A bioinformatics analysis of therapeutic proteins : studying protein partners for the preimplantation factor (PIF) and of SARS-CoV-2 proteins

Mariem Ghoula

Université Paris Cité

École Doctorale Pierre Louis de Santé Publique :
Épidémiologie et Sciences de l'Information Biomédicale (ED393)

Unité de Biologie Fonctionnelle et Adaptative (BFA) (UMR 8251)
Équipe « Modélisation Computationnelle des Interactions Protéine-Ligand »

A bioinformatics analysis of therapeutic proteins: studying protein partners for the preimplantation factor (PIF) and of SARS-CoV-2 proteins

Mariem GHOULA

Thèse de doctorat en Bioinformatique

Dirigée par Pr. Anne-Claude Camproux
et co-dirigée par Dr. Gautier Moroy

Présentée et soutenue publiquement le 10 Juillet 2023
Devant un jury composé de:

Roland STOTE	DR	Université de Strasbourg	Rapporteur
Gwenaëlle ANDRÉ-LEROUX	DR	Université Paris-Saclay	Rapportrice
Samia ACI-SÈCHE	CR	Université d'Orléans	Examinatrice
Claudine MAYER	PU	Université Paris Cité	Examinatrice
Anne-Claude CAMPROUX	PU	Université Paris Cité	Directrice de thèse
Gautier MOROY	MCF-HDR	Université Paris Cité	Co-directeur de thèse
Patrick FUCHS	MCF-HDR	Université Paris Cité	Membre invité
Nathalie JANEL	PR-HDR	Université Paris Cité	Membre invité

Acknowledgements

I would like to express my profound gratitude to my thesis supervisors, Prof. Anne-Claude Camproux and Dr. Gautier Moroy. Their invaluable guidance, unwavering support, and continuous encouragement have played a crucial role in shaping the trajectory of my Ph.D. journey. Their expertise, dedication, and insightful feedback have significantly contributed to the refinement and enhancement of the quality of my research. I am deeply grateful for their mentorship and the profound impact they have had on my academic growth and development.

I thank the Agence Nationale de la Recherche for funding my research. In addition, I express my profound gratitude to the PIF21 project collaborators, especially Pr. Nathalie Janel, for their ongoing contributions and helpful suggestions during my research.

I am deeply grateful for the esteemed individuals who will serve as members of my thesis jury. I thank Dr. Roland Stote and Dr. Gwenaëlle André-Leroux for being the reviewers of my thesis and for their anticipated invaluable participation in evaluating my work and shaping its final outcome. I extend my sincere appreciation to Dr. Patrick Fuchs, Dr. Samia Aci-Sèche, Pr. Nathalie Janel and Pr. Claudine Mayer for their presence and for being part of my thesis jury as well.

I also thank my fellow colleagues, Sarah Nacéri, Pierre Laville, Bryan Dafniet, Vanille Lejal, Mehdi Munim, and Leila Ziani! You guys are the best! I cannot thank you enough for being so kind, engaging in super interesting discussions, and always having my back.

Finally, I would like to express my gratitude to all my friends and to my family who have supported me, provided valuable insights, and offered their assistance along the way. Your encouragement, friendship, and willingness to lend a helping hand have made this endeavor more meaningful and enjoyable. Yasir, thank you for always being there for me no matter what. Menelli, Claudine, and Helena, thank you for your constant support and for being the best friends anyone would wish for.

Completing this Ph.D. thesis would not have been possible without the collective contributions and support of these remarkable individuals. I am truly fortunate to have had the privilege of working with such exceptional mentors, colleagues, and loved ones who have left an indelible mark on my academic and personal growth.

Thank you all for being a part of this incredible journey.

List of abbreviations

PIF: Preimplantation Factor

AA: Amino Acid

Da: Dalton

GnRH: Gonadotropin-releasing hormone

PPI: Protein-Protein Interaction

CCPs: Cell Penetrating Peptides

MD: Molecular Dynamics

CAPRI: Critical Assessment of PRediction of Interaction

MM-GBSA: Molecular Mechanics Generalized Born Surface Area

MM-PBSA: Molecular Mechanics Poisson-Boltzmann Surface Area

FEP: Free Energy Perturbation

ML: Machine Learning

NMR: Nuclear Magnetic Resonance

RBD: Receptor-Binding Domain

ACE2: Angiotensin-Converting Enzyme 2

ANR: Agence Nationale de la Recherche

SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus 2

PDB: Protein Data Bank

Mpro: Main Protease

PLpro: Papain-Like Protease

sPIF: synthetic PIF

FDA: Food and Drug Administration

DS: Down Syndrome

MALDI-TOF: Matrix Assisted Laser Desorption Ionisation/Time Of Flight

HPLC: High Performance Liquid Chromatography

CSP: Circumsporozite protein/plasmodium falciparum

IDE: Insulin-Degrading Enzyme

KCNAB1: K+ voltage-gated channel, shaker-subfamily, beta1

Aβ: Amyloid-β

PDI: Protein Disulfide-Isomerase

LC/MS/MS: Liquid Chromatography tandem Mass Spectrometry

TRX: Thioredoxin

IL-6: Interleukin 6

IL-8: Interleukin 8

ERO1: Endoplasmic Reticulum Oxidoreductase 1

CD4: cluster of differentiation 4

CD8: cluster of differentiation 8

MHC: Major Histocompatibility Complex

HSP: Heat Sock Protein

MYH10: Myosin Heavy Chain 10

HTR: Human Throphoblast

DYRK1A: Dual specificity tyrosine-(Y)-phosphorylation-Regulated Kinase 1A

D1: domain 1

D2: domain 2

D3: domain 3

D4: domain 4

CAMP: Convolutional Attention-based Neural Network for Multi-level Peptide-protein Interaction Prediction

AI: Artificial intelligence

PEST: C-terminal domain enriched in proline (P), glutamic acid (E), serine (S), and threonine (T)

MAPK: Mitogen-Activated Protein Kinase

PI3K: PhosphoInositide 3-Kinase

PKC: Protein Kinase C

PRDX6: Peroxiredoxin-6

BLAST: Basic Local Alignment Search Tool

SPALN: Space-Efficient Spliced Alignment

cDNA: complementary DeoxyriboNucleic Acid

GRCh38: Genome Reference Consortium Human Build 38 Organism: Homo sapiens (human)

hg38: human genome 38

mRNA: messenger RiboNucleic Acid

GRCm38: Genome Reference Consortium Mouse Build 38 Organism: Mus musculus (house mouse)

mm10: mus musculus 10

NRXN3: Neurexin-3

ncRNAs: non-coding RNAs

lncRNAs: long non-coding RNAs

sncRNAs: short non-coding RNAs

nt: nucleotide

smORF: short open reading frame

Content

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction to protein-protein and protein-peptide interactions

Evolution has optimized the interactions of proteins with other proteins and molecules. Biological processes like signaling networks, DNA repair, metabolism, gene expression, replication, transport, and folding are all carried out by proteins at the cellular level in living creatures. Protein interactions with molecules like other proteins, peptides, carbohydrates, lipids, or nucleic acids enable them to accomplish these functions. The most prevalent of these interactions are protein-protein interactions, 15-40% of which are mediated by a short peptide stretch [1]. Signaling, regulatory networks, cell localization, protein degradation, and immunological response are all impacted by protein-peptide interactions. Moreover, many studies demonstrated that protein-peptide interactions could serve as a pharmacological target and that the peptides themselves could make good drug candidates [2]. The molecular association of a protein-protein or a protein-peptide achieved through weak intramolecular forces is a key process in both chemical and biological recognition, with the interaction possibly associated with a conformational change leading to the activation of a biological pathway [3, 4]. Therefore, understanding the molecular interaction between a ligand (e.g., primarily a small molecule or a peptide) and its target (e.g., the surface of a protein) is a strategic approach to understanding the mechanisms triggering biological communication [4] and designing new drugs [5].

1.2 Protein-protein interactions impact for therapeutic uses in Drug Discovery

Protein-protein interactions are an appealing target for therapeutic intervention in drug discovery because they are essential for many biological functions and signaling pathways [6, 7]. New drugs that can be developed to treat a wide range of diseases, such as cancer, viral infections, and autoimmune disorders, may be developed by being able to selectively disrupt or modulate these interactions [8]. The development of anti-cancer drugs that target the interaction between the Bcl-2 family of proteins and pro-apoptotic proteins is one of the most well-known examples that highlights the potential of targeting protein-protein interactions for

therapeutic applications [9]. Recent years have seen a major advancement in our knowledge of protein-protein interactions [10-12]. To find and define prospective therapeutic targets, a number of methods, including high-throughput screening and structural biology, have been built. For the purpose of detecting and developing inhibitors or modulators of protein-protein interactions, the use of computational tools and machine learning algorithms have also become increasingly crucial [11, 12]. Drug discovery continues to be impacted by protein-protein interactions, which has the chance to completely impact how many diseases are treated.

1.3 Peptide potential therapeutic uses in Drug Discovery

Therapeutic peptides are a distinct class of pharmacological drugs consisting of a succession of amino acids (AA) with molecular weights ranging from 500 to 5000 Da (Dalton) [13]. Fundamental studies into therapeutic peptides began with investigating natural hormones such as insulin, oxytocin, vasopressin and GnRH (Gonadotropin-releasing hormone) [14]. These studies shed light on their particular physiological functions in the human body, which sparked interest in therapeutic peptide research. Ultimately, more than 80 therapeutic peptides have been approved globally since the production of the first therapeutic peptide insulin, in 1921, thanks to the significant advancements in the field [15]. Thus, the production of peptide medications became one of the greatest areas of pharmaceutical research due to its evolutionary progress. Common actions of therapeutic peptides include those of hormones, growth factors, neurotransmitters, ion channel ligands and antibiotics. Similarly, to biological agents, such as therapeutic proteins and antibodies, they bind to cell surface receptors with high affinity and specificity and cause intracellular effects. However, in contrast to biological agents, therapeutic peptides exhibit less immunogenicity and cost less to produce than biological agents [16-19]. Therapeutic peptides present several advantages compared to costly small pharmaceutically developed molecules. Apart from their competitive price advantages, therapeutic peptides are better inhibitors of protein-protein interaction (PPI) surfaces. Due to their larger size and flexible backbone, they cover larger PPI regions compared to small molecules [20]. However, therapeutic peptides also come with drawbacks, such as possessing a low membrane permeability and poor *in vivo* stability [14, 16]. One promising approach to overcome the problem of poor cellular membrane penetration of therapeutic peptides is the use of cell penetrating peptides (CPPs), which have shown potential in enhancing the intracellular delivery of proteins due to their ability to efficiently cross cellular membranes [21]. These fundamental benefits and drawbacks of peptides give potential for peptide drug design and optimization as well as challenges for peptide medication development.

1.4 Computational approaches for predicting protein-protein and protein-peptide interactions

There are three major types of computational approaches for predicting protein-protein and protein-peptide interactions: 1) Molecular Docking, 2) Molecular Dynamics (MD) simulations

and 3) Machine Learning. Of course, each method is dependent on the amount of information provided to initiate the prediction. This information includes prior knowledge of a potential interaction between both partners, the three-dimensional structure of the peptide or the protein, details about the binding region or the residues involved in the protein-peptide interaction, and the type of signaling pathway through which this interaction is involved to understand its subsequent effects.

1.4.1 Molecular Docking

Protein and peptide docking is a traditional and efficient method that predicts the conformation and the binding mode of the bound protein or peptide [22, 23]. Because of its success in small molecule-protein docking and ease of use such as through webservers and standalone software, this method for virtual screening in the early stages of drug discovery has grown in popularity [24]. This success has been transferred into the field of protein-protein docking, as demonstrated by the evolution of predictions in the CAPRI (Critical Assessment of PRediction of Interaction) competition [25]. Protein-protein docking is a great approach to fully understand protein interactions when a complex is not available. It also provides the capacity to predict the effects of mutations and screen many compounds once the interactions are identified. Similarly, protein-peptide docking gives important information about the identification of potential candidates for drug development that can mimic the binding of large protein domains. Despite these successes, protein-protein and protein-peptide docking remains a more challenging problem and represents a critical issue in structural biology [11, 26]. Protein-protein docking can face the challenge of the great computational complexity and the difficulty of predicting conformational changes that occurs upon binding. As for protein-peptide docking, many of the issues can be addressed with prior knowledge of the peptide conformation, which can provide more accurate results. Understanding which residues are buried in the binding pocket allows researchers to replicate these contacts during the therapeutic design process. Furthermore, the flexibility of peptides significantly increases the sampling problem in comparison to small molecule docking and limits sampling native-like poses, reducing the docking approach success [27]. In fact, many computational docking approaches have concentrated on small molecules, with peptide docking methods being understudied. As a result, additional steps such as improving the scoring functions and using molecular modeling tools including MODELLER [28] or AlphaFold2 [29] may be required to gather information about the conformation of the biological partners. Indeed, rational design of peptide-based therapeutics necessitates atomistic understanding of protein-peptide complexes. Thus, again, many challenges come across when studying peptides. Although the use of computational pipelines in the early phases of drug discovery for small molecules is well-defined, peptides present some unique challenges that have limited the progress of computational pipelines for inhibitory peptide design. Plus, peptides interact with proteins in a wide range of ways. 1) as coils through specific amino acid interactions; 2) as well-defined secondary structures (e.g., hairpins or alpha helices); and 3) as discontinuous interactions along the peptide chain. Peptide flexibility is important in these interactions because they are often intrinsically disordered in their free form and adopt well-defined structures upon binding - unlike small molecules, which have more limited flexibility.

Nonetheless, several approaches that achieve good accuracy in general biological contexts have recently been developed (e.g., Rosetta FlexPepDock ab initio [30], CABS-dock [31], MDockPeP2 [32], Autodock CranckPep [33], HPEPDOCK [34] etc.). These approaches rely on either template-based [35, 36] or template-free [37, 32] docking. Although Smina [38] which is a molecular docking program that is based on AutoDock Vina and was primarily developed for small molecule docking, it has been modified to handle protein-peptide docking as well. The AlphaFold2-multimer [39] software can also ensure that the complex is accurate and provides good results when results are compared to native-like poses. Another strategy for overcoming the docking limitations is improving the curation of protein-protein and protein-peptide databases which is a critical method for systematic testing, benchmarking, and evaluating these docking methods [40].

1.4.2 Molecular Dynamics simulations

Because of the nature of MD simulations, it is possible to obtain thermodynamics, kinetics, and mechanistic understanding of the protein-protein and protein-peptide binding and unbinding processes [41]. As with other biological applications, MD is limited by the precision of the quantum mechanics model used and the ability to sample the biological complex energy landscape, which typically entails computational resources beyond our existing capacity. Although quantum mechanics can theoretically predict molecule interactions, large biomolecules remain unaffordable in practice [41]. There are important distinctions between small molecule and peptide binding to proteins which makes the latter more computationally challenging. Therefore, usually MD is used as a refinement tool rather than a high-throughput method to routinely assess the structures of protein-peptide complexes when the bound state is unknown. Thus, MD is commonly used to 1) refine docking results [42], 2) estimate binding affinities [43], and 3) use integrative modeling strategies to determine a protein-protein or a protein-peptide complex structure [44]. For instance, short MD simulations are frequently used as the final stage in docking calculations to remove steric overlap, account for local conformational changes, and identify structures based on physicochemical principles. Second, remarkable efforts have been made to determine protein-protein and protein-peptide complexes by combining and integrating computational modelling approaches with experimental data. Using modeling software such as MODELLER [28] or AlphaFold2 [29] can in some cases provide sample protein and peptide conformations that match experimental results when these are limited. When the experimental structure of the complex is known or has been modeled with molecular modeling approaches, MD can be used to determine the binding affinities of the complex using the most common methods of MM-GBSA (Molecular Mechanics Generalized Born Surface Area) and MM-PBSA (Molecular Mechanics Poisson-Boltzmann Surface Area) [45]. It is important to note the robustness of these approaches. However, the results obtained remain estimations as important factors such as flexibility, solvation and entropy can limit the accuracy of the results [46]. Free energy perturbation (FEP), a pathway-based free energy calculation method, has made remarkable accuracy in modeling protein binding with small molecules for a broad range of ligands [47].

1.4.3 Machine Learning

Since AlphaFold2 recent success in protein structure prediction, Machine Learning (ML) approaches are rapidly evolving [48]. Some ML results have shown high quality when compared to models generated with nuclear magnetic resonance (NMR) data [49] or docking approaches [29], demonstrating the success and importance of this field in protein-protein and protein-peptide complex prediction. With these great advances, programs using deep learning methods have emerged to predict protein-peptide complexes [50]. Thus, we can imagine that combining ML pipelines with structure-based approaches is a good strategy to study and analyze in depth a biological complex.

1.5 Thesis organization

Drug discovery requires a thorough understanding of the molecular interactions between protein-protein and protein-peptide. The significance of protein-protein and protein-peptide interactions and their effect on the biological functions of proteins, and the development of potential drugs are examined in this thesis and are thoroughly explained in **the first chapter**. Two biological systems are studied in this thesis. The protein-protein complex involving the Spike protein receptor-binding domain (RBD), and the human angiotensin-converting enzyme 2 (ACE2) receptor was examined in this thesis. The protein-peptide interactions were centered on the Preimplantation Factor (PIF) which is a critical peptide that is necessary for successful implantation and the development in the early stages of pregnancy. It is important to note that this work has been funded by the Agence Nationale de la Recherche (ANR) under the PIF21 project (No. ANR-19-CE18-0023). The research was a collaborative effort, involving multiple contributors who provided valuable insights, experimental analyses, validations, and fruitful exchanges throughout the study (Annex 1). The study included both *in vitro* and *in silico* research. Our team concentrated on *in silico* analysis, using computational approaches to investigate protein-protein and protein-peptide interactions.

In **the second chapter**, we introduce the aim of studying the Spike protein and ACE2 interaction. **The third chapter** examine the interactions of protein-protein interactions, particularly those between the human ACE2 receptor and the Spike protein RBD. A thorough bioinformatics research procedure was applied to study RBD-ACE2 interactions using a range of approaches, including molecular modeling, MD simulations, MM-PBSA analysis and druggable pocket tracking. This methodology was carefully created to be applicable to additional possible protein targets of PIF in addition to the Spike protein RBD and ACE2 receptor. Application of the bioinformatics protocol to study the Spike protein led to the identification of three potential druggable sites, providing potential targets for the development of drugs aimed at inhibiting its activity. The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants of concern are studied in **the fourth chapter** and are their effect on the RBD-ACE2 are analyzed as well. Another focus of the thesis that has been discussed in **the**

fifth chapter is the analysis of PIF-protein interactions and the application of bioinformatics techniques to their solution. Many bioinformatics tools have been used to investigate PIF in detail. One strategy has been to use molecular docking and MD simulations to find possible protein targets of PIF.

The sixth chapter discusses how IDE was recognized as a possible protein target of PIF in earlier studies. Molecular docking and experimental studies, however, failed to support this interaction. Nonetheless, this research inspired a thorough examination of IDE and provided significant new information about its function in a variety of biological processes. Although IDE may not be a protein target of PIF directly, the development in IDE research brought about by this study offers crucial insights into its biological function and prospective therapeutic uses. Due to its function in cellular growth and differentiation as well as its participation in numerous signaling pathways crucial for the control of neuronal development, DYRK1A was also investigated as a potential target of PIF. In order to research the potential relationship between DYRK1A and PIF, experimental and computational methods, like AlphaFold2-multimer [39], were used. The direct relationship between DYRK1A and PIF, however, was unable to be established despite significant efforts. A deep learning tool was also used to look through experimental proteomics data in addition to conventional bioinformatics methods to find possible PIF target proteins. With the use of this technique, it was possible to create an interaction score that identified proteins with a significant possibility of interacting with PIF. Thus, the use of the deep learning method enhanced the possibility of finding novel interactions between PIF and its protein targets as well as gave more prospective protein targets for study. Using a variety of methods from genomics, the genetic origin of PIF has also been carefully studied and explored in **the seventh chapter**. More research is being done to better understand PIF regulation at the transcriptional and post-transcriptional levels. We now have a better understanding of the potential protein targets of PIF thanks to the findings of this thesis, which also served as a foundation for additional experimental verification of these interactions. **The eighth** and final chapter concludes this thesis and offers perspectives that might be interested to explore. Overall, this research has paved the way for future investigations looking into the interactions between PIF and its protein targets and has offered significant new insights into the mechanisms underlying protein-protein interactions.

Chapter 2

Due to the COVID-19 pandemic and the challenges it provided to our research timetable, the primary focus of my thesis, which focused on the study of the PIF peptide and its possible protein partners, encountered delays in obtaining the experimental data that was needed to confirm its protein targets. Consequently, to utilize the available time effectively, a strategic decision was made to shift the research focus towards the structural analysis of SARS-CoV-2 proteins. This decision aimed to develop a methodology that could later be applied to the study of PIF and its potential protein partners. Therefore, collaborating with my Ph.D. colleague, Sarah Naceri, we proceeded on the examination of the interaction between the Spike protein RBD domain and the human receptor ACE2. This collaborative work resulted in the publication of two articles [51, 52], further contributing to the scientific understanding of SARS-CoV-2 and its protein interactions.

2.1 Structural Analysis of SARS-CoV-2 Proteins: Insights, Strategies, and Therapeutic Prospects in the Fight against COVID-19

The emergence of the COVID-19 viral infection, caused by SARS-CoV-2, represented a significant global threat, leading to a devastating impact on public health [53]. Hence, because of the virus's highly infectious nature and severe clinical symptoms, important scientific research was required to understand its biology and develop effective countermeasures [54]. Consequently, structural analysis of SARS-CoV-2 proteins played a crucial role in elucidating the molecular mechanisms underlying this viral infection. As of the most recent available data, an impressive number of protein structures reaching more than 200,000 structures have been resolved and deposited in the Protein Data Bank (PDB) [55], providing valuable insights into the virus's architecture (Figure 1).

Among the key proteins targeted for structural analysis, The Spike protein was one of the primary proteins selected for structural investigation because of its critical involvement in viral entrance into host cells [56]. The Spike protein facilitates viral entrance by mediating the interaction between the viral particle and the human ACE2 receptor [56]. Other proteins, such as the main protease (Mpro) [57] and the papain-like protease (PLpro) [58], were also subjected to thorough structural studies since they played major roles in viral maturation and replication. Thus, numerous articles emerged, outlining diverse strategies for targeting SARS-CoV-2, often complemented by experimental analyses [59]. These studies demonstrated the potential of various molecules, including small molecules [60], monoclonal antibodies [61], and vaccine candidates [62], as promising therapeutic interventions. Notably, molecules such as remdesivir

[63], a nucleotide analog targeting viral RNA synthesis, and monoclonal antibodies, such as bamlanivimab [64] and casirivimab/imdevimab [65], gained significant attention and underwent extensive testing to validate their efficacy against SARS-CoV-2.

Additionally, fighting the COVID-19 pandemic has been made significantly more difficult by the introduction of SARS-CoV-2 mutations. These variations are brought about by genetic modifications in the viral genome, which alter the viral properties of infectivity, transmissibility, and immune evasion [66]. Variants such as the Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2), and Omicron (B.1.1.529) have garnered particular attention due to their global spread and potential impacts on public health [66].

With this in mind, our research focused on investigating the inhibition of the interaction between the Spike protein RBD domain and the human ACE2 receptor as a potential therapeutic strategy. By elucidating the structural details of this interaction and employing computational approaches, we aimed to identify druggable pockets that can be targeted with effective inhibitors disrupting the virus critical viral entry step [51]. Another focus was made on the SARS-CoV-2 variants of concern to better understand their mechanism and their impact on the affinity between RBD and ACE2 [52].

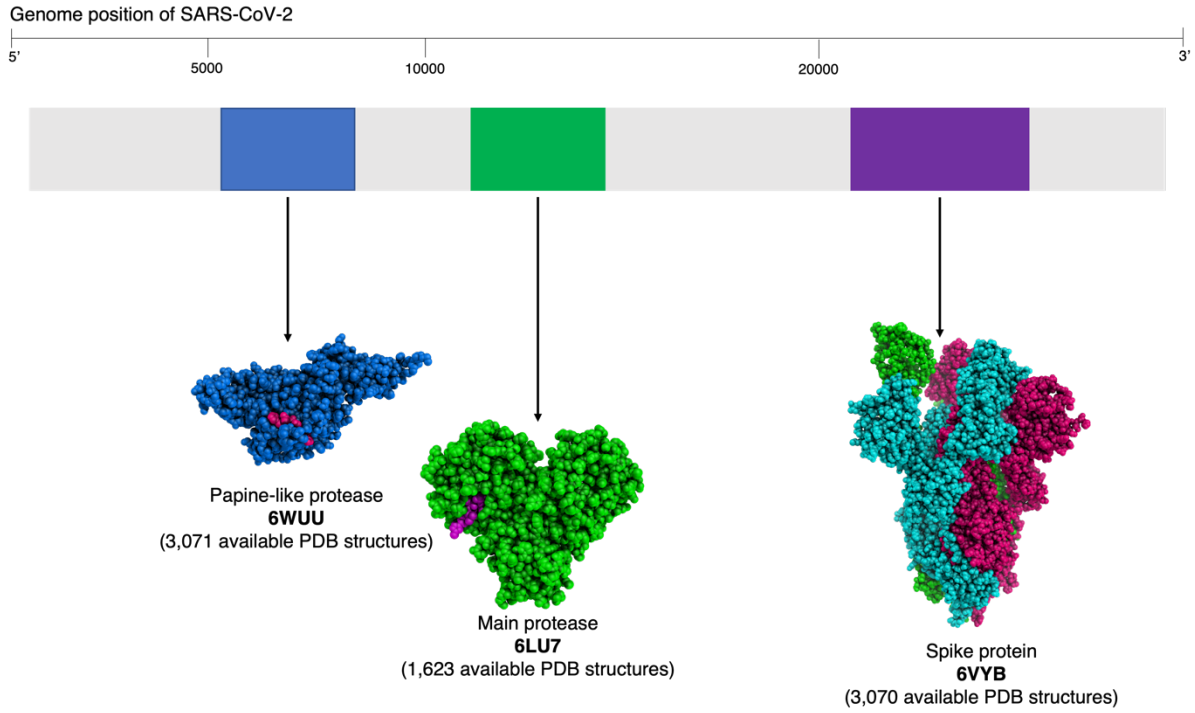


Figure 1. Simple representation of the SARS-CoV-2 architecture including the papain-like protein, the main protease protein, and the Spike protein.

2.2 The Spike protein-ACE2 interaction

The Spike protein is a crucial viral protein that plays a central role in viral entry and infection

[56]. Structurally, it is a trimeric glycoprotein composed of three identical subunits, each consisting of two functional domains: the S1 domain and the S2 domain. The S1 domain contains the RBD, which specifically interacts with the ACE2 receptor (Figure 2). The S2 domain mediates membrane fusion and facilitates viral entry into host cells [56].

Additionally, the Spike protein exists in two distinct conformations: an active form and an inactive form. These conformational states are associated with the “open” and “closed” states of the Spike protein. In the active form, the RBD adopts an “up” conformation, exposing the receptor-binding motif and facilitating interaction with the ACE2 receptor. This open conformation is thought to be critical for efficient viral entry. Conversely, in the inactive form, the RBD is in the “down” conformation, being less accessible for receptor binding (Figure 2) [56].

Studying the interaction between the Spike protein and ACE2 in the context of COVID-19 is of great importance for several reasons. Firstly, understanding the molecular details of this interaction can provide insights into viral entry mechanisms and uncover the key interactions involved between these two proteins. Secondly, targeting this interaction offers potential therapeutic opportunities. By developing inhibitors that disrupt the Spike protein-ACE2 interaction, it may be possible to prevent viral attachment and entry, thereby reducing viral load and mitigating the severity of COVID-19. Moreover, studying this interaction can aid in the design of vaccines and immunotherapies that target the Spike protein, aiming to elicit an immune response against the virus and provide protective immunity.

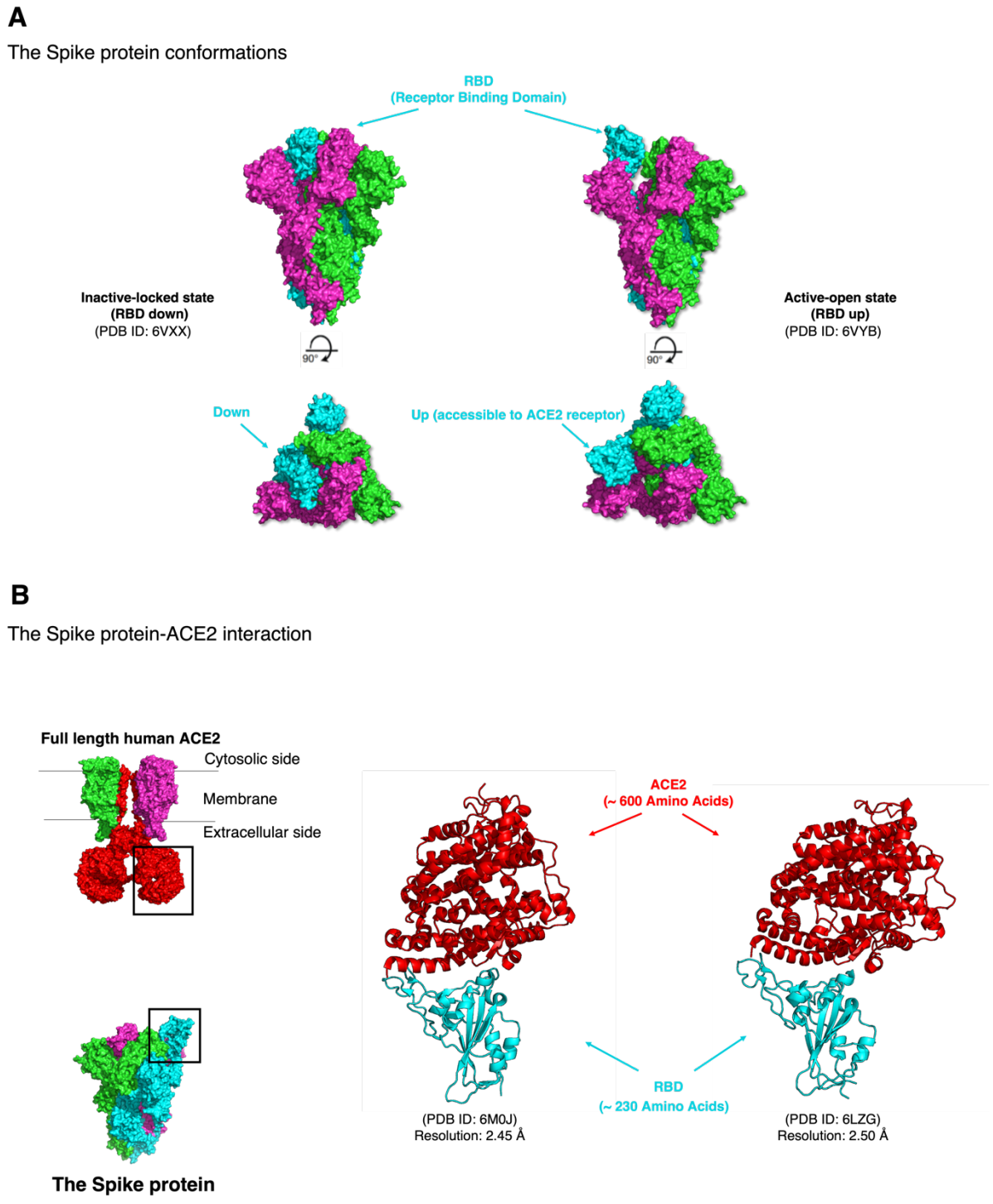


Figure 2. (A) Representation of the two different conformations of the Spike protein in its active and inactive forms. The RBD is also represented in its up and down conformations. (B) Representation of the interaction between RBD and ACE2.

Chapter 3

Protein-protein interaction: Structural exploration of the SARS-CoV-2 RBD and the human receptor ACE2 interaction

We therefore focused on the Spike protein RBD-ACE2 interaction to examine protein-protein interactions. In the following paper [51], we used a variety of bioinformatics methods, such as MD simulations, MM-PBSA free energy calculations and pocket druggability tracking. The purpose of this study was to develop a compelling protocol that considers the structural flexibility of RBD and incorporates machine learning techniques. This would enable us to locate potential druggable binding sites on the RBD for use in drug discovery and design. For this reason, we examined the structural flexibility of both the RBD-ACE2 complex and the unbound RBD using MD simulations. We also identified hot spots, which are important residues involved in the RBD-ACE2 interaction, using MM-PBSA energetics computation. Using a supervised machine learning technique developed in the lab, we assessed the pockets on the protein surface during the MD simulations. The pockets were then classified in terms of druggability scores. We next performed unsupervised hierarchical classification on the large set of estimated pockets using pocket similarity in terms of residue composition. Binding sites are frequently observed along MD simulations and represent primary pocket clusters that group pockets with comparable residue compositions. In terms of the main residue's contribution to the binding site and its druggability, residue localization, frequency, variability, and druggability permitted the identification of three promising druggable pockets. Furthermore, the study covers the potential mechanistic effects of SARS-CoV-2 mutations on the selected druggable binding sites situated in critical areas of the RBD surface and key hot spots. In conclusion, this research presents a promising method for comprehending protein-protein interactions and developing possible therapeutics.

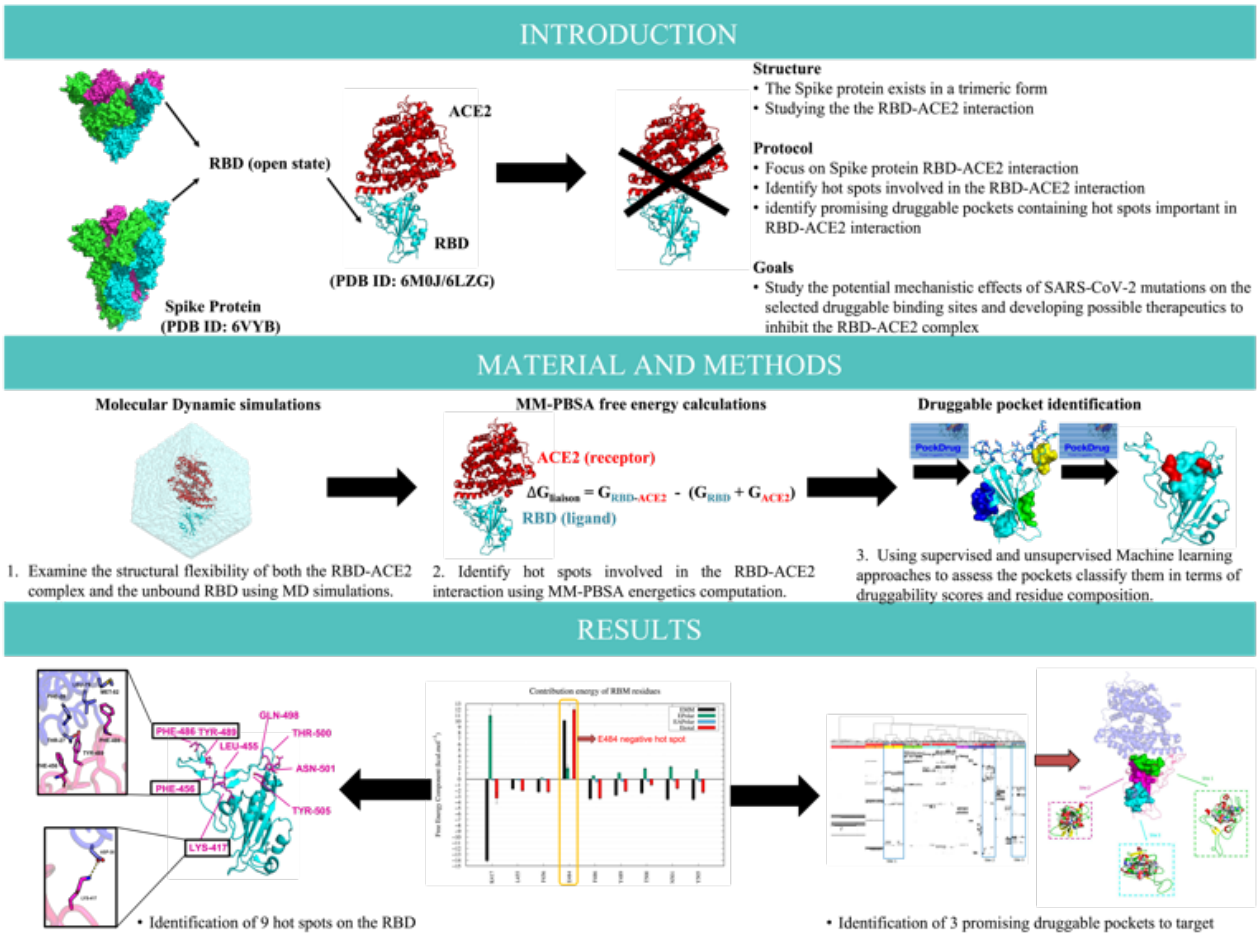


Figure 3. Representation of the bioinformatics protocol used to study RBD-ACE2 interaction.

Computational and Structural Biotechnology Journal 21 (2023) 2339–2351

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

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

Identifying promising druggable binding sites and their flexibility to target the receptor-binding domain of SARS-CoV-2 spike protein

M. Ghoula [1], S. Naceri [1], S. Sitruk, D. Flatters, G. Moroy [*,1], A.C. Camproux [*,1]

Université Paris Cité, CNRS, INSERM, Unité de Biologie Fonctionnelle et Adaptative, F-75013 Paris, France

ARTICLE INFO

Article history:
Received 25 October 2022
Received in revised form 16 March 2023
Accepted 16 March 2023
Available online 18 March 2023

Keywords:
SARS-CoV-2
Spike protein
Structural flexibility
Hot spot residues
Pocket tracking
Binding site flexibility
Druggable binding sites
Key-residues
Molecular dynamics simulation
COVID-19 variants

ABSTRACT

The spike protein of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is crucial for viral infection. The interaction of its receptor-binding domain (RBD) with the human angiotensin-converting enzyme 2 (ACE2) protein is required for the virus to enter the host cell. We identified RBD binding sites to block its function with inhibitors by combining the protein structural flexibility with machine learning analysis. Molecular dynamics simulations were performed on unbound or ACE2-bound RBD conformations. Pockets estimation, tracking and druggability prediction were performed on a large sample of simulated RBD conformations. Recurrent druggable binding sites and their key residues were identified by clustering pockets based on their residue similarity. This protocol successfully identified three druggable sites and their key residues, aiming to target with inhibitors for preventing ACE2 interaction. One site features key residues for direct ACE2 interaction, highlighted using energetic computations, but can be affected by several mutations of the variants of concern. Two highly druggable sites, located between the spike protein monomers interface are promising. One weakly impacted by only one Omicron mutation, could contribute to stabilizing the spike protein in its closed state. The other, currently not affected by mutations, could avoid the activation of the spike protein trimer.

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is responsible for the COVID-19 outbreak [1,2]. It was originally discovered in Wuhan, China, in late December 2019. The World Health Organization (WHO) labeled this epidemic a pandemic in March 2020 and reported nearly 760 million confirmed cases of COVID-19 and 6.8 million deaths by the end of March 2023 (https://covid19.who.int/). Important variants of the SARS-CoV-2 virus have emerged in the UK, Brazil, India, and South Africa from December 2020 to the end of November 2021. Five variants have been recognized by the WHO as variants of concern and are labeled as coronavirus variants: Alpha, Beta, Gamma, Delta, and Omicron [3–5]. SARS-CoV-2 is an extremely unstable virus [6,7], which is favorable for the development of new variants.

COVID-19 vaccines have enabled the reduction of the spread, severity, and death caused worldwide. Eight drugs were approved in May, 2022. Three of these drugs are biologics, which aim to block viral attachment and entry into human cells [8–10]. Two are a combination of two monoclonal antibodies intended to prevent mutational escape: casirivimab/imdevimab (commercialized under the name Ronapreve) [8] and bamlanivimab/etesevimab [9]. The third drug, sotrovimab (Xevudy) [10], is also a monoclonal antibody. Although monoclonal antibodies are an important therapeutic advancement, their manufacturing costs are high, and they are not convenient for patients because they are administered intravenously. Small molecules are often cheaper and easier to produce than protein- or peptide-based drugs [11]. They can withstand a wide range of delivery modalities, including oral administration, making them a preferred choice among pharmaceutical chemists. Currently, four small-molecule drugs are available for the treatment of COVID-19. Remdesivir and molnupiravir exert their antiviral action by perturbing viral RNA replication. Baricitinib attenuates the uncontrollable inflammatory response by the immune system owing to SARS-CoV-2 infection, referred to as a cytokine storm, by specifically inhibiting Janus kinases. From the end of December 2021, a fourth drug, named Paxlovid, has become available for people who are at high risk of developing severe COVID-19 symptoms [12]. Paxlovid is a combination of two antiviral medications, nirmatrelvir

* Corresponding authors.
E-mail addresses: gautier.moroy@u-paris.fr (G. Moroy), anne-claude.camproux@u-paris.fr (A.C. Camproux).
[1] These authors contributed equally

https://doi.org/10.1016/j.csbj.2023.03.029

and ritonavir, administered orally. Several repurposed drugs and new drug candidates are currently in phase III and IV trials. However, it is still crucial to develop drugs that can alleviate the severity of COVID-19 in individuals who are at high risk for progression to severe COVID-19.

The instability of the SARS-CoV-2 genome and the high possibility of new variants emerging make the discovery of new treatments and effective maintenance of already discovered drugs challenging [13]. Therefore, it is crucial to understand the interaction mechanisms of SARS-CoV-2 at the molecular level and the impact of its variants on these interactions.

The SARS-CoV-2 genome encodes four structural proteins: the spike, envelope, membrane, and nucleocapsid proteins, and 16 nonstructural proteins. Spike, envelope, and membrane proteins form the viral envelope, and the nucleocapsid protein binds to the RNA genome [14–16]. The spike protein is a homotrimeric glycoprotein and each monomer is composed of 1273 residues [17]. In coronaviruses, the spike protein can interact with human angiotensin-converting enzyme 2 (ACE2) to initiate fusion with host cells [18]. The receptor-binding domain (RBD) of the spike protein is responsible for the interaction with ACE2 [19]. Therefore, the RBD is a crucial protein target for the development of COVID-19 drugs. The 17 residues on the RBD that interact directly with ACE2 have been grouped under the name of the receptor-binding motif (RBM). Targeting the RBM using small molecules, therapeutic peptides, and neutralizing antibodies was determined to be an attractive method to inhibit the ability of the spike protein to bind ACE2, owing to its low glycosylation [20–22].

Advancements in structural biology and structural bioinformatics methods have enabled the elucidation of molecular and dynamic mechanisms of protein-protein or protein-molecule interactions. It is possible to design molecules capable of disrupting protein-protein interactions using structure-guided approaches. For this purpose, it is important to first understand the flexibility of the structures and to identify the residues essential for stabilizing the interaction, commonly named “hot spots”.

However, the interaction between a drug and a target protein depends on a few key residues as well as on a larger protein cavity or pocket, referred to as the binding site, which must have physicochemical and geometrical properties in agreement with those of the ligand [23]. Therefore, it is also crucial to determine which regions of the protein surface have the suitable druggability profile that can be targeted by therapeutic molecules.

The first SARS-CoV-2 spike protein structure was resolved in February 2020 [24]. The number of available spike protein structures has increased rapidly, explaining the structural mechanisms that allow SARS-CoV-2 entry into the host cell. According to cryo-electron microscopy (cryo-EM) structures of the spike protein, RBD exists in two states: a closed (or “down”) and an open (or “up”) state. In the latter state, the RBD is less buried and can interact directly with ACE2 [24]. At the beginning of this study, three crystal structures of the wild-type ACE2/RBD complex were available: PDB IDs 6M0J [25], 6LZG [26], and 6VW1 [27].

With the release of the experimental structures of the SARS-CoV-2 RBD-ACE2 complex, numerous studies have been performed to elucidate the molecular and dynamic mechanisms involved in this protein-protein interaction.

Considering the flexibility of the protein partners improved the identification of hot spot residues. Spinello et al. identified 12 hot spot residues using SARS-CoV-1 and SARS-CoV-2 PDB structures and the molecular mechanics Poisson Boltzmann surface area (MM-PBSA) method [28,29]. Using structural comparisons between the SARS-CoV-1 RBD-ACE2 X-ray structure and a SARS-CoV2 RBD-ACE2 model built by homology and the MM-PBSA method, the authors identified ten hot spots for the affinity of the SARS-CoV-2 RBD for ACE2 [30]. Recently, a combination of molecular dynamics (MD) simulations and the Molecular mechanics-generalized Born surface area (MM-GBSA) method enabled the identification of 13 hot spot residues [31]. Overall, the hot spot residues assessed using these different methods showed coherency but clear dependency on the employed experimental structures and the protocol used.

Identification of these hot spot residues is a key step in the design of drug molecules capable of disrupting protein-protein interactions. However, the protein surface region or binding site to be targeted by therapeutic compounds also requires characterization in terms of the physicochemical and geometrical properties inherent to its flexibility. The binding sites must be druggable and possess physicochemical and geometrical properties that allow them to bind to drug-like molecules [23].

Some studies have focused on the determination of druggable sites on SARS-CoV-2 spike RBD, in which occupancy could reduce, directly or through an allosteric mechanism, the interaction between the RBD and ACE2 [32–35]. Trigueiro-Louro et al. identified druggable pockets from different spike protein structures using consensus from pocket estimations [32]. They concluded that the RBD is one of the two most promising conserved druggable regions. They identified four to seven RBD pockets from three RBD-ACE2 complex experimental structures (PDB IDs 6VW1, 6LZG, and 6M0J) without considering protein flexibility. They selected two pockets characterized as druggable, observed in the three RBD structures. Carino et al. [33] identified 300 putative pockets in the whole trimeric structure of the spike protein using the Fpocket estimation program [36]. They selected six pockets on the RBD structure based on their potential druggability, structural rigidity, and sequence conservation between the SARS-CoV-2 and SARS-CoV-1 RBD. They identified two continuous pockets in the central β-sheet core of the spike RBD that were targetable by steroidal molecules based on virtual screening of FDA-approved drugs on these six RBD pockets. Using in vitro assays, they confirmed that several compounds highlighted by virtual screening can reduce the ability of the RBD to bind to ACE2.

Although these studies are promising, they have been performed only on static protein structures. The dynamic aspect of the structure, as well as the flexibility of the pockets, were not considered. To identify reliable druggable binding sites for drug design, the importance to consider the presence of cryptic, transitory, and flexible pockets, have been underlined by Stank et al. [37], Abi Hussein et al. [38], and Kuzmanic et al. [39]. Recent research conducted by Dokainish et al. [40] stressed the importance of considering the intrinsic flexibility of the SARS-CoV-2 spike protein receptor domains. They identified cryptic pockets from RBD intermediate states determined using MD simulations and concluded that the intrinsic flexibility of the spike protein must be taken into consideration when developing vaccines or antiviral medications.

In this study, we proposed a protocol considering the structural flexibility of RBD while combining machine learning approaches to identify promising RBD druggable binding sites for drug design development.

For this purpose, we performed a structural flexibility analysis of both unbound RBD and RBD-ACE2 complex using MD simulations. We also identified hot spots (crucial residues involved in the RBD-ACE2 interaction) using MM-PBSA energetics computation. Using a large number of conformations extracted from the MD simulations, we estimated the pockets on the protein surface during the MD simulations and characterized them in terms of the druggability score using a supervised machine learning method developed in the laboratory [41,42]. We then performed unsupervised hierarchical classification on the large set of estimated pockets using pocket similarity in terms of residue composition. Main pocket clusters regrouping pockets with a similar residue composition correspond to binding sites, frequently observed along DM simulations. Analysis of

these pocket clusters in terms of residue localization, frequency, variability, and druggability allowed the identification of promising druggable pockets and of their key residue in terms of contribution to the binding site and its druggability. Additionally, the potential mechanistic impact of mutations of SARS-CoV-2 variants of concern on the selected druggable binding sites located in key regions of the RBD surface and hot spots is discussed.

2. Materials and methods

2.1. Protein preparation

Three crystal structures of the SARS-CoV-2 RBD complexed with the human ACE2 receptor were considered (PDB IDs 6VW1 [25], 6M0J [26], and 6LZG [27]). Overall, the three structures are very similar, with a maximum backbone root mean square deviation (RMSD) of 1.25 Å. However, 6VW1 is based on a chimeric SARS-CoV-2 RBD protein and displays poorer resolution than the 6M0J and 6LZG structures. Coordinates of 6M0J and 6LZG crystal structures were determined at 2.45 Å and 2.50 Å, respectively. Therefore, our study focused only on the 6M0J and 6LZG structures. Each complex was protonated at physiological pH (7.4) using the PROPKA tool [43].

2.2. MD simulations

MD simulations were performed on these two crystallographic structures to study the dynamic behavior and stability of the ACE2-RBD complex. They were also performed to study the dynamic behavior of the unbound state of the RBD initially extracted from the ACE2-RBD complex.

MD simulations were performed using the GROMACS software package [44] with the CHARMM36m all-atom additive protein force field [45] under periodic boundary conditions. A dodecahedron water box of TIP3P water molecules was used to run the simulations. The full simulation system consisted approximately of 12,500 atoms. Non-bonded interactions were truncated at a cut-off distance of ten Å for electrostatic twin-range cut-off and van der Waals cut-off. Neighbor searching was performed every 10 steps. The energy of the system was minimized over approximately 1000 steps using the steepest descent algorithm for energy minimization after the ion addition and neutralization of the systems. Each system was equilibrated with an NVT (Number of particles, Volume, and Temperature) ensemble during 1 ns at a temperature of 300 K and a coupling constant of 0.1 ps. Subsequently, each simulation was performed under number of particles, pressure, and temperature conditions, coupling the system to a heat bath using the Berendsen algorithm and setting the temperature at 300 K and the pressure at 1 bar during 1 ns. For the production step, ten independent runs of 100 ns with different initial velocities were performed. The LINCS algorithm [46] was applied to all the bond lengths to constrain them, and an integration time step of 2 fs was used. The electrostatic interactions were cut off at 1.2 nm and were calculated using the particle mesh Ewald method. As for Lennard-Jones interactions, they were cut off by 1.2 nm by using the potential shift Verlet method [47]. Periodic boundaries conditions were also applied to all systems. The temperature was maintained at 300 K with a V-rescale thermostat [48] and τ_T coupling constant of 0.1 ps. The pressure was maintained at 1 bar with a τ_P constant of 2.0 ps and a compressibility of 4.5×10^{-5} bar^{-1}using the Parrinello-Rahman barostat [49].

For each PDB structure, the trajectories were merged for the analyses, resulting in a total of 1 μs. The RMSD analysis measures the average distances between the starting structure and each structure obtained over time. The root-mean-square fluctuation (RMS Fluctuation) was analyzed to identify flexible residues.

Trajectory analysis was performed using the GROMACS tool (v 2019.5). They were visualized using VMD (v. 1.9.4a38) [50]. The plots were generated using R (v. 3.1.1) [51] and GNUPLOT (v. 5.2) [52]. Figures were generated using PyMOL [53].

2.3. Identification of hot spot residues

The MM-PBSA method is widely used for binding free energy calculations from conformations extracted from the MD trajectory [54]. In this study, MM-PBSA calculations were performed on MD simulations of the RBD-ACE2 complexes (6M0J and 6LZG). The representative conformations of each independent MD simulation were extracted through cluster analysis using the GROMACS gmx cluster tool. The frames were selected in such a way as to cover a wide range of trajectories and to sample different conformational spaces of the complex. Accordingly, representative frames of each of the determined clusters were extracted for every MD simulation using the Gromos algorithm [55] and the best optimal backbone RMSD cutoff was selected. To choose a reasonable cut-off value for each trajectory, we varied the backbone RMSD cutoff between 0.9 and 1.5 Å. Thereby, we found dominant clusters that captured > 70% of the trajectory for each MD simulation. Thus, for every dominant cluster (seven clusters/MD), we extracted one representative conformation, which was subjected to MM-PBSA calculations.

A free energy decomposition analysis was performed using MM-PBSA residue decomposition to retrieve the contribution energy of each amino acid represented on the binding interface of both RBD and ACE2. The total free energy and its individual components for each individual system (6M0J and 6LZG) were averaged and weighted based on the cluster populations, that is, a higher weight was assigned to the conformations extracted from more populated clusters.

We used the *g_mmpbsa* package for MM-PBSA computations [56]. The dielectric relative constant ε was set to eight for proteins and 80 for water [57]. In this approach, the binding free energy ΔG_{bind} between the protein and ligand/protein includes different energy terms and can be calculated as follows:

$$\begin{aligned}\Delta\text{Gbind} &= \text{Gcomplex} - (\text{Gprotein} + \text{Gligand}) \\ &= \Delta\text{EMM} - \text{T}\Delta\text{S} + \Delta\text{Gsol} \\ &= \Delta\text{Evdw} + \Delta\text{Eelec} - \text{T}\Delta\text{S} + \Delta\text{GPB} + \Delta\text{GSA}\end{aligned}$$

where ΔE_{MM} is the gas-phase interaction energy, which is the sum of the van der Waals energy ΔE_{vdw} and the electrostatic energy ΔE_{elec}. ΔG_{sol} is the sum of the polar solvation energy ΔG_{PB} and nonpolar solvation energy ΔG_{SA}. The polar solvation energy was calculated using the Poisson Boltzmann (PB) approximation model, whereas the nonpolar solvation energy was estimated using the solvent-accessible surface area (SASA). The entropy contribution (− TΔS) was ignored in this study because of its high computational cost. After the calculation, the binding free energies were decomposed into each residue. We considered hot spots as residues with binding energy below − 1 (favorable) or above + 1 $kcal.mol^{-1}$ (unfavorable). It is important to note that the more negative the energy, the more favorable is the contribution. In contrast, positive energy values indicate unfavorable interactions and a poor contribution to the complex.

2.4. Pocket estimation, druggability prediction and tracking along the MD simulations

The estimation of the RBD pockets was first performed on the static structures from the PDB.

Pockets were estimated with PockDrug tool based on the automated geometry-based method from Fpocket [36] independent of ligand proximity information. The ensemble of pockets was then characterized in terms of 19 physicochemical and geometrical

properties and druggability score using PockDrug [41,42]. PockDrug provides a prediction of the druggability score which, if greater than 50%, indicates a druggable pocket.

Secondly, pocket tracking was run on the sample of MD generated conformations of the RBD unbound and bound states. This allows the identification of new pockets during dynamics or changes in pocket properties between the PDB conformations or conformations observed during MD. This approach can also detect some separation or fusion of pockets resulting from both local and global alterations, including those occurring in transient and allosteric pockets.

To consider the flexibility of the RBD, we estimated the pockets from conformations obtained from the RBD MD simulations. A total of 1000 RBD conformations were sampled from MD simulations at regular intervals. RBD pockets estimated from this series of 1000 conformations were merged into a pocket set.

2.5. Clustering of pockets and identification of druggable binding sites

We performed unsupervised hierarchical classification using residue pocket similarity to identify binding sites frequently observed along MD simulations, corresponding to main pocket clusters.

The similarity of the RBD pockets can be quantified using binary distance based on common residues. A binary distance of 0 corresponds to two identical pockets in terms of residues. Hierarchical classification of the pockets was performed using the binary distance, Ward metric (ward.D2) [58], and R Hclust package [59]. Dendrogram visualizations were performed using the Heatmaps2 package in R39 [51] to illustrate pocket similarity in terms of residues. The dendrogram lengths between the pockets and/or pocket clusters are proportional to their binary distances.

The resulting pocket classification allows the identification of the main clusters of pockets similar in terms of residues composition, corresponding to binding sites frequently observed along MD simulations. The dissimilarity between the pockets or pocket clusters increased with dendrogram lengths between pockets or clusters. Main pocket clusters were compared i) between pocket sets extracted from two PDB IDs (6LZG and 6M0J) to assess the impact of the initial PDB structure and ii) extracted from bound and unbound RBD conformations to assess the influence of ACE2 interaction.

The number of similar pockets within one cluster indicates the frequency of the appearance of its corresponding binding site. The flexibility of the binding site is described by the residue variability of its corresponding pockets. Analysis of pocket clusters was performed in terms of frequency, residue contributions and variability, RBD localization, and druggability scores. Finally, we combined these statistical and flexibility analyses to select binding sites based on the criteria of frequency and residue stability, localization in key regions of the RBD, accessibility, and druggability scores.

We also crossed the binding site analysis with the hot spot information obtained from the MM-PBSA analysis and evaluated the potential structural impact of mutations observed in the five SARS-CoV-2 variants of concern (Alpha, Beta, Gamma, Delta, and Omicron) on these selected hot spots and binding sites to identify the druggable binding sites that can also be targeted in mutated systems.

2.6. Protocol of identification of druggable binding sites and their flexibility

Combining supervised and unsupervised machine learning techniques with a traditional flexibility study through MD simulations and MM-PBSA energetic computation analysis yields a comprehensive protocol. This three-step protocol identify druggable binding sites frequently observed on the RBD protein and their key residues in terms of druggability and contribution to stability (Fig. S1).

3. Results and discussion

3.1. RBD-ACE2 structures

Several PDB structures have elucidated the interaction network between the ACE2 protein and RBD spike protein [25–27]. The RBM is composed of 17 residues (K417, G446, Y449, Y453, L455, F456, A475, F486, N487, Y489, Q493, G496, Q498, T500, N501, G502, and Y505) that are in contact with ACE2 residues to stabilize the complex interaction [60,61]. With a cutoff distance of 4 Å, the complex was maintained with 13–18 hydrogen bonds and a salt bridge (Fig. 1). SARS-CoV-2 forms distinct patches of residues that spread along the interaction surface. Two hydrogen bonds were formed between Y489 and Y83, and between N487 and Q24 (Fig. 1 A). Q493 interacts with both K31 and E35 ACE2 residues through hydrogen bonding and its two crystallized side chains (Fig. 1B). A single salt bridge formed between K417 and D30 (Fig. 1B). On the other side of the surface, two large patches of residues establish strong intra- and intermolecular interaction networks [62]. This network includes hydrogen bonding, hydrophobic, and π–π interactions [63]. The complex is stabilized by hydrogen bonds between G446-Q42, Y449-D38, Q498-Q42, T500-Y41, N501-Y41, N501-N330, G496-K353, Q498-K353, G502-K353, Y505-E37, D355-R357, and Y505-R393 (Fig. 1 C, Fig. 1D, and Fig. 1E). Residues involved in hydrophobic and van der Waals interactions played an important role in the affinity of the complex (Fig. S2). T27 interacted with Y473, A475, and F456 through hydrophobic packing. F28 interacts with Y489, and H34 with Y453 and Q493 through van der Waals interactions. Other hydrophobic patches were found on both protein surfaces, including L45 with V445, G446, and Q98, L79 with F486 and Y489, and Y505 with K353.

3.2. RBD-ACE2 flexibility

Proteins are dynamic molecules with intrinsic flexibility and often undergo conformational changes upon partner binding. Therefore, it is essential to consider their dynamic behavior to predict which surface regions may be of interest.

RBD flexibility was studied with and without ACE2 by MD simulations. Mean Cα RMSDs have been found to stabilize for both bound proteins systems (6M0J and 6LZG) around 2.5 Å with fewer fluctuations (Fig. S3 and Fig. S4). These results highlight the stability of the SARS-CoV-2 RBD and ACE2 complex throughout all the simulations (Fig. S3A and Fig. S4A). The low mean Cα RMSD values indicated that the unbound RBD was stable (Fig. S3D and S4D). No major flexibility variation was observed between the bound and unbound RBD (Fig. S3C, Fig. S3D, Fig. S4C and Fig. S4D), considering both unbound and bound RBD simulations, have not been reported in other studies. To examine the flexibility and local changes in the complex, Cα RMSF versus the residue number of both RBD systems were investigated (Fig. S5). RMSF analysis revealed that the RBM region flexibility increased to a greater extent in the unbound RBD structure than in the bound RBD and increased from approximately 1.5 to 2.0 Å. This was not surprising, as the RBM mediates contact with ACE2, which tends to be more stable when the complex is formed (Fig. S5A and Fig. S5C) than in unbound RBD (Fig. S5B and Fig. S5D).

3.3. Hot spot residue analysis

MM-PBSA calculations were performed to assess the binding free energies of the ACE2-RBD complexes and to estimate the contribution of residues in this interaction. The MM-PBSA calculations were applied to specific frames of representative states that were extracted from the MD trajectories after clustering analysis. Clustering analysis was based on a specific series of 23 residues for the RBD and

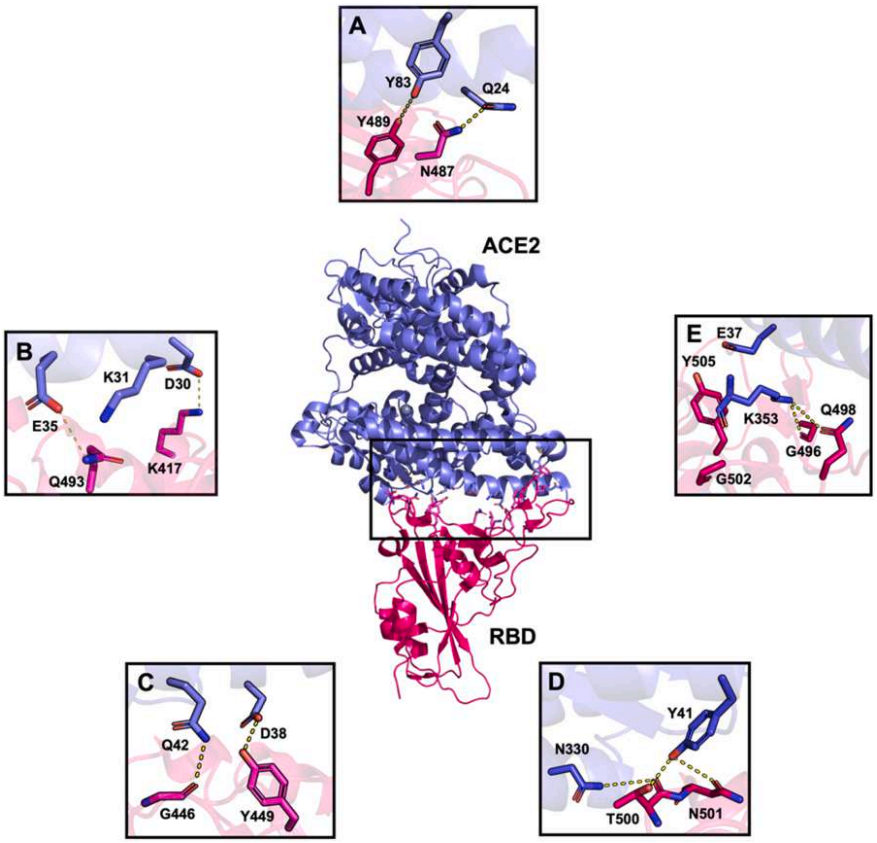


Fig. 1. Interaction networks of the SARS-CoV-2 RBD and ACE2 complex (PDB: 6M0J). (A) Hydrogen bonds with Q24-N487 and Y83-N487. (B) Salt bridge between D30 and K417 and hydrogen bonds with K31-Q493 and E35-Q493. (C) Hydrogen bonds between Q42-G446 and D38-Y449. (D) Hydrogen bonds between Y41-T500, N330-T500, and Y41-N501. (E) Hydrogen bonds between K353-G496, K353-Q498, E37-Y505, and G502-K353. ACE2 is indicated in blue and RBD in pink. Interactions were configured using PyMOL software. Only the side chains of the residues are represented, excluding for glycine residues.(For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

22 residues for ACE2. The residues corresponded to the amino acids spanning the binding interface with a cut-off of 4.0 Å. These residues are K417, G446, G447, Y449, Y453, L455, F456, Y473, A475, G476, S477, E484, F486, N487, Y489, F490, Q493, G496, F497, Q498, T500, N501, and G502 for the RBD and S19, Q24, T27, F28, D30, K31, H34, E35, E37, D38, Y41, Q42, L45, L79, M82, Y83, N330, K353, G354, D355, R357, and R393 for ACE2.

Based on the MM-PBSA calculations, SARS-CoV-2 RBD binds to ACE2 with a ΔG_{bind} of − 41.2 ± 20.2 and − 47.1 ± 21.8 kcal.mol^{-1}, for 6M0J and 6LZG structures, respectively (Table 1). The binding free energies of the two complexes are similar.

Table 1
Average binding free energy and standard deviations in the MD simulations of the RBD-ACE2 complex for PDB IDs 6M0J and 6LZG.

PDB ID	Binding energy components (kcal.mol^{-1})			
	ΔE_{MM}	ΔE_{Polar}	ΔE_{Apolar}	ΔG_{bind}
6M0J	-59.3 ± 3.7	85.1 ± 21.6	-7.7 ± 0.6	-41.2 ± 20.2
6LZG	-59.2 ± 3.9	82.8 ± 22.9	-7.7 ± 0.6	-47.1 ± 21.8

Free energy decomposition analysis was also performed to obtain a detailed insight into the interactions between each residue in the binding interface of the two proteins. The binding interaction of each residue included four terms: molecular mechanics contribution, polar contribution, apolar contribution, and total energy contribution. The individual components of each residue were averaged for each system (6M0J and 6LZG). The hot spot residues on the RBD surface were essentially the same for the core interactions in 6M0J and 6LZG (Fig. 2).

Notably, residues F486 and Y489 inserted into a hydrophobic pocket on the surface of ACE2 formed by residues including T27, F28, L79, and M82. Thus, the presence of aromatic amino acids in the pocket may provide an additional binding force through π-stacking interactions (Fig. 3A) and explain the favorable ΔE_{MM}, which is mainly due to the favorable contribution of ΔE_{vdw}. The residue K417 showed a significant ΔE_{polar} contribution that reaches 11.0 kcal.mol^{-1}. However, the fact that this residue formed a salt bridge and a hydrogen bond with residue D30 led to a more negative ΔE_{MM} contribution of − 14.0 kcal.mol^{-1} (Fig. 3B) and, consequently, it counterbalanced the unfavorable polar solvation energy. Moreover,

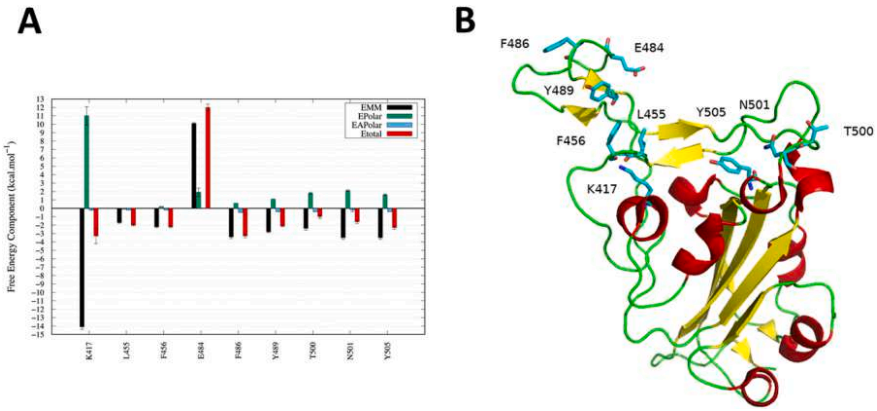


Fig. 2. (A) The energy components of the nine residues on the RBD that contribute significantly to the ΔG of binding with ACE2. The average binding energy components and standard deviations are calculated from the MD simulations of the RBD-ACE2 complex (PDB IDs 6M0J and 6LZG). (B) RBD displayed in cartoon representation. The hot spot residues are indicated by cyan sticks.

ACE2 residue Y41 interacted with Q498 and Y505 through van der Waals interactions and formed two hydrogen bonds with residues T500 and N501 through its side chain (Fig. 3C). Although ACE2 residue H34 was not directly involved in any intermolecular hydrogen bond with RBD, it nonetheless provided favorable polar contacts with the side chain of L455 of the virus protein (Fig. 3D).

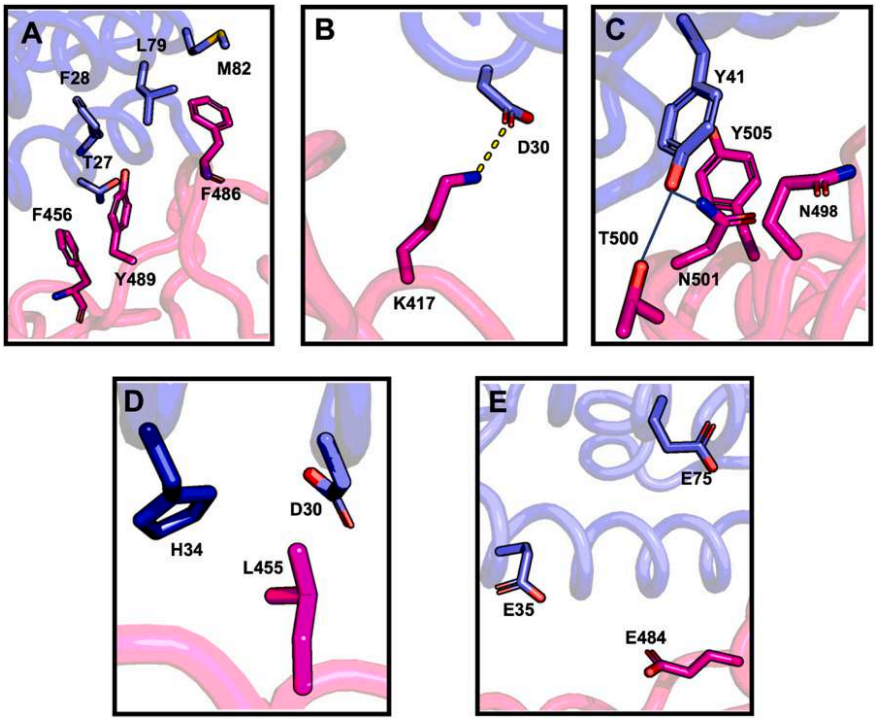


Fig. 3. Contact residues of ACE2-RBD interface issued from MM-PBSA analysis. ACE2 is colored in blue and RBD is indicated in pink. Amino acids are represented by sticks and colored based on their respective proteins. A) Hydrophobic pockets, including F456, F486, and Y489, B) K417 and D30 salt-bridge, and C) ACE2 Y41 interacts with T500 and N501, D) L455 and H34 interactions, and E) E484 negative repulsive charges with E35 and E75. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

L455 also formed intermolecular van der Waals interaction with D30 (Fig. 3D). Interestingly, E484 displayed a large positive E_{MM} contribution that reaches 10.0 kcal.mol^{-1} (Fig. 2). Thus, E484 free energy contribution (E_{TOT}) was found to be disfavoring complex binding and formation. Two negatively charged residues, E35 and E75, surrounded E484, which creates long-range opposing repulsive charges and explains the unfavorable contribution of this residue to the complex (Fig. 3E).

Altogether, our analysis shows that among the residues comprising the ACE2-RBD interface, some critical amino acids are known as the evolutionary signature of RBD, and major hot spots of both proteins contribute favorably to complex stability and formation [27]. In conclusion, eight favorable hot spot residues (K417, L455, F456, F486, Y489, T500, N501, and Y505) and one unfavorable hot spot residue (E484) have been identified as important SARS-CoV-2 residues that are critical for ACE2 binding.

These results are in good agreement with other studies in which the authors identified 10–13 hot spot residues using similar bioinformatics methods on different RBD structures, that is, X-ray structures or models built by homology modelling, from SARS-CoV-1 or SARS-CoV-2 [29–31,64–66].

3.4. Tracking of RBD pockets during the MD simulations

We first examined pockets that were extracted from the unbound RBD system from static PDB. From the initial 6M0J and 6LZG PDB IDs, we estimated four and seven pockets on the unbound RBD, respectively, using the PockDrug program [41,42]. This is consistent with the observation of a variable number of pockets estimated by Trigueiro-Louro et al. on the same PDB structures [32].

We secondly examined pockets that were extracted from the unbound RBD systems considering their flexibility. We extracted two sets of pockets from two datasets of 1000 conformations sampled from each of the unbound RBD structures 6M0J and 6LZG during the MD simulations. The average number of RBD pockets by conformation in these two sets were highly similar (5.1 versus 5.8) despite the different number of pockets estimated on the two static PDBs. The frequency of druggable pockets (50.4% versus 54.0%) and the frequency of pockets located at the interface (31.0% versus 25.0%) were also highly similar within both 6M0J and 6LZG MD trajectories. The consistency of the results obtained from the MD simulations, regardless of the initial structure, shows the importance of integrating the protein and pocket structural flexibility for binding site identification and selection.

The results were consistent for both the 6M0J and 6LZG PDBs; therefore, we decided to focus only on the 6M0J MD trajectory. The average number of 5.1 pockets in the RBD conformation (from one to a maximum of nine pockets) results in a total set of 5065 pockets. Of these pockets, 50.4% were predicted to be druggable (druggability score ≥ 50%), and 40.6% of these pockets were highly druggable (druggability score ≥ 75%). Nearly one-third of the pockets (31.0%) included at least one of the 17 main RBM interacting residues and were located at the RBD-ACE2 interface. We also combined pocket analysis with the hot spot results obtained from the MM-PBSA study to identify the druggable binding sites containing these hot spots. Interestingly, almost all of these interface pockets (98.5%) included at least one of the nine hot spot residues (identified in Section 3.3). This is consistent with the fact that hot spot residues involved in energetic interactions have physicochemical properties and locations that promote pocket accessibility.

To determine how ACE2 affects RBD conformation and binding regions, we subsequently examined pockets extracted from the bound RBD system. The same pocket tracking protocol was applied to 1000 RBD conformations extracted from MD simulations of 6M0J in the RBD-ACE2 complex form. Therefore, a set of 4428 pockets with a pocket druggability frequency of 50.1% was extracted from the 6M0J RBD-ACE2 complex trajectory. This is close to the pocket druggability frequency of 50.4% of the 5065 pockets extracted from the unbound RBD trajectory. Fewer pockets (approximately 640 pockets) were observed in the RBM region of bound RBD. This can be explained by the presence of ACE2, which limits pocket formation in the RBM region. These results clearly show that consideration of the flexibility of the structure leads to the identification of comparable and replicable RBD pockets when utilizing isolated or complex forms, with the exception of the area in contact with ACE2. Comparable steady results were also obtained using 6LZG MD simulations (not shown).

3.5. RBD pockets clustering based on pocket residue similarity

After quantifying the druggable pockets on the two different systems, we performed a hierarchical classification of the pockets extracted from the 6M0J unbound RBD trajectory based on residue similarity. This classification of pockets based on residue composition allows the identification of clusters of pockets with common residues, corresponding to binding sites frequently observed along the MD simulations. Furthermore, the variability in residues within a pocket cluster illustrates the residue flexibility of the corresponding binding site. Highly similar classification results were obtained for pocket sets extracted from MD simulations of the 6M0J and 6LZG PDB. Here, the classification obtained for the 5065 pockets extracted from the 6M0J unbound RBD trajectory is presented. This classification resulted in the identification of eight main pocket clusters, enumerated as clusters I to VIII (Fig. 4). We analyzed these main clusters in terms of frequency, residue localization, stability, and druggability scores.

These eight clusters were observed in at least one-third of the 1000 conformations, and did not correspond to rare pockets (Table S2). They were regularly observed during the MD simulations. The least frequent (cluster VI) was observed in 32.3% of the conformations, whereas the most frequent (cluster I) was observed in approximately 90% of the conformations.

Moreover, these eight clusters were well-characterized in terms of druggability (Fig. 4 and Table S2). Four clusters were mainly druggable (II, III, V, VII, including 68.9–99.2% of druggable pockets), whereas four others identified clusters included a weak or moderate part of druggable pockets (I, IV, VI, VIII including 0.2–27.9% of druggable pockets). Interestingly, each cluster was defined by specific residues that constituted pockets that were well differentiated and localized in different regions of the RBD (Fig. 4). Clusters I, II, III, IV, and VI regrouped very similar pockets, as indicated by the small distances between the pockets within each cluster. Clusters V and VIII showed less pocket similarity in terms of residues, and each could be split into two main homogeneous sub-clusters. Cluster VII was the most variable and included three main subclusters (Fig. 4).

3.6. Selection of RBD druggable binding sites

A more detailed analysis was performed to study the properties and localization of the eight main clusters resulting from RBD pocket classification (Fig. 4) in order to extract druggable binding sites of interest. Six clusters (II to VII) were distant from the RBD-ACE2 interface, and two clusters (I and VIII) were located close to the RDB-ACE2 interface.

Two clusters (IV and VI) were moderately observed in the 1000 RBD conformations (48.8% and 32.3%) and were weakly druggable (27.9% and 7.4%, respectively). The cluster pockets corresponded to two exposed protein cavities in both the closed and open states of the spike protein trimer. These cavities did not establish contact with the ACE2 interface or another RBD monomer. Consequently, they did not form valuable target sites. The other four clusters (II, III, V, and VII) were observed in approximately half or more of the 1000

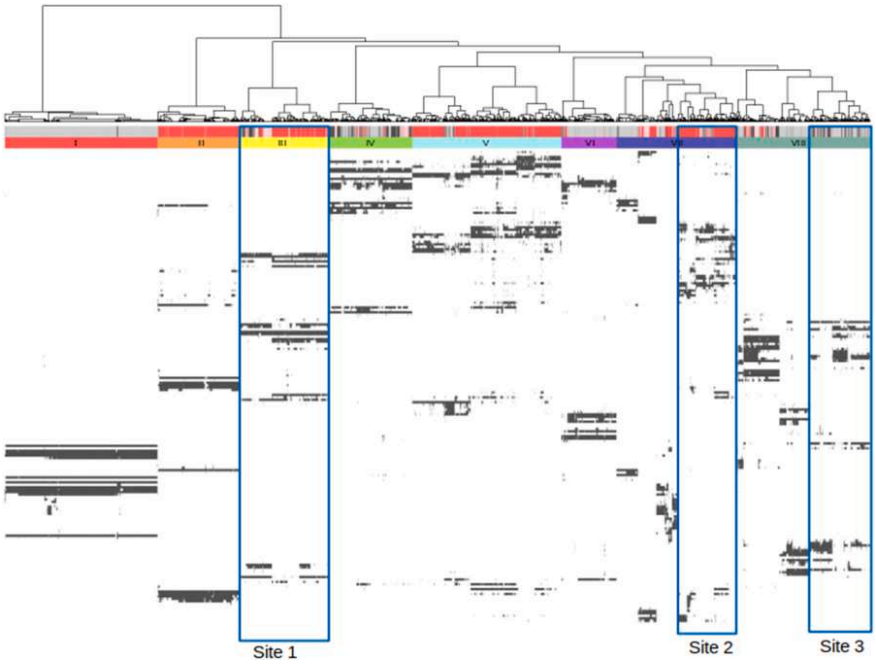


Fig. 4. Hierarchical classification of 5065 RBD pockets extracted from 1000 conformations of unbound RBD (6M0J). Each column corresponds to a pocket. First-color bars are colored according to four levels of druggability scores: respectively in black, grey, pink, and garnet for the non-druggable [0.00, 0.25], less druggable [0.25, 0.50], moderately druggable [0.50, 0.75] and highly druggable [0.75, 1.00] pockets, as predicted using PockDrug [41]. Second-color bars are colored based on the eight main clusters of pockets (noted I to VIII). At the bottom, the residues are ordered according to the numbering of the protein sequence and indicated in black when involved in the pocket. The three blue rectangles indicate the three sub-cluster of pockets corresponding to selected sites of interest.(For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

conformations and included mainly druggable pockets (> 68%). Cluster II was frequent (47.9%) and mainly druggable (99.2%). However, its pockets were located between the two spike protein interfaces in both states and are thus buried. Thus, this cluster cannot be targeted by therapeutic molecules. Cluster V was highly frequent (87.3%) and most of its pockets were druggable (93.0%). The corresponding pockets were highlighted as sites of interest in a previous study [33]. However, this cluster included N343, which is covalently linked to N-acetylglucosamine in the RBD. Notably, N-acetylglucosamines were not used in the MD simulations in this study. Therefore, even without this information, our protocol could identify a region matching an N-linked glycosylation site. We do not consider this cluster of pockets a priority to be targeted by therapeutic molecules due to the presence of N-acetylglucosamine.

Cluster III was observed in half of the conformations (52.4%) and mainly included druggable pockets (85.9%) (Table S2). The pockets were located on the interface region between two RBD proteins in the closed state of the spike protein trimer (Fig. S6). This cluster regroups similar pockets (with 17 residues observed in more than 15% of its pockets) (Table S2). It is noteworthy that Carino et al. highlighted one potential druggable pocket from the static spike protein trimer (PDB ID:6VSB), but it corresponded only to a partial sub-pocket of ten residues of our cluster of pockets [33]. Therefore, it may be of interest to design molecules that bind this cluster of 17 residue pockets to stabilize the closed state and avoid the activation of the spike protein trimer. This frequent and druggable cluster, named site 1 (Table 2), is a potential site of interest for the design of new therapeutic molecules.

Table 2
The three selected potential binding sites with their residue composition (only residue observed in more than 15% of their associated pockets), their occurrence on 1000 RBD conformations and the percentage of druggable pockets within each site. Hot spot residues are in bold.

Selected sites (sub-Cluster)	Residues	occurrence	Druggable pockets frequency (%)
Site 1 (Cluster III)	S375, T376, K378, Y380, G404, E405, V407, R408, I410, A411, Q414, V433, A435, V503, G504, Y508, V510	52.4	85.9
Site 2 (sub-cluster VII)	A363, D364, Y365, S366, L368, Y368, N370, F374, F377, C379, V382, P384, T385, L387, N388, D389, L390, C391, F392, C432, I434, L513, F515, V524, G526	35.0	87.7
Site 3 (sub-cluster VIII)	R403, D405, E406, R408, Q409, T415, G416, **K417**, I418, Y453, **L455**, S494, Y495, G496, F497, Q498, **N501, Y505**	35.0	13.1

Cluster VII was frequent (70.4%) and included pockets mainly structurally located at the bottom end of the RBD and partially buried in the spike protein trimer. However, this cluster regroups pockets that vary in terms of residues; thus, cluster VII can be decomposed into four main pocket sub-clusters. Interestingly, the most frequent sub-cluster, named site 2, which was also observed in 35% of the conformations, corresponded to homogeneous and druggable pockets (87.8%) (Table 2). The pockets point toward another spike protein monomer and are consequently partially buried. This sub-cluster coincides with the pocket identified by Toelzer et al., in which linoleic acid is bound [67]. Linoleic acid appears to stabilize the spike protein in its closed state and induces a reduced ACE2 interaction in vitro. Therefore, our protocol was successfully identified and characterized in terms of druggability and residue stability in this specific region, called site 2, which corresponds to the experimentally known binding site entrance of a small molecule.

We also extracted two frequent clusters, I and VIII, which were located close to the RBD-ACE2 interface. Cluster I was quasi permanent (86.6%), although it displayed no druggability (0.2%). It contains nearby pockets located on the RBD loop between residues T470 and P491 and thus excluded the association of every RBM region. Moreover, cluster I pockets were exposed to the solvent in both the closed and open states of the spike protein. Due to its location and undruggable properties (Table S2), cluster I is not suitable for targeting therapeutic molecules.

Our results showed that cluster VIII was frequent (78.0%) and partially druggable (21.7%). It is relatively variable and regroups three main subclusters located at the ACE2 interface. The most frequent sub-cluster, named site 3, was observed in 35.0% of the conformations and included 13.1% of druggable pockets (Table 2). It is centralized at the interface and includes three hot spots, K417, N501, and Y505, among the 18 residues constituting this pocket. Consequently, site 3 is particularly appealing as a targeted therapeutic molecule.

Considering their occurrence, residue similarity, druggability score, and localization, three sites of interest (sites 1, 2, and 3) were selected for further study (Fig. 5).

These sites are observed in more than 30% of the conformations and are associated with an average druggability score range of 13.1–87.7%. The pockets of these three sites included 20 residues on average (Table 2). The contribution of residues to their pocket cluster, as well as the druggability score, was made available for the three selected sites in Fig. S7. Along with these results, we have shown that the two most frequent sites, 1 and 2, are distant from the RBM, even though they display significant druggable scores. In contrast, site 3 has been classified as less frequent yet druggable.

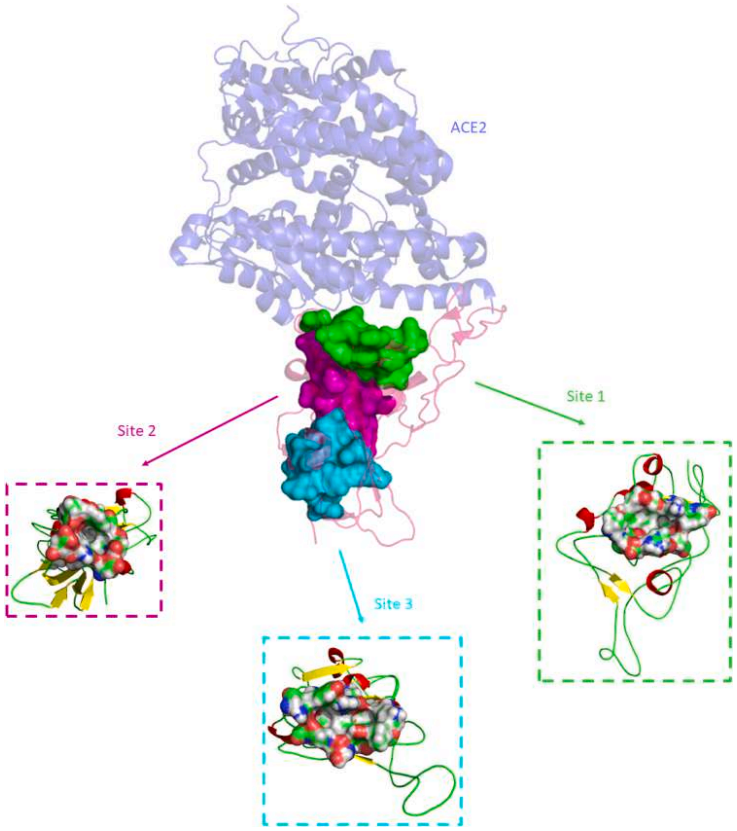


Fig. 5. Representation of the three sites of interest selected on the RBD surface (site 1 represents the trimerization pocket, site 2 indicates a highly druggable pocket capable of holding a drug candidate, and site 3 corresponds to the interface region pocket).

Compared to the other sites, an important feature of site 3 is its location, which is close to the RBD-ACE2 interface, and it includes three highly contributing hot spots: K417, N501, and Y505.

Dokainish et al. identified several pockets with residues overlapping with those of binding site 3 [40]. This site was targeted by nilotinib. The latter study, as many others, validates our protocol because we identified the same site. This site includes four hot spot residues, suggesting the significance of targeting it to disrupt the interaction between RDD and ACE2.

Ten representative pockets were selected for each of the three sites to illustrate the binding site and their flexibility (pocket PDB, residues, and global properties) that could be used for the design of novel compounds that target the spike protein. Table S3 provides descriptions of physicochemical, geometrical, and druggability scores. The ten selected pockets and their corresponding RBD conformations are freely available in the PDB format (https://data.mendeley.com/datasets/xhnjgfbgzr/1).

3.7. Potential impact of mutations on RBD hot spot and binding sites

It is known that the mutations associated with five variants of concern, Alpha, Beta, Gamma, Delta, and Omicron, cause a substantial increase in their transmissibility, virulence, and antigenic escape ability. During the evolution of the SARS-CoV-2 genome, RBD mutations had an impact on spike protein affinity for ACE2 to improve virus transmissibility and/or decrease its affinity for antibodies to improve its antigenic escape ability. The Alpha variant includes only one mutated residue, N501Y, whereas Beta, Gamma, and Omicron variants include both N501Y and K417N/T and E484K/A mutated residues. The Omicron variant also has 12 supplementary mutated residues (G339D, S371L, S373P, S375F, N440K, G446S, S477N, T478K, Q493R, G496S, Q498R, and Y505H). The Delta variant includes two specific mutations (L452R and T478K).

It is of interest to analyze the potential structural impact of RBD mutations of these variants of concern, notably, relative to the nine hot spots we have identified. These mutations directly impacted sites 1 and 3 and three hot spot residues, which correspond to the most frequent mutations, from site 3 (Table 3).

Two mutations (N501 and K417) corresponded to hot spots and played an important role in the stabilization of the interaction between RDB and ACE2. This confirms that these mutations can directly affect the affinity of the RBD for ACE2 and the interactions with antibodies targeting the RBM. The mutation N501Y is observed in Alpha, Beta, Gamma, and Omicron variants, suggesting that it provides SARS-CoV-2 with a selective advantage. E484 and K417 mutations are present in the N501Y mutation in the Beta, Gamma, and Omicron variants. The N501Y mutation counterbalances the decrease in ACE2 affinity due to mutations in K417 and E484, whereas the latter tends to enhance the ability of the variant to escape from neutralizing antibodies. This explains why these three mutations were selected simultaneously during the SARS-CoV-2 evolution. More specifically, the N501 hot spot is stabilized by a hydrogen bond with residue Y41 and contributes favorably to the binding free energy of the complex (Fig. 1D and Figs. 2A and 3C). It is surrounded by a K353 hydrophobic alkyl chain and Y41 benzene ring. Understandably, a tyrosine switch is a better choice because a more favorable interaction can be made with the hydrophobic pocket, particularly π-π stacking with residue Y41. Thus, the mutation of N501 in Y501 can explain the enhanced affinity of the RBD to the ACE2 receptor [68].

Table 3
Mutations in RBD observed in the SARS-CoV-2 variants of concern. The hot spot residues identified by the MM-PBSA analyze are in bolt.

WHO label	Mutations in RBD	Impact on three selected sites
Alpha	**N501Y**	Site 3 (**N501Y**)
Beta	**K417N**, **E484K**, **N501Y**	Site 3 (**K417N**, **N501Y**)
Gamma	**K417T**, **E484K**, **N501Y**	Site 3 (**K417N**, **N501Y**)
Delta	L452R, T478K	No site
Omicron	G339D, S371L, S373P, S375F, **K417N**, N440K, G446S, S477N, T478K, **E484A**, Q493R, G496S, Q498R, **N501Y**, Y505H	Site 3 (**K417N**, **N501Y**, G496S, Q498R, Y505H) Site 1 (S375F)

E484 mutations are present with N501Y and K417 mutations in the Beta, Gamma, and Omicron variants. In the Beta and Gamma variants, the E484K mutation swaps the charge of the side chain. It is a considerable switch from negatively charged glutamic acid (−) to a positively charged lysine (+). In the Omicron variant, the E484A mutation also induces loss of the negative charge carried by the E residue. It is important to note that E484 is located on the RBD loop (amino acids 470–490) and is enclosed by several negatively charged residues, E35 and E75, from the ACE2 receptor. The position of three neighboring glutamic acids could be unfavorable due to repulsive forces; thus, introducing a positively charged residue or a small neutral residue might create a more favorable interaction between the two proteins. This may explain the unfavorable free energy decomposition of E484 during MM-PBSA analysis (Fig. 2 A). Similar to N501, E484 is a critical epitope residue for SARS-CoV-2 neutralizing antibodies. Therefore, charge change can also be a method to alter the electrostatic complementarity of known antibodies binding to this region, leading to better virus adaptation [69].

Concerning the K417 mutations, some studies showed that they may be responsible for increased binding with ACE2 and a decreased affinity for SARS-CoV-2 antibodies when combined with N501Y and E484K [70]. K417 forms a steady salt-bridge with residue D30 and has a favorable energy contribution to the complex (Figs. 1B, 2 A, and 3B). Consequently, abolishing this strong interaction decreases the binding affinity of the RBD and ACE2 complexes. However, the K417 mutation is only present with the N501 and E484 mutations, which may compensate for the loss of affinity with the ACE2 receptor by forming new favorable interactions. Additionally, a recent study showed that K417 is another critical epitope that forms strong salt bridges with SARS-CoV-2 antibody residues [68]. Thus, the reason behind the K417 abrupt change to asparagine is that viral adaptation is vital and overrides the binding affinity. In fact, with this triple mutation, SARS-CoV-2 may be harder to handle and can easily escape antibodies. The explanation for the K417 change to N or T may be a viral adaptation to decrease the affinity of antibodies for spike protein and evade the immune system. Regardless of the reasons for the K417 mutation, the replacement of a residue with a positive charge by residues with a hydrophilic side chain should be considered for drug design.

These mutations directly impacted three hot spot residues, which corresponded to the most frequent mutations. Thus, we analyzed the potential structural impact of RBD mutations of these variants on the three selected binding sites. These mutations directly affected sites 1 and 3. Additionally, site 3 would be greatly impacted by three hot spot residues, which correspond to the most common mutations in the variants of concern (Table 3).

For the Delta variant, L452 and T478 mutations did not occur in residues forming the highlighted sites and were far from the hot spot residues, indicating that inhibitors targeting these sites would not require adaptation to treat this variant.

The Omicron variant was found to be highly mutated. For example, 15 mutated residues were found in the RBD region alone (G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, and Y505H). Only the Omicron S375F mutation has been observed to impact site 1 (Fig. 6A). Currently, there is no available scientific information on how this mutation affects the SARS-CoV-2 life cycle. S375 was exposed to the solvent in both the open and closed states of the spike protein trimer

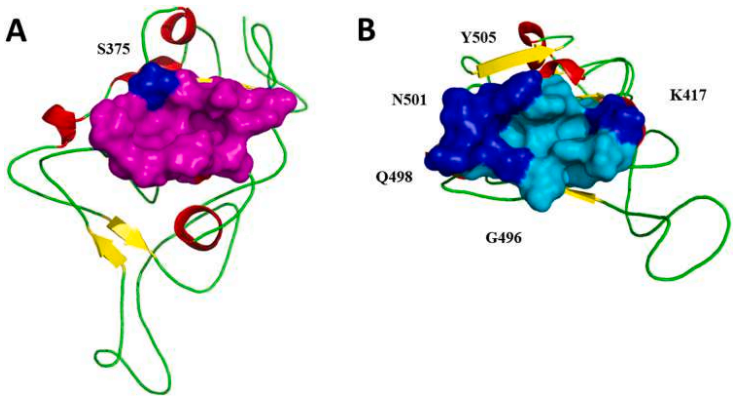


Fig. 6. Mutations, indicated in blue, observed in the variants of concern in A) site 1 and B) site 3. S375 mutation occurs in the Omicron variant. N501 mutation is observed in the Alpha, Beta, Gamma, and Omicron variants; K417 mutation is observed in the Beta, Gamma, and Omicron variants. G496, Q498, and Y505 mutations are specific to the Omicron variant.(For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and did not form a non-covalent bond with any other residue. As a result, it does not appear to play a special role in the spike protein trimer action. Because the alcohol group in the spike residue was replaced by an aromatic cycle in the F residue, the S375F mutation only slightly altered the shape and hydrophobic properties of site 1. Therefore, site 1 seems to be relatively well-conserved among the variants of concern.

In summary, site 3 is located on the RBD-ACE2 interaction interface and is formed by several hot spot residues; therefore, it may be of interest to focus on this region with the aim of targeting it with inhibitory molecules.to focus on this region with the aim of targeting it with inhibitory molecules. However, because of several mutations that occur in different variants of concern, site 3 is difficult to block, in contrast to sites 1 and 2. Even though site 1 seems to be weakly affected by only one mutation, site 2 is currently not affected by any mutation. These two sites can be targeted by inhibitors that would be effective for all currently known variants.

Sites 1 and 2 are contiguous on the RBD surface. Interestingly, a neutralizing antibody isolated from convalescent Covid-19 patients and named CR3022, can bind to the RBD surface corresponding to sites 2 and 3 [71]. This interaction occurs only when two of the three spike proteins in the spike protein trimer are in an open state. Therefore, these sites are accessible to the inhibitory molecules. In the case of mutations that prevent the binding of neutralizing antibodies, targeting these sites by inhibitory molecules seems to be a promising therapeutic approach; notably, site 2 is currently not affected by any mutation.

4. Conclusions

In this study, we analyzed the dynamic behavior of the unbound RBD in complex with the ACE2 protein to identify binding sites that can be targeted by inhibitory molecules. We proposed a protocol combining MD simulations, hot spot identification, pocket tracking along the simulations, and pocket druggability prediction using a supervised method. Then, an unsupervised machine learning analysis was applied to cluster similar RBD pockets in terms of residue composition. This protocol allows the identification of druggable binding sites frequently observed along the MD simulations, the characterization of their residue flexibility and of key residues contributing to the druggability.

The stability of these pocket clusters was verified on different PDBs, in apo and holo forms, which allowed us to select the most pertinent druggable binding sites. Based on their RBD localization and druggability assessments, these three sites seem to be particularly promising. The potential effect of mutations in the variants of concern on these three binding sites was investigated.

Site 3 is located at the RBD-ACE2 protein interaction region and is formed by four hot spot residues. It is therefore an interesting target to disrupt the interaction between the spike protein and ACE2 protein and, consequently, to prohibit SARS-CoV-2 entry into the cell. However, several mutations of residues forming this site have been observed in SARS-CoV-2 variants. This suggests that if an inhibitory molecule can be designed against this site, it should be efficient for only a limited number of variants. Site 2, in which linoleic acid interacts to lock the spike protein in closed form, is relevant to the target. It is highly druggable, undergoes only one mutation in the Omicron variant, and is consequently an interesting site for targeting. Site 1 is highly druggable, accessible and no mutations were observed at this site. To our knowledge, this is the first time it has been characterized using in silico methods. In the spike protein trimer, we can observe that three sites 1 observed on each spike protein monomer are near in space (Fig. S6). This is also a promising target region. This suggests that the design of molecules able to bind to at least two sites 1 located within two RBD monomers may prohibit the transition between the inactive closed form and the active open form of the spike protein.

In summary, our combined protocol provides new insights and highlight opportunities on three binding sites for the development of inhibitors of the RBD of the spike protein.

Ethical approval

Not applicable.

CRediT authorship contribution statement

Mariem Ghoula: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Sarah Naceri:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Samuel Sitruk:** Formal analysis, Investigation. **Delphine Flatters:** Conceptualization,

Methodology, Validation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **Gautier Moroy:** Conceptualization, Methodology, Validation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Anne-Claude Camproux:** Conceptualization, Methodology, Validation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

All authors declare no conflicts of interest.

Acknowledgments

This work was supported by the Agence Nationale de la Recherche (PIF21 Project no. ANR-19-CE18–0023). The authors gratefully acknowledge the financial support of the Université Paris Cité, CNRS Institute, and INSERM Institute. This study was performed using HPC resources from GENCI-CINES. The authors are thankful to Patrick Fuchs, Audrey Deyawe Kongmeneck, and Rachel Blot for their helpful comments on this manuscript.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2023.03.029.

References

[1] Coronaviridae Study Group of the International Committee on Taxonomy of VirusesThe species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2 Nat Microbiol 5 2020 536 544 doi: 10.1038/s41564-020-0695-z.
[2] Chen L, Liu W, Zhang Q, Xu K, Ye G, Wu W, Sun Z, Liu F, Wu K, Zhong B, Mei Y, Zhang W, Chen Y, Li Y, Shi M, Lan K, Liu Y. RNA based mNGS approach identifies a novel human coronavirus from two individual pneumonia cases in 2019 Wuhan outbreak. Emerg Microbes Infect 2020;9:313–9. https://doi.org/10.1080/22221751.2020.1725399
[3] Xie X, Liu Y, Liu J, Zhang X, Zou J, Fontes-Garfias CR, Xia H, Swanson KA, Cutler M, Cooper D, Menachery VD, Weaver SC, Dormitzer PR, Shi PY. Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera. Nat Med 2021;27:620–1. https://doi.org/10.1038/s41591-021-01270-4
[4] Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, Doolabh D, Pillay S, San EJ, Msomi N, Mlisana K, von Gottberg A, Walaza S, Allam M, Ismail A, Mohale T, Glass AJ, Engelbrecht S, Van Zyl G, Preiser W, Petruccione F, Sigal A, Hardie D, Marais G, Hsiao NY, Korsman S, Davies MA, Tyers L, Mudau I, York D, Maslo C, Goedhals D, Abrahams S, Laguda-Akingba O, Alisoltani-Dehkordi A, Godzik A, Wibmer CK, Sewell BT, Lourenço J, Alcantara LCJ, Kosakovsky Pond SL, Weaver S, Martin D, Lessells RJ, Bhiman JN, Williamson C, de Oliveira T. Detection of a SARS-CoV-2 variant of concern in South Africa. Nature 2021;592:438–43. https://doi.org/10.1038/s41586-021-03402-9
[5] Starr TN, Greaney AJ, Dingens AS, Bloom JD. Complete map of SARS-CoV-2 RBD mutations that escape the monoclonal antibody LY-CoV555 and its cocktail with LY-CoV016. Cell Rep Med 2021;2:100255. https://doi.org/10.1016/j.xcrm.2021.100255
[6] Giovanetti M, Benvenuto D, Angeletti S, Ciccozzi M. The first two cases of 2019–nCoV in Italy: Where they come from? J Med Virol 2020;92:518–21. https://doi.org/10.1002/jmv.25699
[7] Sallam M, Ababneh NA, Dababseh D, Bakri FG, Mahafzah A. Temporal increase in D614G mutation of SARS-CoV-2 in the Middle East and North Africa. Heliyon 2021;7:e06035. https://doi.org/10.1016/j.heliyon.2021.e06035
[8] Taylor PC, Adams AC, Hufford MM, de la Torre I, Winthrop K, Gottlieb RL. Neutralizing monoclonal antibodies for treatment of COVID-19. Nat Rev Immunol 2021;21:382–93. https://doi.org/10.1038/s41577-021-00542-x
[9] Hwang YC, Lu RM, Su SC, Chiang PY, Ko SH, Ke FY, Liang KH, Hsieh TY, Wu HC. Monoclonal antibodies for COVID-19 therapy and SARS-CoV-2 detection. J Biomed Sci 2022;29:1. https://doi.org/10.1186/s12929-021-00784-w
[10] Iketani S, Liu L, Guo Y, Liu L, Chan JF, Huang Y, Wang M, Luo Y, Yu J, Chu H, Chik KK, Yuen TT, Yin MT, Sobieszczyk ME, Huang Y, Yuen KY, Wang HH, Sheng Z, Ho DD. Antibody evasion properties of SARS-CoV-2 Omicron sublineages. Nature 2022;604:553–6. https://doi.org/10.1038/s41586-022-04594-4
[11] Choi S, Choi KY. Screening-based approaches to identify small molecules that inhibit protein–protein interactions. Expert Opin Drug Disco 2017;12:293–303. https://doi.org/10.1080/17460441.2017.1280456
[12] Hammond J, Leister-Tebbe H, Gardner A, Abreu P, Bao W, Wisemandle W, Baniecki M, Hendrick VM, Damle B, Simón-Campos A, Pypstra R, Rusnak JM. EPIC-HR investigators, oral nirmatrelvir for high-risk, nonhospitalized adults with covid-19. N Engl J Med 2022;386:1397–408. https://doi.org/10.1056/nejmoa2118542
[13] Duarte CM, Ketcheson DI, Eguíluz VM, Agustí S, Fernández-Gracia J, Jamil T, Laiolo E, Gojobori T, Alam I. Rapid evolution of SARS-CoV-2 challenges human defenses. Sci Rep 2022;12:6457. https://doi.org/10.1038/s41598-022-10097-z
[14] Mariano G, Farthing RJ, Lale-Farjat SLM, Bergeron JRC. Structural Characterization of SARS-CoV-2: Where We Are, and Where We Need to Be. Front Mol Biosci 2020;7:605236. https://doi.org/10.3389/fmolb.2020.605236
[15] Yang H, Rao Z. Structural biology of SARS-CoV-2 and implications for therapeutic development. Nat Rev Microbiol 2021;19:685–700. https://doi.org/10.1038/s41579-021-00630-8
[16] Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. e6 Cell 2020;181:281–92. https://doi.org/10.1016/j.cell.2020.02.058
[17] Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, Bi Y, Ma X, Zhan F, Wang L, Hu T, Zhou H, Hu Z, Zhou W, Zhao L, Chen J, Meng Y, Wang J, Lin Y, Yuan J, Xie Z, Ma J, Liu WJ, Wang D, Xu W, Holmes EC, Gao GF, Wu G, Chen W, Shi W, Tan W. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 2020;395:565–74. https://doi.org/10.1016/s0140-6736(20)30251-8
[18] Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, Somasundaran M, Sullivan JL, Luzuriaga K, Greenough TC, Choe H, Farzan M. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature 2003;426:450–4. https://doi.org/10.1038/nature02145
[19] Tai W, He L, Zhang X, Pu J, Voronin D, Jiang S, Zhou Y, Du L. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. Cell Mol Immunol 2020;17:613–20. https://doi.org/10.1038/s41423-020-0400-4
[20] Ling R, Dai Y, Huang B, Huang W, Yu J, Lu X, Jiang Y. In silico design of antiviral peptides targeting the spike protein of SARS-CoV-2. Peptides 2020;130:170328. https://doi.org/10.1016/j.peptides.2020.170328
[21] Pandey P, Rane JS, Chatterjee A, Kumar A, Khan R, Prakash A, Ray S. 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;39:6306–16. https://doi.org/10.1080/07391102.2020.1796811
[22] Hussain A, Hasan A, Nejadi Babadaei MM, Bloukh SH, Chowdhury MEH, Sharifi M, Haghighat S, Falahati M. Targeting SARS-CoV2 Spike Protein Receptor Binding Domain by Therapeutic Antibodies. Biomed Pharmacother 2020;130:110559. https://doi.org/10.1016/j.biopha.2020.110559
[23] Abi Hussein H, Geneix C, Petitjean M, Borrel A, Flatters D, Camproux AC. Global vision of druggability issues: applications and perspectives. Drug Discov Today 2017;22:404–15. https://doi.org/10.1016/j.drudis.2016.11.021
[24] Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, Graham BS, McLellan JS. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science 2020;367:1260–3. https://doi.org/10.1126/science.abb2507
[25] Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, Zhang Q, Shi X, Wang Q, Zhang L, Wang X. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 2020;581:215–20. https://doi.org/10.1038/s41586-020-2180-5
[26] Wang Q, Zhang Y, Wu L, Niu S, Song C, Zhang Z, Lu G, Qiao C, Hu Y, Yuen KY, Wang Q, Zhou H, Yan J, Qi J. Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. e9 Cell 2020;181:894–904. https://doi.org/10.1016/j.cell.2020.03.045
[27] Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, Geng Q, Auerbach A, Li F. Structural basis of receptor recognition by SARS-CoV-2. Nature 2020;581:221–4. https://doi.org/10.1038/s41586-020-2179-y
[28] Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Disco 2015;10:449–61. https://doi.org/10.1517/17460441.2015.1032936
[29] Spinello A, Saltalamacchia A, Magistrato A. Is the Rigidity of SARS-CoV-2 Spike Receptor-Binding Motif the Hallmark for Its Enhanced Infectivity? Insights from All-Atom Simulations. J Phys Chem Lett 2020;11:4785–90. https://doi.org/10.1021/acs.jpclett.0c01148
[30] Delgado JM, Duro N, Rogers DM, Tkatchenko A, Pandit SA, Varma S. Molecular basis for higher affinity of SARS-CoV -2 spike RBD for human ACE2 receptor. Proteins 2021;89:1134–44. https://doi.org/10.1002/prot.26086
[31] Jawad B, Adhikari P, Podgornik R, Ching WY. Key interacting residues between RBD of SARS-CoV-2 and ACE2 receptor: combination of molecular dynamics simulation and density functional calculation. J Chem Inf Model 2021;61:4425–41. https://doi.org/10.1021/acs.jcim.1c00560
[32] Trigueiro-Louro J, Correia V, Figueiredo-Nunes I, Gíria M, Rebelo-de-Andrade H. Unlocking COVID therapeutic targets: a structure-based rationale against SARS-CoV-2, SARS-CoV and MERS-CoV Spike. Comput Struct Biotechnol J 2020;18:2117–31. https://doi.org/10.1016/j.csbj.2020.07.017
[33] Carino A, Moraca F, Fiorillo B, Marchianò S, Sepe V, Biagioli M, Finamore C, Bozza S, Francisci D, Distrutti E, Catalanotti B, Zampella A, Fiorucci S. Hijacking SARS-CoV-2/ACE2 receptor interaction by natural and semi-synthetic steroidal agents

acting on functional pockets on the receptor binding domain. Front Chem 2020;8:572885. https://doi.org/10.3389/fchem.2020.572885
[34] Gervasoni S, Vistoli G, Talarico C, Manelfi C, Beccari AR, Studer G, Tauriello G, Waterhouse AM, Schwede T, Pedretti A. A comprehensive mapping of the druggable cavities within the SARS-CoV-2 therapeutically relevant proteins by combining pocket and docking searches as implemented in pockets 2.0. Int J Mol Sci 2020;21:5152. https://doi.org/10.3390/ijms21145152
[35] Olotu FA, Omolabi KF, Soliman MES. Leaving no stone unturned: allosteric targeting of SARS-CoV-2 spike protein at putative druggable sites disrupts human angiotensin-converting enzyme interactions at the receptor binding domain. Inform Med Unlocked 2020;21:100451. https://doi.org/10.1016/j.imu.2020.100451
[36] Le Guilloux V, Schmidtke P, Tuffery P. Fpocket: an open source platform for ligand pocket detection. BMC Bioinforma 2009;10:168. https://doi.org/10.1186/1471-2105-10-168
[37] Stank A, Kokh DB, Fuller JC, Wade RC. Protein binding pocket dynamics. Acc Chem Res 2016;49:809–15. https://doi.org/10.1021/acs.accounts.5b00516
[38] Abi Hussein H, Geneix C, Cauvin C, Marc D, Flatters D, Camproux AC. Molecular dynamics simulations of influenza A virus NS1 reveal a remarkably stable RNA-binding domain harboring promising druggable pockets. Viruses 2020;12:537. https://doi.org/10.3390/v12050537
[39] Kuzmanic A, Bowman GR, Juarez-Jimenez J, Michel J, Gervasio FL. Investigating cryptic binding sites by molecular dynamics simulations. Acc Chem Res 2020;53:654–61. https://doi.org/10.1021/acs.accounts.9b00613
[40] 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;11:e75720. https://doi.org/10.7554/elife.75720
[41] Borrel A, Regad L, Xhaard H, Petitjean M, Camproux AC. PockDrug: a model for predicting pocket druggability that overcomes pocket estimation uncertainties. J Chem Inf Model 2015;55:882–95. https://doi.org/10.1021/ci5006004
[42] Hussein HA, Borrel A, Geneix C, Petitjean M, Regad L, Camproux AC. PockDrug-Server: a new web server for predicting pocket druggability on holo and apo proteins. Nucleic Acids Res 2015;43:W436–42. https://doi.org/10.1093/nar/gkv462
[43] Li H, Robertson AD, Jensen JH. Very fast empirical prediction and rationalization of protein pKa values. Proteins 2005;61:704–21. https://doi.org/10.1002/prot.20660
[44] Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 2015;1:19–25. https://doi.org/10.1016/j.softx.2015.06.001
[45] Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, Grubmüller H, MacKerell AD. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods 2017;14:71–3. https://doi.org/10.1038/nmeth.4067
[46] Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: a linear constraint solver for molecular simulations. J Comput Chem 1997;18:1463–72. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12%3C1463::AID-JCC4%3E3.0.CO;2-H
[47] Grubmüller H, Heller H, Windemuth A, Schulten K. Generalized verlet algorithm for efficient molecular dynamics simulations with long-range interactions. Mol Simul) 1991;6(1–3):121–42. https://doi.org/10.1080/08927029108022142
[48] Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J Chem Phys 2007;126(1):014101. https://doi.org/10.1063/1.2408420
[49] Parrinello M, Rahman A. Polymorphic transitions in single crystals: a new molecular dynamics method. J Appl Phys 1981;52:7182–90. https://doi.org/10.1063/1.328693
[50] Humphrey W, Dalke A, Schulten K. VMD: Visual Molecular Dynamics. J Mol Graph 1996;14:33–8. https://doi.org/10.1016/0263-7855(96)00018-5
[51] R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, 2020. https://www.R-project.org.
[52] T. Williams, C. Kelley, many others, GNUPLOT 5.2. An Interactive Plotting Program, 2019. ⟨http://www.gnuplot.info/⟩.
[53] Schrödinger, L.L.C. The PyMOL Molecular Graphics System, Version 2.0, 2015. https://pymol.org/2/.
[54] Vorobjev YN, Almagro JC, Hermans J. Discrimination between native and intentionally misfolded conformations of proteins: ES/IS, a new method for calculating conformational free energy that uses both dynamics simulations with an explicit solvent and an implicit solvent continuum model. Proteins 1998;32:399–413. https://doi.org/10.1002/(SICI)1097-0134(19980901)32:4%3C399::AID-PROT1%3E3.0.CO;2-C
[55] Daura X, Gademann K, Jaun B, van Gunsteren WF, Mark AE. Peptide folding: when simulation meets experiment. Angew Chem Int Ed 1999;38:236–40. https://doi.org/10.1002/(SICI)1521-3773(19990115)38:1/2%3C236::AID-ANIE236%3E3.0.CO;2-M
[56] Kumari R, Kumar R. Open source drug discovery consortium, A. Lynn, *g_mmpbsa* - A GROMACS Tool for High-Throughput MM-PBSA Calculations. J Chem Inf Model 2014;54:1951–62. https://doi.org/10.1021/ci500020m
[57] Kukic P, Farrell D, McIntosh LP, García-Moreno EB, Jensen KS, Toleikis Z, Teilum K, Nielsen JE. Protein dielectric constants determined from NMR chemical shift perturbations. J Am Chem Soc 2013;135:16968–76. https://doi.org/10.1021/ja406995j
[58] Murtagh F, Legendre P. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? J Classif 2014;31:274–95. https://doi.org/10.1007/s00357-014-9161-z
[59] Hahsler M, Hornik K, Buchta C. Getting things in order: an introduction to the R package seriation. J Stat Soft 2008;25:1–34. https://doi.org/10.18637/jss.v025.i03
[60] Mittal A, Manjunath K, Ranjan RK, Kaushik S, Kumar S, Verma V. COVID-19 pandemic: insights into structure, function, and hACE2 receptor recognition by SARS-CoV-2. PLoS Pathog 2020;16:e1008762. https://doi.org/10.1371/journal.ppat.1008762
[61] Yi C, Sun X, Ye J, Ding L, Liu M, Yang Z, Lu X, Zhang Y, Ma L, Gu W, Qu A, Xu J, Shi Z, Ling Z, Sun B. Key residues of the receptor binding motif in the spike protein of SARS-CoV-2 that interact with ACE2 and neutralizing antibodies. Cell Mol Immunol 2020;17:621–30. https://doi.org/10.1038/s41423-020-0458-z
[62] Veeramachaneni GK, Thunuguntla VBSC, Bobbillapati J, Bondili JS. Structural and simulation analysis of hotspot residues interactions of SARS-CoV 2 with human ACE2 receptor. J Biomol Struct Dyn 2021;39:4015–25. https://doi.org/10.1080/07391102.2020.1773318
[63] Delgado Blanco J, Hernandez-Alias X, Cianferoni D, Serrano L. In silico mutagenesis of human ACE2 with S protein and translational efficiency explain SARS-CoV-2 infectivity in different species. PLoS Comput Biol 2020;16:e1008450. https://doi.org/10.1371/journal.pcbi.1008450
[64] Othman H, Bouslama Z, Brandenburg JT, da Rocha J, Hamdi Y, Ghedira K, Srairi-Abid N, Hazelhurst S. Interaction of the spike protein RBD from SARS-CoV-2 with ACE2: similarity with SARS-CoV, hot-spot analysis and effect of the receptor polymorphism. Biochem Biophys Res Commun 2020;527:702–8. https://doi.org/10.1016/j.bbrc.2020.05.028
[65] Chakraborty S. Evolutionary and structural analysis elucidates mutations on SARS-CoV2 spike protein with altered human ACE2 binding affinity. Biochem Biophys Res Commun 2021;534:374–80. https://doi.org/10.1016/j.bbrc.2021.01.035
[66] Williams-Noonan BJ, Todorova N, Kulkarni K, Aguilar MI, Yarovsky I. An active site inhibitor induces conformational penalties for ACE2 recognition by the spike protein of SARS-CoV-2. J Phys Chem B 2021;125(10):2533–50. https://doi.org/10.1021/acs.jpcb.0c11321
[67] Toelzer C, Gupta K, Yadav SKN, Borucu U, Davidson AD, Kavanagh Williamson M, Shoemark DK, Garzoni F, Staufer O, Milligan R, Capin J, Mulholland AJ, Spatz J, Fitzgerald D, Berger I, Schaffitzel C. Free fatty acid binding pocket in the locked structure of SARS-CoV-2 spike protein. Science 2020;370:725–30. https://doi.org/10.1126/science.abd3255
[68] Luan B, Huynh T. Molecular mechanism of the N501Y mutation for enhanced binding between SARS-CoV-2′s spike protein and human ACE2 receptor. J Med Chem 2021;65:2820–6. https://doi.org/10.1021/acs.jmedchem.1c00311
[69] Weisblum Y, Schmidt F, Zhang F, DaSilva J, Poston D, Lorenzi JC, Muecksch F, Rutkowska M, Hoffmann HH, Michailidis E, Gaebler C, Agudelo M, Cho A, Wang Z, Gazumyan A, Cipolla M, Luchsinger L, Hillyer CD, Caskey M, Robbiani DF, Rice CM, Nussenzweig MC, Hatziioannou T, Bieniasz PD. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. eLife 2020;9:e61312. https://doi.org/10.7554/elife.61312
[70] Gobeil SM, Janowska K, McDowell S, Mansouri K, Parks R, Stalls V, Kopp MF, Manne K, Li D, Wiehe K, Saunders KO, Edwards RJ, Korber B, Haynes BF, Henderson R, Acharya P. Effect of natural mutations of SARS-CoV-2 on spike structure, conformation and antigenicity. eabi6226 Science 2021;373. https://doi.org/10.1126/science.abi6226
[71] Yuan M, Wu NC, Zhu X, Lee CD, So RTY, Lv H, Mok CKP, Wilson IA. A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. Science 2020;368:630–3. https://doi.org/10.1126/science.abb7269

Chapter 4

A comparative structural study of the mutations in SARS-CoV-2 RBD variants of concern on the interaction with the ACE2 receptor

The emergence of SARS-CoV-2 variants of concern has significantly impacted the COVID-19 pandemic. This demonstrates the severity of mutations that the wild-type strains have subjected. The Spike protein, which plays a critical role in binding to human cells and pathogenesis, has been identified as a potential therapeutic target to inhibit. In the submitted paper below [52], a common protocol was applied to reconstruct SARS-CoV-2 variant of concern (Alpha, Beta, Gamma, Delta, and Omicron) structures and perform MD simulations to evaluate the ACE2-RBD complex's stability in each variant. MM-PBSA calculations were also carried out to compare the binding and biophysical properties in each SARS-CoV-2 variants of concern when interacting with the human receptor ACE2. In this study, new key residues have been identified playing an important role in maintaining RBD-ACE2 interaction stronger. Our results showed that the variants of concern showed higher affinities for ACE2, with specific mutations such as Q498R or N501Y. These results support experimental data and offer insightful information about how the virus interacts with its receptor in a detailed protocol.

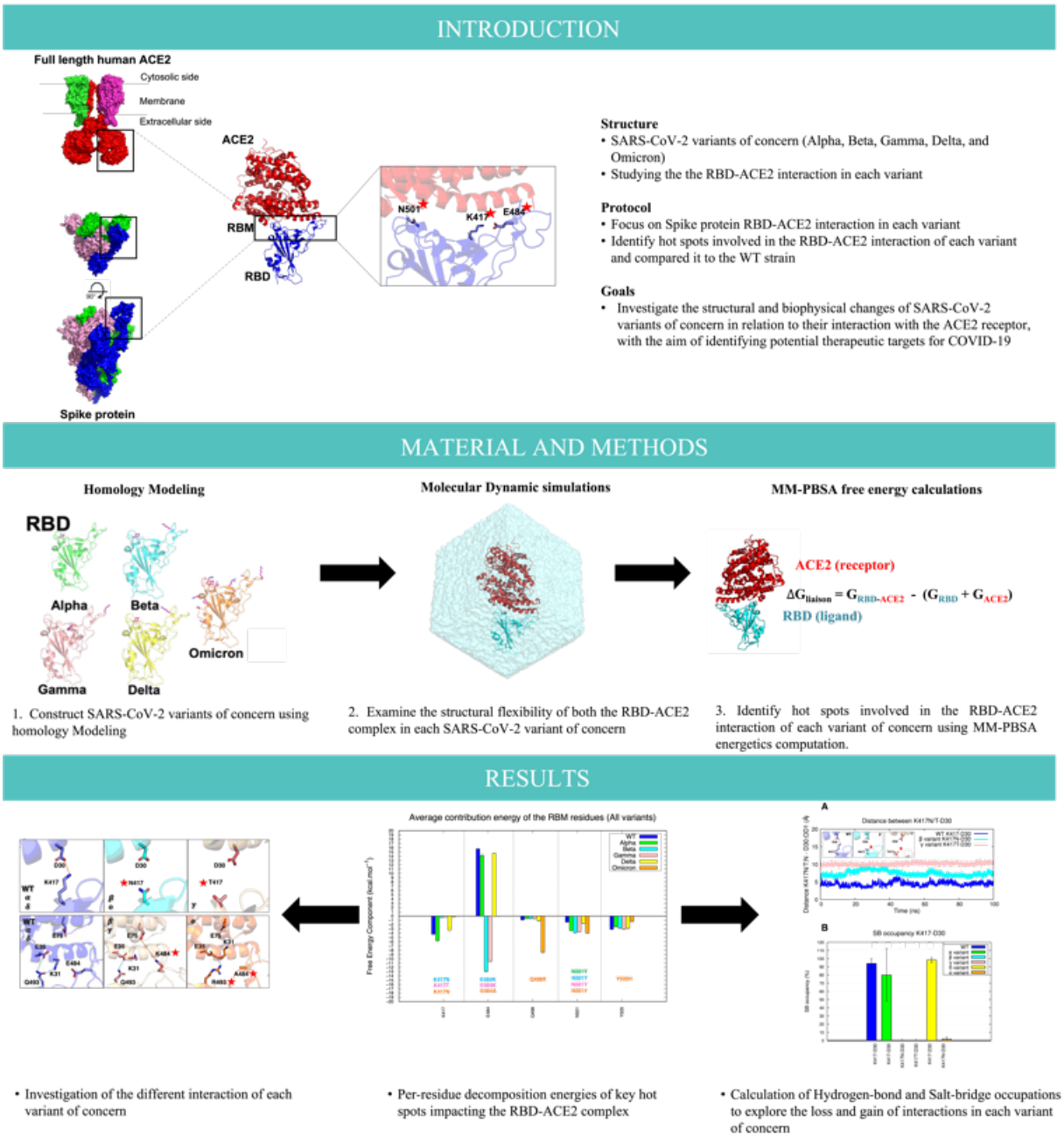


Figure 4. Representation of the bioinformatics protocol used to study SARS-CoV-2 variants of concern and their impact on the RBD-ACE2 interaction.

THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB Article

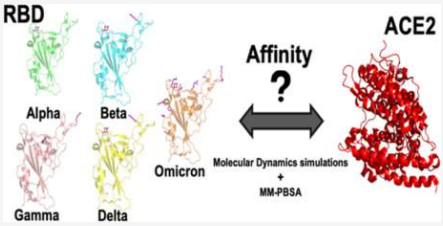
Comparative Study of the Mutations Observed in the SARS-CoV-2 RBD Variants of Concern and Their Impact on the Interaction with the ACE2 Protein

Mariem Ghoula, Audrey Deyawe Kongmeneck, Rita Eid, Anne-Claude Camproux,* and Gautier Moroy*

Cite This: https://doi.org/10.1021/acs.jpcb.3c01467 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: SARS-CoV-2 strains have made an appearance across the globe, causing over 757 million cases and over 6.85 million deaths at the time of writing. The emergence of these variants shows the amplitude of genetic variation to which the wild-type strains have been subjected. The rise of the different SARS-CoV-2 variants resulting from such genetic modification has significantly affected COVD-19's major impact on proliferation, virulence, and clinics. With the emergence of the variants of concern, the spike protein has been identified as a possible therapeutic target due to its critical role in binding to human cells and pathogenesis. These mutations could be linked to functional heterogeneity and use a different infection strategy. For example, the Omicron variant's multiple mutations should be carefully examined, as they represent one of the most widely spread strains and hint to us that there may be more genetic changes in the virus. As a result, we applied a common protocol where we reconstructed SARS-CoV-2 variants of concern and performed molecular dynamics simulations to study the stability of the ACE2−RBD complex in each variant. We also carried out free energy calculations to compare the binding and biophysical properties of the different SARS-CoV-2 variants when they interact with ACE2. Therefore, we were able to obtain consistent results and uncover new crucial residues that were essential for preserving a balance between maintaining a high affinity for ACE2 and the capacity to evade RBD-targeted antibodies. Our detailed structural analysis showed that SARS-CoV-2 variants of concern show a higher affinity for ACE2 compared to the Wuhan strain. Additionally, residues K417N and E484K/A might play a crucial role in antibody evasion, whereas Q498R and N501Y are specifically mutated to strengthen RBD affinity to ACE2 and, thereby, increase the viral effect of the COVID-19 virus.



1. INTRODUCTION

SARS-CoV-2's high transmissibility and mutation rates, combined with a lack of robust preexisting immunity in hosts and a slow rate of immunization through vaccinations, have caused COVID-19 cases to surge to over 572 million worldwide by July 2022, as reported by the WHO.[1] Although numerous antibodies have been shown to neutralize the wild-type (WT) virus or the Wuhan strain, their efficacy against developing variations, particularly those that have been shown to avoid the host immune response and develop an antibody escape mechanism, should be closely studied.[2−5]

The spike or S protein of the SARS-CoV-2 virus facilitates viral entrance into the cell.[6−8] Viral entrance occurs when the spike protein binds to receptors on the host cell, causing the cell membrane to fuse. The S protein is a structural polypeptide with two subunits, S1 and S2, that engage with the ACE2 receptors and fuse the viral and host cell membranes.[9] Moreover, the homo-trimeric spike glycoprotein (residues 1−1273) (Figure 1) has two structural states: active (up) and inactive (down), and it is highly conserved across all human coronaviruses.[10] The S protein has been the main focus of several studies since it is the key protein for the virus's attachment to the cell host.[11] More precisely, most of the studies have focused on the receptor binding domain (RBD; residues 331−524) of the SARS-CoV-2 S protein because of its function in binding the angiotensin-converting enzyme 2 (ACE2) receptor (Figure 1) and immune system recognition.[12−16] Moreover, the RBD (Figure 1) contains the receptor binding motif (RBM) region (residues 437−508), which is responsible for maintaining contact with the ACE2 protein.[8] It is also important to note that treatments against the SARS-CoV-2 virus are developing around the RBD since it is the most crucial protein in the viral process.[17−21] The RBD region is also the most affected part of the virus, where several emerging mutations occurred in different lineages.[17−21] Due to the RBD's variability and its fast mutation rate, most of the treatments are focused on targeting the RBD with numerous

Received: March 2, 2023
Revised: April 20, 2023

https://doi.org/10.1021/acs.jpcb.3c01467
J. Phys. Chem. B XXXX, XXX, XXX−XXX

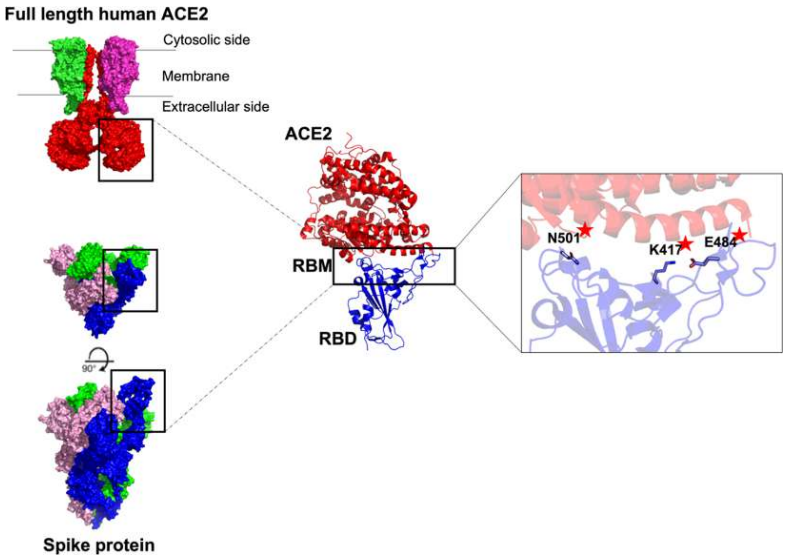


Figure 1. Representation of the interaction between the ACE2 human receptor and the SARS-CoV-2 spike protein. The left panel shows a surface representation of the ACE2 human receptor (upper panel) in a transversal view and the trimeric structure of the SARS-CoV-2 spike protein in a top view (middle panel) and transversal view (lower panel). The middle panel offers a more detailed cartoon representation of the complex formed by the RBD (in blue) and the distal region of ACE2 (in red). The right panel shows the most recurrent mutations in the RBM region. The most recurrent mutations in the five variants of concern are labeled with a red star.

antibodies issued from COVID-19 patients or other antibody therapies under investigation.[22−24] With all this information, we focused on analyzing the RBD to decipher the mutation mechanism and also give more insight into future mutation effects.

Many COVID-19 variants and sub-variants of concern are currently under investigation. However, five variants of concern (Alpha, Beta, Gamma, Delta, and Omicron) have been classified as dangerous and are evolving rapidly all over the world by the WHO.[1] Among these, the Alpha variant was first discovered in the United Kingdom in December 2020 and was declared a variant of concern in December 2020.[25,26] It differs from the original Wuhan strain due to many significant alterations in the spike protein.[27,28] For instance, one is the N501Y mutation located in the RBD region, which increases the virus's contagiousness by improving the spike protein binding to cellular receptors.[28−31] It also has a D614G mutation, which is likely to aid viral replication,[32−37] and a P681H mutation,[38] whose function is unknown, but which has appeared many times spontaneously. Moreover, the Alpha variant is predicted to be roughly 50% more transmissible than the original Wuhan strain.[25] It is also assumed to be linked to worsening illness severity, but this is not clear.[39] However, COVID-19 vaccinations and monoclonal antibody therapies have been reported to be still highly effective against it.[40−48]

The Beta variant was discovered for the first time in South Africa in December 2020.[1] In addition to three of the alterations seen in the Alpha variant,[24,27,28] The Beta variant has a K417N mutation located in the RBD region, which may enable the virus resistance to neutralize antibodies produced by vaccination or previous infection.[49−51] Although it is expected to be 50% more transmissible than prior variants, there is minimal evidence that Beta is linked to more severe diseases including severe gastrointestinal problems, hearing loss, blood clots that cause tissue death, and gangrene.[52,53] Reduced neutralization by antibodies generated by vaccination or as a result of a previous infection is the main source of concern. Despite this, recent vaccines appear to provide effective protection against the Beta variant.[54,55]

The Gamma variant was first discovered in Brazil in January 2021.[1] It contains the E484K, N501Y, and D614G mutations, as do some other variants of concern.[56−58] It also carries a K417T/N mutation, which is linked to greater binding to human cells, potentially making the virus easier to transmit,[59] and a H655Y mutation, whose function is uncertain.[60,61] Moreover, according to a recent study published by Faria and co-workers,[58] the Gamma variant is 1.7−2.4 times more transmissible than non-variants of concern. However, existing COVID-19 vaccinations appear to be effective at preventing the Gamma variant.[54,62]

The Delta variant was first reported in India in May 2021. It has since been confirmed in multiple locations throughout the globe,[1] quickly displacing other variants to become the dominant variant in several countries. Delta has the D614G mutation as well as a few other mutations not reported in other variants of concern.[63] These include a L452R mutation on the RBD, which is expected to enhance infectivity and may help the virus escape immune cell destruction.[64−68] The Delta variant contains the T478K mutation also located on the RBD, which is thought to assist the virus avoid immune detection.[69−72] Finally, this variant of concern has a P681R mutation, which is linked to an increased ability to cause

https://doi.org/10.1021/acs.jpcb.3c01467
J. Phys. Chem. B XXXX, XXX, XXX−XXX

serious illnesses.[73] The Delta variant is thought to be 40−60% more transmissible than the Alpha variant and nearly twice as transmissible as the original Wuhan strain.[68,75] Although data suggest that vaccines are slightly less efficient in avoiding infection with the Delta form,[44,63] they are still quite effective in preventing serious disease.[76,77]

The first verified Omicron infection was discovered in a sample collected on November 9, 2021.[1] The Omicron variant has been authenticated in several areas throughout the world, including sections of North and South America, Europe, Africa, Asia, and Australia.[1] Omicron has a lot of mutations, some of which were identified as dangerous.[78] Fifteen mutations can be already found on the RBD: G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, and Y505H.[79−83] It is unclear whether Omicron is easier to spread from person to person than other variants or if infection with it causes more severe disease. There is currently no evidence that the symptoms associated with Omicron are distinct from those associated with other variations.[84,85] The emergence of the BA.2 sub-lineage, on the other hand, has raised concerns because it looks to be more transmissible.[86] Being currently the most transmissible variant, three other sub-lineages (BA.3, BA.4, and BA.5) of Omicron have emerged and are suspected to become the most dominant variants.[86] The capacity of BA.4 and BA.5 to avoid immune protection brought on by earlier infection and/or vaccination is presumably the cause of their current observed growth advantage, especially if this immunity has weakened over time.[87−89] Currently, there is no evidence that the severity of BA.4 and BA.5 will vary in comparison to earlier Omicron lineage infections.[74]

It is also important to note that many of these variants can contain additional mutations not yet confirmed or only detected in a minority of samples. For instance, the WHO is also monitoring the spread of an Alpha variant with an extra E484K mutation, which could enable the virus to get past the body's immune defenses by evading neutralizing antibodies produced by vaccination or previous infection.[90−92] Another example is the finding of the 'Delta plus' variant, which was first found in Nepal and carries an extra K417N mutation.[77,93−95] These observations show how versatile and adaptable the virus is.

Crystallography[96−102] and cryo-EM[6,103−111] techniques, as well as computational analyses[112−117] and antibody-binding assays[106,118−127] have been used to study the SARS-CoV-2 variants RBD−ACE2 complexes. Deep scanning mutagenesis[128−131] and in vitro evolution[5,132−134] have also been used to investigate SARS-CoV-2 variants effects. Several research papers attempted to map out different classes of antibodies and link them to mutation studies.[135,136] To our knowledge, a detailed comparative investigation regrouping all variants of concern utilizing structure-based simulations and free energy approaches is still lacking to understand and have a general view on the SARS-CoV-2 mechanism and particularly of its variants of concern.

In this paper, we used the same protocol to reproduce reliable results for a more accurate comparison between all systems (WT, Alpha, Beta, Gamma, Delta, and Omicron). We present results from structure-based in silico modeling and a full-atomic molecular dynamics (MD) simulation protocol of the RBD−ACE2 complex. These analyses were performed to evaluate the binding free energies of the five variants of concern on the RBD (Table 1). Our results also showed the differences in interactions formed with the ACE2 protein to give more insight on the variants' mechanisms.

Table 1. Variants of Concerns' Lineages, Date of Emergence, and the Location of Their Mutations on the RBD

variants	lineage	date of emergence	mutations on the RBD
Alpha	B.1.1.7	September 2020	N501Y
Beta	B.1.351	August 2020	K417N, E484K, N501Y
Gamma	P.1	December 2020	K417T/N, E484K, N501Y
Delta	B.1.617.2	October 2020	L452R, T478K
Omicron	B.1.1.529	November 2021	G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H

2. MATERIAL AND METHODS

2.1. Structural Preparation of the Spike Protein Variants. Our study only focuses on WT SARS-CoV-2 RBD with the human ACE2 receptor and its comparison with five different COVID-19 variants of concern (Alpha, Beta, Gamma, Delta, and Omicron) (RBD residues 331−524). The crystal structure of the ACE2−RBD (PDB: 6M0J[97]) was then retrieved from the RCSB Protein Data Bank in PDB format.[137] In fact, here, we focused on the ACE2−RBD complex rather than the glycosylated states of the proteins since we are looking to compare their binding affinities and identify the RBM hotspots. We decided to model Alpha, Beta, Gamma, and Delta mutations since no PDB structures were available at the time the analyses were made. As for the Omicron variant, the respective 7T9L cryo-EM structure was taken from the PDB database.[138] The Omicron PDB was used in this study since several crystal structures of this variant of concern were available at the time our analyses were conducted. Another reason is that the majority of papers have mostly focused on this variant of concern due to its many mutations located on the RBD. Additionally, to validate our approach and verify the consistency of our results, we also constructed an Omicron variant model containing all of the RBD mutations, which we will compare with the experimental 7T9L Omicron variant structure.

Therefore, using the 6M0J structure, the Alpha, Beta, Gamma, Delta, and Omicron mutations were introduced using the "mutations wizard" in the PyMOL molecular modeling package (Schrodinger LLC). All residue rotamers were chosen according to the most probable side chain orientation suggested by the program and displaying the least amount of steric clashes within the structure. As our study only focused on the RBD−ACE2 interaction, all glycans were removed from the complexes. Afterward, each complex was protonated according to a physiological pH (pH = 7.4) using PROPKA.[139]

2.2. MD Simulations. MD simulations were performed on six systems (WT, Alpha, Beta, Gamma, Delta, and Omicron) using the GROMACS tool with the 2020 version.[140] The CHARMM36m force field[141] was selected along with an explicit solvent TIP3P model.[142] A dodecahedron shaped box with an adjusted 12 Å distance between the protein complexes, and the box was used to fit the solvated systems. Then, NA^+ ions were introduced to neutralize the whole system. Once the

https://doi.org/10.1021/acs.jpcb.3c01467
J. Phys. Chem. B XXXX, XXX, XXX−XXX

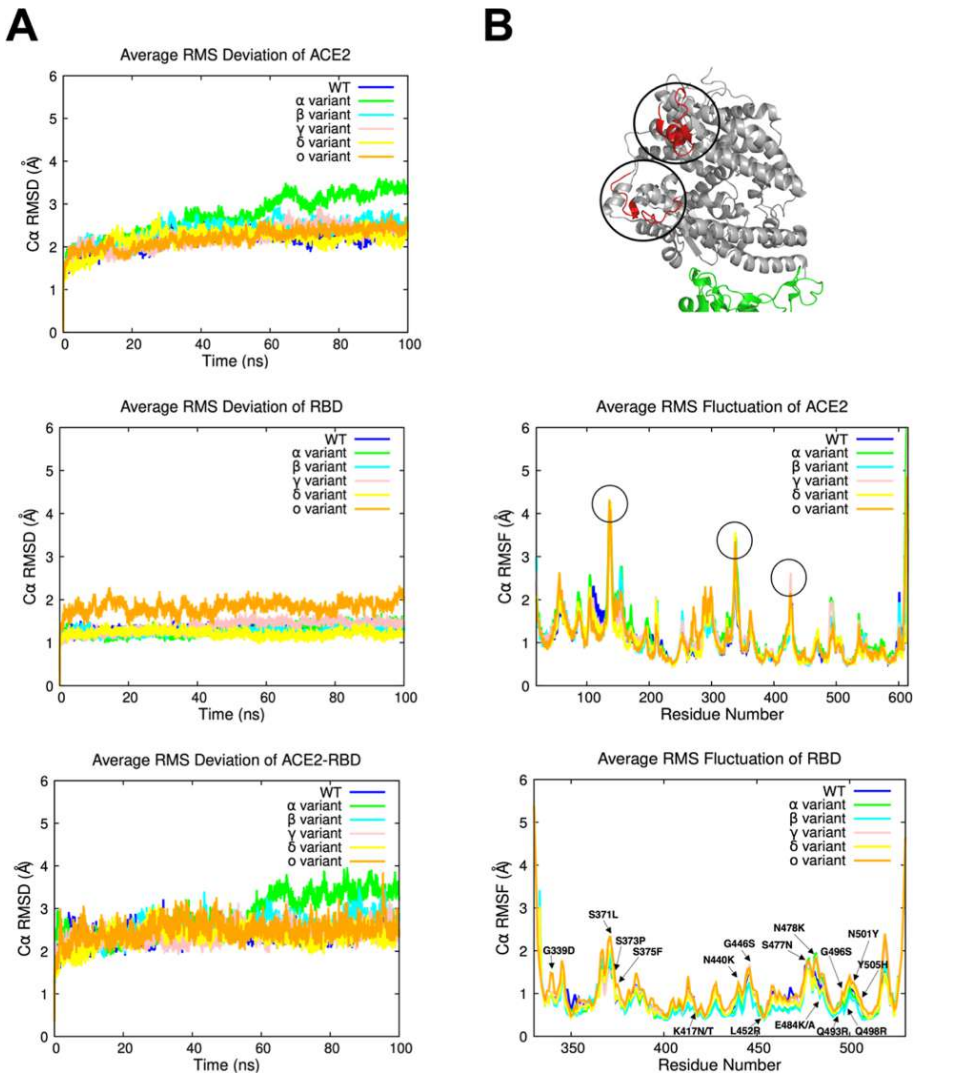


Figure 2. (A) Average C α rmsd values of the ACE2, RBD, and ACE2−RBD complex over 100 ns simulation time for each system. (B) Average C α RMSF values of the ACE2 and the RBD versus the residue number for each system. Highly flexible regions on the ACE2 protein are highlighted with circles and colored in red to easily locate them on the protein. These regions correspond to the circled peaks on the RMSF graph as well.

solvated and electroneutral systems were assembled, a full 50,000-step energy minimization was performed to avoid steric clashes using the "steepest descent" algorithm. Each complex was equilibrated with an *NVT* (number of particles, volume, and temperature) ensemble for 1 ns at a temperature of 300 K and a coupling constant of 0.1 ps. Subsequently, an *NPT* (number of particles, pressure, and temperature) ensemble was running by setting the temperature at 300 K, and the pressure at 1 bar for 1 ns. As for the electrostatic interactions, they were calculated using the particle-mesh Ewald method.[143] Upon the completion of the two equilibration phases, the production phase for each system was performed for 100 ns in triplicate (3 × 100 ns). Originally, 3 runs were performed for each system (WT, Alpha, Beta, Gamma, Delta, and Omicron) for reproducibility purposes (Supporting Information Figures S1−S3). Accordingly, the average C α rmsd and RMSF values for each system were calculated to have an overall view of the protein's stability. The VMD program[144] was used to calculate

https://doi.org/10.1021/acs.jpcb.3c01467
J. Phys. Chem. B XXXX, XXX, XXX−XXX

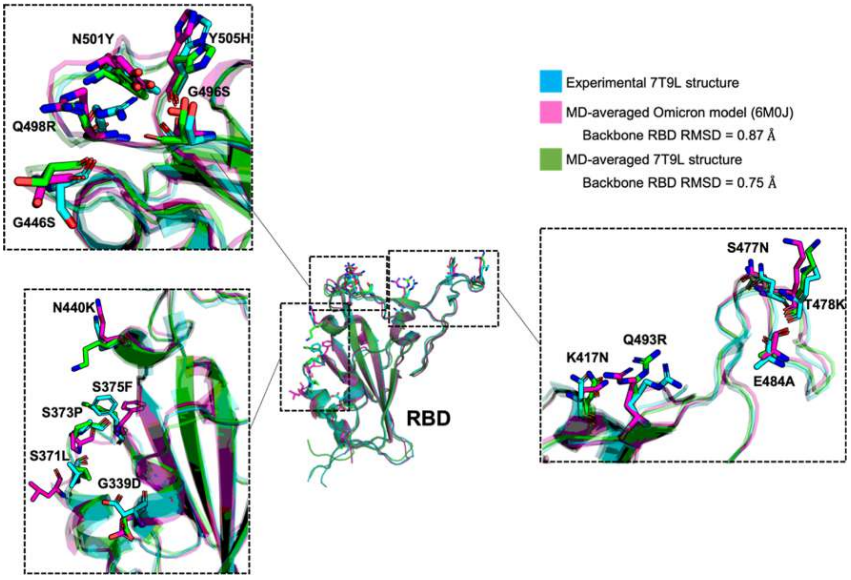


Figure 3. (A) Structural overlay of the MD-averaged Omicron model (magenta) and 7T9L structures (green) with the resolved cryo-EM 7T9L structure (cyan). The 15 mutated residues of the RBD in the Omicron variant have been highlighted and represented in boxes. The backbone RBD rmsd has been specified as well.

all hydrogen-bond (HB) and salt bridge (SB) occupancy rates with the angle and distance between the donor and acceptor set to 30° and 3.5 Å, respectively.

2.3. MM-PBSA Calculations. The MM-PBSA (molecular mechanics-Poisson Boltzmann surface area) methodology and the *g_mmpbsa* program[145] were used to compute the binding free energies of each RBD−ACE2 complex as well as their residue decomposition energies. The *g_mmpbsa* program is a tool that integrates functions from GROMACS[140] and APBS[146] in order to calculate the binding free energies of protein−protein or protein-ligand complexes.

In this approach, the binding free energy G_{bind} between protein and ligand/protein includes different energy terms and could be calculated as

$$G_{bind} = G_{complex} - (G_{protein} + G_{ligand})$$

$$= E_{MM} - TS + G_{sol}$$

$$= E_{vdw} + E_{elec} - TS + G_{PB} + G_{SA}$$

The gas-phase interaction energy E_{MM} is equal to the sum of van der Waals energy E_{vdw} and electrostatic energy E_{elec}. The polar solvation energy G_{PB} and the non-polar solvation energy G_{SA} are added together to form G_{sol}. The Poisson−Boltzmann (PB) approximation model is used to determine the polar solvation energy, whereas the solvent accessible surface area (SASA) is used to estimate the non-polar solvation energy. Owing to the high computational cost and the fact that considering the entropy, even if it improves the agreement with the experimental values of the binding free energy, it does not change the global profile of the energy,[147] the entropy contribution ($-TS$) is omitted in this work. The binding free energies were decomposed to each residue after computation. It is worth noting that the more negative the energy, the more beneficial the contribution. Positive energy values, on the other hand, indicate unfavorable interactions and a low contribution to the complex. For each binding complex, MM-PBSA calculations were carried out on a total of 90 different conformations for each MD simulation, as in the calculations started from 10 ns simulation time and skipped every 1 ns.

3. RESULTS AND DISCUSSION

3.1. Construction of Variant Models. We had to verify that our in silico ACE2−RBD variant models show the root-mean-square deviation (rmsd) values for less than 1 Å (Supporting Information Table S1), which are representative of good models. RBD's Alpha, Beta, Gamma, Delta, and Omicron variants showed a backbone rmsd of only 0.2, 0.1, 0.2, 0.5, and 0.5 Å from the most recently resolved crystallographic and cryo-EM structures (PDB IDs: 7EKF,[97] 7EKG,[97] 7EKC,[97] 7WBQ,[99] and 7T9L[14]) and aligned well against them. Thus, all models are valid to use for further steps.

3.2. Structure Flexibility and Stability of the Simulation Systems. *3.2.1. SARS-CoV-2 WT and Variant Systems.* In this section, we report a thorough analysis to evaluate the stability of the ACE2 protein and the RBD of SARS-CoV-2 in the WT and variant systems. Only the Omicron variant is based on experimental data, but the variant systems comprise all modeled structures for the Alpha, Beta, Gamma, Delta, and Omicron variants. Throughout the 100 ns-simulation runs, the stability profiles of SARS-CoV-2 RBD of each system in complex with human ACE2 were analyzed using GROMACS (Figure 2). C α rmsd values of ACE2 and

https://doi.org/10.1021/acs.jpcb.3c01467
J. Phys. Chem. B XXXX, XXX, XXX−XXX

Table 2. Summary of the Average Energy Terms ΔE_{vdw}, ΔE_{Elec}, ΔE_{Polar}, ΔE_{Apolar}, and ΔG_{bind} of the Six Different ACE2−RBD Complexes (WT, Alpha, Beta, Gamma, Delta, and Omicron)[a]

	binding energy components (kcal mol^{-1})				
ACE2−RBD complex	ΔE_{vdw}	ΔE_{Elec}	ΔE_{Polar}	ΔE_{Apolar}	ΔG_{bind}
WT	−77.3 ± 6.3	−78.3 ± 5.2	114.1 ± 33.0	−10.0 ± 0.9	−51.5 ± 32.0
Alpha	−68.7 ± 15.1	−80.3 ± 12.2	85.6 ± 41.7	−8.5 ± 2.0	−72.0 ± 32.9
Beta	−77.5 ± 6.0	−96.2 ± 4.4	67.7 ± 27.9	−9.1 ± 0.9	−112.3 ± 27.6
Gamma	−75.7 ± 5.6	−96.2 ± 5.4	78.2 ± 31.1	−9.3 ± 0.9	−103.1 ± 29.3
Delta	−77.1 ± 5.4	−121.5 ± 5.4	108.6 ± 35.2	−10.0 ± 0.9	−100.0 ± 33.8
Omicron	−77.8 ± 5.3	−187.4 ± 5.4	129.3 ± 32.8	−10.3 ± 0.9	−142.6 ± 30.4

[a]All computed energies are expressed as mean ± standard deviation.

RBD were also studied separately to have detailed information of each of the proteins' stability (Figure 2A).

ACE2 average C α rmsd values are steady and stable, ranging between 2 and 2.5 Å from the initial structure. The Alpha variant shows more fluctuations, with values increasing progressively until reaching a plateau at 3.5 Å (Figure 2). On the other hand, SARS-CoV-2 RBD showed steady C α rmsd values overall (1−2 Å) with no obvious fluctuations (2−2.2 Å) within the Omicron variant (Figure 2). Similarly, ACE2−RBD complexes have proven to be stable during the MD simulations (2−2.5 Å) apart from the Alpha variant, which displayed the same variations as the ACE2 protein with C α rmsd values reaching a maximum of 3.9 Å, stabilizing at the end of the MD simulation (Figure 2). Therefore, the Alpha variant behavior can be explained by the high flexibility of the ACE2 protein and its constant shifting around the RBD toward the end of the MD simulation trajectory (Supporting Information Figures S1A and S3A). As shown in Supporting Information Figures S4 and S5, there were no significant differences in the rmsd values of the resolved and the constructed Omicron variant structures for the ACE2, RBD, and ACE2−RBD complex, supporting the reliability of our protocol.

Then, we computed the average root-mean-square fluctuation (RMSF) of the C α atoms versus the systems' residue numbers to explore the detailed residual atomic fluctuations (Figure 2). It is also helpful to compare this further with the experimentally obtained *B*-factors from crystallography, which exhibit similar tendencies but with significantly less overall flexibility because of the cryogenic temperatures and lattice conditions that a protein is subjected to in a crystal. Thus, C α RMSF calculations of the 3 MD runs belonging to each of the variants were added in the Supporting Information Figures S6 and S7 for more details. Three major peaks regions can be observed on the ACE2 protein, all corresponding to the highly flexible loops. We discovered a slightly high fluctuation area specific to the Alpha variant, which is a random coil (P490−P500) reaching maximum C α RMSF values of ∼2 Å. Otherwise, no significant changes in terms of structural flexibility were observed in the Alpha variant as both ACE2 and RBD remained stable and showed similar C α RMSF values to the other variants. Therefore, the increased fluctuations of the ACE2 protein belonging to the Alpha variant (Figure 2B) might be modulated by conformational changes within the protein in order to stabilize itself. As for the RBD, we noticed a slight increase in the RMSF values in the Omicron variant, which corresponds to coiled structured regions of the protein. Similar results were observed for the studied cryo-EM structure as well as the constructed Omicron variant model, with RMSF values displaying the same tendency (Supporting Information Figures S8 and S9). The RBM region, which is the binding region of the RBD, remained stable for all variants, and no significant changes were observed. In order to understand potential changes in the conformations, the 7T9L structure was further compared to the MD-averaged Omicron model and cryo-EM resolved structural complexes. The MD-averaged conformations were superimposed with the 7T9L structure, as seen on Figure 3. To properly analyze protein structures, accurate protein side-chain modeling is required. Hence, when comparing the MD-averaged Omicron conformations to the experimental Omicron structure, we computed the rmsd on a basis of all side chain atoms, excluding hydrogen atoms (Table S2). Lower side chain rmsd values produce ensure more precise results.[148] One of the few differences that can be observed is the flexibility of the side chains of the mutated residues located on the RBD surface (G339D, S371L, S373P, S375F, and N440K). The side chain rmsd values of the residues on the RBD surface ranged from 0.37 to 4.08 Å. The conformations' high quality was confirmed by the fact that residues spanning the RBD interaction surface did not exceed side chain rmsd of 2.3 Å. In fact, it's essential to remember that the side chains of the residues are flexible components, which explains why there are a few minor variations in the structures. The backbone RBD rmsd values for the MD-averaged Omicron model and cryo-EM resolved structure complexes, respectively, showed exceptionally low values of 0.75 and 0.87 Å when overlaying both structures to the experimental complex (Figure 3). Hence, by doing a direct structural analysis, agreement with experimental structure data for each complex was verified. Overall, C α rmsd and RMSF results showed stable systems over time and can be used further for additional analyses.

3.3. Binding Free Energy Analysis. The binding affinity values have been reported using the molecular mechanics-generalized Born surface area (MM-GBSA) and MM-PBSA approaches in earlier investigations using the WT protein and one of the many SARS-CoV-2 variants.[80,149−154] In this study, we compared all known variants of concern that have emerged since the beginning of the pandemic using the MM-PBSA method with the aim of studying the effects of the mutations on the binding energy of the ACE2−RBD complex.

The results showed that the binding free energy of the WT system (−51.5 ± 32.0 kcal mol^{-1}) is lower than that in the other systems (Table 2). On the other hand, the Omicron variant displayed the highest free energy with a value of −142.6 ± 31.4 kcal mol^{-1}. This is mostly due to the introduction of several charged residues with favorable interactions making the ΔE_{Elec} energy the key driving factor in the binding process of RBD with ACE2. This is the case of the Alpha variant (N501Y), the Beta variant (E484K and N501Y), the Gamma variant (E484K and N501Y), the Delta variant (L452R and

https://doi.org/10.1021/acs.jpcb.3c01467
J. Phys. Chem. B XXXX, XXX, XXX−XXX

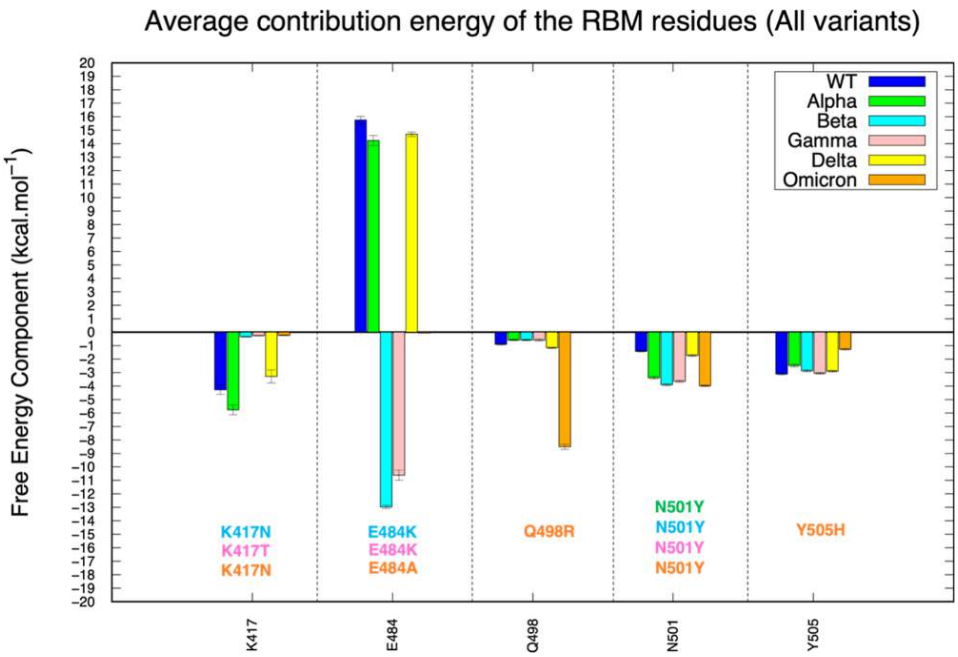


Figure 4. Bar plots of the free energy component of the RBM hot spots in kcal mol^{-1}. Mutated residues are colored according to the variant color (dark blue for the WT system, green for the Alpha variant, cyan for the Beta variant, pink for the Gamma variant, yellow for the Delta variant, and orange for the Omicron variant). Hot spot residues are considered when their decomposition energy is equal to or greater than −1 kcal mol^{-1}. All computed free energy components are displayed with error bars.

T478K), and the Omicron variant (N440K, T478K, Q493R, Q498R, and N501Y). It can be easily proven with the increasing trend of ΔE_{Elec} values of each system reaching −78.3 ± 5.2, −80.3 ± 12.2, −96.2 ± 4.4, −143.6 ± 6.1, −121.5 ± 5.4, and −187.4 ± 5.4 kcal mol^{-1} for the WT, Alpha, Beta, Gamma, Delta, and Omicron variants, respectively. As for the $\Delta E_{vdw,}$ ΔE_{Polar}, and ΔE_{Apolar} energy terms, they all seem to be very similar for all systems with no significant changes. The binding free energy of the constructed Omicron variant and the Omicron variant's cryo-EM structure were also examined. Very similar results have been obtained for both systems, as shown in Supporting Information Table S3. For the Omicron variant model and the resolved Omicron variant structures, respectively, the results showed ΔG_{bind} values of −144.8 28.0 and −142.6 30.4 kcal mol^{-1}. Therefore, these results confirm the reliability of our method of analysis.

Then, we carried out our study with a per-residue decomposition energy comparison analysis (Figure 4). To further ensure the accuracy of our findings, we examined the two systems and applied the same protocol for the 7T9L Omicron variant and the Omicron model variant structures (Supporting Information Figure S11). First, we verified that only RBM residues that are within 6 Å of the ACE2 binding region are examined because these are the essential amino acids that play a role in the RBD−ACE2 interaction. The same analysis was also applied in our previously published paper,[155] where 8 hot spots were identified (K417, L455, F456, F486, Y489, T500, N501, and Y505) on the WT RBM. Thus, we extended this analysis on the mutated complexes we built for the five variants of the RBD to determine whether the WT RBM residues are remaining hot spots on the five variants under study as well as to detect new hot spots brought on by the insertion of the mutations.

Among the residues (K417, L455, F456, F486, Y489, T500, N501, and Y505) that have been detected as hot spots on the WT RBD, five of them (L455, F456, F486, Y489, and T500) were observed to remain key hot spots on the five variants (Alpha, Beta, Gamma, Delta, and Omicron) (Supporting Information Figure S10). Evidently, the same analysis was performed on the cryo-EM Omicron variant and the constructed Omicron variant structures, where both systems displayed identical results (Supporting Information Figure S12). Moreover, these residues are conserved in all variants of concern. Residue L455 forms intermolecular van der Waals interactions and polar contacts with ACE2 residues D30 and H34, respectively. On the other hand, residues F456, F486 and Y489 are inserted in a hydrophobic pocket on the right edge of the RBM region, which is formed by residues T27, F28, L79, and M82. Thus, the presence of aromatic residues in the pocket may provide additional binding force via π-stacking interactions. For instance, F456−T27, F486−L79, and Y498−F28 interact through hydrophobic interactions. Residue F486 forms an additional interaction through a $\pi-\pi$ stacking with residue Y83, while Y489 forms a HB with Y83. Thus, L455, F456, F486, Y489, and T500 can be considered crucial amino

https://doi.org/10.1021/acs.jpcb.3c01467
J. Phys. Chem. B XXXX, XXX, XXX−XXX

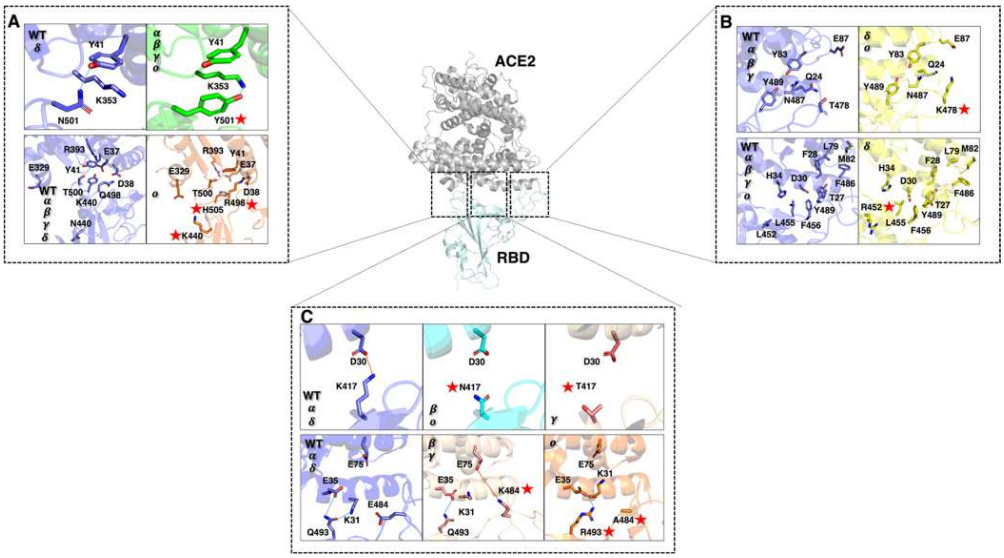


Figure 5. (A) Representation of the conserved key hot spots of the RBM (L455, F456, F486, and Y489) and their interaction with the ACE2 residues within a 6 Å distance. Mutated residues L452R and T478K were also represented to show their absence of interaction with the ACE2 binding region. (B) Representation of the new key hot spots of the RBM (N501Y and Q498R), the conserved T500 hot spot, and their interaction with the ACE2 residues within a 6 Å distance. Mutated residues N440K and Y505H were also represented to show their weak or absent interaction with the ACE2 binding region. (C) Representation of the new positive and negative key hot spots of the RBM (E484K, E484A and K417N/T), the mutated Q493R residue, and their interaction with the ACE2 residues within a 6 Å distance. Mutated residues are specified with a red star.

acids that have been conserved in all RBD variants, with the key role of maintaining tight binding with ACE2.

As shown in Figure 5, the Delta mutations T452R and T478K were also represented as they are located close to the key unchanged hot spots of the RBM. However, very weak to no interaction was made between these residues and the ACE2 binding region. L452R and T478K are located at a distance above 6 Å of the ACE2 binding surface. Moreover, L452R is facing the opposite side of the RBM region, while T478K is mainly exposed to the solvent and too far from ACE2 residue E87 (~15 Å), which makes it difficult to maintain steady interactions with ACE2. Both T478K and L452R roles are still uncertain. We can speculate that both types of mutations can improve RBM contacts with the negatively charged and hydrophilic ACE2 interface by making the RBD more positively charged and hydrophilic. However, as stated previously, T478K cannot interact with any ACE2 residues due to its constant exposure to the solvent and great distance from the ACE2 binding surface. On the other hand, T478K can play an important role in the trimeric conformation of the spike protein, where the introduction of a positive mutated residue, a lysine (K), might tighten the RBD interactions to reinforce the closed/inactive state. The L452R mutation is still debated in several studies[66,156] since it does not play a crucial role in the interaction with ACE2. However, a paper by Forest-Nault and co-workers[157] suggests that the L452R mutation abrogates a hydrophobic patch formed by residues L452, L492, and F490. The loss of this patch could impact the stability of the RBD and possibly its complexation with ACE2, where faster association and dissociation of the RBD have been observed.[157]

New hot spots have also been detected on the RBM surface due to the introduction of several mutations in each variant. The N501Y mutation is common to the Alpha, Beta, Gamma, and Omicron variants, indicating this mutation confers strong advantages to the SARS-CoV-2 virus. As shown in Figure 4, N501 residue has its binding free energy increased from −1.41 kcal mol^{-1} in the WT system or from −1.72 kcal mol^{-1} in the Delta variant to −3.37, −3.88, −3.58, and −3.97 kcal mol^{-1} in the Alpha, Beta, Gamma, and Omicron variants, respectively. In the crystallographic structure of WT RBD−ACE2 complex, N501 residue is known to interact with residue Y41 through a HB. It is also surrounded by the K353 hydrophobic alkyl chain and the Y41 benzene ring (Figure 5). The phenol group on the Y501 side chain in the N501Y mutant can interact through a cation−π interaction with the amine group of the K353 side chain and form a π−π stacking interaction with residue Y41. The higher binding affinity of the N501Y mutant with ACE2 is attributed to these additional stable intermolecular π-interactions with K353 and Y41 (Figure 5). Our results are also in agreement with other studies.[30,31] Thus, we can suggest that the N501Y mutation is responsible for improving the binding affinity of the RBD−ACE2 complex, which is directly linked to the enhanced transmissibility of the virus.[158] On the other hand, N501Y has been proven to have little effect on the neutralization of antibodies.[159]

Among the Omicron mutations, residue Q498R showed a high contributing energy (−8.49 kcal mol^{-1}),[160] which may be due to its interaction with residue D38 through the formation

https://doi.org/10.1021/acs.jpcb.3c01467
J. Phys. Chem. B XXXX, XXX, XXX−XXX

Table 3. Average RBD−ACE2 Non-covalent Interaction Occupancy (%) During MD Simulations (3 Trajectories for Each System of 100 ns Each)[a]

		HB and SB occupancy (%)					
RBD	ACE2	WT	Alpha	Beta	Gamma	Delta	Omicron
417	D30	**SB:94%**	**SB:80.0%**	NI (K417N)	NI (K417T)	**SB:98.0%**	NI (K417N)
484	K31	**SB:58.7%**	**SB:58.3%**	NI (E484K)	NI (E484K)	**SB:59.4%**	NI (E484A)
484	E75	NI	NI	**SB:19.7% (E484K)**	**SB:33.7% (E484K)**	NI	NI (E484A)
498	D38	NI	NI	NI	NI	NI	**SB:60.34% (Q498R)**
498	Q42	**HB:32.1%**	**HB:32.8%**	**HB:48.3%**	**HB:41.07%**	**HB:23.7%**	NI (Q498R)
498	K353	**HB:6.1%**	NI	NI	NI	**HB:11.9%**	NI (Q498R)
505	E37	**HB:71.2%**	**HB:47.1%**	**HB:45.9%**	**HB:34.04%**	**HB:56.4%**	NI (Y505H)
505	K353	NI	NI	NI	NI	NI	**HB:24.6% (Y505H)**

[a]All SB bonds were set to a cutoff distance of 10 Å. NI (no interactions) indicates no interaction was formed between residues or occupancies were lower than 5%. For each pair of residues, the type of interaction is specified (SB or HB). The mutated amino acid of each pair of residues is added between brackets under the occupancy rate.

of an SB (Figure 5). Y505H also had its contribution energy decreased (−1.26 kcal mol^{-1}).[160] The Omicron Y505H mutation causes the hydrogen bonding connections between WT RBD Y505 and ACE2 E37 and R393 to be disrupted, which might explain its lower decomposition energy (Figure 4). As for N440K, the location of the mutated residue is too far (∼15−19 Å) from ACE2 residue E329 to form an SB. It is also often exposed to the solvent and cannot interact correctly with the ACE2 interface region. However, like T478K, it might play a role in making the trimeric form of the spike protein stronger by enhancing the interaction of the neighboring RBDs.

E484 is also a shared mutation between several variants (Beta, Gamma, and Omicron) and a hallmark of numerous SARS-CoV-2 lineages. E484 is drastically mutated to a positively charged residue in the Beta and Gamma variants as a lysine (K). However, E484 is altered to an alanine (A) in the Omicron variant. These findings demonstrate E484's intricate influence and variety of behaviors. As a result, it is not surprising that variants with this mutation have high transmissibility as well as a high rate of antibody escape.[161] E484 is positioned on a highly flexible loop of the RBM region. In prior research[155] as well on other published papers,[150,162−164] we discovered that E484 had extremely positive decomposition energy, as shown in Figure 4 (15.74 kcal mol^{-1}) and was classified as an unfavorable hot spot for the interaction with ACE2. The Alpha (14.22 kcal mol^{-1}) and Delta (14.69 kcal mol^{-1}) variants still show the unfavorable tendency of E484 (Figure 4). This is primarily due to E484's proximity to negatively charged ACE2 residues like E35 and E75 (Figure 5). The E484K mutation, on the other hand, had a completely different effect, as it not only changed the nature of the residue, but also rendered E484K (−12.96 and −12.98 kcal mol^{-1} for the Beta and Gamma variants, respectively) a major positive and contributing hot spot for ACE2 binding. As shown in Table 1, the Beta and Gamma variants free energies increased significantly compared to the WT strain. This is consequently due to the important increase of the ΔE_{Elec} energies. Thus, E484K is able to bind E75 by forming a SB and using the flexible loop to create a more suitable environment (Figure 5). Both the Beta and Gamma versions carry the same E484K mutation, and particular attention has been given to these specific residue alterations. In fact, the E484K mutation may have a stronger transmissibility than the original strain.[161] Additionally, it revealed a reduction in the neutralization activity of some tested antibodies, which could affect how effective the existing vaccines are.[5,161,165] As a result, our findings indicated that the E484K mutation may enhance the RBD−ACE2 complex's binding affinity through more favorable electrostatic forces and tighter interactions on the RBM surface. These findings suggest that the E484K mutation-carrying variants are more transmissible. With several antibodies, other investigations have shown decreased binding affinities, which can result in an immune response escape. This may possibly be because the E484K mutation alters the electrostatic interaction on the RBD surface, decreasing the potency of antibodies. Alternatively, the Omicron E484A mutation did not have any impact on the decomposition energy but rather nullified it (−0.05 kcal mol^{-1}). Thus, the Omicron binding free energy is primarily due to a mixture of additional mutations that function as a compensation for the E484K positive effect.

Similarly, residue K417 is an important hot spot contributing favorably to the ACE2−RBD in the WT SARS-CoV-2 through its SB formation with residue D30 of ACE2 (Figure 5). K417 has been changed to an asparagine (N) in the Beta and Omicron variants, while it was altered to either an asparagine (N) or a threonine (T) in the Gamma variant. As shown in Figure 4, the decomposition energies of K417 are highly favorable compared to the other variants. Moreover, K417 of the Alpha variant shows a higher (−5.76 kcal mol^{-1}) energy than the one observed in the WT strain (−4.26 kcal mol^{-1}) and the Delta variant (−3.27 kcal mol^{-1}). This might be explained by the presence of surrounding mutations that impact the conformation of the RBD and the chemical environment of the residues. It is also clearly shown in Figure 4 that the alteration of the lysine (K) residue considerably reduces the energy of the amino acid. The decomposition energies of K417 have dropped to −0.33, −0.25, and −0.23 kcal mol^{-1}, for the Beta, Gamma, and Omicron variants, respectively. Overall, with the use of other favorable mutations, variants with an altered K417 were able to maintain high binding affinities with ACE2 (Table 2).[151]

In comparison with the WT, the mutated residue, residue Q493R did show a slight increase in its decomposition energy when comparing the WT with Omicron. However, no significant changes were observed in terms of structure. As shown in Figure 5, in both WT and Omicron systems, Q493R forms a HB with ACE2 residue E35. The non-significant decomposition energy might be due to the non-steady bond formed between those pairs of residues or the constant intramolecular SB interaction of E35 and K31, making Q493R−E35 HB less present.

https://doi.org/10.1021/acs.jpcb.3c01467
J. Phys. Chem. B XXXX, XXX, XXX−XXX

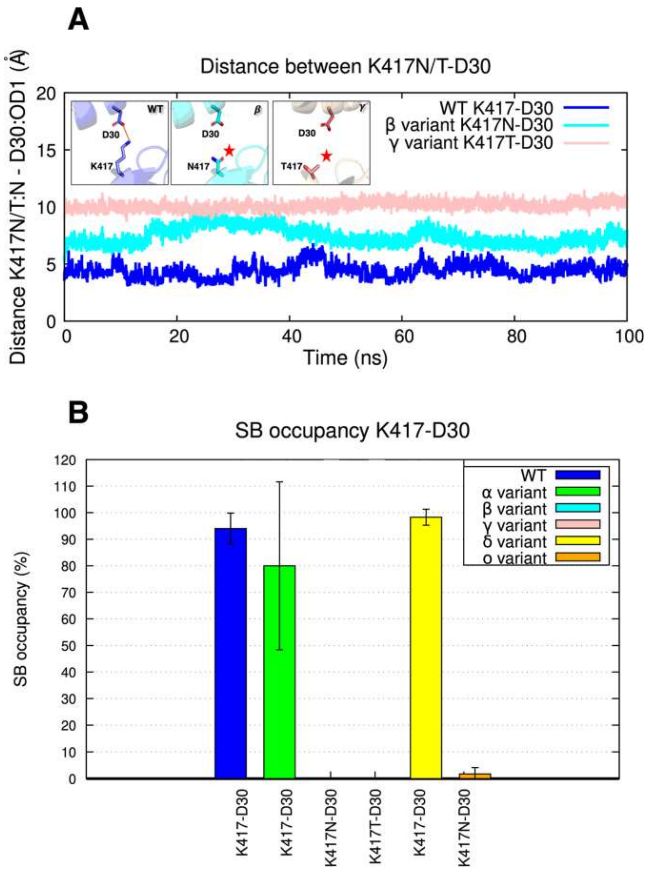


Figure 6. (A) Bar plots of the atomic distance within the formed SB between ACE2 D30 and WT, Beta, and Gamma K417N/T. (B) Histogram of the average SB occupancy (%) of D30-K417N/T in each system (WT, Alpha, Beta, Gamma, Delta, and Omicron). Mutations are marked with a red star. All SB occupancies are displayed with error bars.

In summary, out of all mutations that define the Alpha, Beta, and Gamma variants, N501Y promotes association with ACE2 through π-stacking and hydrogen bonding interactions, while E484K contributes through electrostatic attractions, resulting in a higher affinity for the double Gamma variant and the triple Beta variant. The third main mutation in the Beta and Omicron variants, K417N, acts as an unfavorable residue decreasing the ACE2-binding affinity. On the other hand, K417N might help control the infectivity of the different affected variants rather than focusing on tightening the ACE2 interaction. T478K and L452K mutations might be involved in boosting the Delta variant RBD's affinity for ACE2 by inducing conformational changes on the RBM binding regions, triggering an antibody escape mechanism. As for the rest of the mutations, mostly in the Omicron variant, it is crucial to keep an eye on their impacts because they could be incredibly important for antibody affinities and epitope key site determination. For instance, a combination of these hot spots might increase the strong affinity of the complex formed with ACE2 and RBD. The Delta variant might acquire new mutations like the N501Y or the Q498R, rendering the virus to bind more tightly to human cells. It is, of course, important to also consider other mutations that might be important in the immune escape strategies of SARS-CoV-2. Hence, it remains necessary to monitor the existing mutations and study their effects to predict new possible "gain-of-function" mutations or variants that might gain additional antibody resistance and drastic viral transmission, leading to more aggravating diseases.

3.4. HB and SB Analysis. To analyze the interactions between the ACE2 and the RBD and to highlight the various bonds that were formed and abolished within each system, HB and SB occupancy calculations were performed (Table 3). We chose to illustrate the most significant changes in each system by measuring HB and SB occupancy exceeding 5%.

The first SB formed between ACE2 D30 and RBD K417 is significantly present in the WT (94.0%), Alpha (80.0%), and Delta (98.0%) systems as no mutations were introduced to the 417 residue (Figure 5 and Table 3). However, because K417

https://doi.org/10.1021/acs.jpcb.3c01467
J. Phys. Chem. B XXXX, XXX, XXX−XXX

has been altered to an asparagine (N) or a threonine (T) in the remaining systems, the SB with D30 is no longer present (occupancy < 5%). The average atomic distance between the two residues was also calculated over the simulation time (3 × 100 ns), and the results were compared with those of the WT and Beta systems. Figure 6 illustrates a clear distinction between the two systems, with the distance between D30 and K417 remaining constant with an average distance value of 4 Å, while the distance in Beta is closer to 8−10 Å. As a result, we can assume that the K417N/T mutation, which causes the loss of one of the strongest interactions at the RBD interface, might constitute a convergent strategy employed to avoid being neutralized by antibodies.

Two other SBs are formed between residues E484−K31 and K484−E75 (Table 3). When E484 is non-mutated, a stable SB can be produced with ACE2 residue K31 because the two residues are close to one another. HB occupancy rates of the E484-K31 in the WT, Alpha, and Delta variants are 58.7, 58.3, and 59.4%, respectively. Also, for the majority of the simulation, another stable HB is produced between residues Q493 and K31, which explains why the SB only occurs 50% of the time over the trajectories. On the other hand, a newly formed SB can be observed with ACE2 residue E75 when E484 is changed to a lysine (K). This SB can be seen in both Beta (19.7%) and Gamma (33.7%) systems. This finding supports the dynamic RBD movements that allow for the formation of new bonds to enhance the molecular system's affinity. As for the Omicron variant, all bonds are abolished when residue E484 is mutated to an alanine (A). E484A, being a small and hydrophobic residue and mostly present in a negatively charged environment, has little to no interaction with the other surrounding residues of the binding interface (Figure 5C and Supporting Information Figure S13A,B).

The ACE2 residues D38, Q42, and K353 can form three distinct HB bonds with Q498, as illustrated in Table 3. Interestingly, Q498 interacts with residue Q42 through an HB throughout all systems (WT 32.1%, Alpha 32.8%, 48.3% Beta, 41.0% Gamma, and 23.7% Delta) with the exception of Omicron. Another important observation is that Q498 is able to form another HB with residue K353 in the WT and Delta systems (Table 3 and Supporting Information Figure S35E). These findings highlight how closely WT and Delta interact, as well as how similar their structural conformations are. On the other hand, to compensate for the loss of the HB originally formed with residues K353 and Q42, the Omicron Q498R mutation allowed the addition of a new SB interaction with residue D38 with a steady SB occupancy rate of 60.3% (Supporting Information Figure S13C).

Y505 did not experience any changes, apart from the Omicron variant, where it is mutated to an histidine (H) residue. Y505-E37 appears to be the most common HB, with occupancy rates of 71.2% in WT, 47.1% in Alpha, 45.9% in Beta, 34.0% in Gamma, and 56.4% in Delta (Table 3 and Supporting Information Figure S13F). To overcome the loss of the Y505−E37 HB in the Omicron system, the Y505H mutation interacts with the ACE2 K353 residue with a stable HB occupancy rate of 24.6% (Supporting Information Figure S13G).

A variety of computational studies have been conducted and continue to be performed, comparing the affinities of SARS-CoV-2 variants of concern with the ACE2 receptor. Several methods, such as MM-PBSA or MM-GBSA,[80,149−154] FEP,[166] and neural network models,[167] were applied in this research. These computational methods have made it possible to compare various variants of concern and predict future mutations, as seen in several neural network models.[167]

Nevertheless, despite the advances in this research, none have examined the precise pathogenic circulating variants of concern (WT, Alpha, Beta, Gamma, Delta, and Omicron) with the same structural details using the same approach for compatibility and reproducibility purposes, which makes our study distinctive. In addition, our findings are consistent with experimental data, providing further validation of our computational approach and increasing confidence in our results.[4,98,99,103] Furthermore, it is important to note that while some studies have compared some sets of variants, most of these did not undertake a comprehensive investigation of all variants of concern. Our study, on the other hand, seeks to close the gap by analyzing all the current variants of concern to have a better understanding of the interaction between ACE2 and RBD. When compared to other studies, our study revealed a comparable trend in affinity and residue decomposition energies. Our research could help in the development of new inhibitors by identifying the binding sites for antibodies that can assist in the prevention of the interaction between ACE2 and RBD. As a result, our study provides a critical understanding of the relationship between SARS-CoV-2 variants of concern and the ACE2 receptor, which may aid in the development of strategies to prevent COVID-19.

In summary, our results demonstrate the newly formed HB and SB interactions resulting from various mutations and being introduced at critical hot spots of the RBD interface. Most of the variations were seen in the RBD binding surface residues K417N/T, E484K/A, Q498R, N501Y, and Y505H. It also provides information on the selection of mutations for which SARS-CoV-2 inquired, primarily to increase its affinity for the ACE2 protein and avoid being eliminated by RBD-targeted antibodies.

4. CONCLUSIONS

In this work, we studied the impact of the various mutations of the SARS-CoV-2 RBD WT, Alpha, Beta, Gamma, Delta, and Omicron systems and their affinities with the ACE2 protein. This is the first analysis to overlook all variants and their effects with an identical combination of different computational methods. Molecular modeling enabled the construction of highly similar molecular models of the different RBD−ACE2 systems that fully comply with crystallographic structures. By maintaining the same protocol throughout this study, we were able to ensure the clarity of our work and the reliability of our findings. As a result, even though we were aware that the Omicron model structure would be difficult to accurately model, it contains most of the known mutations in the SARS-CoV-2 virus. Yet, by constructing an Omicron model structure, we were able to confirm the consistency of our approach. In fact, it has been demonstrated that the Omicron model and the resolved cryo-EM Omicron variant structures give results that are very consistent across all the different analyses that we performed. MD simulation results showed the stability of all systems and pointed out flexible regions on both proteins through rmsd and RMSF analyses. The results from the MM-PBSA analysis indicated new hot spots playing crucial roles in enhancing RBD affinity with ACE2. It also showed the impact of the different mutated regions and their ways of increasing the spike protein affinity for ACE2 and mainly escaping RBD-targeted neutralizing antibodies. K417N/T SB has been

https://doi.org/10.1021/acs.jpcb.3c01467
J. Phys. Chem. B XXXX, XXX, XXX−XXX

observed to be abolished when mutated to an asparagine (N) or a threonine (T) residue. The L452R residue is not involved in any interaction; however, it is believed that its mutation is involved in the complex association and dissociation rates. E484K/A mutations are also involved in the formation of a steadier SB with residue E75 when mutated to a lysine (K) in the Beta and Gamma variants and show no interaction when mutated to an alanine (A) in the Omicron variant. The Q498R mutation is involved in the creation of a new SB with residue D38, one of the numerous newly formed bonds in the Omicron variant. The N501Y mutation, which is present in nearly all variants, interacts quite strongly with ACE2 residues through hydrophobic interactions, which increase almost more than twice the decomposition energy of the originally unchanged residue. Therefore, it acts as a crucial hot spot in the mutated ACE2−RBD complex. The Y505H mutation in the Omicron variant shows the abolishment of Y505−R393 and Y505−E37 HB in the rest of the systems and its compensation with the formation of a more dominant interaction with ACE2 residue K353.

Overall, our results demonstrate a strong connection between the computed binding free energies and decomposition energy trends of each variant and its viral effect. It also gives a general view of all variants of concern and their effects at a molecular level, focusing on a variety of key mutations that becoming the new hot spots of the RBD and might be responsible for additional effects in new variants if combined together.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.3c01467.

Protein rmsd on experimental structures and during the MD simulations; RMSF during the MD simulations; MM-PBSA energy terms calculated during the MD simulations; atomic distances extracted during the MD simulations (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Anne-Claude Camproux − *Université de Paris, CNRS, INSERM, Unité de Biologie Fonctionnelle et Adaptative, F-75013 Paris, France*; Phone: +33-1-57-27-83-77; Email: anne-claude.camproux@u-paris.fr

Gautier Moroy − *Université de Paris, CNRS, INSERM, Unité de Biologie Fonctionnelle et Adaptative, F-75013 Paris, France*; orcid.org/0000-0002-8973-0477; Phone: +33-1-57-27-83-85; Email: gautier.moroy@u-paris.fr

Authors

Mariem Ghoula − *Université de Paris, CNRS, INSERM, Unité de Biologie Fonctionnelle et Adaptative, F-75013 Paris, France*

Audrey Deyawe Kongmeneck − *Université de Paris, CNRS, INSERM, Unité de Biologie Fonctionnelle et Adaptative, F-75013 Paris, France*

Rita Eid − *Université de Paris, CNRS, INSERM, Unité de Biologie Fonctionnelle et Adaptative, F-75013 Paris, France*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.3c01467

Author Contributions

This manuscript was written through the contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

The PyMOL software is available to noncommercial users under a distribution-specific license on https://pymol.org/2/. The GROMACS software is free of charge and available on https://www.gromacs.org. The VMD software is available to noncommercial users under a distribution-specific license on http://www.ks.uiuc.edu/Research/vmd/. The g_mmpbsa software is free of charge and available on https://rashmikumari.github.io/g_mmpbsa/.

■ ACKNOWLEDGMENTS

This work was supported by the Agence Nationale de la Recherche (PIF21 project, no. ANR-19-CE18-0023). The authors gratefully acknowledge the financial support of the Université de Paris, the CNRS Institute, and the INSERM Institute.

■ REFERENCES

(1) https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/ (accessed April 10, 2023).

(2) Harvey, W. T.; Carabelli, A. M.; Jackson, B.; Gupta, R. K.; Thomson, E. C.; Harrison, E. M.; Ludden, C.; Reeve, R.; Rambaut, A.; Peacock, S. J.; Robertson, D. L.; Peacock, S. J.; Robertson, D. L. SARS-CoV-2 Variants, Spike Mutations and Immune Escape. *Nat. Rev. Microbiol.* **2021**, *19*, 409−424.

(3) Hu, J.; Peng, P.; Cao, X.; Wu, K.; Chen, J.; Wang, K.; Tang, N.; Huang, A. Increased Immune Escape of the New SARS-CoV-2 Variant of Concern Omicron. *Cell. Mol. Immunol.* **2022**, *19*, 293−295.

(4) McCallum, M.; Walls, A. C.; Sprouse, K. R.; Bowen, J. E.; Rosen, L. E.; Dang, H. V.; De Marco, A.; Franko, N.; Tilles, S. W.; Logue, J.; et al. Molecular Basis of Immune Evasion by the Delta and Kappa SARS-CoV-2 Variants. *Science* **2021**, *374*, 1621−1626.

(5) Weisblum, Y.; Schmidt, F.; Zhang, F.; DaSilva, J.; Poston, D.; Lorenzi, J. C.; Muecksch, F.; Rutkowska, M.; Hoffmann, H.-H.; Michailidis, E.; et al. Escape from Neutralizing Antibodies by SARS-CoV-2 Spike Protein Variants. *eLife* **2020**, *9*, No. e61312.

(6) Wang, Q.; Zhang, Y.; Wu, L.; Niu, S.; Song, C.; Zhang, Z.; Lu, G.; Qiao, C.; Hu, Y.; Yuen, K.-Y.; et al. Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell* **2020**, *181*, 894−904.e9.

(7) Shang, J.; Ye, G.; Shi, K.; Wan, Y.; Luo, C.; Aihara, H.; Geng, Q.; Auerbach, A.; Li, F. Structural Basis of Receptor Recognition by SARS-CoV-2. *Nature* **2020**, *581*, 221−224.

(8) Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; et al. Structure of the SARS-CoV-2 Spike Receptor-Binding Domain Bound to the ACE2 Receptor. *Nature* **2020**, *581*, 215−220.

(9) Zhang, J.; Xiao, T.; Cai, Y.; Chen, B. Structure of SARS-CoV-2 Spike Protein. *Curr. Opin. Virol.* **2021**, *50*, 173−182.

(10) Cai, Y.; Zhang, J.; Xiao, T.; Peng, H.; Sterling, S. M.; Walsh, R. M.; Rawson, S.; Rits-Volloch, S.; Chen, B. Distinct Conformational States of SARS-CoV-2 Spike Protein. *Science* **2020**, *369*, 1586−1592.

(11) 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**, *367*, 1444−1448.

(12) Li, F. Receptor Recognition Mechanisms of Coronaviruses: A Decade of Structural Studies. *J. Virol.* **2015**, *89*, 1954−1964.

(13) Wan, Y.; Shang, J.; Graham, R.; Baric, R. S.; Li, F. Receptor Recognition by the Novel Coronavirus from Wuhan: An Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *J. Virol.* **2020**, *94*, No. e00127.

(14) Li, W.; Zhang, C.; Sui, J.; Kuhn, J. H.; Moore, M. J.; Luo, S.; Wong, S.-K.; Huang, I.-C.; Xu, K.; Vasilieva, N.; et al. Receptor and Viral Determinants of SARS-Coronavirus Adaptation to Human ACE2. *EMBO J.* **2005**, *24*, 1634−1643.
(15) Zhang, H.; Penninger, J. M.; Li, Y.; Zhong, N.; Slutsky, A. S. Angiotensin-Converting Enzyme 2 (ACE2) as a SARS-CoV-2 Receptor: Molecular Mechanisms and Potential Therapeutic Target. *Intensive Care Med.* **2020**, *46*, 586−590.
(16) Scialo, F.; Daniele, A.; Amato, F.; Pastore, L.; Matera, M. G.; Cazzola, M.; Castaldo, G.; Bianco, A. ACE2: The Major Cell Entry Receptor for SARS-CoV-2. *Lung* **2020**, *198*, 867−877.
(17) Yang, Q.; Hughes, T. A.; Kelkar, A.; Yu, X.; Cheng, K.; Park, S.; Huang, W.-C.; Lovell, J. F.; Neelamegham, S. Inhibition of SARS-CoV-2 Viral Entry upon Blocking N- and O-Glycan Elaboration. *eLife* **2020**, *9*, No. e61552.
(18) Huo, J.; Le Bas, A.; Ruza, R. R.; Duyvesteyn, H. M. E.; Mikolajek, H.; Malinauskas, T.; Tan, T. K.; Rijal, P.; Dumoux, M.; Ward, P. N.; et al. Neutralizing Nanobodies Bind SARS-CoV-2 Spike RBD and Block Interaction with ACE2. *Nat. Struct. Mol. Biol.* **2020**, *27*, 846−854.
(19) Linsky, T. W.; Vergara, R.; Codina, N.; Nelson, J. W.; Walker, M. J.; Su, W.; Barnes, C. O.; Hsiang, T.-Y.; Esser-Nobis, K.; Yu, K.; et al. De Novo Design of Potent and Resilient HACE2 Decoys to Neutralize SARS-CoV-2. *Science* **2020**, *370*, 1208−1214.
(20) Cao, L.; Goreshnik, I.; Coventry, B.; Case, J. B.; Miller, L.; Kozodoy, L.; Chen, R. E.; Carter, L.; Walls, A. C.; Park, Y.-J.; et al. De Novo Design of Picomolar SARS-CoV-2 Miniprotein Inhibitors. *Science* **2020**, *370*, 426−431.
(21) Xiong, J.; Xiang, Y.; Huang, Z.; Liu, X.; Wang, M.; Ge, G.; Chen, H.; Xu, J.; Zheng, M.; Chen, L. Structure-Based Virtual Screening and Identification of Potential Inhibitors of SARS-CoV-2 S-RBD and ACE2 Interaction. *Front. Chem.* **2021**, *9*, 740702.
(22) Barton, M. I.; MacGowan, S. A.; Kutuzov, M. A.; Dushek, O.; Barton, G. J.; van der Merwe, P. A. Effects of Common Mutations in the SARS-CoV-2 Spike RBD and Its Ligand, the Human ACE2 Receptor on Binding Affinity and Kinetics. *eLife* **2021**, *10*, No. e70658.
(23) Guruprasad, L. Human SARS CoV -2 Spike Protein Mutations. *Proteins* **2021**, *89*, 569−576.
(24) Mohammadi, M.; Shayestehpour, M.; Mirzaei, H. The Impact of Spike Mutated Variants of SARS-CoV2 [Alpha, Beta, Gamma, Delta, and Lambda] on the Efficacy of Subunit Recombinant Vaccines. *Braz. J. Infect. Dis.* **2021**, *25*, 101606.
(25) Davies, N. G.; Abbott, S.; Barnard, R. C.; Jarvis, C. I.; Kucharski, A. J.; Munday, J. D.; Pearson, C. A. B.; Russell, T. W.; Tully, D. C.; Washburne, A. D.; et al. Estimated Transmissibility and Impact of SARS-CoV-2 Lineage B.1.1.7 in England. *Science* **2021**, *372*, No. eabg3055.
(26) Courjon, J.; Contenti, J.; Demonchy, E.; Levraut, J.; Barbry, P.; Rios, G.; Dellamonica, J.; Chirio, D.; Bonnefoy, C.; Giordanengo, V.; et al. COVID-19 Patients Age, Comorbidity Profiles and Clinical Presentation Related to the SARS-CoV-2 UK-Variant Spread in the Southeast of France. *Sci. Rep.* **2021**, *11*, 18456.
(27) Negi, S. S.; Schein, C. H.; Braun, W. Regional and Temporal Coordinated Mutation Patterns in SARS-CoV-2 Spike Protein Revealed by a Clustering and Network Analysis. *Sci. Rep.* **2022**, *12*, 1128.
(28) Ostrov, D. A. Structural Consequences of Variation in SARS-CoV-2 B.1.1.7. *J. Cell. Immunol* **2021**, *3*, 103.
(29) Bayarri-Olmos, R.; Johnsen, L. B.; Idorn, M.; Reinert, L. S.; Rosbjerg, A.; Vang, S.; Hansen, C. B.; Helgstrand, C.; Bjelke, J. R.; Bak-Thomsen, T.; et al. The Alpha/B.1.1.7 SARS-CoV-2 Variant Exhibits Significantly Higher Affinity for ACE-2 and Requires Lower Inoculation Doses to Cause Disease in K18-HACE2 Mice. *eLife* **2021**, *10*, No. e70002.
(30) Socher, E.; Conrad, M.; Heger, L.; Paulsen, F.; Sticht, H.; Zunke, F.; Arnold, P. Computational Decomposition Reveals Reshaping of the SARS-CoV-2−ACE2 Interface among Viral Variants Expressing the N501Y Mutation. *J. Cell. Biochem.* **2021**, *122*, 1863−1872.
(31) Tian, F.; Tong, B.; Sun, L.; Shi, S.; Zheng, B.; Wang, Z.; Dong, X.; Zheng, P. N501Y Mutation of Spike Protein in SARS-CoV-2 Strengthens Its Binding to Receptor ACE2. *eLife* **2021**, *10*, No. e69091.
(32) Zhang, L.; Jackson, C. B.; Mou, H.; Ojha, A.; Rangarajan, E. S.; Izard, T.; Farzan, M.; Choe, H. The D614G Mutation in the SARS-CoV-2 Spike Protein Reduces S1 Shedding and Increases Infectivity. **2020**, Preprint, bioRxiv: 2020.06.12.148726
(33) Plante, J. A.; Liu, Y.; Liu, J.; Xia, H.; Johnson, B. A.; Lokugamage, K. G.; Zhang, X.; Muruato, A. E.; Zou, J.; Fontes-Garfias, C. R.; et al. Spike Mutation D614G Alters SARS-CoV-2 Fitness. *Nature* **2021**, *592*, 116−121.
(34) Zhang, L.; Jackson, C. B.; Mou, H.; Ojha, A.; Peng, H.; Quinlan, B. D.; Rangarajan, E. S.; Pan, A.; Vanderheiden, A.; Suthar, M. S.; et al. SARS-CoV-2 Spike-Protein D614G Mutation Increases Virion Spike Density and Infectivity. *Nat. Commun.* **2020**, *11*, 6013.
(35) Hou, Y. J.; Chiba, S.; Halfmann, P.; Ehre, C.; Kuroda, M.; Dinnon, K. H.; Leist, S. R.; Schäfer, A.; Nakajima, N.; Takahashi, K.; et al. SARS-CoV-2 D614G Variant Exhibits Efficient Replication Ex Vivo and Transmission in Vivo. *Science* **2020**, *370*, 1464−1468.
(36) Daniloski, Z.; Jordan, T. X.; Ilmain, J. K.; Guo, X.; Bhabha, G.; tenOever, B. R.; Sanjana, N. E. The Spike D614G Mutation Increases SARS-CoV-2 Infection of Multiple Human Cell Types. *eLife* **2021**, *10*, No. e65365.
(37) Korber, B.; Fischer, W. M.; Gnanakaran, S.; Yoon, H.; Theiler, J.; Abfalterer, W.; Hengartner, N.; Giorgi, E. E.; Bhattacharya, T.; Foley, B.; et al. Tracking Changes in SARS-CoV-2 Spike: Evidence That D614G Increases Infectivity of the COVID-19 Virus. *Cell* **2020**, *182*, 812−827.e19.
(38) Lubinski, B.; Fernandes, M. H. V.; Frazier, L.; Tang, T.; Daniel, S.; Diel, D. G.; Jaimes, J. A.; Whittaker, G. R. Functional Evaluation of the P681H Mutation on the Proteolytic Activation of the SARS-CoV-2 Variant B.1.1.7 (Alpha) Spike. *iScience* **2022**, *25*, 103589.
(39) Pascall, D. J.; Vink, E.; Blacow, R.; Bulteel, N.; Campbell, A.; Campbell, R.; Clifford, S.; Davis, C.; Da Silva Filipe, A.; Sakka, N. E.; Fjodorova, L. et al. The SARS-CoV-2 Alpha Variant Is Associated with Increased Clinical Severity of COVID-19 in Scotland: A Genomics-Based Retrospective Cohort Analysis. **2021**, Preprint; Medrxiv, 2021.08.17.21260128 (accessed April 10, 2023).
(40) Bekliz, M.; Adea, K.; Vetter, P.; Eberhardt, C. S.; Hosszu-Fellous, K.; Vu, D.-L.; Puhach, O.; Essaidi-Laziosi, M.; Waldvogel-Abramowski, S.; Stephan, C.; et al. Neutralization Capacity of Antibodies Elicited through Homologous or Heterologous Infection or Vaccination against SARS-CoV-2 VOCs. *Nat. Commun.* **2022**, *13*, 3840.
(41) Bonura, F.; Genovese, D.; Amodio, E.; Calamusa, G.; Sanfilippo, G. L.; Cacioppo, F.; Giammanco, G. M.; De Grazia, S.; Ferraro, D. Neutralizing Antibodies Response against SARS-CoV-2 Variants of Concern Elicited by Prior Infection or MRNA BNT162b2 Vaccination. *Vaccines* **2022**, *10*, 874.
(42) Evans, J. P.; Zeng, C.; Carlin, C.; Lozanski, G.; Saif, L. J.; Oltz, E. M.; Gumina, R. J.; Liu, S.-L. Neutralizing Antibody Responses Elicited by SARS-CoV-2 MRNA Vaccination Wane over Time and Are Boosted by Breakthrough Infection. *Sci. Transl. Med.* **2022**, *14*, No. eabn8057.
(43) Lynch, K. L.; Zhou, S.; Kaul, R.; Walker, R.; Wu, A. H. Evaluation of Neutralizing Antibodies against SARS-CoV-2 Variants after Infection and Vaccination Using a Multiplexed Surrogate Virus Neutralization Test. *Clin. Chem.* **2022**, *68*, 702−712.
(44) Hernández-Luis, P.; Aguilar, R.; Pelegrin-Pérez, J.; Ruiz-Olalla, G.; García-Basteiro, A. L.; Tortajada, M.; Moncunill, G.; Dobaño, C.; Angulo, A.; Engel, P. Decreased and Heterogeneous Neutralizing Antibody Responses Against RBD of SARS-CoV-2 Variants After MRNA Vaccination. *Front. Immunol.* **2022**, *13*, 816389.
(45) Chen, Z.; Zhang, P.; Matsuoka, Y.; Tsybovsky, Y.; West, K.; Santos, C.; Boyd, L. F.; Nguyen, H.; Pomerenke, A.; Stephens, T. et al. Extremely Potent Monoclonal Antibodies Neutralize Omicron

https://doi.org/10.1021/acs.jpcb.3c01467
J. Phys. Chem. B XXXX, XXX, XXX−XXX

and Other SARS-CoV-2 Variants. **2022**, Preprint; Medrxiv, 2022.01.12.22269023 (accessed April 10, 2023).
(46) Dupont, L.; Snell, L. B.; Graham, C.; Seow, J.; Merrick, B.; Lechmere, T.; Maguire, T. J. A.; Hallett, S. R.; Pickering, S.; Charalampous, T.; et al. Neutralizing Antibody Activity in Convalescent Sera from Infection in Humans with SARS-CoV-2 and Variants of Concern. *Nat. Microbiol.* **2021**, *6*, 1433−1442.
(47) Narowski, T. M.; Raphel, K.; Adams, L. E.; Huang, J.; Vielot, N. A.; Jadi, R.; de Silva, A. M.; Baric, R. S.; Lafleur, J. E.; Premkumar, L. SARS-CoV-2 MRNA Vaccine Induces Robust Specific and Cross-Reactive IgG and Unequal Neutralizing Antibodies in Naive and Previously Infected People. *Cell Rep.* **2022**, *38*, 110336.
(48) Shrestha, L. B.; Tedla, N.; Bull, R. A. Broadly-Neutralizing Antibodies Against Emerging SARS-CoV-2 Variants. *Front. Immunol.* **2021**, *12*, 752003.
(49) Moss, D. L.; Rappaport, J. SARS-CoV-2 Beta Variant Substitutions Alter Spike Glycoprotein Receptor Binding Domain Structure and Stability. *J. Biol. Chem.* **2021**, *297*, 101371.
(50) Wang, Z.; Schmidt, F.; Weisblum, Y.; Muecksch, F.; Barnes, C. O.; Finkin, S.; Schaefer-Babajew, D.; Cipolla, M.; Gaebler, C.; Lieberman, J. A.; et al. MRNA Vaccine-Elicited Antibodies to SARS-CoV-2 and Circulating Variants. *Nature* **2021**, *592*, 616−622.
(51) Reincke, S. M.; Yuan, M.; Kornau, H.-C.; Corman, V. M.; van Hoof, S.; Sánchez-Sendin, E.; Ramberger, M.; Yu, W.; Hua, Y.; Tien, H.; et al. SARS-CoV-2 Beta Variant Infection Elicits Potent Lineage-Specific and Cross-Reactive Antibodies. *Science* **2022**, *375*, 782−787.
(52) Abu-Raddad, L. J.; Chemaitelly, H.; Ayoub, H. H.; Yassine, H. M.; Benslimane, F. M.; Al Khatib, H. A.; Tang, P.; Hasan, M. R.; Coyle, P.; AlMukdad, S.; Al Kanaani, Z.; et al. Severity, Criticality, and Fatality of the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Beta Variant. *Clin. Infect. Dis.* **2021**, *75*, e1188−e1191.
(53) Radvak, P.; Kwon, H.-J.; Kosikova, M.; Ortega-Rodriguez, U.; Xiang, R.; Phue, J.-N.; Shen, R.-F.; Rozzelle, J.; Kapoor, N.; Rabara, T.; et al. SARS-CoV-2 B.1.1.7 (Alpha) and B.1.351 (Beta) Variants Induce Pathogenic Patterns in K18-HACE2 Transgenic Mice Distinct from Early Strains. *Nat. Commun.* **2021**, *12*, 6559.
(54) Singer, S. R.; Angulo, F. J.; Swerdlow, D. L.; McLaughlin, J. M.; Hazan, I.; Ginish, N.; Anis, E.; Mendelson, E.; Mor, O.; Zuckerman, N. S.; Erster, O.; et al. Effectiveness of BNT162b2 MRNA COVID-19 Vaccine against SARS-CoV-2 Variant Beta (B.1.351) among Persons Identified through Contact Tracing in Israel: A Prospective Cohort Study. *eClinicalMedicine* **2021**, *42*, 101190.
(55) Corbett, K. S.; Gagne, M.; Wagner, D. A.; O' Connell, S.; Narpala, S. R.; Flebbe, D. R.; Andrew, S. F.; Davis, R. L.; Flynn, B.; Johnston, T. S.; et al. Protection against SARS-CoV-2 Beta Variant in MRNA-1273 Vaccine−Boosted Nonhuman Primates. *Science* **2021**, *374*, 1343−1353.
(56) Souza, P. F. N.; Mesquita, F. P.; Amaral, J. L.; Landim, P. G. C.; Lima, K. R. P.; Costa, M. B.; Farias, I. R.; Belém, M. O.; Pinto, Y. O.; Moreira, H. H. T.; et al. The Spike Glycoprotein of SARS-CoV-2: A Review of How Mutations of Spike Glycoproteins Have Driven the Emergence of Variants with High Transmissibility and Immune Escape. *Int. J. Biol. Macromol.* **2022**, *208*, 105−125.
(57) Hirotsu, Y.; Omata, M. Discovery of a SARS-CoV-2 Variant from the P.1 Lineage Harboring K417T/E484K/N501Y Mutations in Kofu, Japan. *J. Infect.* **2021**, *82*, 276−316.
(58) Faria, N. R.; Mellan, T. A.; Whittaker, C.; Claro, I. M.; Candido, D. d. S.; Mishra, S.; Crispim, M. A. E.; Sales, F. C. S.; Hawryluk, I.; McCrone, J. T.; et al. Genomics and Epidemiology of the P.1 SARS-CoV-2 Lineage in Manaus, Brazil. *Science* **2021**, *372*, 815−821.
(59) Liu, H.; Wei, P.; Kappler, J. W.; Marrack, P.; Zhang, G. SARS-CoV-2 Variants of Concern and Variants of Interest Receptor Binding Domain Mutations and Virus Infectivity. *Front. Immunol.* **2022**, *13*, 825256.
(60) Escalera, A.; Gonzalez-Reiche, A. S.; Aslam, S.; Mena, I.; Laporte, M.; Pearl, R. L.; Fossati, A.; Rathnasinghe, R.; Alshammary, H.; van de Guchte, A.; et al. Mutations in SARS-CoV-2 Variants of Concern Link to Increased Spike Cleavage and Virus Transmission. *Cell Host Microbe* **2022**, *30*, 373.
(61) Baum, A.; Fulton, B. O.; Wloga, E.; Copin, R.; Pascal, K. E.; Russo, V.; Giordano, S.; Lanza, K.; Negron, N.; Ni, M.; et al. Antibody Cocktail to SARS-CoV-2 Spike Protein Prevents Rapid Mutational Escape Seen with Individual Antibodies. *Science* **2020**, *369*, 1014−1018.
(62) Fernández, J.; Bruneau, N.; Fasce, R.; Martín, H. S.; Balanda, M.; Bustos, P.; Ulloa, S.; Mora, J.; Ramírez, E. Neutralization of Alpha, Gamma, and D614G SARS-CoV-2 Variants by CoronaVac Vaccine-induced Antibodies. *J. Med. Virol.* **2022**, *94*, 399−403.
(63) Planas, D.; Veyer, D.; Baidaliuk, A.; Staropoli, I.; Guivel-Benhassine, F.; Rajah, M. M.; Planchais, C.; Porrot, F.; Robillard, N.; Puech, J.; et al. Reduced Sensitivity of SARS-CoV-2 Variant Delta to Antibody Neutralization. *Nature* **2021**, *596*, 276−280.
(64) Motozono, C.; Toyoda, M.; Zahradnik, J.; Saito, A.; Nasser, H.; Tan, T. S.; Ngare, I.; Kimura, I.; Uriu, K.; Kosugi, Y.; et al. SARS-CoV-2 Spike L452R Variant Evades Cellular Immunity and Increases Infectivity. *Cell Host Microbe* **2021**, *29*, 1124−1136.e11.
(65) Yi, C.; Sun, X.; Ling, Z.; Sun, B. Jigsaw Puzzle of SARS-CoV-2 RBD Evolution and Immune Escape. *Cell. Mol. Immunol.* **2022**, *19*, 848−851.
(66) Tchesnokova, V.; Kulasekara, H.; Larson, L.; Bowers, V.; Rechkina, E.; Kisiela, D.; Sledneva, Y.; Choudhury, D.; Maslova, I.; Deng, K.; et al. Acquisition of the L452R Mutation in the ACE2-Binding Interface of Spike Protein Triggers Recent Massive Expansion of SARS- CoV-2 Variants. *J. Clin. Microbiol.* **2021**, *59*, 10.
(67) Kuzmina, A.; Wattad, S.; Khalaila, Y.; Ottolenghi, A.; Rosental, B.; Engel, S.; Rosenberg, E.; Taube, R. SARS CoV-2 Delta Variant Exhibits Enhanced Infectivity and a Minor Decrease in Neutralization Sensitivity to Convalescent or Post-Vaccination Sera. *iScience* **2021**, *24*, 103467.
(68) Tian, D.; Sun, Y.; Zhou, J.; Ye, Q. The Global Epidemic of the SARS-CoV-2 Delta Variant, Key Spike Mutations and Immune Escape. *Front. Immunol.* **2021**, *12*, 751778.
(69) Jhun, H.; Park, H.-Y.; Hisham, Y.; Song, C.-S.; Kim, S. SARS-CoV-2 Delta (B.1.617.2) Variant: A Unique T478K Mutation in Receptor Binding Motif (RBM) of Spike Gene. *Immune Netw.* **2021**, *21*, No. e32.
(70) Cherian, S.; Potdar, V.; Jadhav, S.; Yadav, P.; Gupta, N.; Das, M.; Rakshit, P.; Singh, S.; Abraham, P.; Panda, S.; Team, N. SARS-CoV-2 Spike Mutations, L452R, T478K, E484Q and P681R, in the Second Wave of COVID-19 in Maharashtra, India. *Microorganisms* **2021**, *9*, 1542.
(71) Dhawan, M.; Sharma, A.; Priyanka; Thakur, N.; Rajkhowa, T. K.; Choudhary, O. P. Delta Variant (B.1.617.2) of SARS-CoV-2: Mutations, Impact, Challenges and Possible Solutions. *Hum. Vaccines Immunother.* **2022**, *18*, 2068883.
(72) Goher, S. S.; Ali, F.; Amin, M. The Delta Variant Mutations in the Receptor Binding Domain of SARS-CoV-2 Show Enhanced Electrostatic Interactions with the ACE2. *Med. Drug Discovery* **2022**, *13*, 100114.
(73) Liu, Y.; Liu, J.; Johnson, B. A.; Xia, H.; Ku, Z.; Schindewolf, C.; Widen, S. G.; An, Z.; Weaver, S. C.; Menachery, V. D.; et al. Delta Spike P681R Mutation Enhances SARS-CoV-2 Fitness over Alpha Variant. *Cell Rep.* **2022**, *39*, 110829.
(74) Saito, A.; Irie, T.; Suzuki, R.; Maemura, T.; Nasser, H.; Uriu, K.; Kosugi, Y.; Shirakawa, K.; Sadamasu, K.; Kimura, I.; et al. Enhanced Fusogenicity and Pathogenicity of SARS-CoV-2 Delta P681R Mutation. *Nature* **2022**, *602*, 300−306.
(75) Shiehzadegan, S.; Alaghemand, N.; Fox, M.; Venketaraman, V. Analysis of the Delta Variant B.1.617.2 COVID-19. *Clin. Pract.* **2021**, *11*, 778−784.
(76) Zhan, Y.; Yin, H.; Yin, J.-Y. B. B.1.617.2 (Delta) Variant of SARS-CoV-2: features, transmission and potential strategies. *Int. J. Biol. Sci.* **2022**, *18*, 1844−1851.
(77) Bian, L.; Gao, Q.; Gao, F.; Wang, Q.; He, Q.; Wu, X.; Mao, Q.; Xu, M.; Liang, Z. Impact of the Delta Variant on Vaccine Efficacy and Response Strategies. *Expert Rev. Vaccines* **2021**, *20*, 1201−1209.

https://doi.org/10.1021/acs.jpcb.3c01467
J. Phys. Chem. B XXXX, XXX, XXX−XXX

(78) Callaway, E. Heavily Mutated Omicron Variant Puts Scientists on Alert. *Nature* **2021**, *600*, 21.
(79) Lupala, C. S.; 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**, *590*, 34−41.
(80) da Costa, C. H. S.; de Freitas, C. A. B.; Alves, C. N.; Lameira, J. Assessment of Mutations on RBD in the Spike Protein of SARS-CoV-2 Alpha, Delta and Omicron Variants. *Sci. Rep.* **2022**, *12*, 8540.
(81) Wu, L.; Zhou, L.; Mo, M.; Liu, T.; Wu, C.; Gong, C.; Lu, K.; Gong, L.; Zhu, W.; Xu, Z. SARS-CoV-2 Omicron RBD Shows Weaker Binding Affinity than the Currently Dominant Delta Variant to Human ACE2. *Signal Transduction Targeted Ther.* **2022**, *7*, 8.
(82) Syed, A. M.; Ciling, A.; Taha, T. Y.; Chen, I. P.; Khalid, M. M.; Sreekumar, B.; Chen, P.-Y.; Kumar, G. R.; Suryawanshi, R.; Silva, I.; et al. Omicron Mutations Enhance Infectivity and Reduce Antibody Neutralization of SARS-CoV-2 Virus-like Particles. *Proc. Natl. Acad. Sci. U.S.A.* **2022**, *119*, No. e2200592119.
(83) Miller, N. L.; Clark, T.; Raman, R.; Sasisekharan, R. Insights on the Mutational Landscape of the SARS-CoV-2 Omicron Variant Receptor-Binding Domain. *Cell Rep. Med.* **2022**, *3*, 100527.
(84) Fan, Y.; Li, X.; Zhang, L.; Wan, S.; Zhang, L.; Zhou, F. SARS-CoV-2 Omicron Variant: Recent Progress and Future Perspectives. *Signal Transduction Targeted Ther.* **2022**, *7*, 141.
(85) Menni, C.; Valdes, A. M.; Polidori, L.; Antonelli, M.; Penamakuri, S.; Nogal, A.; Louca, P.; May, A.; Figueiredo, J. C.; Hu, C.; et al. Symptom Prevalence, Duration, and Risk of Hospital Admission in Individuals Infected with SARS-CoV-2 during Periods of Omicron and Delta Variant Dominance: A Prospective Observational Study from the ZOE COVID Study. *Lancet* **2022**, *399*, 1618−1624.
(86) Cao, Y.; Yisimayi, A.; Jian, F.; Song, W.; Xiao, T.; Wang, L.; Du, S.; Wang, J.; Li, Q.; Chen, X.; et al. BA.2.12.1, BA.4 and BA.5 Escape Antibodies Elicited by Omicron Infection. *Nature* **2022**, *608*, 593−602.
(87) Arora, P.; Zhang, L.; Krüger, N.; Rocha, C.; Sidarovich, A.; Schulz, S.; Kempf, A.; Graichen, L.; Moldenhauer, A.-S.; Cossmann, A.; et al. SARS-CoV-2 Omicron Sublineages Show Comparable Cell Entry but Differential Neutralization by Therapeutic Antibodies. *Cell Host Microbe* **2022**, *30*, 1103−1111.e6.
(88) Iketani, S.; Liu, L.; Guo, Y.; Liu, L.; Chan, J. F.-W.; Huang, Y.; Wang, M.; Luo, Y.; Yu, J.; Chu, H.; et al. Antibody Evasion Properties of SARS-CoV-2 Omicron Sublineages. *Nature* **2022**, *604*, 553−556.
(89) Kurhade, C.; Zou, J.; Xia, H.; Liu, M.; Yang, Q.; Cutler, M.; Cooper, D.; Muik, A.; Sahin, U.; Jansen, K. U.; et al. Neutralization of Omicron Sublineages and Deltacron SARS-CoV-2 by Three Doses of BNT162b2 Vaccine or BA.1 Infection. *Emerg. Microb. Infect.* **2022**, *11*, 1828−1832.
(90) Peiffer-Smadja, N.; Bridier-Nahmias, A.; Ferré, V. M.; Charpentier, C.; Garé, M.; Rioux, C.; Allemand, A.; Lavallée, P.; Ghosn, J.; Kramer, L.; et al. Emergence of E484K Mutation Following Bamlanivimab Monotherapy among High-Risk Patients Infected with the Alpha Variant of SARS-CoV-2. *Viruses* **2021**, *13*, 1642.
(91) Lee, H. K.; Knabl, L.; Knabl, L.; Wieser, M.; Mur, A.; Zabernigg, A.; Schumacher, J.; Kapferer, S.; Kaiser, N.; Furth, P. A.; et al. Immune Transcriptome Analysis of COVID-19 Patients Infected with SARS-CoV-2 Variants Carrying the E484K Escape Mutation Identifies a Distinct Gene Module. *Sci. Rep.* **2022**, *12*, 2784.
(92) Yang, W.-T.; Huang, W.-H.; Liao, T.-L.; Hsiao, T.-H.; Chuang, H.-N.; Liu, P.-Y. SARS-CoV-2 E484K Mutation Narrative Review: Epidemiology, Immune Escape, Clinical Implications, and Future Considerations. *Infect. Drug Resist.* **2022**, *15*, 373−385.
(93) Kannan, S. R.; Spratt, A. N.; Cohen, A. R.; Naqvi, S. H.; Chand, H. S.; Quinn, T. P.; Lorson, C. L.; Byrareddy, S. N.; Singh, K. Evolutionary Analysis of the Delta and Delta Plus Variants of the SARS-CoV-2 Viruses. *J. Autoimmun.* **2021**, *124*, 102715.
(94) Chavda, V. P.; Apostolopoulos, V. Global impact of delta plus variant and vaccination. *Expert Rev. Vaccines* **2022**, *21*, 597−600.
(95) Rahman, F. I.; Ether, S. A.; Islam, M. R. The "Delta Plus" COVID-19 Variant Has Evolved to Become the next Potential Variant of Concern: Mutation History and Measures of Prevention. *J. Basic Clin. Physiol. Pharmacol.* **2022**, *33*, 109−112.
(96) Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; et al. Structure of the SARS-CoV-2 Spike Receptor-Binding Domain Bound to the ACE2 Receptor. *Nature* **2020**, *581*, 215−220.
(97) Han, P.; Su, C.; Zhang, Y.; Bai, C.; Zheng, A.; Qiao, C.; Wang, Q.; Niu, S.; Chen, Q.; Zhang, Y.; et al. Molecular Insights into Receptor Binding of Recent Emerging SARS-CoV-2 Variants. *Nat. Commun.* **2021**, *12*, 6103.
(98) Wang, Q.; Zhang, Y.; Wu, L.; Niu, S.; Song, C.; Zhang, Z.; Lu, G.; Qiao, C.; Hu, Y.; Yuen, K.-Y.; et al. Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell* **2020**, *181*, 894−904.e9.
(99) Han, P.; Li, L.; Liu, S.; Wang, Q.; Zhang, D.; Xu, Z.; Han, P.; Li, X.; Peng, Q.; Su, C.; et al. Receptor Binding and Complex Structures of Human ACE2 to Spike RBD from Omicron and Delta SARS-CoV-2. *Cell* **2022**, *185*, 630−640.e10.
(100) Dejnirattisai, W.; Zhou, D.; Supasa, P.; Liu, C.; Mentzer, A. J.; Ginn, H. M.; Zhao, Y.; Duyvesteyn, H. M. E.; Tuekprakhon, A.; Nutalai, R.; et al. Antibody Evasion by the P.1 Strain of SARS-CoV-2. *Cell* **2021**, *184*, 2939−2954.e9.
(101) Nutalai, R.; Zhou, D.; Tuekprakhon, A.; Ginn, H. M.; Supasa, P.; Liu, C.; Huo, J.; Mentzer, A. J.; Duyvesteyn, H. M. E.; Dijokaite-Guraliuc, A.; et al. Potent Cross-Reactive Antibodies Following Omicron Breakthrough in Vaccinees. *Cell* **2022**, *185*, 2116−2131.e18.
(102) Li, L.; Liao, H.; Meng, Y.; Li, W.; Han, P.; Liu, K.; Wang, Q.; Li, D.; Zhang, Y.; Wang, L.; et al. Structural Basis of Human ACE2 Higher Binding Affinity to Currently Circulating Omicron SARS-CoV-2 Sub-Variants BA.2 and BA.1.1. *Cell* **2022**, *185*, 2952−2960.e10.
(103) Dejnirattisai, W.; Zhou, D.; Ginn, H. M.; Duyvesteyn, H. M. E.; Supasa, P.; Case, J. B.; Zhao, Y.; Walter, T. S.; Mentzer, A. J.; Liu, C.; et al. The Antigenic Anatomy of SARS-CoV-2 Receptor Binding Domain. *Cell* **2021**, *184*, 2183.
(104) Du, S.; Liu, P.; Zhang, Z.; Xiao, T.; Yasimayi, A.; Huang, W.; Wang, Y.; Cao, Y.; Xie, X. S.; Xiao, J. Structures of SARS-CoV-2 B.1.351 Neutralizing Antibodies Provide Insights into Cocktail Design against Concerning Variants. *Cell Res.* **2021**, *31*, 1130−1133.
(105) Gobeil, S. M.-C.; Janowska, K.; McDowell, S.; Mansouri, K.; Parks, R.; Stalls, V.; Kopp, M. F.; Manne, K.; Li, D.; Wiehe, K.; et al. Effect of Natural Mutations of SARS-CoV-2 on Spike Structure, Conformation, and Antigenicity. *Science* **2021**, *373*, No. eabi6226.
(106) Liu, L.; Wang, P.; Nair, M. S.; Yu, J.; Rapp, M.; Wang, Q.; Luo, Y.; Chan, J. F.-W.; Sahi, V.; Figueroa, A.; Guo, X. V.; et al. Potent Neutralizing Antibodies against Multiple Epitopes on SARS-CoV-2 Spike. *Nature* **2020**, *584*, 450−456.
(107) McCallum, M.; Bassi, J.; De Marco, A.; Chen, A.; Walls, A. C.; Di Iulio, J.; Tortorici, M. A.; Navarro, M.-J.; Silacci-Fregni, C.; Saliba, C.; et al. SARS-CoV-2 Immune Evasion by the B.1.427/B.1.429 Variant of Concern. *Science* **2021**, *373*, 648−654.
(108) Pymm, P.; Adair, A.; Chan, L.-J.; Cooney, J. P.; Mordant, F. L.; Allison, C. C.; Lopez, E.; Haycroft, E. R.; O'Neill, M. T.; Tan, L. L.; et al. Nanobody Cocktails Potently Neutralize SARS-CoV-2 D614G N501Y Variant and Protect Mice. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, No. e2101918118.
(109) Supasa, P.; Zhou, D.; Dejnirattisai, W.; Liu, C.; Mentzer, A. J.; Ginn, H. M.; Zhao, Y.; Duyvesteyn, H. M. E.; Nutalai, R.; Tuekprakhon, A.; et al. Reduced Neutralization of SARS-CoV-2 B.1.1.7 Variant by Convalescent and Vaccine Sera. *Cell* **2021**, *184*, 2201−2211.e7.
(110) Zhang, J.; Xiao, T.; Cai, Y.; Lavine, C. L.; Peng, H.; Zhu, H.; Anand, K.; Tong, P.; Gautam, A.; Mayer, M. L.; et al. Membrane Fusion and Immune Evasion by the Spike Protein of SARS-CoV-2 Delta Variant. *Science* **2021**, *374*, 1353−1360.
(111) Zhu, X.; Mannar, D.; Srivastava, S. S.; Berezuk, A. M.; Demers, J.-P.; Saville, J. W.; Leopold, K.; Li, W.; Dimitrov, D. S.;

https://doi.org/10.1021/acs.jpcb.3c01467
J. Phys. Chem. B XXXX, XXX, XXX−XXX

Tuttle, K. S.; et al. Cryo-Electron Microscopy Structures of the N501Y SARS-CoV-2 Spike Protein in Complex with ACE2 and 2 Potent Neutralizing Antibodies. *PLoS Biol.* **2021**, *19*, No. e3001237.
(112) Brown, E. E. F.; Rezaei, R.; Jamieson, T. R.; Dave, J.; Martin, N. T.; Singaravelu, R.; Crupi, M. J. F.; Boulton, S.; Tucker, S.; Duong, J.; et al. Characterization of Critical Determinants of ACE2−SARS CoV-2 RBD Interaction. *Int. J. Mol. Sci.* **2021**, *22*, 2268.
(113) Cheng, M. H.; Porritt, R. A.; Rivas, M. N.; Krieger, J. M.; Ozdemir, A. B.; Garcia, G.; Arumugaswami, V.; Fries, B. C.; Arditi, M.; Bahar, I. A Monoclonal Antibody against Staphylococcal Enterotoxin B Superantigen Inhibits SARS-CoV-2 Entry in Vitro. *Structure* **2021**, *29*, 951−962.e3.
(114) Khan, A.; Zia, T.; Suleman, M.; Khan, T.; Ali, S. S.; Abbasi, A. A.; Mohammad, A.; Wei, D. Higher Infectivity of the SARS-CoV-2 New Variants Is Associated with K417N/T, E484K, and N501Y Mutants: An Insight from Structural Data. *J. Cell. Physiol.* **2021**, *236*, 7045−7057.
(115) Celik, I.; Khan, A.; Dwivany, F. M.; Fatimawali; Wei, D.-Q.; Tallei, T. E. Computational Prediction of the Effect of Mutations in the Receptor-Binding Domain on the Interaction between SARS-CoV-2 and Human ACE2. *Mol. Divers.* **2022**, *26*, 3309−3324.
(116) Rath, S. L.; Kumar, K. Investigation of the Effect of Temperature on the Structure of SARS-CoV-2 Spike Protein by Molecular Dynamics Simulations. *Front. Mol. Biosci.* **2020**, *7*, 583523.
(117) 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.
(118) Cao, Y.; Choi, Y. K.; Frank, M.; Woo, H.; Park, S.-J.; Yeom, M. S.; Seok, C.; Im, W. Dynamic Interactions of Fully Glycosylated SARS-CoV-2 Spike Protein with Various Antibodies. *J. Chem. Theory Comput.* **2021**, *17*, 6559−6569.
(119) Cheng, M. H.; Krieger, J. M.; Banerjee, A.; Xiang, Y.; Kaynak, B.; Shi, Y.; Arditi, M.; Bahar, I. Impact of New Variants on SARS-CoV-2 Infectivity and Neutralization: A Molecular Assessment of the Alterations in the Spike-Host Protein Interactions. *iScience* **2022**, *25*, 103939.
(120) Barnes, C. O.; Jette, C. A.; Abernathy, M. E.; Dam, K.-M. A.; Esswein, S. R.; Gristick, H. B.; Malyutin, A. G.; Sharaf, N. G.; Huey-Tubman, K. E.; Lee, Y. E.; et al. SARS-CoV-2 Neutralizing Antibody Structures Inform Therapeutic Strategies. *Nature* **2020**, *588*, 682−687.
(121) Cerutti, G.; Rapp, M.; Guo, Y.; Bahna, F.; Bimela, J.; Reddem, E. R.; Yu, J.; Wang, P.; Liu, L.; Huang, Y.; et al. Structural Basis for Accommodation of Emerging B.1.351 and B.1.1.7 Variants by Two Potent SARS-CoV-2 Neutralizing Antibodies. *Structure* **2021**, *29*, 655−663.e4.
(122) Greaney, A. J.; Starr, T. N.; Barnes, C. O.; Weisblum, Y.; Schmidt, F.; Caskey, M.; Gaebler, C.; Cho, A.; Agudelo, M.; Finkin, S.; et al. Mapping Mutations to the SARS-CoV-2 RBD That Escape Binding by Different Classes of Antibodies. *Nat. Commun.* **2021**, *12*, 4196.
(123) Rogers, T. F.; Zhao, F.; Huang, D.; Beutler, N.; Burns, A.; He, W.; Limbo, O.; Smith, C.; Song, G.; Woehl, J.; et al. Isolation of Potent SARS-CoV-2 Neutralizing Antibodies and Protection from Disease in a Small Animal Model. *Science* **2020**, *369*, 956−963.
(124) Wheatley, A. K.; Pymm, P.; Esterbauer, R.; Dietrich, M. H.; Lee, W. S.; Drew, D.; Kelly, H. G.; Chan, L.-J.; Mordant, F. L.; Black, K. A.; et al. Landscape of Human Antibody Recognition of the SARS-CoV-2 Receptor Binding Domain. *Cell Rep.* **2021**, *37*, 109822.
(125) Xiang, Y.; Nambulli, S.; Xiao, Z.; Liu, H.; Sang, Z.; Duprex, W. P.; Schneidman-Duhovny, D.; Zhang, C.; Shi, Y. Versatile and Multivalent Nanobodies Efficiently Neutralize SARS-CoV-2. *Science* **2020**, *370*, 1479−1484.
(126) Yuan, M.; Liu, H.; Wu, N. C.; Wilson, I. A. Recognition of the SARS-CoV-2 Receptor Binding Domain by Neutralizing Antibodies. *Biochem. Biophys. Res. Commun.* **2021**, *538*, 192−203.
(127) Dong, J.; Zost, S. J.; Greaney, A. J.; Starr, T. N.; Dingens, A. S.; Chen, E. C.; Chen, R. E.; Case, J. B.; Sutton, R. E.; Gilchuk, P.; et al. Genetic and Structural Basis for SARS-CoV-2 Variant Neutralization by a Two-Antibody Cocktail. *Nat. Microbiol.* **2021**, *6*, 1233−1244.
(128) Greaney, A. J.; Loes, A. N.; Crawford, K. H. D.; Starr, T. N.; Malone, K. D.; Chu, H. Y.; Bloom, J. D. Comprehensive Mapping of Mutations in the SARS-CoV-2 Receptor-Binding Domain That Affect Recognition by Polyclonal Human Plasma Antibodies. *Cell Host Microbe* **2021**, *29*, 463−476.e6.
(129) Starr, T. N.; Greaney, A. J.; Hilton, S. K.; Ellis, D.; Crawford, K. H. D.; Dingens, A. S.; Navarro, M. J.; Bowen, J. E.; Tortorici, M. A.; Walls, A. C.; et al. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* **2020**, *182*, 1295.
(130) Starr, T. N.; Greaney, A. J.; Addetia, A.; Hannon, W. W.; Choudhary, M. C.; Dingens, A. S.; Li, J. Z.; Bloom, J. D. Prospective Mapping of Viral Mutations That Escape Antibodies Used to Treat COVID-19. *Science* **2021**, *371*, 850−854.
(131) Starr, T. N.; Greaney, A. J.; Dingens, A. S.; Bloom, J. D. Complete Map of SARS-CoV-2 RBD Mutations That Escape the Monoclonal Antibody LY-CoV555 and Its Cocktail with LY-CoV016. *Cell Rep. Med.* **2021**, *2*, 100255.
(132) Liu, Z.; VanBlargan, L. A.; Bloyet, L.-M.; Rothlauf, P. W.; Chen, R. E.; Stumpf, S.; Zhao, H.; Errico, J. M.; Theel, E. S.; Liebeskind, M. J.; et al. Identification of SARS-CoV-2 Spike Mutations That Attenuate Monoclonal and Serum Antibody Neutralization. *Cell Host Microbe* **2021**, *29*, 477−488.e4.
(133) Muecksch, F.; Weisblum, Y.; Barnes, C. O.; Schmidt, F.; Schaefer-Babajew, D.; Wang, Z.; C Lorenzi, J. C.; Flyak, A. I.; DeLaitsch, A. T.; Huey-Tubman, K. E.; Huey-Tubman, K. E.; et al. Affinity Maturation of SARS-CoV-2 Neutralizing Antibodies Confers Potency, Breadth, and Resilience to Viral Escape Mutations. *Immunity* **2021**, *54*, 1853−1868.e7.
(134) Zahradník, J.; Marciano, S.; Shemesh, M.; Zoler, E.; Harari, D.; Chiaravalli, J.; Meyer, B.; Rudich, Y.; Li, C.; Marton, I.; et al. SARS-CoV-2 Variant Prediction and Antiviral Drug Design Are Enabled by RBD in Vitro Evolution. *Nat. Microbiol.* **2021**, *6*, 1188−1198.
(135) Sánchez-Sendra, B.; Albert, E.; Zulaica, J.; Torres, I.; Giménez, E.; Botija, P.; Beltrán, M. J.; Rodado, C.; Geller, R.; Navarro, D. Neutralizing Antibodies against SARS-CoV-2 Variants of Concern Elicited by the Comirnaty COVID-19 Vaccine in Nursing Home Residents. *Sci. Rep.* **2022**, *12*, 3788.
(136) Stamatatos, L.; Czartoski, J.; Wan, Y.-H.; Homad, L. J.; Rubin, V.; Glantz, H.; Neradilek, M.; Seydoux, E.; Jennewein, M. F.; MacCamy, A. J.; et al. MRNA Vaccination Boosts Cross-Variant Neutralizing Antibodies Elicited by SARS-CoV-2 Infection. *Science* **2021**, *372*, 1413−1418.
(137) 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.
(138) Mannar, D.; Saville, J. W.; Zhu, X.; Srivastava, S. S.; Berezuk, A. M.; Tuttle, K. S.; Marquez, A. C.; Sekirov, I.; Subramaniam, S. SARS-CoV-2 Omicron Variant: Antibody Evasion and Cryo-EM Structure of Spike Protein−ACE2 Complex. *Science* **2022**, *375*, 760−764.
(139) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pK_a Values. *J. Chem. Theory Comput.* **2011**, *7*, 2284−2295.
(140) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43−56.
(141) 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.
(142) Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **2001**, *105*, 9954−9960.

https://doi.org/10.1021/acs.jpcb.3c01467
J. Phys. Chem. B XXXX, XXX, XXX−XXX

(143) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(144) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33−38.
(145) Kumari, R.; Kumar, R.; Lynn, A. Open Source Drug Discovery Consortium; Lynn, A. G_mmpbsa —A GROMACS Tool for High-Throughput MM-PBSA Calculations. *J. Chem. Inf. Model.* **2014**, *54*, 1951−1962.
(146) Jurrus, E.; Engel, D.; Star, K.; Monson, K.; Brandi, J.; Felberg, L. E.; Brookes, D. H.; Wilson, L.; Chen, J.; Liles, K.; et al. Improvements to the APBS Biomolecular Solvation Software Suite. *Protein Sci.* **2018**, *27*, 112−128.
(147) Zoete, V.; Michielin, O. Comparison between Computational Alanine Scanning and Per-Residue Binding Free Energy Decomposition for Protein-Protein Association Using MM-GBSA: Application to the TCR-p-MHC Complex. *Proteins* **2007**, *67*, 1026−1047.
(148) Kimura, S. R.; Brower, R. C.; Vajda, S.; Camacho, C. J. Dynamical View of the Positions of Key Side Chains in Protein-Protein Recognition. *Biophys. J.* **2001**, *80*, 635−642.
(149) Han, P.; Su, C.; Zhang, Y.; Bai, C.; Zheng, A.; Qiao, C.; Wang, Q.; Niu, S.; Chen, Q.; Zhang, Y.; et al. Molecular Insights into Receptor Binding of Recent Emerging SARS-CoV-2 Variants. *Nat. Commun.* **2021**, *12*, 6103.
(150) Socher, E.; Conrad, M.; Heger, L.; Paulsen, F.; Sticht, H.; Zunke, F.; Arnold, P. Computational Decomposition Reveals Reshaping of the SARS-CoV-2−ACE2 Interface among Viral Variants Expressing the N501Y Mutation. *J. Cell. Biochem.* **2021**, *122*, 1863−1872.
(151) Jawad, B.; Adhikari, P.; Podgornik, R.; Ching, W.-Y. Key Interacting Residues between RBD of SARS-CoV-2 and ACE2 Receptor: Combination of Molecular Dynamics Simulation and Density Functional Calculation. *J. Chem. Inf. Model.* **2021**, *61*, 4425−4441.
(152) Rath, S. L.; Padhi, A. K.; Mandal, N. Scanning the RBD-ACE2 molecular interactions in Omicron variant. *Biochem. Biophys. Res. Commun.* **2022**, *592*, 18−23.
(153) Jafary, F.; Jafari, S.; Ganjalikhany, M. R. In silico investigation of critical binding pattern in SARS-CoV-2 spike protein with angiotensin-converting enzyme 2. *Sci. Rep.* **2021**, *11*, 6927.
(154) Mandal, N.; Padhi, A. K.; Rath, S. L. Molecular Insights into the Differential Dynamics of SARS-CoV-2 Variants of Concern. *J. Mol. Graph. Model.* **2022**, *114*, 108194.
(155) Ghoula, M.; Naceri, S.; Sitruk, S.; Flatters, D.; Moroy, G.; Camproux, A. C. Identifying Promising Druggable Binding Sites and Their Flexibility to Target the Receptor-Binding Domain of SARS-CoV-2 Spike Protein. *Comput. Struct. Biotechnol. J.* **2023**, *21*, 2339−2351.
(156) Gan, H. H.; Twaddle, A.; Marchand, B.; Gunsalus, K. C. Structural Modeling of the SARS-CoV-2 Spike/Human ACE2 Complex Interface Can Identify High-Affinity Variants Associated with Increased Transmissibility. *J. Mol. Biol.* **2021**, *433*, 167051.
(157) Forest-Nault, C.; Koyuturk, I.; Gaudreault, J.; Pelletier, A.; L'Abbé, D.; Cass, B.; Bisson, L.; Burlacu, A.; Delafosse, L.; Stuible, M.; et al. Impact of the Temperature on the Interactions between Common Variants of the SARS-CoV-2 Receptor Binding Domain and the Human ACE2. *Sci. Rep.* **2022**, *12*, 11520.
(158) Gu, H.; Chen, Q.; Yang, G.; He, L.; Fan, H.; Deng, Y.-Q.; Wang, Y.; Teng, Y.; Zhao, Z.; Cui, Y.; et al. Adaptation of SARS-CoV-2 in BALB/c Mice for Testing Vaccine Efficacy. *Science* **2020**, *369*, 1603−1607.
(159) Shi, P.-Y.; Xie, X.; Zou, J.; Fontes-Garfias, C.; Xia, H.; Swanson, K.; Cutler, M.; Cooper, D.; Menachery, V.; Weaver, S.; Dormitzer, P. Neutralization of N501Y Mutant SARS-CoV-2 by BNT162b2 Vaccine-Elicited Sera, **2021**. Preprint (In Review), DOI: 10.21203/rs.3.rs-143532/v1.
(160) Sang, P.; Chen, Y.-Q.; Liu, M.-T.; Wang, Y.-T.; Yue, T.; Li, Y.; Yin, Y.-R.; Yang, L.-Q. Electrostatic Interactions Are the Primary Determinant of the Binding Affinity of SARS-CoV-2 Spike RBD to ACE2: A Computational Case Study of Omicron Variants. *Int. J. Mol. Sci.* **2022**, *23*, 14796.
(161) Focosi, D.; Maggi, F. Neutralising Antibody Escape of SARS-CoV-2 Spike Protein: Risk Assessment for Antibody-based Covid-19 Therapeutics and Vaccines. *Rev. Med. Virol.* **2021**, *31*, No. e2231.
(162) Ghorbani, M.; Brooks, B. R.; Klauda, J. B. Critical Sequence Hotspots for Binding of Novel Coronavirus to Angiotensin Converter Enzyme as Evaluated by Molecular Simulations. *J. Phys. Chem. B* **2020**, *124*, 10034−10047.
(163) Cavani, M.; Riofrío, W. A.; Arciniega, M. Molecular Dynamics and MM-PBSA Analysis of the SARS-CoV-2 Gamma Variant in Complex with the HACE-2 Receptor. *Molecules* **2022**, *27*, 2370.
(164) Wang, W. B.; Liang, Y.; Jin, Y. Q.; Zhang, J.; Su, J. G.; Li, Q. M. E484K Mutation in SARS-CoV-2 RBD Enhances Binding Affinity with HACE2 but Reduces Interactions with Neutralizing Antibodies and Nanobodies: Binding Free Energy Calculation Studies. *J. Mol. Graph. Model.* **2021**, *109*, 108035.
(165) Andreano, E.; Piccini, G.; Licastro, D.; Casalino, L.; Johnson, N. V.; Paciello, I.; Dal Monego, S.; Pantano, E.; Manganaro, N.; Manenti, A.; et al. SARS-CoV-2 Escape from a Highly Neutralizing COVID-19 Convalescent Plasma. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, No. e2103154118.
(166) Sergeeva, A. P.; Katsamba, P. S.; Sampson, J. M.; Bahna, F.; Mannepalli, S.; Morano, N. C.; Shapiro, L.; Friesner, R. A.; Honig, B. Free Energy Perturbation Calculations of Mutation Effects on SARS-CoV-2 RBD::ACE2 Binding Affinity. **2022**, Preprint. DOI: 10.1101/2022.08.01.502301
(167) Chen, C.; Boorla, V. S.; Banerjee, D.; Chowdhury, R.; Cavener, V. S.; Nissly, R. H.; Gontu, A.; Boyle, N. R.; Vandegrift, K.; Nair, M. S.; et al. Computational Prediction of the Effect of Amino Acid Changes on the Binding Affinity between SARS-CoV-2 Spike RBD and Human ACE2. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, No. e2106480118.

Chapter 5

Targeting PIF partners for a DS prenatal treatment

The focus of this chapter is a thorough analysis of the PIF peptide, including the study of its functions, origin, and potential protein partners. Despite being a crucial peptide in reproductive biology, little is known about its exact mode of action. The purpose of this chapter is to close this knowledge gap by thoroughly examining a variety of PIF-related topics. Therefore, a variety of bioinformatics methods had to be used to clarify the peptide's characteristics and identify its interactions with other protein targets. The findings of this study will serve as the foundation for the next chapter, which will explore protein-peptide interactions in greater detail and make use of a variety of bioinformatics methods to develop a better understanding of these interactions. Essentially, the main goal of this work is to further our understanding of protein-peptide interactions by the exploration of the PIF potential protein partners and the use of the bioinformatics techniques that have been developed in prior chapters to examine these interactions.

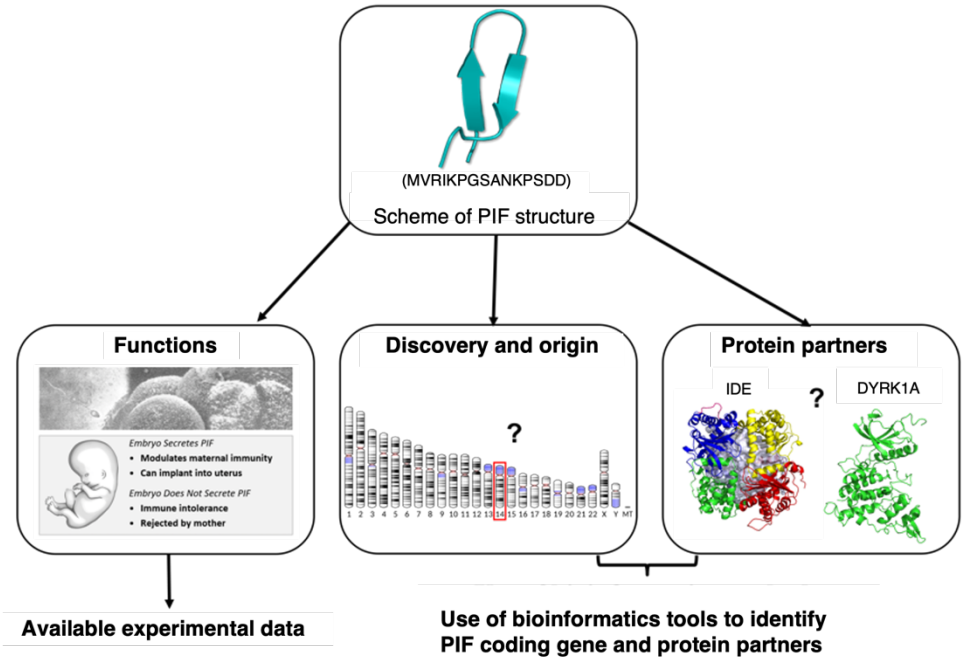


Figure 5. Representation of the protocol used to study PIF functions, origin, and protein partners.

5.1 PIF functions

PIF is a multifunctional peptide that increases embryo viability and functions as a neuroprotective agent [67]. Here, we will focus on the importance of PIF functions, particularly its prenatal roles throughout pregnancy, its neuroprotective benefits, and its implications in cancer.

5.1.1 Prenatal roles during pregnancy

The PIF peptide is one of many significant peptides that the human placenta produces [67, 68]. PIF is observed during the first trimester of pregnancy and is expressed by both the mammalian embryo and placenta [68]. PIF serves as a biomarker for identifying viable pregnancies; its absence indicates no pregnancy [68]. PIF increases endometrial embryo receptivity and the window for implantation, promotes embryo adhesion, controls apoptosis, and stimulates trophoblast invasion. Placentation and pregnancy outcomes are significantly influenced by trophoblast invasion, and PIF controls maternal immunological tolerance [69].

5.1.2 PIF neuroprotective effects

PIF function as a protective factor and its significant impact on postnatal therapy have been suggested by a variety of studies [67-71]. By reducing oxidative stress and protein misfolding, PIF aids in efficient embryonic-maternal communication and completes essential trophic and neuroprotective functions [72]. Additionally, PIF promotes brain regeneration via local and systemic effects and inhibits severe paralysis controlling oxidative stress and protein misfolding [72]. PIF also enhances the expression of genes related to axonogenesis, signal transduction, and neuronal differentiation [73]. The peptide can target actins and the neuronal protein tubulin, both of which are crucial for visceral and muscular development, suggesting that it may also be involved in pathways for postnatal nervous system repair [73].

5.1.3 PIF expression in cancer

PIF controls cell proliferation [69] and apoptosis [73], two processes that are crucial for the progression of cancer. PIF induces apoptosis and cell cycle arrest in many cancer cell types [74, 75]. Moreover, PIF therapy controls cell migration and invasion while preventing tumor growth and invasion. PIF also promotes tumor cell apoptosis through Caspase 3 activation and Bcl-2 downregulation [76].

5.1.4 PIF therapeutic uses

PIF has been substantially investigated as a possible therapeutic agent in both reproductive and

non-reproductive medical contexts due to its broad activity, which includes autoimmune, neuroprotective, and anti-apoptotic actions [67]. A benefit of PIF is its readily replicated biochemical structure. Another would be its ability to cross the blood-brain barrier in intact form, which is critical to achieve its activity [77]. PIF has been investigated as a therapeutic target for infertility in reproductive contexts. The sPIF (synthetic PIF) (MVRIKPGSANKPSDD) which is a synthetic analogue of PIF and commonly used abbreviation in scientific papers [67-77], protects against fetal loss - likely through modulation of inflammatory response in murine models [78]. Another study [79] showed that sPIF acts as a rescue factor in recurrent pregnancy loss patients and limits adverse agents during pregnancy. This was the first study to sponsor trials to use sPIF for treatment of non-pregnant immune disorders and schedule it in the US Food and Drug Administration (FDA) mandated toxicology studies. Additionally, PIF kept demonstrating its benefits in clinically relevant models in the treatment of various immunological and transplantation disorders due to its immune regulatory and regeneration actions [69, 77, 80-84]. These investigations set the path for a FAST-TRACK Phase I study in patients with autoimmune diseases to be completed successfully and for the FDA to grant orphan medicine designation status. Phase I research revealed that PIF is risk-free and free of toxicity or harmful side effects, paving the way for Phase II clinical studies (NCT02239562). Numerous other non-reproductive scenarios have also been used to study PIF. PIF possesses special immune-modulatory qualities that can be used to prevent autoimmune diseases like multiple sclerosis [85] and juvenile diabetes mellitus [80], as well as the emergence of graft-versus-host disease after semi-allogeneic transplant in pre-clinical models [86]. In individuals with neurodegeneration, PIF also helps to revert paralysis and stimulates the formation of new neurons [72].

5.1.5 PIF as a prenatal treatment for DS

Trisomy 21, often known as Down Syndrome (DS), is a genetic condition carried on by an extra copy of chromosome 21 [87]. Reduced cognitive function, developmental problems, congenital heart defects, distinct facial features, and other comorbidities are a few of the clinical symptoms of this condition [87]. Additionally, poor muscle tone is a characteristic of babies with DS that is almost always present [87]. Seizures and dementia-like symptoms as they age are among the many neurodevelopmental issues that DS patients frequently experience [87].

As for its evaluation, DS is diagnosed during pregnancy using a variety of methods, such as ultrasonography, amniocentesis, and chorionic villus sampling [87]. There are non-invasive prenatal diagnostic methods that can also identify DS using fetal cells or cell-free fetal DNA in maternal serum or blood [87].

Despite the lack of specific treatment for DS, persons with this condition can benefit from physical and developmental therapies [88]. Clinical research has revealed that dietary supplements and medications are either dangerous or ineffective despite their development [89]. A recent study revealed the possibility of GnRH injection therapy to improve DS patients' cognitive ability [90]. However, it is crucial to remember that while DS therapy has advanced

significantly, it still does not fully address all the additional health issues that persons with DS may experience. As a result, DS has a variety of symptoms that necessitate lifelong care, and ongoing research is being done to enhance therapies and treatments. Specific to this study, sPIF testing and evaluation as a prenatal treatment in "DS in mice" models also demonstrated great results and opened a new suggestion for the use of PIF as a therapeutic agent for chromosomally affected embryos with DS in early stages of pregnancy. Hence, we assessed the PIF peptide as a prospective and potential prenatal treatment for DS with the aim of reducing the symptoms after birth. Therefore, PIF has been the subject of numerous bioinformatics studies aimed at exploring its potential protein targets and mode of action.

5.2 PIF discovery, origin, and structure

PIF plays a crucial role in early fetal and maternal signaling as early as the two-cell stage in mice, the four-cell stage in humans, and the six-cell stage in bovines [69, 80, 91]. By using a lymphocyte platelet-binding test, PIF was discovered in 1994 by Barnea where its primary role was to be an early biomarker of pregnancy [92]. The lymphocyte platelet-binding assay was performed in pregnant women and non-pregnant women to examine and compare the immune responses and proteins. Consequently, the study led to the discovery of PIF [92]. In order to determine whether the proteins were unique to female reproductive organs, the experiment also compared immune responses with those of men [92]. The preliminary study's findings demonstrated that PIF was only expressed in pregnant women [92]. PIF was also discovered on the fourth day following embryo transfer among women who had successfully undergone *in vitro* fertilization, indicating that it played a part in determining the viability of the embryo. The AA sequence of the native peptide was identified using MALDI-TOF (Matrix Assisted Laser Desorption Ionisation/Time Of Flight), Edman degradation, and HPLC (High Performance Liquid Chromatography) confirmation; it was labeled as PIF and consisted of linear peptides with a length of 9 (PIF_9) (MVRIKPGSA) to 15 AA (PIF_{15}) (MVRIKPGSA-MVRIKPGSANKPSDD) [92]. Moreover, after the last decade, a synthetic version of the peptide (sPIF) was generated and has the same biological activity as the native PIF [67]. With the availability of the sPIF and the fact that it is easily generated due to its simple linear structure, it obviates the need of using the native peptide in either pre-clinical or clinical studies. Interestingly, when both peptides are compared to each other, the shorter peptide is also biologically active *in vitro* [69]. Both PIF_9 and PIF_{15} exerted a similar significant inhibitory effect in human peripheral blood mononuclear cells proliferation, an *in vitro* study where regulatory effects of PIF on global immune response were tested [69].

The genomic origin of PIF is still undetermined compared to the large information about its functions. PIF has undergone multiple purification and characterization processes, however efforts to discover its gene have so far been ineffective [71]. We do know that the peptide sequence is novel, and not homologous to any known protein. Some studies have suggested that the 2-12AA (VRIKPGSANKP) region shares similarity with a significant malarial CSP (Circumsporozite protein/plasmodium falciparum) that is linked to pathogenic invasion [92, 93]. However, we do believe that these results are aberrant since a placental peptide cannot be

obtained from a parasite, let alone perform the same functions. In this manuscript, we do provide directions and some perspectives concerning the origin of PIF. Our results also need experimental analysis to further validate our conclusions. Under the worst-case scenarios, we believe that it is possible that the gene containing the PIF sequence resides within a highly complex and structured repetitive chromosomal region that has not yet been sequenced and annotated in the human genome.

The structure of PIF is currently under investigation. Several studies utilizing *in silico* tools suggest that the peptide sequence confers PIF either a simple linear structure or an alpha-helix folding [94]. However, more in-depth studies have been carried out in our lab, combining different bioinformatics tools like MD simulations, where PIF stable structure takes the form of a beta sheet motif.

5.3 PIF protein targets

Identifying PIF targets is still a challenging task with the little information available on the peptide origin and structure. A few studies [95], majorly conducted by Barnea [72, 73, 96] , have tried to elucidate potential protein targets able to interact with PIF and induce an effect through different modes of interactions. Here, we present some of the potential protein partners of PIF, including those that were chosen for further studies as well as those that were excluded due to feasibility constraints.

5.3.1 Insulin-Degrading Enzyme (IDE)

Using the ProtoArray method (9000 proteins), Barnea and his group discovered that IDE and two transcript variants (1 and 3) of the K+ voltage-gated channel, shaker-subfamily, beta1 (KCNAB1) gene were the top two candidates for protein-peptide interactions with PIF [73]. Given its activities in degrading a variety of peptides as well as its affiliations to numerous neurodegenerative diseases including Alzheimer's, IDE stood out among these proteins as the most promising target to study. IDE, commonly referred to as Insulysin, is a zinc-based metalloprotease of the M16 family [97-100]. *In vitro* insulin degradation and high-affinity insulin binding are mostly carried out by IDE, which was the enzyme's main function when it was first discovered. IDE is crucial in the prevention of type 2 diabetes [101-104] as well as many other illnesses including Alzheimer's [105-111]. It is known that IDE plays a crucial role in the clearance of Alzheimer's amyloid-β (Aβ), and other peptides including insulin, glucagon and amylin [112-117]. Therefore, IDE is considered to be among the most significant enzymes in the human body due to these features. Additionally, as IDE is directly linked to these illnesses, it offers a promising therapeutic target for the development of effective regulators [118, 119]. The large cavity (15,000 A^3) of IDE, where peptides are degraded based on their size, charge distribution, and amyloidogenic potential, is a crucial component and feature of the protein [112-117]. However, IDE remains a challenging target as few information about its mechanism is known until this day. Nonetheless, being a strategic protein involved in

neurodegenerative diseases, where peptides trigger its function and a potential partner to PIF according to research [73, 96, 120], encouraged us to study its mechanism. Plus, IDE pleiotropic action makes it an interesting therapeutic target for both type 2 diabetes and Alzheimer's disease through genetic linkage. IDE inhibitors can act as a therapeutic approach for the treatment of diabetes while effective IDE activators can be used for Alzheimer's disease. In fact, a study by Maianti et al. [119] generated a selective anti-diabetic IDE inhibitor called 6bk that was tested in vivro with the goal of increasing the circulation of insulin and decreasing its degradation.

Since IDE is linked to both diseases, its inhibition may raise a potential issue of an adverse effect which would prevent IDE to cleave misfolded and amyloidogenic peptides, like Aβ. There is no clear answer on which specific strategy should be used on IDE. The key is to develop potent and selective IDE modulators that can be used for therapies of both diseases while evaluating their long-term effects on IDE to avoid these adverse effects. Therefore, our goal was to first understand IDE mechanism through MD simulations and confirm experimentally PIF through western blot and immunoprecipitation analysis to verify the direct association between PIF and IDE. Later, we further analyse their interaction with different docking tools.

5.3.2 Protein Disulfide-Isomerase (PDI)

In 2014, an article by Barnea et al. detailed the use of a microarray assay analysis with both cultured murine and equine embryos [72]. Complementary isolation methods in murine embryo extracts were followed using Liquid Chromatography tandem Mass Spectrometry (LC/MS/MS) and immunohistochemistry to identify protein-protein interactions involving PIF. These methods revealed that PDI and HSP are the PIF major protein interacting groups [72].
One of the most high-ranking group protein's interacting with PIF found by the Barnea group was PDI (PDI, PDIA4, PDIA6-like) [72]. PDI was the first protein folding catalyst discovered and is expressed in almost all mammalian tissues [121] and it has two main biological functions [95]. With its ability to reduce, oxidize, and isomerize disulfide bonds, PDI functions as a dithiol-disulfide oxidoreductase through its antioxidant thioredoxin domain (TRX) [121]. In addition to its redox activity, PDI can function as a chaperone both *in vivo* and *in vitro* conditions [121]. The TRX domain of PDI has been proven to be implicated in oocyte maturation, gamete fusion, monospermy, and proliferation [122], which delays the death of inner cell masses. However, little is known about PDI partners. For example, the 16F16 molecule has been shown to be neuroprotective in cellular and animal models of inflammatory diseases, such as multiple sclerosis, and to inhibit the production of pro-inflammatory cytokines and chemokines, such as IL-6 (Interleukin 6) and IL-8 (Interleukin 8) [123]. A Barnea study performed in 2017 used docking to show that PIF could interact close to the potential PDI binding site of the 16F16 inhibitor [123]. Other protein partners of PDI have been studied such as ubiquilin [124] and ERO1 (Endoplasmic Reticulum Oxidoreductase 1) [125]. However, no confirmed PDI-protein or PDI-peptide crystallographic complex is available to this day. PDI also has the particularity of binding CD4 (cluster of differentiation 4) [126]. In

multiple studies directed by Barnea, it has been proven that PIF targets and interacts with immune cells (CD4/CD8 (cluster of differentiation 8)) and attenuates the severity of several T cell driven autoimmune diseases in animal models [92, 127]. However, the methods through which PIF controls immunological responses and plays a part in maternal circulation are still poorly understood. On the other hand, we do know that during the maternal immune tolerance phase, several antigenic peptides are presented to class I MHC (Major Histocompatibility Complex) molecules through CD4 binding [128]. We speculate that the PIF sequence's similarity with the malarial CSP and supportive findings about its interaction with an oxidative stress inhibiting protein, such as PDI, could trigger an immune response as seen during pregnancy and cancer. However, since little information is provided about this protein target, we decided to focus on other proteins like IDE.

5.3.3 Heat Sock Protein (HSP)

The same study claimed that PIF binds protective HSP proteins (HSP70 and HSP90) [46]. HSP proteins are well known ubiquitously expressed molecular systems [129]. They are crucial components of the cell's machinery for folding proteins, serving as chaperones, and assisting in shielding cells from the damaging effects of physiological stresses [129]. Both HSP70 and HSP90 act as protective protein chaperones and are involved in protein folding, intracellular transport and oxidative stress [129]. They are also investigated as anti-cancer drugs due to their role in apoptosis and stabilization of mutant proteins [130, 131]. HSP proteins are essential modulators of neurotoxicity in Alzheimer's, Parkinson, and Huntington's diseases because they prevent the aggregation of unfolded and misfolded polypeptides, aid in refolding, and contribute to the solubilization of stable protein aggregates [129]. The HSP also shares a similar overall structure [132] and presents a peptide binding domain that recognizes a wide range of unrelated substrates [132]. However, we have little information about the peptide binding mechanism, where to this day, only one crystal structure representing an HPS70-peptide interaction has been resolved [133].

Given this information, HSP proteins could be promising targets to regulate increased conditions of oxidative stress in DS. However, the lack of structural studies and details about the binding mode of peptides, and the lack of peptide specificity from HSP, can be challenging.

5.3.4 Myosin Heavy Chain 10 (MYH10)

In order to understand the molecular mechanisms by which PIF promotes trophoblast invasion, another study was conducted by Yang et al. [95] to identify and create a profile of the proteins that interact with PIF. In this study, inactive PIF and biotin-labeled sPIF were immunoprecipitated with cytoplasmic proteins. MS and isobaric tags for relative and absolute quantification were used to evaluate the protein profiles. The interactions between PIF and MYH10 were evaluated using western blot and immunoprecipitation studies which confirmed their interaction. These results are also concordant with PIF known functions as the interaction between PIF and MYH10 significantly enhanced the invasion and migration capabilities of

HTR(Human Throphoblast)-8 trophoblast cells. However, little is known about the potential partners of MYH10 which makes it difficult to target a specific binding site. Additionally, MYH10 being composed of 1,976AA made the study not feasible, as it will require important computational resources to study PIF-MYH10 interaction through MD simulations.

5.3.5 Dual specificity tyrosine-(Y)-phosphorylation-Regulated Kinase 1A (DYRK1A)

Trisomy 21, also known as DS, is caused by an extra copy of chromosome 21 [86]. Its prevalence makes DS the most prevalent hereditary developmental condition, affecting 1 in 800 live births. All people with DS share intellectual disability and an early beginning of Alzheimer's disease, notwithstanding the complexity and variability of its clinical traits [86]. People with DS have a delayed cognitive development and altered nervous system development, which results in mental retardation and several daily life complications. The understanding of how the additional copy of chromosome 21 affects the DS phenotype has seen significant progress in the last decade. One of the most studied proteins in DS is DYRK1A, due to several of its important roles in neurodevelopment [134]. Due to its location on human chromosome 21's DS crucial area, the human DYRK1A gene was recognized as a potential gene for the condition [134]. Moreover, overexpression of DYRK1A has also been associated with neurodegenerative diseases [134]. Similarly, DYRK1A loss of function is also associated with neurodevelopmental defects and mental retardation [134]. Like most protein kinases, DYRK1A can adopt distinct active and inactive states [134, 135]. These states are regulated by reversible phosphorylation of conserved serine, threonine or tyrosine residues in the centrally located activation loop which is also positioned in the catalytic domain [134, 135]. The phosphorylation of the activation loop is essential to stabilize a conformation with a suitably positioned substrate binding site [135]. Several proteins regulate DYRK1A through phosphorylation events [136, 137]. Also, much progress has been made in the discovery of efficient inhibitors of DYRK1A, such as synthetic and natural compounds, again in the last decade [138]. Moreover, multiple crystal structures of DYRK1A complexed with small inhibitory compounds [139] are available on the PDB [55]. Although, it is important to mention that identifying an inhibitor that specifically or preferentially inhibits DYRK1A, a kinase, remains a challenge. So far, scientific research has focused on the discovery of small compounds to inhibit DYRK1A [139]. It was not until recently, in a study by Soundararajan M., et al. [135], that a natural peptide serving as a DYRK1A substrate and having a significant level of sequence homology with PIF was discovered [135]. Given this information, our study, combined with experimental results, focused on the discovery of the binding mode of PIF with DYRK1A which may also be a substrate of this kinase.

Chapter 6

Protein-peptide interaction: A bioinformatics study of IDE and DYRK1A as potential protein partners of PIF

The focus of this chapter is on the investigation of protein-peptide interactions and the challenges that occur while conducting such studies. We focus on PIF-protein interactions, and more specifically, with its potential protein partners, IDE and DYRK1A. Here, we provide an overview of the current techniques for modeling protein-peptide interactions. We breakdown this challenge into several smaller steps, such as 1) modeling the receptor protein, 2) predicting the peptide binding site, and 3) sampling peptide conformations. We outline each stage's available tools, methods, and reported results. To further validate the potential protein partners of the PIF peptide, we conducted experimental analyses in collaboration with Pr. Janel's team. To evaluate the interaction between PIF and IDE or DYRK1A, co-immunoprecipitation experiments were performed. Parallelly, due to the biological significance of IDE ad complex structure, we also decided to explore its molecular functions through MD simulations. Interestingly, our results suggested that these proteins may not be the correct protein partners of PIF. While studying IDE, the results obtained through docking and co-immunoprecipitation demonstrated that PIF does not bind to IDE. As for DYRK1A, we suggested hypotheses, including the possibility that PIF may be binding and being phosphorylated to DYRK1A, but in another adjacent or further far region on the protein. In fact, DYRK1A structure contains an intrinsically disordered region known as PEST [140]. Due to the absence of a well-defined three-dimensional structure and the dynamic nature of the PEST region make the task of accurate peptide docking and binding site predictions unfeasible. Even though our results did not support the initial hypothesis of PIF binding to IDE and DYRK1A, the findings of our study show the critical importance of thoroughly examining and confirming protein-peptide interactions.

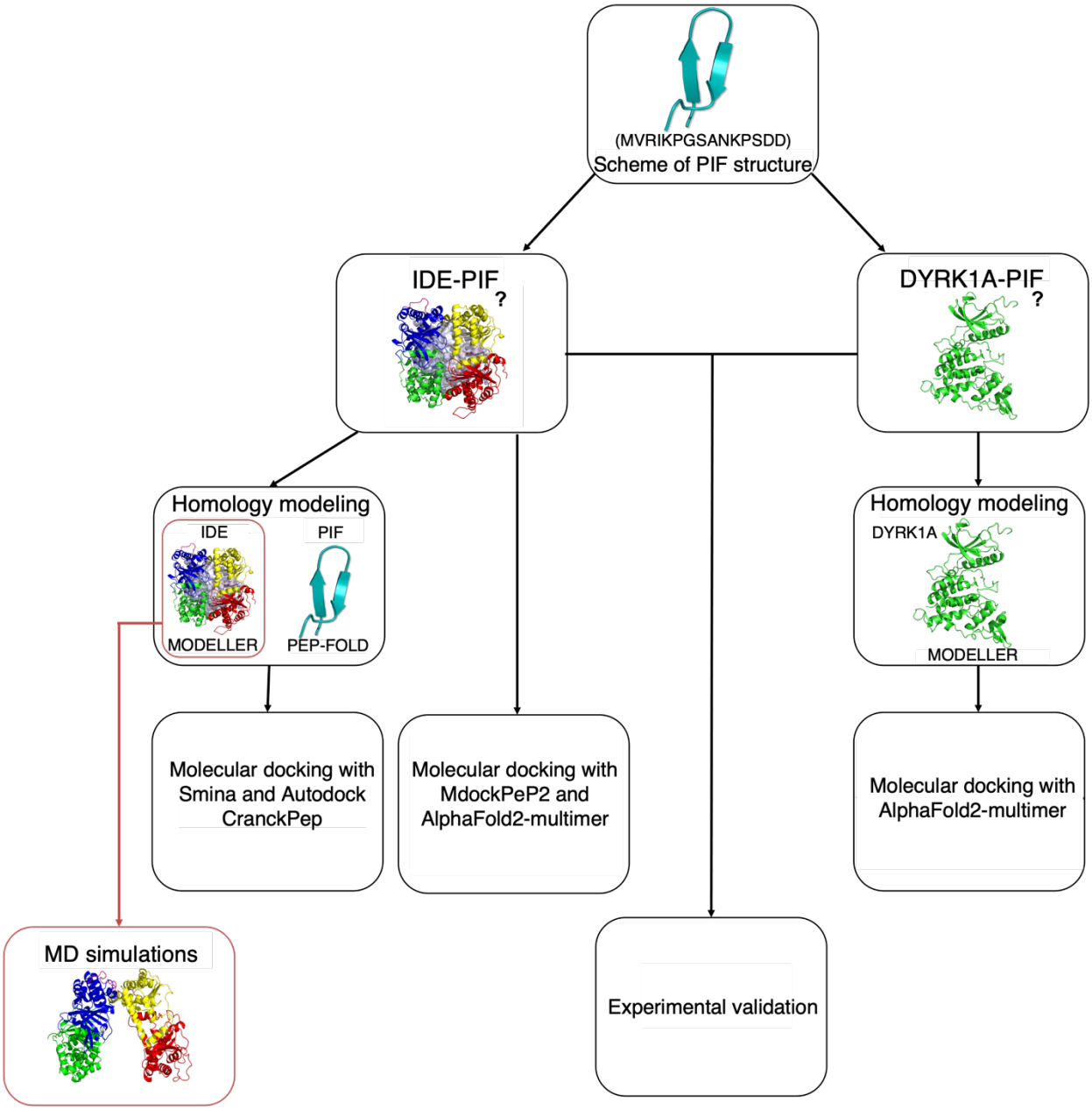


Figure 6. Representation of a simplified protocol to study IDE-PIF and DYRK1A-PIF interactions.

6.1 IDE: a potential PIF target

In a protein array analysis carried out by Barnea et al. [73] in 2010, one of the top candidates for interacting protein-protein interactions was found to be IDE. It was the first publication to mention an interaction between IDE and PIF. In order to find additional PIF targets in extracts from 10-day-old murine embryos, the Barnea group used PIF-based affinity chromatography and an LC/MS/MS study in 2014. Due to the failure of both analytical methods to detect IDE, they turned to immunochemistry, a more accurate technique, to examine the Biotin-PIF positive fractions. According to the findings, out of a total of 96 protein fractions, a PIF interaction was only discovered in 3 of them. To identify an IDE activity, the analysis had to be conducted several times on specific fractions. Because of these findings, it is possible to hypothesize that IDE might not be PIF major protein target, and this hypothesis is supported by the absence of a substantial interaction between IDE and PIF. Another paper in 2018 examined the potential connection between IDE and PIF [96]. *In silico* methods were used in the paper to demonstrate the PIF-IDE interaction mechanism.

IDE being our major track as PIF most promising target, we decided to test their interactions *in silico* and *in vitro*, respectively.

6.2 *In silico* and *in vitro* analyses of IDE-PIF interaction

6.2.1 Exploring IDE-PIF interaction through molecular docking

Before establishing our protocol, we investigated the IDE binding site and the location of its bioactive peptides in its crypt. The homo-dimeric protein IDE was already addressed in section 3.3.1. Each IDE monomer is composed of four structurally similar domains: domain 1 (D1: 43-285), domain 2 (D2: 286-515), domain 3 (D3: 542-768), and domain 4. (D4: 769-1019). The N-terminal domains form D1 and D2 (IDE-N), while the C-terminal domains are constituted of D3 and D4 (IDE-C). These domains have less than 25% sequence similarity and share similar secondary structures. Additionally, IDE has a 26-residue linker (residues 516-541) that joins the IDE-N and IDE-C portions. The IDE-N and IDE-C domains interact strongly and bury a large surface (15,000 A^3) with good shape complementarity. All four domains contribute to the surface of the inner chamber as IDE-N and IDE-C make extensive contact to produce an enclosed catalytic chamber that encloses its substrates [112-117]. Although IDE-N D1 contains IDE catalytic site, IDE-C is also required for the catalytic activity of IDE. Regarding substrate binding IDE can adopt one of two "open" or "closed" conformations. Therefore, IDE needs to transition between these two states in order to trigger the catalytic mechanism. Substrates are able to freely enter and exit the catalytic chamber in the open state. In the close state, the catalytic chamber is sealed, and the previously bound substrates are repositioned to facilitate hydrolysis. In fact, the D2 region possess an exosite (residues 336-342 and 359-363) that

participates in the binding of substrates and which role is to accommodate the N-terminus of the bound peptides. This repositioning of substrates is considered as the initial stage of peptide binding. Once the peptide bound to the exosite, it undergoes conformational changes. However, the peptides interacting with IDE exosite were partially resolved with X-ray crystallization [112-117]. With this understanding, we knew the exosite had to be the PIF N-terminus binding region and that we would need to see PIF establishing some sort of interaction with that particular region.

Therefore, we considered two alternatives for analyzing the IDE-PIF interaction: flexible restricted/binding site-based docking on the IDE protein or accelerated MD simulations to capture the naturally occurring binding of PIF in the IDE cavity. We chose to use the quicker and less time-consuming first alternative because IDE mostly occurs as a homodimer and each monomer is 1,019AA long. However, we also need to mention that peptide-docking has its own limitations. As to this day, peptide-docking remains a challenging task even for short peptides which are highly flexible biomolecules. In fact, addressing the flexibility of peptides is one of the main challenges when predicting the structures of protein-peptide complexes. Due to their high flexibility, peptides are commonly the focus of current approaches for predicting complex structures. These methods typically concentrate on short- or medium-sized peptides. This restriction is mostly brought on by the high degree of flexibility connected to peptides. Another major obstacle is developing an effective scoring function for ranking potential binding modes. An ideal scoring function should efficiently distinguish native or near-native binding configurations from numerous decoys generated by sampling algorithms. Unfortunately, existing scoring functions are still far from achieving perfection. These challenges significantly hinder the application of current computational methods in peptide-based drug discovery and development.

Also, we employed a diverse range of computational methods to obtain complementary insights into the binding mechanisms and dynamics of PIF with IDE. For this protocol, we used several tools such as: Smina [38], Autodock CranckPep [33], MDockPeP2 [32], and AlphaFold2-multimer [39]. By utilizing different software tools with distinct scoring functions, we sought to compare the results obtained and increase the sampling of PIF conformations. This strategy also enhanced the chances of capturing the correct conformation as observed by IDE-bound structures. Additionally, in order to assess and enhance the accuracy of the docking results, a redocking step was additionally performed using the Aβ peptide as a test case, allowing for an evaluation of the software's performance in reproducing known binding poses and interactions.

Therefore, we started the first stage of our research by using the computational tool Smina while we awaited further experimental confirmation of the direct interaction between IDE and PIF from our collaborators and simultaneously with our exploration of IDE structure. Smina is fork of AutoDock Vina that is customized to better support scoring function development and high-performance energy minimization [38]. Smina has been successfully used for peptide docking. We chose Smina because 1) it is faster and performs better than Autodock Vina, 2) possesses multiple scoring functions, and 3) it is free to use and user-friendly. We made the decision to perform binding site-based or restricted docking since it focuses on a predefined

region of the protein surface that is known or suspected to be involved in the interaction. Since we presumed that the exosite could be the main binding site according to the literature, it allowed us for a more targeted exploration providing detailed insights into the binding mechanism and specific residues involved. It is also a good way to compare the predicted binding conformations and interactions of PIF with known bound-IDE peptides.

Autodock CrackPep is a specialized software specifically designed for peptide-protein docking [33]. It focuses on exploring potential binding sites and evaluating the interactions between peptides and proteins at an atomistic level. Autodock CranckPep also considers the flexibility of peptides by allowing the exploration of multiple conformations during the docking process. This software, however, is more suitable for short to medium-length peptides, as the exhaustive sampling of longer peptide sequences may become computationally challenging. It is important to note that the conformation of the peptide is required for both Smina and Autodock CranckPep.

In the subsequent steps, MDockPeP2 and AlphaFold2-multimer were selected since they solely rely on sequence information. These software tools employ distinct methodologies to predict protein-peptide interactions, providing valuable insights into their underlying binding mechanisms. MDockPeP2 utilizes an efficient sampling approach, systematically generating and evaluating binding poses between the protein and peptide. Moreover, it employs a hybrid scoring function, enabling a comprehensive assessment of binding affinity and complementarity. However, a limitation of MDockPeP2 is the restricted AA length, as it is not suitable for peptides exceeding 35AA. On the other hand, AlphaFold2-multimer is a great tool for structure prediction, leveraging deep learning techniques and extensive training on known protein structures to accurately forecast their conformation and, in some instances, their interactions with partners.

In this section, we will also discuss the limits of this protocol as well as the findings that were obtained. In parallel, we also investigated a direct interaction between IDE and PIF with the help of Pr. Janel using the co-immunoprecipitation technique.

6.2.2 PIF flexible docking

We started the analysis by gathering our inputs. The IDE structure was retrieved from the PDB database (PDB ID: 2JG4) [141]. Missing residues and mutations were fixed with the MODELLER software [28] to recover the wild-type conformation of IDE. In this case, we only used IDE monomer as we were only interested in IDE-PIF interaction. To simultaneously compare our results with the literature, we also decided to redock the Aβ peptide to insure the validity of our protocol. As Smina requires three-dimensional structures of both the protein and the ligand, we used the PEP-FOLD software [142] to predict the secondary structure of IDE different potential peptides. Only a fraction of the resolved Aβ conformation has been used for this investigation since the Aβ conformation is not entirely resolved within the IDE-Aβ

complex.

Both the ligands and IDE structures were pre-treated according to Smina requirements and multiple docking scoring functions (Vina [143], Vinardo [144] and dkoes [38]) were used for comparison. Using multiple scoring functions in protein-peptide docking can provide a more comprehensive and accurate assessment of the predicted binding affinity and binding mode of a peptide ligand with its protein receptor. Moreover, this approach can help to identify the most likely binding mode and binding affinity, as well as provide insights into the strengths and weaknesses of different scoring functions in modeling protein-peptide interactions. Additionally, MGLTools [145] scripts were used to generate the correct format of IDE structure, as well as the fully flexible peptides files with added atom types and partial charges for Smina docking. We manually defined the box around the exosite region. The exosite region was enclosed and the box was large enough to hold the peptides. However, the peptides failed to bind properly (Figure 7).

We obtained the same results with the additional tools we used for PIF-IDE docking (Table 1). AutoDock CrankpPep was unable to determine the proper orientation of Aβ, which left ambiguity about the PIF results. MDockPEP2 and AlphaFold2-multimer also failed to capture the right position and the binding conformation of PIF (Figure 1). These results highlight the potential limitations of docking, where the accuracy of the results rely on the IDE structure employed. Hence, in this case, the flexibility of IDE must be thoroughly examined to have a better understanding of its function as well as the peptide mode of action.

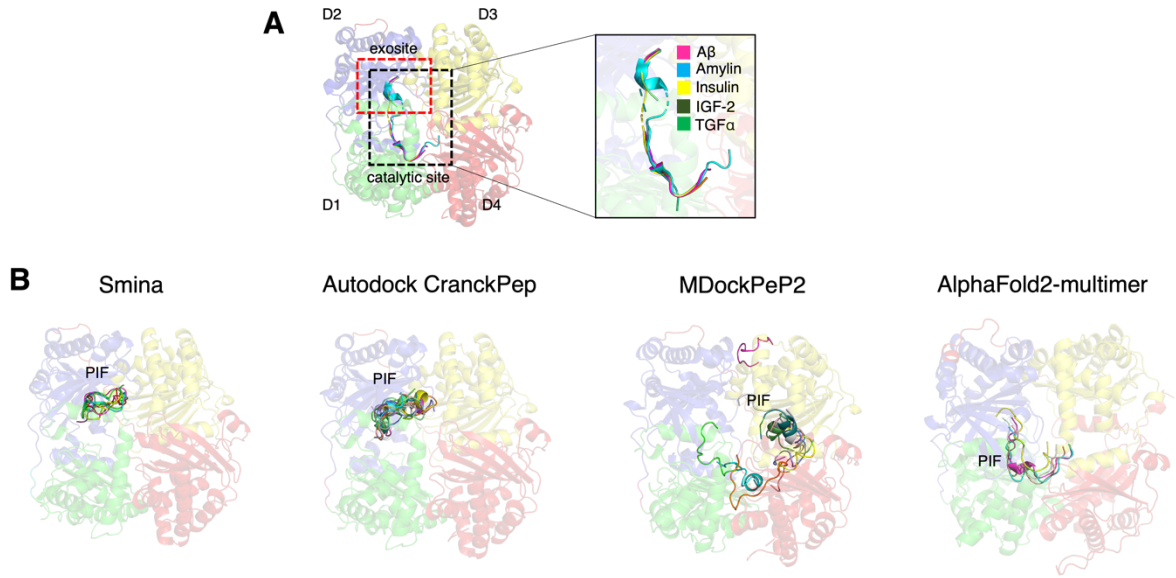


Figure 7. (A) Representation of crystallized IDE bound to its substrates (Aβ (PDB:2G47), Amylin (PDB:2G48), Insulin (PDB:2G54), IGF-2 (PDB:3E4Z), TGFα (PDB:3E50). The exosite and the catalytic site are highlighted. (B) Smina, Autodock CranckPep, MDockPeP2 and AlphaFold2-multimer docking results.

We came to the conclusion that this protocol has different challenges because: 1) the three-

dimensional structure of PIF has been modeled, however we require the resolved structure to guarantee the validity of our results; 2) Flexible docking can be useful in obtaining interesting results about the various conformations a peptide can adopt within its receptor, but it also generates a high number of degrees of freedom, which can be extremely challenging and a time-consuming process; 3) Although the N-terminus must be the initial part of the peptide to engage with the exosite, the binding orientation of PIF into the IDE exosite can still be challenging because there is no experimental evidence to support our findings; 4) Most of the peptides are partially resolved in the crystallographic structures and several portions are missing, which can also lead to uncertainty; 5) The placement of a peptide into the proper location with the correct conformation takes place after a number of events in a highly dynamic and cryptic protein such as IDE. Since molecular mechanisms like peptide binding significantly depend on the dynamic and functional behavior of the receptor, this final element may be one of the most crucial ones to take into account when studying a protein such as IDE.

Docking tool	Scoring function	Result	Limitations
Smina	Vina, Vinardo, dkoes	failed	1,3
MDockPeP2	PepProScore (hybrid)		1
AutoDock CrankPep	Hybrid		2,3
AlphaFold2-multimer	deep learning network		1

Table 1. Summary of the docking tools used to dock PIF with IDE and the limitations encountered for each one of them: 1) need prior information about the peptide binding region, 2) need prior information about the peptide binding region and it is time consuming, 3) accurate binding region but IDE dynamic behavior plays a crucial role leading the impossibility of predicting PIF correct conformation.

Thus, docking may not be the optimal method for predicting the proper conformation of PIF in the IDE cryptic cavity. Considering that IDE is a flexible protein, docking may be one of the initial steps used to produce a set of solutions that need to be refined. MD simulations can then be used to construct a relaxed system with a peptide that can explore various conformations until it finds one that is stable, after which the findings can be analyzed. However, due to several challenges with IDE peptide partners, the time-consuming nature of the task, and the experimental uncertainty of IDE as a PIF protein partner which was confirmed while leading this study, we chose to focus instead on IDE dynamic behavior since the results from this particular research provided critical insights into IDE biological function and potential therapeutic applications.

6.2.3 Experimental analysis of IDE-PIF interaction

Pr. Janel's assistance allowed for the experimental investigation of IDE and PIF interaction. In order to demonstrate or invalidate a direct relationship between IDE and PIF, we chose to combine an experimental analysis. We would then be able to apply alternative techniques or

improve our approach in this respect. Pr. Janel's lab undertaken a co-immunoprecipitation technique which is a powerful technique often used to identify protein-ligands direct interactions. Co-immunoprecipitation concept is very simple. It works by selecting an antibody that targets a known protein (IDE in this case) that is believed to be a member of a larger complex of proteins. By targeting this known member of a complex with an antibody, it is possible to pull the entire protein complex out of solution and thereby identify unknown members of the complex. For this study, an IDE antibody was selected with the presence of IDE and PIF in solution. The study was conducted several times for reproductivity reasons. However, no IDE-PIF complex has been identified which suggests that IDE might not be a potential target of PIF. This result confirms in a way the difficulties encountered in targeting IDE in the 2014 paper of Barnea's group [72] and in our protocol. Thus, the fractions containing the IDE-PIF complexes described in Barnea's paper [72] might be the result of a high concentration of PIF or other various reasons.

6.2.4 IDE: a therapeutic protein in Alzheimer's and type 2 diabetes diseases

Even though our results did not meet the expectations of our original study, IDE has been the subject of extensive research due to its critical role in Alzheimer's disease and Type 2 diabetes. As IDE represents a fundamental research topic in the scientific field, we decided to continue investigating its dynamic behavior and use MD simulations to examine its allosteric activities and cryptic structure. The first paper to be published during this Ph.D. program was focused on this work and was cited twice.

International Journal of
Molecular Sciences

MDPI

Article

Exploring the Structural Rearrangements of the Human Insulin-Degrading Enzyme through Molecular Dynamics Simulations

Mariem Ghoula [1], Nathalie Janel [2], Anne-Claude Camproux [1,*,†] and Gautier Moroy [1,*,†]

[1] Unité de Biologie Fonctionnelle et Adaptative, CNRS, INSERM, Université de Paris, F-75013 Paris, France; mariem.ghoula@inserm.fr
[2] Unité de Biologie Fonctionnelle et Adaptative, CNRS, Université de Paris, F-75013 Paris, France; nathalie.janel@u-paris.fr
* Correspondence: anne-claude.camproux@u-paris.fr (A.-C.C.); gautier.moroy@u-paris.fr (G.M.); Tel.: +33-1-57-27-83-77 (A.-C.C.); +33-1-57-27-83-85 (G.M.)
† These authors contributed equally to this work.

Abstract: Insulin-degrading enzyme (IDE) is a ubiquitously expressed metallopeptidase that degrades insulin and a large panel of amyloidogenic peptides. IDE is thought to be a potential therapeutic target for type-2 diabetes and neurodegenerative diseases, such as Alzheimer's disease. IDE catalytic chamber, known as a crypt, is formed, so that peptides can be enclosed and degraded. However, the molecular mechanism of the IDE function and peptide recognition, as well as its conformation changes, remains elusive. Our study elucidates IDE structural changes and explains how IDE conformational dynamics is important to modulate the catalytic cycle of IDE. In this aim, a free-substrate IDE crystallographic structure (PDB ID: 2JG4) was used to model a complete structure of IDE. IDE stability and flexibility were studied through molecular dynamics (MD) simulations to witness IDE conformational dynamics switching from a closed to an open state. The description of IDE structural changes was achieved by analysis of the cavity and its expansion over time. Moreover, the quasi-harmonic analysis of the hinge connecting IDE domains and the angles formed over the simulations gave more insights into IDE shifts. Overall, our results could guide toward the use of different approaches to study IDE with different substrates and inhibitors, while taking into account the conformational states resolved in our study.

Keywords: molecular dynamics simulation; insulin-degrading enzyme; therapeutic target

Citation: Ghoula, M.; Janel, N.; Camproux, A.-C.; Moroy, G. Exploring the Structural Rearrangements of the Human Insulin-Degrading Enzyme through Molecular Dynamics Simulations. *Int. J. Mol. Sci.* **2022**, *23*, 1746. https://doi.org/10.3390/ijms23031746

Academic Editor: Paulino Gómez-Puertas

Received: 30 December 2021
Accepted: 29 January 2022
Published: 3 February 2022

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

1. Introduction

Insulin-degrading enzyme (IDE), also known as Insulysin, is a zinc protease of the M16 metalloprotease family [1–5]. IDE was originally discovered and named since it is the major enzyme responsible for insulin degradation, in vitro, and insulin binding with high affinity (~10 nM) [6]. IDE plays a major role in preventing type II diabetes [6–9] and other diseases, such as Alzheimer's [10–16]. These characteristics make it one of the most important enzymes in the human body. Moreover, IDE directly links with these diseases, making it a promising therapeutic target to design efficient regulators [17,18]. IDE rapidly breaks down insulin and other peptides to prevent toxic amyloid formation [2,4,6,10]. An important feature of IDE is its large cavity (~15,000 A^3), where peptides are degraded based on their size, charge distribution, and amyloidogenic nature [19–24]. IDE is a catalytic protein known to switch between a closed and an open state. The transition from the closed to the open conformation is triggered by the achievement of a specific interaction between the IDE catalytic chamber and its substrates. Thus, substrates cannot enter the cavity when IDE is closed, which is why IDE needs to undergo an opening state to capture its substrates inside the catalytic chamber. IDE open form is also required for the exit of proteolytic

products. On the other hand, substrates can also lock IDE in the closed conformation to efficiently activate the proteolytic process [24].

Consistent with this, IDE is known to exist as an equilibrium of monomers, dimers, and oligomers [25]. Although IDE majorly exists as a homodimer and is more active than the monomeric form, the latter form does retain its enzymatic activity (Figure 1) [25]. Each monomer has 1019 amino acids and consists of four domains. An extended 26-residue loop connects the N-terminal domains (IDE-N: D1 and D2) and the C-terminal domains (IDE-C: D3 and D4). The N-terminal domains (IDE-N) D1 (residues 43–285) and D2 (residues 286–515) contain several charged, polar, and hydrophobic patches [24]. The surface of IDE-N is also largely neutral or negatively charged [24]. The catalytic site (residues H108, E111, H112, and E189) is located in IDE-N [24]. The exosite (residues 336–342 and 359–363), which is also present in IDE-N, is located approximately 30 Å from the crypt. This exosite is a key site in positioning peptides before degradation takes place. It is also the major site where the N-terminus of substrates are anchored [3,17,24]. On the other hand, the C-terminal domains (IDE-C) D3 (residues 542–768) and D4 (residues 769–1019) have a positively charged surface [24]. Although the catalytic site is situated in IDE-N, IDE-C is crucial for substrate recognition. Studies have shown that both IDE-N and IDE-C are essential for IDE activity and mutations of catalytic residues can severely decrease its function. For instance, IDE E111Q mutation renders the protein nonfunctional [26]. Moreover, site-directed mutagenesis of IDE H108 (H108L and H108Q) inhibits IDE catalytic activity but retains its ability to bind insulin [27]. Similarly, mutation of R824 and Y831 to alanine significantly reduces the catalytic rate of IDE [24].

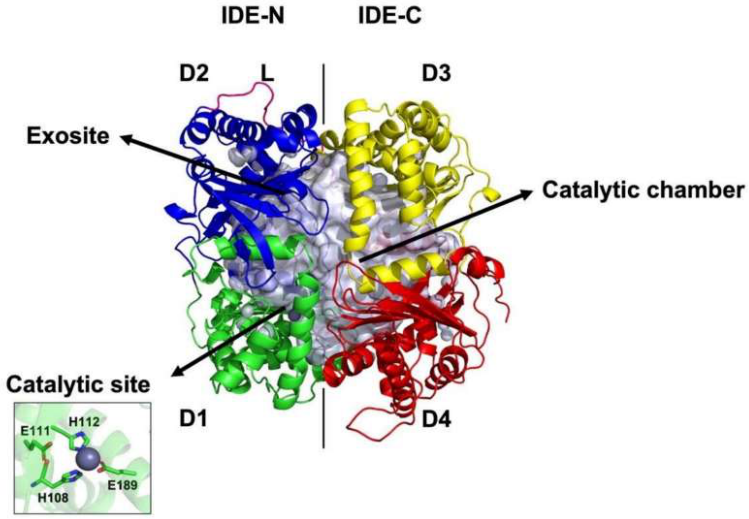


Figure 1. Representation of IDE structure and D1, D2, D3, and D4. IDE-N and IDE-C, as well as the linker (L), the exosite, and the catalytic site, are specified.

IDE is a very challenging protein. Hence, the mechanism through which peptides are recognized is still elusive. The different forms that IDE structure adopts (open-closed) during the catalytic cycle need further investigation as well. Another challenging aspect is the detailed mechanism of IDE allostery that also remains unsolved [24,28]. Solving the detailed and defined allosteric path along the discovery of specific and potent allosteric modulators, triggering the activity of IDE, awaits future work. Moreover, only the closed and holo-conformations of IDE have been solved [19–24]. As for the open state, it has been assessed with Fab-assisted CryoEM at best (Figure S1, Supplementary File 1) [28].

The current information about the conformational changes of IDE, as well as its dynamics information at the atomic level, is not sufficient yet. Moreover, although IDE makes the perfect therapeutic target for both type 2 diabetes and Alzheimer's disease, its dual linkage can also be a challenge for the development of potent modulators. For instance, inhibition of IDE might raise a potential issue of an adverse effect, which would prevent its action to cleave misfolded and amyloidogenic peptides, such as amyloid beta. IDE modulators will also require a long-term evaluation to avoid the adverse effects [18]. Therefore, it is essential to further explore the atomic-level molecular mechanism involved in the structural transitions of IDE for the development of efficient, but also selective inhibitors, and uncover the substrate recognition process that might hold the key to cure many diseases [29,30]. This information would also be crucial to finally decipher the complete role of IDE.

Here, we combine different bioinformatics approaches, such as structural modeling and molecular dynamics (MD) simulations, to address these questions and to complement the existing information concerning IDE and its structural mechanism.

2. Results and Discussion

2.1. MD Simulations Analysis

IDE biological function is directly related to its conformational transitions. With that in mind, we ran 7 MD simulations of the IDE monomer in its unbound state to recover the protein structural changes. Hence, the total simulation time for this system is 7.5 µs.

2.1.1. Root Mean Square Deviation (RMSD) Evaluation of IDE Structure

To analyze the stability of each system, we performed a RMSD analysis of all our MD trajectories (Figure 2). In our study, seven systems were simulated for a total of at least 1 µs. However, run 4 was simulated for a longer time (1.5 µs) since it displayed extremely high RMSD values until 1 µs. As shown in Figure 2A, Cα RMSD was found to stabilize the IDE system with values reaching 2.5 to 4 Å, with fewer fluctuations for most of the trajectories. On the other hand, the fourth and fifth trajectories displayed the largest fluctuations compared to the others.

The RMSD values of the fourth run increased towards 600 ns until they reached their highest values at 1 µs then decreased at 1.1 µs to form a plateau until the end of the trajectory. Thus, the behavior of IDE in run 4 may have been due to the exploration of another state. This state is different from the initial closed structure of IDE, which might explain the excessive fluctuations in the RMSD values. Consequently, the large crypt movements and the flexibility of IDE can be explained through the exploration of an open state.

The RMSD values of the fifth run also increased, starting from 300 ns to reach a plateau with a constant value of 5.0 Å. This result also indicates a conformational change in the structure of IDE. With these observations in mind, for each of the third, fourth, and fifth runs, we extracted a frame at 1 µs, where IDE fluctuates the most. This relates to the results gathered from the RMSD with the changes occurring in the structure of IDE. As shown in Figure 2A, a noticeable twisting motion on D1 and D4 differentiated IDE of the fifth run from the other extracted frames. The helices and loops at the entry of the catalytic chamber witnessed a slight twist while D2 and D3 remained rigid. These movements can be described as rigid-body swing motions of the IDE gate [30]. It was fully explained by McCord et al. [30] that this twisting motion is characterized by a small rigid movement of D1 moving away from D4. Additionally, these movements were further analyzed with a quasi-harmonic analysis to confirm these observations (data not shown). In our case, this distinct state cannot be considered as open enough to enclose short peptides and did not differ significantly from the closed state, as only very small movements occurred in IDE. However, it is an interesting observation, and it enhances the fact that IDE can exist in a mixture of different conformations, and can explore several transient states, while remaining stable.

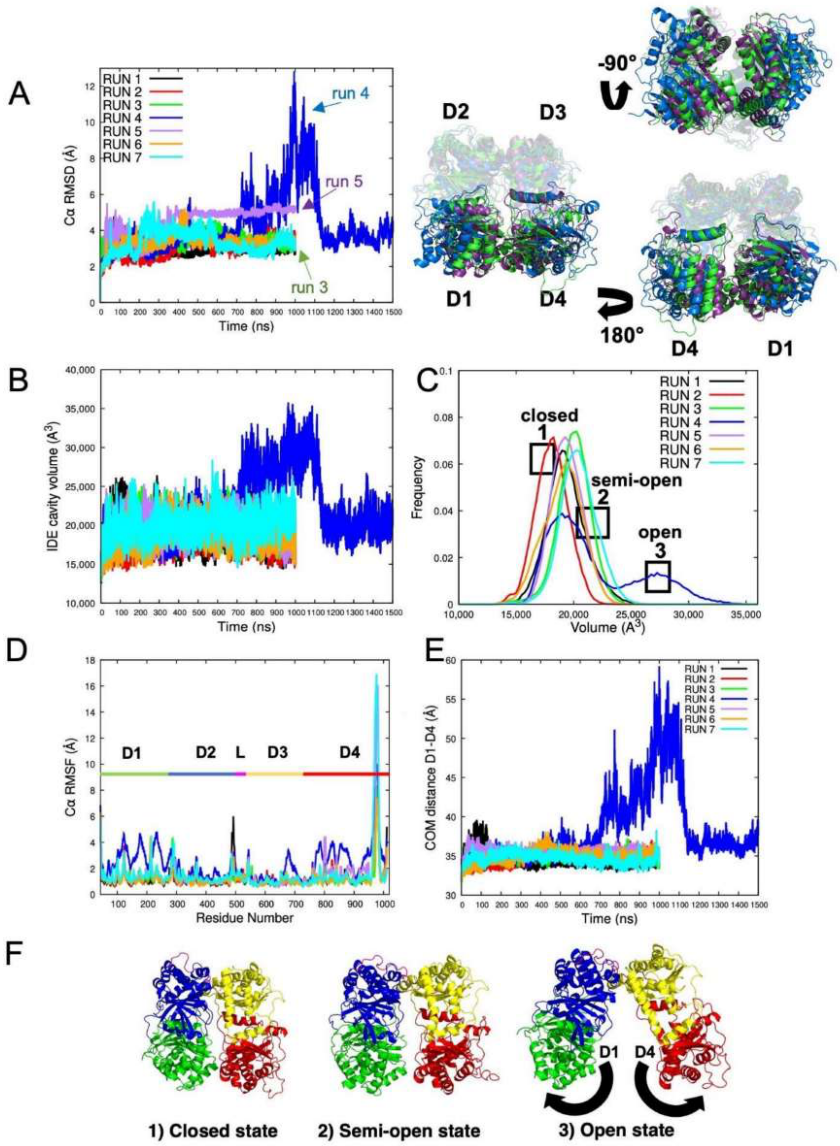


Figure 2. Evaluation of IDE MD simulations and major states of IDE. (**A**) Left: C*α* RMSD evaluation of IDE trajectories and comparison of IDE structures from run 3 (closed), run 4 (open), and run 5 (swinging door) at 1 µs. Right: domain structural analysis: only D1 and D4 are highlighted according to their designed run color, to see the conformational changes of the IDE door. (**B**) Evolution of IDE cavity volumes. (**C**) IDE cavity volume frequency for each MD trajectory. (**D**) C*α* RMSF of IDE different domains. (**E**) Center of mass distance between two domains: D1 and D4. (**F**) Major states of IDE explored during MD simulations.

Altogether, RMSD results showed that IDE exhibited interesting variations. These RMSD fluctuations can be correlated to the structural changes that IDE adopts to be stabilized. Moreover, RMSD fluctuations can be directly linked with IDE cavity changes, which are responsible for the protein structural rearrangements.

2.1.2. IDE Cavity Volume and Hydration Analysis

Accordingly, the cavity volumes of IDE were calculated to examine the expansions of its structural flexibility and potential different states. As shown in Figure 2B, all trajectories, apart from the fourth, explored the same volumes with values ranging between ~15,000 and ~25,000 A^3. The fourth run reached its highest value at 950 ns with a volume of ~35,000 A^3, then a prompt decrease appeared at 1.2 µs, which ended in a plateau until the end of the simulation. Thus, these results are perfectly correlated with the RMSD values previously observed. Considering that the cavity of IDE has an initial volume of ~15,000 A^3 in its closed state [15], the frequency distributions of IDE volumes were calculated and plotted in Figure 2C. Similarly, most of the trajectories displayed volumes corresponding to closed or semi-open states. On the other hand, the fourth run clearly showed the exploration of at least two different states. Indeed, the trajectory is divided into two separate and uneven populations. A more occupied population of volumes defined the closed or semi-open state, whereas a lesser population represented the open state. Therefore, IDE explored at least three different states (closed, semi-open, and open) during the MD simulations. We hypothesize that most of the trajectories have met the closed and semi-open states while the fourth trajectory, which displayed volumes more than twice the initial one, has met the fully open state. With the combination of these results, the IDE open-closed switch represented in the fourth trajectory was displayed (Video S1, Supplementary File 1) to capture IDE movements.

Water molecules are important components in maintaining the functions of proteins. Since the RMSD and the cavity volume analyzes indicated major structural changes in the IDE structure, the solvent molecules, and the total solvent accessible surface area (SASA) were calculated for all trajectories. The SASA analysis stands for the solvent accessible area. Low values or a decrease in the SASA indicate a closed state of the protein structure with very few hydrophobic areas accessible to the solvent. On the other hand, high values, or an increase in the SASA, describe a certain degree of protein's flexibility and the strong exposure of the cavity to the aqueous environment of the system. Thus, the higher the values of SASA, the more the cavity is exposed to the solvent and witnesses several conformational changes. As shown in Figure S1A, SASA values mostly ranged between 39,000 $Å^2$ and 48,000 $Å^2$, with various fluctuations. As expected, the fourth system displayed the most important values of SASA with its highest value reaching ~49,000 $Å^2$ corresponding to the expansion of IDE cavity. The fourth system also witnesses an important decrease of the SASA that correlates with the drop of the solvent molecules (Figure S2B, Supplementary File 1). This event was due to an IDE cavity volume decrease that was accompanied by the simultaneous expulsion of the water molecules. Thus, the SASA analysis, together with the RMSD and the volume cavity calculations, summarize that these results were coupled to protein conformational changes. These results also support the hypothesis that IDE switches from a closed to an open state through an allosteric behavior.

2.1.3. Root Mean Square Fluctuation (RMSF) of IDE Structure

To further analyze the flexibility and local changes in the structure of IDE, Cα RMSF of each residue has been calculated (Figure 2D). RMSF analysis revealed that D1, D4, and an adjacent region to the linker are the most flexible parts of IDE. These fluctuations are mostly observed for the fourth run. Cα RMSF values reached about 4 Å compared to the rest of the MD trajectories (2.0–2.5 Å). These observations agree with experimental data where the swinging door of IDE is mostly carried out by the movements of both D1 and D4, which are the principal regions causing the protein to undergo different

states [28,30]. Figures S2 and S3 illustrates the most flexible regions of the protein along with the residues exhibiting the highest Cα RMSF values. Interestingly, the represented amino acids (Figure 2D) displayed the same pattern of fluctuations in all the trajectories, but with higher Cα RMSF values for the fourth one. As expected, most of the residues are positioned on solvent-exposed regions, such as loops (Figure S4, Supplementary File 1). Therefore, the high fluctuations of these residues can be explained through their intramolecular and intermolecular interactions within the protein and with the solvent. The most flexible residues positioned on D1 and D4 are also solvent exposed but were observed to play major roles in maintaining D1–D4 interactions along the IDE gate. For example, residue Q828 (D4) has been identified to be a key residue interacting with different residues of D1 (R181, E182, and N184) [24]. Residue Q828 exhibits a Cα RMSF value of 4.1 Å when IDE is open against a value of 2.1 Å when IDE is closed. Accordingly, in the closed conformation of IDE, residues R181, E182, and N184 display values of 1 Å, 0.8 Å, and 1 Å, respectively. However, these residues present higher values when IDE is open with values reaching 4.3 Å, 3.8 Å, and 4.0 Å, respectively. Additionally, the flexibility of the exosite and the catalytic site residues were also examined. For all trajectories, both regions exhibited very low Cα RMSF in all trajectories, which did not exceed 1 Å, indicating their structural stability.

To support these results, the distance between the center of mass (COM) of D1 and D4 was plotted in Figure 2E. Interestingly, we observed the same pattern as the previous graphs. D1 and D4 were seen to be moving closer to each other at the beginning of the MD trajectory. Then, both domains moved away to reach their maximum distance value at ~59 Å followed by the recovery of their initial distance towards 1.1 µs. As shown in Figure 2F, the different major states of IDE are illustrated in the complementation of previous results. Furthermore, to rule out the hypothesis that the IDE open state might be an artifact, we analyzed the backbone RMSD of each domain during the fourth trajectory (Figure S5, Supplementary File 1), which appeared to be stable along the simulation time. It is a simple way that serves as an indicator of conformational stability in the system during the simulation.

2.1.4. IDE Hinge Dynamics Analysis

IDE must undergo a hinge-like motion to transition from a closed to an open conformation. This transition is required for the entry of substrates and the release of proteolytic products. IDE possesses a hinge loop or a linker (516–541) connecting D2 and D3. This linker is critical for the proper pivoting motion between the IDE-N and IDE-C. Therefore, the hinge loop is an important region to regulate the allostery of IDE.

To study the hinge loop dynamics and its role in the domains pivoting movements, we used a quasi-harmonic approximation implemented in the gmx anaeig module of GROMACS [31]. Cα atoms of IDE were selected to carry out the analysis. We compared only the movements projected on the first eigenvector as they exposed the major differences. Figure 3A shows the superposed extreme projections of the linker along the first eigenvector. As shown in Figure 3A, we compared the behavior of the major motions of the hinge-loop, both when IDE is open and closed. Surprisingly, the loop remained stable and only slight movements can be noticed in the hinge. In the closed conformation, the hinge displayed minimal back and forth movements coinciding with the protein domain fluctuations. As for the open conformation, the linker represents more fluctuations as it moves along with the extension of the cavity. Accordingly, Figure 3B shows the Cα RMSF evaluation of the linker in each run after extraction of the frames of the quasi-harmonic analysis. The loop exposed higher flexibility in the open conformation, compared to the closed one, with RMSF values ranging between 1.0 and 2.0 Å. The main explanation as to how the loop stays stable is because of the very tight interactions made within the protein. Indeed, the linker is a conserved region of the M16 metalloproteinase proteins [30,32].

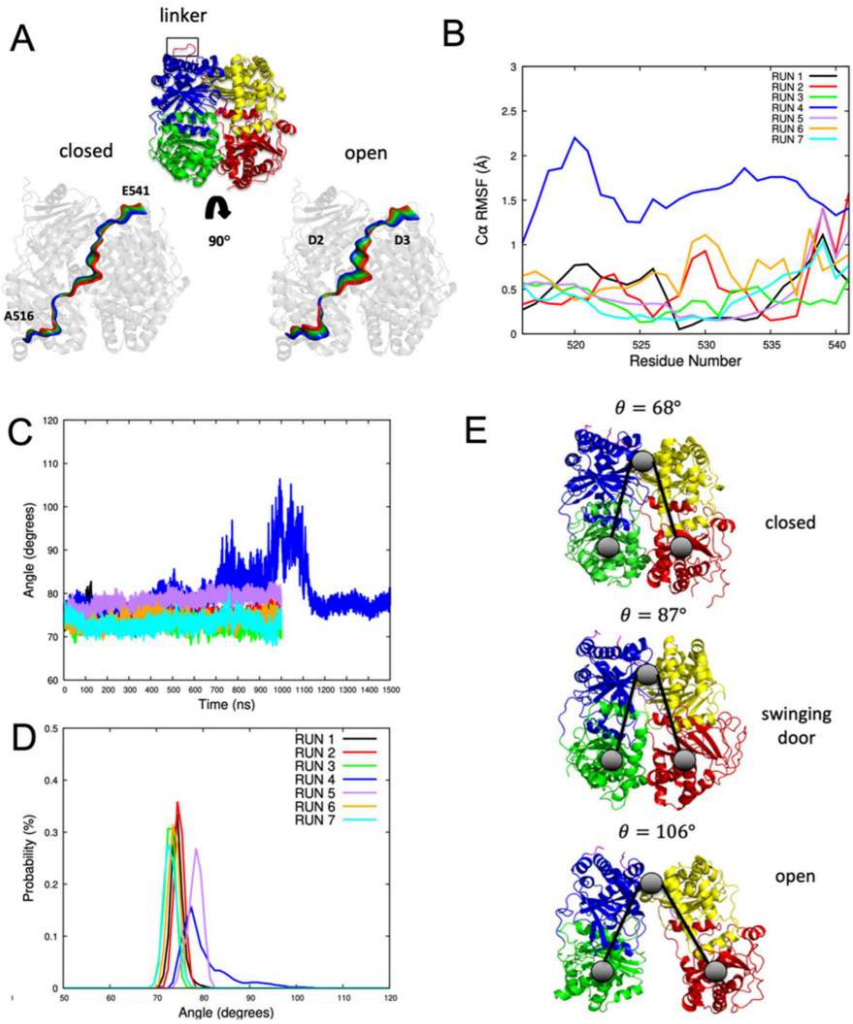


Figure 3. Quasi-harmonic analysis of the IDE hinge-loop. (**A**) The superposed frames of the third (closed) and fourth (open) trajectories are projected onto the first eigenvector. A total of 100 frames were sampled for each trajectory. Colors range from blue (first frame) to red (last frame). (**B**) Cα RMSF evaluation of the linker represented in the 100 sampled frames after the quasi-harmonic analysis. (**C**) Characterizing the hinge-loop motion and angle measurements by taking the COM of D1 and D4 combined with the COM of the linker. The angles were measured along the trajectories. (**D**) Probability of the angle values for each trajectory. (**E**) Representation of the most probable closing/opening angles.

Almost every residue of the loop interacts with either the D2 or the D3 domain through hydrogen bonding, hydrophobic interactions, and salt bridges (Figure S6, Supplementary File 1). For example, K521, K527, and E541 form salt bridges with residues E349, E529,

and K735, respectively. Several hydrogen bonds are formed among N528, E413, and A610. E536 and L538 form two hydrogen bonds with N732. T533 interacts with D636 and K637. Similarly, residue N543 interacts with D636 as well. As for E541, it also interacts with Q563 through hydrogen bonding. There are multiple hydrophobic patches formed with D2 and D3 through the hinge loop. Residue L524 forms hydrophobic interactions with L401 and W409. F530, which is a crucial residue for the maintenance of the catalytic role of IDE [30], interacts with Y607, A611, L616, I640, and L641. Finally, F535 interacts with V420 and F424. These extensive interactions are also conserved in both the closed and open state of IDE; hence, the preservation of a stable linker structure.

Next, we measured the opening angle of IDE to describe, in further detail, the hinge-type motion (Figure 3C–E). Taking the COM of D1 and D4, combined with the COM of the linker, yields an opening angle of a maximum of 106 degrees (run 4) compared to ~68 degrees when it is closed (runs 1, 2, 3, 6, and 7). As expected, the angle values of the fourth trajectory follow the same pattern as the previous results (Figure 2A,F). The angle values fluctuate extensively until reaching their maximum at ~990 ns. A prompt decrease is observed at 1.1 µs, coinciding with the closing of the IDE cavity. Interestingly, the swinging door state (run 5) displayed a distinct angle spanning between 75 and 80 degrees. Therefore, the latter observation confirms the slight opening and twisting motions characterizing the swinging door motion of IDE. This leads to the conclusion that the description of the opening angle, using the hinge-loop as the center point, results in a more accurate distinction between the open and closed states.

2.1.5. Gibbs Free Energy Landscape Analysis

Protein allostery is fundamental to understanding protein functions. Intra-protein atoms distances work via many allosteric processes with a defined path and the catalytic activity of IDE was proposed to be allosterically regulated in several papers [33–35]. However, the detailed mechanism of IDE allostery remains unknown. Here, we describe and retrace the allosteric communication of IDE as a series of local structural changes using the free energy landscape (FEL) approach. To study IDE dynamics movements, all trajectories were concatenated and the final free energy landscape for the first two most contributing principal components (PC) were calculated. To achieve this, a covariance matrix is constructed using the protein backbone coordinates. Then, the diagonalization of this matrix yields a set of eigenvectors and eigenvalues describing the collective modes of the fluctuations of the protein. Generally, the eigenvector with the largest eigenvalue or PC represents the large-amplitude collective motions of the protein. Since we have a system displaying significant movements, we selected the first two PCs characterizing these dominant motions. Moreover, it is very important that protein systems are locally equilibrated, and the determined pathways do not represent artifacts of the chosen coordinates. Thus, given the previous results, the FEL analysis was only used for the converged trajectories, which excluded the fourth MD simulation. These trajectories were concatenated into one single MD simulation to produce the FEL map.

The FEL of IDE is shown in Figure 4A. The lowest energies are represented in blue, whereas high energies are indicated in red. This means that the blue regions represent stable states of the protein, while the red areas describe the unstable states explored during the MD simulations. Moreover, several bins with minimum energies mean that the protein explores different conformations through transition states. Accordingly, the global minimum conformations of IDE were extracted regarding the stable and unstable FEL bins to differentiate them.

The projection of the backbone trajectories along the PCs revealed three major bins ((I), (II), and (III)). The investigation of these bins revealed that the first bin enclosed the structures of the first, third, second, and seventh trajectories, making it the major energy minima and the most occupied basin. Interestingly, the sixth system visited the first bin from 0 to ~47 ns, then jumped off from the major local minimum (I) and moved towards a more distant region of the conformational space, the second basin (II). Additionally, the

fifth system explored the first basin (I) from 0 to 300 ns, then transited to the third basin (III) corresponding to the RMSD jump observed in Figure 2A.

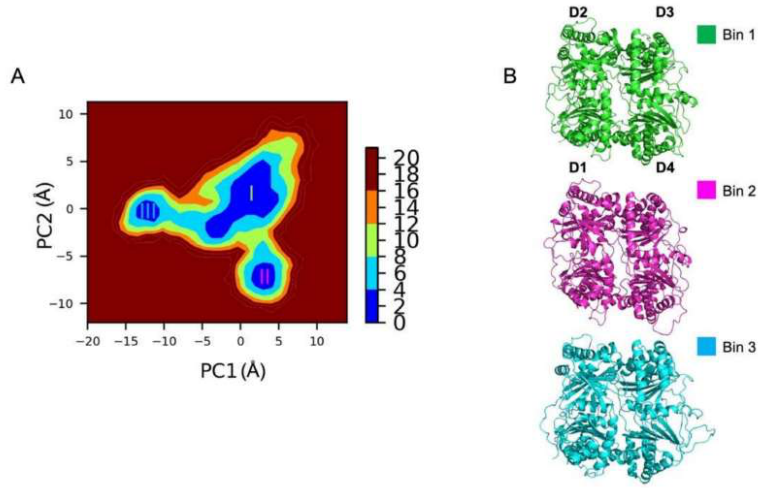


Figure 4. (**A**) Gibbs free energy landscape (FEL) analysis (**B**) with representative IDE structures extracted from the MD simulations. Free energy values are represented in kJ/mol, and their colors are detailed in the color bar. The representative IDE structures are represented in cartoons and display each basin.

Visual examination of the lowest energy structures belonging to their corresponding bins revealed a very similar structure (Figure 4B). Thus, the closed conformations belonged to a single state but their mapping onto different space coordinates shows that structural changes appeared during the MD simulations. Therefore, to understand the conformational changes reflecting the different bins, we extracted one representative IDE structure from each bin (Figure 4B). Compared to most of the MD simulations, the sixth system revealed a slight greater flexibility in the alpha helices of the different domains, describing the constant dynamic movements of the protein. Overall, an RMSD of 2.0 Å was calculated between state (I) and state (II). For the fifth system, a higher RMSD value (2.6 Å) was calculated between state (I) and state (III). The fifth trajectory has already been described in the previous results as a different state on its own with a specific dynamic motion displaying a particular angle between D1 and D4. Therefore, the FEL map regrouped the different structural changes of IDE in terms of energy and highlighted important transitions in the cycle of IDE.

2.1.6. Non-Covalent Interactions

Residue interactions play major roles in IDE dynamic movements and allostery. Knowing that IDE is a flexible protein, it is very important to check the stability of the hydrogen bonds using MD simulations rather than inspecting only the crystal. Thus, the number of hydrogen bonds (HB) between D1 and D4, and between D2 and D3 were calculated (Figures 5 and S7, Supplementary File 1). Notably, the number of HB varied throughout the simulations and significant changes were observed.

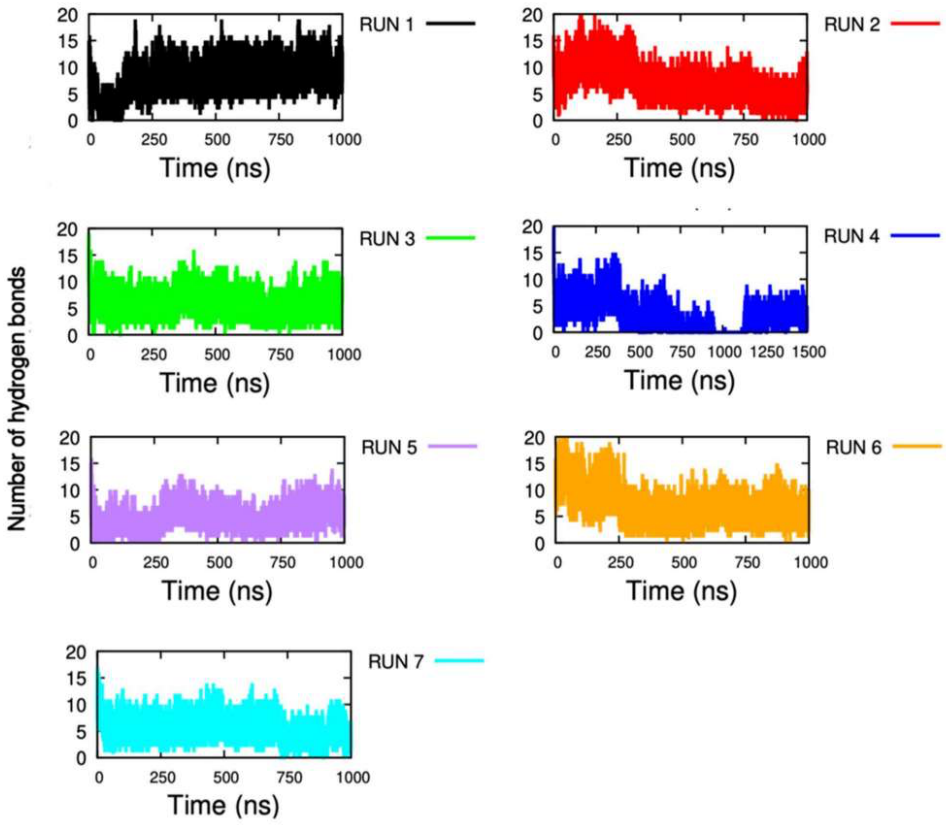


Figure 5. Number of hydrogen bonds formed between D1 and D4 along the MD simulations. Black, red, green, blue, purple, orange, and cyan are respectively for runs 1, 2, 3, 4, 5, 6, and 7.

In most of the trajectories, the number of HB between D1 and D4 remained approximately stable and witnessed only little fluctuations. For example, for all the trajectories, the number of hydrogen bonds averaged between 5 and 7 (Figure 5). However, during the fourth trajectory, we can clearly observe a total disruption coinciding with the frame time of IDE open state exploration (900 ns–1.1 µs). Indeed, the drop of the number of HB, going from ~15 to 0, clearly defined the closed-open switch of IDE that was similarly witnessed in the previous results. Then, a recovery of the number of HB can be seen around 1.1 µs (8 HB) when IDE regains its closed conformation. The same analysis was applied for the D2–D3 complex (Figure S7, Supplementary File 1). Similarly, D2 and D3 HB were stable during the MD simulations. The number of bonds averaged between 7 and 9. As expected, the fourth trajectory also displayed a decrease of the number of HB at the same frame time of D1–D4 HB disruption. However, this decrease was less important than the one observed between D1 and D4. Therefore, these results reinforce the fact that D1 and D4 are the main "gate" domains of the IDE closed-open switch. These results also suggest that D2 and D3 are still sustained during the open switch of IDE to maintain a certain stability of the protein.

To identify the residues involved in HB formation between the different domains of IDE, HB occupancy was calculated. In Table 1, HB and salt-bridges (SB) observed across D1–D4 and D2–D3 binding interfaces are listed, together with their average occupancy

percentage during the simulations. We isolated only the most frequent interactions with a threshold of 10% and with a cutoff distance of 4.0 Å.

Table 1. D1–D4 non-covalent interactions occupancy (%) during MD simulations. HB and SB were reported only if they exist for >= 10% of the investigated period. Backbone (bb) and side chain (sd) interactions were specified.

Domain 1	Domain 4	Non-Covalent Interaction Type	Occupancy (%)						
			Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7
D84 (sd)	K898 (sd)	SB	31.7	73.2	39.7	35.8	92.0	76.3	37.4
E133 (sd)	K884 (sd)	SB	100.0	98.0	100.0	73.1	59.7	100.0	100.0
E182 (sd)	R824 (sd)	SB	57.6	100.0	86.5	26.9	0.0	26.9	88.5
S132 (sd)	E817 (sd)	HB	82.3	39.4	34.1	0.0	0.0	35.8	27.1
K85 (sd)	D895 (sd)	SB	90.8	80.0	60.3	43.3	35.1	52.0	100.0
N184 (sd)	Q828 (bb)	HB	70.6	0.0	0.0	0.0	0.0	0.0	0.0
N184 (sd)	Q828 (sd)	HB	NA	0.0	0.0	0.0	0.0	11.0	0.0
R181 (sd)	Q828 (sd)	HB	NA	0.0	12.9	0.0	0.0	0.0	0.0
N184 (sd)	Q828 (sd)	HB	NA	14.9	0.0	0.0	0.0	0.0	0.0

Four pairs of residues (K898-D84, K884-E133, R824-E182, and K85-D895) formed SB between D1 and D4. The salt bridge formed by K898 and D84 remained stable for all the trajectories (Figure 6). In run 5, the HB occupancy was particularly high due to the angle formed by D1 and D4 (Figure 3). This angle, creating a favorable interaction together with an important lifetime K898–D84 bond, was completely disrupted during run 4. However, it recovered completely when IDE regained its closed conformation at the end of the MD simulation. Residue K884 formed a stable SB with E133. The SB occupancy ranged from 59.7% to 100.0%, attesting to the bond strength and sustainability. The bond created between R824 and E182 was positioned at the main gate of IDE. Interestingly, the R824–E182 bond was observed to be maintained in all trajectories, except for the fifth one. The swinging door motion did not allow the bond to be formed since D1 and D4 might have been too far away from each other and too flexible. As for the fourth trajectory, the R824-E182 interaction witnessed a low percentage of occupancy due to the IDE closed state recovery. The S132-E817 bond was positioned on two helices of IDE that were only accessible through the bottom side of the protein (Figure 6). The HB was well maintained in all simulations (34.1% to 82.3%) except for runs 4 and 5. As expected, in this case, the two domains were not close enough to form the HB. The fourth SB formed with K85 and D895 was also stable with higher occupancy when IDE was closed. Finally, Q828, which is a key residue positioned at the main gate of IDE, was revealed to interact with N184 and R181. Yet, Q828 was observed to form a stable bond only in a few simulations. This appears to be related to the high flexibility of the residue (Figures 1D and S4, Supplementary File 1). Altogether, the interaction patterns across the D1 and D4 interface agree with those known from experiments [24] and the main interactions were recovered in our analysis.

Compared to D1 and D4, more residues were observed to maintain the stability of the D2–D3 domains and 7 SB were formed in the D2–D3 complex (Table S1 and Figure S8, Supplementary File 1). Residue D309 interacted with several residues, excepts in simulations 3 and 7. D309 was found to interact with N672 through its backbone in runs 2, 4, 5, and 6. The absence of this interaction was compensated by the interaction of D309 with either N671 or R668. K657 formed a stable SB with E382 in all the trajectories apart for the fifth run, where K657 interacted with E381 (96.7%) instead. Additionally, K571 connected the D2–D3 complex in all MD simulations with residues D426 and F424, where they occupied approximately 80% and 40% of the simulation time, respectively. It is also important to note that the SB contributed by K351 interplay with E606 and D602 to stabilize the protein. Moreover, the H336-Y609 stable HB interaction was mainly found in all simulations. Interestingly, H336 was replaced by H340 in the second trajectory to connect with Y609 (22.7%). Finally, residue R311 formed a stable SB with E664, the interaction occupied approximately

~70–100% of the trajectories. On the other hand, R311-E664 did not appear possible in the fifth simulation due to the angle formed by the swinging door conformation. As for the sixth run, R311 preferably interacted with R668 (21.7%) and, alternatively, E664 formed an HB with E381 (~37%) (data not shown).

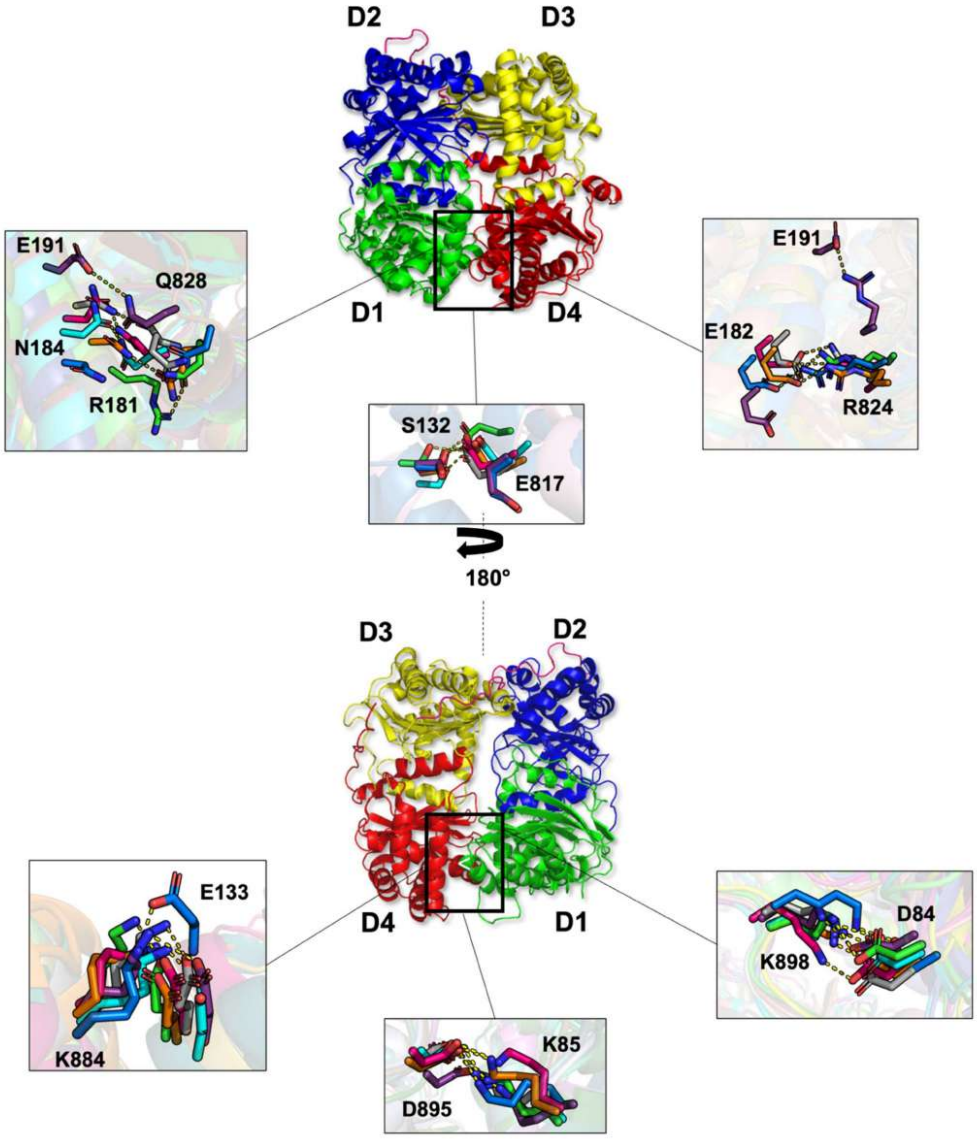


Figure 6. Hydrogen bonds and salt bridges formed between D1 and D4. The color code was conserved, according to the simulation number. Grey, red, green, blue, purple, orange, and cyan are respectively for runs 1, 2, 3, 4, 5, 6, and 7.

To further explore the dynamics and the movements of the residues relative to IDE allostery and support the previous results, we calculated the COM distance of each pair of residues along each trajectory (Figures S9 and S10, Supplementary File 1). As shown in Figure S8, residues K888-E133, D895-K85, and K898-D84 displayed the most stable distances, which agree with the results observed in Table 1 and Figure 6. Apart from the fourth simulation (~30.0–48.0 Å), the COM distances exhibited little fluctuations and did not exceed 10.0 Å. Thus, these observations support the sustainability of K888–E133, D895–K85, and K898–D84 interactions within the D1 and D4 domains. As expected, the fourth simulation displayed a prompt increase representative of the opening of IDE, as tediously described in previous results. As for residues S132–E817, a clear increase of the COM distances can be observed for the fifth trajectory. Indeed, in the fifth MD simulation, residue, E817 does not interact with S132, hence the large distance between the two amino acids. Similarly, in this same trajectory, residue R824 and Q828 interact with E191 instead of E182 and N184, respectively. Thus, these observations explain the larger COM distance in run 5 for those residues.

The same analysis was applied on the D2–D3 complex in Figure S9. A first observation is that compared to the D1 and D4 interactions, little fluctuations are observed on the COM distance plots for D2 and D3. This result support the fact that D2 and D3 tend to follow rigid body movements among the large-scale motions of IDE. However, this does not exclude that a few residues witnessed higher fluctuations and larger COM distances. For example, D309-N672, D309-N672, D309-R668, E382-K657, E381-K657 and R311-E664 are the main residues displaying the largest distances (~20.0–25.0 Å). It should be noted that these amino acids are located on the front surface of IDE. On the other hand, residues D424-K571, F424-K571, K351-E606, K351-D602, H336-Y609 and H340-Y609, exhibited lower and stable COM distances (~10.0 Å). These residues are located on the back surface of IDE. Additionally, these same observations can be noticed for the D1–D4 complex. With these results, we suggest that fluctuations mostly occur in the front side of IDE while the back side and the rest of the protein is mainly stable. These results were interesting as they confirm that the main entrance of peptides and solvent is the most flexible region of the protein while the rest of the protein structure tend to have restricted motions to stabilize the ensemble of the four domains. Altogether, our results prove that the described residues are the main partaker in IDE dynamic movements and forming gate.

3. Materials and Methods

3.1. Protein Modeling

Homology modeling was performed to get the full-length structure and wild type/active form of the IDE protein [24]. For this purpose, the substrate-free 3D crystal structure of human IDE was retrieved from the RCSB Protein Data Bank in PDB format [36] (PDB ID: 2JG4, resolution: 2.80 Å) [37]. We selected the PDB ID: 2JG4 as it represents the only crystallized structure of the closed substrate-free IDE with the lowest resolution. Other IDE closed conformations can be found in the RCSB Protein Data Bank. However, these PDB structures are resolved with either Cryo-EM, FAB antibodies, peptides, or inhibitors, and can have poor resolutions. From this perspective, we fully focused our study on the closed form of the IDE free of substrates to concentrate on its conformational changes through MD simulations.

The crystal structure contains one mutation at the catalytic site (Y831F). The structure also lacks residues 971–978 and 1012–1019 in the fourth domain. Therefore, the homology model building was carried out using the MODELLER software (v10.1) [38] with PDB:2JG4 as a template. Correct side-chain orientations of the catalytic site residues (H108, E111, H112, and E189) were verified to avoid any clash with other residues and the zinc atom. For validation, the model with the lowest value of the DOPE assessment score [39] was selected for further analysis and VERIFY 3D [40] was used for further endorsement (Figure S11, Supplementary File 1). The final model comprised 15,880 atoms corresponding to the 977 residues (43–1019) of the full-length IDE. The protonation states of ionizable residues in the

IDE model were assigned using PROPKA [41] and the pH was set at a physiological value of 7.5.

3.2. Molecular Dynamics Simulations

We ran 7 large-scale simulations of 1 to 1.5 µs each. MD simulations were carried out with the GROMACS (v2019.5) software package [31] and we used the CHARMM36m force field [42] for both the protein and zinc parameters. The protein was placed in a dodecahedron-shaped water box (TIP3P) and a minimum of 10 Å was preserved between each atom of the system and the walls of the box. All runs were run at a 300 K temperature and with a time step of 2 fs. All bonds were constrained using the LINCS algorithm [43] for the protein and the SETTLE algorithm [44] for the water. The energy of the system was minimized over 1000 steps using the Steepest descent algorithm after ion addition and system neutralization. The minimization convergence was set at a maximum force of 1000 kJ/moL/nm. The number of particles, volume, and temperature (NVT) equilibration was performed for 1000 ps at a temperature of 300 K with a coupling constant of 0.1 ps. number of particles, pressure, and temperature (NPT) was run by setting the temperature to 300 K and the pressure to 1 bar. Electrostatic forces were calculated with the particle-mesh Ewald algorithm [45]. For trajectory analysis, we used GROMACS packages, VMD (v. 1. 9. 4a38) [46], PyMOL [47], and GNUPLOT (v. 5. 2) [48]. The volume of the IDE crypt, as well as the number of solvent molecules, were computed using the trj_cavity_v2.0 program [49].

RMSD and RMSF analyses were led with GROMACS. The RMSD analysis measures average distances (in Å) of the studied systems from the corresponding starting structure over the simulation period. RMSD is defined as:

$$\mathrm{RMSD}(t) = \left[\frac{1}{M} \Sigma_{i=1}^{N} m_i \left| r_i(t) - r_i^{ref} \right|^2 \right]^{1/2}$$

where $M = \Sigma_i m_i$ is the total mass, m_i is the mass of atom i, N is the number of atoms, and $r_i(t)$ is the position of atom i at time t after least square fitting the structure to the reference structure. In our study, the calculation and fitting of the protein were done on the Cα atoms.

To examine the flexibility and the local changes in the structure, Cα RMSF versus the number of IDE systems were investigated. The RMSF equation is defined as:

$$\mathrm{RMSF}_i = \left[\frac{1}{T} \Sigma_{t_j=1}^{N} \left| r_i(t_j) - r_i^{ref} \right|^2 \right]^{1/2}$$

where T is the time over which the mean coordinate is calculated regarding the r_i^{ref} (reference position of particle i). The RMSF was computed from the atomic coordinates of the Cα atoms as well.

3.3. Gibbs Free Energy Landscapes

The Gibbs Free energy landscapes were performed to describe IDE motions and different states through the simulations. We applied the gmx covar, gmx anaeig, and gmx sham modules of GROMACS to calculate the two-dimensional representations of the FEL. The FEL of each run was constructed using the projections of their first (PC1) and second principal components (PC2) or eigenvectors.

$$\Delta G(PC1, PC2) = -K_B T \ln P(PC1, PC2)$$

where K_B is the Boltzmann constant, T is the temperature of the simulation, P(PC1, PC2) illustrates the probability distribution of the system along with the first two principal components. PC1 and PC2 usually display the dominant fluctuations in residues for a protein, hence why they were studied in this project.

4. Conclusions

The catalytic activity of IDE is mediated by several structural transitions. These conformational changes are allosterically regulated by its substrates, ATP, and other protein partners [10,33–35,50]. In this paper, we suggest that IDE allostery is conducted by the collective motions of the numerous atoms forming IDE structure. Based on our data, we present a model of IDE in its active state to explain how the allostery of IDE is coordinated by the equilibrium between IDE closed and open states. Our MD simulations revealed that IDE undergoes several states and is a flexible system. RMSD (Figure 2A) and IDE cavity volume calculations have put forth at least three different explored conformations. These different conformations are defined by a closed, a swinging door, semi-open (intermediate state), and an open state. We believe that the probability to recurrently exploring IDE states or new conformational changes can increase and enhance with further computational methods. Notably, future work can focus on the required transition energy to switch from a closed state to an open state with additional MD sampling.

Interestingly, all states were explored during the fourth trajectory with a complete cycle of IDE closed–open–closed transition. Such motions can explain how allostery regulates and governs the IDE biological structure but also suggest additional ways in how IDE may function. Furthermore, the swinging door motion state [30] was explored by one of the simulated trajectories (RMSD: ~4.0–5.0 Å (Figure 2A)). Thus, we completed these results with an RMSF analysis to define the dynamic domains responsible for IDE flexible movements (Figure 2D). RMSF results revealed that D1 (RMSF: ~4.5 Å) and D4 (RMSF: ~4.0 Å) are the main actors of IDE distinct motions, which agree with different published studies [26,30]. These results were supported by the calculation of IDE cavity volumes over time (Figure 2B). The description of the volumes described the different states explored during the MD simulations. The closed and semi-open state volumes ranged between ~15,000 and ~25,000 A^3, while the open state reached higher values (~15,000 A^3 to ~35,000 A^3). Thus, our data agree with the experimental data, where the average cavity volumes (~35,600 A^3) and domain distances reaching their highest values at ~37 Å and ~55 Å are very similar to the simulated semi-open and open states, respectively [26]. Additionally, the cavity analysis was complemented with a SASA study (Figure S2, Supplementary File 1). Moreover, we investigated the hinge dynamics through a quasi-harmonic study (Figure 3). The linker connecting IDE-N and IDE-C was shown to be extensively flexible as the IDE cavity opens, yet stable enough to maintain D2 and D3 connected. Additionally, the angles formed between the D1-linker-D4 in the different simulations were calculated with the aim of supporting the previous results (Figure 3). The determined angles were 68°, 87°, and 106° for the closed, swinging door, and the open sate, respectively. The obtained results confirmed that the angles match with the different structural states observed during MD. It was also observed that IDE is a very dynamic protein and can go through several conformational changes. Therefore, we ran a Gibbs FEL analysis to explore the stability of the explored states of IDE (Figure 4). Only converged MD simulations were used to construct the FEL map. Previous conformation changes observed in IDE protein were confirmed and explored in terms of energy in the FEL analysis. Three major bins were visited by IDE structures, displaying the protein dynamic and its closed state. Moreover, the swinging door motion of IDE was explored as a separate state, which confirms previous results. The open state is still a very challenging conformation to capture. Hence, these observations support the fact that IDE is unstable and needs a substrate to be stabilized in its open conformation. It is also known that extensive contacts are shared between IDE-N and IDE-C, so the IDE structure can be maintained (Figure 5). Thus, the residues maintaining these contacts, as well as the effect of IDE opening, were determined (Tables 1 and S1, Figures S9 and S10, Supplementary File 1). With these results, several mutations, MD simulations, and docking studies with different IDE substrates can be directed to complete the full biological cycle and allostery mechanism of IDE.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/ijms23031746/s1.

Author Contributions: M.G.: conceptualization, methodology, formal analysis, investigation, writing—original draft preparation, writing—review and editing; N.J.: writing—review and editing, project administration, funding acquisition; A.-C.C.: conceptualization, methodology, investigation, writing—original draft preparation, writing—review and editing, supervision, project administration, funding acquisition; G.M.: conceptualization, methodology, investigation, writing—original draft preparation, writing—review and editing, supervision, project administration, funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Agence Nationale de la Recherche (PIF21 project, No. ANR-19-CE18-0023). The authors gratefully acknowledge the financial support of the Université de Paris, the CNRS institute, and the INSERM institute.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rawlings, N.D.; Barrett, A.J. Homologues of insulinase, a new superfamily of metalloendopeptidases. *Biochem. J.* **1991**, *275*, 389–391. [CrossRef] [PubMed]
2. Pivovarova, O.; Höhn, A.; Grune, T.; Pfeiffer, A.F.H.; Rudovich, N. Insulin-degrading enzyme: New therapeutic target for diabetes and Alzheimer's disease? *Ann. Med.* **2016**, *48*, 614–624. [CrossRef] [PubMed]
3. Hulse, R.E.; Ralat, L.A.; Wei-Jen, T. Chapter 22 Structure, Function, and Regulation of Insulin-Degrading Enzyme. In *Vitamins & Hormones*; Elsevier: Amsterdam, The Netherlands, 2009; pp. 635–648.
4. Tundo, G.R.; Sbardella, D.; Ciaccio, C.; Bianculli, A.; Orlandi, A.; Desimio, M.G.; Arcuri, G.; Coletta, M.; Marini, S. Insulin-degrading enzyme (IDE): A novel heat shock-like protein. *J. Biol. Chem.* **2013**, *288*, 2281–2289. [CrossRef] [PubMed]
5. Fernandez-Gamba, A.; Leal, M.; Morelli, L.; Castano, E. Insulin-Degrading Enzyme: Structure-Function Relationship and its Possible Roles in Health and Disease. *Curr. Pharm. Des.* **2009**, *15*, 3644–3655. [CrossRef]
6. Tang, W.-J. Targeting Insulin-Degrading Enzyme to Treat Type 2 Diabetes Mellitus. *Trends Endocrinol. Metab.* **2016**, *27*, 24–34. [CrossRef]
7. González-Casimiro, C.; Merino, B.; Casanueva-Álvarez, E.; Postigo-Casado, T.; Cámara-Torres, P.; Fernández-Díaz, C.; Leissring, M.; Cózar-Castellano, I.; Perdomo, G. Modulation of Insulin Sensitivity by Insulin-Degrading Enzyme. *Biomedicines* **2021**, *9*, 86. [CrossRef]
8. Duckworth, W.C.; Bennett, R.G.; Hamel, F.G. Insulin Degradation: Progress and Potential. *Endocr. Rev.* **1998**, *19*, 608–624. [CrossRef]
9. Farris, W.; Mansourian, S.; Leissring, M.A.; Eckman, E.A.; Bertram, L.; Eckman, C.B.; Tanzi, R.E.; Selkoe, D.J. Partial Loss-of-Function Mutations in In-sulin-Degrading Enzyme that Induce Diabetes also Impair Degradation of Amyloid β-Protein. *Am. J. Pathol.* **2004**, *164*, 1425–1434. [CrossRef]
10. Kurochkin, I.V.; Guarnera, E.; Berezovsky, I.N. Insulin-Degrading Enzyme in the Fight against Alzheimer's Disease. *Trends Pharmacol. Sci.* **2018**, *39*, 49–58. [CrossRef]
11. Zhao, L. Insulin-Degrading Enzyme as a Downstream Target of Insulin Receptor Signaling Cascade: Implications for Alz-heimer's Disease Intervention. *J. Neurosci.* **2004**, *24*, 11120–11126. [CrossRef]
12. Nalivaeva, N.N.; Belyaev, N.D.; Kerridge, C.; Turner, A.J. Amyloid-clearing proteins and their epigenetic regulation as a thera-peutic target in Alzheimer's disease. *Front. Aging Neurosci.* **2014**, *6*, 235. [CrossRef] [PubMed]
13. Li, H.; Wu, J.; Zhu, L.; Sha, L.; Yang, S.; Wei, J.; Ji, L.; Tang, X.; Mao, K.; Cao, L.; et al. Insulin degrading enzyme contributes to the pathology in a mixed model of Type 2 diabetes and Alzheimer's disease: Possible mechanisms of IDE in T2D and AD. *Biosci. Rep.* **2018**, *38*, bsr20170862. [CrossRef] [PubMed]
14. Carrasquillo, M.M.; Belbin, O.; Zou, F.; Allen, M.; Ertekin-Taner, N.; Ansari, M.; Wilcox, S.L.; Kashino, M.R.; Ma, L.; Younkin, L.H.; et al. Concordant Association of Insulin Degrading Enzyme Gene (IDE) Variants with IDE mRNA, Aß, and Alzheimer's Disease. *PLoS ONE* **2010**, *5*, e8764. [CrossRef] [PubMed]
15. Edland, S.D. Insulin-Degrading Enzyme, Apolipoprotein E, and Alzheimer's Disease. *J. Mol. Neurosci.* **2004**, *23*, 213–218. [CrossRef]
16. Zhang, H.; Liu, D.; Huang, H.; Zhao, Y.; Zhou, H. Characteristics of insulin-degrading enzyme in alzheimer's disease: A meta-analysis. *Curr. Alzheimer Res.* **2018**, *15*, 610–617. [CrossRef]

17. Durham, T.B.; Toth, J.L.; Klimkowski, V.J.; Cao, J.X.; Siesky, A.M.; Alexander-Chacko, J.; Wu, G.Y.; Dixon, J.T.; McGee, J.E.; Wang, Y.; et al. Dual Exosite-binding Inhibitors of Insulin-degrading Enzyme Challenge Its Role as the Primary Mediator of Insulin Clearance in Vivo. *J. Biol. Chem.* **2015**, *290*, 20044–20059. [CrossRef]
18. Maianti, J.P.; McFedries, A.; Foda, Z.H.; Kleiner, R.E.; Du, X.Q.; Leissring, M.A.; Tang, W.-J.; Charron, M.J.; Seeliger, M.A.; Saghatelian, A.; et al. Anti-diabetic activity of insulin-degrading enzyme inhibitors mediated by multiple hormones. *Nature* **2014**, *511*, 94–98. [CrossRef]
19. Guo, Q.; Manolopoulou, M.; Bian, Y.; Schilling, A.B.; Tang, W.-J. Molecular Basis for the Recognition and Cleavages of IGF-II, TGF-α, and Amylin by Human Insulin-Degrading Enzyme. *J. Mol. Biol.* **2010**, *395*, 430–443. [CrossRef]
20. Malito, E.; Ralat, L.A.; Manolopoulou, M.; Tsay, J.L.; Wadlington, N.; Tang, W.-J. Molecular Bases for the Recognition of Short Peptide Substrates and Cysteine-Directed Modifications of Human Insulin-Degrading Enzyme. *Biochemistry* **2008**, *47*, 12822–12834. [CrossRef]
21. Manolopoulou, M.; Guo, Q.; Malito, E.; Schilling, A.B.; Tang, W.-J. Molecular Basis of Catalytic Chamber-assisted Unfolding and Cleavage of Human Insulin by Human Insulin-degrading Enzyme. *J. Biol. Chem.* **2009**, *284*, 14177–14188. [CrossRef]
22. Ralat, L.A.; Guo, Q.; Ren, M.; Funke, T.; Dickey, D.M.; Potter, L.R.; Tang, W.-J. Insulin-degrading Enzyme Modulates the Natriuretic Peptide-mediated Signaling Response. *J. Biol. Chem.* **2011**, *286*, 4670–4679. [CrossRef] [PubMed]
23. Ren, M.; Guo, Q.; Guo, L.; Lenz, M.; Qian, F.; Koenen, R.R.; Xu, H.; Schilling, A.B.; Weber, C.; Ye, R.D.; et al. Polymerization of MIP-1 chemokine (CCL3 and CCL4) and clearance of MIP-1 by insulin-degrading enzyme. *EMBO J.* **2010**, *29*, 3952–3966. [CrossRef] [PubMed]
24. Shen, Y.; Joachimiak, A.; Rosner, M.R.; Tang, W.-J. Structures of human insulin-degrading enzyme reveal a new substrate recognition mechanism. *Nature* **2006**, *443*, 870–874. [CrossRef] [PubMed]
25. Song, E.S.; Rodgers, D.W.; Hersh, L.B. A Monomeric Variant of Insulin Degrading Enzyme (IDE) Loses Its Regulatory Properties. *PLoS ONE* **2010**, *5*, e9719. [CrossRef]
26. Perlman, R.K.; Gehm, B.D.; Kuo, W.L.; Rosner, M.R. Functional analysis of conserved residues in the active site of insulin-degrading enzyme. *J. Biol. Chem.* **1993**, *268*, 21538–21544. [CrossRef]
27. Gehm, B.D.; Kuo, W.L.; Perlman, R.K.; Rosner, M.R. Mutations in a zinc-binding domain of human insulin-degrading enzyme eliminate catalytic activity but not insulin binding. *J. Biol. Chem.* **1993**, *268*, 7943–7948. [CrossRef]
28. Zhang, Z.; Liang, W.G.; Bailey, L.J.; Tan, Y.Z.; Wei, H.; Wang, A.; Farcasanu, M.; A Woods, V.; A McCord, L.; Lee, D.; et al. Ensemble cryoEM elucidates the mechanism of insulin capture and degradation by human insulin degrading enzyme. *eLife* **2018**, *7*, e33572. [CrossRef]
29. Leissring, M.A. Insulin-Degrading Enzyme: Paradoxes and Possibilities. *Cells* **2021**, *10*, 2445. [CrossRef]
30. McCord, L.A.; Liang, W.G.; Dowdell, E.; Kalas, V.; Hoey, R.J.; Koide, A.; Koide, S.; Tang, W.-J. Conformational states and recognition of amyloidogenic peptides of human insulin-degrading enzyme. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 13827–13832. [CrossRef]
31. Abraham, M.J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J.C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. [CrossRef]
32. King, J.V.; Liang, W.G.; Scherpelz, K.P.; Schilling, A.B.; Meredith, S.C.; Tang, W.J. Molecular basis of substrate recognition and degradation by human presequence protease. *Structure* **2014**, *22*, 996–1007. [CrossRef] [PubMed]
33. Song, E.S.; Daily, A.; Fried, M.G.; Juliano, M.A.; Juliano, L.; Hersh, L.B. Mutation of Active Site Residues of Insulin-degrading Enzyme Alters Allosteric Interactions. *J. Biol. Chem.* **2005**, *280*, 17701–17706. [CrossRef] [PubMed]
34. Noinaj, N.; Bhasin, S.K.; Song, E.S.; Scoggin, K.E.; Juliano, M.A.; Juliano, L.; Hersh, L.B.; Rodgers, D.W. Identification of the Allosteric Regulatory Site of Insulysin. *PLoS ONE* **2011**, *6*, e20864. [CrossRef]
35. da Cruz, C.H.B.; Seabra, G. Molecular Dynamics Simulations Reveal a Novel Mechanism for ATP Inhibition of Insulin Degrading Enzyme. *J. Chem. Inf. Model.* **2014**, *54*, 1380–1390. [CrossRef] [PubMed]
36. A Structural View of Biology. Available online: http://www.rcsb.org/ (accessed on 30 May 2020).
37. Im, H.; Manolopoulou, M.; Malito, E.; Shen, Y.; Zhao, J.; Neant-Fery, M.; Sun, C.-Y.; Meredith, S.C.; Sisodia, S.S.; Leissring, M.A.; et al. Structure of Substrate-free Human Insulin-degrading Enzyme (IDE) and Biophysical Analysis of ATP-induced Conformational Switch of IDE. *J. Biol. Chem.* **2007**, *282*, 25453–25463. [CrossRef]
38. Eswar, N.; Webb, B.; Marti-Renom, M.A.; Madhusudhan, M.S.; Eramian, D.; Shen, M.; Pieper, U.; Sali, A. Comparative Protein Structure Modeling Using Modeller. *Curr. Protoc. Bioinform.* **2006**, *15*, 5–6. [CrossRef]
39. Shen, M.-Y.; Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **2006**, *15*, 2507–2524. [CrossRef]
40. Bowie, J.U.; Lüthy, R.; Eisenberg, D. A Method to Identify Protein Sequences That Fold into a Known Three-Dimensional Structure. *Science* **1991**, *253*, 164–170. [CrossRef]
41. Olsson, M.H.M.; Søndergaard, C.R.; Rostkowski, M.; Jensen, J.H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K a Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537. [CrossRef]
42. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, G.N.M.; De Groot, B.L.; Grubmüller, H.; MacKerell, A.D. CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2016**, *14*, 71–73. [CrossRef]
43. Hess, B.; Bekker, H.; Berendsen, H.J.; Fraaije, J.G. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1998**, *18*, 1463–1472. [CrossRef]

44. Miyamoto, S.; Kollman, P.A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952–962. [CrossRef]
45. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N\cdot\log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092. [CrossRef]
46. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [CrossRef]
47. *The PyMOL Molecular Graphics System*; Version 2.0 Schrödinger; LLC: New York, NY, USA.
48. Williams, T.; Kelley, C. Gnuplot 5.0: An Interactive Plotting Program. 2015. Available online: http://www.gnuplot.info/ (accessed on 30 May 2020).
49. Paramo, T.; East, A.; Garzón, D.; Ulmschneider, M.B.; Bond, P.J. Efficient Characterization of Protein Cavities within Molecular Simulation Trajectories: Trj_cavity. *J. Chem. Theory Comput.* **2014**, *10*, 2151–2164. [CrossRef]
50. Song, E.S.; Juliano, M.A.; Juliano, L.; Fried, M.G.; Wagner, S.L.; Hersh, L.B. ATP Effects on Insulin-degrading Enzyme Are Mediated Primarily through Its Triphosphate Moiety. *J. Biol. Chem.* **2004**, *279*, 54216–54220. [CrossRef]

6.3 Studying DYRK1A and new potential PIF targets

As previously described in section 5.3.5, DYRK1A plays a crucial role in DS. It also binds a similar peptide with a strong homology sequence (ARPGTPAL) (Figure 5) with PIF. For this reason, we made the decision to determine whether there is a direct interaction between PIF and DYRK1A and, if so, explore a potential inhibitory impact of PIF using both *in vitro* and *in silico* techniques. We simultaneously used AlphaFold2-multimer to reconstruct the DYRK1A-substrate complex, and since the results were reliable, we also used the same method to construct a PIF-DRYK1A complex to inspect if PIF binds at the same region as the DYRK1A substrate.

In the other section of this chapter, we also describe the use of a Convolutional Attention-based Neural Network for Multi-level Peptide-protein Interaction Prediction (CAMP) tool [146] to investigate new potential protein targets of PIF. By utilizing the set of proteins sequences used in the proteomics experiment and the PIF sequence as inputs, we aimed to calculate a binding score for each pair of PIF-protein complex and compare the CAMP results with the proteomics results. This approach was selected to find any possible protein targets that might be commonly identified by both experimental and AI(Artificial intelligence)-based techniques. It is important to note that if there is an overlap of the results from both approaches, it might indicate the likelihood of a potential target that can be further researched as a protein partner of PIF. The use of CAMP as an additional tool for protein-peptide interaction prediction can be another strategy to enhance the accuracy and comprehensiveness of such studies and may lead to the discovery of novel protein targets.

6.3.1 PIF-DYRK1A AlphaFold2-multimer results

In the primary step of our analysis, we retrieved the DYRK1A complex structure in the PDB that has a substrate with a comparable sequence to PIF (PDB:2WO6) [135]. Our goal was to use AlphaFold2-multimer to regenerate the DYRK1A-substrate complex. We also wanted to verify if the reconstruction made with AlphaFold2-multimer reproduces the same position of the peptide in the right binding site of DYRK1A. We also examined if the same interactions were maintained which describes the accuracy of the models created. To provide a more varied selection of models, specific AlphaFold2-multimer characteristics, such as the dropout option, were applied. We set this parameter with an increased number of cycles, as this consistently leads AlphaFold2-multimer to generate better models than running just a single time. We next carried out the same protocol using the PIF sequence to see if it could also bind to the same binding site where the substrate is on DYRK1A and to confirm once again whether the interactions were conserved for the common AA.

AlphaFold2-multimer consequently predicted five models that were highly comparable to the original DYRK1A-substrate complex. All five models essentially replicated the interactions

between the peptide and DYRK1A found in the original structure. However, when the same procedure is applied on PIF, the created models showed that PIF was located on another region of the protein different from the substrate binding area.

Our results suggest two possible hypotheses. First, DYRK1A may not be the appropriate protein partner for PIF, like IDE. This hypothesis is supported by the AlphaFold2-multimer results, which indicate that PIF may not be located in the same binding region as the substrate and do not maintain the same interactions. Secondly, PIF may bind to DYRK1A, but the binding site may be located in a region that is either adjacent or further away from the substrate binding site. To test the latter hypothesis, additional experimental evidence is required. Hence, experimental analyses to examine these ideas were conducted in cooperation with Pr. Janel's team.

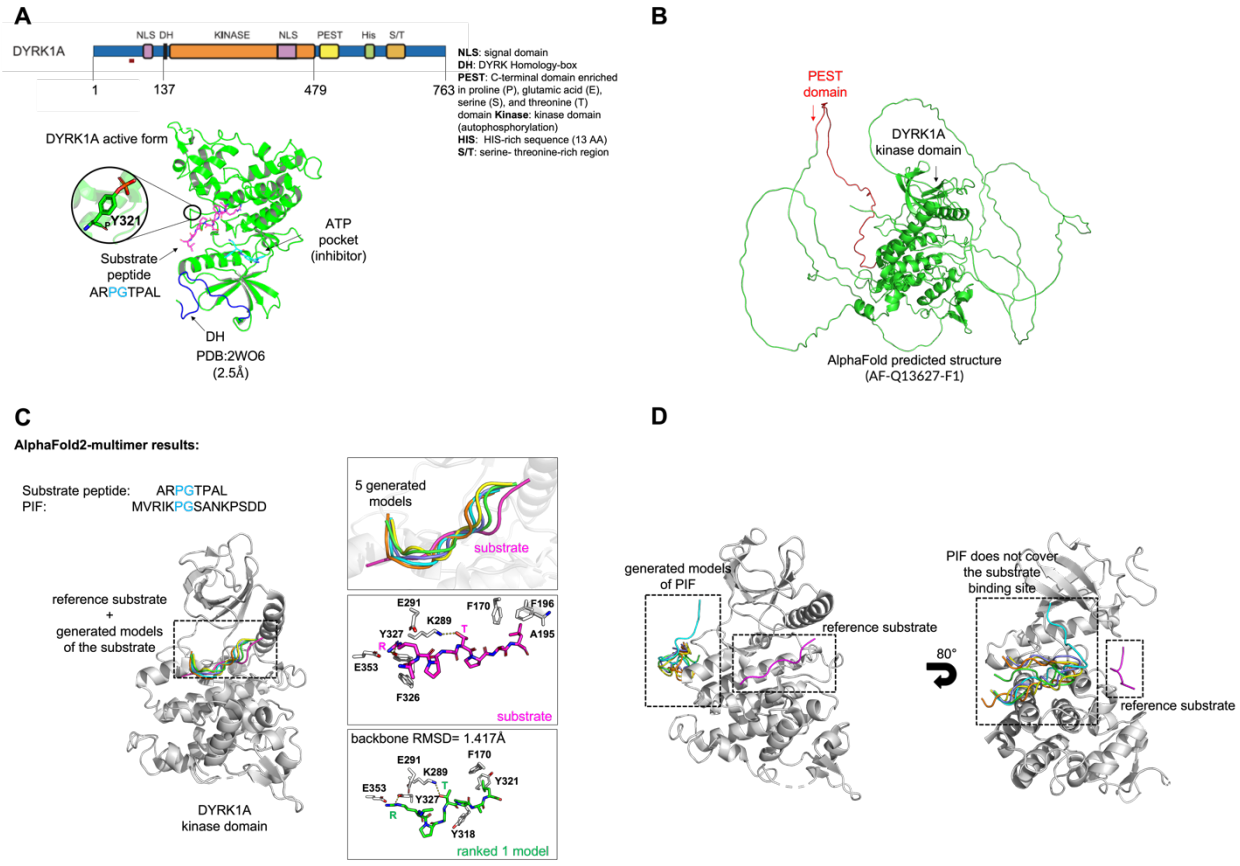


Figure 8. A) Structure of DYRK1A and its domain organization. Autophosphorylation of Y321 in the catalytic domain is critical for full activity of DYRK1A and is highlighted. B) The AlphaFold structure prediction of the full structure of DYRK1A. The PEST domain is represented in red. C) AlphaFold2-multimer results of the reference substrate and DYRK1A complex. D) AlphaFold2-multimer results of PIF-DYRK1A complex.

6.3.2 Co-immunoprecipitation results

The results of the co-immunoprecipitation experiment suggested that DYRK1A may phosphorylate PIF which implies that PIF does bind to DYRK1A. However, this event takes place in a disordered region of the protein, the PEST (a C-terminal domain enriched in proline (P), glutamic acid (E), serine (S), and threonine (T)) domain, which makes this task even more challenging even when using the latest techniques like AlphaFold2-multimer. Disorder regions in proteins are frequently unstructured portions that cannot be modeled, making it difficult to demonstrate a potential interaction. However, these results support earlier findings made with the AlphaFold2-multimer. Therefore, PIF do not bind on the kinase domain of DYRK1A, but it might play another role by interacting with the PEST domain. Unfortunately, these results require further studies with the use of experimental approaches.

6.4 Using a deep-learning framework to identify new PIF partners

With the help of Pr. Janel's lab, MS-based proteomics was used for identifying and quantifying proteins interacting with PIF. In the study, proteins from mice brains along with the addition of biotinylated PIF were analyzed to pull down potential interacting proteins. Then, the resulting protein mixture was subjected to LC/MS/MS analysis to identify and quantify the proteins partners of PIF. For reproducibility purposes, the experiment has been performed twice and several samples have been analyzed (Tables 2 and 3). A couple of factors were looked at in order to find potential PIF-protein interactions. These parameters include the coverage percentage and the number of unique peptides identified for a given protein. The coverage percentage represents the proportion of the protein sequence that is covered by peptides identified through MS. Also, proteins with higher coverage percentage and more unique peptides are generally considered to be more confidently identified. Another parameter that has been considered is the -10LgP or the score which represents the statistical significance of the peptide identification. This score is commonly used as a cutoff to filter out low-confidence identifications. Therefore, higher -10LgP values indicate a more significant peptide identification.

On the other hand, the CAMP tool predicts protein-peptide interactions using a deep learning algorithm. It takes in the AA sequences of both the peptide and the protein and converts them into high-dimensional vectors called dense vectors, using an embedding layer that captures the underlying relationships between the AA. The sequences are then processed by convolutional layers that detect local patterns and features primarily generated by third-party software. The output of these layers is then passed through an attention mechanism that assigns weights to different parts of the input sequence based on their importance for predicting the interaction. This allows the algorithm to focus on the most relevant parts of the protein-peptide complex.

The attention mechanism is essentially a mechanism that allows the algorithm to selectively emphasize important regions of the input. The outputs of the attention mechanism are then transformed using fully connected layers that perform non-linear operations on the input features. Non-linear operations are essentially mathematical operations that enable the algorithm to capture complex and non-linear relationships between the input features. The fully connected layers are trained to map the input sequences to a single output value, which represents the predicted likelihood of interaction between the peptide and protein.

Therefore, for the PIF study, the inputs we used in CAMP were the 796 protein sequences from the proteomics experimental data and the PIF sequence. After obtaining the results from CAMP, we decided to combine them with the results obtained from MS-based proteomics to further investigate potential interactions between PIF and proteins. The CAMP tool scripts were then modified to fit our specific requirements, and the output displayed the prediction score for each protein-peptide complex. It is worth noting that the prediction score ≥ 0.5 indicates that there is an interaction between the given protein-peptide pair, and <0.5 indicates no interaction. This threshold value was selected based on the ROC curve analysis of the model's performance on a validation set [146].

A set of proteins with the highest scores, coverage percentages, number of unique peptides, and prediction scores above 0.5 were identified (Tables 2 and 3). Among them, the Serine-threonine kinase receptor-associated protein [147], Dynein light chain 2 [148], Platelet-activating factor acetylhydrolase IB subunit beta [149], Nucleolin [150], Calcium/calmodulin-dependent protein kinase type II subunit beta [151], Receptor of activated protein C kinase 1 [152], Protein phosphatase 1 regulatory subunit 7 [153], Heterogeneous nuclear ribonucleoprotein A/B [154], and Dual specificity mitogen-activated protein kinase kinase 1 [155] stood out. Interestingly, most of the highly ranked proteins belong to the kinase family. Further analysis revealed that three of these proteins, the dynein protein, the Calcium/calmodulin-dependent protein kinase, and the Heterogeneous nuclear ribonucleoprotein A/B have known bound peptides and available PDB structures [156-165]. These results and the availability of PDB structures of these proteins with known bound peptides can highly facilitate further docking analyses with PIF and comparison studies. Additionally, these proteins have been linked to a variety of signaling pathways, including the Mitogen-Activated Protein Kinase (MAPK), the PhosphoInositide 3-Kinase (PI3K), and the Protein Kinase C (PKC) pathways. In fact, these pathways are directly related to the biological functions of the PIF peptide, which is known to regulate immune responses, promote tolerance, and reduce inflammation. A further network protein analysis can also be performed to gain a better understanding of the relationship and interaction between the identified proteins and their potential roles in cellular pathways related to PIF potential functions.

Interestingly, the Peroxiredoxin-6 (PRDX6) [166] protein which shares the same gene as the PDI protein and has an antioxidant role as well exhibits good results regarding the MS-based proteomics experiment but an average CAMP score of 0.39. These results can be explained by the fact that the PRDX6 protein could have unique characteristics or structural features that are not well-represented in the existing neural network-based model of CAMP. It is also possible

that the CAMP tool has limitations or biases in predicting certain types of proteins and should be explored further. However, this result remains interesting and the PRDX6 could be a potential protein partner of PIF as it is also linked to DS [167].

On the other hand, DYRK1A and IDE exhibited poor CAMP scores of 0.31 and 0.03, respectively. Additionally, DYRK1A has not been identified through the MS-based proteomics experiment. These results compared to the co-immunoprecipitation and AlphaFold2-multimer results might explain that the binding site of PIF may not reside in the kinase domain of DYRK1A. As for IDE, these experimental results also complement the results obtained with docking and co-immunoprecipitation and prove that IDE might not be the correct protein partner of PIF.

This study combined an AI-based computational tool with an experimental technique to evaluate the biological roles of a set of proteins and their connection to PIF. The results obtained from this analysis were consistent with what is currently known about the proteins biological functions and how they can be possibly interacting with PIF. These findings may lead to new insights into the structural mechanisms between PIF and these proteins with the aim of underlaying their biological functions in various physiological processes.

Accession	Name	Score (-10LgP)	Coverage % (Sample 1)	Coverage % (Sample 2)	Number of unique peptides	CAMP prediction score
Q9Z1Z2	Serine-threonine kinase receptor-associated protein	241.6	31.71	32.29	10	0.89
Q9D0M5	Dynein light chain 2	212.76	64.04	14.61	4	0.88
P63005	Platelet-activating factor acetylhydrolase IB subunit beta	244	38.78	3.66	12	0.85
P09405	Nucleolin	289.57	27.72	26.03	18	0.79
P28652	Calcium/calmodulin-dependent protein kinase type II subunit beta	235.63	23.8	10.15	4	0.75
P68040	Receptor of activated protein C kinase 1	165.29	18.93	8.83	5	0.68
Q3UM45	Protein phosphatase 1 regulatory subunit 7	183.66	20.22	2.49	6	0.65
Q99020	Heterogeneous nuclear ribonucleoprotein A/B	171.4	19.65	12.98	3	0.51
P31938	Dual specificity mitogen-activated protein kinase kinase 1	187.16	23.16	8.4	9	0.5
O08709	**Peroxiredoxin-6**	220.76	57.14	32.59	11	0.39
Q61214	**Dual specificity tyrosine-phosphorylation-regulated kinase 1A**	x	x	x	x	0.31
Q9JHR7	**Insulin-degrading Enzyme**	61.57	1.18	0	1	0.03

Table 2. Summary results of the first MS-based proteomics experiment with biotin-PIF in two samples and bound proteins, along with CAMP results. Identified bound proteins are ordered with the CAMP scores. PRDX6, IDE and DYRK1A proteins are added for score comparison.

Accession	Name	Score (-10LgP)	Coverage % (Sample 1)	Coverage % (Sample 2)	Coverage % (Sample 3)	Coverage % (Sample 4)	Number of unique peptides	CAMP prediction score
Q9Z1Z2	Serine-threonine kinase receptor-associated protein	203.28	20.57	10.29	22.57	0	6	0.89
Q9D0M5	Dynein light chain 2	244.65	51.69	26.97	65.17	26.97	6	0.88
P63005	Platelet-activating factor acetylhydrolase IB subunit beta	253.94	3.66	0	33.66	36.83	14	0.85
P09405	Nucleolin	230.64	1.98	0	16.83	19.09	12	0.79
P28652	Calcium/calmodulin-dependent protein kinase type II subunit beta	251.14	7.75	7.75	29.34	18.82	5	0.75
P68040	Receptor of activated protein C kinase 1	166.295	5.05	0	18.3	0	5	0.68
Q3UM45	Protein phosphatase 1 regulatory subunit 7	188.92	0	5.54	29.64	5.82	9	0.65
Q99020	Heterogeneous nuclear ribonucleoprotein A/B	233.81	11.23	12.98	35.79	24.91	8	0.51
P31938	Dual specificity mitogen-activated protein kinase kinase 1	171.35	15.01	11.45	20.36	8.14	6	0.5
O08709	**Peroxiredoxin-6**	303.86	58.48	59.82	74.11	46.88	19	0.39
Q61214	**Dual specificity tyrosine-phosphorylation-regulated kinase 1A**	x	x	x	x	x	x	0.31
P68040	**Insulin-degrading Enzyme**	x	x	x	x	x	x	0.03

Table 3. Summary results of the second MS-based proteomics experiment with biotin-PIF in two samples and bound proteins, along with CAMP results. Identified bound proteins are ordered with the CAMP scores. PRDX6, IDE and DYRK1A proteins are added for score comparison.

Chapter 7

A large-scale genomic study of PIF-encoding gene

7.1 PIF-encoding gene search

The search of PIF-encoding gene has been one of the most challenging task. In fact, a study by Mueller M. et al. [71] also failed to identify the coding gene of the PIF peptide. Our work focused on testing different tools in the aim of identifying a region that could be labeled as PIF coding location. We also specified the limitations of our protocol as well as the lack of the available information to characterize this type of peptide.

7.1.1 Basic Local Alignment Search Tool (BLAST) search

The first step to undertake when looking to identify an unknown AA sequence is to perform a BLAST search [168]. However, the BLAST search only suggests a 100% homology with the malarial CSP for the 2-12AA (VRIKPGSANKP) PIF region. Moreover, the region coincides with a T-cell epitope of the Th3R domain of the malarial circumsporozoite protein [169]. Plus, the 3-11AA PIF sequence (RIKPGSANK) has been described as an HLA A03-restricted class I epitope that is part of the CSP long synthetic peptide vaccine candidate that elicited potent T cell responses in malaria naïve adults [170]. Despite the fact that these findings are « interesting », it is difficult to envision a placental peptide being classified as part of a malarial protein area. Therefore, due to the lack of the BLAST tool sensitivity for short peptides and considering these results as aberrant, we decided to move forwards with more specified tools and designed protocols to identify PIF-encoding gene.

7.1.2 Space-Efficient Spliced Alignment (SPALN) search

Not being able to find a matching sequence to PIF due its unique and short sequence, we decided to perform a large-scale genomic mapping of PIF onto the human genome. To do so, PIF sequence was fragmented into different AA lengths. The resulting sequences were fragmented in a way where we obtained different combinations of continuous sequences of PIF going from 3 to 15AA. SPALN [171] is a tool that is able to map and align a set of cDNA (complementary DeoxyriboNucleic Acid) and protein sequences onto a whole genome. Thus, we used SPALN and the latest version of the human genome (GRCh38 (Genome Reference

Consortium Human Build 38 Organism: Homo sapiens (human))/hg38 (human genome 38)) [172] to perform the genome search. 171 transcript segments homologous to the PIF sequence were identified. Then, among these transcripts, we filtered our data so that only those that were distinct and had PIF-unique sequences with a maximum sequence length of 6AA were retained.

We identified 21, 9 and 2 transcript segments corresponding to 6AA, 7AA, and 8AA, respectively that perfectly match the PIF sequence. As a result, we focused on the transcripts with the largest number of AA matches. The first transcript (PGSANKPS) is situated in an mRNA (messenger RiboNucleic Acid) region on an identified area of chromosome 8 (18q11.22). This chromosomal position lacked relevant information. We attempted to find this transcript on the mouse genome (GRCm38 (GRCm38: Genome Reference Consortium Mouse Build 38 Organism: Mus musculus (house mouse)/mm10 (mus musculus 10)) [173], but we were unsuccessful. The second transcript, however, (VRIKPGSA), corresponds to a non-coding portion of the NRXN3 (Neurexin-3) gene located on chromosome 14 (14q24.3). Interestingly, we were also able to identify this transcript in the mouse and gorilla [174] genomes which might suggest the conservation of this region.

The identification of the native PIF peptide in previous studies included peptides with lengths ranging from 9 to 15AA, as stated in section 3.2 (MVRIKPGSA-MVRIKPGSANKPSDD). We were successful in identifying a crucial physiologically active region of the PIF peptide despite the fact that the discovered transcript lacks the M residue. These findings led to a number of explanations as to why it was challenging to find PIF in databases of known expressed peptides or proteins, as well as to numerous hypotheses regarding the identification and production of the peptide. Additional research is currently being conducted to verify these results.

7.1.3 The genetic origin of PIF

Our results were successful in identifying a biologically active region of PIF in the human genome. However, this determined sequence corresponds to a non-coding region of chromosome 14 (14q24.3). The majority of the mammalian genome is pervasively transcribed, as is well known. Plus, the majority of these transcripts are categorized as non-coding RNAs (ncRNAs) and over two-thirds of human genes are practically only translated into ncRNAs, which do not encode any known proteins [175]. Long non-coding RNAs (lncRNAs) and short non-coding RNAs (sncRNAs) are two prevalent categories for ncRNAs. LncRNAs are distinguished from sncRNAs by having a length of >200 nt (nucleotide). An ncRNA that codes for peptides is often defined as a transcript that was initially annotated as non-coding but was later found to include a short open reading frame (smORF) that encodes a peptide with less than 100AA [176, 177]. While peptides from coding areas are translated from well-known short coding sequences, smORFs, the biogenesis of non-coding peptides appears to be different from that of other peptides. Plus, in terms of biology, it is currently unclear how non-coding peptides vary from traditional coding region-derived peptides in terms of their use, biogenesis, or structural characteristics.

Currently, ribosome profiling and MS are the two main high throughput whole-genome laboratory-based techniques that are being widely used to empirically confirm smORFs. However, these techniques also have their limitations like biased results or the high cost of the performed procedure making the discovery of new smORFs a challenging task in the field of genomics. On the other hand, some studies have also proven that certain lncRNAs had smORFs (< length 300 nt) that might encode a short peptide with essential biological functions [178]. The existence of functional short peptides encoded by lncRNAs raises the possibility that these lncRNAs might serve two functions, interacting with both RNA and peptides, and should be classed as bifunctional RNAs. In the case of PIF, the two hypothesis could be valid. Thus, PIF can be labeled as an ncRNA-encoded smORF peptide or a portion of an lncRNA-encoded smORF peptide. A recent paper of Yang, L., et al. 2013 [179] looks at performing single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Among their results, they determined a novel lncRNA that matches the PIF non-coding region on chromosome 14 which might also support our hypothesis. Overall, these findings explain the absence of PIF in multiple databases of known expressed proteins and peptides.

We do not know exactly the reason behind why only a portion of PIF has been determined in the human genome. We might speculate that the lack of the genome annotation or the existence of PIF variants may be the reason. Therefore, a focused analysis should be conducted to take into account these limitations.

7.1.4 Differential expression of PIF during the preimplantation phase

Human embryonic genome activation starts between the two and eight-cell stages [154]. PIF is a peptide that trophoblast cells secrete in the early stages of embryonic development before the formation of the placenta [69, 72, 90]. PIF expression in human embryos starts at the 4-cell stage, increases by the morula stage, and continues through the first trimester [90]. Additionally, PIF expression in the blastocyst was found to be an early predictor of the viability of multiple pregnancies. Therefore, we decided to identify critical genes to embryonic implantation that might coincide with PIF expression in early, middle and late stages of human blastocytes. The aim of this study is to determine which genes and gene pathways are differentially expressed during preimplantation and determine transcript-level differential expressions of PIF. This future study can also be applied on other different stages of the preimplantation period. For further validation and analysis, this protocol can also be used on RNA-seq or single-cell data of hyperplastic endometriotic lesions and advanced uterine or prostate cancer since PIF has been detected by immunohistochemistry under those conditions.

Chapter 8

Conclusion

This thesis focused on the study and contributed significantly to the understanding of protein-protein and protein-peptide interactions, particularly the interactions between the Spike protein RBD domain and the ACE2 receptor, and the interactions of the PIF peptide with its protein targets.

The development of a bioinformatics protocol including molecular modeling, MD simulations, free energy calculations and druggable pocket tracking approaches has enabled the identification of three potential druggable pockets in the RBD-ACE2 binding surface. These results offer promising targets for the development of therapeutic molecules that can inhibit the viral reaction caused by the Spike protein. Additionally, a thorough study focusing on the COVID-19 variants of concern has provided valuable insights into the molecular interactions of the mutated RBD and its receptor ACE2. It also highlighted several hot spots that can be targeted with small molecules or antibodies in order to block the Spike protein with ACE2.

Protein-peptide interactions are known to be challenging molecular systems. Our work focused on the PIF peptide and used several bioinformatics tools to identify its potential protein partners. Using different techniques such as molecular modeling, MD simulations and deep learning, we were able to investigate potential PIF targets such as IDE and DYRK1A, but also identify new potential partners. While limitations were encountered in some of these interactions and some targets have been overruled with experimental validation, these results have provided crucial information on the biological functions of these proteins and offered new potential targets to consider. Additionally, the genomic origin of PIF has also been highlighted, providing promising results for future research to better understand the regulation of PIF during pregnancy at the transcriptional and post-transcriptional levels.

Overall, we are confident that these results have the potential to inform and guide future drug discovery and development research. Also, we hope that this work will inspire and inform further investigations into protein-protein and protein-peptide interactions, ultimately leading to the development of more effective and targeted therapies for COVID-19 and DS.

Perspectives

There is substantial scope for further work to complement and support the findings in this thesis. The exploration of protein-protein interactions, specifically the Spike protein RBD and ACE2 complex, has uncovered three druggable pockets on the binding surface of the RBD-ACE2 complex. Another study has revealed regions that are susceptible to the effects of mutations on RBD-ACE2 affinity. Consequently, several steps can be undertaken to advance this research. One approach involves docking multiple inhibitors onto the identified druggable pockets to identify promising candidates capable of efficiently inhibiting the RBD-ACE2 complex. Also, identifying common druggable pockets for different RBD systems (Wild-type, Alpha, Beta, Gamma, Delta, and Omicron) and evaluating the possibility of small compounds to target these pockets and disrupt the RBD-ACE2 complex might be an additional topic of research. To complement these results, collaboration with experimental teams would be crucial to validate the results obtained and identify effective inhibitors.

In the case of the PIF peptide, additional investigations can be interesting to complete and validate the findings of this thesis. It can be extremely interesting to perform a thorough analysis of PIF protein partners using additional different methodologies. For instance, after employing the CAMP tool to obtain protein candidates, *in vitro* and *in silico* investigations testing the direct interactions between PIF and these proteins may offer insightful information. Cross docking, where PIF is docked against multiple receptor structures obtained from the PDB, could also enable a comprehensive exploration of PIF binding modes and affinities by considering a diverse range of receptor conformations. An ambitious yet promising approach would involve developing a machine learning model utilizing a large dataset of experimentally validated protein-peptide complexes to predict the specific amino acids or protein surfaces to which PIF could bind. Furthermore, completing the genomics analysis would significantly contribute to determining the differentially expressed genes and gene pathways during preimplantation, providing a better understanding on transcript-level differential expressions of PIF. This comprehensive investigation would enhance our understanding of the mode of action of PIF and provide greater control over its functions.

References

[1] Petsalaki, E.; Russell, R. Peptide-Mediated Interactions in Biological Systems: New Discoveries and Applications. *Curr. Opin. in Biotechnol.* **2008**, *19*, 344–350.

[2] Dutta, S.; Chen, T. S.; Keating, A. E. Peptide Ligands for Pro-Survival Protein Bfl-1 from Computationally Guided Library Screening. *ACS Chem. Biol.* **2013**, *8* (4), 778–788.

[3] Rebek, J. Introduction to the Molecular Recognition and Self-Assembly Special Feature. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106* (26), 10423–10424.

[4] van Dun, S.; Ottmann, C.; Milroy, L.-G.; Brunsveld, L. Supramolecular Chemistry Targeting Proteins. *J. Am. Chem. Soc.* **2017**, *139* (40), 13960–13968.

[5] Jin, X.; Zhu, L.; Xue, B.; Zhu, X.; Yan, D. Supramolecular Nanoscale Drug-Delivery System with Ordered Structure. *Natl. Sci.* **2019**, *6* (6), 1128–1137.

[6] Peng, X.; Wang, J.; Peng, W.; Wu, F.-X.; Pan, Y. Protein–Protein Interactions: Detection, Reliability Assessment and Applications. *Brief Bioinform* **2016**, bbw066.

[7] Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. Unraveling the Importance of Protein−Protein Interaction: Application of a Computational Alanine-Scanning Mutagenesis to the Study of the IgG1 Streptococcal Protein G (C2 Fragment) Complex. *J. Phys. Chem. B* **2006**, *110* (22), 10962–10969.

[8] Lu, H.; Zhou, Q.; He, J.; Jiang, Z.; Peng, C.; Tong, R.; Shi, J. Recent Advances in the Development of Protein–Protein Interactions Modulators: Mechanisms and Clinical Trials. *Sig Transduct Target Ther* **2020**, *5* (1), 213.

[9] Thomas, S.; Quinn, B. A.; Das, S. K.; Dash, R.; Emdad, L.; Dasgupta, S.; Wang, X.-Y.; Dent, P.; Reed, J. C.; Pellecchia, M.; Sarkar, D.; Fisher, P. B. Targeting the Bcl-2 Family for Cancer Therapy. *Expert Opin Ther Targets* **2013**, *17* (1), 61–75.

[10] Sedov, I. A.; Zuev, Y. F. Recent Advances in Protein–Protein Interactions. *IJMS* **2023**, *24* (2), 1282.

[11] Durham, J.; Zhang, J.; Humphreys, I. R.; Pei, J.; Cong, Q. Recent Advances in Predicting and Modeling Protein–Protein Interactions. *TIBS* **2023**, *48* (6), 527–538.

[12] Träger, T.; Kastritis, P. L. Cracking the Code of Cellular Protein–Protein Interactions: Alphafold and whole-cell Crosslinking to the Rescue. *Mol. Syst. Biol.* **2023**, *19* (4), e11587.

[13] Henninot, A.; Collins, J. C.; Nuss, J. M. The Current State of Peptide Drug Discovery: Back to the Future? *J. Med. Chem.* **2018**, *61* (4), 1382–1414.

[14] Craik, D. J.; Fairlie, D. P.; Liras, S.; Price, D. The Future of Peptide-Based Drugs: **Peptides in Drug Development**. *Chem. Biol. Drug Des.* **2013**, *81* (1), 136–147.

[15] Wang, L.; Wang, N.; Zhang, W.; Cheng, X.; Yan, Z.; Shao, G.; Wang, X.; Wang, R.; Fu, C. Therapeutic Peptides: Current Applications and Future Directions. *Sig. Transduct. Target Ther*. **2022**, *7* (1), 48.

[16] Fosgerau, K.; Hoffmann, T. Peptide Therapeutics: Current Status and Future Directions. *Drug Discov. Today* **2015**, *20* (1), 122–128.

[17] Giordano, C.; Marchiò, M.; Timofeeva, E.; Biagini, G. Neuroactive Peptides as Putative Mediators of Antiepileptic Ketogenic Diets. *Front. Neurol.* **2014**, *5*.

[18] Davda, J.; Declerck, P.; Hu-Lieskovan, S.; Hickling, T. P.; Jacobs, I. A.; Chou, J.; Salek-Ardakani, S.; Kraynov, E. Immunogenicity of Immunomodulatory, Antibody-Based, Oncology Therapeutics. *J. immuno. cancer* **2019**, *7* (1), 105.

[19] Waldmann H. Human monoclonal antibodies: the residual challenge of antibody immunogenicity. *Methods Mol. Biol*. **2014**, 1060, 1-8.

[20] Petta, I.; Lievens, S.; Libert, C.; Tavernier, J.; De Bosscher, K. Modulation of Protein–Protein Interactions for the Development of Novel Therapeutics. *Mol. Ther.* **2016**, *24* (4), 707–718.

[21] Abrigo, N. A.; Dods, K. K.; Makovsky, C. A.; Lohan, S.; Mitra, K.; Newcomb, K. M.; Le, A.; Hartman, M. C. T. Development of a Cyclic, Cell Penetrating Peptide Compatible with In Vitro Selection Strategies. *ACS Chem. Biol.* **2023**, *18* (4), 746–755.

[22] Meng, X.-Y.; Zhang, H.-X.; Mezei, M.; Cui, M. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *CAD* **2011**, *7* (2), 146–157.

[23] Agrawal, P.; Singh, H.; Srivastava, H. K.; Singh, S.; Kishore, G.; Raghava, G. P. S. Benchmarking of Different Molecular Docking Methods for Protein-Peptide Docking. *BMC Bioinformatics* **2019**, *19* (S13), 426.

[24] Taylor, R. D. A review of protein-small molecule docking methods. *J. Comput. Aided Mol. Des.* **2002**, 19, 151–166.

[25] Lensink, M. F.; Velankar, S.; Wodak, S. J. Modeling Protein-Protein and Protein-Peptide Complexes: CAPRI 6th Edition: Modeling Protein-Protein and Protein-Peptide Complexes. *Proteins* **2017**, *85* (3), 359–377.

[26] Ciemny, M.; Kurcinski, M.; Kamel, K.; Kolinski, A.; Alam, N.; Schueler-Furman, O.; Kmiecik, S. Protein–Peptide Docking: Opportunities and Challenges. *Drug Discov..Today* **2018**, *23* (8), 1530–1537.

[27] Rentzsch, R.; Renard, B. Y. Docking Small Peptides Remains a Great Challenge: An Assessment Using AutoDock Vina. *Brief. in Bioinformatics* **2015**, *16* (6), 1045–1056.

[28] Webb, B.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *CP Protein Science* **2016**, *86* (1).

[29] Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A., *et al*. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589.

[30] Raveh, B.; London, N.; Zimmerman, L.; Schueler-Furman, O. Rosetta FlexPepDock Ab-Initio: Simultaneous Folding, Docking and Refinement of Peptides onto Their Receptors. *PLoS ONE* **2011**, *6* (4), e18934.

[31] Kurcinski, M.; Jamroz, M.; Blaszczyk, M.; Kolinski, A.; Kmiecik, S. CABS-Dock Web Server for the Flexible Docking of Peptides to Proteins without Prior Knowledge of the Binding Site. *Nucleic Acids Res.* **2015**, *43* (W1), W419–W424.

[32] Xu, X.; Zou, X. Predicting Protein–Peptide Complex Structures by Accounting for Peptide Flexibility and the Physicochemical Environment. *J. Chem. Inf. Model.* **2022**, *62* (1), 27–39.

[33] Zhang, Y.; Sanner, M. F. AutoDock CrankPep: Combining Folding and Docking to Predict Protein–Peptide Complexes. *Bioinformatics* **2019**, *35* (24), 5121–5127.

[34] Zhou, P.; Jin, B.; Li, H.; Huang, S.-Y. HPEPDOCK: A Web Server for Blind Peptide–Protein Docking Based on a Hierarchical Algorithm. *Nucleic Acids Res.* **2018**, *46* (W1), W443–W450.

[35] Lee, H.; Heo, L.; Lee, M. S.; Seok, C. GalaxyPepDock: A Protein–Peptide Docking Tool Based on Interaction Similarity and Energy Optimization. *Nucleic Acids Res.* **2015**, *43* (W1), W431–W435.

[36] Obarska-Kosinska, A.; Iacoangeli, A.; Lepore, R.; Tramontano, A. PepComposer: Computational Design of Peptides Binding to a given Protein Surface. *Nucleic Acids Res.* **2021**.

[37] Dominguez, C.; Boelens, R.; Bonvin, A. M. J. J. HADDOCK: A Protein−Protein Docking

Approach Based on Biochemical or Biophysical Information. *J. Am. Chem. Soc.* **2003**, 125 (7), 1731–1737.

[38] Koes, D. R.; Baumgartner, M. P.; Camacho, C. J. Lessons Learned in Empirical Scoring with Smina from the CSAR 2011 Benchmarking Exercise. *J. Chem. Inf. Model.* **2013**, *53* (8), 1893–1904.

[39] Evans, R.; O'Neill, M.; Pritzel, A.; Antropova, N.; Senior, A.; Green, T.; Žídek, A.; Bates, R.; Blackwell, S.; Yim, J.; *et al*. Protein Complex Prediction with AlphaFold-Multimer; preprint; *Bioinformatics*, **2021**.

[40] Hauser, A. S.; Windshügel, B. LEADS-PEP: A Benchmark Data Set for Assessment of Peptide Docking Performance. *J. Chem. Inf. Model.* **2016**, *56* (1), 188–200.

[41] De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* **2016**, *59* (9), 4035–4061.

[42] Guterres, H.; Im, W. Improving Protein-Ligand Docking Results with High-Throughput Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2020**, *60* (4), 2189–2198.

[43] Govind Kumar, V.; Polasa, A.; Agrawal, S.; Kumar, T. K. S.; Moradi, M. Binding Affinity Estimation from Restrained Umbrella Sampling Simulations. *Nat Comput Sci* **2022**, *3* (1), 59–70.

[44] Karaca, E.; Bonvin, A. M. J. J. Advances in Integrative Modeling of Biomolecular Complexes. *Methods* **2013**, *59* (3), 372–381.

[45] Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin. on Drug Discov.*. **2015**, *10* (5), 449–461.

[46] Zoete, V.; Michielin, O. Comparison between Computational Alanine Scanning and Per-Residue Binding Free Energy Decomposition for Protein-Protein Association Using MM-GBSA: Application to the TCR-p-MHC Complex. *Proteins* **2007**, *67* (4), 1026–1047.

[47] Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J., *et al*. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015**, *137* (7), 2695–2703.

[48] Senior, A. W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Žídek, A.; Nelson, A. W. R.; Bridgland, A., *et al*. Improved Protein Structure Prediction Using Potentials from Deep Learning. *Nature* **2020**, *577* (7792), 706–710.

[49] Tejero, R.; Huang, Y. J.; Ramelot, T. A.; Montelione, G. T. AlphaFold Models of Small Proteins Rival the Accuracy of Solution NMR Structures. *Front. Mol. Biosci.* **2022**, *9*, 877000.

[50] Lei, Y.; Li, S.; Liu, Z.; Wan, F.; Tian, T.; Li, S.; Zhao, D.; Zeng, J. A Deep-Learning Framework for Multi-Level Peptide–Protein Interaction Prediction. *Nat. Commun.* **2021**, *12* (1), 5465.

[51] Ghoula, M.; Naceri, S.; Sitruk, S.; Flatters, D.; Moroy, G.; Camproux, A. C. Identifying Promising Druggable Binding Sites and Their Flexibility to Target the Receptor-Binding Domain of SARS-CoV-2 Spike Protein. *CSBJ* **2023**, *21*, 2339–2351.

[52] Ghoula M.; Deyawe Kongmeneck D.; Camproux A-C.; Moroy G. A comparative structural study of the mutations in SARS-CoV-2 RBD variants of concern on the interaction with ACE2 protein. *J. Phys*. Chem. B, accepted for publication, **(2023)**.

[53] Li, J.; Lai, S.; Gao, G. F.; Shi, W. The Emergence, Genomic Diversity and Global Spread of SARS-CoV-2. *Nature* **2021**, *600* (7889), 408–418.

[54] Lubin, JH.; Zardecki, C.; Dolan, EM.; *et al*. Evolution of the SARS-CoV-2 proteome in three dimensions (3D) during the first 6 months of the COVID-19 pandemic. *Proteins* **2022**, 90 (5), 1054-1080.

[55] 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* (1), 235-242.

[56] Jackson, C. B.; Farzan, M.; Chen, B.; Choe, H. Mechanisms of SARS-CoV-2 Entry into Cells. *Nat Rev Mol Cell Biol* **2022**, *23* (1), 3–20.

[57] Hu, Q.; Xiong, Y.; Zhu, G.; Zhang, Y.; Zhang, Y.; Huang, P.; Ge, G. The SARS-CoV-2 Main Protease (M^{pro}): Structure, Function, and Emerging Therapies for COVID-19. *MedComm* **2022**, *3* (3).

[58] Osipiuk, J.; Azizi, S.-A.; Dvorkin, S.; Endres, M.; Jedrzejczak, R.; Jones, K. A.; Kang, S.; Kathayat, R. S.; Kim, Y.; Lisnyak, V. G.; *et al*. Structure of Papain-like Protease from SARS-CoV-2 and Its Complexes with Non-Covalent Inhibitors. *Nat Commun* **2021**, *12* (1), 743.

[59] Errico, J. M.; Zhao, H.; Chen, R. E.; Liu, Z.; Case, J. B.; Ma, M.; Schmitz, A. J.; Rau, M. J.; Fitzpatrick, J. A. J.; Shi, P.-Y.; *et al*. Structural Mechanism of SARS-CoV-2 Neutralization by Two Murine Antibodies Targeting the RBD. *Cell Reports* **2021**, *37* (4), 109881.

[60] Narayanan, A.; Narwal, M.; Majowicz, S. A.; Varricchio, C.; Toner, S. A.; Ballatore, C.; Brancale, A.; Murakami, K. S.; Jose, J. Identification of SARS-CoV-2 Inhibitors Targeting Mpro and PLpro Using in-Cell-Protease Assay. *Commun Biol* **2022**, *5* (1), 169.

[61] Cox, M.; Peacock, T. P.; Harvey, W. T.; Hughes, J.; Wright, D. W.; COVID-19 Genomics UK (COG-UK) Consortium; Willett, B. J.; Thomson, E.; Gupta, R. K.; Peacock, S. J.; *et al*. SARS-CoV-2 Variant Evasion of Monoclonal Antibodies Based on in Vitro Studies. *Nat Rev Microbiol* **2023**, *21* (2), 112–124.

[62] Joshi, G.; Borah, P.; Thakur, S.; Sharma, P.; Mayank; Poduri, R. Exploring the COVID-19 Vaccine Candidates against SARS-CoV-2 and Its Variants: Where Do We Stand and Where Do We Go? *Human Vaccines & Immunotherapeutics* **2021**, *17* (12), 4714–4740.

[63] Kokic, G.; Hillen, H. S.; Tegunov, D.; Dienemann, C.; Seitz, F.; Schmitzova, J.; Farnung, L.; Siewert, A.; Höbartner, C.; Cramer, P. Mechanism of SARS-CoV-2 Polymerase Stalling by Remdesivir. *Nat Commun* **2021**, *12* (1), 279.

[64] Choudhary, M. C.; Chew, K. W.; Deo, R.; Flynn, J. P.; Regan, J.; Crain, C. R.; Moser, C.; Hughes, M. D.; Ritz, J.; Ribeiro, R. M.; *et al*. Emergence of SARS-CoV-2 Escape Mutations during Bamlanivimab Therapy in a Phase II Randomized Clinical Trial. *Nat Microbiol* **2022**, *7* (11), 1906–1917.

[65] Yoshida, J.; Shiraishi, K.; Tamura, T.; Otani, K.; Kikuchi, T.; Mataga, A.; Ueno, T.; Tanaka, M. Casirivimab-Imdevimab Neutralizing SARS-CoV-2: Post-Infusion Clinical Events and Their Risk Factors. *J Pharm Health Care Sci* **2022**, *8* (1), 1.

[66] Carabelli, A. M.; Peacock, T. P.; Thorne, L. G.; Harvey, W. T.; Hughes, J.; COVID-19 Genomics UK Consortium; De Silva, T. I.; Peacock, S. J.; Barclay, W. S.; De Silva, T. I.; Towers, G. J.; Robertson, D. L. SARS-CoV-2 Variant Biology: Immune Escape, Transmission and Fitness. *Nat Rev Microbiol* **2023**.

[67] Zare, F.; Seifati, S. M.; Dehghan-Manshadi, M.; Fesahat, F. Preimplantation Factor (PIF): A Peptide with Various Functions. *JBRA Assist. Reprod.* **2020**.

[68] Ramu, S.; Stamatkin, C.; Timms, L.; Ruble, M.; Roussev, R. G.; Barnea, E. R. PreImplantation Factor (PIF) Detection in Maternal Circulation in Early Pregnancy Correlates with Live Birth (Bovine Model). *Reprod. Biol. Endocrinol.* **2013**, *11*, 105.

[69] Barnea, E. R.; Kirk, D.; Paidas, M. J. PreImplantation Factor (PIF) Promoting Role in Embryo Implantation: Increases Endometrial Integrin-A2β3, Amphiregulin and Epiregulin While Reducing Betacellulin Expression via MAPK in Decidua. *Reprod. Biol. Endocrinol.* **2012**, *10* (1), 50.

[70] Weiss, L.; Or, R.; Jones, R. C.; Amunugama, R.; JeBailey, L.; Ramu, S.; Bernstein, S. A.; Yekhtin, Z.; Almogi-Hazan, O.; Shainer, R.; *et al*. Preimplantation Factor (PIF*) Reverses Neuroinflammation While Promoting Neural Repair in EAE Model. *J. Neurol. Sci.* **2012**, *312* (1–2), 146–157.

[71] Mueller, M.; Zhou, J.; Yang, L.; Gao, Y.; Wu, F.; Schoeberlein, A.; Surbek, D.; Barnea, E. R.; Paidas, M.; Huang, Y. PreImplantation Factor Promotes Neuroprotection by Targeting MicroRNA Let-7. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111* (38), 13882–13887.

[72] Barnea, E. R.; Lubman, D. M.; Liu, Y.-H.; Absalon-Medina, V.; Hayrabedyan, S.; Todorova, K.; Gilbert, R. O.; Guingab, J.; Barder, T. J. Insight into PreImplantation Factor (PIF*) Mechanism for Embryo Protection and Development: Target Oxidative Stress and Protein Misfolding (PDI and HSP) through Essential RIPK Binding Site. *PLoS ONE* **2014**, *9* (7), e100263.

[73] Paidas, M. J.; Krikun, G.; Huang, S. J.; Jones, R.; Romano, M.; Annunziato, J.; Barnea, E. R. A Genomic and Proteomic Investigation of the Impact of Preimplantation Factor on Human Decidual Cells. *Am. J. Obstet. Gynecol.* **2010**, *202* (5), 459.e1-459.e8.

[74] Raspollini, M. R.; Montagnani, I.; Cirri, P.; Baroni, G.; Cimadamore, A.; Scarpelli, M.; Cheng, L.; Lopez-Beltran, A.; Montironi, R.; Barnea, E. R. PreImplantation Factor Immunohistochemical Expression Correlates with Prostate Cancer Aggressiveness. *Int. J. Biol. Markers* **2020**, *35* (2), 82–90.

[75] Altwerger G, Buza N, Hui P, et al. Preimplantation factor and its expression in aggressive uterine cancer. *Reprod. Sci.* **2018**, *25*, 285A.

[76] Moindjie, H.; Santos, E. D.; Gouesse, R.-J.; Swierkowski-Blanchard, N.; Serazin, V.; Barnea, E. R.; Vialard, F.; Dieudonné, M.-N. Preimplantation Factor Is an Anti-Apoptotic Effector in Human Trophoblasts Involving P53 Signaling Pathway. *Cell Death Dis.* **2016**, *7* (12), e2504–e2504.

[77] Chen, YC.; Rivera, J.; Fitzgerald, M.; Hausding, C.; Ying, YL.; Wang, X.; Todorova, K.; Hayrabedyan, S.; Barnea, ER.; Peter, K. PreImplantation factor prevents atherosclerosis via its immunomodulatory effects without affecting serum lipids. *Thromb. Haemost*. **2016**, *115* (5), 1010-1024.

[78] Di Simone, N.; Di Nicuolo, F.; Marana, R.; Castellani, R.; Ria, F.; Veglia, M.; Scambia, G.; Surbek, D.; Barnea, E.; Mueller, M. Synthetic PreImplantation Factor (PIF) Prevents Fetal Loss by Modulating LPS Induced Inflammatory Response. *PLoS ONE* **2017**, *12* (7), e0180642.

[79] Goodale, L. F.; Hayrabedyan, S.; Todorova, K.; Roussev, R.; Ramu, S.; Stamatkin, C.; Coulam, C. B.; Barnea, E. R.; Gilbert, R. O. PreImplantation Factor (PIF) Protects Cultured Embryos against Oxidative Stress: Relevance for Recurrent Pregnancy Loss (RPL) Therapy. *Oncotarget* **2017**, *8* (20), 32419–32432.

[80] Weiss, L.; Bernstein, S.; Jones, R.; Amunugama, R.; Krizman, D.; JeBailey, L.; Hazan, O.; Yachtin, J.; Shiner, R.; Reibstein, I.; Triche, E.; Slavin, S.; Or, R.; Barnea, E. R.

Preimplantation Factor (PIF) Analog Prevents Type I Diabetes Mellitus (TIDM) Development by Preserving Pancreatic Function in NOD Mice. *Endocrine* **2011**, *40* (1), 41–54.

[81] Roussev, R. G.; Dons'koi, B. V.; Stamatkin, C.; Ramu, S.; Chernyshov, V. P.; Coulam, C. B.; Barnea, E. R. Preimplantation Factor Inhibits Circulating Natural Killer Cell Cytotoxicity and Reduces CD69 Expression: Implications for Recurrent Pregnancy Loss Therapy. *Reprod. BioMed. Online* **2013**, *26* (1), 79–87.

[82] Barnea, E. R.; Almogi-Hazan, O.; Or, R.; Mueller, M.; Ria, F.; Weiss, L.; Paidas, M. J. Immune Regulatory and Neuroprotective Properties of Preimplantation Factor: From Newborn to Adult. *Pharmacol. Ther.* **2015**, *156*.

[83] Mueller, M.; Schoeberlein, A.; Zhou, J.; Joerger-Messerli, M.; Oppliger, B.; Reinhart, U.; Bordey, A.; Surbek, D.; Barnea, E. R.; Huang, Y.; Paidas, M. PreImplantation Factor Bolsters Neuroprotection via Modulating Protein Kinase A and Protein Kinase C Signaling. *Cell Death Differ.* **2015**, *22* (12), 2078–2086.

[84] Migliara, G.; Mueller, M.; Piermattei, A.; Brodie, C.; Paidas, M. J.; Barnea, E. R.; Ria, F. PIF* Promotes Brain Re-Myelination Locally While Regulating Systemic Inflammation-Clinically Relevant Multiple Sclerosis *M.Smegmatis* Model. *Oncotarget* **2017**, *8* (13), 21834–21851.

[85] Barnea, E. R. Applying Embryo-Derived Immune Tolerance to the Treatment of Immune Disorders. *Ann. N. Y. Acad. Sci.* **2007**, *1110* (1), 602–618.

[86] Antonarakis, S. E.; Skotko, B. G.; Rafii, M. S.; Strydom, A.; Pape, S. E.; Bianchi, D. W.; Sherman, S. L.; Reeves, R. H. Down Syndrome. *Nat. Rev. Dis. Primers.* **2020**, *6* (1), 9.

[87] Mann, S.; Spiric, J.; Mitchell, C.; Hilgenkamp, T. I. M. Development of a Physical Therapy-Based Exercise Program for Adults with Down Syndrome. *IJERPH* **2023**, *20* (4), 3667.

[88] Salman M. Systematic review of the effect of therapeutic dietary supplements and drugs on cognitive function in subjects with Down syndrome. *Eur. J. Paediatr. Neurol.* **2002**, *6* (4), 213-9.

[89] Manfredi-Lozano, M.; Leysen, V.; Adamo, M.; Paiva, I.; Rovera, R.; Pignat, J.-M.; Timzoura, F. E.; Candlish, M.; Eddarkaoui, S.; Malone, S. A.; *et al*. GnRH Replacement Rescues Cognition in Down Syndrome. *Science* **2022**, *377* (6610), eabq4515.

[90] Stamatkin, C. W.; Roussev, R. G.; Stout, M.; Absalon-Medina, V.; Ramu, S.; Goodman, C.; Coulam, C. B.; Gilbert, R. O.; Godke, R. A.; Barnea, E. R. PreImplantation Factor (PIF) Correlates with Early Mammalian Embryo Development-Bovine and Murine Models. *Reprod. Biol. Endocrinol.* **2011**, *9* (1), 63.

[91] Barnea, ER.; Lahijani, KI.; Roussev, R.; Barnea, JD.; Coulam, CB. Use of lymphocyte platelet binding assay for detecting a preimplantation factor: a quantitative assay. *Am. J. Reprod. Immunol.* **1994**, *32* (3), 133-138. xs

[92] Barnea, ER.; Kirk, D.; Ramu, S.; Rivnay, B.; Roussev, R.; Paidas, MJ. PreImplantation Factor (PIF) orchestrates systemic antiinflammatory response by immune cells: effect on peripheral blood mononuclear cells. *Am. J. Obstet. Gynecol*. **2012**, *207*(4).

[93] Perez, R. R. G. (75) Inventors: Eytan R. Barnea, Cherry Hill, NJ (US). ASSAYSFORPREIMPLANTATIONFACTOR AND PREMPLANTATION FACTOR PEPTIDES, 2010.

[94] Hayrabedyan, S.; Kirk, D.; Todorova, K.; Barnea, E. Preimplantation Factor (PIF*) Shares a Common RIPK Target to Regulate Global Immune Function Required for Maintained Homeostasis. *J. Reprod. Immunol.* **2014**, Vol. 101–102.

[95] Yang, M.; Yang, Y.; She, S.; Li, S. Proteomic Investigation of the Effects of Preimplantation Factor on Human Embryo Implantation. *Mol. Med. Report* **2017**.

[96] Hayrabedyan, S.; Todorova, K.; Spinelli, M.; Barnea, E. R.; Mueller, M. The Core Sequence of PIF Competes for Insulin/Amyloid β in Insulin Degrading Enzyme: Potential Treatment for Alzheimer's Disease. *Oncotarget* **2018**, *9* (74), 33884–33895.

[97] Rawlings, N. D.; Barrett, A. J. Homologues of Insulinase, a New Superfamily of Metalloendopeptidases. *Biochem. J.* **1991**, *275* (2), 389–391.

[98] Pivovarova, O.; Höhn, A.; Grune, T.; Pfeiffer, A. F. H.; Rudovich, N. Insulin-Degrading Enzyme: New Therapeutic Target for Diabetes and Alzheimer's Disease? *Ann. Med.* **2016**, *48* (8), 614–624.

[99] Hulse, R. E.; Ralat, L. A.; Wei-Jen, T. Chapter 22 Structure, Function, and Regulation of Insulin-Degrading Enzyme. In *Vitamins & Hormones*; Elsevier, 2009; Vol. 80, pp 635–648.

[100] Fernandez-Gamba, A.; Leal, M.; Morelli, L.; Castano, E. Insulin-Degrading Enzyme: Structure-Function Relationship and its Possible Roles in Health and Disease. *Curr. Pharm. Des.* **2009**, *15*, 3644–3655.

[101] Tang, W.-J. Targeting Insulin-Degrading Enzyme to Treat Type 2 Diabetes Mellitus. *Trends Endocrinol. Metab.* **2016**, *27*, 24–34.

[102] González-Casimiro, C.; Merino, B.; Casanueva-Álvarez, E.; Postigo-Casado, T.; Cámara-Torres, P.; Fernández-Díaz, C.; Leissring, M.; Cózar-Castellano, I.; Perdomo, G. Modulation of Insulin Sensitivity by Insulin-Degrading Enzyme. *Biomedicines* **2021**, *9*, 86.

[103] Duckworth, W.C.; Bennett, R.G.; Hamel, F.G. Insulin Degradation: Progress and Potential. *Endocr. Rev.* **1998**, *19*, 608–624.

[104] Farris, W.; Mansourian, S.; Leissring, M.A.; Eckman, E.A.; Bertram, L.; Eckman, C.B.; Tanzi, R.E.; Selkoe, D.J. Partial Loss-of- Function Mutations in In-sulin-Degrading Enzyme that Induce Diabetes also Impair Degradation of Amyloid β-Protein. *Am. J. Pathol.* **2004**, *164*, 1425–1434.

[105] Kurochkin, I.V.; Guarnera, E.; Berezovsky, I.N. Insulin-Degrading Enzyme in the Fight against Alzheimer's Disease. *Trends Pharmacol. Sci.* **2018**, *39*, 49–58.

[106] Zhao, L. Insulin-Degrading Enzyme as a Downstream Target of Insulin Receptor Signaling Cascade: Implications for Alz-heimer's Disease Intervention. *J. Neurosci.* **2004**, *24*, 11120–11126.

[107] Nalivaeva, N.N.; Belyaev, N.D.; Kerridge, C.; Turner, A.J. Amyloid-clearing proteins and their epigenetic regulation as a thera-peutic target in Alzheimer's disease. *Front. Aging Neurosci.* **2014**, *6*, 235.

[108] Li, H.; Wu, J.; Zhu, L.; Sha, L.; Yang, S.; Wei, J.; Ji, L.; Tang, X.; Mao, K.; Cao, L.; et al. Insulin degrading enzyme contributes to the pathology in a mixed model of Type 2 diabetes and Alzheimer's disease: Possible mechanisms of IDE in T2D and AD. *Biosci. Rep.* **2018**, *38*, bsr20170862.

[109] Carrasquillo, M.M.; Belbin, O.; Zou, F.; Allen, M.; Ertekin-Taner, N.; Ansari, M.; Wilcox, S.L.; Kashino, M.R.; Ma, L.; Younkin, L.H.; et al. Concordant Association of Insulin Degrading Enzyme Gene (IDE) Variants with IDE mRNA, Aß, and Alzheimer's Disease. *PLoS ONE* **2010**, *5*, e8764.

[110] Edland, S.D. Insulin-Degrading Enzyme, Apolipoprotein E, and Alzheimer's Disease. *J. Mol. Neurosci.* **2004**, *23*, 213–218.

[111] Zhang, H.; Liu, D.; Huang, H.; Zhao, Y.; Zhou, H. Characteristics of insulin-degrading enzyme in alzheimer's disease: A meta-analysis. *Curr. Alzheimer Res.* **2018**, *15*, 610–617.

[112] Guo, Q.; Manolopoulou, M.; Bian, Y.; Schilling, A.B.; Tang, W.-J. Molecular Basis for the Recognition and Cleavages of IGF-II, TGF-α, and Amylin by Human Insulin-Degrading Enzyme. *J. Mol. Biol.* **2010**, *395*, 430–443.

[113] Malito, E.; Ralat, L.A.; Manolopoulou, M.; Tsay, J.L.; Wadlington, N.; Tang, W.-J. Molecular Bases for the Recognition of Short Peptide Substrates and Cysteine-Directed Modifications of Human Insulin-Degrading Enzyme. *Biochem. J.* **2008**, *47*, 12822–12834.

[114] Manolopoulou, M.; Guo, Q.; Malito, E.; Schilling, A.B.; Tang, W.-J. Molecular Basis of Catalytic Chamber-assisted Unfolding and Cleavage of Human Insulin by Human Insulin-degrading Enzyme. *J. Biol. Chem.* **2009**, *284*, 14177–14188.

[115] Ralat, L.A.; Guo, Q.; Ren, M.; Funke, T.; Dickey, D.M.; Potter, L.R.; Tang, W.-J. Insulin-degrading Enzyme Modulates the Natriuretic Peptide-mediated Signaling Response. *J. Biol. Chem.* **2011**, *286*, 4670–4679.

[116] Ren, M.; Guo, Q.; Guo, L.; Lenz, M.; Qian, F.; Koenen, R.R.; Xu, H.; Schilling, A.B.; Weber, C.; Ye, R.D.; et al. Polymerization of MIP-1 chemokine (CCL3 and CCL4) and clearance of MIP-1 by insulin-degrading enzyme. *EMBO J.* **2010**, *29*, 3952–3966.

[117] Shen, Y.; Joachimiak, A.; Rosner, M.R.; Tang, W.-J. Structures of human insulin-degrading enzyme reveal a new substrate recognition mechanism. *Nature* **2006**, *443*, 870–874.

[118] Durham, T.B.; Toth, J.L.; Klimkowski, V.J.; Cao, J.X.; Siesky, A.M.; Alexander-Chacko, J.; Wu, G.Y.; Dixon, J.T.; McGee, J.E.; Wang, Y.; et al. Dual Exosite-binding Inhibitors of Insulin-degrading Enzyme Challenge Its Role as the Primary Mediator of Insulin Clearance in Vivo. *J. Biol. Chem.* **2015**, *290*, 20044–20059.

[119] Maianti, J.P.; McFedries, A.; Foda, Z.H.; Kleiner, R.E.; Du, X.Q.; Leissring, M.A.; Tang, W.-J.; Charron, M.J.; Seeliger, M.A.; Saghatelian, A.; et al. Anti-diabetic activity of insulin-degrading enzyme inhibitors mediated by multiple hormones. *Nature* **2014**, *511*, 94–98.

[120] Sánchez-Cruz, A.; Hernández-Fuentes, M. D.; Murillo-Gómez, C.; De La Rosa, E. J.; Hernández-Sánchez, C. Possible Role of Insulin-Degrading Enzyme in the Physiopathology of Retinitis Pigmentosa. *Cells* **2022**, *11* (10), 1621.

[121] Ali Khan, H.; Mutus, B. Protein Disulfide Isomerase a Multifunctional Protein with Multiple Physiological Roles. *Front. Chem.* **2014**, *2*.

[122] Ohashi, Y.; Hoshino, Y.; Tanemura, K.; Sato, E. Distribution of Protein Disulfide Isomerase during Maturation of Pig Oocytes: PDI CLUSTER IN MATURE PORCINE OOCYTES. *Anim. Sci. J.* **2013**, *84* (1), 15–22.

[123] Kaplan, A.; Gaschler, M. M.; Dunn, D. E.; Colligan, R.; Brown, L. M.; Palmer, A. G.; Lo, D. C.; Stockwell, B. R. Small Molecule-Induced Oxidation of Protein Disulfide Isomerase Is Neuroprotective. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112* (17).

[124] Ko, H. S.; Uehara, T.; Nomura, Y. Role of Ubiquilin Associated with Protein-Disulfide Isomerase in the Endoplasmic Reticulum in Stress-Induced Apoptotic Cell Death. *J. Biol. Chem.* **2002**, *277* (38), 35386–35392.

[125] Shergalis, A. G.; Hu, S.; Bankhead, A.; Neamati, N. Role of the ERO1-PDI Interaction in Oxidative Protein Folding and Disease. *Pharmacol. Ther.* **2020**, *210*, 107525.

[126] Ryser, H. J.-P.; Flückiger, R. Keynote Review: Progress in Targeting HIV-1 Entry. *Drug Discov. Today* **2005**, *10* (16), 1085–1094.

[127] Barnea, E. R.; Hayrabedyan, S.; Todorova, K.; Almogi-Hazan, O.; Or, R.; Guingab, J.; McElhinney, J.; Fernandez, N.; Barder, T. PreImplantation Factor (PIF*) Regulates Systemic Immunity and Targets Protective Regulatory and Cytoskeleton Proteins. *Immunobiol.* **2016**, *221* (7), 778–793.

[128] Comiskey M, Warner CM, Schust DJ. MHC Molecules of the Preimplantation Embryo and Trophoblast. In: Madame Curie Bioscience Database. Austin (TX): Landes Bioscience; **2000-2013**. Available from: https://www.ncbi.nlm.nih.gov/books/NBK6277/

[129] Stetler, R. A.; Gan, Y.; Zhang, W.; Liou, A. K.; Gao, Y.; Cao, G.; Chen, J. Heat Shock Proteins: Cellular and Molecular Mechanisms in the Central Nervous System. *Prog. Neurobiol.* **2010**, *92* (2), 184–211.

[130] Albakova, Z.; Armeev, G. A.; Kanevskiy, L. M.; Kovalenko, E. I.; Sapozhnikov, A. M. HSP70 Multi-Functionality in Cancer. *Cells* **2020**, *9* (3), 587.

[131] Park, H.-K.; Yoon, N. G.; Lee, J.-E.; Hu, S.; Yoon, S.; Kim, S. Y.; Hong, J.-H.; Nam, D.; Chae, Y. C.; Park, J. B.; Kang, B. H. Unleashing the Full Potential of Hsp90 Inhibitors as Cancer Therapeutics through Simultaneous Inactivation of Hsp90, Grp94, and TRAP1. *Exp. Mol. Med.* **2020**, *52* (1), 79–91.

[132] Sharma, D.; Masison, D. Hsp70 Structure, Function, Regulation and Influence on Yeast Prions. *PPL* **2009**, *16* (6), 571–581.

[133] Zhang, P.; Leu, J. I.-J.; Murphy, M. E.; George, D. L.; Marmorstein, R. Crystal Structure of the Stress-Inducible Human Heat Shock Protein 70 Substrate-Binding Domain in Complex with Peptide Substrate. *PLoS ONE* **2014**, *9* (7), e103518.

[134] Feki, A.; Hibaoui, Y. DYRK1A Protein, A Promising Therapeutic Target to Improve Cognitive Deficits in Down Syndrome. *Brain Sciences* **2018**, *8* (10), 187.

[135] Soundararajan, M.; Roos, A. K.; Savitsky, P.; Filippakopoulos, P.; Kettenbach, A. N.; Olsen, J. V.; Gerber, S. A.; Eswaran, J.; Knapp, S.; Elkins, J. M. Structures of Down Syndrome Kinases, DYRKs, Reveal Mechanisms of Kinase Activation and Substrate Recognition. *Structure* **2013**, *21* (6), 986–996.

[136] Skurat, A. V.; Dietrich, A. D. Phosphorylation of Ser640 in Muscle Glycogen Synthase by DYRK Family Protein Kinases. *J. Biol. Chem.* **2004**, *279* (4), 2490–2498.

[137] Murakami, N.; Bolton, D. C.; Kida, E.; Xie, W.; Hwang, Y.-W. Phosphorylation by Dyrk1A of Clathrin Coated Vesicle-Associated Proteins: Identification of the Substrate Proteins and the Effects of Phosphorylation. *PLoS ONE* **2012**, *7* (4), e34845

[138] Stotani, S.; Giordanetto, F.; Medda, F. DYRK1A Inhibition as Potential Treatment for Alzheimer's Disease. *Future Med. Chem.* **2016**, *8* (6), 681–696.

[139] Miyazaki, Y.; Kikuchi, M.; Umezawa, K.; Descamps, A.; Nakamura, D.; Furuie, G.; Sumida, T.; Saito, K.; Kimura, N.; Niwa, T.; Sumida, Y.; Umehara, T.; Hosoya, T.; Kii, I. Structure-Activity Relationship for the Folding Intermediate-Selective Inhibition of DYRK1A. *Eur. J. Med. Chem.* **2022**, *227*, 113948.

[140] Arbones, M. L.; Thomazeau, A.; Nakano-Kobayashi, A.; Hagiwara, M.; Delabar, J. M. DYRK1A and Cognition: A Lifelong Relationship. *Pharmacology & Therapeutics* **2019**, *194*, 199–221.

[141] Im, H.; Manolopoulou, M.; Malito, E.; Shen, Y.; Zhao, J.; Neant-Fery, M.; Sun, C.-Y.; Meredith, S. C.; Sisodia, S. S.; Leissring, M. A.; Tang, W.-J. Structure of Substrate-Free Human Insulin-Degrading Enzyme (IDE) and Biophysical Analysis of ATP-Induced Conformational Switch of IDE. *J. Biol. Chem.* **2007**, *282* (35), 25453–25463.

[142] Maupetit, J.; Derreumaux, P.; Tuffery, P. PEP-FOLD: An Online Resource for de Novo Peptide Structure Prediction. *Nucleic Acids Res.* **2009**, *37* (Web Server), W498–W503.

[143] Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2009.**

[144] Quiroga, R.; Villarreal, M. A. Vinardo: A Scoring Function Based on Autodock Vina Improves Scoring, Docking, and Virtual Screening. *PLoS ONE* **2016**, *11* (5), e0155183.

[145] Sanner MF, Olson AJ, Spehner JC. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* **1996**, *38* (3), 305-20.

[146] Lei, Y.; Li, S.; Liu, Z.; Wan, F.; Tian, T.; Li, S.; Zhao, D.; Zeng, J. A Deep-Learning Framework for Multi-Level Peptide–Protein Interaction Prediction. *Nat. Commun*. **2021**, *12* (1), 5465.

[147] Jung, H.; Seong, H.-A.; Manoharan, R.; Ha, H. Serine-Threonine Kinase Receptor-Associated Protein Inhibits Apoptosis Signal-Regulating Kinase 1 Function through Direct Interaction. *J. Biol. Chem.* **2010**, *285* (1), 54–70.

[148] Day, C. L.; Puthalakath, H.; Skea, G.; Strasser, A.; Barsukov, I.; Lian, L.-Y.; Huang, D. C. S.; Hinds, M. G. Localization of Dynein Light Chains 1 and 2 and Their Pro-Apoptotic Ligands. *Biochem. J.* **2004**, *377* (3), 597–605.

[149] Escamez, T.; Bahamonde, O.; Tabares-Seisdedos, R.; Vieta, E.; Martinez, S.; Echevarria, D. Developmental Dynamics of PAFAH1B Subunits during Mouse Brain Development. *J. Comp. Neurol.* **2012**, *520* (17), 3877–3894.

[150] Schwab, M.; De Trizio, I.; Ghobrial, M.; Shiu, J.-Y.; Sürücü, O.; Girolamo, F.; Errede, M.; Yilmaz, M.; Haybaeck, J.; Moiraghi, A.; *et al*. Nucleolin Promotes Angiogenesis and Endothelial Metabolism along the Oncofetal Axis in the Human Brain Vasculature. *JCI Insight* **2023**, *8* (8), e143071.

[151] Zalcman, G.; Federman, N.; Romano, A. CaMKII Isoforms in Learning and Memory: Localization and Function. *Front. Mol. Neurosci.* **2018**, *11*, 445.

[152] Li, X.; Li, J.; Qian, J.; Zhang, D.; Shen, H.; Li, X.; Li, H.; Chen, G. Loss of Ribosomal RACK1 (Receptor for Activated Protein Kinase C 1) Induced by Phosphorylation at T50 Alleviates Cerebral Ischemia-Reperfusion Injury in Rats. *Stroke* **2019**, *50* (1), 162–171.

[153] Goswami, S.; Korrodi-Gregório, L.; Sinha, N.; Bhutada, S.; Bhattacharjee, R.; Kline, D.; Vijayaraghavan, S. Regulators of the Protein Phosphatase PP1γ2, PPP1R2, PPP1R7, and PPP1R11 Are Involved in Epididymal Sperm Maturation. *J. Cell. Phys.* **2019**, *234* (3), 3105–3118.

[154] Low, Y.-H.; Asi, Y.; Foti, S. C.; Lashley, T. Heterogeneous Nuclear Ribonucleoproteins: Implications in Neurological Diseases. *Mol. Neurobiol.* **2021**, *58* (2), 631–646.

[155] Pérez-Sen; Queipo; Gil-Redondo; Ortega; Gómez-Villafuertes; Miras-Portugal; Delicado. Dual-Specificity Phosphatase Regulation in Neurons and Glial Cells. *IJMS* **2019**, *20* (8), 1999.

[156] Rapali, P.; Radnai, L.; Süveges, D.; Harmat, V.; Tölgyesi, F.; Wahlgren, W. Y.; Katona, G.; Nyitray, L.; Pál, G. Directed Evolution Reveals the Binding Motif Preference of the LC8/DYNLL Hub Protein and Predicts Large Numbers of Novel Binders in the Human Proteome. *PLoS ONE* **2011**, *6* (4), e18818.

[157] Benison, G.; Karplus, P. A.; Barbar, E. Structure and Dynamics of LC8 Complexes with KXTQT-Motif Peptides: Swallow and Dynein Intermediate Chain Compete for a Common Site. *J. Mol. Biol.* **2007**, *371* (2), 457–468

[158] Gallego, P.; Velazquez-Campoy, A.; Regué, L.; Roig, J.; Reverter, D. Structural Analysis of the Regulation of the DYNLL/LC8 Binding to Nek9 by Phosphorylation. *J. Biol. Chemi.* **2013**, *288* (17), 12283–12294.

[159] Liang, J.; Jaffrey, S.; Guo, W.; Snyder, H. S.; Clardy, J. Structure of the PIN/LC8 dimer with a bound peptide. *Nat. Struct. Mol. Biol*. 1999, **6**, 735–740.

[160] Romes, E. M.; Tripathy, A.; Slep, K. C. Structure of a Yeast Dyn2-Nup159 Complex and Molecular Basis for Dynein Light Chain-Nuclear Pore Interaction. *J. Biol. Chem.* **2012**, *287* (19), 15862–15873.

[161] Rao, L.; Romes, E. M.; Nicholas, M. P.; Brenner, S.; Tripathy, A.; Gennerich, A.; Slep, K. C. The Yeast Dynein Dyn2-Pac11 Complex Is a Dynein Dimerization/Processivity Factor: Structural and Single-Molecule Characterization. *MBoC* **2013**, *24* (15), 2362–2377.

[162] Lightcap, C. M.; Sun, S.; Lear, J. D.; Rodeck, U.; Polenova, T.; Williams, J. C. Biochemical and Structural Characterization of the Pak1-LC8 Interaction. *J.Biol. Chem.***2008**, *283* (40), 27314–27324.

[163] Benison, G.; Karplus, P. A.; Barbar, E. The Interplay of Ligand Binding and Quaternary Structure in the Diverse Interactions of Dynein Light Chain LC8. *J. Mol. Biol.* **2008**, *384* (4), 954–966.

[164] Özden, C.; Sloutsky, R.; Mitsugi, T.; Santos, N.; Agnello, E.; Gaubitz, C.; Foster, J.; Lapinskas, E.; *et al*. CaMKII Binds Both Substrates and Activators at the Active Site. *Cell Reports* **2022**, *40* (2), 111064.

[165] Bhattacharya, S.; Wang, S.; Reddy, D.; Shen, S.; Zhang, Y.; Zhang, N.; Li, H.; Washburn, M. P.; Florens, L.; Shi, Y.; *et al*. Structural Basis of the Interaction between SETD2 Methyltransferase and HnRNP L Paralogs for Governing Co-Transcriptional Splicing. *Nat. Commun.* **2021**, *12* (1), 6452.

[166] Power, J. H. T.; Asad, S.; Chataway, T. K.; Chegini, F.; Manavis, J.; Temlett, J. A.; Jensen, P. H.; Blumbergs, P. C.; Gai, W.-P. Peroxiredoxin 6 in Human Brain: Molecular Forms, Cellular Distribution and Association with Alzheimer's Disease Pathology. *Acta. Neuropathol*. **2008**, *115* (6), 611–622.

[167] Perluigi, M.; Butterfield, D. A. The Identification of Protein Biomarkers for Oxidative Stress in Down Syndrome. *Exp. Rev. Prot.* **2011**, *8* (4), 427–429.

[168] Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215* (3), 403–410.

[169] Mohamed, N. S.; Ali Albsheer, M. M.; Abdelbagi, H.; Siddig, E. E.; Mohamed, M. A.; Ahmed, A. E.; Omer, R. A.; Muneer, M. S.; Ahmed, A.; Osman, H. A.; *et al*. Genetic Polymorphism of the N-Terminal Region in Circumsporozoite Surface Protein of Plasmodium Falciparum Field Isolates from Sudan. *Malar. J.* **2019**, *18* (1), 333.

[170] López, J. A.; Weilenman, C.; Audran, R.; Roggero, M. A.; Bonelo, A.; Tiercy, J. M.; Spertini, F.; & Corradin, G. A synthetic malaria vaccine elicits a potent CD8(+) and CD4(+) T

lymphocyte immune response in humans. Implications for vaccination strategies. *Eur. J. Immunol.* **2001**, *31* (9), 2839.

[171] Gotoh, O. A space-efficient and accurate method for mapping and aligning cDNA sequences onto genomic sequence. *Nucleic Acids Res***. 2008**, *36* (8) 2630-2638.

[172] Schneider, V. A.; Graves-Lindsay, T.; Howe, K.; Bouk, N.; Chen, H.-C.; Kitts, P. A.; Murphy, T. D.; Pruitt, K. D.; Thibaud-Nissen, F.; Albracht, D.; *et al*. Evaluation of GRCh38 and de Novo Haploid Genome Assemblies Demonstrates the Enduring Quality of the Reference Assembly. *Genome Res.* **2017**, *27* (5), 849–864.

[173] Gobet, N.; Jan, M.; Franken, P.; Xenarios, I. Towards Mouse Genetic-Specific RNA-Sequencing Read Mapping. *PLoS Comput. Biol.* **2022**, *18* (9), e1010552

[174] Gordon, D.; Huddleston, J.; Chaisson, M. J. P.; Hill, C. M.; Kronenberg, Z. N.; Munson, K. M.; Malig, M.; Raja, A.; Fiddes, I.; Hillier, L. W.; *et al*. Long-Read Sequence Assembly of the Gorilla Genome. *Science* **2016**, *352* (6281), aae0344–aae0344.

[175] Ma, L.; Bajic, V. B.; Zhang, Z. On the Classification of Long Non-Coding RNAs. *RNA Biology* **2013**, *10* (6), 924–933.

[176] Pauli, A.; Valen, E.; Schier, A. F. Identifying (Non-)Coding RNAs and Small Peptides: Challenges and Opportunities: Prospects & Overviews. *BioEssays* **2015**, *37* (1), 103–112.

[177] Ruiz-Orera, J.; Messeguer, X.; Subirana, J. A.; Alba, M. M. Long Non-Coding RNAs as a Source of New Peptides. *eLife* **2014**, *3*, e03523.

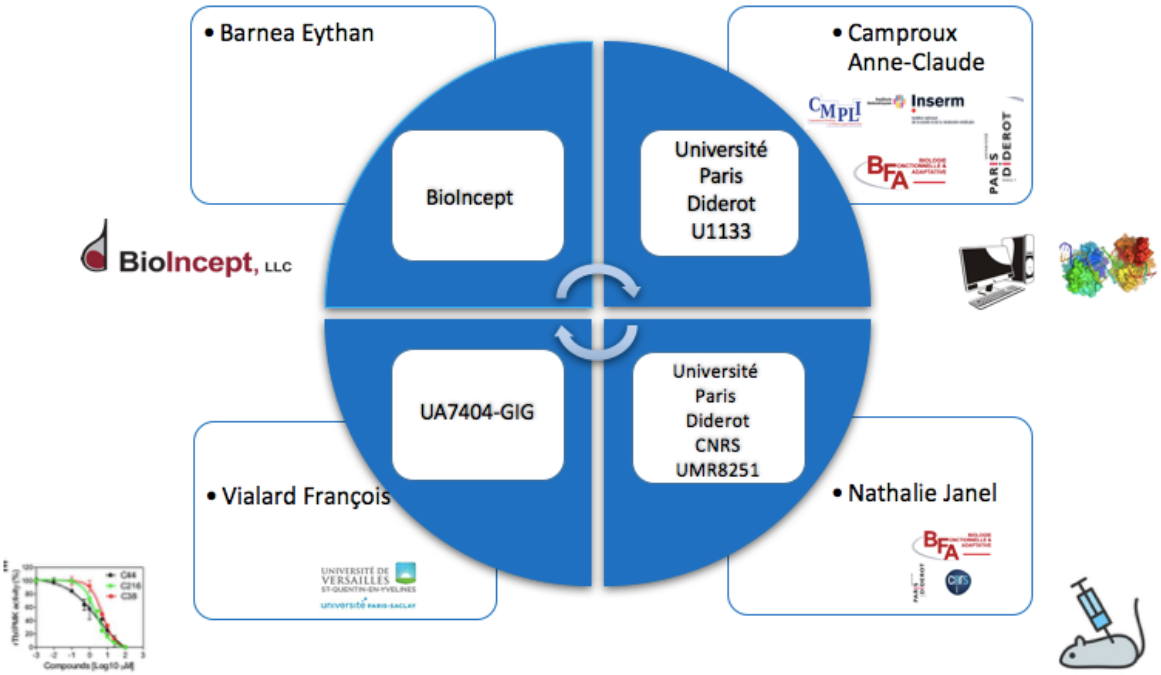
[178] Choi, S.-W.; Kim, H.-W.; Nam, J.-W. The Small Peptide World in Long Noncoding RNAs. *Briefings in Bioinformatics* **2019**, *20* (5), 1853–1864.

[179] Yan, L.; Yang, M.; Guo, H.; Yang, L.; Wu, J.; Li, R.; Liu, P.; Lian, Y.; Zheng, X.; Yan, J.; *et al*. Single-Cell RNA-Seq Profiling of Human Preimplantation Embryos and Embryonic Stem Cells. *Nat. Struct. Mol. Biol.* **2013**, *20* (9), 1131–1139.

[180] Asami, M.; Lam, B. Y. H.; Ma, M. K.; Rainbow, K.; Braun, S.; VerMilyea, M. D.; Yeo, G. S. H.; Perry, A. C. F. Human Embryonic Genome Activation Initiates at the One-Cell Stage. *Cell Stem Cell* **2022**, *29* (2), 209-216.e4.

Annex 1:

Representation of the different teams working on the PIF21 ANR project.



Annex 2:

Résumé détaillé

L'étude détaillée des interactions moléculaires entre les protéines et leurs partenaires est nécessaire pour une meilleure compréhension des phénomènes biologiques et faciliter la conception de molécules thérapeutiques. Ainsi, la présente thèse porte sur l'importance de la caractérisation des interactions protéine-protéine et protéine-peptide et leur impact sur le développement de potentiels médicaments. Au cours de cette thèse, nous nous sommes concentrés sur deux systèmes moléculaires différents. Le premier est une interaction protéine-protéine impliquant la protéine Spike du SARS-CoV-2 (*Severe Acute Respiratory Syndrome-Coronavirus 2*) et son récepteur humain ACE2 (*Angiotensin-Converting Enzyme 2*). Le second porte sur l'identification des interactions entre le peptide PIF (*Preimplantation Factor*) et de potentielles protéines partenaires.

L'étude de l'interaction entre la protéine Spike du SARS-CoV-2 et le récepteur humain ACE2 est d'une importance capitale puisque cette interaction est nécessaire pour l'entrée du virus dans les cellules humaines. L'étude structurale approfondie de ce complexe, comprenant le RBD (*Receptor Binding Domain*) de la protéine Spike et ACE2, s'est concentrée sur deux axes essentiels ayant abouti à deux publications. Premièrement, une analyse du complexe ACE2-RBD a été entreprise, afin d'identifier les résidus les plus importants pour la mise en place et la stabilité de ce complexe. Dans ce but, un protocole de recherche de bioinformatique structurale a été mis au point en ayant recours à plusieurs approches complémentaires, de modélisation moléculaire, telles que des simulations de dynamique moléculaire (DM) et des calculs d'énergies libres avec la méthode MM-PBSA (*Molecular Mechanics/Poisson-Boltzmann Surface Area*), mais également à des outils d'analyse statistique afin de caractériser et de suivre au cours des simulations de DM les régions à la surface du RBD pouvant être ciblées par des molécules médicamenteuses, appelées poches druggables. Deuxièmement, une étude approfondie a été menée pour évaluer l'impact des variants préoccupants du SARS-CoV-2 (Alpha, Beta, Gamma, Delta et Omicron) sur l'affinité du RBD avec ACE2, ainsi que sur la stabilité de ce complexe, permettant ainsi une compréhension plus complète des mécanismes moléculaires mis en jeu.

La modélisation moléculaire a permis de construire des modèles tridimensionnels précis du RBD et des mutations impliquées dans les différents variants. Les simulations de DM ont été essentielles pour étudier le comportement des protéines et capturer les fluctuations et les mouvements moléculaires qui peuvent influencer les interactions entre RBD et ACE2. Des simulations de DM ont permis d'explorer les états conformationnels d'ACE2 et de RBD, offrant

ainsi un aperçu détaillé de leur flexibilité respective à l'état isolée ou en complexe, ainsi que de leur impact sur les interactions étudiées.

Les calculs d'énergies libres MM-PBSA ont permis d'évaluer quantitativement l'affinité entre RBD et ACE2. Cette méthode a fourni des informations détaillées sur la force et la stabilité des interactions étudiées. De plus, 9 résidus clés (K417, L455, F456, E484, F486, Y489, T500, N501 et Y505), ou "*hot spots*", ont été identifié à la surface du RBD comme étant cruciaux pour son interaction avec ACE2. Ces résidus clés ont été identifiés en évaluant les contributions énergétiques individuelles de chaque résidu dans l'interaction entre le RBD et ACE2. Cette analyse a donc permis de mettre en évidence les régions du RBD qui sont essentielles pour la reconnaissance et la liaison à ACE2. La méthode MM-PBSA a également été utilisée pour évaluer l'effet des résidus mutés sur l'interaction entre RBD et ACE2. En introduisant des mutations observées dans le RBD chez les variants les plus préoccupants du SARS-CoV-2, il a été possible d'évaluer comment ces changements affectent l'affinité et la stabilité de l'interaction avec ACE2. Nos résultats montrent que le variant Omicron présente une affinité accrue avec ACE2 en raison de ses nombreuses mutations au niveau du RBD. Ces constatations sont cohérentes avec les données expérimentales, ce qui confirme le caractère hautement virulent et contagieux du variant Omicron. Cette analyse a fourni d'autres informations utiles sur l'importance fonctionnelle de certains résidus et leur rôle dans la liaison entre RBD et ACE2. Grâce à cette étude, nous avons pu observer la présence de plusieurs interactions hydrogène ou de ponts salins, qui ont été abolis ou créées pour compenser et conférer une spécificité à chaque variant. Par exemple, le résidu K417, qui est l'un des *hot spots* les plus importants et qui contribue positivement à l'affinité du complexe ACE2-RBD, ne présente plus d'interaction avec son résidu partenaire D30 situé à la surface d'interaction d'ACE2. Cependant, malgré cette perte d'interaction, la présence d'autres mutations, notamment au sein d'Omicron, a démontré leur capacité à compenser cette perte en possédant des propriétés physico-chimiques favorables à la création d'un environnement plus propice à l'interface ACE2-RBD. Dans l'ensemble, nos résultats mettent en évidence une corrélation significative entre les énergies de liaison calculées, les variations dans la décomposition énergétique de chaque résidu muté et leur effet viral au sein du variant concerné. Cette étude permet également d'obtenir une vision globale de tous les variants préoccupants et de comprendre leurs effets sur l'interaction entre le RBD et ACE2 au niveau moléculaire.

A partir d'un échantillonnage représentatif des différentes conformations du domaine RBD issues des simulations de DM, une recherche de poches druggables a été effectuée à l'aide de l'outil PockDrug développé au sein de notre équipe. Une analyse approfondie des poches et de leur évolution au cours des simulations de DM a été effectuée sur le RBD, qu'il soit lié ou non à ACE2, afin d'identifier et de classer les poches druggables qui pourraient être ciblées par des molécules inhibitrices. Les poches extraites lors des simulations de DM ont été caractérisées en termes de propriétés physico-chimiques, de géométrie et de druggabilité. Elles ont également été classées en fonction de la similarité des résidus afin d'identifier des poches stables et fréquemment druggables tout au long des simulations. PockDrug a ainsi permis d'identifier trois sites dans lesquels des molécules inhibitrices pourraient se lier de manière spécifique et perturber l'interaction avec ACE2 ou inhiber l'activation de la protéine Spike. Ces

sites sont prometteurs pour le développement de médicaments car ils sont prédits druggables et se situent à des endroits à la surface de la protéine Spike qui sont importants pour son activité biologique. En effet, les sites 1 et 2 sont situés entre 2 monomères de la protéine Spike. Cette région joue un rôle crucial dans l'activation de Spike grâce à un changement conformationnel qui permet à RBD de se lier à ACE2. Par conséquent, il serait possible de maintenir Spike dans un état inactif en ciblant ces 2 sites. Il est à noter que notre étude met en évidence pour la première fois que le site 1 est une région à fort potentiel thérapeutique. Enfin, le site 3, lui, se situe sur la région qui permet au RBD d'interagir avec ACE2.

La seconde partie de ce manuscrit porte sur les interactions impliquant le peptide PIF (MVRIKPGSANKPSDD) et des protéines partenaires. Cette étude nous a permis de confirmer ou d'infirmer les cibles potentielles de PIF, tout en ouvrant la voie à l'exploration de nouveaux partenaires potentiels.

Le peptide PIF a été démontré comme étant essentiel au cours des premiers stades de la grossesse. La sécrétion de PIF est importante pour l'implantation et le développement de la grossesse. De plus, PIF possède des propriétés neuroprotectrices et est capable d'atténuer certains phénomènes inflammatoires observés dans le cerveau. Il pourrait donc potentiellement jouer un rôle thérapeutique de premier plan pour aider à traiter la trisomie 21. Cependant, les cibles protéiques potentielles du peptide PIF restent largement inconnues, bien que certaines d'entre elles ont été discutées dans des articles antérieurs. Parmi ces cibles, l'une des plus prometteuses semblait être la protéine IDE (*Insulin-Degrading Enzyme*). IDE joue un rôle crucial dans la dégradation de plusieurs peptides tels que l'insuline et les amyloïdes bêta, et est donc impliquée dans des maladies telles que la maladie d'Alzheimer et le diabète. De plus, la structure dynamique de la protéine IDE est étroitement liée à son rôle biologique. Des études ont révélé que IDE peut adopter des conformations différentes, ce qui lui permet d'interagir avec différents substrats de manière spécifique. Les changements conformationnels de l'IDE en forme ouverte ou fermée sont essentiels pour son activité enzymatique et sa capacité à reconnaître et à dégrader les peptides cibles. La combinaison de différentes méthodes *in silico*, telles que les DM et le *docking* moléculaire, et d'approches *in vitro*, comme l'immunoprécipitation, nous a amené à conclure qu'il n'existe pas d'interaction directe entre le peptide PIF et IDE. Cependant, nous avons décidé de poursuivre nos études de DM sur cette protéine pour mieux comprendre sa flexibilité, car elle représente une cible importante dans plusieurs maladies et joue un rôle important dans le développement de stratégies thérapeutiques. Une étude approfondie du comportement dynamique et de la flexibilité d'IDE a été menée à l'aide de techniques comme la modélisation moléculaire et les simulations de DM. Les mouvements d'ouverture et de fermeture d'IDE liés à sa fonction d'enzyme cryptique ont ainsi pu être décrits pour la première fois, ce qui a conduit à une publication de ces résultats. Nous avons réussi non seulement à capturer les différents états d'IDE mais aussi à identifier de manière précise les résidus impliqués dans ces changements conformationnels. En effet, comprendre comment ces mouvements sont impliqués dans la liaison des inhibiteurs ou des modulateurs de l'IDE peut aider à concevoir des stratégies de développement de médicaments ciblant cette protéine. Ainsi, bien que l'interaction directe entre le peptide PIF et IDE n'ait pas

été confirmée, nos travaux ont contribué à une meilleure compréhension de la flexibilité d'IDE directement liée à sa fonction biologique. Par conséquent, ces résultats ouvrent la voie à de futures études pour explorer davantage le rôle d'IDE dans divers processus biologiques et développer des solutions thérapeutiques pour cette protéine clé.

En raison de sa fonction dans la croissance et la différenciation cellulaires, ainsi que sa participation à de nombreuses voies de signalisation cruciales pour le contrôle du développement neuronal, DYRK1A (*Dual Specificity Tyrosine Phosphorylation Regulated Kinase 1A*) également été considérée comme une cible potentielle du peptide PIF. DYRK1A joue un rôle significatif dans la trisomie 21, faisant de cette protéine une cible d'intérêt. De manière intéressante, lors de notre étude structurale de DYRK1A, nous avons identifié une importante similarité de séquence entre un peptide substrat de DYRK1A et PIF. Ce peptide substrat se lie spécifiquement au domaine kinase de DYRK1A. Par conséquent, nous avons cherché à examiner plus en détail la relation potentielle entre DYRK1A et PIF afin de déterminer s'il existe une interaction fonctionnelle entre ces deux systèmes moléculaires. Dans ce but, nous avons utilisé le programme AlphaFold2-multimer qui est capable de prédire avec une grande précision la structure de complexes protéiques. AlphaFold2-multimer utilise le score pLDDT pour évaluer ses prédictions structurales. Ce score peut prendre des valeurs entre 0 et 100 et plus il est important, plus AlphaFold2-multimer accorde une grande confiance dans sa prédiction. D'après nos résultats, AlphaFold2-multimer est capable de retrouver la structure connue expérimentalement de DYRK1A en interaction avec un peptide qui possède une taille proche de celle du peptide PIF. Concernant l'interaction entre DYRK1A et le peptide PIF, AlphaFold2-multimer propose des modèles ayant des scores pLDDT élevés, c'est-à-dire entre 50 et 90, qui laissent à penser que nous pouvons être confiants dans ses prédictions structurales du complexe DYRK1A-peptide PIF. Cependant, nous n'avons pas pu obtenir, malgré nos multiples tentatives, une structure satisfaisante de l'interaction entre PIF et DYRK1A qui aurait dû, d'après les résultats expérimentaux de nos collaborateurs, impliquer le domaine PEST (*C-terminal domain enriched in proline (P), glutamic acid (E), serine (S), and threonine*) une région intrinsèquement désordonnée de DYRK1A qui est, par essence, complexe à modéliser et donc difficile à étudier. Toutefois, ces résultats ont permis de conclure que PIF ne présente pas une spécificité de substrat envers le domaine kinase de DYRK1A.

Étant donné que deux des cibles partenaires potentielles de PIF identifiées précédemment n'ont pas été confirmées par les méthodes *in silico* et expérimentales, nous avons entrepris une étude approfondie de PIF en utilisant diverses approches afin de rechercher de nouveaux partenaires potentiels. Ces approches intègrent deux méthodes, à savoir la protéomique quantitative par spectrométrie de masse effectuée par nos collègues expérimentalistes du projet ANR (Agence Nationale de Recherche) PIF et de *deep learning*. En combinant ces deux approches, notre objectif était de croiser les résultats, les filtrer et identifier des protéines d'intérêts pour PIF. À l'aide du programme CAMP (*Convolutional Attention-based Neural Network for Multi-level Peptide-protein interaction Prediction*), basé sur du *deep learning*, nous avons identifié plusieurs cibles protéiques présentant des scores de prédiction élevés suggérant une possible interaction avec PIF. Par la suite, ces résultats ont été croisés avec les données obtenues expérimentalement par une méthode de protéomique quantitative par spectrométrie de masse.

Le croisement de ces méthodes a permis l'identification de 10 protéines d'intérêt, dont plusieurs kinases, qui sont impliquées dans les mêmes processus biologiques que PIF. Ainsi, plusieurs cibles susceptibles d'interagir avec PIF ont été identifiées : la protéine STRAP (Serine-threonine kinase receptor-associated protein), DYNLL2 (Dynein light chain 2), PAFAH1B2 (Platelet-activating factor acetylhydrolase IB subunit beta), Nucléoline, CAMKB2 (Calcium/calmodulin-dependent protein kinase type II subunit beta), RACK1 (Receptor of activated protein C kinase 1), PPP1R7 (Protein phosphatase 1 regulatory subunit 7), HNRNPAB (Heterogeneous nuclear ribonucleoprotein A/B), and MAP2K1 (Dual specificity mitogen-activated protein kinase kinase 1) et PRDX6 (Peroxiredoxin-6). En outre, il est intéressant de noter que plusieurs de ces protéines sont connues expérimentalement pour interagir aves des peptides pour exercer leur activité biologique. Cette caractéristique nous donne la possibilité de comparer ultérieurement ces résultats avec ceux prédits pour les interactions impliquant PIF et valider une cible thérapeutique.

Malgré les rôles importants de PIF, son origine génétique n'a jamais pu être identifiée. Par le biais de méthodes génomiques, l'origine génétique du PIF a été soigneusement étudiée, et des recherches supplémentaires sont en cours pour mieux comprendre la régulation du PIF aux niveaux transcriptionnels et post-transcriptionnels. En effet, PIF se distingue par sa petite taille et une combinaison unique d'acides aminés, ce qui le rend non homologue à tout autre peptide humain répertorié dans les bases de données disponibles comme BLAST (*Basic Local Alignment Search Tool*). Par conséquent, pour identifier le gène codant de PIF dans le génome humain, nous avons utilisé le programme SPALN (*Space-Efficient Spliced Alignment*). Nous avons également refait l'étude sur le génome du gorille et de la souris. Cette analyse approfondie a révélé la présence d'une partie du peptide PIF (VRIKPGSA) dans une région non codante du gène NRXN3 (Neurexin-3-alpha) située sur le chromosome 14 humain. En conclusion, la séquence unique et non homologue du PIF par rapport aux peptides humains connus dans les bases de données disponibles soulève des questions sur son origine génétique. Nos résultats indiquent que PIF pourrait être un peptide codé par une région non codante du chromosome 14, soit en tant que smORF (*small open Reading Frame*) d'un ARN non-codant, soit en tant que partie d'un ARN long non-codant bifonctionnel. Pour mieux comprendre l'origine et les fonctions biologiques de PIF, des études futures devraient se concentrer sur l'exploration de ces hypothèses.

Grâce aux découvertes de cette thèse, nous avons maintenant une meilleure compréhension des protéines cibles potentielles du PIF, ce qui a également servi de base à une vérification expérimentale supplémentaire de ces interactions. Dans l'ensemble, cette recherche ouvre la voie à de futures investigations sur les interactions entre le PIF et ses protéines cibles, offre de nouvelles perspectives importantes sur les mécanismes sous-jacents aux interactions protéine-protéine, ainsi qu'à l'identification de cibles thérapeutiques potentielles dans divers contextes pathologiques. Pour compléter ces résultats, une collaboration avec des équipes expérimentales serait cruciale pour valider les résultats obtenus et identifier des inhibiteurs efficaces, notamment dans le contexte du SARS-CoV-2. Cette approche permettrait de confirmer l'importance potentielle des poches identifiées en tant que cible thérapeutique pour lutter contre

les infections virales. De plus, l'application du protocole utilisé pour étudier le complexe ACE2-RBD pourrait être appliqué à PIF et à l'étude de ses interactions avec de nouvelles cibles protéiques pour les quantifier et les caractériser, afin d'aboutir à une meilleure compréhension des mécanismes moléculaires impliqués dans le contexte d'interactions protéine-peptide.

Abstract

A bioinformatics analysis of therapeutic proteins: studying protein partners for the preimplantation factor (PIF) and of SARS-CoV-2 proteins

Protein-protein and protein-peptide interactions are the basis for many biological processes. Therefore, understanding the molecular and structural mechanisms of these interactions can aid in the design of therapeutic molecules.

This thesis examines the importance of protein-protein and protein-peptide interactions in drug discovery and biological processes. The aim is to explore protein-protein interactions between the human ACE2 receptor and the Spike protein RBD, as well as protein-peptide interactions between the PIF peptide and its protein targets. A structural bioinformatics protocol including molecular modeling, MD simulations, MM-PBSA calculations and druggable pocket tracking was developed to study RBD-ACE2 interactions. This methodology led to the identification of three potential druggable sites on the Spike protein. A second study investigated the impact of SARS-CoV-2 variants of concern mutation and their effects on ACE2 affinity. Furthermore, we also used molecular docking, MD simulations and genomics approaches to investigate potential PIF protein partners such as IDE and DYRK1A. This work let to the examination of the IDE protein and offered crucial insights into its biological function and prospective therapeutic uses. Moreover, the AlphaFold2-multimer program was used to predict the structural arrangement of the interaction between DYRK1A and PIF.

Keywords: molecular dynamics simulations, molecular docking, molecular modeling, MM-PBSA calculations, therapeutic targets, protein-protein interactions, protein-peptide interactions, SARS-CoV-2, Preimplantation Factor peptide (PIF)

Résumé

Analyse bioinformatique des protéines thérapeutiques : étude des protéines partenaires du facteur de préimplantation (PIF) et des protéines du SARS-CoV-2

Les interactions protéine-protéine et protéine-peptide sont essentielles pour de nombreux phénomènes biologiques. Ainsi, comprendre les mécanismes de ces interactions pourrait aider à la conception des molécules thérapeutiques.

L'objectif de cette thèse est d'étudier les interactions protéine-protéine entre le récepteur ACE2 humain et le domaine RBD de la protéine Spike du SARS-CoV-2, ainsi que les interactions protéine-peptide entre le peptide PIF et ses cibles protéiques. Un protocole de bioinformatique structurale comprenant des approches de modélisation moléculaire, de simulations de dynamique moléculaire, de calculs MM-PBSA et de suivi des poches druggables a été appliqué pour étudier les interactions RBD-ACE2. Cette approche a principalement permis l'identification de trois sites druggables sur la protéine Spike. Un second travail a permis d'étudier les impacts des mutations des principaux variants de la protéine Spike et leurs conséquences sur son affinité avec ACE2. Concernant l'étude du peptide PIF et ses partenaires potentiels tels qu'IDE et DYRK1A, nous avons utilisé des approches de modélisation moléculaire, de *docking* moléculaire, de simulations de dynamique moléculaire et de génomique. Ce travail portant sur la protéine IDE a permis de mieux comprendre les mouvements à l'origine de sa fonction biologique et pourrait être la base du développement de molécules thérapeutiques.

Mots-clefs : simulation de dynamique moléculaire, *docking* moléculaire, modélisation moléculaire, calculs MM-PBSA, cibles thérapeutiques, interactions protéine-protéine, interaction protéine-peptide, SARS-CoV-2, facteur de préimplantation (PIF)