytic state act within the extracellular space, where interleukin 1 beta (IL-1β) and tumour necrosis factor-alpha (TNF-α) promote an increased expression of pro-inflammatory factors and angiogenesis via HIF-mediated vascular endothelial growth factor (VEGF) overexpression with increased M1 macrophage recruitment (12–14). It might also restore physiological lactate levels, reducing expression levels of phosphofructokinase 1 (PFK1), histone acetylation, inactivation of p53, and the overexpression of the multi-drug resistance (MDR) factors and genes (15), responsible for drug resistance (16). In conditions of reduced oxidative phosphorylation and fatty acid oxidation for ATP production, the prominent resource left to sustain the hyperactive cell metabolism is glycolysis, making glycolysis an ideal target for therapeutic approaches (17). All the interconnected pathways mentioned and their influence on oncogenic phenotype are summarised in Figure 1.

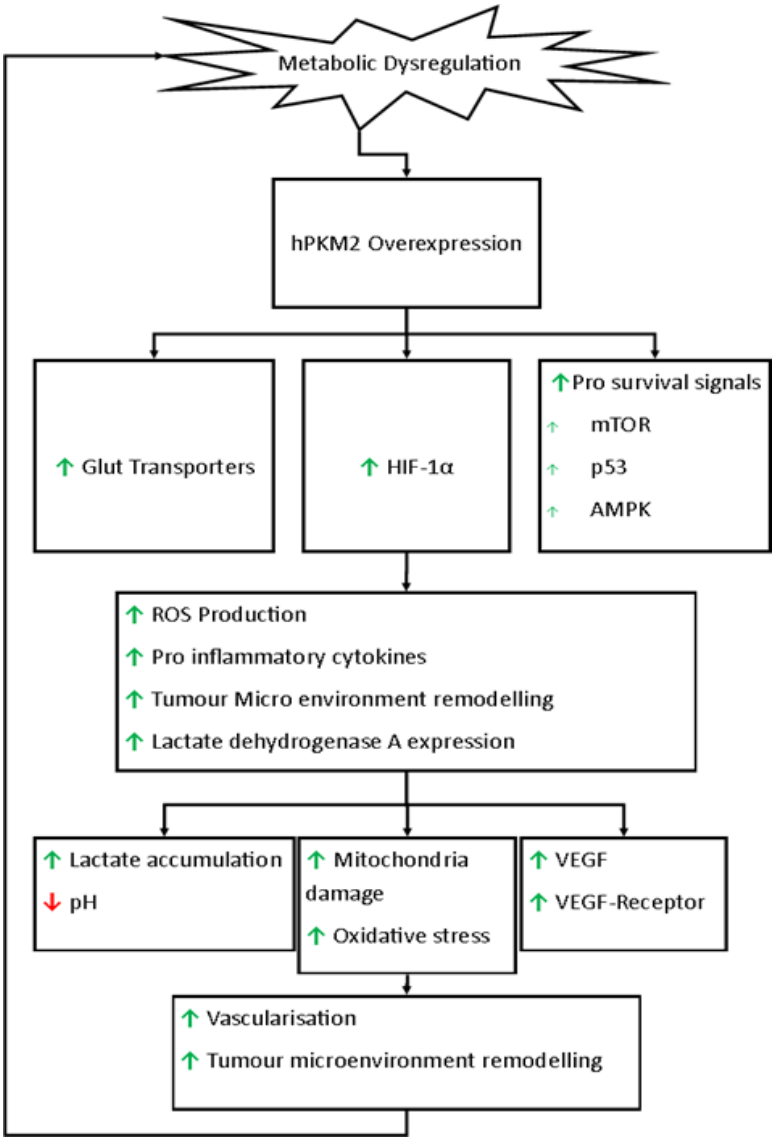


Figure 1: The central role of hPKM2 in conjunction with metabolic adjustments occurring in oncotic phenotypes. The positive feedback contributes to establishing new pathological homeostasis with distinctive hallmarks which act in TME. The endpoint of all the transformations leads to tumour development and increased therapy resistance, which is linked to poor prognosis.

Pyruvate kinase (hPK) is a rate-limiting glycolysis enzyme, responsible for the irreversible transphosphorylation of a phosphate group from phosphoenolpyruvate (PEP) to adenosine diphosphate (ADP) to produce ATP (18). There are four human isoforms of PK, encoded by three tissue-specific genes: the L isoform is expressed in the liver, R in erythrocytes, and M in muscles as a result of alternative splicing of exon 9 and 10 (M1 and M2) which differ by 22 positions, located near the allosteric regulation site for 1,6 fructose bisphosphate (FBP) and specific deletions in tumour cells (19). Human pyruvate kinase M2 (hPKM2) increased level of expression is found in a variety of cancers such as gastric, colorectal, and bile duct,

rheumatic diseases, lung cancer, breast cancer, neuroendocrine tumours, urological malignancies, renal cell carcinoma, diabetic nephropathy, haematological malignancies, prostate cancer, and thyroid carcinoma (20). hPKM2 is expressed in tissues with anabolic functions, including proliferating cells and cancer (21), suggesting that it could be a suitable target in cancer therapy. The development of safe and effective treatments targeting the glycolytic pathway provides an opportunity to interrupt the dysregulated homeostasis observed in cancer. Numerous studies related hPKM2 inhibition with a positive effect on cancer progression (22–26), however the specific mechanism of action is still not fully understood. A promising clinical trial for the hPKM2 inhibitor TLN-232/CAP-232 was conducted by Thallion Pharmaceuticals in 2007 but was terminated due to legal reasons (10). Shikonin and its analogues have been shown to effectively inhibit hPKM2, demonstrating selectivity for hPKM2 (25–28), impairing the glycolytic pathway for ATP production.

Using hPKM2 crystallographic information as a starting point, we employed extensive molecular docking and molecular dynamics (MD) simulations to highlight, for the first time, the binding mode of PEP inside the hPKM2 catalytic site. Additionally, we describe the binding mode for a set of experimentally tested PEP bio-isosteres compounds (29,30), then propose and test *in silico* a modification of the enol moiety of PEP as a bio-isosteric substitution to improve the inhibitors' stability to possibly reduce chemotherapy resistance (27). We evaluate the binding energy of the compounds with extensive molecular mechanics energy combined Poisson-Boltzmann energy surface area (MM-PBSA). This work reinvigorates the research of small molecules able to inhibit hPKM2, with the aim of depleting cancerous cells of their main ATP source.

Methods

Protein Preparation and Force Field Settings

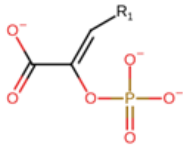
All systems were prepared using the CHARMM36/CGenFF 3.0.1 (31,32) force field combination. hPKM2 structure tetramer was retrieved from PDB 1T5A (20) and modelled, restricting our system to a single chain. The Na^{+} and Mg^{2+} ions were kept, while FBP was removed to study the binding capacity of the enzyme in its unbound pre-activation state, where the Mg^{2+}-ADP complex is not present in the binding site. FBP was not considered, as we aimed to study PEP inside the catalytic site in the pre-catalytic state. The protonation state of residues was calculated by Propka (33) at a simulated pH of 6.5 to match the tumour intracellular pH (34,35), and edited by pdb2pqr (36), while disulfide bonds were identified by HTMD (37), visually inspected, and patched manually through VMD (38). The protein

potential energy was minimised through 1000 steps of the conjugate gradient algorithm using ACEMD (39).

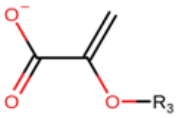
Ligand Preparation and Parameter Settings

PEP structure (PubChem: 348274362) was built as a template for building all the bio isosteric molecules considered (Figure 2A, Table 1) using the VMD Molefacture plugin (https://www.ks.uiuc.edu/Research/vmd/plugins/molefacture/) while functional group substitutions or ligand modifications were performed with Chimera (40). ADMET prediction was computed using ADMETlab 2.0 (41) and included the toxicological profiles . Ligand descriptors were calculated with RDkit (42,43). For each ligand, the initial docking poses and scores were obtained using Autodock Vina (44,45) by centring an isometric grid box on residue K270 with a side of 25 Å. To validate the docking protocol we performed the self-docking of the only inhibitor with structural information available, oxalacetate (OXL) (PDB entry 1T5A). The OXL self-docking displayed an RMSD of 0.93 Å in agreement with the crystallography resolution. Results that displayed conformations similar to the X-ray crystallography in terms of the orientation of the phosphate moiety toward the PO_4^- group as resolved in 1T5A were used as starting coordinates for MD simulations. For each ligand, the initial parameters were obtained through the CGenFF server (46,47). Atomic charges were calculated with the restrained electrostatic potential (RESP) using the Antechamber package (48), while angles and dihedrals with high penalty scores were optimised using Gaussian 09 (49) using the MP2/6-31g(d) level of theory. All ligands were simulated individually in water in short MDs for a visual evaluation of the reliability of the parameters.

Table 1. List of hPKM2 inhibitors



Phosphoenolpyruvate = R_1 H (PEP)
Compound 2 = R_1 (Z)-F (FEP)
Compound 3 = R_1 (Z)-Cl (CEP)



Compound 4 = R_3 SO_3 (SEP)

Name	Molecular Weight (Da)	pK_d or pIC_{50}
PEP	164.96	5.6 [1]
Compound 2	182.92	7 [2]
Compound 3	198.92	7.3 [2]
Compound 4	165.95	2.3 [2]

The table above shows the different functional group addition and modifications as experimentally performed by Garcìa-Alles *et al* (29)

1: Experimental pK_d reported by Duffy *et al* (50).

2: pIC_{50} as reported by Garcìa-Alles *et al* (29).

System Preparation for Molecular Dynamics (MD)

hPKM2 structure and topology files were prepared using VMD's Psfgen plugin (https://www.ks.uiuc.edu/Research/vmd/plugins/psfgen/), and the resulting structures were visually inspected. The systems were simulated for a total time of 1000 ns with TIP3P water molecules (51) added to the simulation box using the Solvate plugin 1.5 (http://www.ks.uiuc.edu/Research/vmd/plugins/solvate/) to give a 10 Å padding in every direction. The charge neutrality was achieved by adding Na^+/Cl^- to the concentration of 0.150 M using the Autoionize plugin 1.3 (http: //www.ks.uiuc. edu/Research/vmd /plugins/autoionize/). ACEMD was used for both the equilibration and the productive MD trajectories. The energy of the systems was reduced through 1000 conjugate-gradient minimization steps to eliminate possible clashes and optimize atomic distances. Equilibration was reached in isothermal-isobaric conditions (NPT) using the Berendsen barostat (52) (target pressure 1 atm) and the Langevin thermostat (53) (target temperature 310 K) during a 4 ns long MD simulation (integration time step 2 fs). During the equilibration, a positional restraint of 1 kcal/ mol $Å^2$ was applied on the alpha carbons of hPKM2 for the first 3 ns, while ligands' restraints were kept through the whole equilibration to avoid unwanted displacements. Positional restraints of 1 kcal/mol $Å^2$ were also applied on protein side chains for the first 2 ns. Productive trajectories were produced with an integration time step of 4 fs, using hydrogen mass repartition (54) in the canonical ensemble (NVT), with no positional restraints. The cut-off distance for electrostatic interactions was set at 9 Å, with a switching function applied beyond 7.5 Å. Long-range Coulomb interactions were handled using the particle mesh Ewald summation method (PME) (55) with default ACEMD settings.

MD trajectories analysis

The root mean square deviation (RMSD) and root mean square fluctuation (RMSF) analyses were computed using VMD and MDTraj (56). Ligand-protein contacts, including hydrogen bonds, were detected using the GetContacts scripts tool (https://getcontacts.github.io), with a threshold distance and angle of 3.5 Å and 120°, respectively. Contacts and HB were expressed as occupancy (% of total MD frames). The Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) was computed with the MMPBSA.py (57–59) script (AmberTools20 suite at http://ambermd.org/), converting the CHARMM psf topology files to Amber prmtop format through ParmEd (http://parmed.github.io/ParmEd/html/index.html).

Results

Assessing the dynamics of apo hPKM2

The X-ray diffraction structure of hPKM2 (PDB: 1T5A) reports the tetrameric hPKM2 (Figure 2A) in complex with the inhibitor oxalate (OXL) and the positive allosteric modulator FBP. hPKM2 forms a tetrameric structure through the C and N domains, while the catalytic site is situated between the A and B domains (Figure 2B). In the catalytic site of each monomer, OXL engages T328 and K270 side chains in hydrogen bonds and forms bidentate coordination with an Mg^{2+} ion, which in turn coordinates with D296, E272, and a water molecule (Figure 2C). A potassium ion occupies the inner part of the pocket, coordinating with the former water molecule and S243, D113, N75, and T114 (backbone). Situated in a more external site between the A and B domains (20), a phosphate group interacts with R120 and H78, in correspondence with the site putatively occupied by ADP (Figure 2C).

During a 1μs MD simulation, the extremities of domains N and B exhibited the highest flexibility (Figure 2D), in line with previous work (60). We assessed the dynamics of the apo hPKM2 (obtained by removing OXL and FBP) by measuring the RMSD to 1T5A (Figure 2E), which resulted in 2.8 ± 0.9 Å, in agreement with the nominal resolution (2.8 Å) of 1T5A. The flexibility displayed by the FBP pocket indicated that this site is stabilised by the allosteric modulator. The catalytic site residues S243, T114, and D113 oriented their sidechains toward K270 and remained stable throughout the simulation (Video S1). Intriguingly, the Mg^{2+} ion remained interlocked between D296 and E272, in line with its crucial role in coordinating the substrate binding as well as the catalytic mechanism (20). The catalytic residues are encompassed between domain A and the mobile domain B which has been reported to close in the presence of Mg^{2+}, ADP (18), and K^{+}. We did not observe any spontaneous closing of the B domain onto the A domain, due to the absence of the Mg^{2+}-ADP complex connecting the two domains (61). However, the residues R73, K270, S240, D113, T114, E272, and D296, involved in the enzymatic reaction, kept their original orientation during the simulation, indicating minimal conformational changes.

Taken together, these results indicate that, the stability of the catalytic site is secured by the tight inter-residue network of contacts and hydrogen bonds and suggest that the catalytic site could be a template for the structure-based drug design of novel inhibitors, in the absence of the FBP from the allosteric site. Although the potassium ion may be present in the ground state hPKM2 (61), it was soon displaced in the absence of the stabilizing substrate, while the Mg^{2+} was completely and independently stabilized by the D296 and E272 active site residues, suggesting that Mg^{2+} should be considered for drug design purposes.

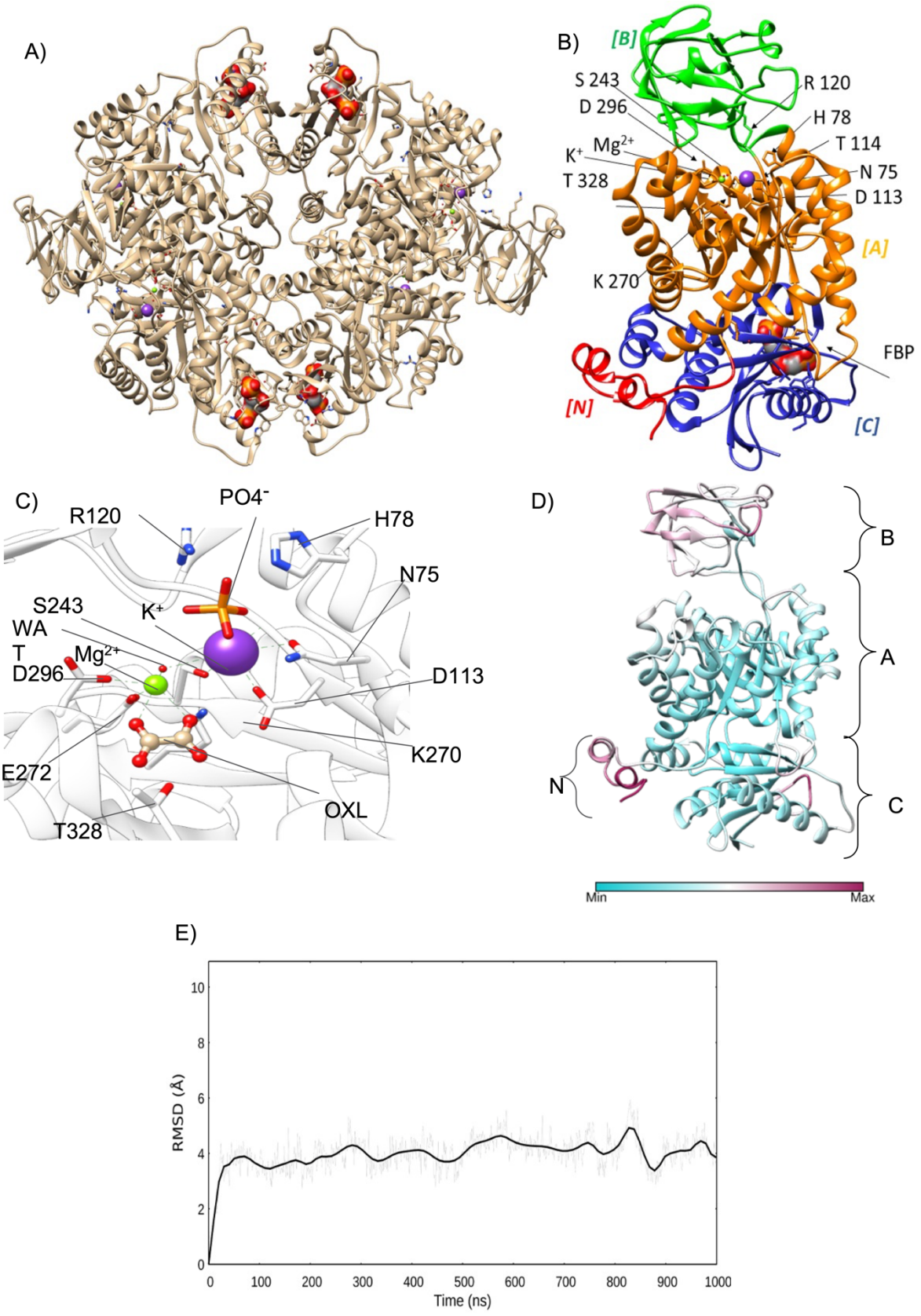


Figure 2. Crystal structure of hPKM2 and its active site domains. A) The X-ray diffraction structure of tetrameric hPKM2 (PDB 1T5A, tan ribbon) with FBP is represented as a molecular surface in the allosteric pocket. Ions and phosphate are represented as a ball

and sticks; **B**) closeup ribbon representation of hPKM2 monomer with domain classification is shown (62): the N-terminal domain (residues I13-A42 in red), the A domain hosting the binding site where ions are present (residues R43-K115, K224-A387), the B domain (G116-E223), and C domain (A388 -P531). **C)** Single chain closeup ribbon structure of hPKM2 monomer in complex with OXL. The OXL position is stabilized by the Mg^{2+} ion which is suggested to support the phosphoryl transition. A phosphate group is shown as a ball and stick to indicate where the Mg^{2+} - ADP should be **D)** RMSF values plotted on the ribbon representation of hPKM2 with domains classification: N-terminal domain (residues I13-A42), the A domain hosting the binding (residues R43-K115, K224-A387), the B domain (G116-E223), and C domain (A388-P531). **E)** RMSDs of hPKM2 alpha carbon atoms over time, which averaged 3.9 ± 0.63 Å, indicating that the monomer remained stable.

Phosphoenolpyruvate binding mode

Although the catalytic mechanism of PKM2 is well known, to the best of our knowledge no information is available about PEP interactions within the active site, as there are no reported structural or computational studies addressing the binding mode of PEP before its phosphate group is transferred to ADP (pre-catalytic state). To tackle this knowledge gap, we docked PEP to hPKM2 and performed post-docking MD simulation. PEP was overall stable throughout the MD trajectory (RMSD = 2.31 ± 0.47 Å, Figure S2A, RMSF = 2.09 ± 0.63 Å Figure S2B) although it completely stabilised in the second half of the simulations thanks to the coordination of Na+ and Mg^{2+} ions by the carboxylic and phosphate groups, with the aid of D178, E272, and D296 to complete the coordination (Video S2). Contacts and hydrogen bonds analysis between PEP and the hPKM2 active site (Figure 3) highlights the interaction of the PEP phosphate group with the Mg^{2+}ion corroborating the crucial role of the cation in coordinating PEP, as previously reported (20) (61). M291 and T328 directly engaged the enol pyruvate in interatomic contacts, while R73 and K270 formed hydrogen bonds with PEP carboxyl moiety (Figure 3A, Video S2). The MM-PBSA calculated binding free energy for compound 2 was -52.98 ± 23.08 kcal/mol, with E272 and D296 being the major contributors to the stability.

The stability of PEP during the MD simulation suggests that this is the binding mode the substrate adopts in the pre-catalytic hPKM2. We identified a set of interactions that stabilize PEP within the catalytic site. A network of hydrogen bonds is formed between the carboxyl group of PEP and R73 and K270. D178 interacts with the phosphate group through the Na^{+} ion, while E272 and D296 engaged the phosphate group through the Mg^{2+} ion, locking the orientation of the phosphate toward E272 and D296 while orienting the carboxyl group toward K115 and R120 (Figure 3B).

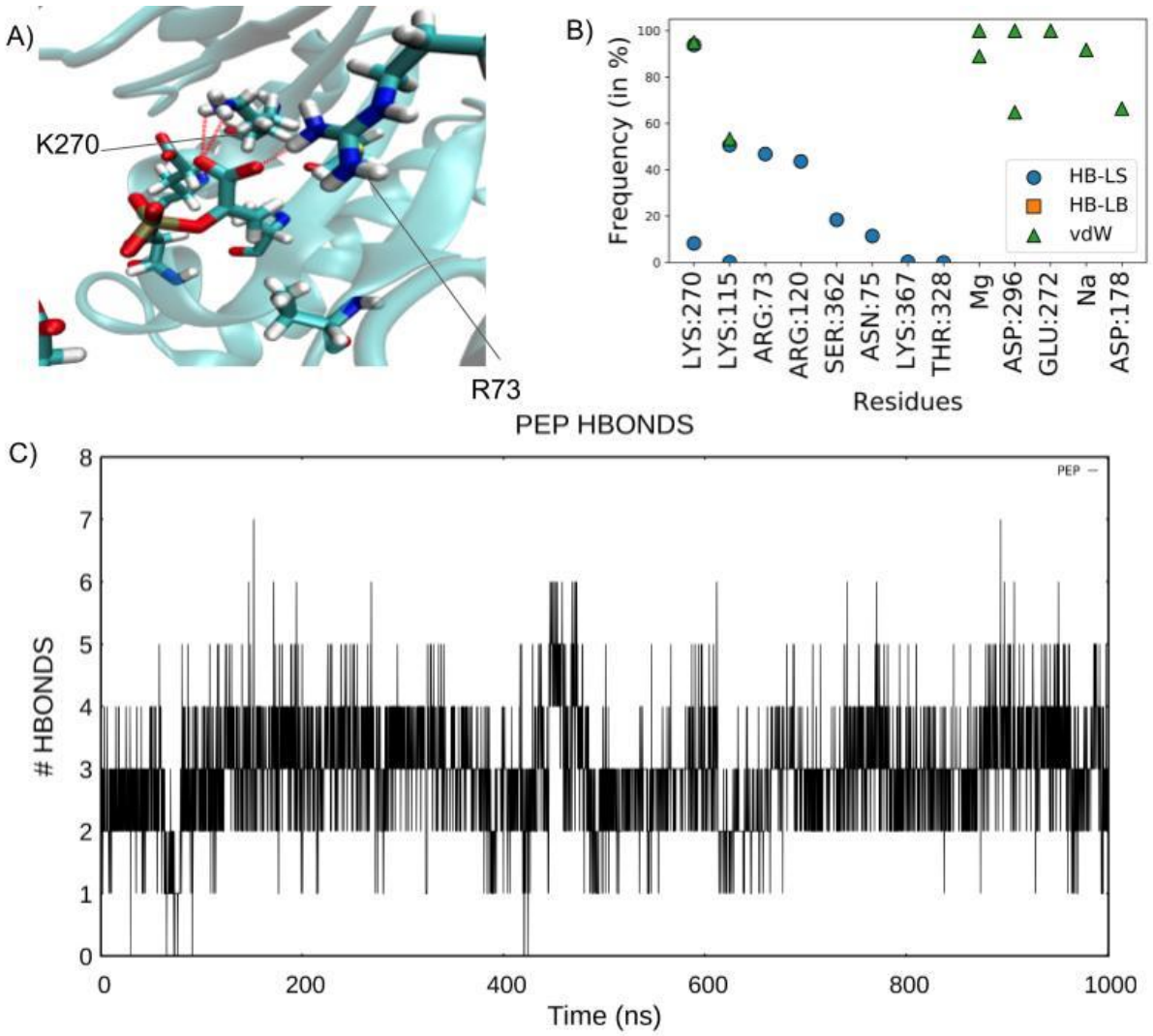


Figure 3. PEP Binding Site, RMSD, and Hydrogen Bond Network. A) PEP forms a hydrogen bond network throughout the simulation mainly with R73 and K270 (ions not shown) **B)** Contacts occupancy between PEP and hPKM2 are shown with the chain identifier, the residue name, and its number. The double vdW interactions for Mg^{2+} and D296 are due to individual interaction with each oxygen atom of the PEP phosphate group. PEP engaged with a Na^{+} ion through the phosphate group. No contact with the ligand's backbone was made during the simulation. C) hydrogen bond count during the simulation (2.93 ± 0.88). HB-LS: hydrogen bond between the ligand and the protein's sidechain; HB-LB: hydrogen bonds between the ligand and the protein's backbone alpha carbon atoms. vdW: ligand-protein interatomic contacts.

Halogenated PEP-derivatives are Stable in the Catalytic Site

As previously reported by Garcìa-Alles *et al* (29), the modification of the phosphate or the carboxylate groups reduces the binding affinity of PEP bio-isosteres toward hPKM2,

suggesting these moieties as crucial for binding, in accordance with our MD simulation of PEP. Compounds 2 and 3 (Table 1) bear both the carboxylate and phosphate with the addition of a halogen atom on the ethylene scaffold to deactivate the substrate thanks to an electronic effect. hPKM2 possesses Z-stereoselectivity for halogenated PEP-analogue inhibitors such as the (Z)-phosphoenol-3-fluoropyruvate (compound 2) and (Z)-phosphoenol-3-chloropyruvate (compound 3) (50,63). While fluorinated PEP derivatives inhibit the phosphotransferase reaction, the chlorinated counterparts have a modest effect. As the binding mode of these PEP derivatives s have not been described yet, we investigated their mechanism of action.

Compounds 2 and 3 Binding Mode

Compounds 2 and 3 molecular docking results showed both the phosphate group and the scaffold orientation in line with the X-ray structure. (Figures S1B, S1C). In the best binding pose, the fluorine atom of compound 2 was predicted between E272 and D296, while the carboxyl group oriented towards residue T238 and the phosphate group towards the position occupied by the potassium ion in the X-ray structure. Compound 2 established a set of hydrogen bond interactions with R73, K115, K270, and S362 (Figure 4A). Post-docking MD simulation displayed stabilizing interactions with residues K73, S362, K115, K270, T328, R120, T175, and E272 (Figure 4B). During the second half of the MD simulation, compound 2 rearranged within the active site S362 (Video S3) to remain in a stable configuration (RMSD of 2.8 ± 0.91 Å, RMSF of 2.1 ± 0.41 Å Figure S2A, Figure S2B respectively). The MM-PBSA computed binding affinity of compound 2 was -53.41 ± 18.6 kcal/mol, with R73 and K270 providing the best interactions.

Compound 3 molecular docking predicted a divergent binding mode compared to compound 2, probably due to the bulkier chlorine atom. The phosphate group engaged R73, K270 and S362 through a set of hydrogen bonds for the totality of the simulation (Figure 4C), while the carboxyl group coordinated with the Mg^{2+} ion along with E272 and D296. During the simulation, the chlorine atom position varied between A293, M291, and M360 (Video S3) with the molecule being in a stable position between residues R73, K270, S362, K115, D296, E272, M291, T328, A293, and A327 (Figure 4D), displaying a similar set of interactions as PEP (Figure 3B).

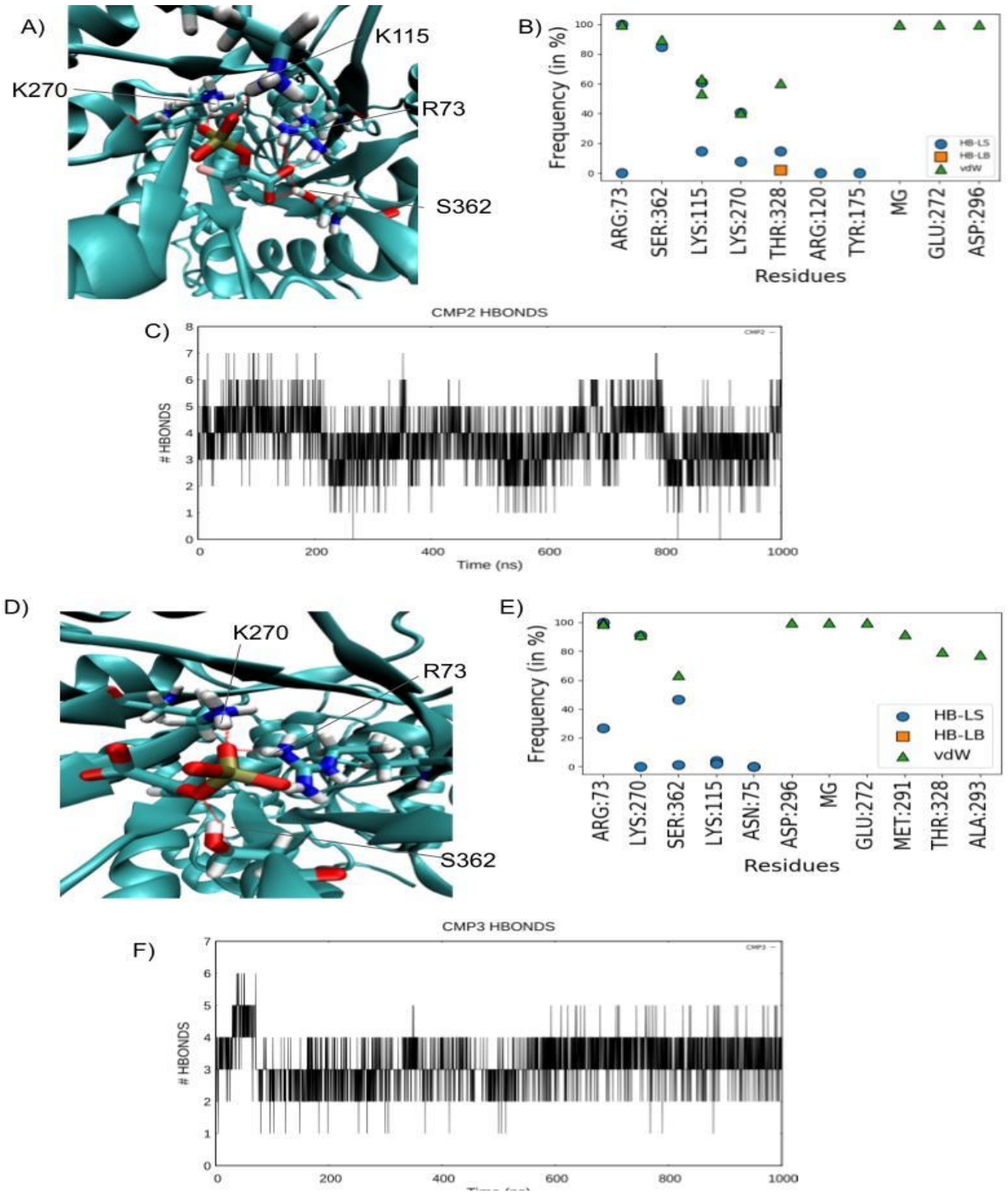


Figure 4. Compound 2 and Compound 3 Binding Analyses. A) Hydrogen bond network representation of compound 2. Ions are not shown for clarity **B)** Contact plot of compound 2 **C)** Compound 2 hydrogen bond formation plot over time shows 3.8 ± 1.03 bonds D) Hydrogen bond network representation of compound 3. Ions are not shown for clarity. **E)** Contact plot of compound 3. F) Compound 3 hydrogen bonds count over simulation (3.2 ± 0.63). HB-LS: hydrogen bond between the ligand and the protein's sidechain; HB-LB: hydrogen bonds between the ligand and the protein's backbone alpha carbon atoms. vdW: ligand-protein interatomic contacts.

Despite the relative mobility of compound 3 (average RMSD of 2.51 ± 0.5 Å and average RMSF of 1.4 ± 0.16 Å, Figure S2A, S2B respectively), the PBSA binding energy was -84.99 ± 9.5 kcal/mol, with R73, R120, K270, and K367 contributing the most to the binding affinity. Compound 3 phosphate engaged with K270, R73, and S362, suggesting the crucial role of this group in the binding mechanism (Figure 4C). Compound 3 differed from compound 2 by the stability of the scaffold, locked between R73, K270, M291, E272, and D296. For both compounds, Mg^{2+} and Na^{+} ions interacted with the phosphate group favoring the orientation of the ligand inside the catalytic site (Video S3).

In summary, the presence of the phosphate and the carboxyl groups is required to achieve favourable coordination of the ligands inside the active site. PEP and compound 2 engaged in a similar set of interactions, while compound 3 probably has a different binding mode, although still effective in inhibiting hPKM2 (Table 1) Overall, the shared scaffold between compounds 2, compound 3, and PEP favoured a similar stabilizing network of hydrophobic and hydrogen bond interactions.

A Sulfate Moiety Alters the Binding Mode within hPKM2

Compound 4 activity was the lowest of all the compounds tested (Table 1). The reason for this is that hPKM2 can still catalyse the sulfonyl transfer from the solfoenolpyruvate (compound 4) to ADP to yield adenosine 5'-sulfatopyrophosphate, with a 250-600 fold reduced reaction rate compared to PEP (64). Compound 4 best docking pose was used for post-docking MD simulation, which highlighted a diverged interaction fingerprint from PEP, although still comprising residues R73, K270, S362, T328, and D296, E272 through the Mg^{2+} ion (Figure 5A). In contrast to PEP, compound 4 engaged R73 and K270 through the carboxyl group for the first 600 ns of the simulation, while the sulfate moiety transitorily interacted with R73 and S362 (Video S3).

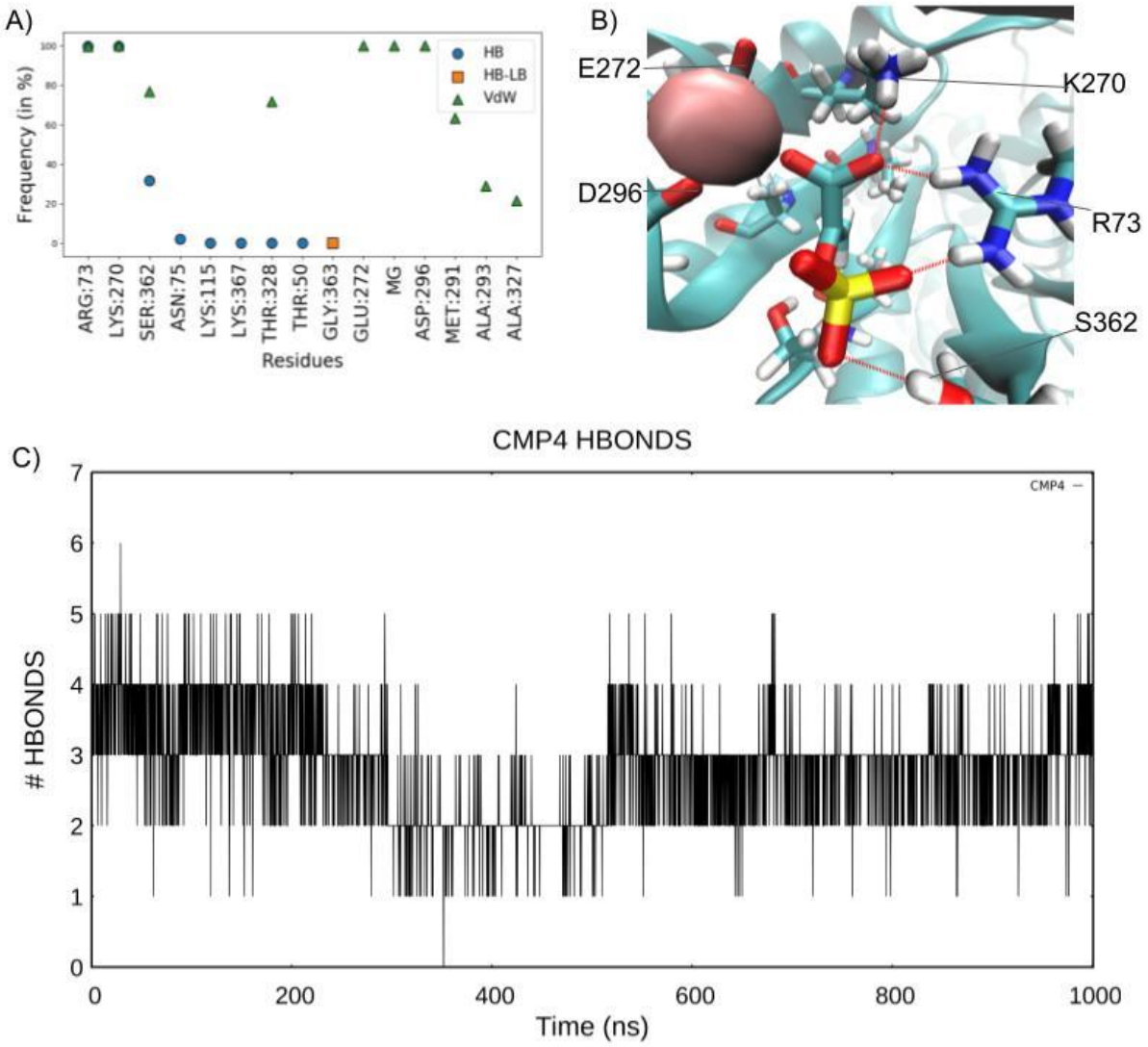


Figure 5. Compound 4 Binding Mode. A) Contact plot of compound 4 indicates a similar set of interactions with PEP, especially with R73, K270, and T328 **B)** Hydrogen bond network representation of compound 4. C) Compound 4 hydrogen bond count (2.84 ± 0.76) throughout the simulation. HB-LS: hydrogen bond between the ligand and the protein's sidechain; HB-LB: hydrogen bonds between the ligand and the protein's backbone alpha carbon atoms. vdW: ligand-protein interatomic contacts.

The subsequent sulfate displacement from the Mg^{2+} ion briefly produced a rearrangement of the scaffold orientation as indicated by the RMSD (RMSD = 2.18 ± 0.69 Å, RMSF= 1.83 ± 0.35 Å Figure S2A, S2B). The scaffold reverted to the original position with the sulfate group oscillating between R73 and S362. Compound 4 established a bidentate hydrogen bond interaction with R73 with the sulphate and the carboxyl group, which established a hydrogen bond also with K270. S362 engaged with the sulfate group, while on the other side of the compound, the carboxyl group coordinated with the Mg^{2+} along with E272 and D296 (Figure

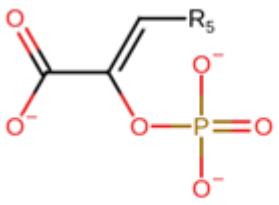
5A). The PBSA binding energy for compound 4 was -37.85 ± 6.3 kcal/mol, in line with its modest pIC_{50} (Table 1, Figure S6).

In summary, compound 4 displayed a different binding mode compared to PEP (Video S3) although engaged in the same set of residues as PEP, as the PEP phosphate interactions (i.e., the chelation of the Mg^{2+}) were sustained by compound 4 carboxyl group.

Extended Bio Isosteres Expand Residue Engagement in hPKM2 binding site

Our previous simulations highlighted the importance of both the carboxyl and phosphate groups for hPKM2 binding. The ethylene moiety of the scaffold did not engage significantly with any residue, suggesting that modifications or extensions on the scaffold could improve the binding. In addition, the presence of halogen atoms in compounds 2 and 3 negatively impacted the hepatotoxicity profile of the molecules (65,66), suggesting potential liver injury and poor clearance (1.865 ml/min/Kg and 1.773 ml/min/Kg). Compound 4 was presented with a better ADMET profile, sharing, however, a poor estimated clearance rate (1.896 ml/min/Kg). We, therefore, replaced the halogen atom in compounds 2 and 3 with a primary amide (Compound 5, Table 2) and docked it within the enzyme.

Table 2. Proposed Substitution



Compound 5 = R_5 $CONH_2$

Name	Molecular Weight (Da)	cLogP
Compound 5	207.96	-0.7016

During a post-docking MD simulation, compound 5 resided in the active site of hPKM2, resembling PEP interactions and binding mode. The carboxyl group engaged with R73, S362, K270, K115, T328, E272, and D296 (Figure 6A) locking the scaffold in a favourable orientation inside the active site. An Mg^{2+} cation coordinated both the phosphate and the carbonyl oxygen of the secondary amide, with D296 and E272, further stabilising the ligand (Video S4). The secondary amide contributed to the stabilization and orientation of the scaffold toward T328. This suggests that additional extensions of compound 5 which include T328 target residue, might be well-tolerated, as indicated by the computed binding energy for compound 5 of -79.36 ± 6.78 kcal/mol, with numerous sets of residues involved in the binding interaction such as R73, K115, R120, K207, K270 and K367. Compound 5, (RMSD = 2.96 ± 0.69 Å RMSF = 2.45 ± 0.41 Å) displayed a similar set of hydrogen as PEP (Video S4), with R73, K115, K270, T328, and S362 (Figure 6B).

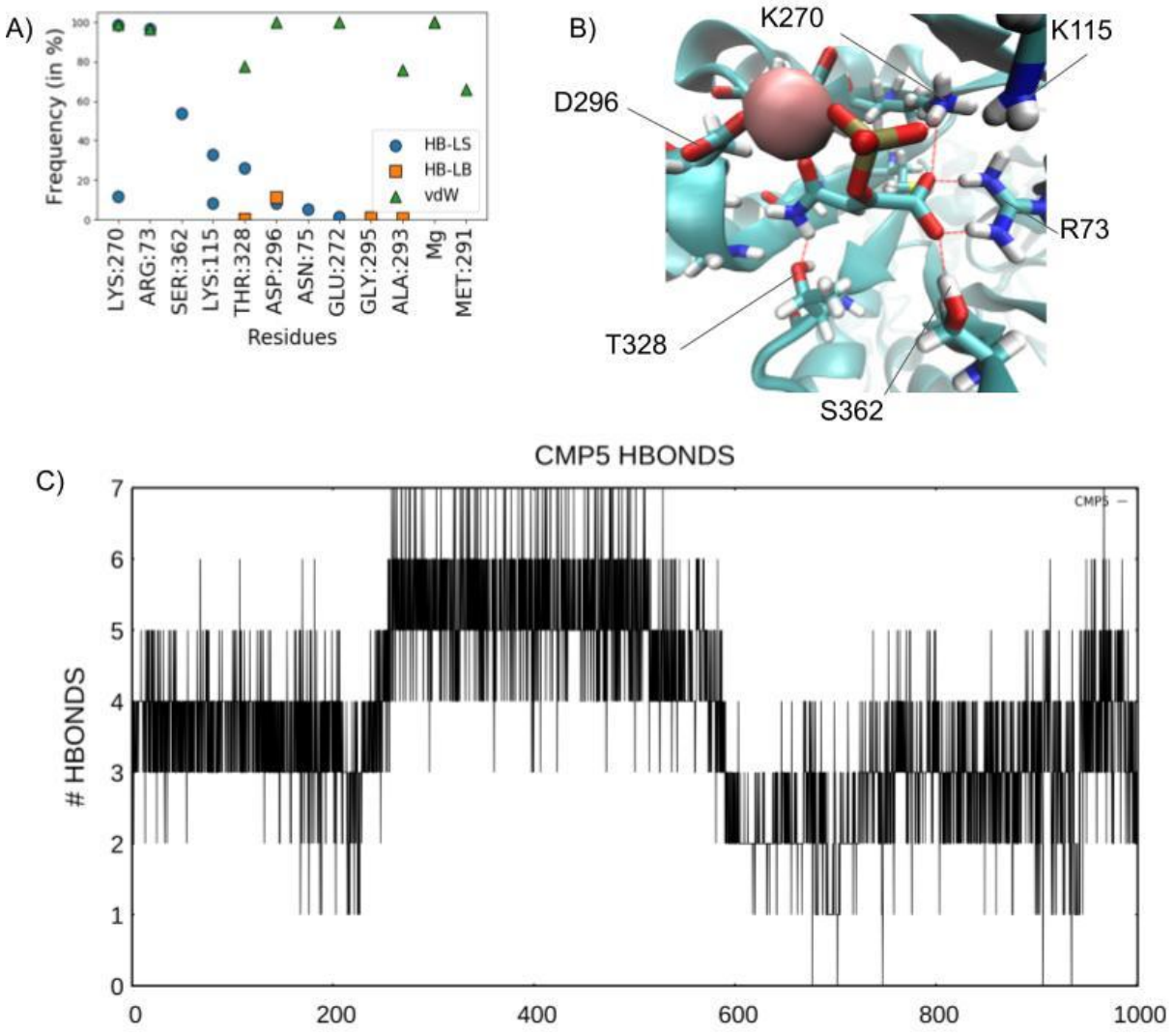


Figure 6. Compound 5 Binding Mode. A) Contact plot of compound 5 **B)** Hydrogen bond network representation of compound 5. C) Hydrogen bond count (3.76 ± 1.33) over the

simulation time. HB-LS: hydrogen bond between the ligand and the protein's sidechain; HB-LB: hydrogen bonds between the ligand and the protein's backbone alpha carbon atoms. vdW: ligand-protein interatomic contacts.

Taken together, these results confirm the necessity of negatively charged groups (i.e., carboxylate or phosphate) to participate in the coordination of Mg^{2+} by E272 and D296. The presence of the phosphate group with the addition of a secondary amide on the enol pyruvate scaffold contributed to the stabilization of the ligand in accordance with PEP binding mode, with the addition of an improved set of hydrogen bonds and hydrophobic interactions. Furthermore, the ADMET profile of compound 5 indicated no risk of liver injury and an overall safer toxicology profile including the carcinogenicity output.

Discussion

The impact of aerobic glycolysis in cancer cells remains one of the hallmarks of cancers. The metabolic shift that sustains the unregulated proliferative capabilities of cancerous cells suggests the main role of hPKM2 in unregulated ATP production. In concert with HIF-1 α expression, an increased hPKM2 activity favours a glycolytic state compared to the oxidative phosphorylation in the cancerous hypoxic TME. The increased glucose uptake with the biochemical pathway shifts initiates a signalling cascade that gives glycolytic cancer types an energetic advantage over regular cells as well as increasing chemotherapeutic resistance. Targeting the overexpressed hPKM2 is a viable, yet scarcely explored therapeutic approach to impact a large variety of cancers. Disrupting the glycolytic ATP production might result in a consistent energy depletion for hypoxic ATP-dependant cancerous cells, thus reducing cell replication rate and possibly causing the activation of the autophagic biochemical pathway. Furthermore, considering the hPKM2 role in drug resistance, the combined hPKM2 inhibition in conjunction with a chemotherapic regime might favour tumour regression with a combined effect derived from ATP depletion, autophagic mechanism, and a reduction in chemotherapy resistance. Disrupting the final irreversible limiting-step of the glycolytic pathway could trigger selective damage to the highly glucose-dependant cancerous cells while having a lower effect on healthy tissues that mainly rely on oxidative phosphorylation.

We provided for the first time insights on the binding mode of PEP within the hPKM2 catalytic site and rationalised the activity of bio-isosteres compounds. The co-presence of the phosphate and carboxyl groups plays an essential function in the orientation of the scaffold by chelating the Mg^{2+} ion, stabilising the ligand inside the active site. With compound 5, we aimed to design a PEP isostere by keeping the carboxyl and the phosphate groups, crucial elements for enzyme-ligand molecular recognition, expanding the set of interactions between

the compound and the active site. We also built the new molecule with no halogen atoms, considering their modest inhibitory impact, improving also the predicted toxicological profile of the molecule. As expected, compound 5 displayed a stable simulated binding mode against hPKM2. We believe it is possible to design new and more potent inhibitors based on compound 5, in which the primary amide is substituted to grow in the catalytic site and form further interactions, besides increasing the drug-likeness of the compounds.

Conclusion

We propose the importance of the bio-isosteric approach to developing new inhibitors to reduce hPKM2 activity in glycolytic cancer types. Further studies are required to expand on this current work and explore the chemical space of PEP analogues, as well as experimental data to validate our hypothesis.

Videos:

https://www.sciencedirect.com/science/article/pii/S0010482523003177?via%3Dihub#appsec1

References:

1. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011 Mar 4;144(5):646–74.

2. Fritz V, Fajas L. Metabolism and proliferation share common regulatory pathways in cancer cells. Oncogene. 2010 Aug 5;29(31):4369–77.

3. Ghobrial IM, Witzig TE, Adjei AA. Targeting apoptosis pathways in cancer therapy. CA Cancer J Clin. 2005 Jun;55(3):178–94.

4. Ganapathy-Kanniappan S, Geschwind J-FH. Tumor glycolysis as a target for cancer therapy: progress and prospects. Mol Cancer. 2013 Dec 3;12:152.

5. Warburg O. On the Origin of Cancer Cells. Science. 1956 Feb 24;123(3191):309–14.

6. Schwartz L, Supuran CT, Alfarouk KO. The warburg effect and the hallmarks of cancer. Anticancer Agents Med Chem. 2017;17(2):164–70.

7. Liberti MV, Locasale JW. The warburg effect: how does it benefit cancer cells? Trends Biochem Sci. 2016 Mar;41(3):211–8.

8. Vaupel P, Schmidberger H, Mayer A. The Warburg effect: essential part of metabolic reprogramming and central contributor to cancer progression. Int J Radiat Biol. 2019 Jul;95(7):912–9.

9. Muz B, de la Puente P, Azab F, Azab AK. The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy. Hypoxia (Auckl). 2015 Dec 11;3:83–92.

10. Gill KS, Fernandes P, O'Donovan TR, McKenna SL, Doddakula KK, Power DG, et al. Glycolysis inhibition as a cancer treatment and its role in an anti-tumour immune response. Biochim Biophys Acta. 2016 Aug;1866(1):87–105.

11. Hayashi Y, Yokota A, Harada H, Huang G. Hypoxia/pseudohypoxia-mediated activation of hypoxia-inducible factor-1α in cancer. Cancer Sci. 2019 May;110(5):1510–7.

12. Hirschhaeuser F, Sattler UGA, Mueller-Klieser W. Lactate: a metabolic key player in cancer. Cancer Res. 2011 Nov 15;71(22):6921–5.

13. San-Millán I, Brooks GA. Reexamining cancer metabolism: lactate production for carcinogenesis could be the purpose and explanation of the Warburg Effect. Carcinogenesis. 2017 Feb 1;38(2):119–33.

14. Ying M, You D, Zhu X, Cai L, Zeng S, Hu X. Lactate and glutamine support NADPH generation in cancer cells under glucose deprived conditions. Redox Biol. 2021 Oct;46:102065.

15. Bukowski K, Kciuk M, Kontek R. Mechanisms of multidrug resistance in cancer chemotherapy. Int J Mol Sci. 2020 May 2;21(9).

16. Icard P, Fournel L, Wu Z, Alifano M, Lincet H. Interconnection between Metabolism and Cell Cycle in Cancer. Trends Biochem Sci. 2019 Jun;44(6):490–501.

17. Macheda ML, Rogers S, Best JD. Molecular and cellular regulation of glucose transporter (GLUT) proteins in cancer. J Cell Physiol. 2005 Mar;202(3):654–62.

18. Israelsen WJ, Vander Heiden MG. Pyruvate kinase: Function, regulation and role in cancer. Semin Cell Dev Biol. 2015 Jul;43:43–51.

19. Israelsen WJ, Dayton TL, Davidson SM, Fiske BP, Hosios AM, Bellinger G, et al. PKM2 isoform-specific deletion reveals a differential requirement for pyruvate kinase in tumor cells. Cell. 2013 Oct 10;155(2):397–409.

20. Dombrauckas JD, Santarsiero BD, Mesecar AD. Structural basis for tumor pyruvate kinase M2 allosteric regulation and catalysis. Biochemistry. 2005 Jul 12;44(27):9417–29.

21. Dang CV. PKM2 tyrosine phosphorylation and glutamine metabolism signal a different view of the Warburg effect. Sci Signal. 2009 Nov 17;2(97):pe75.

22. Park JH, Kundu A, Lee SH, Jiang C, Lee SH, Kim YS, et al. Specific Pyruvate Kinase M2 Inhibitor, Compound 3K, Induces Autophagic Cell Death through Disruption of the Glycolysis Pathway in Ovarian Cancer Cells. Int J Biol Sci. 2021 May 5;17(8):1895–908.

23. Apostolidi M, Vathiotis IA, Muthusamy V, Gaule P, Gassaway BM, Rimm DL, et al. Targeting pyruvate kinase M2 phosphorylation reverses aggressive cancer phenotypes. Cancer Res. 2021 Aug 15;81(16):4346–59.

24. Zhou Y, Huang Z, Su J, Li J, Zhao S, Wu L, et al. Benserazide is a novel inhibitor targeting PKM2 for melanoma treatment. Int J Cancer. 2020 Jul 1;147(1):139–51.

25. Chen J, Xie J, Jiang Z, Wang B, Wang Y, Hu X. Shikonin and its analogs inhibit cancer cell glycolysis by targeting tumor pyruvate kinase-M2. Oncogene. 2011 Oct 20;30(42):4297–306.

26. Ning X, Qi H, Li R, Li Y, Jin Y, McNutt MA, et al. Discovery of novel naphthoquinone derivatives as inhibitors of the tumor cell specific M2 isoform of pyruvate kinase. Eur J Med Chem. 2017 Sep 29;138:343–52.

27. Wang X, Zhang F, Wu X-R. Inhibition of pyruvate kinase M2 markedly reduces chemoresistance of advanced bladder cancer to cisplatin. Sci Rep. 2017 Apr 5;7:45983.

28. Ning X, Qi H, Li R, Jin Y, McNutt MA, Yin Y. Synthesis and antitumor activity of novel 2, 3-didithiocarbamate substituted naphthoquinones as inhibitors of pyruvate kinase M2 isoform. J Enzyme Inhib Med Chem. 2018 Dec;33(1):126–9.

29. García-Alles LF, Erni B. Synthesis of phosphoenol pyruvate (PEP) analogues and evaluation as inhibitors of PEP-utilizing enzymes. Eur J Biochem. 2002 Jul;269(13):3226–36.

30. Chen JJ, Schmucker LN, Visco DP. Virtual high-throughput screens identifying hPK-M2 inhibitors: Exploration of model extrapolation. Comput Biol Chem. 2019 Feb;78:317–29.

31. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J Comput Chem. 2010 Mar;31(4):671–90.

32. Huang J, MacKerell AD. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. J Comput Chem. 2013 Sep 30;34(25):2135–45.

33. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK Predictions. J Chem Theory Comput. 2011 Feb 8;7(2):525–37.

34. Boedtkjer E, Pedersen SF. The acidic tumor microenvironment as a driver of cancer. Annu Rev Physiol. 2020 Feb 10;82:103–26.

35. Swietach P. What is pH regulation, and why do cancer cells need it? Cancer Metastasis Rev. 2019 Jun;38(1–2):5–15.

36. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res. 2004 Jul 1;32(Web Server issue):W665-7.

37. Doerr S, Harvey MJ, Noé F, De Fabritiis G. HTMD: High-Throughput Molecular Dynamics for Molecular Discovery. J Chem Theory Comput. 2016 Apr 12;12(4):1845–52.

38. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996 Feb;14(1):33–8, 27.

39. Harvey MJ, Giupponi G, Fabritiis GD. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. J Chem Theory Comput. 2009 Jun 9;5(6):1632–9.

40. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem. 2004 Oct;25(13):1605–12.

41. Xiong G, Wu Z, Yi J, Fu L, Yang Z, Hsieh C, et al. ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. Nucleic Acids Res. 2021 Jul 2;49(W1):W5–14.

42. Landrum G, Tosco P, Kelley B, Sriniker, Gedeck, NadineSchneider, et al. rdkit/rdkit: 2020_03_1 (Q1 2020) Release. Zenodo. 2020;

43. rdkit.Chem.rdMolDescriptors module — The RDKit 2022.03.1 documentation [Internet]. [cited 2022 Jun 3]. Available from: https://www.rdkit.org/docs/source/rdkit.Chem.rdMolDescriptors.html#rdkit.Chem.rdMolDescriptors.GetMorganFingerprint

44. Eberhardt J, Santos-Martins D, Tillack AF, Forli S. Autodock vina 1.2.0: new docking methods, expanded force field, and python bindings. J Chem Inf Model. 2021 Aug 23;61(8):3891–8.

45. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010 Jan 30;31(2):455–61.

46. Vanommeslaeghe K, MacKerell AD. Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing. J Chem Inf Model. 2012 Dec 21;52(12):3144–54.

47. Vanommeslaeghe K, Raman EP, MacKerell AD. Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. J Chem Inf Model. 2012 Dec 21;52(12):3155–68.

48. Wang J, Wang W, Kollman PA, DA Case. Antechamber: an accessory software package for molecular mechanical calculations. J Am Chem Soc. 2001;222.

49. Frisch A. Gaussian 09w reference. Wallingford, USA, 25p. 2009;

50. Duffy TH, Nowak T. Stereoselectivity of interaction of phosphoenolpyruvate analogues with various phosphoenolpyruvate-utilizing enzymes. Biochemistry. 1984 Feb 14;23(4):661–70.

51. 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.

52. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys. 1984;81(8):3684.

53. Loncharich RJ, Brooks BR, Pastor RW. Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. Biopolymers. 1992 May;32(5):523–35.

54. Hopkins CW, Le Grand S, Walker RC, Roitberg AE. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. J Chem Theory Comput. 2015 Apr 14;11(4):1864–74.

55. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. J Chem Phys. 1995;103(19):8577.

56. McGibbon RT, Beauchamp KA, Harrigan MP, Klein C, Swails JM, Hernández CX, et al. Mdtraj: A modern open library for the analysis of molecular dynamics trajectories. Biophys J. 2015 Oct 20;109(8):1528–32.

57. Miller BR, McGee TD, Swails JM, Homeyer N, Gohlke H, Roitberg AE. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. J Chem Theory Comput. 2012 Sep 11;8(9):3314–21.

58. Wang E, Sun H, Wang J, Wang Z, Liu H, Zhang JZH, et al. End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. Chem Rev. 2019 Aug 28;119(16):9478–508.

59. Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discov. 2015 May;10(5):449–61.

60. Donovan KA, Zhu S, Liuni P, Peng F, Kessans SA, Wilson DJ, et al. Conformational dynamics and allostery in pyruvate kinase. J Biol Chem. 2016 Apr 22;291(17):9244–56.

61. Larsen TM, Benning MM, Rayment I, Reed GH. Structure of the bis(Mg2+)-ATP-oxalate complex of the rabbit muscle pyruvate kinase at 2.1 A resolution: ATP binding over a barrel. Biochemistry. 1998 May 5;37(18):6247–55.

62. Muirhead H, Clayden DA, Barford D, Lorimer CG, Fothergill-Gilmore LA, Schiltz E, et al. The structure of cat muscle pyruvate kinase. EMBO J. 1986 Mar;5(3):475–81.

63. Liu J, Peliska JA, O'Leary MH. Synthesis and study of (Z)-3-chlorophosphoenolpyruvate. Arch Biochem Biophys. 1990 Feb 15;277(1):143–8.

64. Peliska JA, O'Leary MH. Sulfuryl transfer catalyzed by pyruvate kinase. Biochemistry. 1989 Feb 21;28(4):1604–11.

65. Safari S, Motavaf M, Seyed Siamdoust SA, Alavian SM. Hepatotoxicity of halogenated inhalational anesthetics. Iran Red Crescent Med J. 2014 Sep 5;16(9):e20153.

66. Wang Y, Ming X-X, Zhang C-P. Fluorine-Containing Inhalation Anesthetics: Chemistry, Properties and Pharmacology. Curr Med Chem. 2020;27(33):5599–652.

8.2 Hidden GPCR structural transitions addressed by multiple walkers supervised molecular dynamics (mwSuMD)

Supervised molecular dynamics (SuMD) is a robust method for investigating the pathways of ligand-receptor binding and unbinding. In this project, I contributed to the development of the multiple walker SuMD (mwSuMD). mwSuMD allows for the exploration of a broader spectrum of conformational transitions relevant to drug design. The method has been first benchmarked to G protein-coupled receptors (GPCRs) for validation. These GPCRs constitute the most abundant family of membrane receptors in eukaryotes and serve as both fundamental drug targets and well-established test systems, representing more than one-third of drugs approved for human use. The new mwSuMD greatly enhances the sampling capacities of the SuMD method in exploring the binding pathways and the ligand-target molecular recognition mechanism. mwSuMD supervises the evolution of different metrics of the system such as the RMSD, distance, number of hydrogen bonds and contacts between two molecular entities to reveal the binding or unbinding pathway of the two selections, as well as the conformational changes over time, using energetically-unbiased MD.

I also introduced significant improvements to the existing SuMD, such as multi-engine versatility to run independently with NAMD, ACEMD, GROMACS, and OpenMM, multiprocessing parallelization, code refactorization and algorithm improvement for faster calculations, as well as remodelling the general architecture for a more user-friendly application programming interface (API). The code has been deposited on my GitHub (https://github.com/pipitoludovico) public repository and is freely accessible from my profile. The work and results of mwSuMD are published as the preprint: “Hidden GPCR structural transitions addressed by multiple walkers supervised molecular dynamics (mwSuMD)”.

Authors: Giuseppe Deganutti, Ludovico Pipitò, Roxana M. Rujan, Tal Weizmann, Peter Griffin, Antonella Ciancetta, Stefano Moro, and Christopher A. Reynolds

Abstract

G protein-coupled receptors (GPCRs) are the most abundant membrane proteins and the target of about 35% of approved drugs. Despite this, the structural basis of GPCR pharmacology is still a matter of intense study. Molecular dynamics (MD) simulations aim to expand our knowledge of GPCR dynamics by building upon the recent advances in structural biology. However, the timescale limitations of classic MD hinder its applicability to numerous structural processes happening in time scales longer than microseconds (hidden structural transitions). For this reason, the overall MD impact on the study of GPCRs pharmacology

and drug design is still limited. To overcome this, we have developed an unbiased adaptive sampling algorithm, namely multiple walker supervised MD (mwSuMD), and tested it on different hidden transitions involving GPCRs. By increasing the complexity of the simulated process, we report the binding and unbinding of the vasopressin peptide, the inactive-to-active transition of the glucagon-like peptide-1 receptor (GLP-1R), the stimulatory G protein (G_s) and inhibitory G_i binding to the adrenoreceptor b_2 (b_2 AR) and the adenosine 1 receptor (A_1R) respectively, and the heterodimerization between the adenosine receptor A_2 ($A_{2A}R$) and the dopamine receptor D_2 (D_2R). We demonstrate that mwSuMD is a helpful tool for studying at the atomic level GPCR transitions that are challenging to address with classic MD simulations.

Introduction

Supervised molecular dynamics (1,2) (SuMD) is a powerful technique for studying ligand-receptor binding and unbinding pathways; here we present a significant enhancement to the method, namely multiple walker supervised molecular dynamics (mwSuMD) that permits a wider range of conformational transitions relevant to drug design to be studied. We validated the method by applying it to G protein-coupled receptors (GPCRs), as these are both fundamental drug targets and well-validated test systems. GPCRs are the most abundant family of membrane receptors in eukaryotes (3) and the target for more than one-third of drugs approved for human use (4). Vertebrate GPCRs are subdivided into five subfamilies (Rhodopsin or class A, Secretin or class B, Glutamate or class C, Adhesion, and Frizzled/Taste2) according to function and sequence (5,6). Common features of all GPCRs are seven transmembrane (TM) helices connected by three extracellular loops (ECLs) and three intracellular loops (ICLs), while an extended and structured N-terminus extracellular domains (ECD) is found in all subtypes, but class A. The primary function of GPCRs is transducing extracellular chemical signals into the cytosol by binding and activating four G protein families ($G_{s/olf}$, $G_{i/o}$, $G_{12/13}$ and $G_{q/11}$) responsible for decreasing ($G_{i/o}$) or increasing ($G_{s/olf}$) the cyclic adenosine-3',5'-monophosphate (cAMP), and generating inositol-1,4,5-triphosphate (IP_3) and diacylglycerol (DAG) to increase Ca^{2+} intracellular levels (G_q) (7).

GPCR structures have been solved by X-ray and cryo-electron microscopy (cryo-EM) at an increasing pace since the first X-ray structures in 2000 (8) and 2007 (9). However, many aspects of their pharmacology remain elusive. For example, the structural determinants of the selectivity displayed towards specific G proteins or the ability of certain agonists to drive a preferred intracellular signaling pathway over the others (i.e. functional selectivity or bias) (10). What makes GPCRs challenging proteins to characterize with standard techniques is

their inherent flexibility and the transitory nature of the complexes formed with extracellular and intracellular effectors. One of the possible approaches to integrate or sometimes overcome the limits of experimental conditions is performing molecular dynamics (MD) simulations. MD is a computational methodology that predicts the movement and interactions of (bio)molecules in systems of variable complexity, at atomic detail, enabling useful working hypotheses and rationalization of experimental data. However, standard MD sampling is limited to the microsecond or, in the best conditions, the millisecond time scale (11,12). For this reason, different algorithms have been designed to speed up the simulation of rare events such as ligand (un)binding and conformational transitions. Amongst the most popular and effective ones, there are metadynamics (13), accelerated MD (aMD) (14), and Gaussian-accelerated MD (GaMD) (15). Such methods, which introduce an energy potential to overcome the energy barriers preventing the complete exploration of the free energy surface, thus de facto biasing the simulation, have been used to propose activation mechanisms of GPCRs (16,17). Energetically unbiased MD protocols, on the other hand, comprise the weighted ensemble MD (weMD) (18) and SuMD (1,19). SuMD has been successfully applied to the (un)binding mechanism of both small molecules, peptides, and small proteins (1,19–23). Since SuMD is optimized only for (un)bindings, we have designed a new version of the software, namely multiple walker SuMD (mwSuMD), that extends the applicability of the method to conformational transitions and protein:protein binding.

We tested mwSuMD on a series of increasingly complex hidden structural transitions involving both class A and class B1 GPCRs. Firstly, we validated the method on the nonapeptide arginine vasopressin (AVP) by simulating binding (dynamic docking) and unbinding paths from the vasopressin 2 receptor (V_2R). AVP is an endogenous hormone that mediates antidiuretic effects on the kidney by signaling through three class A GPCR subtypes: V_{1a} and V_{1b} receptors activate phospholipases via $G_{q/11}$ protein, while the V_2 receptor (V_2R) activates adenylyl cyclase by interacting with G_s protein (24) and is a therapeutic target for hyponatremia, hypertension, and incontinence (25). Dynamic docking, although more computationally demanding than standard molecular docking, provides insights into the binding mode of ligands in a fully hydrated and flexible environment. Moreover, it informs about binding paths and the complete mechanism of formation leading to an intermolecular complex, delivering in the context of binding kinetics (26) and structure-kinetics relationship (SKR) studies (27).

We then show that mwSuMD can be employed to simulate the receptor activation of the class B1 GPCR glucagon-like peptide-1 receptor (GLP-1R) upon binding of the small molecule PF06882961. GLP-1R is a validated target in type 2 diabetes and probably the best-characterized class B1 GPCR from a structural perspective. GLP-1R is the only class

B1 receptor with structurally characterized non-peptidic orthosteric agonists, which makes it a model system for studying the druggability of the entire B1 subfamily.

The further case studies we report are the G_s and G_i proteins binding to the adrenoreceptor b_2 (b_2 AR) and the adenosine 1 receptor ($A_1R_)$, starting from different conditions. GPCRs preferentially couple to very few G proteins out of 23 possible counterparts (28,29). More importantly, agonists can modify the receptor selectivity profile by imprinting unique intracellular conformations from the orthosteric binding site. The mechanism behind these phenomena is one of the outstanding questions in the GPCR field (28). It is increasingly accepted that dynamic and transient interactions determine whether the encounter between a GPCR and a G protein results in productive or unproductive coupling (30). MD simulations are considered a useful orthogonal tool for providing working hypotheses and rationalizing existing data on G protein selectivity. However, so far, it has not delivered as expected. Attempts so far have employed energetically biased simulations or have been confined to the Ga subunit (16,17).

The last GPCR key process simulated through mwSuMD is the heterodimerization in the membrane between the adenosine receptor A_2 ($A_{2A}R$) and the dopamine receptor D_2 (D_2R). The $A_{2A}R$:D_2R heterodimer (31) is a therapeutic target for neurodegenerative diseases, Parkinson's disease, and schizophrenia (32–34) due to the reciprocal antagonistic allosteric effect between monomers (35). $A_{2A}R$ activation reduces the binding affinity of D_2R agonists, while $A_{2A}R$ antagonists enhance the dopaminergic tone by decreasing the adenosine negative allosteric modulation on D_2R. Heterobivalent ligands able to inhibit $A_{2A}R$ and activate D_2R represent a valuable pharmacological tool (36) and, in principle, therapeutic options for conditions characterized by reduction of dopaminergic signaling in the central nervous system. The successive dynamic docking of the heterobivalent ligand compound 26 (37) to the heterodimer suggested by mwSuMD produced a ternary complex stabilized by lipids.

Results and Discussion

Short mwSuMD time windows improve the AVP dynamic docking prediction

AVP has an amphipathic nature and interacts with both polar and hydrophobic V_2R residues located on both TM helices and ECLs. Although AVP presents an intramolecular C1-C6 disulfide bond that limits the overall conformational flexibility of the backbone, it has a high number of rotatable bonds, making dynamic docking complicated (38). We assessed the performance of mwSuMD and the original version of SuMD in reconstructing the experimental V_2R:AVP complex using different settings, simulating a total of 92 binding events . As a reference, the AVP RMSD during a classic (unsupervised) equilibrium MD

simulation of the AVP:V_2R complex was 3.80 ± 0.52 Å. SuMD (1,19) produced a minimum root mean square deviation (RMSD) to the cryo-EM complex of 4.28 Å, with most of the replicas (distribution mode) close to 10 Å (Figure 1a). MwSuMD, with the same settings (Figure 1b, Table S1) in terms of time window duration (600 ps), metric supervised (the distance between AVP and V_2R), and acceptance method (slope) produced slightly more precise results (distribution mode RMSD = 7.90 Å) but similar accuracy (minimum RMSD = 4.60). Supervising the AVP RMSD to the experimental complex rather than the distance (Figure 1c) and using the SMscore (Equation 1) as the acceptance method (Figure 1d) worsened the prediction. Supervising distance and RMSD at the same time (Figure 1e), employing the DMscore (Equation 2), recovered accuracy (minimum RMSD = 4.60 Å) but not precision (distribution mode RMSD = 12.40 Å). Interestingly, decreasing the time window duration from 600 ps to 100 ps impaired the SuMD ability to predict the experimental complex (Figure 2a), but enhanced mwSuMD accuracy and precision (Figure 2b-d). The combination of RMSD as the supervised metric and SMscore produced the best results in terms of minimum RMSD and distribution mode RMSD, 3.85 Å and 4.40 Å, respectively (Figure 2d, Video S1), in agreement with the AVP deviations in the equilibrium MD simulation of the AVP:V_2R complex.

These results suggest that short time windows can dramatically improve the dynamic docking performance of mwSuMD. However, it is necessary to know the final bound state to employ the RMSD, while the distance as the supervised metric is required to dynamically dock ligands with unknown bound conformation. Both distance and RMSD-based simulations delivered insights into the binding path and the residues involved along the recognition route. For example, mwSuMD suggested V_2R residues E184^{ECL2}, P298^{ECL3}, and E303^{ECL3} as involved during AVP binding, although not in contact with the ligand in the orthosteric complex.

Further to binding, a SuMD approach was previously employed to reconstruct the unbinding path of ligands from several GPCRs (1,2), (39). We assessed mwSuMD capability to simulate AVP unbinding from V_2R. Five mwSuMD and five SuMD replicas were collected using 100 ps time windows (Table 1). Overall, mwSuMD outperformed SuMD in terms of time required to complete a dissociation (Video S2), producing dissociation paths almost 10-fold faster than SuMD. Such rapidity in dissociating inherently produces a limited sampling of metastable states along the pathway, which can be compensated by seeding classic (unsupervised) MD simulations from configurations extracted from the unbinding route (40,41). Here, the set of V_2R residues involved during the dissociation was comparable to the binding, though ECL2 and ECL3 were slightly more involved during the association than the dissociation, in analogy with other class A and B GPCRs (21,40).

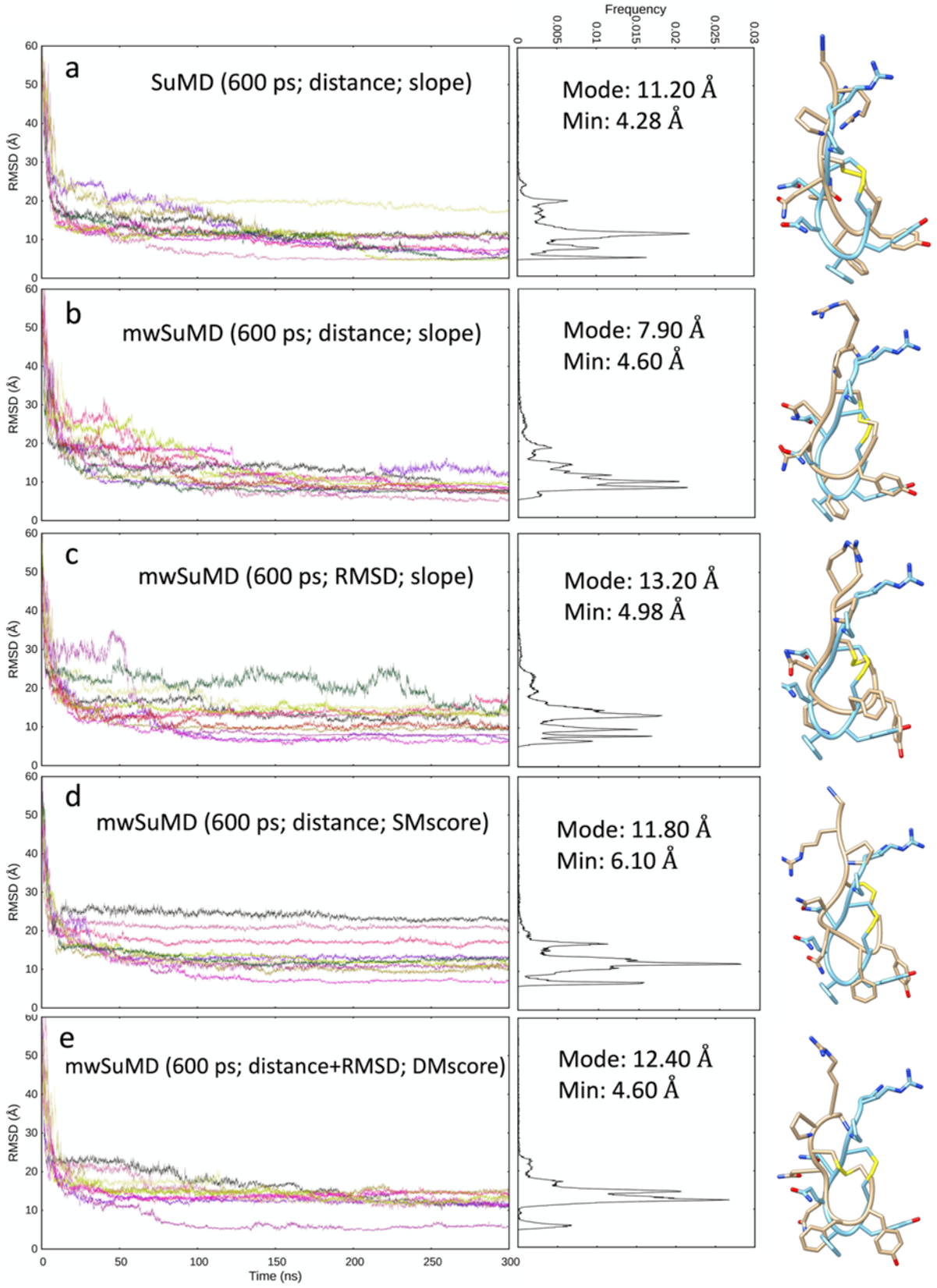


Figure 1. AVP SuMD and mwSuMD binding simulations to V_2R (600 ps time windows). For each set of settings **(a-e),** the RMSD of AVP Ca atoms to the cryo-EM structure 7DW9 is reported during the time course of each SuMD (**a**) or mwSuMD (**b-e**) replica alongside the

RMSD values distribution and the snapshot corresponding to the lowest RMSD values (AVP from the cryo-EM structure 7DW9 in cyan stick representation, while AVP from simulations in a tan stick). A complete description of the simulation settings is reported in Table 1 and the Methods section.

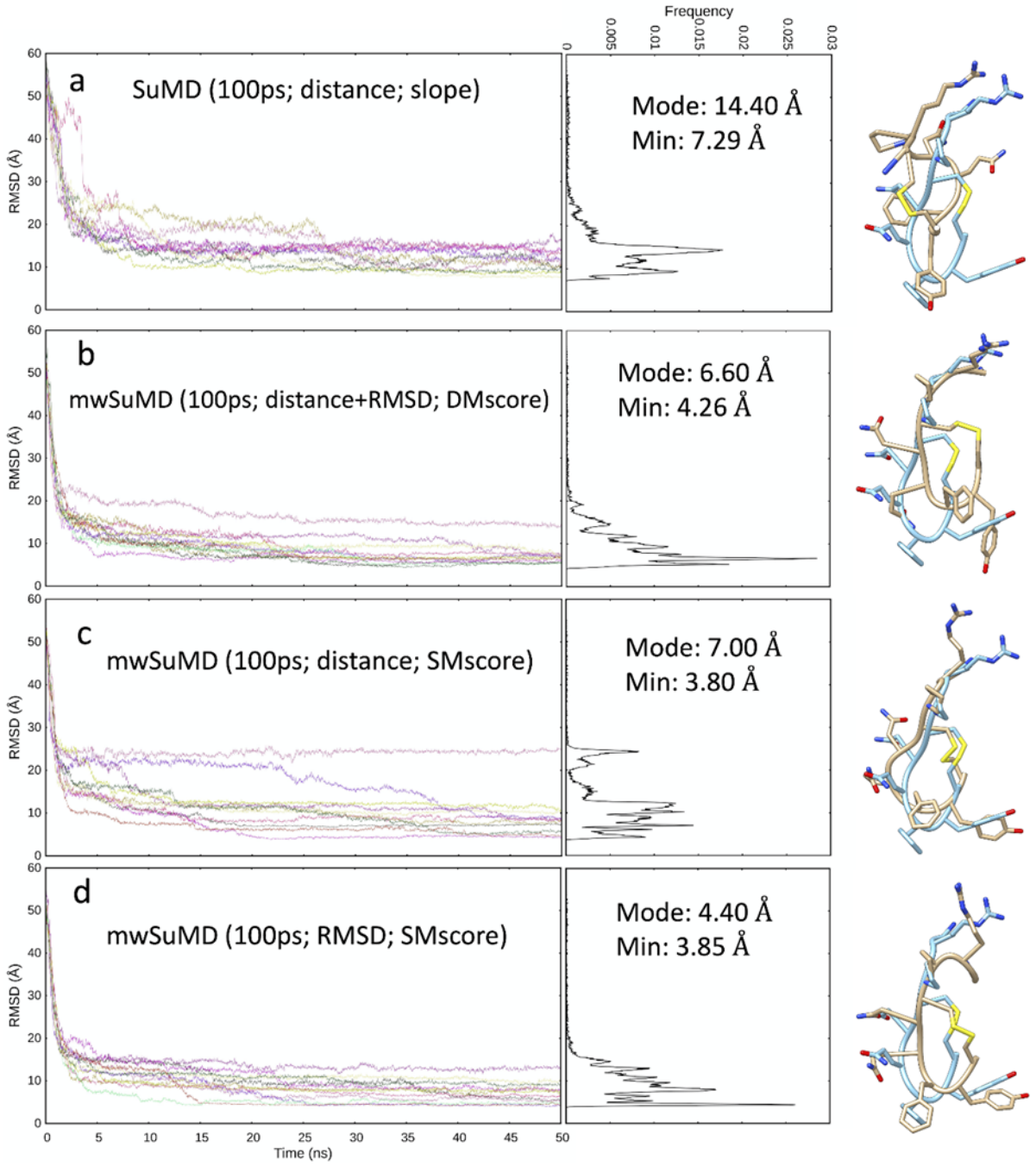


Figure 2. AVP SuMD and mwSuMD binding simulations to V_2R (100 ps time windows). For each set of settings **(a-d)** the RMSD of AVP Ca atoms to the cryo-EM structure 7DW9 is reported during the time course of each SuMD (**a**) or mwSuMD (**b-d**) replica alongside the

RMSD values distribution and the snapshot corresponding to the lowest RMSD values (AVP from the cryo-EM structure 7DW9 in cyan stick representation, while AVP from simulations in a tan stick). A complete description of the simulation settings is reported in Table 1 and the Methods section.

PF06882961 binding and GLP-1R activation

The GLP-1R has been captured by cryo-EM in both the inactive apo (ligand-free) and the active (G_s-bound) conformations, and in complex with either peptides or non-peptide agonists (42–47). In the inactive apo GLP-1R, residues forming the binding site for the non-peptide agonist PF06882961 are dislocated and scattered due to the structural reorganization of the transmembrane domain (TMD) and extracellular domain (ECD) that occurs on activation. Moreover, GLP-1R in complex with GLP-1 or different agonists present distinct structural features, even amongst structurally related ligands. This complicates the scenario and suggests divergent recognition mechanisms amongst different agonists. We simulated the binding of PF06882961 using multistep supervision on different metrics of the system (Figure 3) to model the structural hallmark of GLP-1R activation (Video S3, Video S4).

Several metrics were supervised in a consecutive fashion. Firstly, the distance between PF06882961 and the TMD as well as the RMSD of the ECD to the active state (stage 1); secondly, the RMSD of ECD and ECL1 to the active state (stage 2); thirdly, the RMSD of PF06882961 and ECL3 to the active state (stage 3); lastly, only the RMSD of TM6 (residues I345-F367, Ca atoms) to the active state (stage 4). The combination of these supervisions produced a conformational transition of GLP-1R towards the active state (Figure 3, Video S4). Noteworthy, the sequence of these supervisions was arbitrary and does not necessarily reflect the right order of the steps involved in GLP-1R activation. This kind of planned multistep approach is feasible when the end-point receptor inactive and active structures are available, and the inherent flexibility of different domains is known. In class B GPCRs, the ECD is the most dynamic sub-structure, followed by the ECL1 and ECL3 which display high plasticity during ligand binding (21,48). For this reason, we first supervised these elements of GLP-1R, leaving the bottleneck of activation, TM6 outward movement, as the last step. However, the protocol employed can be tweaked to study how each conformational transition takes place and influences the receptor domains. Structural elements not directly supervised, such as TM1 or TM7, displayed an RMSD reduction to the active state because they were influenced by the movement of supervised helixes or loops. For example, the supervision of ECL3 (stage 3) and TM6 (stage 4) facilitated the spontaneous rearrangement

of the ECD to an active-like conformation after the ECD had previously experienced transient high flexibility during stages 2 and 3 (Figure 3).

During the supervision of ECL3 and PF06882961 (stage 3), we observed a loosening of the intracellular polar interactions that stabilize GLP-1R TM6 in the inactive state. As a result, the subsequent supervision of TM6 (residues I345-F367, Ca atoms) rapidly produced the outward movement towards the active state, in the last step of the mwSuMD simulation (stage 4). Taken together, these results suggest a concerted conformational transition for ECD and ECL1 during the binding of PF06882961 and an allosteric effect between ECL3 and the bottom of TM6. Interestingly, while the intracellular polar interactions were destabilized by the ECL3 transition to an active-like conformation (stages 2 and 3), the outward movement of TM6 (stage 4) did not favor the closure of ECL3 towards PF06882961, which appear to be driven by direct interactions between the ligand and $R310^{5.40}$ or $R380^{7.35}$. Since we were interested in reconstructing the binding of PF06882961 to GLP-1R and the successive receptor structural transitions to prepare the intracellular G protein binding site, our mwSuMD simulation did not include G_s. Therefore, any allosteric effect triggered by the binding of the effector could have been overlooked, as well as the complete stabilization of TM6 in the active conformation, which is known to be achieved only when the intracellular effector is bound (49).

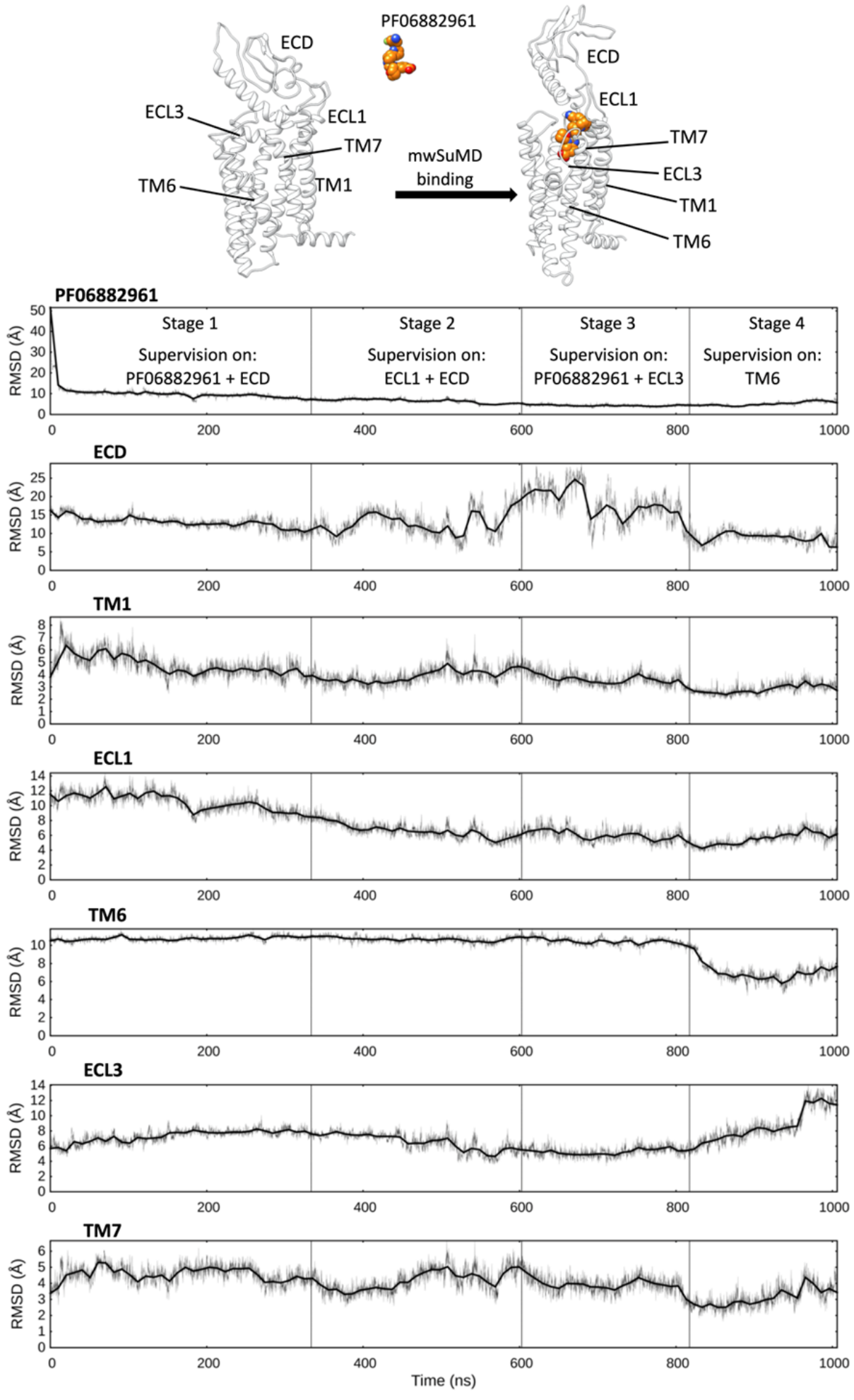

PF06882961
ECD
ECL3
ECL1
TM7
TM6
TM1
mwSuMD
binding
ECD
ECL1
TM7
ECL3
TM1
TM6
PF06882961
Stage 1
Supervision on:
PF06882961 + ECD
Stage 2
Supervision on:
ECL1 + ECD
Stage 3
Supervision on:
PF06882961 + ECL3
Stage 4
Supervision on:
TM6
ECD
TM1
ECL1
TM6
ECL3
TM7
RMSD (Å)
Time (ns)

Figure 3. MwSuMD simulation of PF06882961 binding to GLP-1R and receptor activation. Each panel reports the root mean square deviation (RMSD) to a GLP-1R structural element or the position of the ligand in the active state (top panel), over the time course (all but ECL3 converging to the active state). ECD: extracellular domain; TM: transmembrane helix; ECL: extracellular loop. The mwSuMD simulation was performed with four different settings over 1 microsecond in total.

G proteins ¾ class A GPCR binding simulations

We tested the ability of mwSuMD to simulate the binding between the prototypical class A receptor, the b_2 adrenoreceptor (b_2 AR), and the stimulatory G protein (G_s), without energy input. mwSuMD simulations started from the intermediate, agonist-bound conformation of b_2 AR and the inactive G_s to resemble pre-coupling conditions. Three mwSuMD replicas were performed by supervising the distance between G_s helix 5 (H5) and b_2 AR as well as the RMSD of the intracellular end of TM6 to the fully-active state of the receptor . To monitor the progression of the simulations, we computed the RMSD of the Ca atoms of the Ga and Gb subunits to the experimental complex (50) (Video S5, Figure 4ab). During two out of three replicas, both Ga and Gb reached values close to 5 Å (minimum RMSD = 3.94 Å and 3.96 Å respectively), in good agreement with the reference (the b_2 AR:G_s complex, PDB 3SN6, Figure 4c). The flexibility of G_sb is backed by both MD and cryo-EM data suggesting G protein rocking motions around G_sa:receptor interactions (21,51).

According to the model of G protein activation, the binding to the receptor allosterically stabilizes the orthosteric agonist, adrenaline in our simulations, and destabilizes the guanosine 5'-diphosphate (GDP) within Ga, resulting in the exchange with the ribonucleoside guanosine 5'triiphosphate (GTP) upon opening of the G protein alpha-helical domain (AHD). triggering the subsequent dissociation of Ga from Gbg. In our simulations, adrenaline was not further stabilized in the timescale of the simulations (Figure 4d), probably because the simulations sampled intermediate states, therefore, suboptimal b_2 AR:G_s interactions that were unable to allosterically stabilize the agonist. Upon receptor activation by the orthosteric agonist, TM6 undergoes an outward movement to accommodate the G protein that is accompanied by an anticlockwise rotation. We did not observe this rotation, which suggests that mwSuMD did not sample the complete G_s coupling. One of the b_2 AR residues undergoing rotation upon receptor activation is E268$^{6.30}$, involved in the conserved salt bridge (named ionic lock) with R131$^{3.50}$ that stabilizes the inactive state. Interestingly, during simulations, E268$^{6.30}$ formed hydrogen bonds with the G_s residues R385$^{H5.17}$, and R389$^{H5.21}$, both conserved across G protein subfamilies G_s, $G_{i/o}$, and $G_{q/11}$. We speculate that these interactions, not observed in any GPCR active state cryo-EM or X-ray structure, stabiles the

early stage of G_s binding and that the TM6 full rotation occurs at a late stage of the coupling as a rate-limiting step of the process. GDP, instead, was slightly destabilized by G_s binding to b_2 AR (Figure 4e), although a complete dissociation requires the opening of the AHD, the first step for GDP release, which requires timescales longer than our simulations (52).

Usually, ICL3 of the GPCR and the G protein loop hgh4 are masked out from deposited cryo-EM structures due to their high flexibility and therefore low resolution. During our simulations, these two loops formed polar intermolecular interactions through $R239^{ICL3}$, $R260^{ICL3}$, $K235^{ICL3}$, and $E322^{hgh4.12}$, $D323^{hgh4.13}$. Further transient interactions not visible in the experiential structures, involved a mix of conserved and unique residues forming hydrogen bonds : $R63^{ICL1}$-$E392^{H5.24}$, $K232^{5.71}$-$D378^{H5.10}$, $K235^{5.74}$- $D378^{H5.10}$, $K235^{ICL3}$-$D343^{H4.13}$, $K267^{6.29}$-$L394^{c}$, $R239^{ICL3}$-$E314^{hgh4.04}$, and $S137^{3.56}$-$D381^{H5.13}$. None of the interactions reported in Table S2 is evident from the experimental b_2 AR:G_s complex, implying that mwSuMD can deliver useful working hypotheses for mutagenesis and spectroscopic experiments from out-of-equilibrium simulations. Results also suggest that the G_s binding is driven by a combination of conserved and unique transitory interactions with b_2 AR, possibly contributing to G protein selectivity. The conserved interactions would be necessary for the binding regardless of the receptor:G protein couple involved, while the transitory interactions should produce an effective engagement of the G protein.

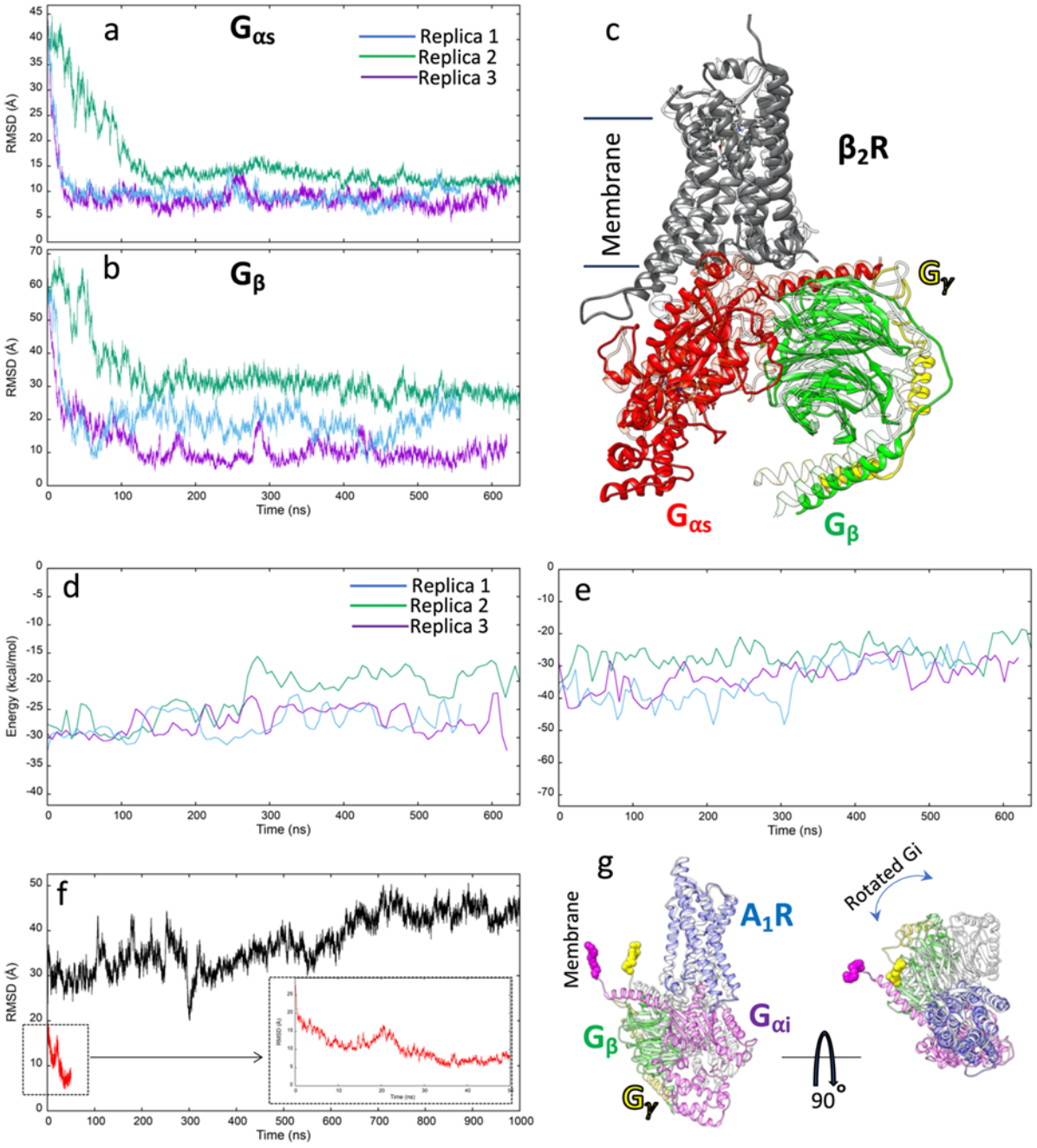


Figure 4. G protein binding simulations to b_2AR and A_1R. a) RMSD of G_sa to the experimental complex (PDB 3NS6) during three mwSuMD replicas; **b)** RMSD of G_sb to the experimental complex (PDB 3NS6) during three mwSuMD replicas; **c)** superposition of the experimental G_s: **b_2** AR complex (transparent ribbon) and the MD frame with the lowest G_sa RMSD **(**3.94 Å**); d**) adrenaline MM-GBSA binding energy during three mwSuMD replicas; **e**) GDP MM-GBSA binding energy during three mwSuMD replicas; **f**) RMSD of G_ia (residues 243-355) to the experimental complex (PDB 6D9H) during a mwSuMD simulation (red, magnified in the box) and a 1000-ns long classic MD simulation (black); **g**) two-view

superposition of the experimental G_i:**A_1** R complex (transparent ribbon) and the MD frame with the lowest G_ia RMSD **(**4.82 Å**).**

A possible pitfall of the above reported G_s:b_2 AR mwSuMD binding simulation is that G proteins bear potential palmitoylation and myristoylation sites that can anchor the inactive trimer to the plasma membrane (53,54), de facto restraining possible binding paths to the receptor. To address this point and test the possible system dependency of mwSuMD, we prepared a different class A GPCR, the adenosine A1 receptor (A_1R), and its principal effector, the inhibitory G protein (G_i) considering G_{ia} residue C3 and G_g residue C65 as palmitoylated and geranylgeranylated respectively and hence inserted in the membrane. Both classic (unsupervised) and mwSuMD simulations were performed on this system (Video S6, Figure 4f). In about 50 ns of mwSuMD, the G_{ia} subunit engaged its intracellular binding site on A_1R and formed a complex in close agreement with the cryo-EM structure (PDB 6D9H, RMSD » 5 Å). The membrane anchoring affected the overall G_i binding and the final complex, which was rotated compared to the experimental structure due to the lipidation of G_{ia} and G_g (Figure 4g). This suggests that future, more comprehensive studies of G protein binding and activation should consider several G protein orientations around the receptor as the starting points for mwSuMD simulations, to evaluate as many binding paths as possible. For comparison, 1 ms of cMD did not produce a productive engagement as the G_{ia} remained at RMSD values > 40 Å, suggesting the effectiveness of mwSuMD in sampling G protein binding rare events without the input of energy.

The heterodimerization between A_{2A} and D_2R, and binding simulations of the heterobivalent ligand compound 26.

The current structural model of the A_{2A}R:D_2R heterodimer is that TM4 and TM5 from both the two receptors contribute to form the primary interface of the dimer, although the involvement of TM7 is not ruled out (55). Following this interaction model, we first dynamically docked A_{2A}R and D_2R in an explicit 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) membrane model, then simulated the binding of the heterobivalent compound 26 (37) (CP26) to the preformed A_{2A}R:D_2R heterodimer (Video S7). Since membrane proteins are characterized by slow lateral diffusion (56), we favored the encounter between A_{2A}R and D_2R by input energy as metadynamics and adiabatic MD, during mwSuMD (hybrid metadynamics/aMD/mwSuMD), followed by 1.5 ms of classic MD (cMD) to relax the system and check the stability of the A_{2A}R:D_2R interactions.

During the first 200 ns of simulation with energy bias (Figure 5a,c and Figure S7a), A_{2A}R and D_2R rapidly moved close to each other and reached a distance of about 30 Å (computed between centroids), before stabilizing at around 40 Å (Figure 5a). The computed

molecular mechanics combined with the Poisson–Boltzmann and surface area continuum solvation (MM-PBSA) binding energy suggested two energy minima (Figure 5c) at these distances. The successive cMD simulation did not produce remarkable changes in the distance between receptors (Figure 4b), although the energy fluctuated before reaching about -10 kcal/mol, at the end of the simulation (Figure 5d). The sharp energy minima after 25 and 150 ns were due to the high number of direct contacts between $A_{2A}R$ and D_2R, favored by the energy added to the system. When the input of energy bias was stopped (Figure 5b,d) the POPC residues re-equilibrated at the interface between proteins and mediated intracellular polar interactions between $R150^{4.40\ D2R}$, $Y146^{4.36\ D2}$ and $R199^{5.60\ A2A}$, $Y103^{3.51\ A2A}$ as well as extracellular polar interactions between the top of $TM4^{D2}$, $TM5^{D2}$ and $TM5^{A2A}$, $TM6^{A2A}$ (Figure 5f), suggesting that the $A_{2A}R$:D_2R heterodimerization relies on lipids to mediate short-range interactions between receptors.

The dynamic docking of the herobivalent ligand C26 further stabilized the $A_{2A}R$:D_2R dimer (Figure 5e), in line with experimental data (37). C26 reached the bound state rapidly inserting the agonist pharmacophore within the D_2R orthosteric site (Video S7), while the pyrazole-triazole-pyrimidine scaffold remained in metastable complex with $A_{2A}R$, before completely binding the orthosteric site at the end of the simulation (Video S7). In the final state, the long linker between pharmacophores extended over the top of the interface formed by $A_{2A}R$ and D_2R at the level of the receptors' ECL2 (Figure 5g). A network of polar interactions between POPC, $Y179^{A2A}$, and $Y192^{D2}$ contributed to stabilizing this ternary complex. Interestingly, the latter residues were pinpointed as important for $A_{2A}R$:D_2R interactions (55). From a binding energy perspective, C26 reached the most stable configurations between 80 and 100 ns, before the pyrazole-triazole-pyrimidine component of the ligand completed the binding to $A_{2A}R$. This suggests some contribution of the linker to the overall stability of the ternary complex with $A_{2A}R$ and D_2R. Two out of four mwSuMD replicas produced $A_{2A}R$:D_2R:C26 ternary complexes with C26 engaged both by the orthosteric site of $A_{2A}R$ and D_2R, while in the remaining two replicas the A_2AR pharmacophore remained stacked on the extracellular vestibule of the receptor, although in the proximity of the binding site.

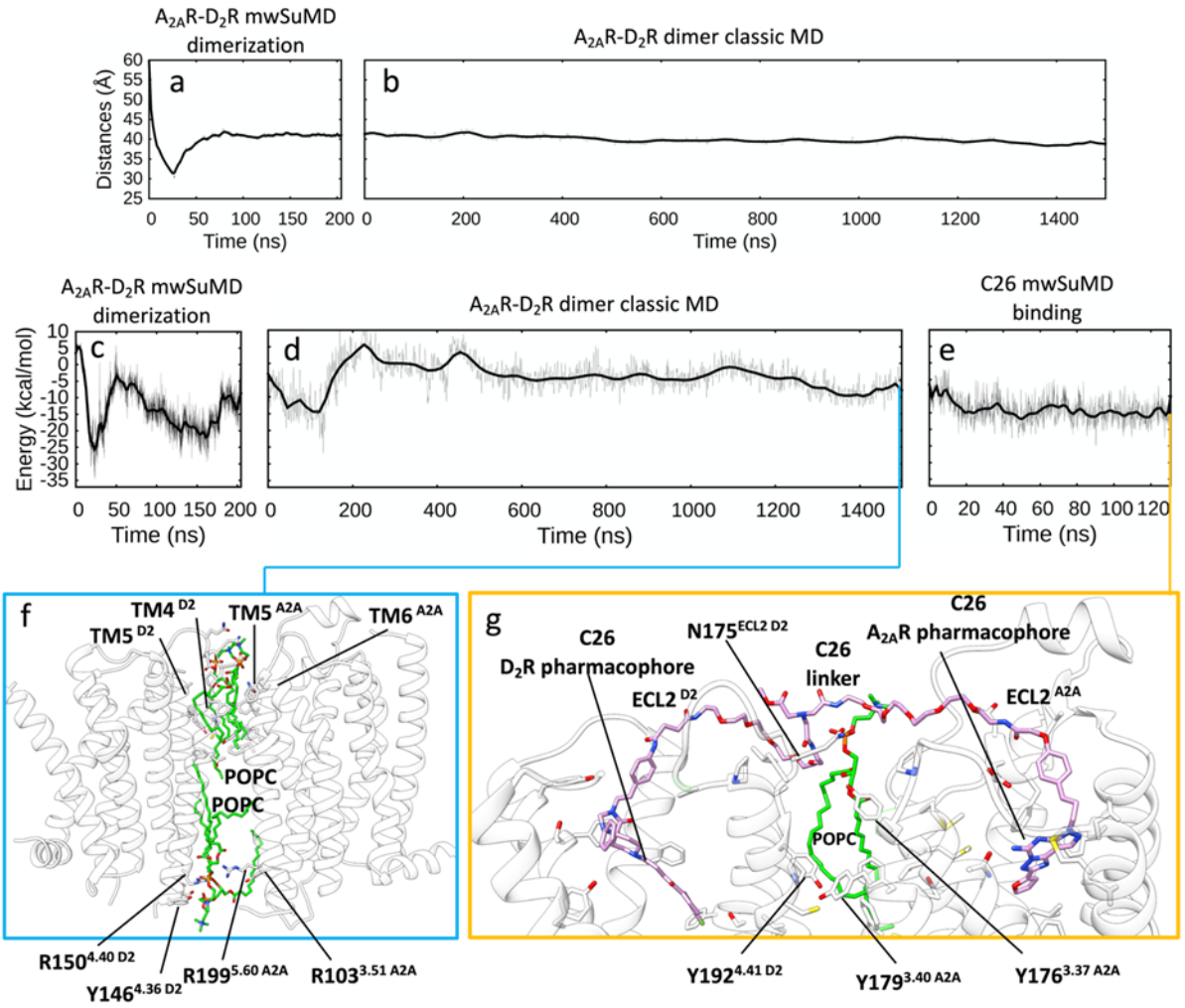


Figure 5. $A_{2A}R$:D_2R heterodimerization and formation of the ternary complex with C26. **a)** Distance between the centroids of $A_{2A}R$ and D_2R during the hybrid metadynamics/aMD/mwSuMD simulation; **b**) distance between the centroids of A2AR and D2R during the successive cMD simulation; **c**) MM-PBSA binding energy between A2AR and D2R during the hybrid metadynamics/aMD/mwSuMD simulation; **d**) MM-PBSA binding energy between A2AR and D2R during the successive cMD simulation; **e**) MM-PBSA binding energy between A2AR and D2R during the mwSUMD binding of C26. **f**) $A_{2A}R$:D_2R heterodimer (white ribbon) after 1.5 ms of cMD; POPC residues (green stick) were involved in polar and hydrophobic interactions; **g**) extracellular view of the $A_{2A}R$:D_2R:C26 ternary complex (D_2R TM2 and TM3 removed for clarity).

Conclusion

Classic MD simulations sample the phase space with an efficiency that depends on the energy barrier between neighboring minima. Processes like (un)binding and protein activation require the system to overcome numerous energy barriers, some of which create a bottleneck that slows the transition down to the millisecond, or second, time scale. To overcome some of these limits, we have developed an energetically-unbiased adaptive sampling algorithm, namely multiple walker mwSuMD, which is based on traditional SuMD, while drawing on parallel multiple replica methods (57,58), and tested it on complex structural events characterizing GPCRs.

MwSuMD performed similarly to SuMD for the dynamic docking of AVP to V2R when time windows of 600 ps were employed. Time windows of 100 ps remarkably improved mwSuMD. Usually, dynamic docking is performed to predict the geometry of complexes or sample the binding path of an already known intermolecular complex, or both. The RMSD of AVP to the experimental coordinates as the supervised metric produced the best results. Consequently, the RMSD should be the metric of choice to study the binding path of well-known intermolecular complexes. The distance, on the other hand, is necessary when limited structural information about the binding mode is available. In the absence of structural information regarding the final bound state, it is possible to sample numerous binding events employing mwSuMD and evaluate the final bound states rank by applying end-point free energy binding methods like the molecular mechanics energies combined with the Poisson–Boltzmann or generalized Born and surface area continuum solvation (MM/PBSA and MM/GBSA) models. Our simulations suggested a remarkable predictivity of distance-driven mwSuMD, as demonstrated by the lowest deviation from the experimental AVP:V2R complex. Remarkably, the dissociation of AVP from V2R was simulated much more rapidly by mwSuMD than by SuMD, suggesting it is an efficient tool for studying the dissociation of ligands from GPCRs.

We increased the complexity of binding simulations by considering GLP-1R and the non-peptide agonist PF06882961. Using mwSuMD, we obtained a binding of the ligand in good agreement with the cryo-EM structure, followed by an active-like conformational transition of GLP-1R. The choice of the metrics supervised was driven by structural data available (45) and extensive preparatory MD simulations, however, alternative binding routes are possible from either the bulk solvent or the membrane (40,59,60). Future studies on GLP-1R and other class B1 GPCR should consider different starting points for the ligand and alternative apo receptor conformations to improve the sampling.

MwSuMD was further tested on the G_s and G_i binding to b_2 AR and A_1R, respectively. MwSuMD produced G protein:GPCR complexes in remarkable agreement with experimental

structural data without the input of energy in a few hundred nanoseconds when starting from inactive G_s and the intermediate active b_2 AR, or a few tens of nanoseconds when considering the active-state A_1R and G_i was anchored to the plasma membrane through the palmitoylation and the geranylgeranylation of Gag (53,54,61).

The final case study was the dimerization process between $A_{2A}R$ and D_2R in a membrane model. To speed up the encounter between receptors, we introduced an energy bias in the form of abMD and MetaD. Although mwSuMD is an unbiased adaptive sampling method, it can be easily coupled to many forms of bias to favor the simulation of energy-requiring processes. Our results suggest a fundamental contribution of the phospholipids on the stabilization of the heterodimer, in agreement with experiments (62,63) and in disagreement with X-ray or protein-protein molecular docking results frequently predicting extended interfaces between monomers (64). MwSuMD was able to dynamically dock the heterobivalent ligand CP26, supporting a stabilizing effect on the $A_{2A}R$:D_2R heterodimer. A complete characterization of the possible interfaces between GPCR monomers, which falls beyond the goal of the present work, should be achieved by preparing different initial unbound states characterized by divergent relative orientations between monomers to dynamically dock in an explicit membrane.

In summary, we showcased the extended applicability domain of mwSuMD to key aspects of GPCRs structural biology. However, given the generality and simplicity of its implementation, we anticipate that mwSuMD can be employed to study a wide range of phenomena characterizing membrane and cytosolic proteins.

Methods

Force field, ligands parameters, and general systems preparation

The CHARMM36 (65,66)/CGenFF 3.0.1 (67–69) force field combination was employed in this work. Initial ligand force field, topology and parameter files were obtained from the ParamChem webserver (67). Restrained electrostatic potential (RESP) (70) partial charges were assigned to all the non-peptidic small molecules but adrenaline and guanosine-5'-diphosphate (GDP) using Gaussian09 (HF/6-31G* level of theory) and AmberTools20.

Six systems were prepared for MD . Hydrogen atoms were added using the pdb2pqr (71) and propka (72) software (considering a simulated pH of 7.0); the protonation of titratable side chains was checked by visual inspection. The resulting receptors were separately inserted in a 1-palmitoyl-2-oleyl-sn-glycerol-3-phosphocholine (POPC) bilayer (previously built by using the VMD Membrane Builder plugin 1.1, Membrane Plugin, Version 1.1. at:

http://www.ks.uiuc.edu/Research/vmd/plugins/membrane/), through an insertion method (73). Receptor orientation was obtained by superposing the coordinates on the corresponding structure retrieved from the OPM database (74). Lipids overlapping the receptor transmembrane helical bundle were removed and TIP3P water molecules (75) were added to the simulation box by means of the VMD Solvate plugin 1.5 (Solvate Plugin, Version 1.5. at <http://www.ks.uiuc.edu/Research/vmd/plugins/solvate/). Finally, overall charge neutrality was reached by adding Na^+/Cl^- counter ions up to the final concentration of 0.150 M), using the VMD Autoionize plugin 1.3 (Autoionize Plugin, Version 1.3. at <http://www.ks.uiuc.edu/Research/vmd/plugins/autoionize/).

System equilibration and general MD settings

The MD engine ACEMD 3 (76) was employed for both the equilibration and productive simulations. The equilibration was achieved in isothermal-isobaric conditions (NPT) using the Berendsen barostat (77) (target pressure 1 atm) and the Langevin thermostat (78) (target temperature 300 K) with low damping of 1 ps^{-1}. For the equilibration (integration time step of 2 fs): first, clashes between protein and lipid atoms were reduced through 1500 conjugate-gradient minimization steps, then a positional constraint of 1 kcal mol^{-1} $Å^{-2}$ on all heavy atoms was gradually released over different time windows: 2 ns for lipid phosphorus atoms, 60 ns for protein atoms other than alpha carbon atoms, 80 ns for alpha carbon atoms; a further 20 ns of equilibration was performed without any positional constraints.

Productive trajectories were computed with an integration time step of 4 fs in the canonical ensemble (NVT). The target temperature was set at 300 K, using a thermostat damping of 0.1 ps^{-1}; the M-SHAKE algorithm (79,80) was employed to constrain the bond lengths involving hydrogen atoms. The cut-off distance for electrostatic interactions was set at 9 Å, with a switching function applied beyond 7.5 Å. Long-range Coulomb interactions were handled using the particle mesh Ewald summation method (PME) (81) by setting the mesh spacing to 1.0 Å.

Vasopressin binding simulations

The vasopressin 2 receptor (V_2R) in complex with vasopressin (AVP) and the G_s protein (82) was retrieved from the Protein Data Bank (83) (PDB ID 7DW9). The G_s was removed from

the system and the missing residues on ECL2 (G185-G189) were modeled from scratch using Modeller 9.19 (84). AVP was placed away from V_2R in the extracellular bulk and the resulting system was prepared for MD simulations and equilibrated as reported above.

During SuMD simulations, the distance between the centroids of AVP residues C1-Q4 and V_2R residues Q96, Q174, Q291, and L312 (Ca atoms only) was supervised over time windows of 600 ps or 100 ps . MwSuMD simulations considered the same distance, the RMSD of AVP residues C1-Q4 to the experimental bound complex or the combination of the two during time windows of 600 ps (3 walkers) or 100 ps (10 walkers) . Slope, SMscore, or DMscore (see Methods section **MwSuMD protocol**) was used in the different mwSuMD replicas performed . Simulations were stopped after 300 ns (time window duration = 600 ps) or 50 ns (time window duration = 100 ps) of total SuMD or mwSuMD simulation time.

Vasopressin unbinding simulations

The V_2R:AVP complex was prepared for MD simulations and equilibrated as reported above. During both SuMD and mwSuMD simulations , the distance between the centroids of AVP residues C1-Q4 and V_2R residues Q96, Q174, Q291, and L312 (Ca atoms only) was supervised over time windows of 100 ps (10 walkers seeded for mwSuMD simulations). Replicas were stopped when the AVP-V_2R distance reached 40 Å.

GLP-1R:PF06882961 binding simulations

The inactive, ligand-free glucagon-like peptide receptor (GLP-1R) was retrieved from the Protein Data Bank (83) (PDB ID 6LN2) (85). Missing residues in the stalk and ICL2 were modeled with Modeller 9.29. The PF06882961 initial conformation was extracted from the complex with the fully active GLP-1R (86) (PDB ID 7LCJ) and placed away from GLP-1R in the extracellular bulk. The resulting system was prepared for MD simulations and equilibrated as reported above. CGenFF dihedral force field parameters of PF06882961 with the highest penalties (dihedrals NG2R51-CG321-CG3C41-CG3C41 (penalty=143.5) and NG2R51-CG321-CG3C41-OG3C51 (penalty=152.4)) were optimised employing Gaussian09 (geometric optimization and dihedral scan at HF/6-31g (d) level of theory) and the VMD force field toolkit plugin (87).

Four classic MD replicas, for a total of 8 ms, were performed on the inactive, ligand-free receptor (prepared for MD simulations and equilibrated as reported above) to assess the possible binding path to the receptor TMD and therefore decide the initial position of

PF06882961 in the extracellular bulk of the simulation box. A visual inspection of the trajectories suggested three major conformational changes that could allow ligand access to the TMD. Transitory openings of the ECD (distance $Q47^{ECD}$ - $S310^{ECL2}$), TM6-TM7 (distance $H363^{6.52}$ - $F390^{7.45}$), and TM1-ECL1 (distance $E138^{1.33}$ and $W214^{ECL1}$) were observed. Since the opening of TM1-ECL1 was observed in two replicas out of four, we placed the ligand in a favorable position for crossing that region of GLP-1R.

MwSuMD simulations were performed stepwise to dock the ligand within GLP-1R first and then relax the receptor towards the active state. The PF06882961 binding was obtained by supervising at the same time the distance between the ligand and GLP-1R TM7 residues L379-F381, which are part of the orthosteric site (Ca atoms only), and the RMSD of the ECD (residues W33-W120, Ca atoms only) to the active state (PDB ID 7LCJ) until the former distance reached 4 Å. In the second phase of mwSuMD, the RMSD of the ECD (residues W33-W120, Ca atoms only) and the ECL1 to the active state (PDB ID 7LCJ) Ca atoms of residues M204-L224) were supervised until the latter reached less than 4 Å. During the third phase, the RMSD of PF06882961, as well as the RMSD of ECL3 (residues A368-T378, Ca atoms), were supervised until the former reached values lower than 3 Å. In the last mwSuMD step, only the RMSD of TM6 (residues I345-F367, Ca atoms) to the active state (PDB ID 7LCJ) was supervised until less than 5 Å.

Membrane-anchored Gi protein:A_1R simulations

Since the full-length structure of the inactive human Gi protein has not been yet resolved by X-ray or cryo-EM, it was modeled by superimposing the AlphaFold2 (88) models of the Gai (P63096-F1), Gb (Q9HAV0-F1), and Gg (P50151-F1) subunits to the PDB file 6EG8 (a Gs heterotrimer). The resulting homotrimer (without GDP) was processed through Charmm-GUI (89) to palmitoylate residue $C3^{Gai}$ and geranylgeranylate residue $C65^{Gg}$ (53,90). The side chains of these two lipidated residues were manually inserted into a 120 x 120 Å POPC membrane and the resulting system was (previously built by using the VMD Membrane Builder plugin 1.1, Membrane Plugin, Version 1.1. at: http://www.ks.uiuc.edu/Research/vmd/plugins/membrane/). Lipids overlapping the palmitoyl and geranylgeranyl groups were removed and TIP3P water molecules (75) were added to the simulation box by means of the VMD Solvate plugin 1.5 (Solvate Plugin, Version 1.5. at <http://www.ks.uiuc.edu/Research/vmd/plugins/solvate/). Finally, overall charge neutrality was reached by adding Na^+/Cl^- counter ions up to the final concentration of 0.150 M), using the VMD Autoionize plugin 1.3 (Autoionize Plugin, Version 1.3. at <http://www.ks.uiuc.edu/Research/vmd/plugins/autoionize/). The first stage of equilibration was performed as reported above (Methods section **System equilibration and general MD**

settings) for 120 ns, followed by a second stage in the NVT ensemble for a further 1 ms without any restraints to allow the membrane-anchored heterotrimeric Gi protein to stabilize within the intracellular side of the simulation box. After this two-stage, long equilibration, the active state A_1R in complex with adenosine (PDB 6D9H) was manually inserted into the equilibrated membrane above the Gi protein using the corresponding structure retrieved from the OPM database as a reference, and the system further equilibrated for 120 ns as reported above (Methods section **System equilibration and general MD settings**). The A_1R-Gi system was then subjected to both a 1 ms-long classic MD simulation and a mwSuMD simulation . During the mwSuMD simulation, the RMSD of helix 5 (H5) G_as residues 329-354 to the PDB 6D9H was supervised, seeding three walkers of 100 ps each until the productive simulation time reached 50 ns (total simulation time 150 ns).

A2A:D2R heterodimerization

The inactive state $A_{2A}R$ and D_2R were retrieved from the Protein Data Bank (83) (PDB ID 5NM4 and 6LUQ, respectively) (91,92). Antagonists bound to the orthosteric site were removed and no modeling of the missing IC loops was attempted. $A_{2A}R$ and D_2R were manually placed roughly 40 Å away from each other, on the plane of the membrane, orienting the two receptors to favor the dimerization through the interface formed by TM5 and TM6, as suggested by Borroto-Esquela D. O. *et al*. (55) The resulting system was prepared for MD simulations and equilibrated as reported above.

The heterodimerization between $A_{2A}R$ and D_2R was simulated with mwSuMD, seeding batches of three walkers with a duration of 100 ps each **.** During each walker, the distance between TM5 of $A_{2A}R$ and D_2R was supervised. At the same time, the distance between the centroids of $A_{2A}R$ and D_2R was used as a collective variable for adiabatic MD (93) (abMD) and well-tempered metadynamics (94,95) (wtMetaD) performed with Plumed 2.6 (96). For abMD, a distance target of 30 Å and a force constant of 10000 $kJ*mol^{-1}*Å^{-1}$) was used, while mwMetaD was performed by seeding gaussian functions every 1 ps (sigma=1 Å; height=0.837 kJ/mol; T=310K) with a bias factor of 30. When the $A_{2A}R$ - D_2R distance reached values lower than 40 Å and the first contacts between proteins were formed, the abMD was stopped and wtMetaD continued with an harmonic energy wall at 30 Å to avoid artificial crushing between the receptors due to the added energy bias. When the distance between $A_{2A}R$ and D_2R was stable at about 30 Å, the collective variable biased by wtMetaD was set as the number of atomic contacts between $A_{2A}R$ and D_2R, until reaching 200 ns of simulation. Finally, to relax the system and challenge the stability of the heterodimer formed during the biased mwSuMD simulation, a 1.5 ms classic MD simulation was performed.

A_{2A}R-D_2R heterobitopic ligand binding simulations

The A_2AR-D_2R heterobivalent ligand compound 26 (37) was parameterized as reported above and placed in the bulk solvent of the A_2AR:D_2R complex from the classic MD. Four mwSuMD replicas were collected supervising at the same time the distance between the A_{2A} antagonist pyrazole-triazole-pyrimidine scaffold and the centroid of A_{2A}R residues F168, N253, and A277 (Ca atoms) as well as the distance between the D_2 antagonist 4-fluorobenzyl scaffold and the centroids of the Ca of D_2R residues C118, F198, and V115 (Ca atoms). Ten walkers of 100 ps were simulated for every mwSuMD batch of replicas.

Gs protein:b_2 AR binding simulations

The model of the adrenergic b_2 receptor (b_2 AR) in an intermediate active state was downloaded from GPCRdb (https://gpcrdb.org/). The full agonist adrenaline (ALE) was inserted in the orthosteric site by superposition with the PDB ID 4LDO (fully-active b_2 AR) (97). The structure of the inactive, GDP bound G_s protein (98) was retrieved from the Protein Data Bank (83) (PDB ID 6EG8) and placed in the intracellular bulk. The resulting system (G_s > 50 Å away from (b_2 AR) was prepared for MD simulations and equilibrated as reported above. The PDB ID 3SN6 (fully-active b_2 AR in complex with G_s (50)) was used as the reference for RMSD computations. Three mwSuMD replicas were performed supervising at the same time the distance between the helix 5 (H5) G_as residues R385-L395 and the b_2 AR residues V31-P330 as well as the RMSD of b_2 AR TM6 residues C265-I278 (Ca atoms only) to the fully active state, during 100 ps time windows (5 walkers).

Multiple walker SuMD (mwSuMD) protocol

The supervised MD (SuMD) is an adaptive sampling method (99) for speeding up the simulation of binding events between small molecules (or peptides (100,101)) and proteins (1,19) without the introduction of any energetic bias. Briefly, during the SuMD a series of short unbiased MD simulations are performed, and after each simulation, the distances between the centers of mass (or the geometrical centers) of the ligand and the predicted binding site (collected at regular time intervals) are fitted to a linear function. If the resulting

slope is negative (showing progress towards the target) the next simulation step starts from the last set of coordinates and velocities produced, otherwise, the simulation is restarted by randomly assigning the atomic velocities.

In the implementation for AceMD, mwSuMD needs as input the initial coordinates of the system as a pdb file, the coordinates, and the atomic velocities of the system from the equilibration stage, the topology file of the system, and all the necessary force filed parameters. The user can decide to supervise one (X) or two metrics (X', X'') of the simulated system over short simulations seeded in batches, called walkers. In the former case, either the slope of the linear function interpolating the metric values or a score can be adopted to decide whether to continue the mwSuMD simulation. When the user decides to supervise two metrics, then a specific score is used. In the present work, distances between centroids, RMSDs, or the number of atomic contacts between two selections were supervised . The choice of the metrics is system and problem dependent, as the RMSD should be most useful when the final state is known, while the distance is required when the target state is unknown; details on the scores are given below. The decision to restart or continue mwSuMD after any short simulation is postponed until all the walkers of a batch are collected. The best short simulation is selected and extended by seeding the same number of walkers, with the same duration as the step before.

For each walker, the score for the supervision of a single metric (SMscore) is computed as the square root of the product between the metric value in the last frame ($X_{last\ frame}$) and the average metric value over the short simulation ($\overline{X}$):

$$SMscore = \sqrt{X_{last\ frame} * \overline{X}} \quad (1)$$

If the metric is set to decrease (e.g. binding or dimerization) the walker with the lowest SMscore is continued, otherwise (e.g. unbinding or outwards opening of domains), it is the walker with the highest score to be extended. Using the SMscore rather than the slope should give more weight to the final state of each short simulation, as it is the starting point for the successive batch of simulations. Considering the average of the metric should favor short simulations consistently evolving in the desired direction along the metric.

If both X' and X'' are set to increase during the mwSuMD simulations, the score for the supervision of two metrics (DMscore) on each walker is computed as follows:

$$DMscore = \left(\left(\frac{X'_{last\ frame}}{\bar{X}'_{batch\ walkers}} - 1 \right) + \left(\frac{X''_{last\ frame}}{\bar{X}''_{batch\ walkers}} - 1 \right) \right) * 100 \quad (2)$$

Where $X'_{last\ frame}$ and $X''_{last\ frame}$ are the metrics values in the last frame, while $\bar{X}'_{batch\ walkers}$ and $\bar{X}''_{batch\ walkers}$ represent the average value of the two metrics over all the walkers in the batch. Subtracting the value 1 to the metric ratio ensures that if one of the two metrics from the last frame ($X'_{last\ frame}$ or $X''_{last\ frame}$) is equal to the average ($\bar{X}''_{batch\ walkers}$ or $\bar{X}''_{batch\ walkers}$) then that metric addend is null and DMscore depends only on the remaining metric. If any of the two metrics is set to decrease, then the corresponding component in Equation 2 is multiplied by -1 to maintain a positive score. Considering the average value of the two metrics over all the walkers rather than only over the considered walker should be more representative of the system evolution along the defined metric. In other words, the information about the metric is taken from all the walkers to better describe the evolution of the system.

The DMScore is designed to preserve some degree of independence between the two metrics supervised. Indeed, if the variation of one of them slows down and gets close to zero, the other metric is still able to drive the system's evolution. It should be noted that DMScore works at its best if the two metrics have similar variations over time, as it is in the case of distance and RMSD (both of which are distance-based). Notably, when a walker is extended by seeding a new batch of short simulations and the remaining walkers are stopped, the atomic velocities are not reassigned. This allows the simulations to be as short as a few picoseconds if desired without introducing artifacts due to the thermostat latency to reach the target temperature (usually up to 10-20 ps when a simulation is restarted reassigning the velocities of the atoms).

The current implementation of mwSuMD is for python3 and exploits MDAnalysis (102) and MDTRaj (103) modules.

MD Analysis

Interatomic distances were computed through MDAnalysis (102); root mean square deviations (RMSD) were computed using VMD (104) and MDAnalysis (102).

Interatomic contacts and ligand-protein hydrogen bonds were detected using the GetContacts scripts tool (https://getcontacts.github.io), setting a hydrogen bond

donor-acceptor distance of 3.3 Å and an angle value of 120° as geometrical cut-offs. Contacts and hydrogen bond persistency are quantified as the percentage of frames (over all the frames obtained by merging the different replicas) in which protein residues formed contacts or hydrogen bonds with the ligand.

The MMPBSA.py (105) script, from the AmberTools20 suite (The Amber Molecular Dynamics Package, at http://ambermd.org/), was used to compute molecular mechanics energies combined with the generalized Born and surface area continuum solvation (MM/GBSA) method or the molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) approach, after transforming the CHARMM psf topology files to an Amber prmtop format using ParmEd (documentation at <http://parmed.github.io/ParmEd/html/index.html).

Supplementary Videos were produced employing VMD and avconv (at https://libav.org/avconv.html). Molecular graphics images were produced using the UCSF Chimera (106) (v1.14).

Numbering system

Throughout the manuscript, the Ballesteros-Weinstein residues numbering system for class A GPCRs (107) and the Wootten residues numbering system for class B GPCRs (108) are adopted.

Videos:

https://www.biorxiv.org/content/10.1101/2022.10.26.513870v2.article-info

References

1. Cuzzolin, A.; Sturlese, M.; Deganutti, G.; Salmaso, V.; Sabbadin, D.; Ciancetta, A.; Moro, S. Deciphering the Complexity of Ligand-Protein Recognition Pathways Using Supervised Molecular Dynamics (SuMD) Simulations. *J. Chem. Inf. Model.* **2016**, *56*, 687–705.
2. Deganutti, G.; Moro, S.; Reynolds, C. A. A Supervised Molecular Dynamics Approach to Unbiased Ligand-Protein Unbinding. *J. Chem. Inf. Model.* **2020**, *60*, 1804–1817.
3. de Mendoza, A.; Sebé-Pedrós, A.; Ruiz-Trillo, I. The Evolution of the GPCR Signaling System in Eukaryotes: Modularity, Conservation, and the Transition to Metazoan Multicellularity. *Genome Biol. Evol.* **2014**, *6*, 606–619.
4. Sriram, K.; Insel, P. A. G Protein-Coupled Receptors as Targets for Approved Drugs: How Many Targets and How Many Drugs? *Mol. Pharmacol.* **2018**, *93*, 251–258.
5. Schiöth, H. B.; Fredriksson, R. The GRAFS Classification System of G-Protein Coupled Receptors in Comparative Perspective. *Gen. Comp. Endocrinol.* **2005**, *142*, 94–101.
6. G protein-coupled receptors | G protein-coupled receptors | IUPHAR/BPS Guide to PHARMACOLOGY https://www.guidetopharmacology.org/GRAC/FamilyDisplayForward?familyId=694 (accessed May 18, 2022).
7. Syrovatkina, V.; Alegre, K. O.; Dey, R.; Huang, X.-Y. Regulation, Signaling, and Physiological Functions of G-Proteins. *J. Mol. Biol.* **2016**, *428*, 3850–3868.
8. Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A.; Le Trong, I.; Teller, D. C.; Okada, T.; Stenkamp, R. E.; Yamamoto, M.; Miyano, M. Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor. *Science* **2000**, *289*, 739–745.
9. Cherezov, V.; Rosenbaum, D. M.; Hanson, M. A.; Rasmussen, S. G. F.; Thian, F. S.; Kobilka, T. S.; Choi, H.-J.; Kuhn, P.; Weis, W. I.; Kobilka, B. K.; Stevens, R. C. High-Resolution Crystal Structure of an Engineered Human Beta2-Adrenergic G Protein-Coupled Receptor. *Science* **2007**, *318*, 1258–1265.
10. Tan, L.; Yan, W.; McCorvy, J. D.; Cheng, J. Biased Ligands of G Protein-Coupled Receptors (GPCRs): Structure-Functional Selectivity Relationships (SFSRs) and Therapeutic Potential. *J. Med. Chem.* **2018**, *61*, 9841–9878.
11. Hollingsworth, S. A.; Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99*, 1129–1143.
12. Durrant, J. D.; McCammon, J. A. Molecular Dynamics Simulations and Drug Discovery. *BMC Biol.* **2011**, *9*, 71.
13. Bussi, G.; Laio, A. Using Metadynamics to Explore Complex Free-Energy

Landscapes. *Nat. Rev. Phys.* **2020**.

14. Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919–11929.
15. Miao, Y.; Feher, V. A.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theory Comput.* **2015**, *11*, 3584–3595.
16. Mattedi, G.; Acosta-Gutiérrez, S.; Clark, T.; Gervasio, F. L. A Combined Activation Mechanism for the Glucagon Receptor. *Proc. Natl. Acad. Sci. USA* **2020**.
17. Miao, Y.; McCammon, J. A. Mechanism of the G-Protein Mimetic Nanobody Binding to a Muscarinic G-Protein-Coupled Receptor. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 3036–3041.
18. Zuckerman, D. M.; Chong, L. T. Weighted Ensemble Simulation: Review of Methodology, Applications, and Software. *Annu. Rev. Biophys.* **2017**, *46*, 43–57.
19. Sabbadin, D.; Moro, S. Supervised Molecular Dynamics (SuMD) as a Helpful Tool to Depict GPCR-Ligand Recognition Pathway in a Nanosecond Time Scale. *J. Chem. Inf. Model.* **2014**, *54*, 372–376.
20. Deganutti, G.; Atanasio, S.; Rujan, R.-M.; Sexton, P. M.; Wootten, D.; Reynolds, C. A. Exploring Ligand Binding to Calcitonin Gene-Related Peptide Receptors. *Front. Mol. Biosci.* **2021**, *8*, 720561.
21. Dong, M.; Deganutti, G.; Piper, S. J.; Liang, Y.-L.; Khoshouei, M.; Belousoff, M. J.; Harikumar, K. G.; Reynolds, C. A.; Glukhova, A.; Furness, S. G. B.; Christopoulos, A.; Danev, R.; Wootten, D.; Sexton, P. M.; Miller, L. J. Structure and Dynamics of the Active Gs-Coupled Human Secretin Receptor. *Nat. Commun.* **2020**, *11*, 4137.
22. Deganutti, G.; Prischi, F.; Reynolds, C. A. Supervised Molecular Dynamics for Exploring the Druggability of the SARS-CoV-2 Spike Protein. *J. Comput. Aided Mol. Des.* **2021**, *35*, 195–207.
23. Cary, B. P.; Deganutti, G.; Zhao, P.; Truong, T. T.; Piper, S. J.; Liu, X.; Belousoff, M. J.; Danev, R.; Sexton, P. M.; Wootten, D.; Gellman, S. H. Structural and Functional Diversity among Agonist-Bound States of the GLP-1 Receptor. *Nat. Chem. Biol.* **2021**.
24. Birnbaumer, M. Vasopressin Receptors. *Trends Endocrinol. Metab.* **2000**, *11*, 406–410.
25. Ball, S. G. Vasopressin and Disorders of Water Balance: The Physiology and Pathophysiology of Vasopressin. *Ann Clin Biochem* **2007**, *44*, 417–431.
26. Guo, D.; Heitman, L. H.; IJzerman, A. P. Kinetic Aspects of the Interaction between Ligand and G Protein-Coupled Receptor: The Case of the Adenosine Receptors.

Chem. Rev. **2017**, *117*, 38–66.

27. Guo, D.; Heitman, L. H.; IJzerman, A. P. The Role of Target Binding Kinetics in Drug Discovery. *ChemMedChem* **2015**, *10*, 1793–1796.
28. Flock, T.; Hauser, A. S.; Lund, N.; Gloriam, D. E.; Balaji, S.; Babu, M. M. Selectivity Determinants of GPCR-G-Protein Binding. *Nature* **2017**, *545*, 317–322.
29. Wall, M. J.; Hill, E.; Huckstepp, R.; Barkan, K.; Deganutti, G.; Leuenberger, M.; Preti, B.; Winfield, I.; Carvalho, S.; Suchankova, A.; Wei, H.; Safitri, D.; Huang, X.; Imlach, W.; La Mache, C.; Dean, E.; Hume, C.; Hayward, S.; Oliver, J.; Zhao, F.-Y.; Spanswick, D.; Reynolds, C. A.; Lochner, M.; Ladds, G.; Frenguelli, B. G. Selective Activation of Gαob by an Adenosine A1 Receptor Agonist Elicits Analgesia without Cardiorespiratory Depression. *Nat. Commun.* **2022**, *13*, 4150.
30. Culhane, K. J.; Gupte, T. M.; Madhugiri, I.; Gadgil, C. J.; Sivaramakrishnan, S. Kinetic Model of GPCR-G Protein Interactions Reveals Allokairic Modulation of Signaling. *Nat. Commun.* **2022**, *13*, 1202.
31. Canals, M.; Marcellino, D.; Fanelli, F.; Ciruela, F.; de Benedetti, P.; Goldberg, S. R.; Neve, K.; Fuxe, K.; Agnati, L. F.; Woods, A. S.; Ferré, S.; Lluis, C.; Bouvier, M.; Franco, R. Adenosine A2A-Dopamine D2 Receptor-Receptor Heteromerization: Qualitative and Quantitative Assessment by Fluorescence and Bioluminescence Energy Transfer. *J. Biol. Chem.* **2003**, *278*, 46741–46749.
32. Fuxe, K.; Guidolin, D.; Agnati, L. F.; Borroto-Escuela, D. O. Dopamine Heteroreceptor Complexes as Therapeutic Targets in Parkinson's Disease. *Expert Opin. Ther. Targets* **2015**, *19*, 377–398.
33. Rivera-Oliver, M.; Díaz-Ríos, M. Using Caffeine and Other Adenosine Receptor Antagonists and Agonists as Therapeutic Tools against Neurodegenerative Diseases: A Review. *Life Sci.* **2014**, *101*, 1–9.
34. Richardson, P. J.; Kase, H.; Jenner, P. G. Adenosine A2A Receptor Antagonists as New Agents for the Treatment of Parkinson's Disease. *Trends Pharmacol. Sci.* **1997**, *18*, 338–344.
35. Fuxe, K.; Borroto-Escuela, D. O. Heteroreceptor Complexes and Their Allosteric Receptor-Receptor Interactions as a Novel Biological Principle for Integration of Communication in the CNS: Targets for Drug Development. *Neuropsychopharmacology* **2016**, *41*, 380–382.
36. Soriano, A.; Ventura, R.; Molero, A.; Hoen, R.; Casadó, V.; Cortés, A.; Fanelli, F.; Albericio, F.; Lluís, C.; Franco, R.; Royo, M. Adenosine A2A Receptor-Antagonist/Dopamine D2 Receptor-Agonist Bivalent Ligands as Pharmacological Tools to Detect A2A-D2 Receptor Heteromers. *J. Med. Chem.* **2009**, *52*, 5590–5602.

37. Pulido, D.; Casadó-Anguera, V.; Gómez-Autet, M.; Llopart, N.; Moreno, E.; Casajuana-Martin, N.; Ferré, S.; Pardo, L.; Casadó, V.; Royo, M. Heterobivalent Ligand for the Adenosine A2A-Dopamine D2 Receptor Heteromer. *J. Med. Chem.* **2022**.
38. Gioia, D.; Bertazzo, M.; Recanatini, M.; Masetti, M.; Cavalli, A. Dynamic Docking: A Paradigm Shift in Computational Drug Discovery. *Molecules* **2017**, *22*.
39. Atanasio, S.; Deganutti, G.; Reynolds, C. A. Addressing Free Fatty Acid Receptor 1 (FFAR1) Activation Using Supervised Molecular Dynamics. *J. Comput. Aided Mol. Des.* **2020**, *34*, 1181–1193.
40. Deganutti, G.; Barkan, K.; Preti, B.; Leuenberger, M.; Wall, M.; Frenguelli, B. G.; Lochner, M.; Ladds, G.; Reynolds, C. A. Deciphering the Agonist Binding Mechanism to the Adenosine A1 Receptor. *ACS Pharmacol. Transl. Sci.* **2021**, *4*, 314–326.
41. Deganutti, G.; Barkan, K.; Ladds, G.; Reynolds, C. A. A Multisite Model of Allosterism for the Adenosine A1 Receptor. *BioRxiv* **2020**.
42. Zhao, P.; Liang, Y.-L.; Belousoff, M. J.; Deganutti, G.; Fletcher, M. M.; Willard, F. S.; Bell, M. G.; Christe, M. E.; Sloop, K. W.; Inoue, A.; Truong, T. T.; Clydesdale, L.; Furness, S. G. B.; Christopoulos, A.; Wang, M.-W.; Miller, L. J.; Reynolds, C. A.; Danev, R.; Sexton, P. M.; Wootten, D. Activation of the GLP-1 Receptor by a Non-Peptidic Agonist. *Nature* **2020**, *577*, 432–436.
43. Kawai, T.; Sun, B.; Yoshino, H.; Feng, D.; Suzuki, Y.; Fukazawa, M.; Nagao, S.; Wainscott, D. B.; Showalter, A. D.; Droz, B. A.; Kobilka, T. S.; Coghlan, M. P.; Willard, F. S.; Kawabe, Y.; Kobilka, B. K.; Sloop, K. W. Structural Basis for GLP-1 Receptor Activation by LY3502970, an Orally Active Nonpeptide Agonist. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 29959–29967.
44. Ma, H.; Huang, W.; Wang, X.; Zhao, L.; Jiang, Y.; Liu, F.; Guo, W.; Sun, X.; Zhong, W.; Yuan, D.; Xu, H. E. Structural Insights into the Activation of GLP-1R by a Small Molecule Agonist. *Cell Res.* **2020**, *30*, 1140–1142.
45. Zhang, X.; Belousoff, M. J.; Zhao, P.; Kooistra, A. J.; Truong, T. T.; Ang, S. Y.; Underwood, C. R.; Egebjerg, T.; Šenel, P.; Stewart, G. D.; Liang, Y.-L.; Glukhova, A.; Venugopal, H.; Christopoulos, A.; Furness, S. G. B.; Miller, L. J.; Reedtz-Runge, S.; Langmead, C. J.; Gloriam, D. E.; Danev, R.; Sexton, P. M.; Wootten, D. Differential GLP-1R Binding and Activation by Peptide and Non-Peptide Agonists. *Mol. Cell* **2020**, *80*, 485–500.e7.
46. Cong, Z.; Chen, L.-N.; Ma, H.; Zhou, Q.; Zou, X.; Ye, C.; Dai, A.; Liu, Q.; Huang, W.; Sun, X.; Wang, X.; Xu, P.; Zhao, L.; Xia, T.; Zhong, W.; Yang, D.; Eric Xu, H.; Zhang, Y.; Wang, M.-W. Molecular Insights into Ago-Allosteric Modulation of the Human Glucagon-like Peptide-1 Receptor. *Nat. Commun.* **2021**, *12*, 3763.

47. Cong, Z.; Zhou, Q.; Li, Y.; Chen, L.-N.; Zhang, Z.-C.; Liang, A.; Liu, Q.; Wu, X.; Dai, A.; Xia, T.; Wu, W.; Zhang, Y.; Yang, D.; Wang, M.-W. Structural Basis of Peptidomimetic Agonism Revealed by Small Molecule GLP-1R Agonists Boc5 and WB4-24. *BioRxiv* **2022**.

48. Cong, Z.; Liang, Y.-L.; Zhou, Q.; Darbalaei, S.; Zhao, F.; Feng, W.; Zhao, L.; Xu, H. E.; Yang, D.; Wang, M.-W. Structural Perspective of Class B1 GPCR Signaling. *Trends Pharmacol. Sci.* **2022**, *43*, 321–334.

49. Hilger, D.; Kumar, K. K.; Hu, H.; Pedersen, M. F.; O'Brien, E. S.; Giehm, L.; Jennings, C.; Eskici, G.; Inoue, A.; Lerch, M.; Mathiesen, J. M.; Skiniotis, G.; Kobilka, B. K. Structural Insights into Differences in G Protein Activation by Family A and Family B GPCRs. *Science* **2020**, *369*.

50. Rasmussen, S. G. F.; DeVree, B. T.; Zou, Y.; Kruse, A. C.; Chung, K. Y.; Kobilka, T. S.; Thian, F. S.; Chae, P. S.; Pardon, E.; Calinski, D.; Mathiesen, J. M.; Shah, S. T. A.; Lyons, J. A.; Caffrey, M.; Gellman, S. H.; Steyaert, J.; Skiniotis, G.; Weis, W. I.; Sunahara, R. K.; Kobilka, B. K. Crystal Structure of the B2 Adrenergic Receptor-Gs Protein Complex. *Nature* **2011**, *477*, 549–555.

51. Liang, Y.-L.; Belousoff, M. J.; Fletcher, M. M.; Zhang, X.; Khoshouei, M.; Deganutti, G.; Koole, C.; Furness, S. G. B.; Miller, L. J.; Hay, D. L.; Christopoulos, A.; Reynolds, C. A.; Danev, R.; Wootten, D.; Sexton, P. M. Structure and Dynamics of Adrenomedullin Receptors AM1 and AM2 Reveal Key Mechanisms in the Control of Receptor Phenotype by Receptor Activity-Modifying Proteins. *ACS Pharmacol. Transl. Sci.* **2020**, *3*, 263–284.

52. Dror, R. O.; Mildorf, T. J.; Hilger, D.; Manglik, A.; Borhani, D. W.; Arlow, D. H.; Philippsen, A.; Villanueva, N.; Yang, Z.; Lerch, M. T.; Hubbell, W. L.; Kobilka, B. K.; Sunahara, R. K.; Shaw, D. E. SIGNAL TRANSDUCTION. Structural Basis for Nucleotide Exchange in Heterotrimeric G Proteins. *Science* **2015**, *348*, 1361–1365.

53. Linder, M. E.; Middleton, P.; Hepler, J. R.; Taussig, R.; Gilman, A. G.; Mumby, S. M. Lipid Modifications of G Proteins: Alpha Subunits Are Palmitoylated. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 3675–3679.

54. Zhang, Z.; Melia, T. J.; He, F.; Yuan, C.; McGough, A.; Schmid, M. F.; Wensel, T. G. How a G Protein Binds a Membrane. *J. Biol. Chem.* **2004**, *279*, 33937–33945.

55. Borroto-Escuela, D. O.; Rodriguez, D.; Romero-Fernandez, W.; Kapla, J.; Jaiteh, M.; Ranganathan, A.; Lazarova, T.; Fuxe, K.; Carlsson, J. Mapping the Interface of a GPCR Dimer: A Structural Model of the A2A Adenosine and D2 Dopamine Receptor Heteromer. *Front. Pharmacol.* **2018**, *9*, 829.

56. Ramadurai, S.; Holt, A.; Krasnikov, V.; van den Bogaart, G.; Killian, J. A.; Poolman, B. Lateral Diffusion of Membrane Proteins. *J. Am. Chem. Soc.* **2009**, *131*, 12650–12656.

57. Dickson, A.; Brooks, C. L. WExplore: Hierarchical Exploration of High-Dimensional Spaces Using the Weighted Ensemble Algorithm. *J. Phys. Chem. B* **2014**, *118*, 3532–3542.
58. Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
59. Stanley, N.; Pardo, L.; Fabritiis, G. D. The Pathway of Ligand Entry from the Membrane Bilayer to a Lipid G Protein-Coupled Receptor. *Sci. Rep.* **2016**, *6*, 22639.
60. Bokoch, M. P.; Jo, H.; Valcourt, J. R.; Srinivasan, Y.; Pan, A. C.; Capponi, S.; Grabe, M.; Dror, R. O.; Shaw, D. E.; DeGrado, W. F.; Coughlin, S. R. Entry from the Lipid Bilayer: A Possible Pathway for Inhibition of a Peptide G Protein-Coupled Receptor by a Lipophilic Small Molecule. *Biochemistry* **2018**.
61. Vögler, O.; Barceló, J. M.; Ribas, C.; Escribá, P. V. Membrane Interactions of G Proteins and Other Related Proteins. *Biochim. Biophys. Acta* **2008**, *1778*, 1640–1652.
62. Goddard, A. D.; Dijkman, P. M.; Adamson, R. J.; Watts, A. Lipid-Dependent GPCR Dimerization. *Methods Cell Biol.* **2013**, *117*, 341–357.
63. Prasanna, X.; Sengupta, D.; Chattopadhyay, A. Cholesterol-Dependent Conformational Plasticity in GPCR Dimers. *Sci. Rep.* **2016**, *6*, 31858.
64. Meng, X.-Y.; Mezei, M.; Cui, M. Computational Approaches for Modeling GPCR Dimerization. *Curr. Pharm. Biotechnol.* **2014**, *15*, 996–1006.
65. Huang, J.; MacKerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **2013**, *34*, 2135–2145.
66. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods* **2017**, *14*, 71–73.
67. Vanommeslaeghe, K.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J. Chem. Inf. Model.* **2012**, *52*, 3144–3154.
68. Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* **2012**, *52*, 3155–3168.
69. Yu, W.; He, X.; Vanommeslaeghe, K.; MacKerell, A. D. Extension of the CHARMM General Force Field to Sulfonyl-Containing Compounds and Its Utility in Biomolecular Simulations. *J. Comput. Chem.* **2012**, *33*, 2451–2468.
70. Woods, R. J.; Chappelle, R. Restrained Electrostatic Potential Atomic Partial Charges for Condensed-Phase Simulations of Carbohydrates. *Theochem* **2000**, *527*, 149–156.

71. Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: An Automated Pipeline for the Setup of Poisson-Boltzmann Electrostatics Calculations. *Nucleic Acids Res.* **2004**, *32*, W665-7.
72. Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical PK Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.
73. Sommer, B. Membrane Packing Problems: A Short Review on Computational Membrane Modeling Methods and Tools. *Comput Struct Biotechnol J* **2013**, *5*, e201302014.
74. Lomize, M. A.; Lomize, A. L.; Pogozheva, I. D.; Mosberg, H. I. OPM: Orientations of Proteins in Membranes Database. *Bioinformatics* **2006**, *22*, 623–625.
75. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926.
76. Harvey, M. J.; Giupponi, G.; Fabritiis, G. D. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *J. Chem. Theory Comput.* **2009**, *5*, 1632–1639.
77. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684.
78. Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin Dynamics of Peptides: The Frictional Dependence of Isomerization Rates of N-Acetylalanyl-N'-Methylamide. *Biopolymers* **1992**, *32*, 523–535.
79. Forester, T. R.; Smith, W. SHAKE, Rattle, and Roll: Efficient Constraint Algorithms for Linked Rigid Bodies. *J. Comput. Chem.* **1998**.
80. Kräutler, V.; van Gunsteren, W. F.; Hünenberger, P. H. A Fast SHAKE Algorithm to Solve Distance Constraint Equations for Small Molecules in Molecular Dynamics Simulations. *J. Comput. Chem.* **2001**, *22*, 501–508.
81. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577.
82. Zhou, F.; Ye, C.; Ma, X.; Yin, W.; Croll, T. I.; Zhou, Q.; He, X.; Zhang, X.; Yang, D.; Wang, P.; Xu, H. E.; Wang, M.-W.; Jiang, Y. Molecular Basis of Ligand Recognition and Activation of Human V2 Vasopressin Receptor. *Cell Res.* **2021**, *31*, 929–931.
83. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
84. Fiser, A.; Sali, A. Modeller: Generation and Refinement of Homology-Based Protein

Structure Models. *Meth. Enzymol.* **2003**, *374*, 461–491.

85. Wu, F.; Yang, L.; Hang, K.; Laursen, M.; Wu, L.; Han, G. W.; Ren, Q.; Roed, N. K.; Lin, G.; Hanson, M. A.; Jiang, H.; Wang, M.-W.; Reedtz-Runge, S.; Song, G.; Stevens, R. C. Full-Length Human GLP-1 Receptor Structure without Orthosteric Ligands. *Nat. Commun.* **2020**, *11*, 1272.
86. Zhang, X.; Johnson, R. M.; Drulyte, I.; Yu, L.; Kotecha, A.; Danev, R.; Wootten, D.; Sexton, P. M.; Belousoff, M. J. Evolving Cryo-EM Structural Approaches for GPCR Drug Discovery. *Structure* **2021**, *29*, 963–974.e6.
87. Mayne, C. G.; Saam, J.; Schulten, K.; Tajkhorshid, E.; Gumbart, J. C. Rapid Parameterization of Small Molecules Using the Force Field Toolkit. *J. Comput. Chem.* **2013**, *34*, 2757–2770.
88. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.
89. Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859–1865.
90. Mystek, P.; Rysiewicz, B.; Gregrowicz, J.; Dziedzicka-Wasylewska, M.; Polit, A. Gγ and Gα Identity Dictate a G-Protein Heterotrimer Plasma Membrane Targeting. *Cells* **2019**, *8*.
91. Weinert, T.; Olieric, N.; Cheng, R.; Brünle, S.; James, D.; Ozerov, D.; Gashi, D.; Vera, L.; Marsh, M.; Jaeger, K.; Dworkowski, F.; Panepucci, E.; Basu, S.; Skopintsev, P.; Doré, A. S.; Geng, T.; Cooke, R. M.; Liang, M.; Prota, A. E.; Panneels, V.; Nogly, P.; Ermler, U.; Schertler, G.; Hennig, M.; Steinmetz, M. O.; Wang, M.; Standfuss, J. Serial Millisecond Crystallography for Routine Room-Temperature Structure Determination at Synchrotrons. *Nat. Commun.* **2017**, *8*, 542.
92. Fan, L.; Tan, L.; Chen, Z.; Qi, J.; Nie, F.; Luo, Z.; Cheng, J.; Wang, S. Haloperidol Bound D2 Dopamine Receptor Structure Inspired the Discovery of Subtype Selective Ligands. *Nat. Commun.* **2020**, *11*, 1074.
93. Marchi, M.; Ballone, P. Adiabatic Bias Molecular Dynamics: A Method to Navigate the Conformational Space of Complex Molecular Systems. *J. Chem. Phys.* **1999**, *110*, 3697.
94. Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12562–12566.

95. Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, 020603.
96. Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput Phys Commun* **2014**, *185*, 604–613.
97. Ring, A. M.; Manglik, A.; Kruse, A. C.; Enos, M. D.; Weis, W. I.; Garcia, K. C.; Kobilka, B. K. Adrenaline-Activated Structure of B2-Adrenoceptor Stabilized by an Engineered Nanobody. *Nature* **2013**, *502*, 575–579.
98. Liu, X.; Xu, X.; Hilger, D.; Aschauer, P.; Tiemann, J. K. S.; Du, Y.; Liu, H.; Hirata, K.; Sun, X.; Guixà-González, R.; Mathiesen, J. M.; Hildebrand, P. W.; Kobilka, B. K. Structural Insights into the Process of GPCR-G Protein Complex Formation. *Cell* **2019**, *177*, 1243–1251.e12.
99. Deganutti, G.; Moro, S. Estimation of Kinetic and Thermodynamic Ligand-Binding Parameters Using Computational Strategies. *Future Med. Chem.* **2017**, *9*, 507–523.
100. Salmaso, V.; Sturlese, M.; Cuzzolin, A.; Moro, S. Exploring Protein-Peptide Recognition Pathways Using a Supervised Molecular Dynamics Approach. *Structure* **2017**, *25*, 655–662.e2.
101. Bower, R. L.; Yule, L.; Rees, T. A.; Deganutti, G.; Hendrikse, E. R.; Harris, P. W. R.; Kowalczyk, R.; Ridgway, Z.; Wong, A. G.; Swierkula, K.; Raleigh, D. P.; Pioszak, A. A.; Brimble, M. A.; Reynolds, C. A.; Walker, C. S.; Hay, D. L. Molecular Signature for Receptor Engagement in the Metabolic Peptide Hormone Amylin. *ACS Pharmacol. Transl. Sci.* **2018**, *1*, 32–49.
102. Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **2011**, *32*, 2319–2327.
103. McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. Mdtraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109*, 1528–1532.
104. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J Mol Graph* **1996**, *14*, 33–38, 27.
105. Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.Py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 3314–3321.
106. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—a Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.
107. Ballesteros, J. A.; Weinstein, H. [19] Integrated Methods for the Construction of

Three-Dimensional Models and Computational Probing of Structure-Function Relations in G Protein-Coupled Receptors. In *Receptor Molecular Biology*; Methods in Neurosciences; Elsevier, 1995; Vol. 25, pp. 366–428.

108. Wootten, D.; Simms, J.; Miller, L. J.; Christopoulos, A.; Sexton, P. M. Polar Transmembrane Interactions Drive Formation of Ligand-Specific and Signal Pathway-Biased Family B G Protein-Coupled Receptor Conformations. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 5211–5216.

9 Concluding Remarks and Future Research Directions

The structural characterisation of both the SARS-CoV-2 RBD and S2 domains is crucial for SBDD. My studies clarified the role of the heparinoids heparin and heparan sulphate, highlighting the importance of the sulfonation in the recognition by positively charged residues (i.e. arginine, lysine, and histidine) on the heparin-binding motifs (HBM) on the RBD. The HBM-HP interactions at the base of RBD provide a rationale behind the stabilising role of fully sulphated HP. On the other hand, I determined multiple sites on the N-terminal domain (NTD) where HS is more likely to interact, compared to HP. My definition of the respective roles of HP and HS possibly clarified the role of heparinoids in SARS-CoV-2:ACE2 molecular recognition. These observations inspired a high-throughput virtual screening (HTVS) using HP's molecular fingerprint to identify molecular candidates that could saturate the HBM on RB to interfere with ACE2 binding or RBD opening.

However, the most promising molecular candidates available from the ZINC database were displaced during longer MD simulations, suggesting docking false positives. The implementation of generative recurrent neural networks (RNN) for the hunting of HP-inspired binders will be addressed as the next step for this pipeline and an RNN is under training at the time of this writing. The conserved region of the SARS-CoV-2 S2 domain (i.e. the stalk of the spike protein) was identified as a promising target for fragment-based drug design. Inspired by the criticalities of mutation-prone antibodies and treatment, I identified a highly conserved region of the stalk against which a set of optimized fragments was tested, revealing the potential and limits of targeting the flexible section of the SARS-CoV-2 S2 domain. Planarity, a molecular size < 200 Da and a limited steric hindrance were the stringent requirements for successful fragments to overcome the distribution of glycans along the stalk and intersect with the flexible chains of the homotrimer. The stalk's intrinsic flexibility, paired with the glycan's sweeping motion restricted the area of the investigation to a portion right below the S1 domain and the flexible loop closer to the viral membrane. The results, although preliminary, suggested that an alternative strategy is required to overcome the structural challenges of targeting the S protein stalk such as the use of a different type of drugs (peptides, aptamers or antibodies) or specifically targeting the glycans that protect that region. The importance of MD to validate the stability of a ligand was highlighted during the different high throughput protocols implemented, to bridge molecular docking to MD and refine the results for more accurate results.

In the last paper presented for the SARS-2 section (Chapter 6.4), by applying the SuMD method in collaboration with the Mechanistic and Structural Biology department of AstraZeneca, the effect of the RBD mutations characterising Delta and Omicron variants' was putatively revealed to address kinetic and SPR experiments reported in the literature.

My work proposed differences in the binding pathway and affinity between the RBD of Alpha, Delta, and Omicron variants. Specifically, the changes in the electrostatic landscape of the Omicron variant determined a series of metastable conformations and a sequence of interactions resulting in a slower optimal positioning for ACE2 binding. This observed behaviour matched with the description of Omicron's delayed binding kinetic. On the other hand, quantitative evaluation of the binding affinity, upon complex formation, indicated Omicron as the most prominent binder between the variants, in agreement with the experimental results.

The MD approaches deployed in this led the drug design to prevent the Nef-Nef dimerization mechanism. HIV's nonstructural proteins are promising targets for the development of novel therapeutics, due to the difficulties that highly mutant viruses pose for the design of effective vaccines. Nef is an important player in viral pathogenesis, and replication, and participates in the host immune mechanism suppression by altering macrophages' kinases. Structural studies revealed that Nef dimers are formed with Fyn kinase SH3 domain and Hck kinase SH3-SH2 domain, necessary for Nef activation. It follows that the Nef dimerization mechanism could be a viable target for drug development. The first work in collaboration with the University of Kolkata revealed possible binding pockets at the Nef-Nef binding interface. This crucial site was targeted by generative fragment-based methods through molecular docking and dynamics. The most promising compounds were then optimized for better access to the lipophylic binding interface and the extension of the hydrogen bond network.

mwSuMD is an advancement of the existing SuMD method, with applications that range from protein-ligand to protein-protein binding or unbinding predictions for both cytosolic and membrane systems. The power of mwSuMD lies in its parallelized architecture, which allows the simultaneous observation and comparison of multiple states, speeding up stochastic sampling typical of binding and unbinding events. mwSuMD has been tested on GPCRs challenging structural transitions such as the activation of the G protein upon binding and Nef antagonists' binding pathways within the nanosecond scale, drastically reducing the computational cost. The implementation of quantitative affinity scores in relation to the observed binding pathways should be the next milestone addressed by this method, possibly through a combination of QM/MM approaches.

The recent advancements in the field of artificial intelligence (AI) for discriminative and generative algorithms are a growing trend in medicinal research. AI is a powerful tool that finds its natural application in the data-rich field of drug discovery (1) and molecular representation (2) and aims to improve the predictions of physical-chemical properties of compounds (3), as well as the identification of binding pockets in proteins (4) or predicting the efficacy of compounds against a set of biological targets (5). Generative algorithms were

deployed through this PhD project to expand the exploration of the chemical space beyond the available databases and to refine the prediction of chemical properties of databases of compounds compared to computational libraries. Unfortunately, the massive enthusiasm for AI yet falls short when it comes to drug discovery results (6,7) with much room for improvement. At this stage, the field of AI research and computational chemistry need to work together for the development and refinement of more accurate FF parameters (8), the acceleration of QM calculations (9,10) and the prediction of protein structures (11) and their function. The advantages of AI's data analysis and speed, if paired with molecular modelling, might bring new life to drug discovery by enriching the existing chemical databases with additional, undiscovered features.

At the time of writing this text (January 2024), I am developing several software applications and computational tools. These include a long-to-short-term recurrent neural model (12, 13) that generates novel compounds using the Simplified Molecular Linear Insertion System (SMILES), a bitvector molecular fingerprint-based database scraper (14), a molecular dynamics-based pipeline for the search for cryptic antibody epitopes. This pipeline explores different angles and rotations in the formation of complexes, clustering and comparing their binding energies for comparative screening. I'm also working on a refined water-based docking protocol, which uses water occupations to refine Vina scores by exploring large databases of compounds. In addition, I am developing combinatorial software that uses molecular fragments to generate novel drugs and constitutional isomers, as well as many other publicly available tools on https://github.com/pipitoludovico.

References:

1. Paul, D. *et al.* Artificial intelligence in drug discovery and development. *Drug Discov. Today* **26**, 80–93 (2021).
2. David, L., Thakkar, A., Mercado, R. & Engkvist, O. Molecular representations in AI-driven drug discovery: a review and practical guide. *J. Cheminform.* **12**, 56 (2020).
3. Wang, Z., Li, H., Carpenter, C. & Guan, Y. Challenge-Enabled Machine Learning to Drug-Response Prediction. *AAPS J.* **22**, 106 (2020).
4. Bai, Q. *et al.* MolAICal: a soft tool for 3D drug design of protein targets by artificial intelligence and classical algorithm. *Brief. Bioinformatics* **22**, (2021).
5. Dhakal, A., McKay, C., Tanner, J. J. & Cheng, J. Artificial intelligence in the prediction of protein-ligand interactions: recent advances and future directions. *Brief. Bioinformatics* **23**, (2022).
6. AI's potential to accelerate drug discovery needs a reality check. *Nature* **622**, 217 (2023).
7. First AI-designed drugs fall short in the clinic, following years of hype – Endpoints News. https://endpts.com/first-ai-designed-drugs-fall-short-in-the-clinic-following-years-of-hype/.
8. Unke, O. T. *et al.* Machine learning force fields. *Chem. Rev.* **121**, 10142–10186 (2021).
9. Zheng, P., Zubatyuk, R., Wu, W., Isayev, O. & Dral, P. O. Artificial intelligence-enhanced quantum chemical method with broad applicability. *Nat. Commun.* **12**, 7022 (2021).
10. Hermann, J. *et al.* Ab initio quantum chemistry with neural-network wavefunctions. *Nat. Rev. Chem.* **7**, 692–709 (2023).
11. Skolnick, J., Gao, M., Zhou, H. & Singh, S. Alphafold 2: why it works and its implications for understanding the relationships of protein sequence, structure, and function. *J. Chem. Inf. Model.* **61**, 4827–4831 (2021).
12. Segler MHS, Kogej T, Tyrchan C, Waller MP. Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks. ACS Cent Sci. 2018 Jan 24;4(1):120–31.
13. Merk D, Friedrich L, Grisoni F, Schneider G. De novo design of bioactive small molecules by artificial intelligence. Mol Inform. 2018 Jan 10;37(1–2).
14. Capecchi A, Probst D, Reymond J-L. One molecular fingerprint to rule them all: drugs, biomolecules, and the metabolome. J Cheminform. 2020 Jun 12;12(1):43.