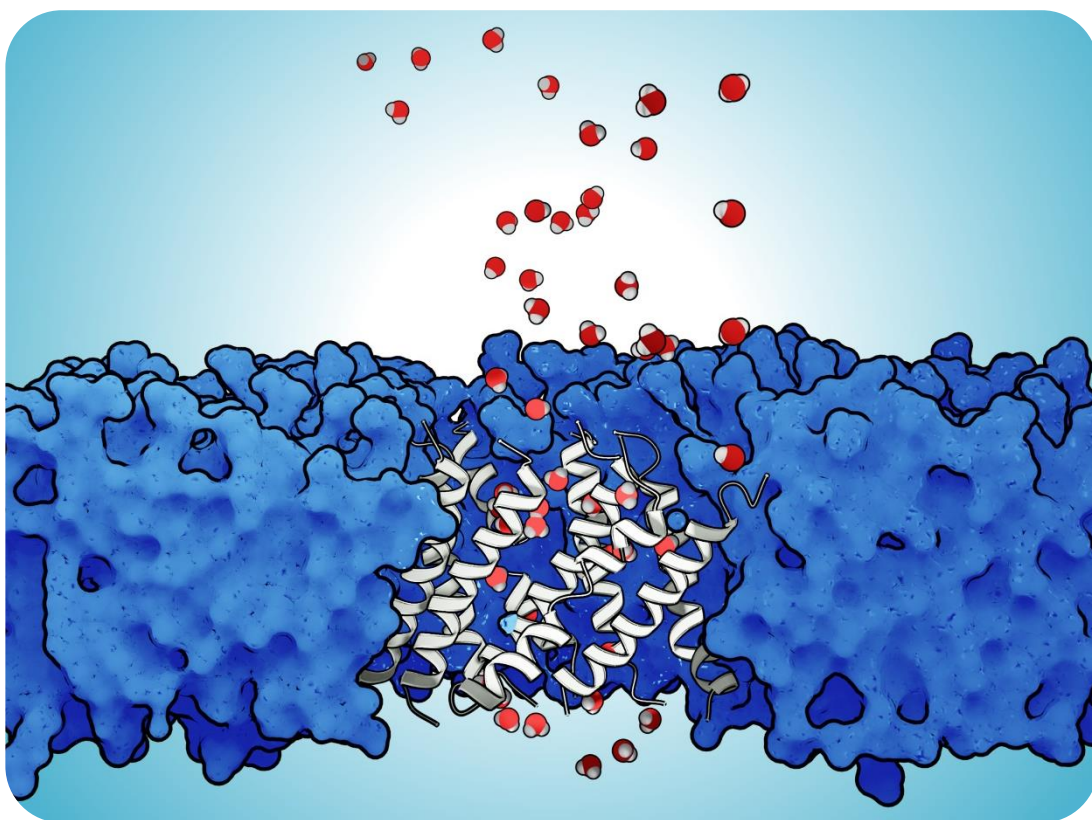


Unlocking Viral Entry: The Key Role of Fusion Proteins

The importance of viral fusion protein in target cell recognition and viral fusion

Mariana de Oliveira Farinha Fidalgo Valério



Dissertation presented to obtain the **Ph.D degree in
Molecular Biosciences**, speciality in Biochemistry

Oeiras, December, 2023

Unlocking Viral Entry: The Key Role of Fusion Proteins

Mariana de Oliveira Farinha Fidalgo Valério

Graduated in Biochemistry for Health

Dissertation presented to obtain the **Ph.D degree in Molecular Biosciences**

Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Supervisor:

Prof. Dr. Cláudio M. Soares

Co-Supervisors:

Dr. Diana Lousa

Dr. Manuel N. Melo

Members of the Jury:

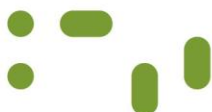
Prof. Dr. Siewert-Jan Marrink

Prof. Dr. Nuno Fernando Duarte Cordeiro Correia dos Santos

Prof. Dr. Rita Alexandra do Nascimento Cardoso Guedes

Prof. Dr. Bruno Lourenço da Silva Victor

December 2023



INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
/UNL

Knowledge Creation



ACKNOWLEDGEMENTS

It has been a transformative four-year journey, and, looking back, I can see that it wouldn't have been the same without a couple of people who contributed directly or indirectly to making this an amazing journey.

First of all, I would like to thank my supervisors, Professor Cláudio M. Soares, Dr. Diana Lousa, and Dr. Manuel N. Melo, for providing me with the opportunity to develop this thesis at ITQB NOVA. Their continuous support, guidance, patience, and availability throughout these years have been instrumental to my growth and success. A special thanks for all their devotion and time spent teaching me and the insightful comments, always "spot on."

A special and profound thanks to everyone who was or still is a part of the Protein Modeling, Molecular Simulation, Multiscale Modeling and Synthetic Biology groups, especially Sofia, Lucie, Luís, and Alexandra for their patience, friendship, and support in the toughest moments. You have become my friends and without you, this phase would have been much harder and less fun.

To my supervisors, I also express my sincere appreciation for granting me the freedom to imagine and pursue new ideas, particularly the project that led me to Dr. Paulo C. T. Souza's lab in Lyon. I extend my heartfelt gratitude to the DAMM lab at ENS Lyon for hosting me, providing an enriching environment, and sharing their invaluable insights and teachings. The experience gained within the DAMM lab has not only shaped my perspective but has significantly deepened my understanding of the field.

Special thanks go to Dr. Paulo C. T. Souza, Dr. Luís Borges-Araújo, and Dr. Lisbeth Kjølbye, whose generous sharing of expertise, guidance, and the creation of a stimulating learning environment has been integral to the successful completion of my project. The collaborative and inclusive atmosphere within the DAMM lab has turned this academic journey into not only an educational pursuit but also an immensely enjoyable one.

Additionally, I extend my sincere appreciation to the Fundação para a Ciência e Tecnologia (FCT), a pivotal supporter of my academic journey. Their funding has played a crucial role in the realization of this thesis and has been instrumental in facilitating my research endeavors. Beyond the financial assistance, FCT has allowed me to attend

various congresses, enhancing my academic exposure and fostering connections within the scientific community. Moreover, their financial support to my stay in Lyon, allowed me to benefit from the collaborative and intellectually stimulating environment of the DAMM lab at ENS Lyon. This support has not only shaped the trajectory of my academic pursuits but has also contributed significantly to the success of my overall scholarly experience.

To all my lifelong friends, for their unwavering support and friendship, and to two special people who couldn't stay by my side all along this journey. I hope I will see you again.

Finally, I reserve my deepest thanks to my family. You have been the pillars, consistently setting the bar high and never allowing me to fall face down. Your unwavering support in every choice and the freedom you've granted me to "fly and make my own mistakes" are the bedrock of my determination to work hard every day to become a better person.

”

“Nothing is lost, nothing is created, everything is transformed.”

— **Antoine Lavoisier**, *Traité élémentaire de chimie*, 1789
(Chemist)

ABSTRACT

Viral infections are a substantial global health concern, affecting millions of individuals annually. The initial stage of infection involves viral entry, a process characterized by intricate mechanisms that remain poorly understood. In this thesis, we employ molecular dynamics simulations to elucidate the complex mechanisms that allow for the entry of enveloped viruses into our cells. To achieve this we focused on the existing gaps in the literature concerning the role of fusion proteins in ensuring efficient viral infection, with specific attention directed towards events associated with target cell receptor recognition and membrane fusion. This was studied by analyzing both the SARS-CoV-2 and parainfluenza 5 viruses.

In **Chapter 3** we start by scrutinizing the impact of naturally occurring mutations on the SARS-CoV-2 receptor binding domain (RBD) dynamics. To investigate this, we employed atomistic molecular dynamics simulations of the wild-type RBD and four variants of concern (Alpha, Beta, Delta, and Omicron). Our results revealed that the wild-type RBD can exist in two forms: "open" for ACE2 binding and "closed", where the binding surface is blocked. The Alpha, Beta, Delta and Omicron variants strongly favor the open state, with Delta showing an additional alternative conformation that may aid antibody evasion (the "reverse open"). These findings suggest that variants of concern (VOCs), especially Omicron and Delta, influence RBD conformation to facilitate ACE2 binding and potentially assist antibody evasion.

In a sub-section of this chapter, we describe our contribution to a study aimed at identifying the effects of single and synergistic amino acid substitutions in promoting SARS-CoV-2 immune escape. This work combined antibody neutralization assays on pseudotyped viruses, bearing the SARS-CoV-2 fusion protein, and computational analysis to identify residues frequently interacting with antibodies. Our contribution consisted of the analysis of 57 PDB structures, to discern the key residues of the SARS-CoV-2 fusion protein engaged in antibody interactions, an analysis complemented with RBD-ACE2 simulations. This work provided insights into the E484K immune evasion, showing that E484 is frequently involved in interactions with antibodies, but not with ACE2. Additionally, we identified a novel amino acid of concern, S494, exhibiting a similar

interaction pattern.

In **Chapter 4** we shift our focus to the parainfluenza virus and we delve into the role of the fusion peptide in mediating membrane fusion. We conducted molecular dynamics simulations, employing a hybrid approach that combines atomistic and coarse-grain detail. We then compared the simulation outcomes with biophysical assays conducted by our research partners. From the simulation results, we were able to identify crucial amino acid residues in the parainfluenza fusion peptide (PIFP) that are essential for membrane fusion. Additionally, at high concentrations, the PIFPs induced the assembly of water-permeable pore-like structures, facilitating lipid head intrusion and tail protrusion. Biophysical experiments confirm these findings, revealing that PIFP can promote fusion and/or membrane leakage depending on the peptide/lipid ratio.

To build upon the data described in Chapter 4, in **Chapter 5**, we introduced alanine mutations to the PIFP amino acids identified as crucial for membrane fusion, namely Q120 and F103. As before, we conducted molecular dynamics simulations of these systems and this data was complemented by biophysical assays conducted by our research collaborators. The biophysical assays demonstrated that the F103A mutation significantly affects lipid mixing and membrane leakage, contrasting with the marginal impact of the Q120A mutation. *In silico* data corroborated this trend, showing a lower flux of water through the F103A pore-like structure and fewer lipid tail protrusion events.

Lastly, in **Chapter 6** we conduct exploratory research on the influence of SARS-CoV-2 and parainfluenza fusion proteins on viral fusion. We employ coarse-grained molecular dynamics simulations to study fusion between simplified viral vesicles, both with and without fusion proteins, and cell membrane models. Notably, in systems containing a fusion protein, we observed a significant increase in membrane-membrane interactions in both SARS-CoV-2 and parainfluenza systems. Extensive mixing of lipids from vesicles and cell membranes occurred, accompanied by the formation of more stalk structures compared to simulations without proteins. Despite the interesting results of this study, there is still room for improvement as none of the simulations reached full fusion.

By integrating insights from these comprehensive studies on SARS-CoV-2 and parainfluenza fusion proteins, our research contributes to a deeper understanding of viral fusion dynamics. These findings hold promise for developing therapies aimed at inhibiting viral entry and combating viral infections, with profound implications for global health outcomes.

Keywords: Fusion proteins, Molecular Dynamics simulations, Fusion peptide, Receptor binding domain, Enveloped viruses

RESUMO

As infecções virais são uma preocupação a nível global, afetando milhões de pessoas anualmente. A fase inicial de uma infecção envolve a entrada do vírus na célula hospedeira, um processo caracterizado por mecanismos complexos e ainda pouco compreendidos. Nesta tese, utilizámos simulações de dinâmica molecular para melhor compreender os mecanismos que permitem a entrada de vírus encapsulados nas células. Focamo-nos principalmente no estudo do papel das proteínas de fusão numa infecção viral, mais especificamente nos eventos relacionados com o reconhecimento dos recetores da célula-alvo e da fusão da membrana. Para isto, utilizamos os vírus SARS-CoV-2 e parainfluenza 5 como modelos.

No **Capítulo 3**, examinamos o impacto de mutações que ocorrem naturalmente, na dinâmica do domínio de ligação ao recetor (RBD) do SARS-CoV-2. Para esta investigação, realizamos simulações atomísticas de dinâmica molecular do RBD "*wild-type*" e quatro variantes (Alpha, Beta, Delta e Omicron). Os nossos resultados mostraram que o RBD "*wild-type*" pode existir em duas formas: "aberta" para a ligação com o recetor ACE2 e "fechada", onde a superfície de ligação está bloqueada. As variantes Alpha, Beta, Delta e Omicron favorecem fortemente o estado aberto, enquanto que a Delta mostra uma conformação alternativa adicional, que pode auxiliar na evasão aos anticorpos (a "abertura inversa"). Estas descobertas sugerem que as variantes, especialmente Omicron e Delta, influenciam a conformação do RBD para facilitar a ligação ao ACE2 e potencialmente auxiliar na evasão a anticorpos.

Numa sub-secção deste capítulo, descrevemos um estudo sobre o efeito que substituições pontuais e combinadas na proteína de fusão do SARS-CoV-2 podem ter na evasão do vírus a anticorpos. Este trabalho combinou ensaios de neutralização de anticorpos em pseudovírus, contendo a proteína de fusão do SARS-CoV-2, e análises computacionais para identificar resíduos que interagem mais frequentemente com anticorpos. A nossa contribuição para este trabalho envolveu a análise computacional de 57 estruturas PDB, para discernir os resíduos-chave da proteína de fusão do SARS-CoV-2 envolvidos em interações com anticorpos, uma análise complementada por simulações RBD-ACE2. A nossa contribuição para este estudo mostra que o E484 está frequentemente envolvido

em interações com anticorpos, mas não com a proteína recetora ACE2. Adicionalmente, identificamos um novo aminoácido que exibe um padrão de interação semelhante, o S494.

No **Capítulo 4**, direcionamos o nosso foco para o vírus parainfluenza e investigamos o papel do seu péptido de fusão na mediação da fusão de membranas. Realizamos simulações de dinâmica molecular, utilizando uma abordagem híbrida que combina detalhes atomísticos e de "*coarse-grain*". Em seguida, comparamos os resultados da simulação com ensaios biofísicos conduzidos pelos nossos colaboradores experimentais. A partir dos resultados da simulação, identificamos os resíduos de aminoácido cruciais no péptido de fusão do parainfluenza (PIFP) para a fusão de membranas. Além disso, a concentrações elevadas, os PIFPs induziram a formação de estruturas semelhantes a poros permeáveis à água, facilitando a intrusão das cabeças dos lípidos e a protrusão da sua cauda. Os dados biofísicos confirmam essas descobertas, revelando que o PIFP pode promover a fusão e/ou passagem de água pela membrana, dependendo da razão péptido/lípido.

Para adicionar os dados descritos no Capítulo 4, no **Capítulo 5**, alteramos para aliana os aminoácidos do PIFP identificados como cruciais para a fusão de membranas, nomeadamente Q120 e F103. Como descrito anteriormente, realizamos simulações de dinâmica molecular destes sistemas, complementando estes dados com ensaios biofísicos levados a cabo pelos nossos colaboradores. Estes demonstraram que a mutação F103A afeta significativamente a fusão de duas membranas e a passagem de água através delas, contrastando com o impacto marginal da mutação Q120A. Os dados *in silico* corroboraram essa tendência, revelando uma redução na passagem de água pela membrana e menos eventos de protrusão nos sistemas com o PIFP F103A.

Por fim, no **Capítulo 6**, investigamos a influência das proteínas de fusão do SARS-CoV-2 e parainfluenza na fusão viral. Utilizamos simulações de dinâmica molecular de "*coarse-grain*" para estudar a fusão entre vesículas virais simplificadas e modelos de membrana celular, com e sem proteínas de fusão. Na presença da proteína de fusão, observamos um aumento significativo na interação das membranas tanto nos sistemas virais de SARS-CoV-2, quanto parainfluenza. Nestes sistemas, verifica-se uma mistura extensiva dos lípidos das vesículas virais e membranas celulares, acompanhados pela formação de mais estruturas de "*stalk*" em comparação com simulações sem proteínas. Apesar dos resultados interessantes deste estudo, ainda há espaço para melhorias, uma vez que nenhuma das simulações atingiu a fusão total.

Ao integrar os conhecimentos obtidos sobre as proteínas de fusão do SARS-CoV-2 e parainfluenza, esta tese contribui para uma compreensão mais profunda da dinâmica da fusão viral. Estas descobertas potenciam o desenvolvimento de terapias destinadas a inibir a entrada viral e a combater infeções virais, com implicações profundas para a saúde global.

Palavras-chave: Proteínas de fusão, Simulações de Dinâmica Molecular, Péptido de fusão, Domínio de ligação ao receptor, Vírus encapsulados

LIST OF ABBREVIATIONS AND SYMBOLS

Abbreviations

AA	Atomistic Simulations
ACE2	Angiotensin-Converting Enzyme 2
BLI	Bilayer Interferometry
CAC	Critical Aggregation Concentration
CD	Circular dichroism
CG	Coarse Grain
CV	Collective Variable
COVID-19	Coronavirus disease 2019
DH	Hydrodynamic Diameter
DLS	Dynamic light scattering
DOPC	1,2-dioleoyl-sn-glycero-3-phosphocholine
DOPE	1,2-dioleoyl-sn-glycero-3-phosphoethanolamine
DOPG	1,2-dioleoyl-sn-glycero-3-phosphoglycerol
F	Fusion glycoprotein
FP	Fusion Peptide
FRET	Förster resonance energy transfer
HIV	Human Immunodeficiency Virus
HN	Hemagglutinin-Neuraminidase
HPIV3	Parainfluenza Virus 3
LUVs	Large Unilamellar Vesicles
NPT	Isothermal–Isobaric Ensemble
PBC	Periodic Boundary Conditions
PCA	Principal Component Analysis
PCs	Principal Components
PIFP	Parainfluenza Fusion Peptide
PIVs	Parainfluenza Viruses
POPC	1-palmitoyl-2-oleoyl-glycero-3-phosphocholine
POPG	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol

MD	Molecular dynamics
MERS-CoV	Middle East Respiratory Syndrome-Related Coronavirus
RBD	Receptor Binding Domain
RBM	Receptor Binding Motif
RIN	Residue Interaction Network
RG	Radius of Gyration
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SASA	Solvent Accessible Surface Area
SPR	Surface Plasmon Resonance
SUVs	Small Unilamellar Vesicles
SV5	Parainfluenza Virus 5
VOCs	Variants of Concern
2D	Two-dimensional
5,6-CF	5,6-carboxyfluorescein
6HB	Six-Helix Bundle
Symbols	
k_{on}	Association Rate Constant
k_{off}	Dissociation Rate Constant
λ_{exc}	Donor Excitation Wavelength
K_P	Partition Constant
R	Fusion Efficiency
$S_L(t)$	Dissociation Fractional Response
I_F	Fluorescence Emission Intensity

CONTENTS

Acknowledgements	iii
Abstract	vii
Resumo	ix
List of abbreviations and symbols	xi
List of Figures	xvii
List of Tables	xxi
1 General Introduction	1
1.1 Viral Infections: A Global Health Challenge	1
1.1.1 Viral Entry: The door to Infection	2
1.1.2 Enveloped Viruses: The Fusion Protein is key	5
1.1.3 Target Cell Recognition: Finding the lock	7
1.1.4 Viral membrane fusion: the role of Fusion Peptide	12
1.2 Simulation as a tool to better understand viral fusion	15
1.3 Dissertation objectives and outline	16
2 General Theory and Methods	19
2.1 Molecular Modelling	19
2.2 Molecular Dynamics Simulations	19
2.2.1 Integration algorithms	20
2.2.2 Force field	21
2.2.3 Periodic boundary conditions	23
2.2.4 Constant temperature and/or pressure	24
2.3 The challenge of scale: A multiscale approach	25

3	Exploring the impact of mutations on the SARS-CoV-2 RBD conformational dynamics and antibody neutralization	29
3.1	SARS-CoV-2 variants impact RBD conformational dynamics and ACE2 accessibility	32
3.1.1	Abstract	32
3.1.2	Introduction	32
3.1.3	Methods	35
3.1.4	Results and discussion	37
3.1.5	Conclusion	46
3.2	Signatures in SARS-CoV-2 spike protein conferring escape to neutralizing antibodies	48
3.2.1	Abstract	48
3.2.2	Introduction	48
3.2.3	Methods	49
3.2.4	Results	51
3.2.5	Conclusion	54
3.3	Supplementary Information	55
4	Parainfluenza Fusion Peptide Promotes Membrane Fusion by Assembling into Oligomeric Porelike Structures	73
4.1	Abstract	75
4.2	Introduction	75
4.3	Materials and Methods	78
4.3.1	Instrumentation	78
4.3.2	Lipid Vesicles Preparation	78
4.3.3	PIFP-induced Vesicles Fusion Assay	78
4.3.4	Vesicles Leakage	79
4.3.5	Circular Dichroism	79
4.3.6	Dynamic Light Scattering	79
4.3.7	Surface Plasmon Resonance	80
4.3.8	Coarse Grain MD Simulations of the PIFP in a POPC/POPG membrane	80
4.3.9	Atomistic MD Simulations of the PIFP in a POPC/POPG	82
4.3.10	Metadynamics Simulations in aqueous solution	82
4.4	Results	83
4.4.1	PIFP Induces Fusion and Perturbation on Negatively Charged Lipid Vesicles	83
4.4.2	PIFP–Lipid Unbinding Follows a Second-Order Kinetics	83
4.4.3	PIFP Helix Content Depends on the Membrane Saturation Level	86
4.4.4	Vesicle Aggregation Triggering Is Dependent on the PIFP Concentration	86

4.4.5	CG Simulations Reveal the PIFP Aggregates When Inside a POPC:POPG Membrane	87
4.4.6	PIFP That Can Adopt Moderately Helical Structures in Water and Helicity in the Membrane Is Influenced by the Peptide/Lipid Ratio	89
4.4.7	PIFP Peptide–Peptide Interactions Stabilize Porelike Structures	92
4.4.8	Water Can Pass through the PIFP Porelike Structures	94
4.4.9	PIFP Induces Concentration-Dependent Lipid Tail Protrusion	95
4.5	Discussion	97
4.6	Supplementary Information	98
4.6.1	Molecular Dynamics Simulations Analysis Tools	99
5	Substituting the Parainfluenza Fusion Peptide N-terminal residue impacts its fusogenicity	113
5.1	Abstract	115
5.2	Introduction	115
5.3	Methods	117
5.3.1	Instrumentation	117
5.3.2	Lipid Vesicles Preparation	117
5.3.3	PIFPs-induced lipid mixing	118
5.3.4	PIFPs-induced Vesicle Leakage	118
5.3.5	Surface Plasmon Resonance	118
5.3.6	Dynamic Light Scattering	119
5.3.7	Coarse-Grain MD Simulations of the PIFP in a POPC:POPS membrane	119
5.3.8	Atomistic MD Simulations of the PIFP in a POPC/POPS	120
5.4	Results	121
5.4.1	F103A and Q120A substitutions do not affect the PIFP ability to induce vesicle aggregation in a concentration-dependent manner	121
5.4.2	PIFP-Q120A has a higher partition coefficient than PIFP-F103A	122
5.4.3	The N-terminal phenylalanine residue is critical for the PIFP’s ability to induce lipid mixing and membrane leakage	122
5.4.4	MD simulations reveal that F103A and Q120A assemble into oligomeric structures stabilized by distinct interaction networks	124
5.4.5	Q120A and F103A substitution lead to decrease in water flux through the membrane	128
5.4.6	Substitution of the N-terminal phenylalanine residue significantly affects lipid tail protrusion	129
5.5	Conclusion	130
6	Computational microscopy of the viral entry process and the role of the fusion machinery	133

6.1	Abstract	135
6.2	Introduction	135
6.3	Methods	137
6.3.1	Test simulations to define the system size, composition, and simulation conditions	137
6.3.2	Simulations of viral fusion without fusion proteins	139
6.4	Results	141
6.4.1	Test simulation to define the system size, composition and simulation conditions	141
6.4.2	Simulations of viral fusion without fusion proteins	143
6.4.3	Simulations of viral fusion promoted by fusion proteins	145
6.5	Conclusions and future perspectives	146
6.6	Supplementary Information	149
7	General conclusions	155
	Bibliography	159

LIST OF FIGURES

1.1	Schematic representation of the viral life cycle of the SARS-CoV-2 virus . . .	3
1.2	Graphical representation of the two most common viral entry pathways . .	4
1.3	Schematic representation of the major structural features of the membrane fusion processes across the three canonical classes of fusion protein	6
1.4	Schematic representation of the SARS-CoV-2 viral particle binding the receptor cell ACE2 protein, which is attached to the target cell membrane.	8
1.5	ACE2 bound to glycosylated SARS-CoV-2 S protein	10
1.6	Amino acid changes to the spike (S) protein in SARS-CoV-2 variants of concern (VOCs)	11
1.7	Conformations adopted by the IFP and their effect in the membrane	14
1.8	PIFP secondary structure in perfusion state and inserted in membrane . . .	16
2.1	Molecular simulation methods used to probe different scales and levels of description of biomolecular systems	20
2.2	Schematic illustration of the bonded and nonbonded interactions in the potential force field systems	22
2.3	Periodic boundary conditions	24
2.4	Representation of Martini mapping examples	26
3.1	SARS-CoV-2 receptor binding domain structure	33
3.2	Two-dimension principal component analysis of SARS-CoV-2 RBD conformational dynamics in water	38
3.3	Closeup snapshots of SARS-CoV-2 RBD intramolecular interactions that stabilize the various conformations	39
3.4	S protein-antibody contact analysis.	52
3.5	RBD-ACE2 contact analysis correlated with the antibody binding data . . .	53
S3.1	ACE2 bound to glycosylated SARS-CoV-2 S protein	56
S3.2	RBD open – closed dynamics over simulation time	56
S3.3	Two-dimension principal component analysis of SARS-CoV-2 RBD’s mutants conformational dynamics in water	59

S3.4	Residue interaction networks (RINs) for the “open” and “closed” conformations of wt RBD	60
S3.5	Residue interaction networks (RINs) for the “open” and “closed” conformations of alpha RBD	60
S3.6	Residue interaction networks (RINs) for the “open” and “closed” conformations of beta RBD	61
S3.7	Residue interaction networks (RINs) for the “open” and “closed” conformations of delta RBD	61
S3.8	Residue interaction networks (RINs) for the “open” and “closed” conformations of omicron RBD	62
S3.9	Secondary structure difference between closed and open conformations of wt, alpha, beta, omicron and delta SARS-CoV-2 RBD simulated in water	63
S3.10	Snapshots representative of all wt RBD PCA basins	64
S3.11	Snapshots representative of all alpha RBD PCA basins	65
S3.12	Snapshots representative of all beta RBD PCA basins	66
S3.13	Snapshots representative of all delta RBD PCA basins	67
S3.14	Snapshots representative of all omicron RBD PCA basins	68
S3.15	RBD C α root-mean-square deviation (RMSD) moving average in solution	69
S3.16	Representation of the RBD of the three Omicron lineages and their specific mutations	70
S3.17	S protein-antibody contact frequency histogram	70
S3.18	Relevant residues for antibody and ACE2 binding	71
4.1	Parainfluenza fusion peptide features	76
4.2	Characterization of PIFP–lipid membrane interactions in vitro	85
4.3	Density profiles obtained from CG simulations of the PIFP	88
4.4	Free-energy surface of the PIFP in water obtained from Metadynamics calculations	90
4.5	PIFP dynamic helicity during atomistic MD simulations	91
4.6	Prevalence of peptide–peptide contacts	92
4.7	PIFP pore characterization	93
4.8	Lipid tail protrusion	96
S4.1	PIFP-induced LUV aggregates polydispersity index	100
S4.2	CG systems created to study the PIFP’s behavior in a POPC:POPG membrane	101
S4.3	Lipid order parameters for the lipid tails in the CG system	101
S4.4	Average number of contacts formed by each residue of the PIFPs and the water molecules	102
S4.5	Average number of contacts formed by each residue of the PIFPs and the membrane lipids	103
S4.6	Average number of contacts formed by each residue of the PIFPs and the heads of the POPG lipids	104

S4.7	PIFP helix geometry in AA simulations	105
S4.8	RMSD for the PIFP	106
S4.9	Superimposed snapshots of the nonameric AA (replicate 3) and CG simulations last frames	106
S4.10	Lipid head position with probability higher than 40%	107
S4.11	Water density in AA simulations	108
S4.12	Lipid order parameters for the lipid tails in the AA systems	109
S4.13	Bar plot of average number of contacts formed between the PIFP and the membrane lipids	109
S4.14	Average number of contacts formed by each residue of the PIFPs peptides and the membrane lipids	110
S4.15	Residue root mean square fluctuation in AA MD simulations	111
5.1	Parainfluenza fusion peptide in F protein context and after cleavage	117
5.2	PIFPs and PIFPs-vesicles aggregation profiles	122
5.3	Characterization of PIFPs-lipid membrane (un)binding in vitro	123
5.4	Characterization of PIFPs fusogenic properties in vitro	124
5.5	Snapshots of the initial and final frames of the CG and atomistic MD simulations	125
5.6	Residue interaction networks (RINs) for the peptide-peptide interactions . .	126
5.7	PIFP pore characterization	128
5.8	Lipid tail protrusion	131
6.1	Representation of the classical viral fusion process	136
6.2	Representation of the stages of membrane fusion in a sequential manner . .	137
6.3	Lipid and Protein Composition of Fusion Systems	138
6.4	Snapshots of the preliminary fusion system components	140
6.5	Initial and final frames of the preliminary viral fusion simulation	142
6.6	Schematic view of SARS-CoV-2 and parainfluenza viral fusion simulations without fusion proteins	144
6.7	Schematic view of SARS-CoV-2 and parainfluenza viral fusion simulations with fusion proteins	145
S6.1	Building the pre-bundle state of the SARS-CoV-2 and PIV fusion proteins . .	150
S6.2	Snapshots of the SARS-CoV-2 fusion simulations without fusion proteins . .	151
S6.3	Snapshots of the parainfluenza fusion simulations without fusion proteins . .	152
S6.4	Snapshots of the SARS-CoV-2 fusion simulations with fusion proteins . . .	153
S6.5	Snapshots of the parainfluenza fusion simulations with fusion proteins . . .	154

LIST OF TABLES

1.1	Examples of fusion peptides from different virus families. Adapted with permission from [35]. Copyright 2021 Frontiers Media S.A.	13
2.1	Advantages and limitations of using a coarse-grained approach, in molecular dynamics simulations	27
S3.1	Energy surface landscape analysis from 2D PCA of SARS-CoV-2 RBD conformational dynamics in water	57
S3.2	Energy surface landscape analysis from 2D PCA of SARS-CoV-2 RBD conformational dynamics in water	58
S3.3	Compilation of ACE2-RBD binding kinetics data from recent studies.	59
S3.4	Summary of antibodies studied in the S protein-antibody complexes	62
S4.1	List with detailed composition of the systems to be studied	100

GENERAL INTRODUCTION

1.1 Viral Infections: A Global Health Challenge

Viral infections present an important global health issue, impacting billions of people annually, with both immediate and long-lasting effects [2]. A wide range of viral diseases currently impact the global population, involving various pathogens with diverse transmission modes. Some well-known pathogenic viruses examples include influenza, HIV, measles, dengue, Zika, Ebola, and SARS-CoV-2 [3–5].

Viruses can target different organs, leading to a broad spectrum of symptoms and health implications. For instance, respiratory viruses are responsible for respiratory tract infections that can lead to issues ranging from common cold to severe respiratory diseases [6]. Hepatitis viruses primarily affect the liver, leading to chronic liver disease, cirrhosis, or even liver cancer [7], while neurotropic viruses target the central nervous system, resulting in conditions such as meningitis, encephalitis, or long-term neurological complications [8]. Viral infections are more prevalent and severe in young children, elderly people, pregnant women and individuals with compromised immune systems [9]. Regions with lower sanitation standards, crowded living conditions, and limited vaccination coverage are also more impacted by viral infections than more developed settings [10].

A wide range of hosts can be targeted by viruses, including animals, plants, and bacteria. In particular, viral outbreaks in other animals can have consequences equal to or more significant than viral infections in humans, both for the animals that are infected, as well as for human populations that are indirectly affected by these outbursts. The spread of these diseases can result in substantial economic losses, especially in the livestock industry, where outbreaks like avian influenza and foot-and-mouth disease can lead to food scarcity. Additionally, some animal viruses can cross over to humans, impacting both human health and the global economy [11]. An example of this is the COVID-19 pandemic, caused by a virus that initially infected bats and then spread to humans [12], demonstrating the potential of such viruses to not only affect animal health, but also pose significant threats to human populations.

Furthermore, new viral outbreaks continue to appear frequently. As of September

2023 a new outbreak of the Nipah virus has raised concerns in Kerala, southern India. The Nipah virus was initially identified more than two decades ago during an outbreak among pig farmers in Malaysia [13], which later spread to Singapore, resulting in nearly 300 cases and over 100 deaths. In Bangladesh, these outbreaks have become almost an annual occurrence, with studies linking them to the consumption of fermented date-palm sap contaminated with bat urine [14]. However, the precise origins of the virus and the mechanisms behind its transmission from bats to humans during the Kerala outbreak remain uncertain [15]. Moreover, the strain of the virus found in India and Bangladesh differs from the one in Malaysia; it can spread among humans and is more deadly. This emphasizes the need for vigilant monitoring of viral infections in animals and swift responses to such outbreaks.

Current global efforts to combat viral infections employ a complex approach that includes prevention, diagnosis, treatment, and public health interventions. Vaccination programs play a crucial role in these efforts by reducing their viral transmission, and protecting communities through group immunity [16, 17]. Yet, to this day, humans are still affected by several viral diseases, making it crucial to research innovative strategies to combat these infections. By deepening our understanding of the molecular details of viral infections, we can aim for a future, where prevention, diagnosis, treatment, and control measures are in place to safeguard our well-being.

1.1.1 Viral Entry: The door to Infection

For a virus to spread, it must come into contact with a susceptible host. This can happen through respiratory droplets, direct contact with infected individuals or surfaces (an example focusing on SARS-CoV-2 is shown in figure 1.1), ingestion of contaminated food or water, and vector-borne transmission (via insects or other organisms). Once the virus reaches the host, it attaches to specific receptors on the host cell's surface. This attachment is typically highly specific to the virus and host protein/receptor. After attachment, the virus enters the host cell, often through endocytosis or direct fusion with the plasma membrane. Inside the host cell, the virus's genetic material (DNA or RNA) undergoes replication and transcription to generate the necessary components for viral replication. The replication process varies among viruses, some rely on the host cell's machinery for replication, while others carry their own enzymatic tools. As a result, new viral particles are assembled, and these newly formed viruses accumulate within the host cell. Ultimately, they are released from the host cell, either through cell lysis, causing cell rupture, or, in certain instances, via budding from the cell membrane, a process that does not immediately lead to the death of the host cell [18–20].

Viral release from the host cell depends largely on how their genetic material is shielded from the outside. When the viral capsid or nucleocapsid is surrounded by a lipid bilayer envelope derived from the host cell or endoplasmic membrane we consider this an enveloped virus (as seen in figure 1.2b). The membrane allows enveloped viruses to leave

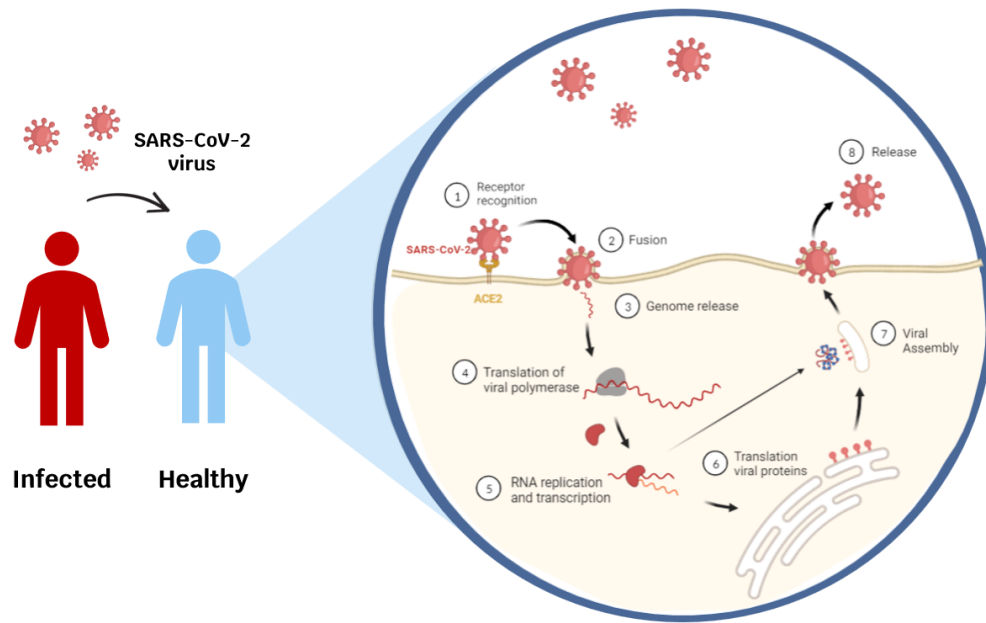


Figure 1.1: Schematic representation of the viral life cycle of the SARS-CoV-2 virus. A healthy individual enters into contact with an infected person. When the virus reaches a susceptible host, it binds to angiotensin-converting enzyme 2 (ACE2) receptors on the host cell's surface. Following attachment, the virus typically enters the host cell through membrane fusion. Within the host cell, the virus leverages the cellular replication machinery, leading to the assembly of new viral particles using duplicated genetic material and viral proteins. These recently formed viruses accumulate within the host cell and are subsequently released. Figure created with BioRender . com.

host cells by budding, being discharged without causing rapid cell lysis [21]. Examples of enveloped viruses include influenza, HIV, SARS-CoV-2, parainfluenza and dengue. Conversely, when the viruses lack a lipid envelope, and are made up solely of a protein capsid that protects the viral DNA, they are called non-enveloped viruses (figure 1.2a). These viruses usually leave host cells by lysis, damaging the host cell membrane [22]. Examples of non-enveloped viruses include adenoviruses, noroviruses, and rhinoviruses.

Shifting the focus to the initial stages of viral entry, this is a crucial phase in the establishment and progression of the infection. Upon binding to receptors, which can be proteins, carbohydrates, or lipids, viruses utilize two main entry routes: endocytic or non-endocytic pathways. The endocytic route involves the recognition of cell surface receptors, a step that triggers the internalization of the virus into the host cell (Figure 1.2b). This is the preferential mode of viral entry for non-enveloped viruses, however several enveloped viruses such as influenza, dengue and ebola utilize this mechanism to enter the host cell [23–26]. The non-endocytic route enables viruses to directly penetrate the plasma membrane under neutral pH conditions (Figure 1.2b) [27–30]. Enveloped viruses such as parainfluenza utilize this mechanism to enter the host cell, and some, such as the SARS-CoV-2 or HIV [31, 32], can utilize both the non-endocytic and endocytic pathways

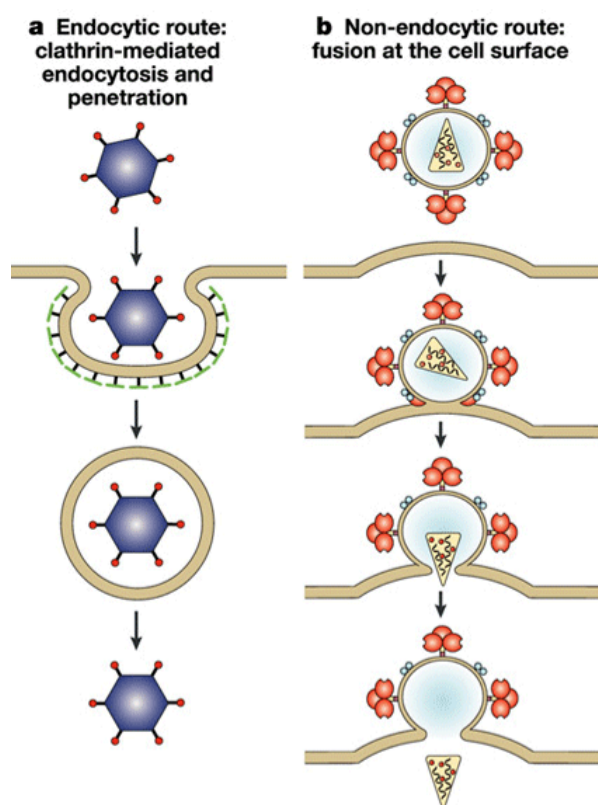


Figure 1.2: Graphical representation of the two most common viral entry pathways: Endocytic (A) and non-endocytic (B). The virus depicted in the endocytic pathway is non-enveloped and on the non-endocytic pathway an enveloped virus is illustrated. Reprinted with permission from [18]. Copyright 2004 American Chemical Society.

for cellular entry.

An essential process involved in the entry of enveloped viruses both through endocytic and non-endocytic routes is membrane fusion. During viral entry, once the viral and cellular or endocytic membranes approach each other, specific membrane proteins facilitate their fusion.

Successful entry into host cells requires significant conformational changes in the viral entry proteins or the host-cell receptors. These conformational changes can be triggered by various factors, such as the acidic environment within endosomes or proteolytic cleavage, in cases of non-endocytic entry pathways. Ultimately, these conformational alterations in the viral entry proteins lead to either the penetration of the virus into the host cell (for non-enveloped viruses) or the fusion of the viral envelope with the host cell membrane (for enveloped viruses). This fusion process marks a critical point in the virus's journey, enabling it to establish infection within the host cell.

1.1.2 Enveloped Viruses: The Fusion Protein is key

In this thesis, we focus on the study of enveloped viruses, which are characterized by the presence of a lipid envelope surrounding their genetic material. This envelope is acquired during the process of viral budding or release from the host and is covered by viral glycoproteins, which play crucial roles in viral entry and immune evasion [33].

Amongst these viral glycoproteins we can find the fusion proteins that are responsible for catalyzing the fusion of viral and host cell membranes during viral entry. As mentioned earlier, these fusion proteins undergo conformational changes essential for their function, when triggered by specific conditions, which result in a transition from a pre-fusion to a short-lived fusion-competent, and finally to post-fusion state. A fusion protein, initially in a metastable pre-fusion state, is primed for activation upon encountering the appropriate cues: When the virus approaches the host cell and conditions such as a reduction in pH within endosomes or proteolytic cleavage (particularly in non-endocytic entry) trigger these changes, the fusion proteins rapidly transition to their fusion-competent state, which allow them to insert the fusion peptide into the host membrane, and finally they undergo another change that leads to the post-fusion state. The transition from the pre-fusion to the fusion-competent state involves the reorganization of protein domains, exposing hydrophobic regions that were previously shielded. The exposed hydrophobic regions, also known as fusion peptides, then interact with the host membrane initiating the fusion process. This interaction lowers the kinetic barrier that must be overcome for the two membranes to fuse. The final conformational transition where the protein folds back on itself forming the post-fusion state leads to the merger of the virus and host cell membranes, effectively creating a pore or channel through which the viral genome can be released into the interior of the host cell.

Viral fusion proteins differ mainly in their structural organization and mechanisms of membrane fusion and can be divided into three classes:

Class I fusion proteins are found in several viruses such as influenza, HIV, and parainfluenza. They are trimers in both the pre- and post-fusion states (Figure 1.3A), and share common structural features [34]. In the post-fusion state, they possess a central N-terminal trimeric α -helical coiled coil, surrounded by three C-terminal helices, forming what is known as a six-helix bundle (6HB). The specific size and position of the 6HB can vary among different Class I fusion proteins. The fusion peptide region is commonly located at or near the N-terminus of each monomer, but the precise location can also differ between viruses. In the case of influenza, for example, the fusion process starts with the recognition of the target protein, followed by cleavage and internalization through the endocytic pathway. The low pH of the endosome then triggers conformational changes, which drive fusion. In the case of the SARS-CoV-2 virus, as mentioned previously, it can also enter the cell through the endocytic pathway, but the most common mechanism is by fusing directly with the cell membrane. In this case, there is recognition of target ACE2 protein followed by cleavage and exposure of the fusion peptide region [31]. Despite

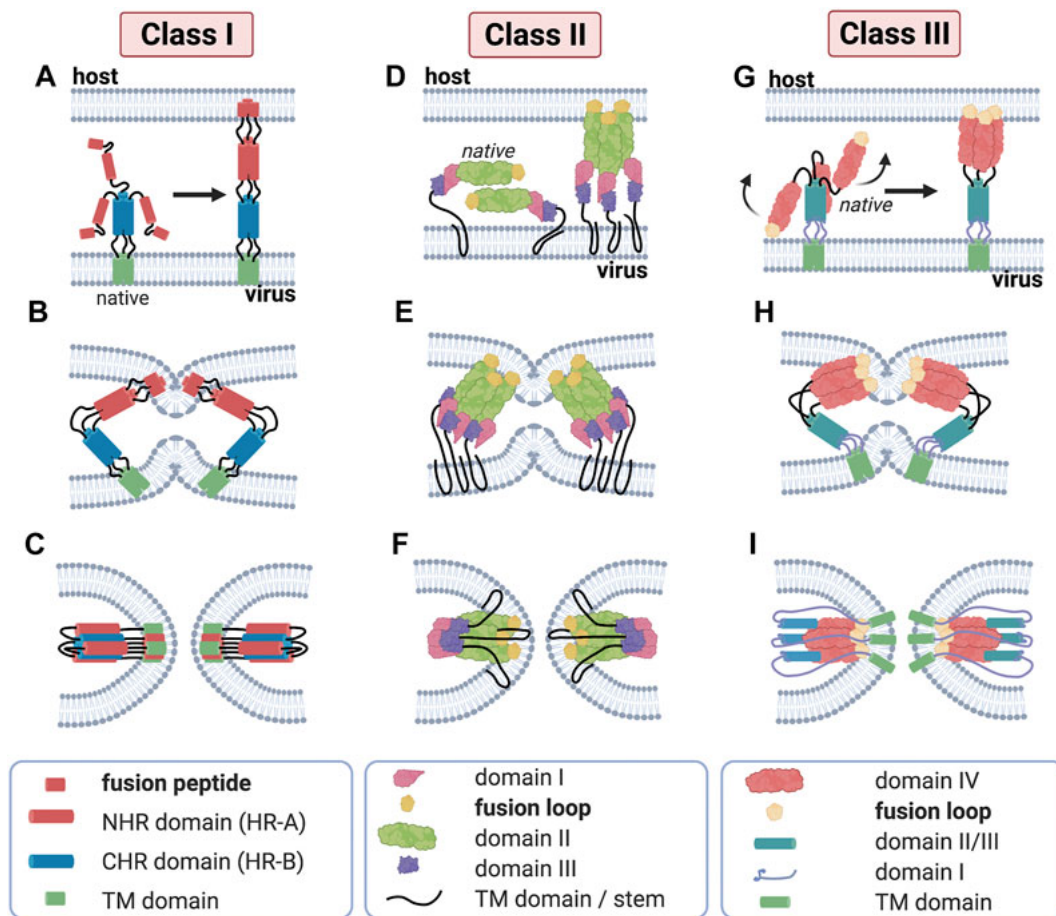


Figure 1.3: Schematic representation of the major structural features of the membrane fusion processes across the three canonical classes of fusion protein - Class I, Class II, and Class III. Reprinted with permission from [35]. Copyright 2021 Frontiers Media S.A.

slight differences between various Class I fusion proteins, they are all metastable trimers, with the fusion peptide protected until triggered by an external factor. Upon exposure, the fusion peptide inserts into the target cell membrane and conformational changes in the fusion protein facilitate fusion between the target and viral membranes (Figure 1.3B).

Class II fusion proteins include those found on flaviviruses and alphaviruses. They are initially associated with a chaperone protein, that is cleaved during or soon after viral assembly. Unlike class I fusion proteins, which are already in trimeric form before fusion, class II proteins are dimers (Figure 1.3D) that consist primarily of β -sheet structures and the fusion peptide region is a loop found at the tips of the β -strands. In their pre-fusion state, the ectodomains of Class II fusion proteins exist as anti-parallel dimers that lie flat along the virion surface. They rely on the acidic environment of the endosome to rearrange the dimers into trimers (Figure 1.3D, right side), leading to the exposure of the fusion peptide region and subsequent membrane fusion (Figure 1.3E,F) [36].

Class III proteins share features from both Class I and Class II fusion proteins. They exist as trimers in their pre-fusion state and contain a central α -helical coiled-coil, similar

to Class I proteins (Figure 1.3G). Their fusion domains are similar to those of Class II proteins, as their fusion loops are located at the tips of extended β -strands. Upon acidification, the protein undergoes a series of conformational changes that position the fusion loops in the vicinity of the target membrane (Figure 1.3H). Subsequently, the protein bends back, reorienting the C-terminal portion to become anti-parallel to the N-terminal segment, bringing the viral and target membranes together for fusion (Figure 1.3I) [37]. The structural characteristics and conformational changes of Class III fusion proteins make them an intriguing group of viral fusion proteins with unique features compared to Class I and Class II proteins.

Given their essential role in viral entry, fusion proteins are attractive targets for antiviral drugs. Various strategies have been developed to inhibit fusion protein-mediated membrane fusion, including the design of molecules that interfere with the conformational changes required for fusion. For example, fusion inhibitor drug enfuvirtide, or T20 [38], blocks structural changes in the HIV fusion protein and has been widely used in combination with other drugs for the treatment of HIV since March 2003. Another example is the IPB02 fusion inhibitor, which showed the ability to inhibit SARS-CoV-2 S protein-mediated cell-cell fusion and pseudovirus transduction in *in vitro* studies [39].

1.1.3 Target Cell Recognition: Finding the lock

The first crucial stage in viral infection is the recognition process between fusion proteins and specific receptors located on the surface of target cells. The strength and selectivity of the fusion protein-receptor interplay has a large impact on the efficiency of viral entry into host cells. A strong binding affinity between fusion proteins and receptors enhances the efficiency of viral entry, making it easier for the virus to invade host cells [40, 41]. The main cell receptors involved in viral entry are proteins or other molecules, such as sialic acid molecules and heparan sulfate proteoglycans, located on the cell surface. These receptors are often utilized by viruses to initiate the process of attachment and entry into host cells [20, 42]. The efficiency of the fusion protein-receptor interaction can be influenced by various factors, including the abundance and distribution of receptors on target cells [43–45], the conformational changes induced in the fusion protein upon receptor binding [44–46], and the availability of co-receptors or other molecules that assist in the fusion process [47]. Furthermore, viral pathogens have evolved diverse strategies to enhance the fusion protein-receptor interaction and increase their infectivity. For example, some viruses employ multiple fusion proteins or utilize co-receptors to improve the efficiency of viral entry [28, 48].

The cell receptors that a virus can bind to, play a critical role in determining the range of cell types and tissues it can target within a host organism. Different viruses may utilize distinct receptors, or even the same receptor family, to achieve entry. This diversity in receptor usage is a key factor in understanding how viruses establish infections and the spectrum of cells and tissues they can affect. For example:

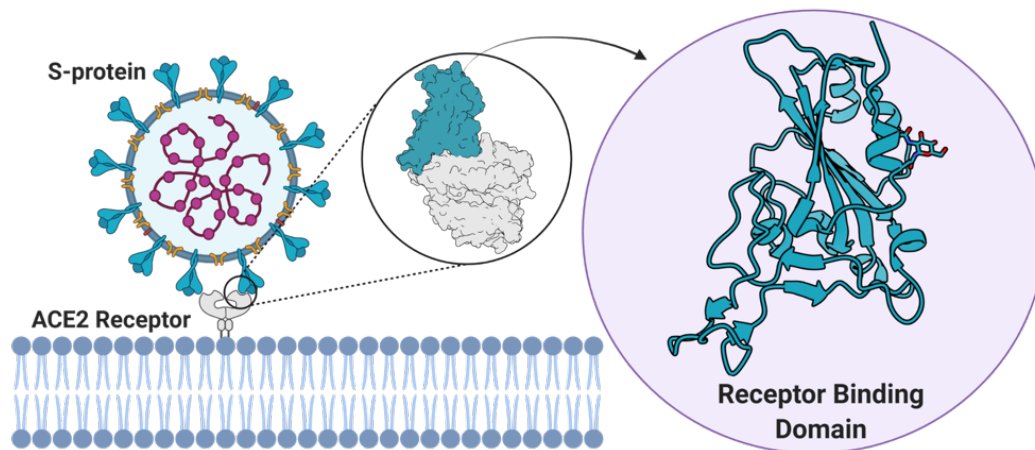


Figure 1.4: Schematic representation of the SARS-CoV-2 viral particle binding the receptor cell ACE2 protein, which is attached to the target cell membrane. Target cell recognition is achieved by the RBD region, which is shown in higher detail on the right side of the scheme. Figure created with BioRender.com.

- Influenza viruses enter target cells through receptor-mediated endocytosis, following binding to sialic acid-containing receptors [49, 50]. Although interactions with these receptors were initially thought to be the sole factor in entry, recent discoveries reveal contributions from C-type lectin receptors (CLRs) such as dendritic cell-specific intercellular adhesion molecular 3-grabbing non-integrin (DC-SIGN or CD209) and L-SIGN (CD209L) [51, 52].

- Parainfluenza viruses are also known to bind sialic acid residues on the host cell surface [53]. This role is played by the hemagglutinin-neuraminidase, that upon binding to its cell receptor, triggers conformational changes in the fusion protein F, allowing it to reach its fusion state [54, 55].

- The SARS-CoV-2 virus utilizes the ACE2 receptor for entry into human cells, primarily those of the respiratory tract (Figure 1.4) [56, 57]. In addition to ACE2, several molecules have been suggested to serve as alternative receptors for SARS-CoV and SARS-CoV-2. These include C-type lectins, DC-SIGN and L-SIGN [58–60].

These three viruses target primarily respiratory epithelial cells due to their specific cell tropism. However, in other viruses such as the HIV, its primary target cells are CD4⁺ T cells and macrophages. This specificity arises from the requirement of HIV to bind to the CD4 receptor [61], which is highly expressed on CD4⁺ T cells, along with co-receptors like CCR5 or CXCR4.

Even though the precise receptors and co-receptors involved may differ, the specificity for certain cell types is a fundamental aspect of their infection strategies. This interaction between viral surface proteins and host cell surface molecules determines their cell tropism and plays a pivotal role in the pathogenesis of these viral infections.

1.1.3.1 The case of the SARS-CoV-2 receptor recognition

In recent years, SARS-CoV-2 has been the subject of intensive study, serving as a noteworthy example of the initial steps in target cell recognition by fusion proteins. The first structures of its fusion protein in its prefusion [46, 62] and post-fusion [63] states were elucidated through Cryo-EM. The SARS-CoV-2 fusion protein, also known as S protein, belongs to the class I fusion protein family and is synthesized as a single 1273 amino acid polypeptide chain. It forms a trimeric structure, with each monomer consisting of two subunits, S1 and S2 [64]. These subunits serve distinct functions in receptor binding (S1) and membrane fusion (S2). Specifically, the S1 subunit comprises an N-terminal domain (residues 14–305) and a receptor binding domain (residues 319–541). In its prefusion state, the S protein assembles into a homotrimer and undergoes significant structural alterations to control the exposure and accessibility of the RBD. This involves a dynamic "up" and "down" mechanism, wherein the RBD transitions from a receptor-accessible "up" conformation to a receptor-inaccessible "down" conformation [46, 62, 65].

The RBD plays a pivotal role in facilitating the interaction between SARS-CoV-2 and host cells through its binding to angiotensin-converting enzyme 2 (ACE2), a crucial regulator within the renin-angiotensin system [56, 64, 66, 67]. This binding event represents one of the initial steps in the primary pathway of viral entry for SARS-CoV-2. Consequently, the dynamic transition of the RBD's position from "down" to "up" holds paramount significance in this process [62, 68, 69]. Simulation studies have elucidated the conformational transition process of the SARS-CoV-2 spike glycoprotein, shifting between its "down" and "up" states [70]. These studies have identified a "semi-up" intermediate state and suggest that this intermediate state is essential for the recognition of the host cell target ACE2.

Another key structural feature that is relevant for ACE2 binding is the protein's extensive glycosylation [71] (Figure 1.5 A). Protein glycosylation plays a crucial role in viral pathogenesis [72–74], as demonstrated by the characteristic N-glycan coating of the viral fusion proteins [75–77]. The SARS-CoV-2 spike has 22 predicted N-glycosylation sites per monomer [46, 77], and 20 out of 22 SARS-CoV-2 N-linked glycosylation sites are conserved in the SARS-CoV spike protein [46] (Figure 1.5A). While most of the S protein surface is densely glycosylated, shielding it from host defense mechanisms, the RBD itself contains only a single glycosylation site [78], N343, which is located relatively distant from the ACE2-RBD interface (Figure 1.5 B).

In addition to the Cryo-EM structures that reveal the location of the glycosylation on the S protein [46, 62], several works have contributed to the understanding of their role in the protein dynamics. In Casalino *et al.* long all-atom molecular dynamics simulations reveal the essential role of N-glycans at sites, N165 and N234, in modulating the conformational dynamics of the S protein RBD. N234A and N165A mutations show increased instability of RBD in the "up" state and are supported by biolayer interferometry results, where the deletion of these glycosylation sites reduces binding to ACE2 [78].

Furthermore, since 2020, several new SARS-CoV-2 variants of concern (VOCs) have

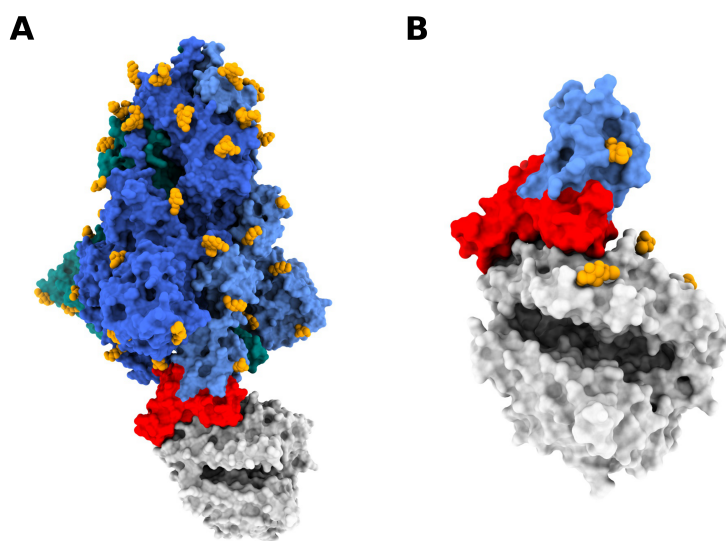


Figure 1.5: ACE2 bound to glycosylated SARS-CoV-2 S protein. (A) ACE2 (White) bound to SARS-CoV2 S glycoprotein (blue). (B) Closeup of ACE2 bound RBD. RBM is rendered in red. Glycosylations are rendered in orange. Reprinted from Valério *et al.*, 2022 [79].

emerged, including: variant B.1.1.7 (Alpha) [80]; B.1.351 (Beta) [81]; P.1 (Gamma) [82]; B.1.617.2 (Delta) [83]; B.1.1.529 (Omicron) [84, 85] and Omicron subvariants BA.1, BA.2, BA.3, BA.4. and BA.5 [86]. Knowing that certain mutations may affect glycosylation attachment sites, several studies have compared the site-specific glycosylation between the variants. Both N- and O-glycosylation sites were found to be highly conserved among the S protein mutant variants, particularly at the sites on the RBD [87–89]. Also, newer variants displayed a slight increase in oligomannose and a decrease in highly complex glycans, which can be correlated with the increased transmissibility of the VOCs [88].

The SARS-CoV-2 RBD-ACE2 complex has also been extensively studied, with structural biology studies revealing which residues in the SARS-CoV-2 RBD are essential for ACE2 binding [90–92]. The majority of these are either highly conserved or share similar side chain properties with those in the SARS-CoV RBD [90].

These are supported by computational studies that have shed light on the intricate mechanisms underlying the binding of the RBD to the ACE2 cell receptor [93–96], as well as potential applications in the pharmaceutical realm [97–102]. This knowledge is invaluable in the quest to develop effective antiviral treatments for SARS-CoV-2, particularly identifying amino acid pairs that interact strongly at the interface between the S protein and ACE2.

On this matter, several works have contributed to uncovering critical hotspots within the ACE2 receptor, notably K31 and K353 [103–105]. On the RBD side, interestingly, some of hotspot residues involved in binding, including K417, F486, Q493, and N501 [105, 106], have been subject to mutations in the variants (Figure 1.6):

- **N501Y**: Is present in the alpha, beta, and gamma variants and has been shown to

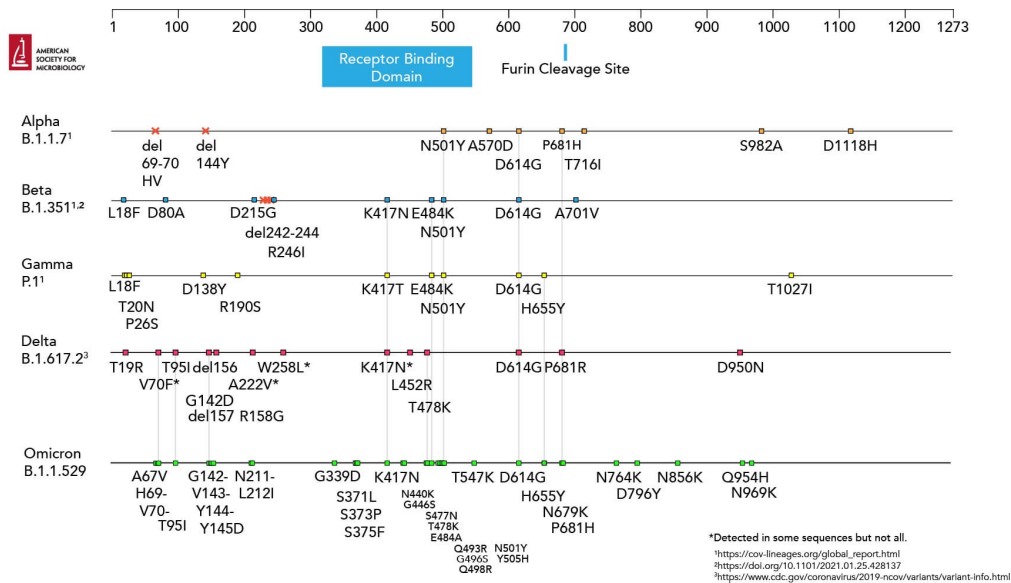


Figure 1.6: Amino acid changes to the S protein in SARS-CoV-2 variants of concern (VOCs). Reprinted from American Society of Microbiology [110].

enhance the interaction between the RBD and ACE2 receptor [106, 107].

- **K417N**: Found in the beta variant, the K417N mutation is associated with a reduction in the binding strength between the RBD and ACE2 receptor [105].

- **K417T**: Identified in the gamma variant, is also linked to a decrease in binding strength between the RBD and ACE2 receptor [108].

Other common mutations found in the RBD of SARS-CoV-2 variants include:

- **E484K**: Found in the beta and gamma variants. This mutation has been described to increase the binding strength of the RBD-ACE2 complex [108] and, since position 484 is part of the most important site for virus recognition by neutralizing antibodies, it also significantly reduces the binding affinities for some neutralizing antibodies/nanobodies [109].

- **L452R and T478K**: Both mutations were identified in the Delta variant. Their simultaneous occurrence has been showed to increase the stability and intra-chain interactions in the S protein, which may change the interaction ability of neutralizing antibodies [101].

With the dawn of the AI field, a recent study has demonstrated the remarkable capability of machine learning algorithms to be trained using molecular dynamics (MD) simulations of the binding process between the S protein and ACE2 proteins [111]. These algorithms can effectively predict the binding affinity of different SARS-CoV-2 variants and the ACE2 protein with a high degree of accuracy.

Even though the RBD-ACE2 binding dynamics have been well studied, the dynamics of the RBD by itself have not received as much scrutiny. From the available experimental structural data, it is known that the core β -sheet structure is very stable, but the receptor binding motif (RBM) seems to be quite dynamic and not as structurally defined, unless bound to other proteins, like ACE2 [30, 67, 90, 112, 113] or antibody fragments [114–120].

The existing MD simulation research of the RBD by itself are limited. Some studies focus on short simulations [121–123], while others did not delve into the RBM dynamics in depth [78, 124–126]. Consequently, there is a lack of knowledge regarding the conformational dynamics of this motif and the influence of point mutations when the RBD is not bound to its receptor.

1.1.4 Viral membrane fusion: the role of Fusion Peptide

Following target cell recognition, fusion proteins undergo structural changes that lead to the unveiling of the fusion peptide. This region is typically located at the N-terminus of the fusion protein and plays a pivotal role in facilitating the insertion of the fusion protein into the host cell membrane [33, 127, 128]. For class I and III fusion proteins the proteolytic cleavage event is crucial for activating the fusion protein. This allows for the removal of inhibitory regions or structural constraints, allowing the fusion peptide to be exposed and interact with the target cell membrane [127, 128]. This cleavage step is often triggered by the target cell recognition or environmental conditions encountered during viral entry.

While the sequence of the fusion peptide can differ among various viruses, there are certain noteworthy characteristics that are often conserved. For instance, the fusion peptide commonly contains hydrophobic amino acid residues, allowing it to engage with the lipid bilayer of the host cell membrane [33] (Table 1.1). This hydrophobic nature is essential for the fusion peptide to insert into the membrane and initiate the process of membrane fusion. The amino acid conservation within the fusion peptide among viruses of the same family is also a common feature and highlights its functional importance in viral fusion [33]. Consequently, mutations or alterations to the fusion peptide sequence can have a significant impact on viral infectivity.

Another characteristic shared by fusion peptides is the abundance of alanine and glycine residues [143] (Table 1.1). These residues are more abundant in structures with intrinsic conformational flexibility than in hydrophobic protein domains. This may indicate that sequences of fusion peptides have evolved under selective pressures that extend beyond the constraints imposed by mere hydrophobicity. Aromatic residues are also commonly found in FP sequences. These residues may aid in overcoming the energy cost of peptide partitioning into membranes, especially in non N-terminal fusion peptides (Table 1.1). The lack of a free N-terminus constrains conformational changes in interaction with the membrane surface in these cases [144]. The favourable interactions of aromatic side chains with phospholipids may also help to stabilize the FP sequence insertion into the top bilayer leaflet [145].

Fusion peptides are generally assumed to migrate from the aqueous solution to target cell membranes through spontaneous partitioning, driven by their hydrophobic nature. However, when synthetic fusion peptides are studied in aqueous environments, they tend to aggregate. Under typical experimental conditions, the aggregation process in solution generally proceeds more slowly than the binding to membranes, thus enabling the

1.1. VIRAL INFECTIONS: A GLOBAL HEALTH CHALLENGE

Class I	Retroviridae	HIV-1	AVGIGALFLGFLGAAGSTMGARS	gp41	[129–132]	
		FIV	AAIHVMLALATVLSIAGAGTGATA	gp36	[129]	
		HTLV-1	AVPVAVWLVSALAMGAGVAGGITGS	gp21	[131]	
		SIV	GVFVLGFLGLATAGSAMGAAS	gp32	[131, 133]	
		ASLV	GPTARIFASILAPGVAAAQALREIERLA	EnvA	[131]	
		BLV	SPVAALTLGLALSVGLTGINVAVS	gp30	[131]	
	Orthomyxoviridae	Influenza A	GLFGAIAGFIENGWEGMIDG	HA	[129–131]	
		Influenza B	GFFGAIAGFIEGGWEGMIAHGHY	HA	[129]	
	Filoviridae	Ebola	GAAIGLAWIPYFGPAAEGIYTEGL	GP2	[130–132, 134]	
		MARV	LAAGLSWIPFFGPGI	GP2	[131]	
	Paramyxoviridae	Sendai	FAGVIGTIALGVATSAQITAGIA	F1	[129, 130, 135]	
		SV5	FAGVVIGLAALGVATAAQVTAAVA	F1	[129, 130, 135]	
		SV41	VSANQAGSRKRREFAGVVVGLAALGVATAAQ	F1	[129, 130, 135]	
		Measles	FAGVVLAGAALGVATAAQITAGIA	F1	[129, 130, 135]	
		HeV	LAGVVMAGIAIGIATAAQITAGV	F	[131]	
		NDV	FIGAIIGSVALGVATAAQITAA	F	[131]	
		PIV5	FAGVVIGLAALGVATAAQVTAVALV	F	[131]	
		Coronaviridae	SARS-CoV	MYKTPTLKYFGGFNFSQILPDPFL (N-ter)	S	[129, 131]
			GAALQIPFAMQMAYRF (internal)			
	SARS-CoV-2		SFIEDLLFNKVTADAGFMKQY (N-ter)	S2	[136–138]	
			SFIEDLLFNKVTADAGFIKQYGDCLGDIAARDLICAQK (N-ter)	S2	[137, 139, 140]	
	Class II	Togaviridae		MIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRF (internal)	S2	[136, 141]
SFV			VYTGVPFMWGGAYCFDCDS	E1	[130, 131]	
Flaviviridae		CHIKV	VYPFMWGGAYCFCDTENT	E1	[131]	
		DNV	DRGWGNGCGLFGKGS	E	[130–132, 135]	
		WNE	VDRGWGNGCGLFGKGSIDTCAKFACSTKAIGR	E	[135]	
		JE	TDRGWGNGCGLFGKGSIDTCAKFSCSTKAIGR	E	[135]	
		YF	SDRGWGNGCGLFGKGSIVACAKFTCAKSMSLF	E	[135]	
		SLE	VDRGWGNGCGLFGKGSIDTCAKFTCKNKATGK	E	[135]	
Herpesviridae		TBEV	CGLFGKGSIVACVKAAC	E	[135, 142]	
		HSV	WFGHRY	gB	[130–132]	
Class III		Rhabdoviridae		VEAF		
			VSV	WY	G	[131, 132]
	Baculoviridae		YA			
		AcMNPV	YAYNGGSLDPNTRV	gp64	[131, 132]	
		VKRQNNNHFAHHTCNK				

Table 1.1: Examples of fusion peptides from different virus families. Adapted with permission from [35]. Copyright 2021 Frontiers Media S.A.

efficient incorporation of these peptides into the desired target membranes. Nevertheless, the formation of peptide aggregates presents a challenge when experimentally evaluating the equilibrium between water and membrane partitioning for these peptides [146–148]. To address this challenge, Han and Tamm pioneered an approach involving influenza fusion peptides linked to a polar tetra-lysine moiety [147, 148].

The insertion depth and angle of fusion peptides inside the membrane can vary due to factors such as peptide length and membrane lipid composition, especially the presence of anionic phospholipids or cholesterol [149–151].

1.1.4.1 Influenza fusion peptide

The effect of peptide length on the FPs membrane insertion is particularly evident in the case of the influenza fusion peptide (IFP), a widely studied fusion peptide that serves as a benchmark for understanding other N-terminal fusion peptides.

NMR studies of the 20-residue long IFP in a lipid environment have provided insights into its structural behavior. Under fusogenic conditions (pH 5), the peptide adopts a helical inverted V structure when placed in detergent micelles. However, at pH 7.4, it takes on a less structured conformation in the C-terminal region [152]. In contrast, recent research has shown that a IFP with only 3 residues more (23 residues instead of 20), forms

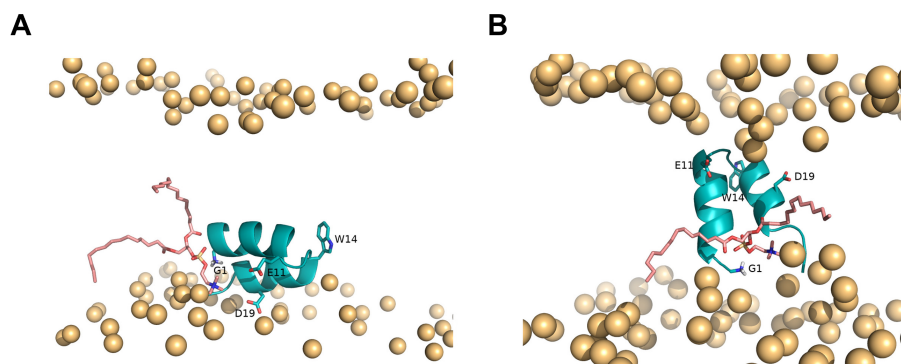


Figure 1.7: Conformations adopted by the IFP and their effect in the membrane. IFP conformations obtained in constant-pH MD simulations [159], starting from two distinct conformations (labeled as horizontal and vertical) obtained using a self-assembly approach [160]. (A) Illustration of a lipid tail protrusion event promoted by interaction of a lipid with the peptide N terminus, observed in the constant-pH MD simulations performed at pH 5 starting from the horizontal conformation [159]. (B) Illustration of a lipid tail protrusion event promoted by interaction of a lipid with the peptide N terminus, observed in the constant-pH MD simulations performed at pH 5 starting from the vertical conformation [159]. Reprinted with permission from [161]. Copyright 2021 FEBS Press.

a tightly wound helical hairpin structure at both pH 4 and 7.4 [153]. This highlights how the addition of just three amino acid residues, which are conserved within the fusion peptide family [154], can significantly influence the fusion peptide's structure and its interaction with membranes.

Molecular dynamics simulations have also played a pivotal role in exploring the structural characteristics and orientation of IFP within lipid membranes. In many studies, the fusion peptide is initially positioned at the interface between lipid head groups and lipid tails, following NMR suggestions. In these simulations, the peptide generally remains in this region throughout the simulation duration [155–158]. However, this approach may not fully capture the peptide's dynamic behavior because it restricts its ability to explore alternative states due to substantial energy barriers associated with such transitions.

To overcome this limitation, Victor *et al.* employed a self-assembly approach. In this approach, lipids and water molecules start the simulation from random positions and spontaneously assemble around the peptide during the simulation [160]. In most replicates, the peptide adopted a membrane-spanning helix-turn-helix configuration (Figure 1.7B), although one replicate remained at the interface between lipid head groups and lipid tails (see Figure 1.7A). Replica-exchange molecular dynamics simulations further supported the idea that the influenza FP can exist in two distinct configurations, with the membrane-spanning conformation representing the lowest free energy state for the 23-residue fusion peptide with a charged N-terminus [162, 163].

The membrane-spanning conformation of the IFP also leads to a stronger perturbation of the membrane and the splaying of the lipid acyl chains, known as lipid tail protrusion

event [160, 164] (Figure 1.7, lipid in pink).

1.1.4.2 Parainfluenza fusion peptide

Another similar fusion peptide, this one less studied, is the parainfluenza fusion peptide (PIFP). Studies show that when concealed within the its fusion protein this region exhibits a combination of β -sheet and α -helical features [165] (Figure 1.8B). Upon activation and insertion into the target membrane, the PIFP extends into a helix (Figure 1.8). Other studies also underline the functional role of the α -helical conformation. When mutating the glycine residues of the PIFP there is an increase in the fusion activity of transfected cells [166], and augmented helical content of the FP [167].

Further details of the conformational dynamics of the PIFP emerged from solid-state nuclear magnetic resonance investigations. These studies indicate that the PIFP can adopt either α -helical or β -strand conformations, depending on the lipid tail composition of the membrane in which it is inserted [149] (Figure 1.8A).

Despite these insights, a crucial question remains: how does the PIFP facilitate membrane fusion? Data from interactions between six PIFPs in phospholipid micelles indicate a cooperative and specific assembly of PIFPs into hexamers within membranes [168]. Short molecular dynamics simulations of the hexameric structure further revealed the penetration of water into its core from the viral side of the membrane [168]. To gain deeper insights, it is essential to explore whether these water molecules fully cross the membrane and how the number of peptides inserted into the membrane influences their oligomerization.

1.2 Simulation as a tool to better understand viral fusion

Simulation studies have emerged as valuable tools for unraveling the molecular mechanisms underlying viral fusion. By utilizing computational models and simulations, researchers have gained insights into the dynamic behavior of fusion proteins and their essential role in the viral fusion process.

Simulations offer a unique opportunity to investigate the intricate details of fusion protein conformational changes [170, 171], and membrane fusion events at the molecular level, complementing experimental studies effectively [172, 173].

As previously mentioned, molecular dynamics simulations have played a pivotal role in understanding the involvement of the FP in the fusion process. These simulations have shed light on how the fusion peptide interacts with a lipid bilayer, elucidating the mechanisms behind its insertion into the host cell membrane, the destabilization it induces, and its role in facilitating membrane fusion [158–160, 168, 174, 175].

Regarding the full-size fusion protein, numerous simulation studies have played an important part in deepening our understanding of its role in viral fusion and giving

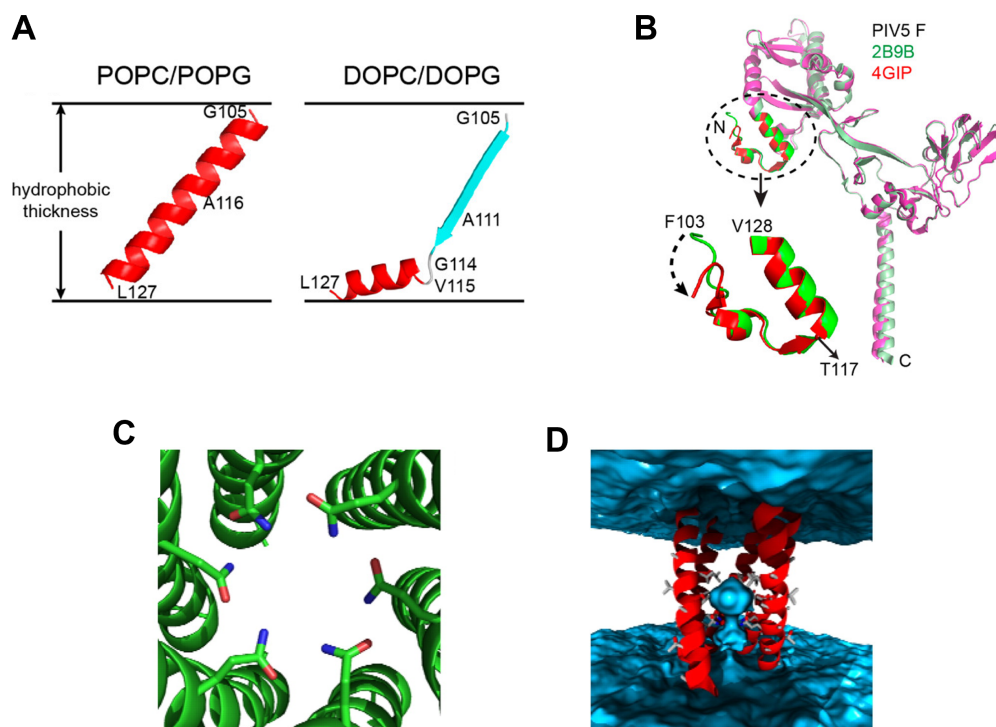


Figure 1.8: PIFP secondary structure in perfusion state and inserted in membrane. (A) PIFP structure determined by SS-NMR in POPC/POPG and DOPC/DOPG bilayers. The peptide is fully α -helical in POPC/POPG bilayers, but adopts a mixed strand/helix conformation in DOPC/DOPG bilayers [149]. (B) Prefusion crystal structures of the PIV5 F protein in the uncleaved (green) [165] and cleaved (red) [169] states. Figures (A) and (B) were reprinted with permission from [149] Copyright © 2014 American Chemical Society. (C) Snapshot of PIFP hexameric bundle obtained from a short MD simulation (50 ns). The Q120 residues form hydrogen bonds with one another as well as waters on the interior [168]. (D) Side view of the transmembrane PIFP bundle with the N-terminal side up. Waters are shown in blue, highlighting the penetration of water into the core of the 6HB from the viral side of the membrane [168]. Figures (C) and (D) were reprinted with permission from [168]. Copyright © 2011 National Academy of Sciences.

insights into the dynamic structural changes that they undergo during the fusion process [170, 171].

In summary, simulation studies have proven an essential asset in unraveling the intricate details of viral fusion and will be employed in this thesis with the same goal.

1.3 Dissertation objectives and outline

This dissertation aims to fill gaps in the existing literature on viral fusion, with a particular focus on events associated with target cell receptor recognition and membrane

fusion. To achieve this, we focus on two viruses: SARS-CoV-2 and parainfluenza. In the case of SARS-CoV-2, we examine the influence of naturally occurring mutations on the dynamics of the SARS-CoV-2 RBD and antibody recognition. In parallel, the parainfluenza virus served as a model to study the role played by the fusion peptide in facilitating the fusion of viral and host cell membranes, dissecting the impact of specific point mutations on this crucial step. To gain a more integrated perspective of fusion, we constructed (near full-scale) models of viral fusion for both SARS-CoV-2 and parainfluenza and studied the function of fusion proteins in this process.

This research journey unfolds over five chapters, each offering distinct insights, as outlined below.

In **Chapter 2**, we start with a broad general introduction to Molecular Dynamics, the main method used and explored in all other chapters throughout this thesis.

In **Chapter 3** we describe how the occurrence of naturally occurring mutations in the SARS-CoV-2 fusion protein may affect target cell recognition and subsequent cell infection. This process is dependent on the recognition of the angiotensin-converting enzyme 2 (ACE2) receptor by the fusion protein's receptor binding domain (RBD). Crucially, a significant portion of these mutations has been observed within the RBD region, with some of these playing a pivotal role in the emergence of dominant viral variants, largely due to their heightened fitness advantages [83, 176]. These mutations can influence the conformational dynamics of the RBD. To investigate this, we employed atomistic molecular dynamics simulations of the wild-type RBD and four VOCs (Alpha, Beta, Delta, and Omicron).

In this chapter, we also present our contribution to a study aimed at identifying the effects of single and synergistic amino acid substitutions in promoting immune escape. We analyzed 57 PDB structures, and we determined which residues of the SARS-CoV-2 fusion protein most frequently engaged in interactions with antibodies and complemented this data with RBD-ACE2 simulations to pinpoint hotspot residues for immune evasion.

In **Chapter 4** we shift our focus to the parainfluenza virus. Similarly to other enveloped viruses, its membrane fusion process starts with the recognition and binding of the receptor-binding glycoprotein to receptors at the host cell surface. This binding triggers the fusion protein (or F glycoprotein) to undergo major conformational rearrangements and the proteolytic cleavage of the F protein, exposing the fusion peptide (FP) region [177, 178]. Once exposed, this region becomes embedded in the target membrane, as observed in previous studies [179].

To gain a deeper understanding of its role in membrane fusion, we conducted molecular dynamics simulations, employing a hybrid approach that combines atomistic and coarse-grain detail. We then compared the simulation outcomes with biophysical assays conducted in collaboration with our research partners.

In **Chapter 5** to build upon the data described in Chapter 4, we introduced alanine mutations to the PIFP amino acid residues identified as crucial for membrane fusion, namely Q120 and F103. As before, we conducted molecular dynamics simulations and

biophysical assays.

Finally, in **Chapter 6** we embark on an exploratory study of the influence of SARS-CoV-2 and parainfluenza fusion proteins on viral fusion. We employ coarse-grained molecular dynamics simulations to study fusion between simplified viral vesicles, both with and without fusion proteins, and cell membrane models.

In summary, delving into the fusion mechanisms of both SARS-CoV-2 and parainfluenza, two class I viruses, offers a unique vantage point in viral fusion research. It allows us to draw valuable parallels and distinctions between these systems, leading to a more comprehensive understanding of viral fusion processes and potential therapeutic avenues.

GENERAL THEORY AND METHODS

2.1 Molecular Modelling

Predicting the molecular properties of biological systems can be challenging, not only due to their complexity but also when it comes to the costly time and resources required. Even though experimental techniques can be used to study the molecular properties of a system, information, such as the conformations that characterize the system over time, is usually not obtainable [180]. This is where computer simulations come in as a valuable tool. Besides providing information on the system's dynamics, they also enable the study of the system under conditions that are difficult to probe experimentally.

A diverse array of computational models can be employed to understand the conformational dynamics of biomolecules and their interactions. According to the level of detail required and the properties under study, various molecular simulation methods can be used (Figure 2.1). For example, quantum mechanical calculations allow for the study of chemical reactions in electronic detail within single molecules. At larger scales, all-atom and coarse-grained molecular dynamics (MD) simulations provide the capacity to examine biological phenomena comprehensively, ranging from the individual protein level to the vast domain of large subcellular organelles. In this thesis, we leveraged the power of MD simulations.

2.2 Molecular Dynamics Simulations

In MD simulations, Newton's equations of motion (equation 2.1) are integrated numerically in order to generate the time-dependent behavior of a system:

$$F_i = m_i a_i = m_i \frac{dv_i}{dt} = m_i \frac{d^2 r_i}{dt^2} \quad (2.1)$$

where r_i represents the Cartesian coordinates of a particle i with mass m_i at a specific time t . The acceleration of each particle is the first derivative of its velocity (v_i) and the second derivative of its position (r_i). F_i is the force acting on the particle which is given by

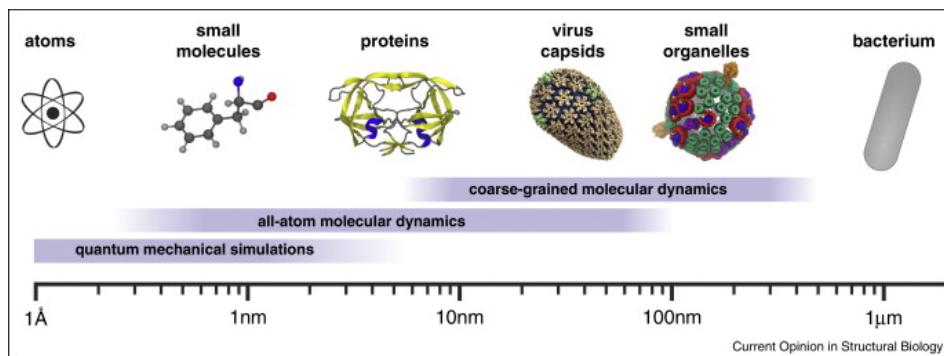


Figure 2.1: Molecular simulation methods used to probe different scales and levels of description of biomolecular systems. Reprinted with permission from [181]. Copyright © 2015 Elsevier.

the negative gradient of the potential energy function U , which in turn depends on the coordinates of all atoms in the system:

$$F_i = \frac{-\partial U(r_1, r_2 \dots r_N)}{\partial r_i} \quad (2.2)$$

2.2.1 Integration algorithms

Several algorithms can be used to integrate Newton’s laws of motion and update the positions and velocities of particles over time. The leap-frog algorithm is one of these algorithms and uses the following relationships [182, 183]:

$$v(t + \frac{\Delta t}{2}) = v(t - \frac{\Delta t}{2}) + \frac{\Delta t}{m_i} F_i(t) \quad (2.3)$$

$$\vec{r}(t + \Delta t) = \vec{r} + \vec{v}(t + \frac{\Delta t}{2})\Delta t \quad (2.4)$$

The atomic positions, r , at time t , and the velocities, v , at time $t + \frac{\Delta t}{2}$ are used to obtain the new positions and velocities using the forces, $F_i(t)$, on each atom. This will allow for the velocities to “leap over” the positions and then the positions to “leap over” the velocities, hence the name leap-frog [184].

The choice of the time interval, Δt , must be done carefully, taking into account the properties of the molecular system and it must be considerably smaller than the characteristic time of the fastest motion studied [184]. On the other hand, the time step should be as large as possible, in order to speed up the simulation, enabling it to sample a significant fraction of the system’s configurational space. A strategy that is frequently used to increase the integration time step is to constrain bond lengths to their reference values, using constraint algorithms.

2.2.2 Force field

The potential energy function is based on a set of interaction functions and parameters, called the force field. Most physics-based MD force fields are modeled by a potential energy function, U^{system} (for example, as described in equations 2.5, 2.6 and 2.7), that only depends on the position of the particles that make up the system:

$$U^{system} = U^{bonded} + U^{non-bonded} \quad (2.5)$$

$$U^{bonded} = \underbrace{\frac{1}{2} \sum_{bond} K_r (r - r_0)^2}_{U_{bond}} + \underbrace{\frac{1}{2} \sum_{ang} K_\theta (\theta - \theta_0)^2}_{U_{angle}} + \underbrace{\frac{1}{2} \sum_{n,dih} K_\varphi [1 + \cos(n\varphi - \delta)]}_{U_{dihedral}} \quad (2.6)$$

$$U^{non-bonded} = \underbrace{\sum_{i,j} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]}_{U_{Lennard-Jones}} + \underbrace{\sum_{i,j} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}}}_{U_{electrostatic}}. \quad (2.7)$$

U^{system} consists of a sum of terms that describe the interactions between each particle. The first set of terms describes the 2-, 3-, and 4-body interactions of covalently bonded particles, i.e. bonds, angles, and dihedral angles (U^{bonded} , equation 2.6, Figure 2.2). The second set of terms describes the van der Waals and electrostatic interactions, which are commonly represented by a combination of Lennard-Jones (LJ) and Coulomb potentials ($U^{non-bonded}$, equation 2.7). Some variations to this potential energy function may occur between forcefields. In this thesis, we employed both the Amber14sb[185] and Charmm36m [186] atomistic force fields, as well as the Martini 3 coarse-grained forcefield [187].

2.2.2.1 Bonded interactions

Bonded interactions describe the 2-, 3-, and 4-body interactions of covalently bonded particles, i.e. bonds, angles, and torsion angles (U^{bonded} , equation 2.6, Figure 2.2). Many potentials are available within MD engines to accomplish this, and the type of potential used highly depends on the forcefield and each particular case. GROMACS alone — the MD engine used in this thesis — contains over 6 bond, 10 angle, and 8 dihedral bonded potentials that can be implemented in any forcefield [188].

Nonetheless, in most cases these terms are described by harmonic potentials, to maintain bonds, angles, or dihedrals close to their reference values (r_0 , θ_0 and δ), with their stiffness determined by the potential's force constant (K_r , K_θ and K_φ). This is the case for the bond, angle, and dihedral potentials described in equation 2.6, and the most common case in the three forcefields used in this thesis.

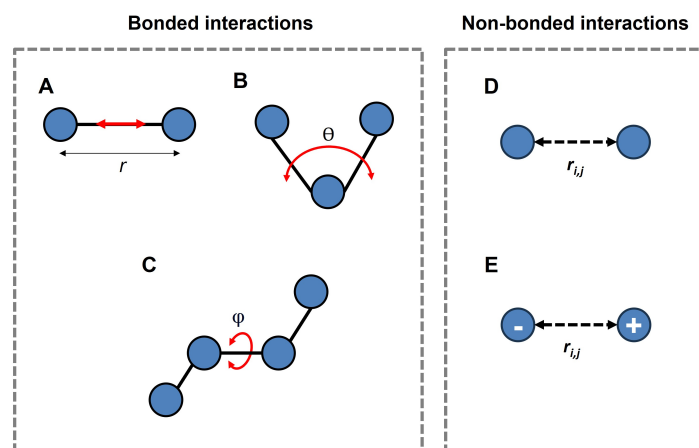


Figure 2.2: Schematic illustration of the bonded and nonbonded interactions in the potential force field systems. (A) Bond stretching, (B) angle bending, (C) dihedrals, (D) van der Waals interactions and (E) electrostatic interactions.

2.2.2.2 Non-bonded interactions

Atoms not separated by bonds, or those not belonging to the same molecule, engage in nonbonded interactions: van der Waals and electrostatic forces.

Van der Waals interactions, first described by Johannes van der Waals, encompass attractive forces resulting from instantaneous dipoles arising from electron cloud fluctuations and repulsive forces explained by Pauli’s exclusion principle. Despite being weaker than electrostatic interactions, accurately modeling van der Waals forces is vital in molecular mechanics force fields (FFs). Typically, these forces are described using a Lennard-Jones potential (equation 2.7), comprising an attractive term varying with the sixth power of the particle distance (r) and a repulsive term decaying with the twelfth power of r . Due to their short-range nature, it is often reasonable to neglect van der Waals forces when the distance between particles exceeds a specific cutoff value, resulting in substantial computational time savings.

When the charge distribution is represented by assigning partial charges to each atom, electrostatic interactions between two atoms are calculated using Coulomb’s law (equation 2.7). Here, q_i and q_j denote the atomic partial charges, r_{ij} represents the interatomic distance, ϵ_0 represents the permittivity of vacuum, and ϵ_r is the relative permittivity of the medium. Electrostatic interactions can be truncated for computational efficiency, but their strong and long-ranged nature (decay with $1/r$) prevents abrupt truncation. Two main strategies are used to handle the long-range effects of electrostatic interactions:

- Lattice sum methods like the **Particle Mesh Ewald (PME)** [189] method to compute long-range potential energy contributions in periodic systems. These methods calculate interactions between each particle and all other particles, including periodic images, within the simulation box. Long-range interactions are computed in reciprocal space using a Fourier transform, significantly accelerating calculations.

- Continuum methods, such as the **Reaction Field method** [190], model the impact of long-range dipole-dipole interactions. In this approach, each molecule is enclosed by a sphere defined by a specified cutoff radius. Interactions between a molecule and particles within this sphere are explicitly calculated. Beyond the cutoff, the medium is considered a homogeneous environment with a defined dielectric constant (ϵ_r). The molecule induces polarization in this medium, creating a reaction field.

Like in the case of the bonded interactions, the choice of electrostatic treatment method is forcefield dependent. For example, AA force fields such as Amber14sb and CHARMM36m typically use PME, while the Martini CG forcefield typically uses Reaction field.

2.2.3 Periodic boundary conditions

In molecular dynamics simulations, when working with a finite system, some thought must be given to how to deal with the boundaries of the system. If we were to simulate a closed box of a given substance in vacuum, forces acting on atoms near the edges of the box would be different from the forces that act on atoms near the center of the box. On the other hand, if there are no walls, atoms could get out of the box over time and the number of particles is not maintained [191, 192].

A solution to these problems is to use periodic boundary conditions (PBC). When doing so, the simulated system is considered to be replicated in all possible directions with equivalent boxes such that molecules can move between adjacent box replicates (Figure 2.3). Hence, every atom in the central cell is under the influence of every other atom surrounding it, and so the atoms in the simulation box are not affected by physical boundaries [191, 192]. In reality, these replicate boxes are not explicitly simulated and, instead, mathematical operations mimic this periodicity.

In MD simulations using PBC, simulation boxes have to be large enough to avoid interactions between the same molecule over periodic boundaries. Given that very large simulation boxes hinder computational performance, one has to find a compromise between these two conditions. Simulation boxes can have different geometries, e.g., cubic, rhombic dodecahedron, truncated octahedron. The rhombic dodecahedron and truncated octahedron, exhibit shapes that are closer to a sphere than a cube. Consequently, they prove to be more suitable for studying approximately spherical macromolecules in a solution. One key advantage is that these shapes require fewer solvent molecules to fill the simulation box while maintaining a minimum distance between images of the macromolecule. Hence, depending on the system in study, there will be an optimal choice of simulation box type. Simulations of soluble proteins are typically performed using truncated dodecahedron boxes, whereas simulations of membrane bilayers are typically performed in cubic boxes.

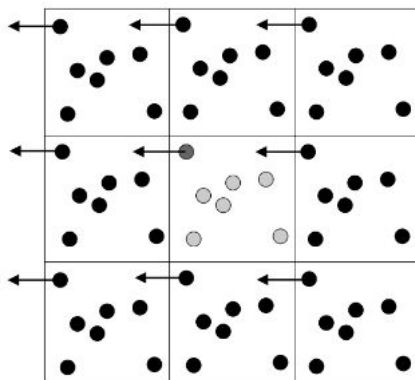


Figure 2.3: A slice through a small portion of the system (the central box with atoms in grey) and the copies (with atoms are shown black). Each copy is identical, and each atom undergoes the same time development as its image in every other copy. As the dark grey atom (top left, central cell) leaves the central cell, its image enters from an adjoining copy, shown by the vector displacements in the figure. This keeps the density constant. Adapted with permission from [193]. Copyright © 1999 John Wiley & Sons.

2.2.4 Constant temperature and/or pressure

MD simulation methods aim to simulate the collective behaviors of systems by considering key factors like the number of particles (N), temperature (T), and container volume (V). To compute thermodynamic averages, various system configurations are sampled, and an average is calculated from this array of conformations. This type of average is called an ensemble average because it encompasses all conceivable states of the system in a macroscopic context [184].

During an MD simulation, if both energy (E) and volume (V) remain constant, the simulation is said to sample from a microcanonical ensemble (NVE). In this ensemble, the system is isolated from its surroundings and all the microstates that compose the ensemble have the same energy and, therefore, the same probability of occurring.

Most real systems are not completely isolated from thermal influences, making simulations that maintain constant values for N , T , and pressure (P) or those that uphold constant values for V , T , and chemical potential more realistic [191]. In the NVT ensemble, the probability distribution for the microscopic states is given by the Boltzmann distribution [194]:

$$P_i(N, V, T) = \frac{e^{-E_i(N,V)/kT}}{Q(N, V, T)} \quad (2.8)$$

Where E_i is the energy of state i , k is the Boltzmann constant and T is the absolute temperature.

Most biomolecular MD simulations are performed at a constant temperature. This is important when one wants to compare MD results with *in vivo* or *in vitro* experiments performed at a given constant temperature. Temperature is a thermodynamic property

given by the average kinetic energy of a system over simulation time. The most common way to control temperature in an MD simulation is to couple the system to an external heat bath that can provide or remove energy from the system [191].

An example of a temperature coupling algorithm is the Berendsen bath [195], in which the deviation of the system's temperature (T) from the bath's temperature (T_{bath}) is slowly corrected. This correction guarantees that the rate of temperature change is proportional to the difference between the bath and system temperatures [195]:

$$\frac{dT(t)}{dt} = \frac{T_{\text{bath}} - T(t)}{\tau} \quad (2.9)$$

The coupling parameter (τ), determines how the bath temperature and system temperature are coupled (the smaller the τ , the stronger the coupling). In practice, at each time step, the velocities are multiplied by a time-dependent scaling factor (λ), given by [195]:

$$\lambda = \left[1 + \frac{\Delta t}{\tau} \frac{T_{\text{bath}}}{T(t)} - 1 \right]^{1/2} \quad (2.10)$$

In some cases, it is desirable to perform the simulation in an NPT ensemble, i.e., at constant temperature and pressure. Just like temperature, the pressure can be controlled by coupling the system to a pressure bath. For instance, the Berendsen pressure algorithm [195] scales the coordinates and box vectors at every time step, so that the pressure can relax towards a given reference pressure (P_{bath}) [195]:

$$\frac{dP(t)}{dt} = \frac{P_{\text{bath}} - P(t)}{\tau} \quad (2.11)$$

To rescale the coordinates and box vectors, the algorithm uses a scaling factor (μ), given by the following expression (in which β is the isothermal compressibility of the system) [195]:

$$\mu = 1 - \frac{\beta \Delta t}{3\tau_p} (P_{\text{bath}} - P) \quad (2.12)$$

There are many barostat and thermostat methods that can be used. These typically differ on the accuracy of the canonical ensemble reproduction and their capability to quickly and smoothly reach the desired temperature/pressure. In practice, different forcefields will call for different barostat and thermostat methods. These can even differ between different simulation stages, such as the equilibration and the production stage. The barostats and thermostats used in this thesis are described in each chapter-specific method section.

2.3 The challenge of scale: A multiscale approach

Computational models can define the functional dynamics of complex systems in exceptional detail. For instance, atomistic and coarse-grained approaches are the most commonly used models to study the dynamics of biological systems:

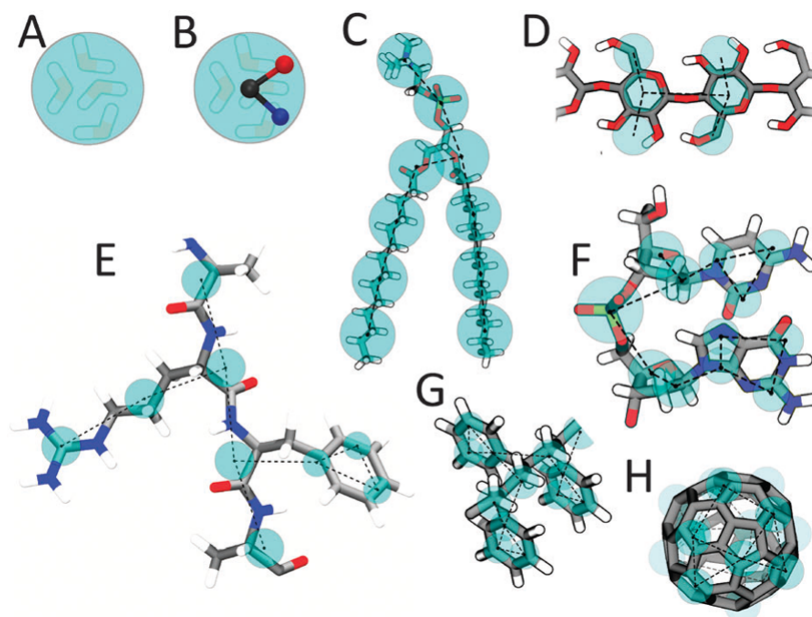


Figure 2.4: Martini mapping examples: (A) A water particle representing four water molecules, (B) Polarizable water molecule with embedded charges, (C) DMPC lipid, (D) Polysaccharide fragment, (E) Peptide, (F) DNA fragment, (G) Polystyrene fragment, (H) Fullerene molecule. Martini CG beads are shown as cyan transparent beads overlaying the atomistic structure. Reprinted with permission from [196]. Copyright © 2013 Royal Society of Chemistry.

Atomistic models: The parameters in these models are derived from a combination of high-level quantum chemistry calculations and experimental data on small molecular systems [197, 198]. The detailed nature of these models entails considering a large number of particles and employing a small time step (1-2 fs). Consequently, full atomistic models are primarily restricted to the study of small systems, often in conjunction with sampling enhancement techniques like replica exchange [199] or metadynamics [200]. The benefit of a fully atomistic simulation high molecular detail we can obtain from it and the biological questions it can answer, such as protein folding.

Coarse-grained models: Coarse-grained models offer a simplified representation of proteins and other biomolecules when compared to atomistic models. In these models, groups of atoms are averaged, leading to smoothed potential energy functions and faster simulations. Coarse-grained models have been developed to investigate longer timescales and larger scales, critical for understanding various biological processes, including interactions between lipids and proteins. Due to the smoothed potential, larger time steps (10-50 fs) are commonly used in these simulations. The Martini model is one of the most popular coarse-grained models for simulating lipid membranes and proteins. It was developed to study lipid bilayers and has since been extended to model proteins, nucleic acids, and carbohydrates. In the Martini model, four heavy atoms are grouped into a single bead,

2.3. THE CHALLENGE OF SCALE: A MULTISCALE APPROACH

Approach	Advantages	Limitations	Timescale
Coarse-Grained	- Faster simulations due to reduced degrees of freedom. - Suitable for studying large systems and long timescales. - Allows for the study of mesoscale phenomena.	- Loss of atomic-level detail. - Limited accuracy for some properties (e.g., bond lengths). - Limited applicability to small molecules or specific interactions.	Up to milliseconds or longer
Atomistic	- Provides high-resolution details of molecular structures and interactions. - Accurate for small molecules and specific interactions.	- Computationally expensive for large systems and long timescales. - Restricted to shorter simulations. - May miss mesoscale phenomena.	Up to microseconds

Table 2.1: Advantages and limitations of using a coarse-grained approach, in molecular dynamics simulations. Comparison with atomistic approaches, highlighting their differences in terms of timestep, timescale, and key features.

and nonpolar and polar interactions are represented by different types of interaction sites (Figure 2.4) [187, 201].

However, these models are commonly faced with conflicting demands: to gain insight into certain high-resolution phenomena that evolve over large length and time-scales. Multiscale modeling has emerged as a crucial strategy to bridge this gap effectively.

An example of this is, for instance, combining atomistic and coarse-grain scales in MD simulations. This offers a powerful approach to studying complex biological systems. Atomistic simulations are highly detailed and provide accurate representations of molecular interactions, but they are computationally expensive (Table 2.1). Coarse-grain models reduce the number of degrees of freedom, allowing for simulations over longer timescales and larger systems (Table 2.1), which may be impractical with atomistic models alone. By starting with a coarse-grain model and transitioning to atomistic details when necessary, you can strike a balance between accuracy and computational cost.

Also, several biological processes involve phenomena that occur on different lengths and time scales. Combining atomistic and coarse-grain scales allows for the capture of relevant details at each scale. In summary, the combination of atomistic and coarse-grain scales in MD simulations presents a versatile and efficient approach to investigating complex systems. By integrating the extended timescales of coarse-grained simulations with the molecular precision of atomistic simulations, a deeper understanding of the dynamics of larger and more intricate systems can be achieved without an unreasonable increase in computational demand. Recent publications have successfully applied this approach [172, 202, 203].

EXPLORING THE IMPACT OF MUTATIONS ON
THE SARS-CoV-2 RBD CONFORMATIONAL
DYNAMICS AND ANTIBODY NEUTRALIZATION

This chapter contains published data from:

Valério, M., Borges-Araújo, L., Melo, M. N., Lousa, D., & Soares, C. M. (2022). SARS-CoV-2 variants impact RBD conformational dynamics and ACE2 accessibility. *Frontiers in Medical Technology*, 4, 1009451.

Author contributions: MV and LB-A contributed equally to this work. DL and CS designed the simulation setup. MV and CS performed the simulations. MV, LB-A, MM, DL and CS designed the analysis and MV and LB-A performed them.

This chapter contains published data from:

Alenquer, M., Ferreira, F., Lousa, D., Valério, M., Medina-Lopes, M., Bergman, M. L., Gonçalves, J., Demengeot, J. , Leite, R. B., Lilue, J., Ning, Z., Penha-Gonçalves, C., Soares, H., Soares, C. M., & Amorim, M. J. (2021). Signatures in SARS-CoV-2 spike protein conferring escape to neutralizing antibodies. *PLoS pathogens*, 17(8), e1009772.

Author contributions: MJA and CS conceived and supervised the study. All authors designed and conducted experiments; DL prepared the systems' topologies and CS performed the simulations. MV, DL and CS designed the analysis and MV performed them.

3.1 SARS-CoV-2 variants impact RBD conformational dynamics and ACE2 accessibility

3.1.1 Abstract

Coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has killed over 6 million people and is having a devastating social and economic impact around the world. The rise of new variants of concern (VOCs) represents a difficult challenge due to the loss of vaccine and natural immunity, as well as increased transmissibility. All VOCs contain mutations in the spike glycoprotein, which mediates fusion between the viral and host cell membranes. The spike glycoprotein binds to angiotensin-converting enzyme 2 (ACE2) via its receptor binding domain (RBD) initiating the infection process. Attempting to understand the effect of RBD mutations in VOCs, a lot of attention has been given to the RBD-ACE2 interaction. However, this type of analysis ignores more indirect effects, such as the conformational dynamics of the RBD itself. Observing that some mutations occur in residues that are not in direct contact with ACE2, we hypothesized that they could affect the RBD conformational dynamics. To test this, we performed long atomistic molecular dynamics (MD) simulations to investigate the structural dynamics of wt RBD, and that of four VOCs (Alpha, Beta, Delta, and Omicron). Our results show that the wt RBD presents two distinct conformations: an “open” conformation where it is free to bind ACE2; and a “closed” conformation, where the RBM ridge blocks the binding surface. The Alpha and Beta variants shift the open/closed equilibrium towards the open conformation by roughly 20%, likely increasing ACE2 binding affinity. Simulations of the Delta and Omicron variants showed extreme results, with the closed conformation being rarely observed. The Delta variant also differed substantially from the other variants, alternating between the open conformation and an alternative “reversed” one, with a significantly changed orientation of the RBM ridge. This alternate conformation could provide a fitness advantage due to increased availability for ACE2 binding, and by aiding antibody escape through epitope occlusion. These results support the hypothesis that VOCs, and particularly the Omicron and Delta variants, impact RBD conformational dynamics in a direction that promotes efficient binding to ACE2 and, in the case of Delta, may assist antibody escape.

3.1.2 Introduction

Coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [204–206], is a global pandemic with higher mortality than that of seasonal influenza [207]. As of July 2022, over 6 million lives had been claimed by this disease [208]. Infection by SARS-CoV-2 requires the fusion of viral and host cell membranes, at either the cell surface or the endosomal membrane [31]. As for the severe acute respiratory syndrome coronavirus (SARS-CoV) and the Middle East respiratory syndrome-related coronavirus (MERS-CoV), the SARS-CoV-2 fusion process is mediated

3.1. SARS-COV-2 VARIANTS IMPACT RBD CONFORMATIONAL DYNAMICS AND ACE2 ACCESSIBILITY

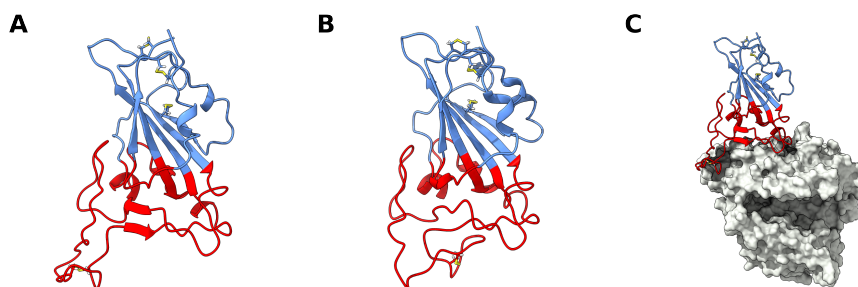


Figure 3.1: SARS-CoV-2 receptor binding domain (RBD) structure. Structure of wt RBD in the open (A) and closed (B) conformations. Snapshots obtained from the AA MD simulations. Disulfide bonds are represented in yellow sticks. Structure of wt RBD bound to ACE2 is also shown (C). The RBM region is colored red and the ridge in dark red, with the rest of the protein being colored in blue. ACE2 is in grey.

by the viral envelope spike glycoprotein [31]. Upon viral attachment or uptake, host factors trigger large-scale conformational rearrangements in the S protein, including a refolding step that leads directly to membrane fusion and viral entry [56, 66, 209–212].

The SARS-CoV-2 S protein is composed of a signal peptide located at the N-terminus (residues 1–13) and 2 subunits, S1 (residues 14–685) and S2 (residues 686–1,273) [64]. The S1 and S2 subunits are responsible for receptor binding and membrane fusion, respectively [64]. The S1 subunit consists of a N-terminal domain (residues 14–305) and a receptor binding domain, or RBD (residues 319–541). In its prefusion state, the S protein exists as a homotrimer and undergoes large conformational changes to control the exposure and accessibility of the RBD. This is done via an “up” and “down” mechanism, where the RBD changes from a receptor-accessible “up” conformation to a receptor-inaccessible “down” conformation [62, 65, 213].

The RBD is responsible for the interaction of SARS-CoV-2 with host cells via binding to the angiotensin-converting enzyme 2 (ACE2) [56, 64, 66, 67], a regulator of the renin-angiotensin system. Binding to ACE2 is one of the first steps in what is considered to be the main mode of SARS-CoV-2 viral entry, hence the importance of the RBD positional change from “down” to “up” [62, 213].

A lot of attention has been given to the SARS-CoV-2 RBD—ACE2 complex due to both its mechanistic implications [93–96, 214, 215] and pharmaceutical potential [97–102, 216]. However, not much attention has been given to the dynamics of the RBD by itself. The RBD core structure when bound to ACE2 (Figure 3.1A) consists of a twisted five stranded antiparallel β sheet ($\beta 1$, $\beta 2$, $\beta 3$, $\beta 4$ and $\beta 7$), with short connecting helices and loops [90]. While most of the S protein surface is densely glycosylated, shielding it from host defense mechanisms, the RBD itself contains only a single glycosylation site [78], N343, which is located relatively distant from the ACE2-RBD interface (Supplementary Figure S3.1B).

This core β sheet structure is further stabilized by 3 disulfide bonds. Between the core $\beta 4$ and $\beta 7$ strands (residues 438–506), there is an extended region containing 2 short β strands ($\beta 5$ and $\beta 6$), the alpha 4 and alpha 5 helices and loops. This region is the

receptor-binding motif (RBM), which contains most of the residues that are responsible for interacting with ACE2 [62, 90]. When complexed with ACE2, the RBM folds into a concave surface, that accommodates the N-terminal α -helix of ACE2, with a ridge (residues 471–491) on one side, formed by a disulfide-bridge-stabilized loop (Cys480–Cys488). It is in this surface that several RBM residues establish specific and non-specific interactions with ACE2 residues [90].

From the available experimental structural data the core β -sheet structure is very stable, but the RBM seems to be quite dynamic and not as structurally defined, unless bound to other proteins, like ACE2 [30, 67, 90, 112, 113] or antibody fragments [114–120]. Molecular dynamics (MD) simulation studies have also mostly focused on RBD complexed with these proteins, and while there are MD simulation studies of the free RBD, they either focus on short simulations [121–123] or do not explore the RBM dynamics in detail [78, 124–126]. As such, not much is known about the conformational dynamics of this motif when unbound. This is relevant because the conformational dynamics of the SARS-CoV-2 RBD and RBM might not only play an important role in receptor recognition and binding, but also provide important information for the development of newer improved pharmaceuticals.

Recently, a significant number of naturally occurring mutations to the SARS-CoV-2 S protein have also been reported [81, 83, 217]. Many of these mutations have been identified in the RBD, some of which have given rise to the dominant viral variant in certain regions due to their significant fitness advantage [80, 81, 217]. Many of these RBD mutations are thought to increase fitness by increasing the binding affinity for ACE2 or by escaping neutralization by anti-SARS-CoV-2 monoclonal antibodies [218]. Still, the impact of these mutations on the structural dynamics of the RBD and the RBM have not yet been investigated.

In this work, we use atomistic molecular dynamics (MD) simulation methods to investigate the structural dynamics of the SARS-CoV-2 RBD, and that of four naturally occurring variants of concern (VOCs): variant B.1.1.7, or Alpha [80] (N501Y); variant B.1.351, or Beta [81] (K417N, E484K and N501Y); variant B.1.617.2, or Delta [83] (L452R, T478K); and variant B.1.1.529, or Omicron [84, 85] (G339D, S371I, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y and Y505H). Our results show that the RBM dynamics of the wt RBD are such that it is not always in a conformation competent for ACE2 binding (Figure 3.1). Conversely, all variants, particularly Delta and Omicron, stabilize binding-competent configurations which could increase ACE2 binding efficiency. Besides impacting binding, the large conformational space visited by the variants may also hinder antibody recognition of the RBM region, thus providing a fitness advantage by facilitating antibody escape.

3.1.3 Methods

3.1.3.1 Molecular dynamics simulations

All atomistic simulations were performed with the GROMACS 2020.3 [219, 220] package and modelled using the Amber14sb forcefield [185], alongside the TIP3P water model [221]. The initial wt RBD structure was obtained from PDB ID: 6M0J [90], which corresponds to an ACE2 bound conformation of RBD; ACE2 was excluded from this structure. The different RBD variants were generated by mutating the appropriate residues in the wt RBD using PyMOL [222].

It is worth noting that glycosylations were not included in our simulation systems. Despite most of the S protein surface being densely glycosylated, the RBD itself contains only a single glycosylation site far from the RBM region [78], where the dynamics reported in this work are observed. Additionally, while it has been reported that other neighboring glycosylation sites can effectively shield the RBD, this glycan shield is paired with its down-to-up conformational change in the complete S-protein: when the RBD is down the glycan shield camouflages the RBD and RBM, however, when the RBD is up it emerges from the glycan shield and presents a fully accessible RBM [78]. It is in this up state, when the RBM is fully accessible to the solvent and glycans no longer play a relevant role, that the dynamics we observe may play a role in modulating binding to ACE2. Additionally, accounting for glycans would inevitably introduce degrees of freedom that would complicate sampling. Given that we do not expect glycans to play a relevant role in RBM dynamics, we opted for a reductionist approach by simulating the RBD without glycosylations.

Simulations were performed on each RBD protein structure in water. Each structure was inserted in a truncated dodecahedron box filled with water molecules (considering a minimum distance of 1.2 nm between protein and box walls). The total charge of the system was neutralized with the required number of Na^+ ions, with additional Na^+ and Cl^- ions added to the solution to reach an ionic strength of 0.1 M.

The system was energy-minimized using the steepest descent method for a maximum of 50,000 steps with position restraints on the heteroatom positions by restraining them to the crystallographic coordinates using a force constant of 1,000 kJ/mol in the X, Y and Z positions. Before performing the production runs, an initialization process was carried out in 5 stages of 100 ps each. Initially, all heavy-atoms were restrained using a force constant of 1,000 kJ/mol/nm, and at the final stage only the $\text{C}\alpha$ atoms were position-restrained using the same force constant. In the first stage, the Berendsen thermostat [195] was used to initialize and maintain the simulation at 300 K, using a temperature coupling constant of 0.01 ps, without pressure control. The second stage continued to use the Berendsen thermostat but now with a coupling constant of 0.1 ps. The third stage kept the same temperature control, but introduced isotropic pressure coupling with the Berendsen barostat [195], with a coupling constant of 5.0 ps. The fourth stage changed the thermostat to V-rescale [223], with a temperature coupling constant of 0.1 ps, and the

barostat to Parrinello-Rahman [224], with a pressure coupling constant of 5.0 ps. The fifth stage is equal to the fourth stage, but position restraints are only applied on $C\alpha$ atoms. For production simulations, conditions were the same as for the fifth stage, but without any restraints. In all cases, 2 fs integration steps were used. Long-range electrostatic interactions were treated with the PME [189, 225] scheme, using a grid spacing of 0.12 nm, with cubic interpolation. The neighbor list was updated every twenty steps with a Verlet cutoff with a 0.8 nm radius. All bonds were constrained using the LINCS algorithm [226].

Simulations of each system were performed for at least 7 μ s over 5 replicates (the wt was simulated for 15 μ s, and the Alpha, Beta, Delta and Omicron variants for 7 μ s each). The first 3 μ s of simulation were considered as equilibration time and the remaining frames were used for analysis. Visualization and rendering of simulation snapshots was performed with the molecular graphics viewers VMD [227], PyMOL [222] and UCSF Chimera [228].

3.1.3.2 Principal component analysis

PCA is a standard dimensionality reduction method that we apply here to the $(3N-6)$ -dimensional space of possible RBD conformations (in our case, N being the number of RBD residues). PCA consists of a linear transformation that changes a set of possibly correlated dimensions into a set of linearly uncorrelated, mutually orthogonal ones, called principal components (PCs). The first PC can be defined as the direction that accounts for as much of the variance in the data as possible, with each successive PC accounting for as much of the remaining variance as possible. Reduction of data dimensionality is achieved by retaining only a few of the first PCs — which represent the strongest correlations in the data, in our case, the most important conformational motions—, thus sacrificing some information for simplicity. Discussions of the mathematical and computational backgrounds can be found elsewhere [229–232].

In this work, PCA was applied to sets of conformational coordinates obtained from MD simulations. Prior to PCA, each conformation was translationally and rotationally fitted to the RBD core $C\alpha$ carbons of the wt crystal structure (hence the -6 in the dimensionality). PCs were determined using MDAnalysis [233], from the entire pool of simulation trajectories, considering only the coordinates of the RBD's $C\alpha$ carbons. The dimensionality was reduced to the 2 most representative PCs, preserving a large part of the variance. RBD structures for each simulation frame, for each variant, could then be projected as points in this two-dimensional space, enabling a simplified visual representation of the conformation space explored by the RBD in each case.

The probability density function for each trajectory projection was estimated using a gaussian kernel estimator [231, 234] implemented in LandscapeTools' `get_density` software as described elsewhere [231, 235]. This procedure defines a probability density function $P(r)$, with the values of $P(r)$ being stored for the position of each data point and for the nodes of a two-dimensional uniform grid, with a mesh size of 0.5 Å. These values

were used to define an energy surface, calculated as [231]:

$$E(r) = -kBT \ln \left(\frac{P(r)}{P_{\max}} \right) \quad (3.1)$$

Where $P_{\max}$ is the maximum of the probability density function, $P(r)$. The energy surface landscapes were analyzed by determining the energy minima and respective basins. The basins were defined as the set of all conformations whose steepest descent path along the energy surface leads to a particular minimum [231, 236, 237]. Here, the steepest descent paths for each grid cell were computed, with each conformation inheriting the path of its corresponding grid cell. Landscape regions with $E > 6$ kBT were discarded, resulting in the final set of basins for each data set.

3.1.3.3 Residue interaction network analysis

Residue interaction networks (RINs) are graph representations of protein structures, where the nodes represent amino acid residues, and the edges represent interactions between residues. Pairwise residue interactions were analyzed for the 5,000 lowest energy conformations obtained for the most populated open, closed and reversed conformation basins of the energy surface landscapes of each RBD variant, using RIP-MD [238]. Several types of interactions between AAs were probed: $C\alpha$ contacts, hydrogen bonds, salt bridges, disulfide bonds, cation- π , π - π , Arg-Arg, Coulomb and van der Waals. The parameters defining each interaction, as well as their mathematical formulation can be found elsewhere [238]. Once the interactions were determined, the interaction networks were visualized using Cytoscape [239].

3.1.4 Results and discussion

Our aim was to study the conformational dynamics of the SARS-CoV-2 RBD, as well as that of several other SARS-CoV-2 VOCs in solution. To this effect, we simulated the wt, Alpha, Beta, Delta and Omicron variants of the SARS-CoV-2 RBD. The Gamma variant was not studied due to its similarity to the Beta variant: in the RBD both variants share the E484K and N501Y mutations; the single difference is the K417N mutation in the Beta variant vs. K417T in Gamma [240, 241]. In either case, K417 is mutated to a residue with a polar uncharged side chain, which should impact the interaction network similarly. As such, the conformational dynamics specific to the RBD and RBM are expected to be similar.

3.1.4.1 WT RBD presents two distinct RBM conformations in aqueous solution

Visual inspection of the trajectories obtained in the simulation of wt RBD in water revealed that large dynamic conformational changes occur in the RBM region (Figure 3.2A). The dynamics observed appear to show an opening and closing of the ACE2 binding surface of the RBM. To better characterize these conformational dynamics, we performed

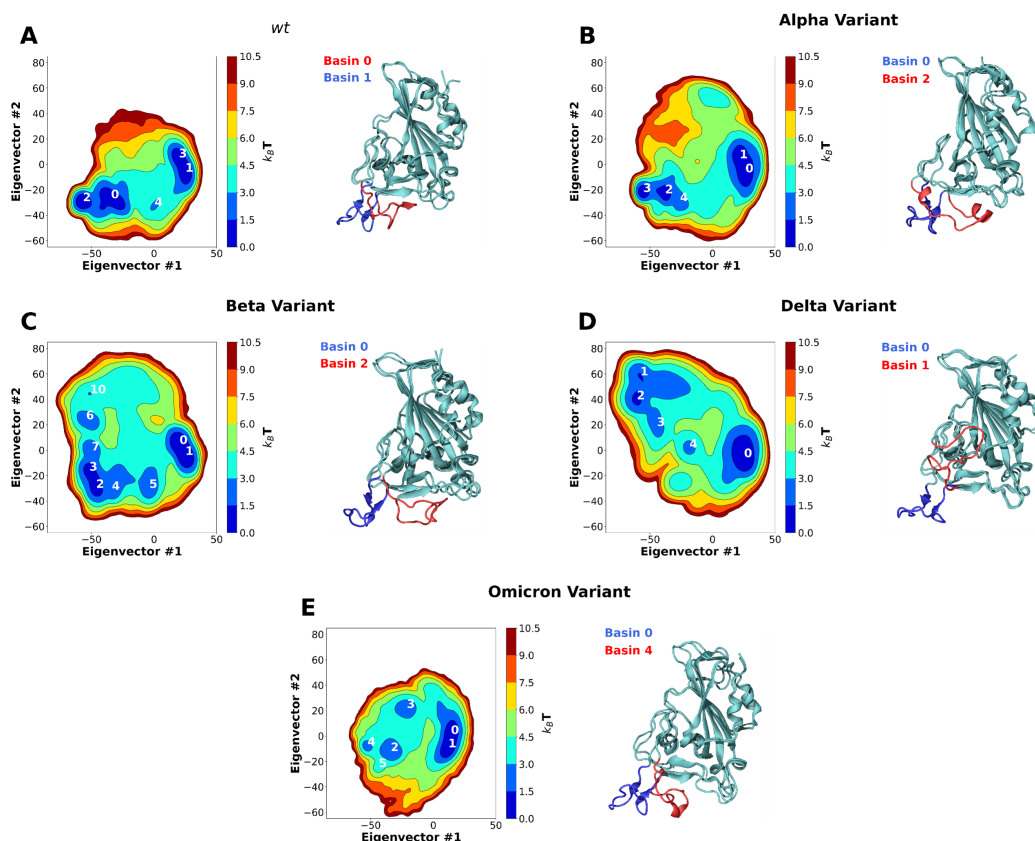


Figure 3.2: Two-dimension principal component analysis (PCA) of SARS-CoV-2 RBD conformational dynamics in water. Plots of the first two principal components determined from the $C\alpha$ backbone of the wt RBD (A) as well as the Alpha (B), Beta (C), Delta (D) and Omicron (E) variants. Basins with $k_B T < 3$ are numbered in each figure. Snapshots of the lowest energy structures for selected open and closed basins are also shown. The ridge regions of the open and closed snapshots are colored in blue and red, respectively.

principal component analysis (PCA) on the coordinates recovered from these simulations, reducing them to 2 principal components; this 2D configuration space sampling was expressed as free energy landscapes (Figure 3.2).

For wt RBD, we observe two deep basin clusters (Figure 3.2A), as well as several other lesser populated basins. Closer analysis of the RBD conformations that make up each basin shows that wt basins 1 and 3 correspond to conformations close to the ACE2-bound one determined by x-ray crystallography [90] (Figures 3.1A, 3.2A). We named these “open” configurations. In contrast, the second basin cluster (basins 0 and 2) was made up by conformations quite distinct from the open ones. In these basins, the loop that makes up the RBM is twisted and collapsed over the region that binds ACE2, effectively hiding it from the solvent (Figures 3.1B, 3.2A). We named these conformations “closed”. Further analysis of the PCA results reveals that the wt RBD is in a closed state for more than half of the simulation time ($\approx 55.5\%$, Supplementary Table S3.1). Given that in these conformations the RBM closes on itself, hiding the ACE2 binding surface, we can speculate that the RBD

3.1. SARS-COV-2 VARIANTS IMPACT RBD CONFORMATIONAL DYNAMICS AND ACE2 ACCESSIBILITY

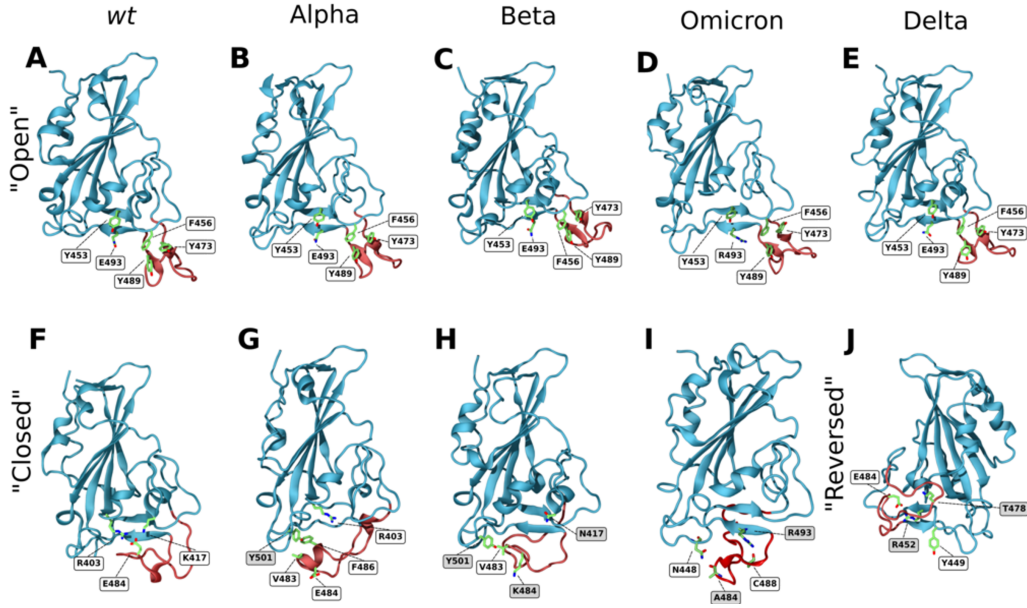


Figure 3.3: Closeup snapshots of SARS-CoV-2 RBD intramolecular interactions that stabilize the various conformations. Snapshots from AA MD simulations showcasing crucial intramolecular interactions responsible for stabilizing the open, closed, and reversed conformations for the wt (A,F), Alpha (B,G), Beta (C,H), Omicron (D,I) and Delta (E,J) RBD variants. The ridge region of the RBD is colored in red and residues of interest in green. Text labels indicate relevant residues, with shaded labels indicating mutations relative to wt. All figures are rotated 180 degrees relative to Figures 3.1, 3.2, apart from snapshot J.

would be unable to effectively bind to ACE2 and initiate an ACE2-dependent infection process. Moreover, the open and closed states were visited reversibly (Supplementary Figure S3.2), indicating that our simulations were not kinetically trapped in either basin. The open and closed RBD conformations reported here should not be confused with the “up” and “down” S protein conformational states which control the exposure of the RBD in the context of the S protein homotrimer. The open/closed dynamics likely act as an additional RBM exposure control, which would be particularly important for RBDs in the “up” S protein conformational state where they are fully exposed to the solvent.

Residue interaction network (RIN) analysis was performed for the 5,000 lowest energy structures of basins 1 (open) and 0 (closed). From the identified interactions, we selected those that were present in over 50% of the simulation frames (Supplementary Figure S3.4). We also only considered interactions that are established by RBM residues, or those in their immediate vicinity. These RINs were then used to probe the different intramolecular interactions established in each of the conformations.

In the open conformation, the RBD ridge is stabilized by a triple π -stacking interaction between residues Y489–F456–Y473 and a hydrogen bond between Y489–Y473. Additionally, two hydrogen bonds are established between residues Y453 and E493, which help stabilize the formation of a small β -sheet (Figure 3.3).

In the closed conformation, however, the π -stacking interactions are broken, and

new interactions with RBD core residues are formed in their place. F456 forms a stable π -stacking with Y421, Y489 forms a transient π -stacking interaction with F486 and Y473 forms a hydrogen bond with the backbone of Y451. Moreover, E484 forms a salt bridge with R403, that is found in the RBD core, and a hydrogen bond with K417 (Figure 3.3F). This hydrogen bond does not show up in the RIN, as K417 can establish a bond with each of the two glutamate oxygens, each with $\approx 40\%$ prevalence (each thus below our 50% selection cutoff). These two interactions, together with the formation of three hydrogen bonds (C480–S494–G482–Q493) are responsible for the closing of the ridge and consequent shielding of the ACE2 binding surface. The importance of the E484–R403 and E484–K417 interactions for the closing of the loop was confirmed by simulating the E484K and K417N mutants. Either of these single mutations were enough to completely deplete the closed conformation (Supplementary Figures S3.3A,B for E484K and K417N, respectively). This shows that both these interactions are crucial for the stabilization of the wt closed state. Still, several other transient hydrogen bonds, formed between residues L492, G493 and S494 of strand β_6 , and T478, C480, N481, G482 and E484 of the RBM ridge, assist in stabilizing the structure. The closed conformation does not seem to substantially impact the RBD's secondary structure (Supplementary Figure S3.9). The largest impact appears to be limited to residues 473–474 and 488–489, that in the open state display a slight β -sheet character. However, upon closing, this β -sheet character disappears. This effect comes from residues 473 and 489 no longer participating in the triple π -stacking that was likely stabilizing this region. Apart from impacting ACE2 accessibility, the closing of the RBM ridge also decreases the solvent accessible surface area (SASA) of the RBD by slightly over 3% (Supplementary Table S3.2). Although other studies have noted the high flexibility in the RBM region of the RBD [121, 122, 242], this is, as far as we know, the first report of this hinge mechanism that can effectively hide the ACE2 binding surface of the RBD from binding partners, which could only be observed through the analysis of μ s-long MD simulations. While it is likely that induced fit interactions might assist in opening a closed conformation for binding to ACE2, it is safe to assume that the closed conformation will have its binding to ACE2 substantially hindered when compared to an open conformation. Other studies have also observed RBM concealment mechanisms in the context of the spike protein [242]. It has been shown that “down” state RBDs can conceal their RBM by interacting with the neighboring RBDs, in a temperature dependent manner. This further showcases the tendency of the RBM to conceal its hydrophobic surface, either by closing in on itself (as observed in the present study) or by interacting with neighbouring RBDs [as observed by Rath et al. [242]].

3.1.4.2 SARS-CoV-2 alpha and beta variants impact RBM conformational dynamics and exposure

The first SARS-CoV-2 variant of concern to be identified was first detected in the UK. It is often referred to as B.1.1.7 or Alpha variant and has only one mutation in the RBD

3.1. SARS-COV-2 VARIANTS IMPACT RBD CONFORMATIONAL DYNAMICS AND ACE2 ACCESSIBILITY

region—N501Y. A second variant emerged soon after in South Africa, independently of B.1.1.7, referred to as B.1.351 or Beta variant. In the RBD region, this variant shares the N501Y mutation with the Alpha variant and includes two others: K417N and E484K [80].

In line with what was observed for the wt RBD, MD simulations of the RBDs from the Alpha and Beta variants also showed the prevalence of two sets of RBM conformations, corresponding to open and closed conformations. PCA analysis of the Alpha variant trajectory shows two deep basin clusters (Figure 3.2B), basins 0 and 1, and basins 2 and 3, which correspond to open and closed conformations respectively. However, unlike the wt variant, the Alpha variant remains most of the simulation time in an open conformation ($\approx 72.64\%$, Supplementary Table S3.1). The Beta variant (Figure 3.2C) also has two deep basin clusters (basins 0 and 1, and basins 2 and 3), corresponding to open and closed conformations, respectively. Like the Alpha variant, Beta remains in an open conformation for substantially longer time than the wt ($\approx 69\%$, Supplementary Table S3.1). In both cases, and as for wt, our simulations were able to reversibly visit both states (Supplementary Figure S3.2).

Both Alpha and Beta variants shift the open/closed equilibrium towards more open conformations by roughly 20%. An opening $\Delta\Delta G$ was calculated from the ratio between the time spent in the open and closed states, where the time spent in each individual open and closed basin was added together (Supplementary Table S3.1). The equilibrium shift led to a decrease in the opening $\Delta\Delta G$ from 0.55 ± 0.17 kJ/mol, in the case of wt RBD, to -2.44 ± 0.22 and -2.09 ± 0.14 kJ/mol, for the Alpha and Beta variants, respectively. As mentioned previously, it is likely that only the open conformations are fully available to bind to ACE2, meaning that these mutations substantially increase the accessibility of RBD to ACE2, and probably impact ACE2-RBD binding.

By analyzing the intramolecular residue interactions for both variants, we observe that the interactions which stabilize the open conformation in the wt RBD are conserved in both Alpha and Beta variants, namely the triple π -stacking between residues Y489–F456–Y473, as well as the hydrogen bond between Y489 and Y473. An additional hydrogen bond between Q493 and Y453 assists in stabilizing the $\beta 6$ strand (Figures 3.3B,C).

Interestingly, in both the open and closed conformations of the Alpha variant, the interactions established by residue Y501 (Alpha's only mutation in the RBD) that were previously present in the wt variant are maintained in the Alpha variant (two hydrogen bonds established through the residue backbones: Q458–Y501 and Y501–Q506). However, the main interactions that stabilize the closed conformations differ between the Alpha variant and wt (although some transient hydrogen bonds between strand $\beta 6$ and the RBM ridge do remain). Instead of the E484–R403 salt bridge seen for wt, in the Alpha variant the closed conformation is promoted by the formation of hydrophobic interactions between the mutated Y501, V483 and F486 (Figure 3.3G). This arrangement hinders the establishment of the E484–R403 salt-bridge while being itself less stable than the open conformations. This is the likely cause for the decrease in percentage of closed state observed for Alpha. Progression to the E484–R403 salt-bridge may also be prevented in

part by the establishment of a short α -helix, discussed ahead.

In the Beta variant, the closed conformation is notably impacted by both the E484K and the N501Y mutations. The E484K mutation prevents the formation of the E484–R403 salt bridge that was crucial for the stability of the closed conformation in the wt protein. However, unlike the single E484K mutant (Supplementary Figure S3.3), the Beta variant can still reach a closed conformation. This is because it can establish the same hydrophobic interaction between Y501 and V483 as the Alpha variant (Figure 3.3H). This closed state is also stabilized by the same transient hydrogen bonds between strand $\beta 6$ and the RBM ridge seen in the wt and Alpha variants.

Concerning the secondary structure, there are no substantial differences between the Alpha or Beta open states and the wt open state (Supplementary Figure S3.9). However, upon closing, both Alpha and Delta form a small α -helix between residues 475 and 490, for roughly 30% of the simulation time. This helical character might be relevant for the Alpha variant, as it assists in facing the E484 sidechain away from R403 (Figures 3.3G,H), hindering the formation of the salt-bridge. Additionally, the Alpha variant also shows some helicity in residues 482–489, which likely arises from contacts between residues in this helix and the mutated N501Y.

Curiously, while the Alpha variant also shows a considerable decrease in SASA upon closing ($\approx 5\%$), the Beta variant shows no substantial change.

Overall, these results showcase a possible alternative mechanism for how the Alpha and Beta variants might facilitate viral entry into the host cells. By shifting the open/closed equilibrium towards the ACE2-accessible open conformation, both variants are facilitating ACE2–RBD binding, which will inevitably lead to an increase in binding affinity and enhanced receptor-dependent infection.

3.1.4.3 SARS-CoV-2 delta variant shows conformational dynamics distinct from the other variants

During the second half of 2021 the global dominant SARS-CoV-2 variant was B.1.617.2 (or Delta) [83]. It contains two mutations in the RBD region: L452R and T478K. Like the wt, Alpha and Beta variants, MD simulations of the Delta RBD show the prevalence of two sets of RBM conformations, one of which corresponds to the wt open conformation and is stabilized by the same interactions observed for the three other variants (Figure 3.3E). However, unlike those variants, MD simulations of the Delta RBD do not show the occurrence of a closed conformation at all. Instead, an alternative open conformation is present, which we refer to as “reversed”. PCA analysis of the Delta variant trajectory, shows two deep basins, 0 and 2 in Figure 3.2D, which correspond to the open and reversed conformations, respectively. Similarly to the other variants, simulations were able to reversibly visit the two states (Supplementary Figure S3.2).

The reversed conformation showcases the incredible flexibility of the RBM region, which not only opens and closes over the ACE2 binding surface of the RBD but acts as a

3.1. SARS-COV-2 VARIANTS IMPACT RBD CONFORMATIONAL DYNAMICS AND ACE2 ACCESSIBILITY

two-way hinge that leans to the side of the RBD. This alternative conformation might also prove significant advantages over the wt open state: RBD-targeting antibodies are known to bind via recognition of the RBM ridge region [93, 243]; the reversed state putatively hides this region from antibody recognition, while still providing an open ACE2 binding surface for infection.

A hydrogen bond between the mutated R452 on strand $\beta 5$ and Y449 appears to be one of the main driving forces folding the Delta variant's ridge region backwards. This interaction destabilizes the $\beta 5$ strand and enables the ridge to move up and interact with the core. Transient interactions between ridge residues G476, S477 as well as the mutated K478 with residues R346, F347 and N354 of strand $\beta 1$ stabilize the contact between the ridge loop and the RBD core, keeping it locked in place (Figure 3.3J).

Regarding the secondary structure, much like the other variants, the Delta open conformation is very similar to that of the wt (Supplementary Table S3.3). However, as expected, the reversed conformation shows substantial differences. In this state, the two small beta strands formed by residues 473–474 and 488–489, present in the open conformation, are completely lost. Additionally, the beta-sheet formed by strands $\beta 5$ and $\beta 6$ becomes less prevalent, likely due to the L452R mutation (one of the $\beta 5$ strand residues that destabilizes the β -sheet by establishing a new interaction with Y449). Curiously, like in the Alpha and Beta variants, there is also a significant alpha helical character between residues 490 and 475.

As for the closed conformations of the wt, Alpha and Beta variants, the Delta reversed conformation also leads to a decrease in SASA ($\approx 3\%$). Unlike the closed conformations, however, this alternative open conformation still presents a fully accessible ACE2 binding surface.

3.1.4.4 SARS-CoV-2 omicron variant further improves RBM accessibility compared to alpha and beta

Towards the end of 2021 a new VOC—B.1.1.529 or Omicron—overtook Delta as the dominant variant in most world regions [244]. The Omicron variant is highly distinct from other VOCs [245], containing 15 mutations in the RBD region (G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y and Y505H), 10 of which are concentrated in the RBM (N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y and Y505H). Some of these mutations are also observed, or are similar to those, in the Alpha, Beta and Delta variants: K417N, T478K, E484A and N501Y.

Unlike the other variants, MD simulations of the Omicron RBD do not show a clear prevalence of two distinct sets of RBM conformations (Figure 3.2E). While PCA analysis shows a deep basin cluster corresponding to the wt open conformation, there are only shallow basins corresponding to the closed conformation. No alternative reversed conformation is observed. The open conformation accounted for almost the entire set

of configurations sampled during the simulations of the Omicron variant ($\approx 95\%$, Supplementary Table S3.1), substantially larger than that of the wt, Alpha or Beta variants. The simulations were able to reversibly visit either state (Supplementary Figure S3.2). When compared to the wt variant, Omicron resulted in a 50% shift in the open/closed equilibrium towards more open conformations, with an opening $\Delta\Delta G$ of -8.38 ± 0.5 kJ/mol.

The triple π -stacking interactions and hydrogen bonds that stabilize the open conformation in the Omicron variant are common to those present in the wt, Alpha, Beta and Delta variants (Figure 3.3D). However, the closed conformation is quite distinct from the one seen with the other variants (Figure 3.3I). Only two transient hydrogen bonds—one between the sidechain amide of N448 and the backbone of A484, and the other between the sidechain of R493 and the backbone of C488—are in place to stabilize the closed conformation. This contrasts with the stronger interactions present in the wt, Alpha and Beta closed conformations. The wt closed conformation was stabilized by the formation of two salt bridges between K417, E484 and R403. In Omicron, E484 is mutated to an alanine preventing these two interactions. In the Alpha and Beta variants, instead of these salt bridges, several hydrophobic interactions between Y501, V483 and F486 promoted the closing of the loop. The Omicron Y501N mutation disrupts this hydrophobic core. It is clear that several mutations to the Omicron RBM actively hinder the closing of the loop while promoting the open conformation, which ultimately facilitates ACE2-RBD binding.

No substantial differences in secondary structure between the Omicron and wt variants in either the open or closed states were observed (Supplementary Figure S3.9).

It is worth noting that the Omicron variant has recently been classified into five different lineages based on their mutations, BA.1, BA.2, BA.3 [84, 246], BA.4 and BA.5 [81]. The BA.1 lineage was used in this work. Regarding the RBD region, all five lineages have 11 mutations in common (G339D, S373P, S375F, K417N, N440K, S477N, T478K, E484A, Q498R, N501Y and Y505H). BA.1 has three specific mutations (S371I, G446S and G496S), BA.2 has four (S371F, T376A, D405N and R408S), BA.3 has a combination of BA.1 and BA.2 mutations (G446S, S371F and D405N) and BA.4 and BA.5 have several mutations in common with BA.2 (S371F, T376A, D405N and R408S), plus two others (L452R and F486V).

According to our simulations, the mutations most associated with the open/closing dynamics of the RBD ridge are, for the most part, common to the BA.1, BA.2, BA.3, BA.4 and BA.5 lineages (E484A, Q493R, N501Y, and K417N). The only exception to this is the Q493R mutation in the BA.4 and BA.5 lineages, which does not occur and instead, the wt Q493 residue is maintained. Additionally, the mutations specific to the BA.2 and BA.3 lineages are relatively far from the RBM loop region and as such are unlikely to play a role in this mechanism (Supplementary Figures S3.16B,C). Two of the mutations specific to BA.1 are present in this region (G446S and G496S, Supplementary Figure S3.16A). However, these mutations did not play an obvious role in the RBD ridge dynamics, and we speculate that reverting these residues back to glycines would not impact these dynamics

significantly. Taking all of this into consideration, we expect that the RBD dynamics of the other Omicron lineages to be fairly similar to that of BA.1.

3.1.4.5 Impact of SARS-CoV-2 variants on ACE2 binding affinity

To find experimental basis for our results, we compiled ACE2-RBD binding kinetics data from recent studies [123, 247–256] (Supplementary Table S3.3). These results were obtained by surface plasmon resonance (SPR) and biolayer interferometry (BLI) and encompass data regarding both the wt and studied variants. Additionally, we compiled results obtained for just the RBD as well as for the entire S protein. While the binding kinetics values recovered from these studies are not fully consistent with each other, likely due to differences in particular experimental setups, they are mostly in the same range, and appear to follow similar trends. Regarding the equilibrium dissociation constant (K_d), all variants have an increased binding affinity when compared to the wt. However, with the currently available data, it is hard to distinguish between the efficiency of the several variants, with the Alpha and Beta variants showing a slightly better affinity than Delta.

To get more information, we analyzed both the association (k_{on}) and dissociation rate constants (k_{off}). k_{off} reflects the lifetime of the protein-protein complex and as such, the strength of the interaction. We observe a consistent decrease in k_{off} for the variants in comparison to the wt. The Alpha and Beta variants stand out from Delta and Omicron in this regard, with substantially lower k_{off} values. These results hint at the VOCs interacting more strongly with ACE2 than the wt, with the Alpha and Beta complexes being substantially more stable than those of Delta and Omicron. Several other MD studies have studied the impact of these mutations on ACE2-RBD contacts, binding affinity and binding modes, showcasing how the substantially altered ACE2-RBD interaction of the Alpha and Beta variants might be outperforming that of the wt variant [257–262]. The Delta variant does not contain mutations to the RBD ACE2 binding surface and, as such, the interactions established are not substantially different from those of wt. This is reflected in a k_{off} that is closer to, if still lower than, that of the wt.

The variants also substantially impact k_{on} . This rate constant reflects the efficiency with which protein-protein collisions lead to a bound state. While a couple of studies show no significant impact [123, 247], most show that the variants lead to a substantial increase in k_{on} , reflecting an increase in RBD accessibility to ACE2 [249–253]. We propose that this can be explained by the significant changes in RBM conformational dynamics that we have here described, where mutations lead to a decrease in the prevalence of the closed state, favoring binding. As such, our results point to an alternative mechanism for enhancing RBD-ACE2 binding, not by directly strengthening ACE2-RBD interactions, but rather by boosting, via modulation of ridge dynamics, the ACE2 binding competence.

3.1.5 Conclusion

In this work we performed AA MD simulations of the SARS-CoV-2 RBD, as well as that of the Alpha, Beta, Delta, and Omicron VOCs, to characterize the impact of the mutations on RBD conformational dynamics in solution. Our results show that the wt RBD adopts two distinct conformations in equilibrium: an open conformation where the RBD is free to bind ACE2; and a closed conformation, where the RBM ridge blocks the ACE2 binding surface and likely hinders binding to ACE2. We characterized the two states and showed that they originate from specific intramolecular interactions between residues of the RBM ridge and those of the surface that binds ACE2. As far as we know, this is the first report of this “hinge-like” mechanism, which can effectively shield the surface of RBD binding ACE2 from the solvent and binding partners. This mechanism is yet to be seen in experimentally solved RBD structures, which have thus far struggled to fully resolve the unbound RBM region [46, 62, 96]. The RBM is found unresolved in most structures due to the large flexibility of the region, and those that are fully resolved are often structures of RBD complexed with either ACE2 [30, 67, 90, 112, 113], antibodies [114–120] or itself by dimerizing via the surface binding ACE2 [263, 264].

The four variants tested in this work significantly impacted the open/closed equilibrium we observed for wt RBD. Both Alpha and Beta variants shifted the equilibrium towards more open conformations by roughly 20%. In Omicron the open conformation accounted for 96% of simulation time while the Delta variant did not show the presence of a closed conformation at all. This shift towards more open conformations likely enhances ACE2 binding affinity by increasing accessibility to the RBM and facilitating binding. Several experimental binding studies have shown that these variants lead to a substantial increase in ACE2-RBD binding association rate constant, reflecting an increased ACE2 accessibility, in agreement with our findings.

Additionally, the Delta variant showed an alternative open conformation, distinct from that of the other variants. This alternative conformation keeps the ACE2 binding surface open and accessible for binding, but significantly alters the conformation of the RBM ridge. This state presents a substantially altered ridge region, which bends backwards towards the RBD core, shielding some of it from exposure. We hypothesize that this may provide a fitness advantage by aiding in antibody escape since many RBD-targeting antibodies bind to the RBM ridge region [115, 243, 265, 266]. These RBD-targeting antibodies are also more sensitive to viral evolution than antibodies that bind other regions of the RBD [267]. In the alternative open conformation, the ridge may be not as easily recognized, while the ACE2 binding surface remains unobstructed for infection. The substantially different conformational dynamics of the RBM region between the variants, correlates well with the hypothesis proposed by Quaglia et al. that mutations are enriched at intrinsically disordered regions of the SARS-CoV-2 proteome and that they may contribute towards immune evasion [268]. These results show that the mutations found in the four VOCs studied impact RBD conformational dynamics in a direction that promotes efficient

3.1. SARS-COV-2 VARIANTS IMPACT RBD CONFORMATIONAL DYNAMICS AND ACE2 ACCESSIBILITY

binding to ACE2 and (in the case of the Delta variant) antibody escape, an effect that has thus far been disregarded. In this context, our findings can also help explain some of the antibody-evading characteristics of the emergent Omicron variant.

3.2 Signatures in SARS-CoV-2 spike protein conferring escape to neutralizing antibodies

3.2.1 Abstract

Understanding SARS-CoV-2 evolution and host immunity is critical to control COVID-19 pandemics. At the core it is an arms-race between SARS-CoV-2 antibody and angiotensin-converting enzyme 2 (ACE2) recognition, a function of the viral protein spike. Mutations in spike impacting antibody and/or ACE2 binding are appearing worldwide, imposing the need to monitor SARS-CoV2 evolution and dynamics in the population. Determining signatures in SARS-CoV-2 that render the virus resistant to neutralizing antibodies is critical. We performed *In silico* analysis of the hotspot residues for S protein-antibody and S protein-ACE2 interactions. This highlighted the importance of the receptor binding domain (RBD), especially the receptor binding motif (RBM) in the S protein for antibody binding, aligning with its role in preventing ACE2 interaction and viral infection. Notably, E484 stands out as a critical residue for antibody binding, consistent with its presence in highly prevalent variants capable of reinfection. Additionally, we predict that mutations at residue 494 may facilitate antibody escape without significantly altering ACE2 binding affinity. While our study has limitations in representing the full diversity of patient antibodies and biases associated with structural focus, our results strongly correlate with experimental evidence of antibody escape and variant emergence. This analysis provides a valuable tool for identifying mutation hotspots requiring close monitoring as the pool of available structural data continues to expand.

3.2.2 Introduction

The emergence of SARS-CoV-2, the virus responsible for the COVID-19 pandemic, has posed a fundamental question: Can it evolve to evade immunity, whether acquired through natural infection or vaccination? The answer to this multi-layered and complex question depends on the type, duration and heterogeneity of the selective pressures imposed by the host and by the environment to the virus, but also on the rate and phenotypic impact of the mutations the virus acquires. The central question is whether the virus can select mutations that escape host immunity while remaining efficient to replicate in the host [206, 269–271]. Identifying immune evasion signatures in the virus is critical for preparedness of future interventions to control COVID-19.

Viral mutations have multifaceted impacts, affecting transmission, virulence, immunity, therapies, detection, and host range [272, 273]. Monitoring these mutations is crucial, especially considering the dynamic evolution of SARS-CoV-2. Initially, the virus displayed neutral genetic drift, with the D614G mutation being a notable exception, enhancing transmissibility [274]. However, since late 2020, several concerning variants have emerged globally, including B.1.1.7, B.1.1.7 [United Kingdom (UK)] [80], B.1.351 [South Africa (SA)] [81], P.1 (Brazil) [82], B.1.617 (India) [83], as well as several lineages of interest that

carry the mutation E484K. These variants exhibit increased transmission, virulence, and immune evasion [275, 276]. Mutations in the spike protein's receptor binding domain (RBD), like D614G, N501Y, and L452R, enhance ACE2 binding and transmission [277]. Some mutations, such as E484K, reduce neutralization [278]. Shared mutations in different lineages suggest possible fitness advantages. Understanding the collective impact of these mutations in spike is crucial.

The alarm caused by the emergence of SARS-CoV-2 variants with the potential to escape host immunity acquired through infection or vaccination has been confirmed by a series of reports for lineages B.1.351 and P.1. [279–283]. These raise concerns on whether a reduction in vaccine efficacy could result in re-infections and delay the reduction in mortality caused by circulating SARS-CoV-2. There is, therefore, an urgent need to develop mutation-tolerant vaccines (and biopharmaceuticals) targeting the spike protein. In this sense, it is critical to determine what makes mutations well tolerated for the viral lifecycle whilst efficiently escaping immunity. In this work, we then used available structures of the complexes spike-antibodies and spike-ACE2 to determine the frequency of interactions of each amino acid residue of the RBD with either ACE2 or antibodies. We show that the amino acid residues at positions 484 and 494 frequently engage in binding to antibodies but not in binding to the receptor ACE2. Our work suggests that the amino acid residues at the RBD that are more dispensable for binding to ACE2 can more promptly evolve immune escape mutants if the amino acid residue substitution severely alters the binding specificity of the antibody.

3.2.3 Methods

3.2.3.1 S protein-antibody contact analysis

In order to determine which residues of the SARS-CoV-2 S protein contribute the most to the binding of antibodies, we studied 57 PDB structures containing the S protein (or only the RBD region) bound to antibodies (Table S3.4). These complexes were chosen based on their availability in the PDB repository [284] in March 2022 and are all neutralizing antibodies. We first used the MDAnalysis library [233, 285] to identify the residues of the antibody and the S protein located in the interface between the two proteins. This was done for all PDB structures using in-house python scripts. To determine the relevance of each S protein/RBD residue in the binding, a distance cut-off value of 4.5 Å was applied as a criterion (i.e. a contact is observed when two residues have a minimum distance lower than 4.5 Å). Finally, the numerical python library (NumPy [286]) was used to calculate the frequency of contact for each S protein residue with an antibody residue, from the sample of PDB structures analysed.

3.2.3.2 MD simulation of the RBD-ACE2 complex

Molecular dynamics (MD) simulations of the RBD bound to the ACE2 protein were performed with the GROMACS 2020.3 package [220], using the Amber14sb [185] forcefield, starting from the 6m0j structure [90], in a truncated dodecahedron box filled with water molecules (minimum of 1.2 nm between protein and box walls). The TIP3P water model [221] was used and the total charge of the system (-23, including the constitutive Zn^{2+} and Cl^{-} ions bound to ACE2) was neutralized with 23 Na^{+} ions. Additional Na^{+} and Cl^{-} ions were added to the solution to reach an ionic strength of 0.1M. The system was energy-minimized using the steepest descent method for a maximum of 50000 steps using position restraints on the heteroatom positions by restraining them to the crystallographic coordinates using a force constant of 1000 kJ/mol in the X, Y and Z positions. Before performing the production runs, an initialization process was carried out in 5 stages of 100 ps each. Initially all heavy-atoms were restrained using a force constant of 1000 kJ/mol/nm, and at the final stage only the only C- α atoms were position-restrained using the same force constant. In the first stage, the Berendsen temperature algorithm [195] was used to initialise and maintain the simulation at 300K, using a temperature coupling constant of 0.01ps, without pressure control. The second stage continue to use the Berendsen temperature algorithm but with a force constant of 0.1ps. The third stage kept the same temperature control settings, but introduced pressure coupling with the Berendsen pressure algorithm [195] with a pressure coupling constant of 5.0 ps and applied isotropically. The fourth stage changed the temperature algorithm to V-rescale [223], with a temperature coupling constant of 0.1ps, and the pressure algorithm to Parrinello-Rahman [224] with a temperature coupling constant of 5.0ps. The fifth stage is equal to the fourth stage, but position restraints are only applied to C- α atoms. During the simulation, the equations of motion were integrated using a timestep of 2 fs. The temperature was maintained at 300 K, using the V-rescale [223] algorithm with a time constant of 0.1 ps, and the pressure was maintained at 1 bar using the Parrinello-Rahman [224] pressure coupling algorithm, with a time constant of 5 ps; pressure coupling was applied isotropically. Long-range electrostatic interactions were treated with the PME [189, 225] method, using a grid spacing of 0.12 nm, with a cubic interpolation. The neighbour list was updated every twenty steps and the cutoff scheme used was Verlet with 0.8nm as the cut-off radius. All bonds were constrained using the LINCS algorithm [226]. The system was simulated for 8 μ s in 5 replicates (to a total of 40 μ s).

3.2.3.3 RBD-ACE2 contacts frequency

In order to determine which residues contribute the most for the interaction between RBD and ACE2, and the persistence of these interactions, we performed a contact analysis throughout the simulation. We started by eliminating the first 2 μ s of equilibration of all replicates. The MDAnalysis library [233, 285] was then used to pinpoint the residues of the RBD that are in contact with the ACE2 protein. A distance cut-off value of 4.5 Å

3.2. SIGNATURES IN SARS-COV-2 SPIKE PROTEIN CONFERRING ESCAPE TO NEUTRALIZING ANTIBODIES

was applied as a criterion (i.e. a contact is observed when two residues have a minimum distance lower than 4.5 Å). Finally, we determined the percentage of time for which a given RBD residue is at less than 4.5 Å of ACE2.

3.2.4 Results

To determine which S protein residues are more frequently targeted by neutralizing antibodies, we analysed 57 PDB structures containing the S protein (or only the RBD region) bound to antibodies (Table S3.4). Our results revealed that the receptor binding motif (RBM), in the receptor binding domain (RBD), is the region with the highest frequency of contacts (Figure 3.4A). This region is important for the SARS-CoV-2 infection since it mediates contacts with the angiotensin converter enzyme (ACE2) [91, 287, 288] and, for this reason, the binding of antibodies to this region would hinder ACE2 binding and consequently the entry of the virus in the cell. In addition to contacts with the RBD, we also identified other regions of the S protein with antibody binding sites. The antibodies FC05 and DH1050.1 bind to segments in the N-terminal region (between residues 143-152 and 246-257) and 2G12 binds to residues 936-941 in the stalk (Figure 3.4 C).

To choose the cut-off above which to select the most important residues for antibody binding, we built a histogram (Figure S3.17). Three distinct groups appear: the first includes residues that have an interaction probability below 20%, the second includes residues whose interaction is between 20% and 45% and the last corresponds to frequent binders (interaction probability over 45%). Within this last group we find Y449, L455, F456, E484, F486, N487, Y489, Q493, S494, and Y505 residues (Figure 3.4A and B). More than half of these residues have hydrophobic side chains, showing that, within the batch of antibodies we studied, the most common binding mode is through hydrophobic contacts. These residues are important for antibody binding, which means that mutations within this group may enable the virus to escape antibodies from patients previously exposed to the WT or another variant, or people vaccinated using WT sequences of the S protein. In fact, prior studies have shown that mutations on most of these residues influence antibody escape [269, 289]. For residues L455, F456 and F486, mutations for less hydrophobic residues have been shown to reduce binding by polyclonal antibodies [269] and mutants at site 487 escape human monoclonal antibodies COV2-2165 and COV2-2832 [289, 290]. Interestingly, our results also show that there is a high prevalence of antibodies binding to E484 (50% - Figure 3.4B). Mutations in this residue have been demonstrated to lead to antibody escape [93, 271, 289]. Importantly, this mutation is present in both the South African [81] and Brazilian variants [291] and there is evidence that it contributes to antibody escape [269].

To be able to escape antibodies while maintaining the ability to efficiently infect cells, the mutations introduced must destabilize the interaction with antibodies while keeping a high binding affinity for ACE2. Thus, we reason that mutations that occur in residues that are relevant for antibody binding but not very relevant for the interaction with ACE2

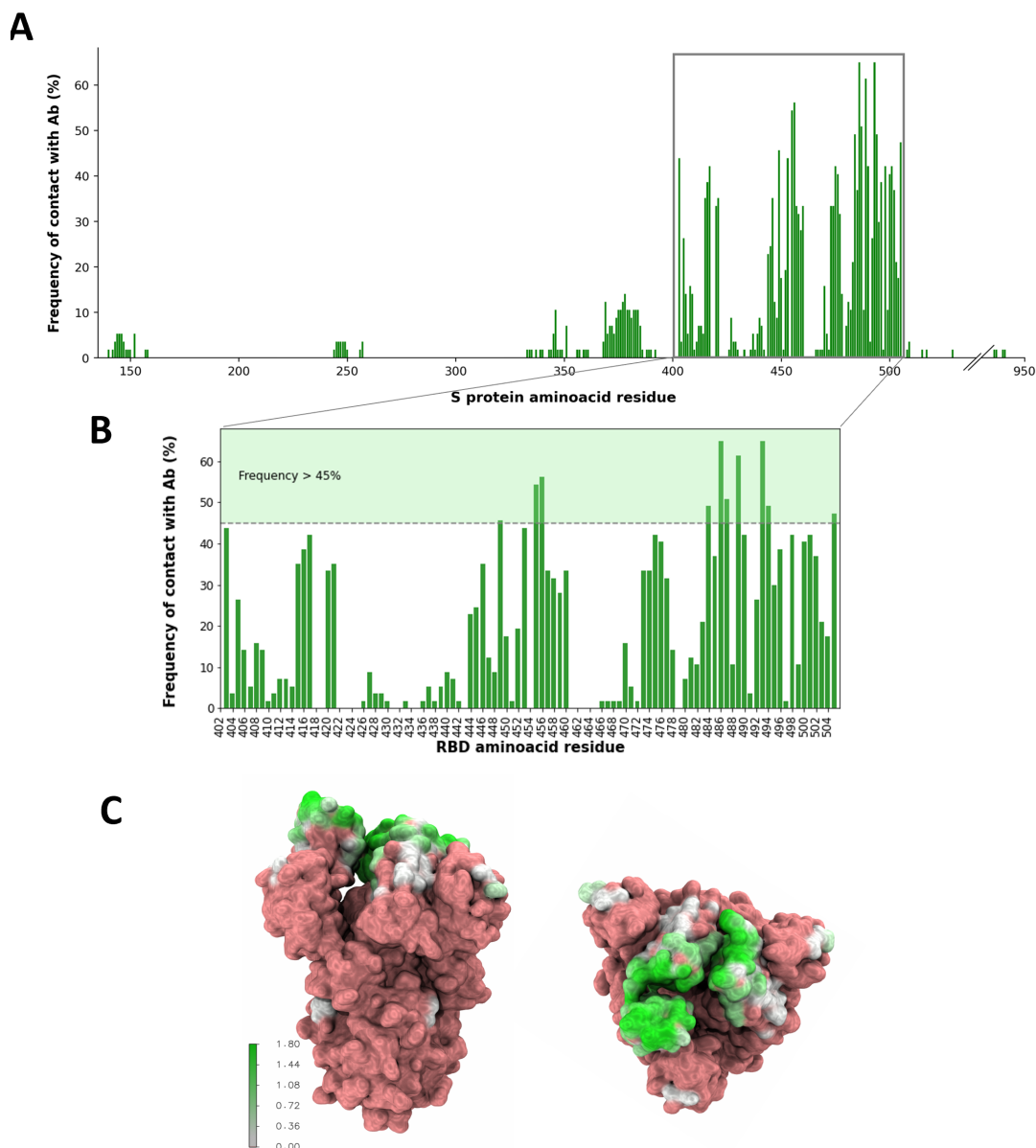


Figure 3.4: S protein-antibody contact analysis. (A) Histogram of the contact frequency of every S protein residue with antibodies. using a cut-off value $4.5 \text{ \AA}$ to define a contact (B) Zoom in of the histogram shown in (A) in the region comprising the residues with the highest frequency of antibody contacts (located in the RBD). (C) Antibody contact map represented on the S protein surface, seen from front (left side) and from top (right side); the map was built with VMD[227] and the colour gradient is used to represent the log of the frequency of contact with Ab as shown in the scale bar.

can be advantageous for the virus. To gain further insights into this subject, we performed MD simulations of the RBD bound to ACE2. These allowed us to determine which RBD residues are relevant for binding to ACE2, and the persistence of these interactions (Figure 3.5A). Combining the information on Antibody and ACE2 binding allows us to predict which are the most relevant hotspots (high affinity for antibodies and low affinity for

3.2. SIGNATURES IN SARS-COV-2 SPIKE PROTEIN CONFERRING ESCAPE TO NEUTRALIZING ANTIBODIES

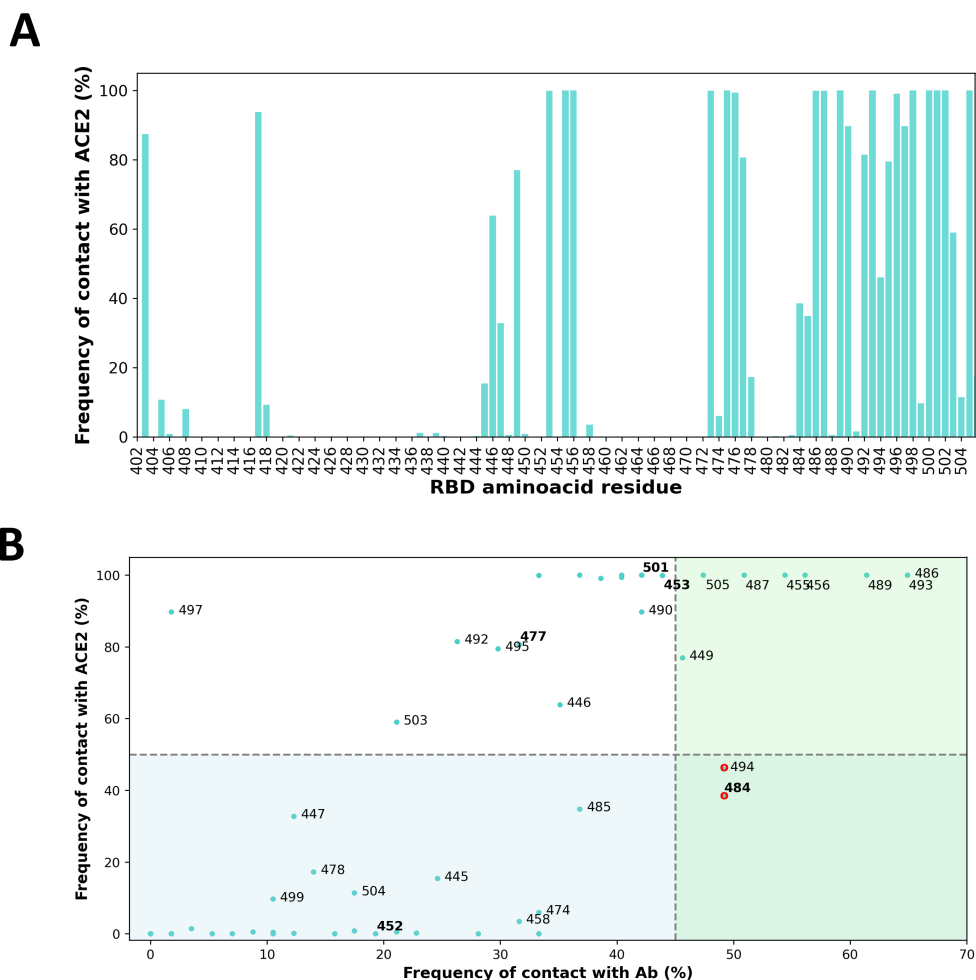


Figure 3.5: RBD-ACE2 contact analysis correlated with the antibody binding data. (A) Histogram of the contact frequency of RBD residues with ACE2, obtained from MD simulations, using a cut-off value 4.5 Å to define a contact. (B) Scatter plot correlating the RBD residues' frequency of contact with ACE2 and antibodies; each point represents one residue and those that were studied experimentally are highlighted in bold; the quadrants are filled with different colours to highlight the relevance of different residues as potential mutation hotspots, with the most relevant residues located in quadrant IV and shown as red points.

ACE2). Using these criteria, two residues stand out: E484 and S494 (Figure 3.5B, IV quadrant circled in red). These results are consistent with the evidence that E484 is an important mutation site and brings to light a new relevant site: S494. Mutations in this residue have been shown to reduce the binding by polyclonal plasma antibodies [269] and have no visible effect in the RBD-ACE2 binding [271]. This leads us to hypothesize that a mutation in this residue can, not only facilitate the escape of the virus to antibodies, but also maintain the ability to bind to the receptor and enter into host cells. The mutation S494P has been found in nature with a considerable prevalence, which is in agreement with this hypothesis.

On the other hand, some residues that are important for antibody binding are also involved in frequent interactions with ACE2. Seven of these residues remained bound to ACE2 throughout the whole simulation in all the replicates: L455, F456, F486, N487, Y489, Q493 and Y505 (Figure 3.5B, top residues in quadrant II). This group includes residue F456, which was found to be a hotspot for escape mutations [269]. Although this may seem contradictory, the same study shows that mutations in this residue were very rarely found in nature [269]. We hypothesise that the low mutation rate for this site may be due to the high relevance of this residue for RBD-ACE2 binding. In fact, deep mutational analysis shows that when mutating F456 to any aminoacid residue besides itself or its SARS-CoV counterpart (a leucine) the ACE2 binding affinity is reduced [271]. Even though a viral strain with a single mutation at the 456 site could escape antibodies, it would be unfit to subsist in nature, due to a reduced ability to infect host cells.

3.2.5 Conclusion

Our analysis of the S protein-antibody contacts revealed that (within the available data set) the RBM, in the RBD, is an important region for antibody binding, which makes sense, since antibody binding to this region can prevent the interaction with ACE2 and ultimately impair viral infection. We also observed that the E484 is relevant for antibody binding, in agreement with our and previous experimental results and with the occurrence of this mutation in two variants that are becoming highly prevalent and were shown to be able to re-infect people previously exposed to other SARS-CoV-2 variants. Additionally, we predict that mutations in residue 494 may be problematic, as they could facilitate antibody escape without substantially altering the affinity for the ACE2 binding.

The pool of complexes between the S protein/RBD analysed here is not necessarily representative of the universe of antibodies found in COVID-19 patients, since it is limited by the availability of structural information. All the antibodies analysed are neutralizing antibodies, which is not always the case in nature. Additionally, many structural studies have focused on the interaction with the RBD, while other regions have not been so exhaustively analysed, which creates a bias in terms of the number of RBD binding antibodies with known structure. Limitations of the methods used, such as crystallographic contacts, can also influence the results. Nevertheless, its combination with data from molecular simulation and experimental analysis of mutations helps to predict and explain the emergence of new mutations in nature.

Our results are in very good agreement with experimental data on mutation-driven escape to polyclonal antibodies and with the emergence of natural variants, which are able to re-infect previously exposed people. Thus, this analysis represents a simple strategy to predict mutation hotspots that should be kept under surveillance. The predictions will become even better as the pool and diversity of available structures increase.

3.3 Supplementary Information of Exploring the impact of mutations on the SARS-CoV-2 RBD conformational dynamics and antibody neutralization

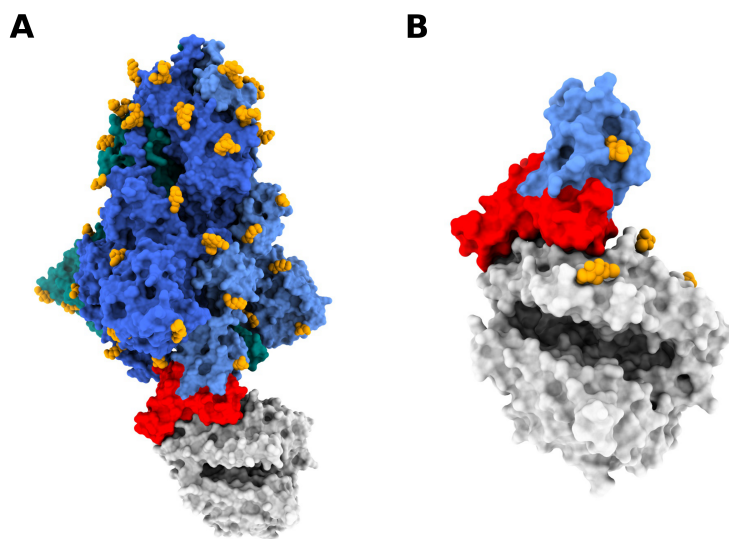


Figure S3.1: ACE2 bound to glycosylated SARS-CoV-2 S protein. (A) ACE2 (White) bound to SARS-CoV2 S glycoprotein (blue). (B) Closeup of ACE2 bound RBD. RBM is rendered in red. Glycosylations are rendered in orange.

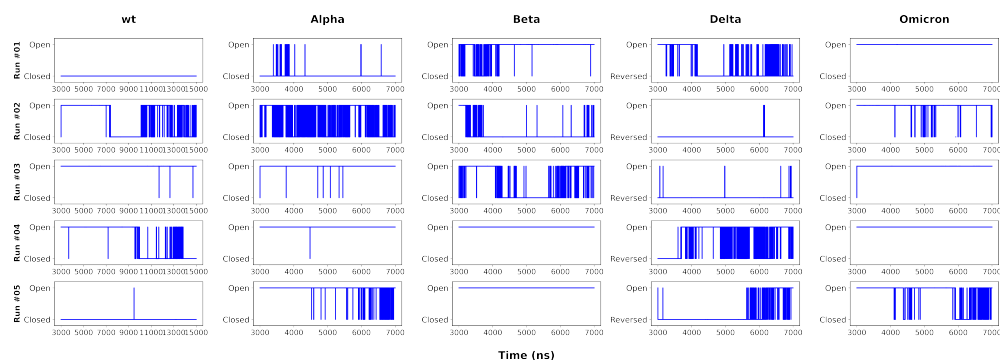


Figure S3.2: RBD open – closed dynamics over simulation time. RBD open – closed dynamics as determined by analysis of the conformational basins (Supplementary Table S1) recovered by PCA. Data shown for the five replicas for each variant tested. The first 3 μ s of simulation were used for equilibration.

Table S3.1: Energy surface landscape analysis from 2D PCA of SARS-CoV-2 RBD conformational dynamics in water. Energy surface landscape analysis and defined basins for each of the tested RBD variants. Energy minima, frame percentage and loop conformation for each of the basins is also given. Overall analysis of “open” vs. “closed” conformation is also shown. 95 % Confidence intervals (CI) were calculated with bootstrap resampling from the frame percentages recovered from the individual simulation replicas. Representative structures can be seen in figures S3.8, S3.9, S3.10 and S3.11.

	Basin	Free Energy	$\langle E \rangle / k_B T$	S/R	$E_{min} / k_B T$	Frame percentage (%)	Loop Conformation	Time In "closed" state (%) $\pm$ CI (95%)	Time In "open" state (%) $\pm$ CI (95%)	"closing" $\Delta\Delta G$ (kJ/mol)
WT	0	-5.16	1.36	6.52	0.44	37.69	"Closed"	55.50 $\pm$ 2.12 - 2.93	44.49 $\pm$ 0.93 - 1.12	0.5 $\pm$ 0.21 - 0.13
	1	-4.67	0.99	5.67	0.26	23.01	"Open"			
	2	-4.41	0.57	4.98	0.00	17.66	"Closed"			
	3	-4.20	1.04	5.24	0.44	14.42	"Open"			
	4	-3.28	3.62	6.90	2.86	5.77	"Open"			
	5	-1.57	4.07	5.64	3.61	1.04	"Open"			
	6	0.79	5.30	4.51	5.02	0.13	"Open"			
	7	0.83	5.42	4.60	5.21	0.14	"Closed"			
	8	0.91	5.39	4.48	5.01	0.12	"Open"			
	9	2.70	5.06	2.37	5.01	0.01	"Closed"			
Total						99.99				
Alpha	0	-5.20	1.02	6.22	0.00	39.18	"Open"	27.31 $\pm$ 1.79 - 2.04	72.64 $\pm$ 1.16 - 1.43	-2.44 $\pm$ 0.21 - 0.23
	1	-4.64	1.47	6.11	0.35	22.41	"Open"			
	2	-4.37	1.84	6.21	0.83	17.23	"Closed"			
	3	-3.82	0.96	4.78	0.35	9.85	"Closed"			
	4	-3.10	2.77	5.86	1.83	4.82	"Open"			
	5	-2.71	3.88	6.58	3.01	3.28	"Open"			
	6	-2.34	4.10	6.44	3.11	2.41	"Open"			
	7	-0.28	5.21	5.49	4.93	0.29	"Open"			
	8	-0.11	4.43	4.54	4.05	0.25	"Open"			
	9	0.11	5.46	5.35	5.10	0.23	"Closed"			
Total						99.95				
Beta	0	-4.95	0.58	5.53	0.00	23.73	"Open"	30.18 $\pm$ 1.24 - 1.74	69.81 $\pm$ 0.66 - 0.73	-2.09 $\pm$ 0.13 - 0.15
	1	-4.70	0.46	5.16	0.09	18.41	"Open"			
	2	-4.38	1.31	5.69	0.90	13.43	"Closed"			
	3	-3.98	1.32	5.30	0.90	9.03	"Closed"			
	4	-3.83	2.70	6.53	1.93	7.72	"Closed"			
	5	-3.78	2.86	6.64	2.02	7.36	"Open"			
	6	-3.36	2.93	6.30	2.30	4.86	"Open"			
	7	-3.20	2.20	5.40	1.68	4.14	"Open"			
	8	-2.93	4.08	7.02	3.32	3.21	"Open"			
	9	-2.71	3.66	6.37	3.11	2.54	"Open"			
	10	-2.58	3.53	6.12	2.98	2.23	"Open"			
	11	-2.14	3.70	5.85	3.28	1.44	"Open"			
	12	-1.74	3.78	5.52	3.48	0.95	"Open"			
	13	-1.72	3.99	5.71	3.70	0.94	"Open"			
Total						99.99				
Omicron	0	-4.77	1.33	6.11	0.17	37.27	"Open"	3.35 $\pm$ 0.33 - 0.55	96.59 $\pm$ 2.01 - 2.22	-8.38 $\pm$ 0.50 - 0.29
	1	-4.71	0.88	5.59	0.00	35.10	"Open"			
	2	-3.78	3.06	6.83	1.84	13.87	"Open"			
	3	-3.49	3.12	6.61	2.11	10.35	"Open"			
	4	-2.19	3.12	5.31	2.27	2.83	"Closed"			
	5	-0.46	4.37	4.83	3.89	0.52	"Closed"			
Total						99.94				

Table S3.2: Surface Accessible Surface Area (SASA) analysis of SARS-CoV-2 RBD in water. SASA values were calculated using the GROMACS tool gmx sasa for the whole trajectory (Entire trajectory) and for the two major conformations ("open"/"closed" and "open"/"reversed"). Results were also divided in the contribution of hydrophobic atoms, which are the ones with charges $[-0.2, 0.2]$, and hydrophilic, those outside of this range. 95 % Confidence intervals (CI) were calculated with bootstrap resampling.

Variant	Protein SASA (nm ²)	Hydrophobic atoms SASA (nm ²)	Hydrophilic atoms SASA (nm ²)
	Average $\pm$ CI (95%)	Average $\pm$ CI (95%)	Average $\pm$ CI (95%)
WT			
<i>Average</i>	107.72 $\pm$ ^{0.07} _{0.07}	55.07 $\pm$ ^{0.05} _{0.05}	52.62 $\pm$ ^{0.05} _{0.05}
<i>"Open"</i>	109.08 $\pm$ ^{0.05} _{0.05}	55.11 $\pm$ ^{0.04} _{0.04}	53.97 $\pm$ ^{0.04} _{0.05}
<i>"Closed"</i>	105.52 $\pm$ ^{0.05} _{0.05}	53.68 $\pm$ ^{0.04} _{0.04}	51.57 $\pm$ ^{0.04} _{0.04}
Alpha			
<i>Average</i>	108.16 $\pm$ ^{0.04} _{0.04}	55.13 $\pm$ ^{0.03} _{0.03}	53.04 $\pm$ ^{0.03} _{0.03}
<i>"Open"</i>	108.62 $\pm$ ^{0.06} _{0.06}	55.26 $\pm$ ^{0.04} _{0.04}	53.36 $\pm$ ^{0.04} _{0.04}
<i>"Closed"</i>	102.99 $\pm$ ^{0.07} _{0.07}	52.69 $\pm$ ^{0.04} _{0.04}	50.31 $\pm$ ^{0.05} _{0.05}
Beta			
<i>Average</i>	108.10 $\pm$ ^{0.01} _{0.01}	54.95 $\pm$ ^{0.01} _{0.01}	53.15 $\pm$ ^{0.01} _{0.01}
<i>"Open"</i>	108.58 $\pm$ ^{0.06} _{0.06}	53.70 $\pm$ ^{0.04} _{0.04}	53.88 $\pm$ ^{0.04} _{0.04}
<i>"Closed"</i>	107.65 $\pm$ ^{0.08} _{0.08}	56.32 $\pm$ ^{0.05} _{0.05}	51.33 $\pm$ ^{0.05} _{0.06}
Delta			
<i>Average</i>	109.62 $\pm$ ^{0.11} _{0.11}	55.46 $\pm$ ^{0.08} _{0.08}	54.16 $\pm$ ^{0.08} _{0.07}
<i>"Open"</i>	109.58 $\pm$ ^{0.05} _{0.05}	54.87 $\pm$ ^{0.04} _{0.04}	54.70 $\pm$ ^{0.04} _{0.04}
<i>"Reversed"</i>	106.08 $\pm$ ^{0.07} _{0.07}	53.65 $\pm$ ^{0.04} _{0.04}	52.43 $\pm$ ^{0.04} _{0.04}
Omicron			
<i>Average</i>	110.97 $\pm$ ^{0.04} _{0.04}	56.84 $\pm$ ^{0.03} _{0.03}	54.13 $\pm$ ^{0.02} _{0.02}
<i>"Open"</i>	110.83 $\pm$ ^{0.05} _{0.05}	56.68 $\pm$ ^{0.04} _{0.04}	54.16 $\pm$ ^{0.04} _{0.04}
<i>"Closed"</i>	109.73 $\pm$ ^{0.07} _{0.07}	57.07 $\pm$ ^{0.07} _{0.07}	52.66 $\pm$ ^{0.05} _{0.05}

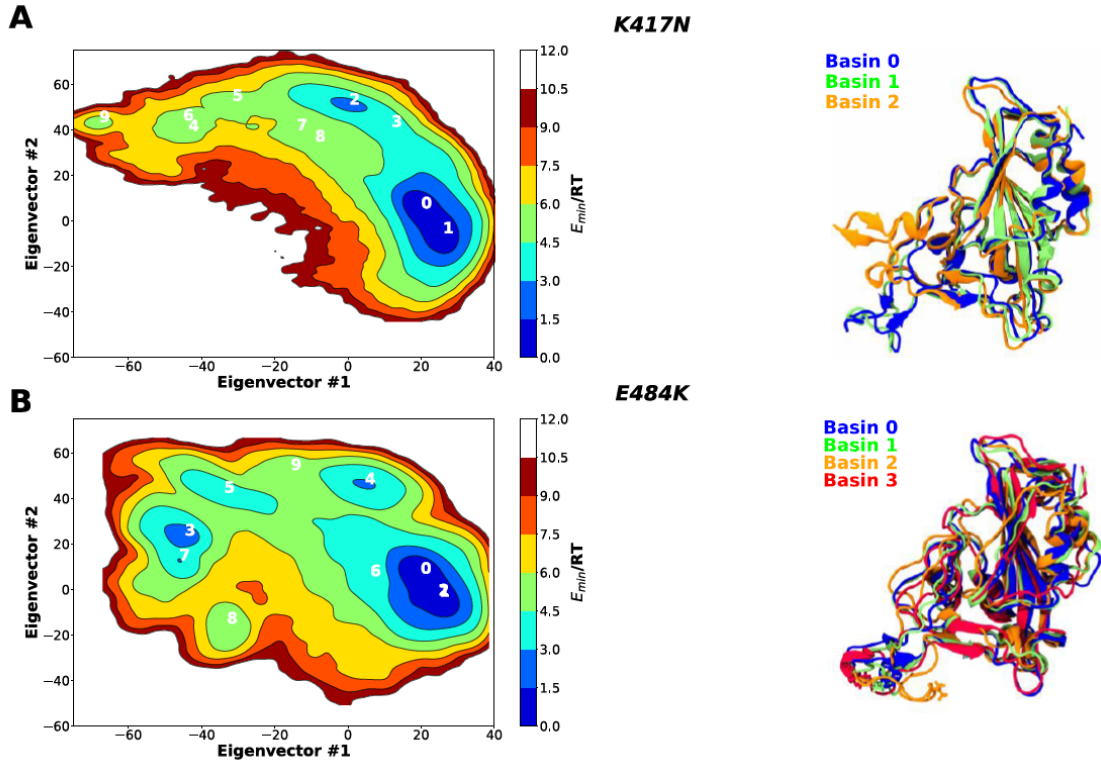


Figure S3.3: Two-dimension principal component analysis (PCA) of SARS-CoV-2 RBD's mutants conformational dynamics in water. Plots of the first two principal components determined from the $C\alpha$ backbone of the (A) K417N and (B) E484K RBD mutants. Snapshots of the lowest energy structures for selected basins are also shown.

Reference	RBD/Spike	Technique	WT			Alpha			Beta			Delta			Omicron		
			$K_{on}(M^{-1}s^{-1})$	$K_{off}(s^{-1})$	$K_d(nM)$	$K_{on}(M^{-1}s^{-1})$	$K_{off}(s^{-1})$	$K_d(nM)$	$K_{on}(M^{-1}s^{-1})$	$K_{off}(s^{-1})$	$K_d(nM)$	$K_{on}(M^{-1}s^{-1})$	$K_{off}(s^{-1})$	$K_d(nM)$	$K_{on}(M^{-1}s^{-1})$	$K_{off}(s^{-1})$	$K_d(nM)$
McCallum et al. 2021	RBD	SPR	7.70×10^4	6.70×10^{-3}	8.3	7.50×10^4	1.20×10^{-3}	15	-	-	-	5.90×10^6	4.30×10^{-3}	63	-	-	-
Tian et al. 2021	RBD	SPR	2.50×10^4	2.10×10^{-4}	8.3	7.50×10^4	0.37×10^{-4}	0.5	6.40×10^4	0.30×10^{-4}	0.5	-	-	-	-	-	-
LaBédre et al. 2021	RBD	SPR	4.50×10^4	7.80×10^{-3}	17	5.70×10^4	1.30×10^{-4}	2.4	7.60×10^4	4.30×10^{-3}	5.8	-	-	-	-	-	-
Supasa et al. 2021	RBD	SPR	3.88×10^4	2.92×10^{-3}	7.51	6.38×10^4	6.85×10^{-4}	10.7	-	-	-	-	-	-	-	-	-
Wörnberger et al. 2021	RBD	SPR	6.86×10^4	11.0×10^{-3}	16.2	4.36×10^4	1.59×10^{-3}	3.72	6.86×10^4	4.42×10^{-3}	6.5	10.1×10^4	8.23×10^{-3}	8.61	-	-	-
de Souza et al. 2021	RBD	SPR	0.90×10^6	91.6×10^{-4}	10.3	1.30×10^6	15.5×10^{-4}	1.2	1.20×10^6	39.4×10^{-4}	3.3	-	-	-	-	-	-
Lan et al. 2022	RBD	SPR	1.24×10^6	7.35×10^{-3}	59	-	-	-	-	-	-	-	-	-	1.73×10^6	4.86×10^{-3}	28
Dejnattasi et al. 2022	RBD	SPR	7.00×10^4	5.10×10^{-3}	7.3	1.10×10^6	1.60×10^{-3}	1.5	1.30×10^6	4.00×10^{-3}	3.2	1.20×10^6	5.80×10^{-3}	4.9	8.20×10^4	6.4×10^{-3}	7.8
Meng et al. 2022	RBD	BLI	1.10×10^5	1.40×10^{-2}	127	-	-	-	-	-	-	5.90×10^6	1.1×10^{-2}	190	1.10×10^5	4.70×10^{-3}	44
Junker et al. 2022	RBD	BLI	1.26×10^5	1.72×10^{-4}	1.26	-	-	-	-	-	-	1.24×10^5	1.19×10^{-4}	0.96	1.85×10^5	1.45×10^{-4}	0.78
de Souza et al. 2021	Spike	SPR	0.09×10^6	5.80×10^{-4}	6.4	0.10×10^6	1.70×10^{-4}	0.1	0.30×10^6	3.00×10^{-4}	0.3	-	-	-	-	-	-
Saville et al. 2021	Spike	BLI	1.40×10^5	7.09×10^{-4}	5.06	-	-	-	-	-	-	1.51×10^5	4.01×10^{-4}	2.65	-	-	-
Yang et al. 2021	Spike	BLI	3.43×10^4	10.6×10^{-5}	3.1	3.86×10^4	5.25×10^{-5}	1.36	7.32×10^4	1.79×10^{-5}	0.25	4.39×10^6	1.71×10^{-5}	0.39	-	-	-

Table S3.3: Compilation of ACE2-RBD binding kinetics data from recent studies. Kinetic parameters of ACE2 binding to wt, alpha, beta, delta and omicron RBD/Spike variants data obtained from SPR and BLI.

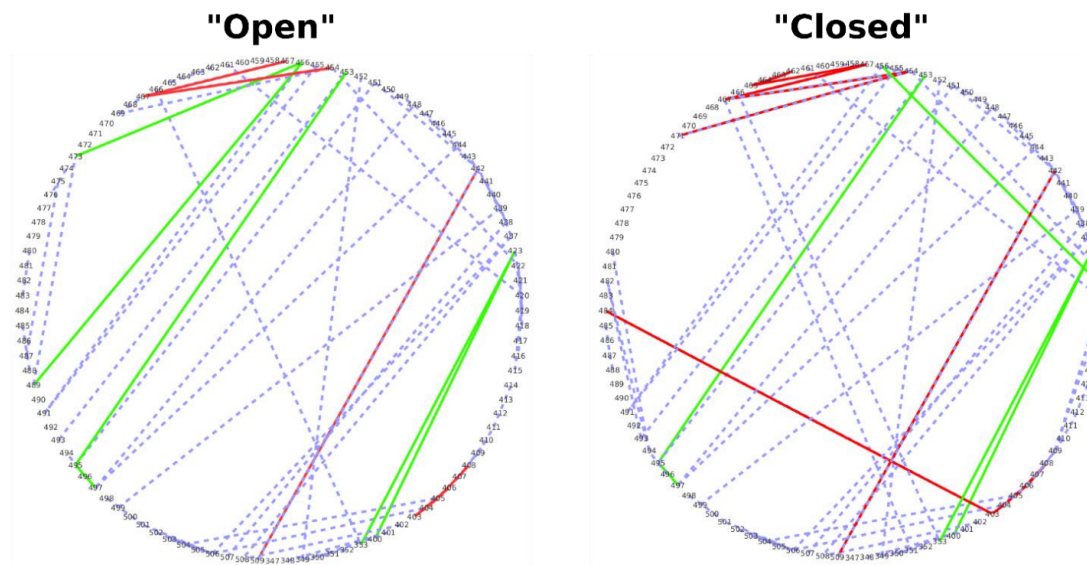


Figure S3.4: Residue interaction networks (RINs) for the “open” and “closed” conformations of wt RBD. RINs determined using RIP-MD for the 5000 lowest energy conformations obtained for the most populated “open” and “closed” basins of wt RBD. Hydrogen bonds, salt bridges and pi-pi interactions are shown in blue, red and green, respectively.

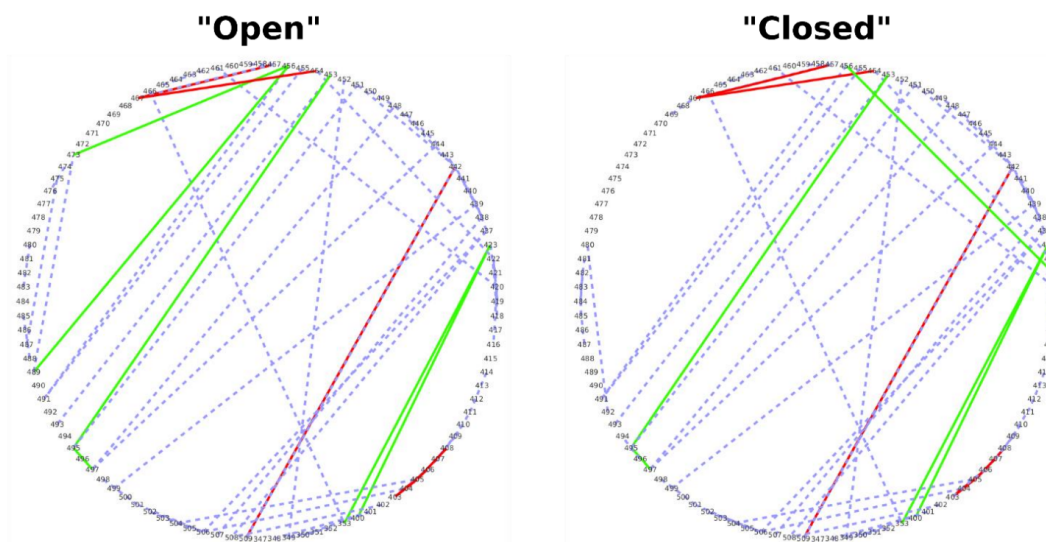


Figure S3.5: Residue interaction networks (RINs) for the “open” and “closed” conformations of alpha RBD. RINs determined using RIP-MD for the 5000 lowest energy conformations obtained for the most populated “open” and “closed” basins of wt RBD. Hydrogen bonds, salt bridges and pi-pi interactions are shown in blue, red and green, respectively.

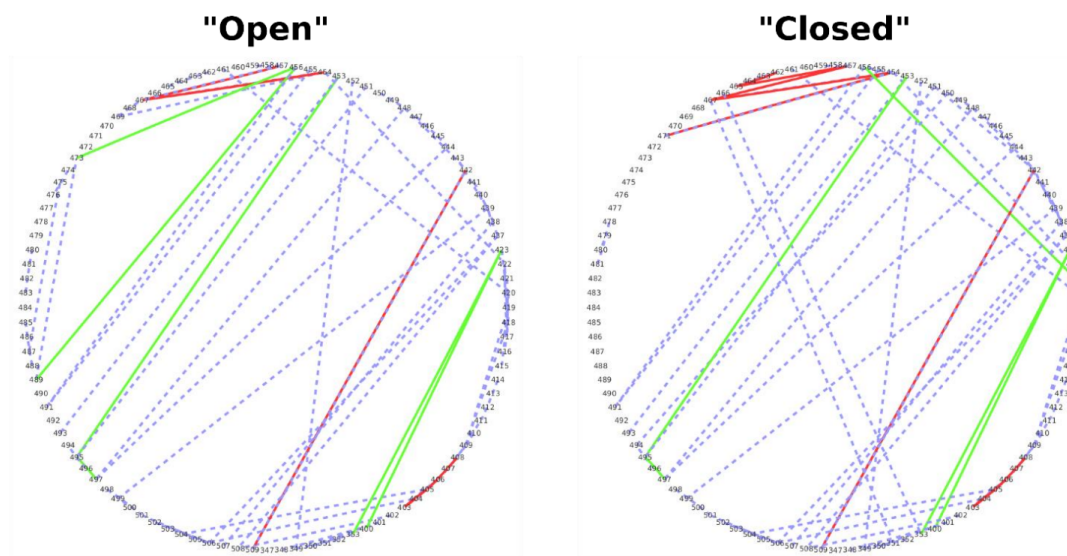


Figure S3.6: Residue interaction networks (RINs) for the “open” and “closed” conformations of beta RBD. RINs determined using RIP-MD for the 5000 lowest energy conformations obtained for the most populated “open” and “closed” basins of wt RBD. Hydrogen bonds, salt bridges and pi-pi interactions are shown in blue, red and green, respectively.

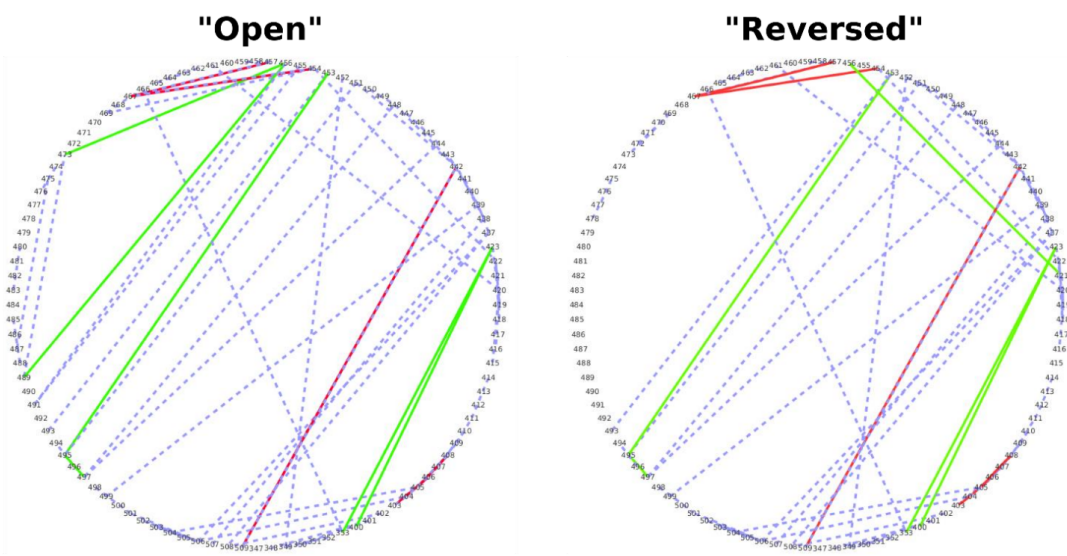


Figure S3.7: Residue interaction networks (RINs) for the “open” and “closed” conformations of delta RBD. RINs determined using RIP-MD for the 5000 lowest energy conformations obtained for the most populated “open” and “closed” basins of wt RBD. Hydrogen bonds, salt bridges and pi-pi interactions are shown in blue, red and green, respectively.

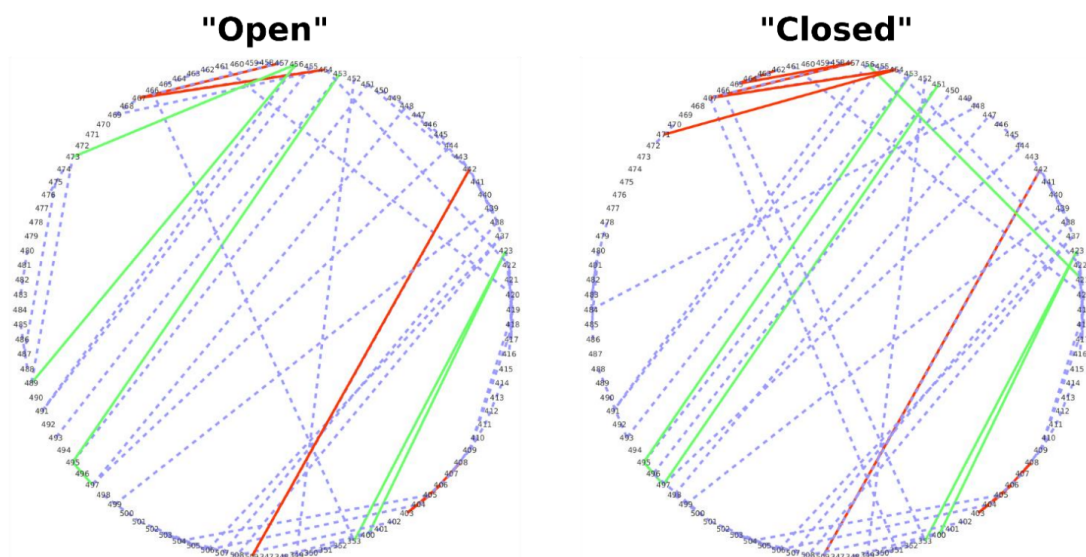


Figure S3.8: Residue interaction networks (RINs) for the “open” and “closed” conformations of omicron RBD. RINs determined using RIP-MD for the 5000 lowest energy conformations obtained for the most populated “open” and “closed” basins of wt RBD. Hydrogen bonds, salt bridges and pi-pi interactions are shown in blue, red and green, respectively.

SARS-CoV-2 antibodies studied						
15033	2-4	2-15	15033-7	2G12	2H2	3C1
4A8	B38	BD-236	BD-368-2	BD-604	BD-629	BD23
C002	C102	C104	C105	C110	C119	C121
C135	C144	C1A-B12	C1A-B3	C1A-C2	C1A-F10	CB6
CV30	DH1047	DH1050.1	EY6A	FC05	H014	LY-CoV555
CC12.3	COVA1-16	COVA2-04	COVA2-39	CT-P59	CV07-250	CV07-270
P17	P2B-2F6	P2C-1A3	P2C-1F11	REGN10933	REGN10987	S2A4
S2E12	S2H13	S2H14	S2M11	S304	S309	STE90-C11
No name (7cjf)						

Table S3.4: Summary of antibodies studied in the S protein-antibody complexes. These antibodies were chosen due to the availability of their structure resolved together with the S protein (or just the RBD region) in the PDB repository [284].

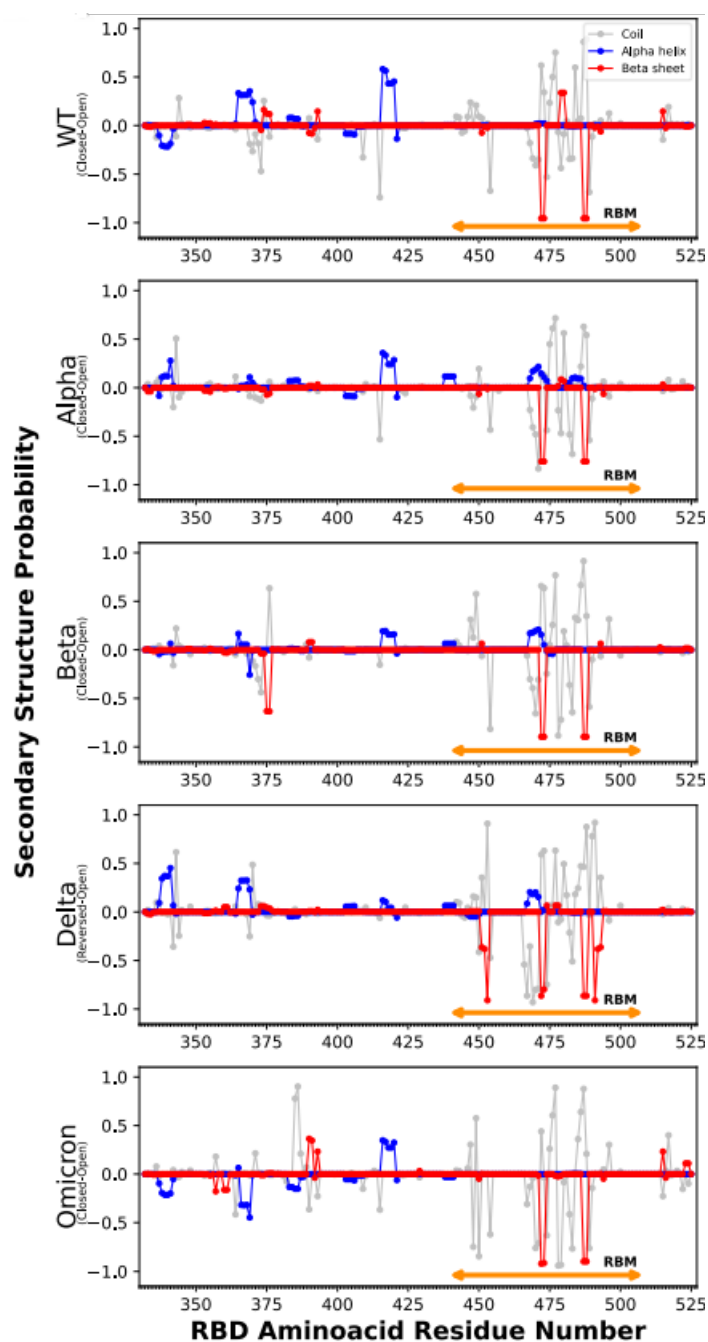


Figure S3.9: Secondary structure difference between closed and open conformations of wt, alpha, beta, omicron and delta SARS-CoV-2 RBD simulated in water. The probability of coil, α -helix and β -sheet secondary structures was obtained using the GROMACS for all conformations (open, closed and reversed) of all four variants. RBM residues are shown with an orange arrow.

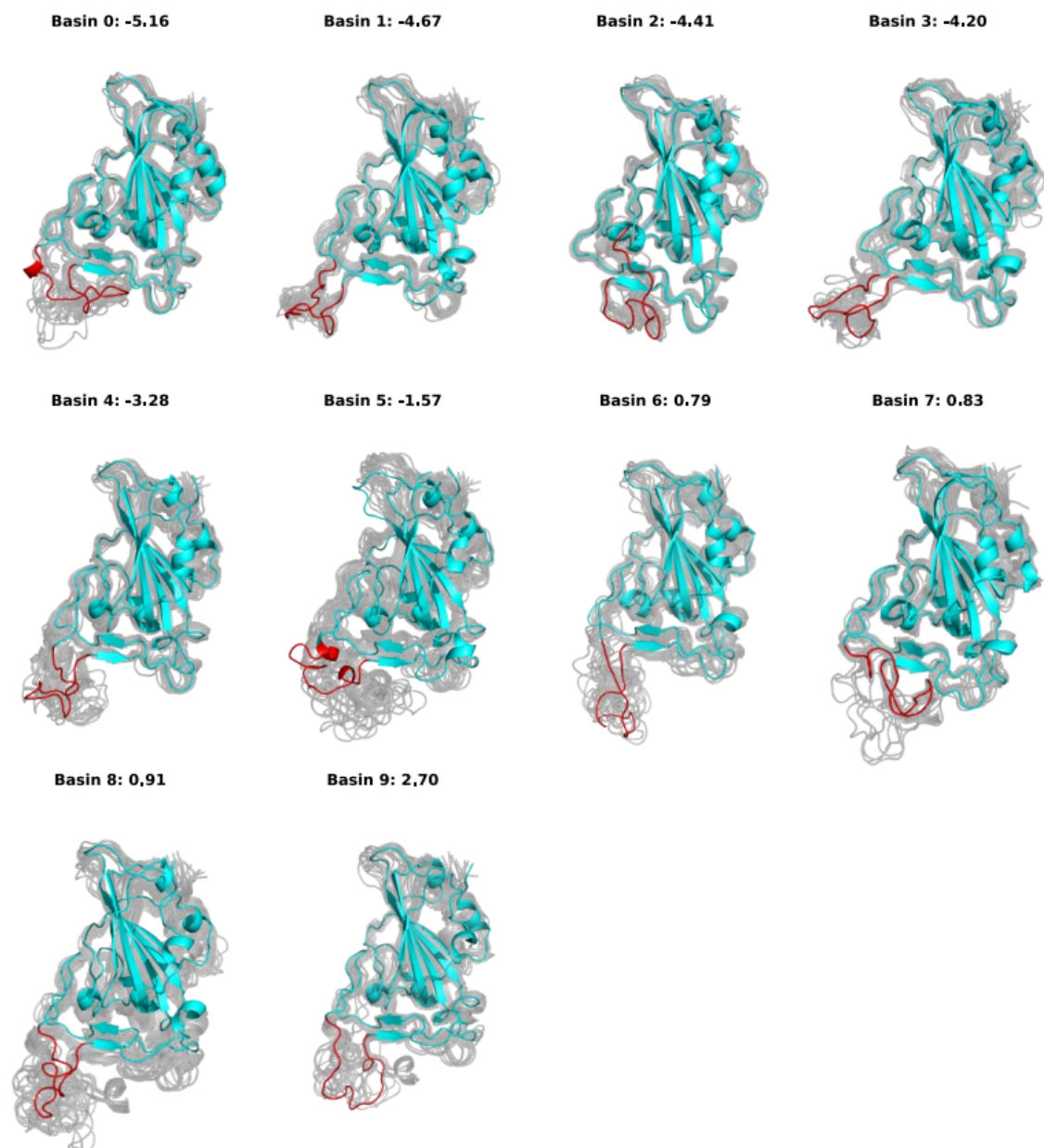


Figure S3.10: Snapshots representative of all wt RBD PCA basins. The structures corresponding to the free energy minima of all conformational basins are represented in blue, with the ridge region highlighted in red, together with structures sampled from the same basin (background, gray colored).

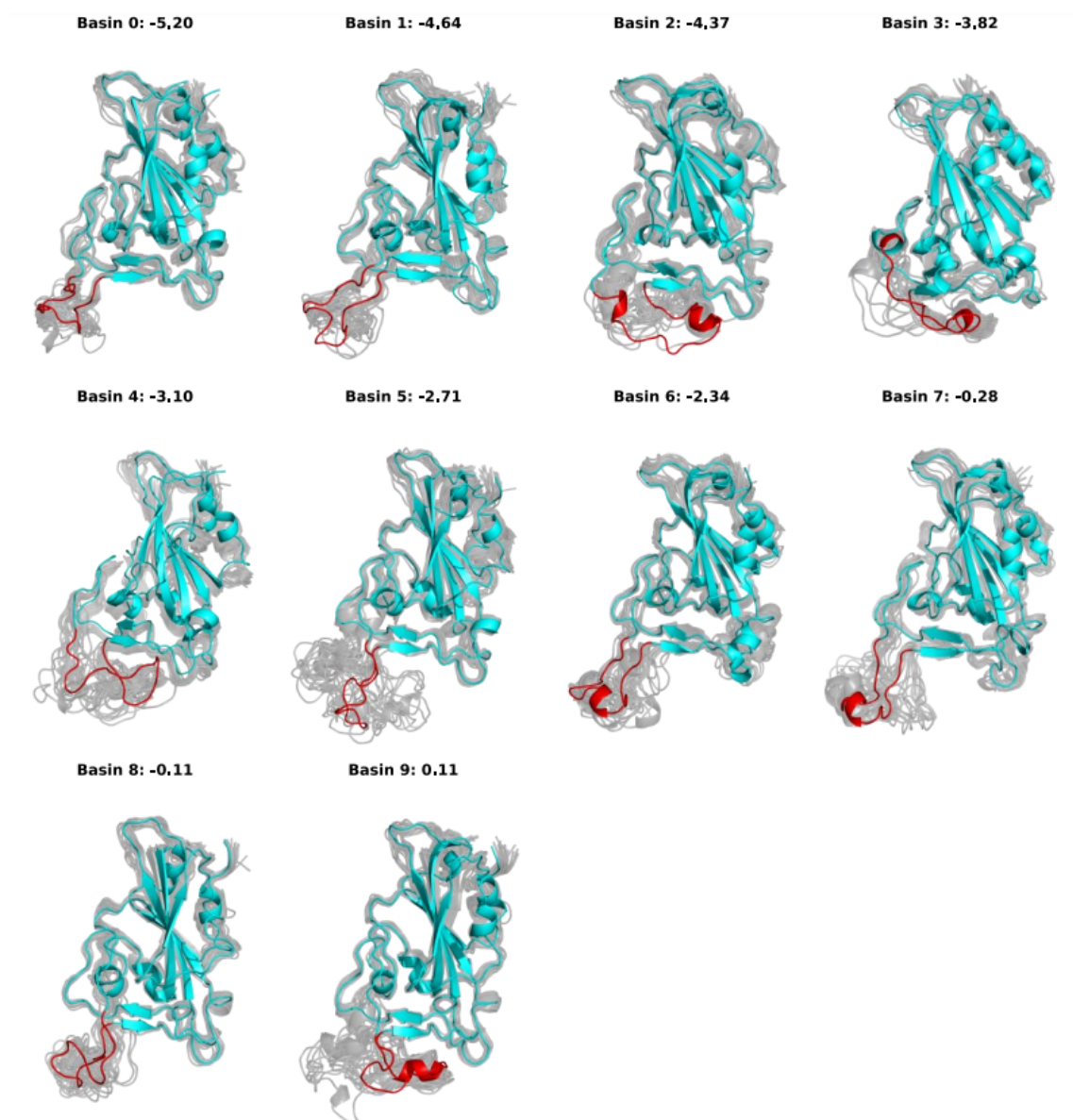


Figure S3.11: Snapshots representative of all alpha RBD PCA basins. The structures corresponding to the free energy minima of all conformational basins are represented in blue, with the ridge region highlighted in red, together with structures sampled from the same basin (background, gray colored).

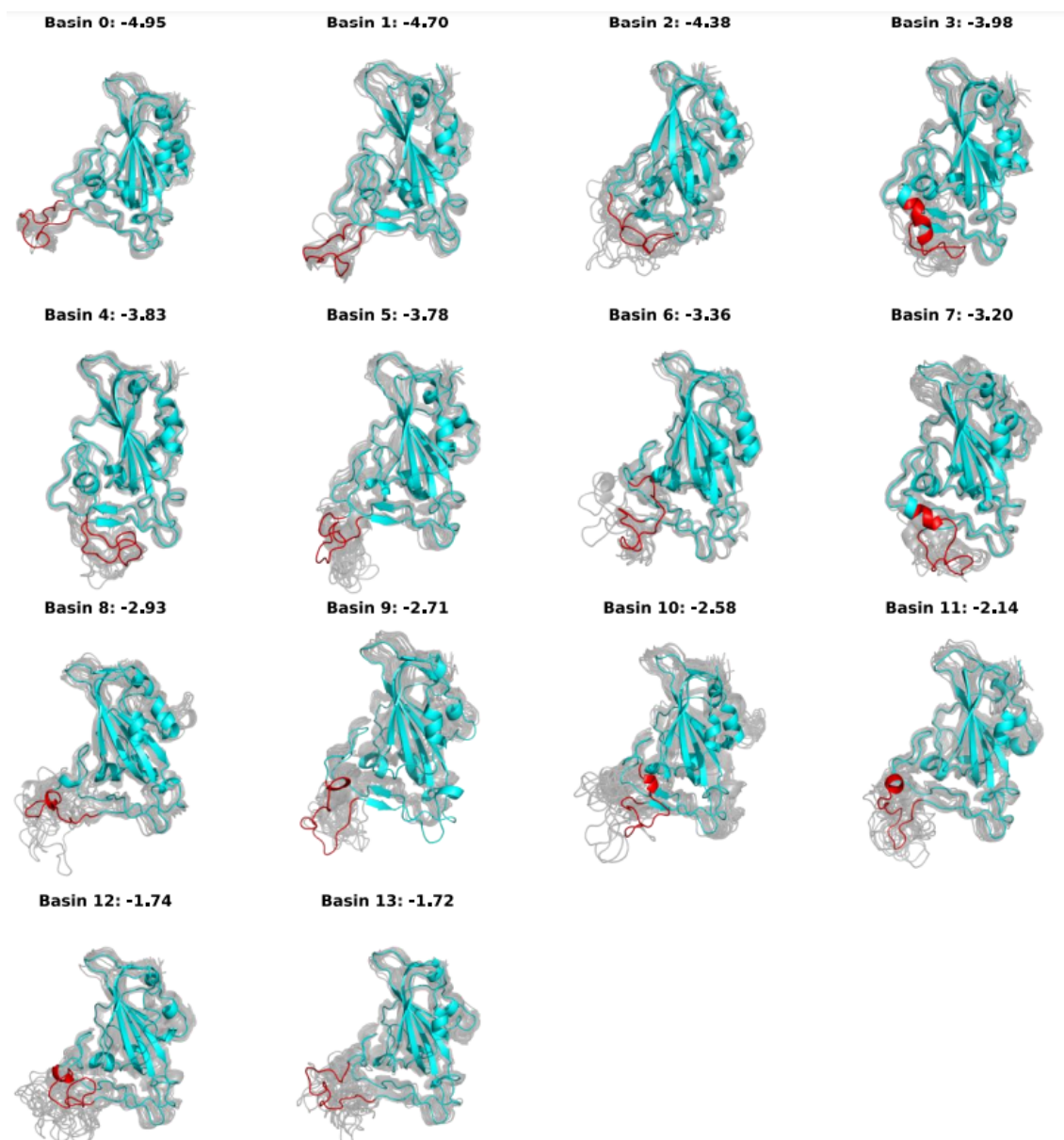


Figure S3.12: Snapshots representative of all beta RBD PCA basins. The structures corresponding to the free energy minima of all conformational basins are represented in blue, with the ridge region highlighted in red, together with structures sampled from the same basin (background, gray colored).

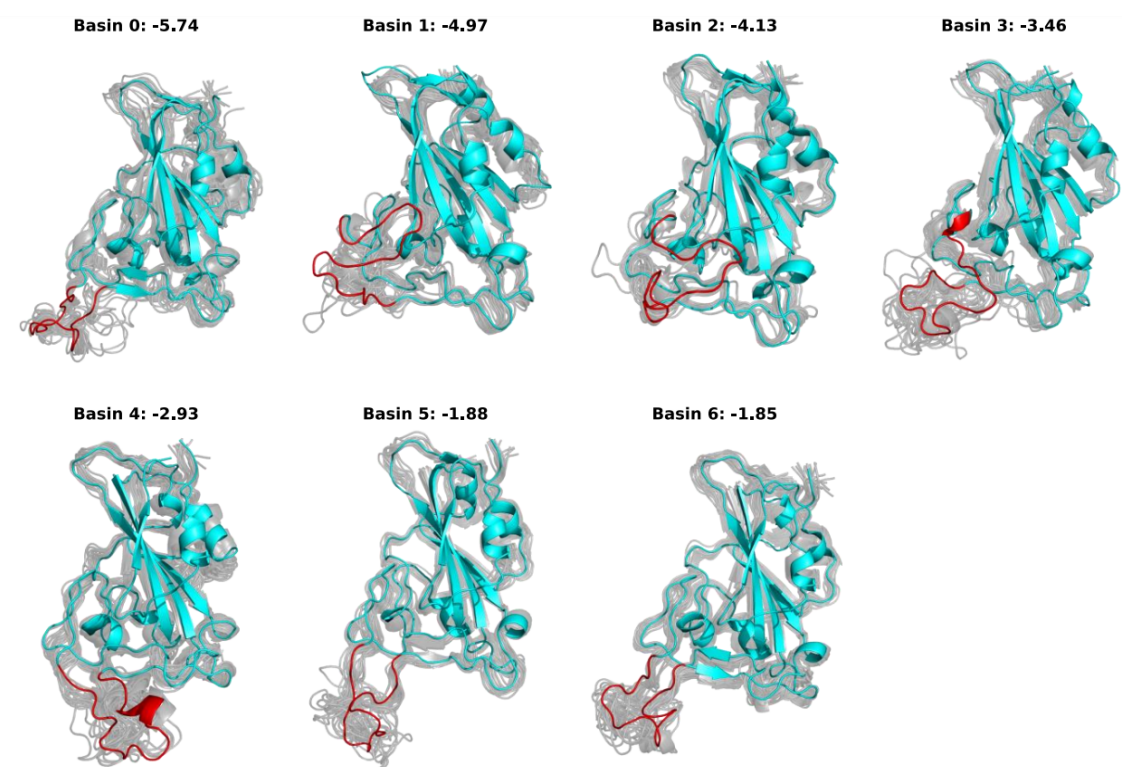


Figure S3.13: Snapshots representative of all delta RBD PCA basins. The structures corresponding to the free energy minima of all conformational basins are represented in blue, with the ridge region highlighted in red, together with structures sampled from the same basin (background, gray colored).

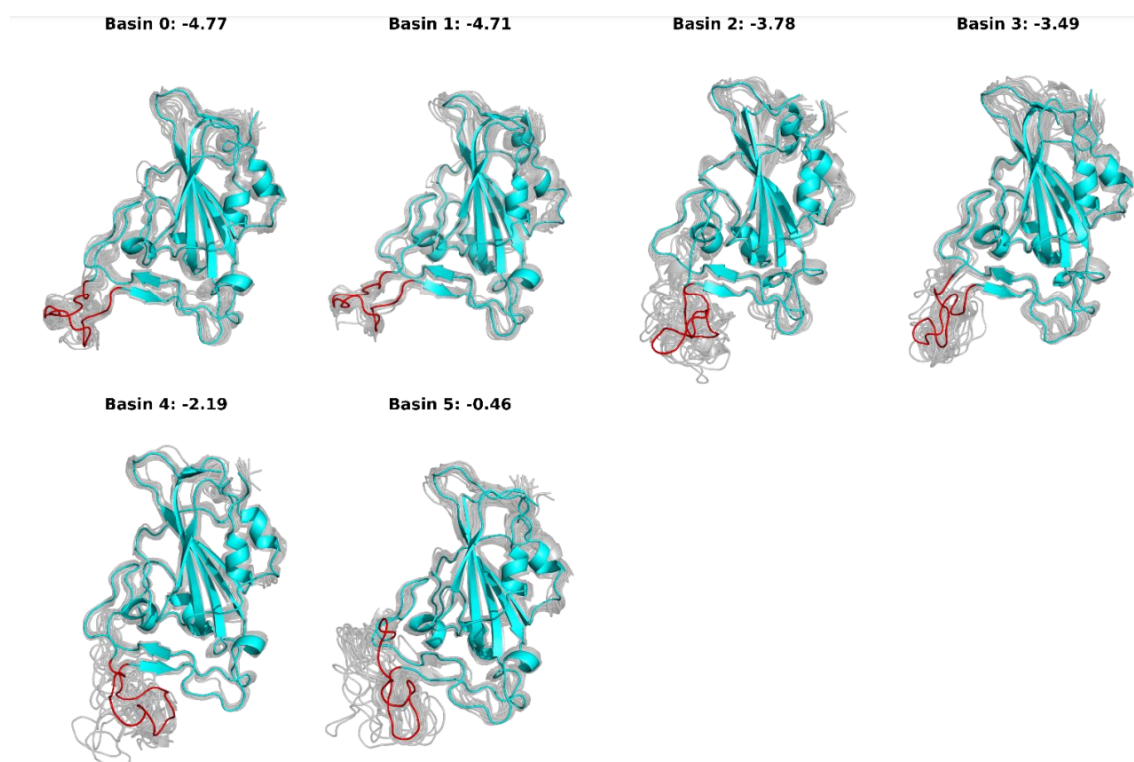


Figure S3.14: Snapshots representative of all omicron RBD PCA basins. The structures corresponding to the free energy minima of all conformational basins are represented in blue, with the ridge region highlighted in red, together with structures sampled from the same basin (background, gray colored).

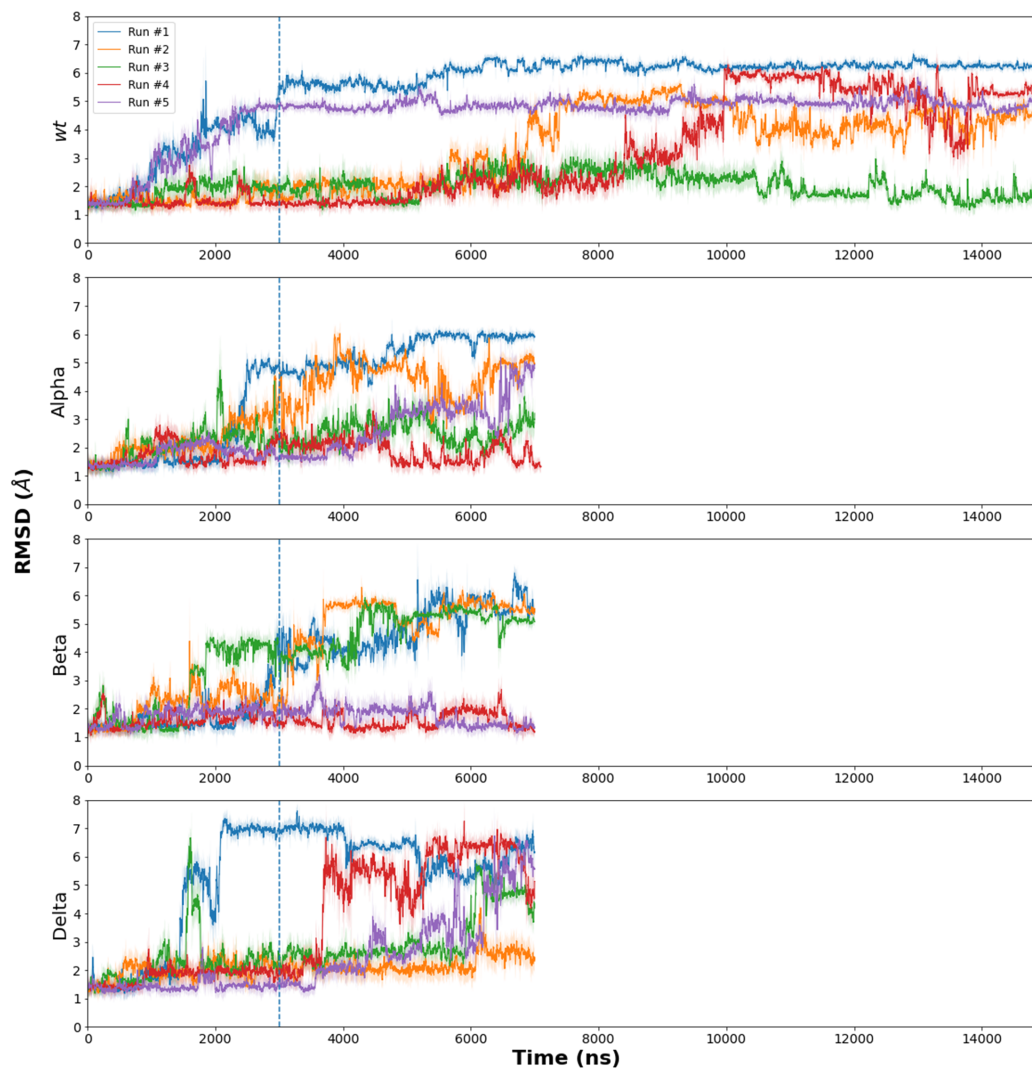


Figure S3.15: RBD $C\alpha$ root-mean-square deviation (RMSD) moving average in solution. Data shown for the five replicas for each variant tested. C were fitted against the RBD X-ray structure from PDB ID: 6M0J. The moving average was calculated using the neighboring 50 frames. The first 3 μ s of simulation were used for equilibration (blue dashed line) and the remaining frames were used for further PCA and RIN analysis.

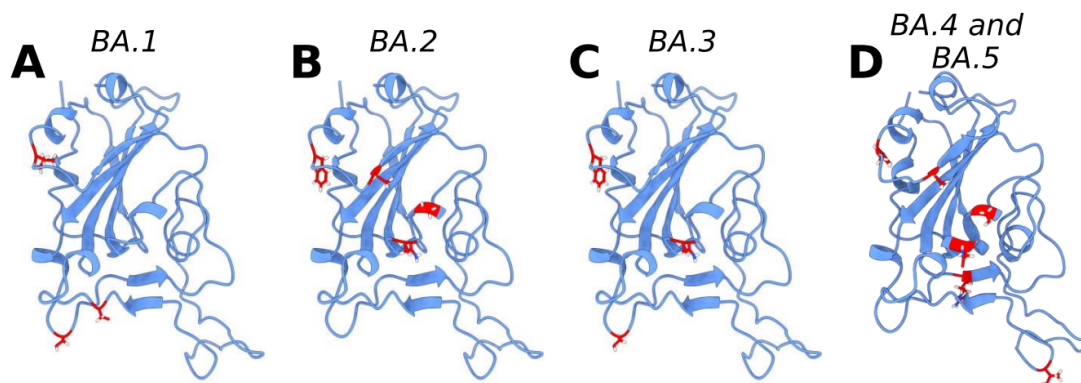


Figure S3.16: Representation of the RBD of the three Omicron lineages and their specific mutations. The RBD regions of all three lineages are shown side-by-side in blue. Specific mutations of lineages (A) BA.1 [S371L, G446S, G496S], (B) BA.2 [S371F, T376A, D405N, R408S], (C) BA.3 [S371F, D405N, G446S] and (D) BA.4 and BA.5 [S371F, T376A, D405N, R408S, L452R, F486V] are shown in sticks and highlighted in red.

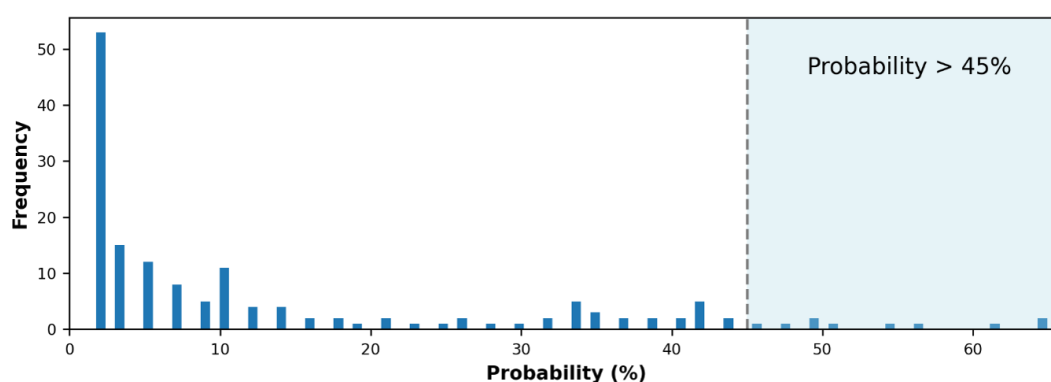


Figure S3.17: S protein-antibody contact frequency histogram. The bar plot shows the distribution of probabilities for the S protein-antibody contacts.

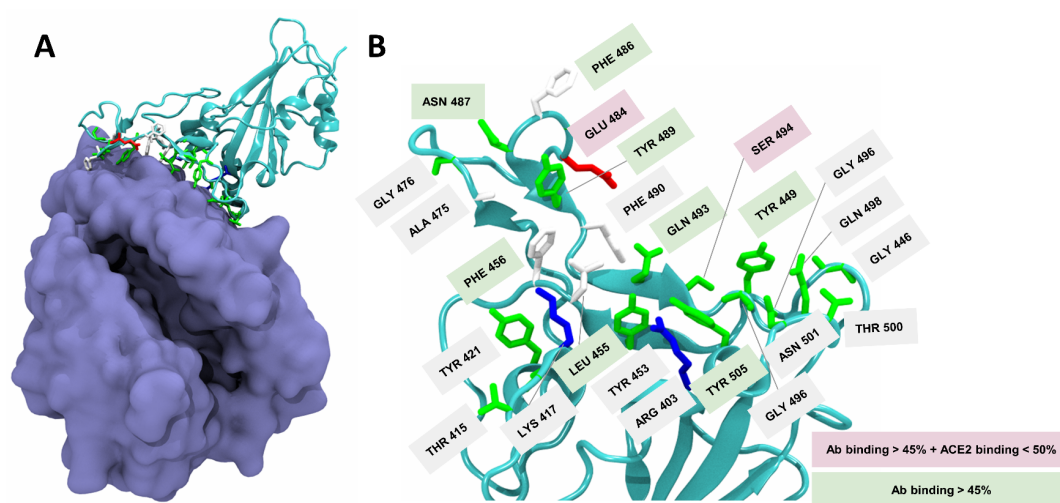


Figure S3.18: Relevant residues for antibody and ACE2 binding. (A) Crystal structure of the SARS-CoV-2 RBD complexed with ACE2. ACE2 is shown as a purple surface and the RBD is shown in a cyan cartoon representation with key antibody interacting residues depicted as sticks. (B) Zoom in to the RBM region of the RBD. Residues relevant for antibody binding (>35% frequency of contact) are depicted as sticks. Of these, the ones with an antibody binding probability higher than 45% have a green label, and those that also have a low frequency of binding to ACE2 (<50%) are labelled in pink.

PARAINFLUENZA FUSION PEPTIDE
PROMOTES MEMBRANE FUSION BY
ASSEMBLING INTO OLIGOMERIC PORELIKE
STRUCTURES

This chapter contains published data from:

Valério, M., Mendonca, D. A., Morais, J., Buga, C. C., Cruz, C. H., Castanho, M. A., Melo, N. M., Soares, M. S. and Lousa, D. (2022). Parainfluenza fusion peptide promotes membrane fusion by assembling into oligomeric porelike structures. *ACS Chemical Biology*, 17(7), 1831-1843.

Author contributions: M.V. and D.A.M. contributed equally to this work. M.V., C.H.C., M.N.M., C.M.S., and D.L. designed the computational systems and analysis. M.V. performed the atomistic and coarse-grain simulations. C.H.C performed metadynamics simulations. D.A.M., J.M., C.C.B., M.A.R.B.C., and A.S.V. designed the experimental assays and D.A.M., J.M., and C.C.B. performed the experiments.

4.1 Abstract

Paramyxoviruses are enveloped viruses harboring a negative-sense RNA genome that must enter the host's cells to replicate. In the case of the parainfluenza virus, the cell entry process starts with the recognition and attachment to target receptors, followed by proteolytic cleavage of the fusion glycoprotein (F) protein, exposing the fusion peptide (FP) region. The FP is responsible for binding to the target membrane, and it is believed to play a crucial role in the fusion process, but the mechanism by which the parainfluenza FP (PIFP) promotes membrane fusion is still unclear. To elucidate this matter, we performed biophysical experimentation of the PIFP in membranes, together with coarse grain (CG) and atomistic (AA) molecular dynamics (MD) simulations. The simulation results led to the pinpointing of the most important PIFP amino acid residues for membrane fusion and show that, at high concentrations, the peptide induces the formation of a water-permeable porelike structure. This structure promotes lipid head intrusion and lipid tail protrusion, which facilitates membrane fusion. Biophysical experimental results validate these findings, showing that, depending on the peptide/lipid ratio, the PIFP can promote fusion and/or membrane leakage. Our work furthers the understanding of the PIFP-induced membrane fusion process, which might help foster development in the field of viral entry inhibition.

4.2 Introduction

Parainfluenza viruses (PIVs) contribute significantly to the annual global disease burden in humans [292, 293]. It is estimated that annually, in the USA alone, parainfluenza infections are responsible for more than 325,000 hospitalizations of children under 5 years old [294]. Although it is difficult to estimate the exact number of infections caused by the PIVs worldwide, these viruses have the potential to infect millions of individuals, leading to an undetermined number of deaths, especially in areas with inadequate health-care resources. To make matters worse, there are no specific treatments for infections caused by PIVs [292, 293]. PIVs belong to the Paramyxoviridae family, consisting of enveloped viruses harboring a negative-sense RNA genome that, like all enveloped viruses, depend on the fusion of the viral envelope with the host cell membrane to infect the host. In the PIVs, the machinery responsible for viral fusion is composed of the receptor-binding glycoprotein hemagglutinin-neuraminidase (HN) and the fusion glycoprotein (F), that, together, form the viral fusion complex [177]. Opposite to what occurs in the cell entry of the influenza virus [295], PIVs are not endocytosed by the target cells and instead fuses its membrane with the host membrane at the cell surface (and neutral pH)[296]. The membrane fusion process starts with the recognition and binding of the receptor-binding glycoprotein to receptors at the host cell surface. This binding triggers the F glycoprotein to undergo major conformational rearrangements and the proteolytic cleavage of the F protein, exposing the fusion peptide (FP) region (Figure 4.1A; the PIFPs are highlighted in

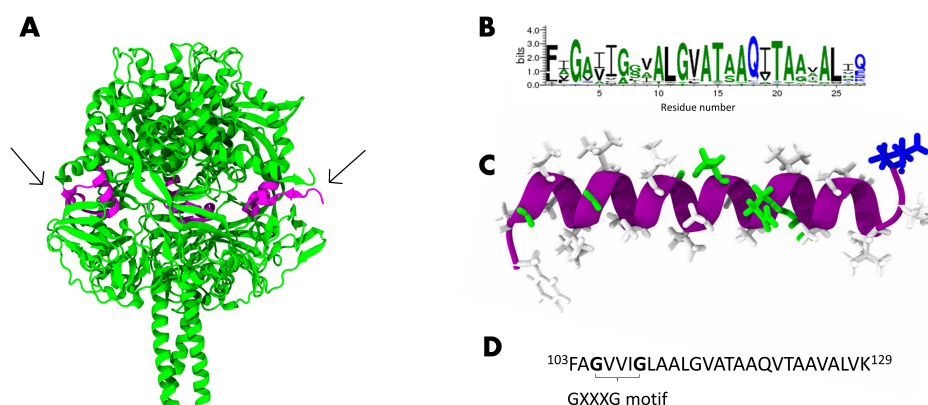


Figure 4.1: Parainfluenza fusion peptide features. (A) Cartoon representation of the crystal structure of the cleaved prefusion F protein trimer, 4GIP, viewed from the side, with the FP highlighted colored in pink. Arrows indicate the protease cleavage sites. (B) Sequence conservation of glycoprotein F among members of the Paramyxoviridae family. Color is used to show hydrophobicity of the residues (blue: hydrophilic, green: neutral, and black: hydrophobic). (C) Molecular image of the PIFP with the backbone in purple and the residues colored according to their biophysical properties: nonpolar (white), basic (blue), and polar residues (green). (D) PIFP amino acid sequence with the GXXXG motif highlighted.

pink), promoting viral entry [178, 296]. After being exposed, this region is inserted into the target membrane [179] and the F protein refolds to form a six-helix bundle (6HB), placing the fusion peptides and the transmembrane domains in proximity, facilitating membrane fusion [297, 298].

Regions with functional requirements for membrane fusion and viral entry, such as the PIFP, are, in general, highly conserved among viral families and contain several hydrophobic residues (Figure 4.1B,C, respectively). Another common feature among these sequences is their location within the fusion protein. In class I fusion proteins, they are usually located at their N-terminal end or close to it [166]. All of these traits are also found in the PIFP, together with the presence of GXXXG motifs (Figure 4.1D). The GXXXG motifs are the hallmark of many high-affinity associations between transmembrane helices [299, 300]. The presence of these motifs in the PIFP alludes to the possibility that inter-PIFP peptide interactions are common, a hypothesis that was investigated in this work. Interestingly, mutations to these glycine residues in the FP from the paramyxovirus simian parainfluenza virus 5 (SV5) resulted in mutant F proteins with hyperactive fusion phenotypes but not in the human parainfluenza virus 3 (HPIV3) [301]. Most studies regarding the mechanisms of the viral FPs and their interaction with model membranes have focused on influenza virus and human immunodeficiency virus (HIV). Hypotheses

for the mechanism by which the FP from these viruses induces membrane fusion suggest different modes of action, including altering the membrane curvature[302–305], increasing [306–308] or decreasing lipid order [160, 164], and inducing pore formation and/or its stabilization [305, 309, 310]. Additionally, a mechanism proposed on the basis of molecular dynamics (MD) simulations has shown that lipid tail protrusion (i.e., a lipid acyl chain that extends to and beyond the corresponding phosphate group) is a determinant step in membrane fusion [158, 309]. The occurrence of lipid tail protrusion has been observed in several simulation studies of the influenza FP in membrane bilayers, which indicates that the peptide increases the probability of lipid protrusion events [159, 160, 164, 172]. For the influenza virus, it has also been shown that the peptide interacts with the lipid headgroups, mainly through the N-terminal group, which induces these headgroups to penetrate deeper into the membrane (headgroup intrusion) [157, 159, 160, 172]. Recent experimental data has elucidated the orientation adopted by the PIFP in a membrane bilayer, a determinant for the peptide-induced perturbation of lipid bilayers [311]. NMR structural analysis in 1-palmitoyl-2-oleoyl-glycerol-3-phosphocholine (POPC)/1-palmitoyl-2-oleoyl-sn-glycerol-3-phosphoglycerol (POPG), 1,2-dioleoyl-sn-glycerol-3-phosphocholine (DOPC)/1,2-dioleoyl-sn-glycerol-3-phosphoglycerol (DOPG), and 1,2-dioleoyl-sn-glycerol-3-phosphoethanolamine (DOPE) membranes showed that the peptide is fully α -helical when inside a POPC/POPG membrane, adopting a mixed strand/helix conformation in the DOPC/DOPG membrane, and is primarily a β -strand in the DOPE membranes [149]. Regarding the peptide insertion in the membrane, 1H spin diffusion experiments in liquid crystal and gel phases showed that in a POPC/POPG membrane the N-terminal half of the peptide is more exposed to the membrane surface, whereas in a DOPC/DOPG membrane, the C-terminal residues are more exposed to water than the N-terminal residues [149]. A crucial question remains unanswered: What is the mechanism by which the PIFP promotes membrane fusion? Data on the interaction between PIFPs in phospholipid micelles has shed some light on this matter. When inside a membrane, there seems to be a cooperative and specific assembly of the PIFP into hexamers [168]. Short molecular dynamics (MD) simulations of the hexameric structure also showed that there is penetration of water into its core from the viral side of the membrane [168]. In this work, we conducted a comprehensive analysis of the effect of peptide concentration on the PIFP properties by combining experimental biophysical data with results from coarse grain (CG) and all-atom (AA) MD simulations. Förster resonance energy transfer (FRET)-based assays revealed that the percentage of peptide-induced lipid mixing and leakage is higher at high peptide/lipid ratio conditions. Circular dichroism (CD) results showed that there is also an effect of concentration on the PIFP secondary structure. The simulation studies elucidated the peptide–peptide interactions that stabilize PIFP inside a mixed bilayer to allow the formation of a porelike structure. Finally, we characterized the peptide–membrane interactions and pinpointed the amino acid residues that are crucial for the peptide’s ability to interact with and perturb the host membrane. This work provides a valuable contribution to understanding the membrane fusion process induced

by the PIFP, which can be very useful in the future development of antiviral therapies targeting this virus.

4.3 Materials and Methods

4.3.1 Instrumentation

For steady-state fluorescence measurements, a FLS920 series Edinburgh Instruments spectrofluorometer (Livingston, UK), and Varian Cary Eclipse spectrofluorometer (Crawley, UK) was used. CD spectroscopy was performed in a J-815 spectrophotometer from Jasco (Oklahoma City, OK, USA). Dynamic light scattering (DLS) measurements were performed in a Malvern Instruments ZetasizerNano ZS (Worcestershire, UK). Surface plasmon resonance (SPR) experiments were performed in a GEHealthcare Biacore X100 (Little Chalfont, UK). Extrusion of liposomal vesicles were carried out in a LiposoFast-Basic plus Stabilizer setup from Avestin (Mannheim, Germany) with Hamilton (Bonaduz, Switzerland) syringe through a 100 and 50 nm pore Nucleopore polycarbonate membrane (Whatman/GE Healthcare, UK).

4.3.2 Lipid Vesicles Preparation

Large unilamellar vesicles (LUVs) and small unilamellar vesicles (SUVs) of POPC and POPC:POPG (4:1) were used as membrane model systems. The lipids were solubilized in chloroform in a round-bottom flask and the organic solvent was evaporated under a gentle nitrogen flow, forming a thin lipidic film, which was placed under vacuum overnight. The lipidic film was rehydrated with 10 mM HEPES, 150 mM NaCl pH, 7.4 buffer – sample buffer - and ten freeze/thaw cycles were performed resulting in a suspension of multilamellar vesicles. LUV and SUV were obtained by extrusion of the multilamellar vesicles through 100-nm and 50-nm pore size polycarbonate filters, respectively.

4.3.3 PIFP-induced Vesicles Fusion Assay

Fusion/Lipid mixing of vesicles induced by PIFP was assessed by FRET methodology. Briefly, LUV pre-doped with both 1% N-NBD-PE (donor) and 1% N-Rh-PE (acceptor) and unlabeled vesicles were mixed in a 1:3 ratio in sample buffer. Two conditions were tested: Low Peptide/Lipid ratio - 18 μ M PIFP + 3.5 mM mixed vesicles; and High Peptide/Lipid ratio - 18 μ M PIFP + 0.77 mM mixed vesicles. Fluorescence emission spectra were recorded between 500 and 650 nm, with the donor excitation wavelength (λ_{exc}) at 470 nm. For each condition, fusion efficiency (R) was calculated:

$$R = \frac{I_F 530 \text{ nm}}{I_F 588 \text{ nm}} \quad (4.1)$$

In which $I_F 530 \text{ nm}$ is the fluorescence emission intensity collected at 530 nm (donor emission maximum wavelength) and $I_F 588 \text{ nm}$ is the fluorescence emission intensity

collected at 588 nm (acceptor emission maximum wavelength). Fusion (%) was then calculated using the following equation:

$$Fusion (\%) = \frac{R_S - R_0}{R_{max} - R_0} \times 100 \quad (4.2)$$

In which R_S is the R of the sample, R_0 is the R for the negative control condition (mixed vesicles in the absence of peptide - FRET maximum) and R_{max} is the R for the positive control (mixed vesicles in the presence of 1% (v/v) Triton X-100 – absence of FRET).

4.3.4 Vesicles Leakage

Content leakage from POPC or POPC:POPG (4:1) vesicles was quantified by 5,6-carboxy - fluorescein (5,6-CF) fluorescence dequenching using LUVs loaded with 50 mM 5,6-CF as described previously[312]. Briefly, 2-fold dilutions of peptides (starting at 30 μ M) were incubated for 15 min with LUVs at 0.77 mM in sample buffer. Fluorescence emission intensity ($\lambda_{exc} = 493$ nm and $\lambda_{emi} = 520$ nm) was measured and the leakage percentage was calculated as described previously[312].

4.3.5 Circular Dichroism

PIFP and the vesicles were prepared in 10 mM phosphate, 150 mM NaF, pH 7.4 buffer. CD spectra of the peptide at 30 μ M, in the absence or presence of LUV, were acquired between 190 and 260 nm at 25 degrees celsius. Each final spectrum corresponds to the averaged accumulation of 10 scans. Spectra were plotted using mean molar residue ellipticity ($[\theta]$) values:

$$[\theta] = \frac{\theta}{aa \times l \times c} \quad (4.3)$$

in which θ is the ellipticity, aa the number of amino acid residues of the peptide sequence, l is the cuvette path length, and c is the molar concentration. $[\theta]$ values were corrected for the buffer background noise.

4.3.6 Dynamic Light Scattering

DLS measurements were performed to quantify lipid vesicles aggregation induced by PIFP. Titration of PIFP, with increasing peptide concentrations from 0.9375 up to 30 μ M, was performed in a 0.77 mM POPC:POPG (4:1) LUV sample in sample buffer. After each individual addition of peptide, the sample was incubated for 5 min at 25 degrees Celsius before measurement. Measurements comprised ten individual runs, each corresponding to an averaged autocorrelation curve obtained from at least ten replicate sample scans. Diffusion coefficient (D) values were calculated from autocorrelation curves using a CONTIN-based method[313]. D values were used to determine the particles Z-average hydrodynamic diameter and derived count rate through the Stokes-Einstein-Sutherland equation[314].

4.3.7 Surface Plasmon Resonance

To investigate the affinity of the PIFP peptide toward lipid membranes, SPR experiments were performed. The flow system was primed three-times before initiating an experiment in sample buffer. L1 sensor chips were used in all experiments. Each sensor chip was rinsed with three injections of 20 mM CHAPS before starting an assay. To prepare the lipid surface, 1 mM POPC or POPC:POPG (4:1) SUV sample was injected over the sensor chip at a 2 $\mu\text{L}/\text{min}$ flow speed for 2400 s. Loosely bound vesicles were removed with a 36 s injection of 10 mM NaOH at 50 $\mu\text{L}/\text{min}$ flow. PIFP peptide samples in concentration ranging from 1 to 50 μM were injected over the deposited SUV surface at 5 $\mu\text{L}/\text{min}$ flow speed during 200 s (association time). Each sample was allowed 800 s dissociation time after injection stopped. After each run, the L1 sensor chip was regenerated with sequential injections of 20 mM CHAPS (5 $\mu\text{L}/\text{min}$ for 60 s), 0.5% (v/v) SDS (5 $\mu\text{L}/\text{min}$ for 60 s), 10 mM NaOH containing 20% (v/v) methanol (50 $\mu\text{L}/\text{min}$ for 36 s) and 10 mM NaOH (50 $\mu\text{L}/\text{min}$ for 36 s). Response values were monitored to ensure effective surface regeneration. Response values were corrected for the peptide molecular weight. The partition constant (K_p) was then calculated using the following equation:

$$\frac{RU_S}{RU_L} = \frac{\gamma_L K_p \frac{M_S}{M_L} [S]_W}{1 + \sigma \gamma_L K_p [S]_W} \quad (4.4)$$

In which RU_S is the total sample response, RU_L is the total lipid deposition response, γ_L the lipid molar volume, M_S is peptide molecular weight, M_L is the lipid vesicles molecular weight, $[S]_W$ is the sample concentration and σ is the lipid to solute ratio at the saturation point.

PIFP unbinding kinetics fitted the following equation:

$$S_L(t) = \alpha e^{-k_{off,\alpha} t} + \beta e^{-k_{off,\beta} t} + S_{L,r} \quad (4.5)$$

In which $S_L(t)$ is the dissociation fractional response (response after 200 s) through time, α and β are the fractional contributions of each population to the measured response of each solute population; $k_{off,\alpha}$ and $k_{off,\beta}$ are the dissociation constants of each solute population and $S_{L,r}$ is the fraction of retained solute. The weight-averaged dissociation constant ($\langle k_{off} \rangle$) was then calculated by the following equation:

$$\langle k_{off} \rangle = \frac{\alpha k_{off,\alpha} + \beta k_{off,\beta}}{\alpha + \beta} \quad (4.6)$$

4.3.8 Coarse Grain MD Simulations of the PIFP in a POPC/POPG membrane

Solid-state NMR studies of PIFP revealed that the peptide adopts a fully beta-helical conformation inside POPC/POPG membranes[149]. Based on these data, a fully α -helical PIFP was created using the *remodeling* tool from Rosetta[315] and converted to a Martini

coarse grained topology[201, 316] with the martinize 2.0 program[187]. This program was also used to generate the topology and structure files based on the atomistic pdb file.

CG simulations of the PIFP at different peptide/lipid ratios were performed in a membrane system containing POPC and POPG lipids at an 80:20 mol ratio. The peptides were inserted vertically crossing both leaflets, in trimer, hexamer and nonameric assemblies (Figures S4.2A, B, C and D respectively), and the multimers were constructed by multiplying the structure of the 27-residue-long helical FP. The N termini of the monomers were protonated and the C termini's amidated to reproduce the state of peptides in the context of the F protein. The axes of the α -helices in each dimer were oriented parallel to each other and separated by at least 3 nm.

All simulations were performed with the GROMACS 2020.3 package[220], using the Martini 3.0 force field[187] and lipid topologies released with it. The system was built using the INSANE script, in a 10 x 15 x 14 nm box. The total charge of the systems was first neutralized with either Na^+ or Cl^- ions, and then an additional 140 mM NaCl was added.

In order to maintain the peptides inside the membrane bilayer, flat-bottomed harmonic well potential with a force constant of $500 \text{ kJ nm}^{-2} \text{mol}^{-1}$ and a flat radius of 0.5 and 0.2 nm was applied to the peptide's terminal residues and to the lipid heads, respectively. Preliminary simulations of the PIFPs inside the membrane showed that, when inserted in the membrane without any restraints, the peptides quickly moved to the outer leaflet. Since previous NMR studies showed that PIFPs should remain in the membrane[149], and to better replicate the biological behavior, a potential was employed keeping the peptide's termini at their initial (transmembranar) position. To minimize this bias and allow some degree of adjustment, a $500 \text{ kJ nm}^{-2} \text{mol}^{-1}$ harmonic flat-bottomed potential was employed in the z direction, centered at the termini's initial position and with a flat (zero) region for 0.5 nm above and below this reference point. To prevent a membrane displacement in z that would result in the expulsion of the peptides from its core, a potential of the same type was employed on the lipid phosphates, but with a flat region of 0.25 nm.

The systems were energy-minimized using the steepest-descent method for 5000 steps. In this step the peptide's backbone atoms were position-restrained using a force constant of $500 \text{ kJ mol}^{-1} \text{nm}^{-2}$. Before performing production runs, an equilibration step was carried out for 20 ns. The simulations were run in the isothermal-isobaric (NPT) ensemble, coupling the system to a temperature and pressure bath using the V-rescale and Parrinello-Rahman[224] schemes, respectively (the Berendsen barostat was employed during equilibration, for its robustness). The target temperature was 310 K and semi-isotropic pressure coupling was set to a reference pressure of 1 atm using a compressibility of $3 \times 10^{-4} \text{ bar}^{-1}$.

The target temperature was set to 310 K and semi-isotropic pressure coupling also set to 1 atm using a compressibility of $3 \times 10^{-4} \text{ bar}^{-1}$. Reaction-field electrostatics were employed, and Lennard-Jones and Coulombic interactions were both cut-off at 1.1 nm[317]. Each system was run for 20 μs .

4.3.9 Atomistic MD Simulations of the PIFP in a POPC/POPG

MD simulations of the fusion peptide (FP) were performed using the last frame of the CG simulations as the starting structure. The four systems (with one, three, six and nine copies of the PIFP) were converted from coarse-grain to atomistic detail using the backward.py script [318]. All atomistic simulations were performed with the GROMACS package, version 2020.3[220], using the CHARMM36m force field[186], with the TIP3P water model. The minimization and equilibration protocols used default parameters generated by CHARMM-GUI for GROMACS [319–321]. In short, after 5000 steps of minimization using the steepest-descent algorithm, the system temperature was maintained at 310 K by coupling to a Nosé-Hoover thermostat[322] with a 1 ps coupling constant. The system was then coupled semi-isotropically to a Parinello-Rahman barostat[224] and further equilibrated for 125 ps using 1 ps timestep, and 500 ps with a 1 ps timestep. During the equilibration, atomic positions restraints on the protein heavy atoms were gradually relaxed in six steps from 4000 to 2000 kJ nm⁻²mol⁻¹ and from 500 to 200 kJ nm⁻²mol⁻¹ the protein backbone and side-chains, respectively. The lipid's phosphate atoms were also gradually relaxed from 1000 to 40 kJ nm⁻²mol⁻¹. After the sixth step, the simulations were performed without any restraints. Three replicates were simulated for each system with all simulations being run for 1 μ s.

4.3.10 Metadynamics Simulations in aqueous solution

Metadynamics simulations[200] were performed to analyze the conformational properties of the PIFP in aqueous solution. We started by performing standard MD simulations using an extended conformation of the peptide (with the same protonation state as mentioned above) that was placed in a water box, considering a minimum distance of 2 nm between the peptide and the box walls. The total charge of the system was first neutralized with Cl⁻ and then an additional 140 mM was added. This system was simulated for 1 μ s with the GROMACS package, version 2020.3[220], using the CHARMM36m force field[186] and the TIP3P model. The simulation parameters were identical to the ones used in the atomistic simulations performed in the membrane. The final structure of the unbiased simulation was used to initialize the metadynamics simulation, which was performed with the PLUMED plugin version 2.6[323, 324] for GROMACS 2018.8, using the same simulation parameters. Metadynamics simulations were performed in a 2D collective variable (CV) space defined by ALPHARMSD (CV1) and RADIUS of GYRATION (CV2). Gaussian potentials with a height of 0.1 kJ/mol were added at every 10 ps. The Gaussian widths were 0.1 and 0.05 for CV1 and CV2, respectively. To avoid systematic errors at the boundaries, lower and upper limits were defined, beyond which the bias potential was set to 0. The lower limits for CV1 and CV2 were set at 0.5 and 0.27, respectively, and the upper limits for CV1 and CV2 were set at 21 and 0.34, respectively. The metadynamics simulation was extended for 400 ns, until reasonable convergence was achieved in the free energy profiles.

4.4 Results

4.4.1 PIFP Induces Fusion and Perturbation on Negatively Charged Lipid Vesicles

PIFP conformational properties and effects on lipid membranes are still unclear. To fill this gap, we synthesized the PIFP that corresponds to residues 103–129 of the PIV5 F protein bound to an 8-amino-3,6-dioxaoctanoic acid conjugated to a polylysine tail (FAGVVIGLAALGVATAAQVTAVALVK-AEEAc-KKKK) and characterized its structure and interaction with lipid membranes using an array of biophysical methodologies. Given its potential fusogenic properties, we first evaluated the PIFP ability to induce fusion and perturbation on lipid vesicles. In this study, we wish to understand the role played by lipid charge in the PIFP ability to induce membrane fusion, and for that reason, we compared POPC vesicles with POPC:POPG (4:1) vesicles. Fusion was followed by PIFP-induced FRET efficiency variations in a mixture containing NBD-RhB-doped vesicles (donor and acceptor, respectively) and nondoped vesicles [325], while the membrane perturbation was followed by the PIFP-induced leakage of 5,6-carboxyfluorescein (5,6-CF)-loaded vesicles. Despite the inability to induce fusion or leakage from zwitterionic POPC vesicles, PIFP induces fusion and leakage on anionic POPC:POPG (4:1) vesicles (Figure 4.2A,B). Besides its hydrophobic nature, PIFP is also cationic, with a net charge of +5 at pH 7.4, which contributes to the membrane charge-action dependence observed, not unique within fusion peptides [326]. Moreover, PIFP induces 9.83% of lipid fusion at the low peptide/lipid ratio condition (membrane nonsaturated condition) and 20.3% of lipid fusion at the high peptide/lipid ratio condition (membrane saturation condition), an approximately 2-fold increase, suggesting that the fusion efficiency is also influenced by the local concentration of the peptide on the membrane. The same pattern is observed for the membrane perturbation phenomena, on which 5,6-CF leakage increases in a PIFP-concentration-dependent fashion. These results may indicate the presence of peptide–peptide interactions, such as self-association/clustering or conformational changes, which eventually lead to a pore formation, that may play a major role in the fusion process of PIFP. To answer this question, PIFP–lipid interactions were thoroughly characterized *in vitro*.

4.4.2 PIFP–Lipid Unbinding Follows a Second-Order Kinetics

After demonstrating that PIFP is fusogenic, it was of utmost importance to unravel its fusion mechanisms. As the fusion process is guided by the interactions between the peptide and the lipid molecules present in the membrane [327], we first characterized the lipid binding of PIFP. Anionic POPC:POPG (4:1) vesicles were used for peptide–membrane interaction characterization assays since PIFP is only capable of inducing the fusion of vesicles with this composition. Peptide–lipid binding was evaluated through surface plasmon resonance (SPR). This technique allows real-time detection of bound molecules

CHAPTER 4. PARAINFLUENZA FUSION PEPTIDE PROMOTES MEMBRANE FUSION BY ASSEMBLING INTO OLIGOMERIC PORELIKE STRUCTURES

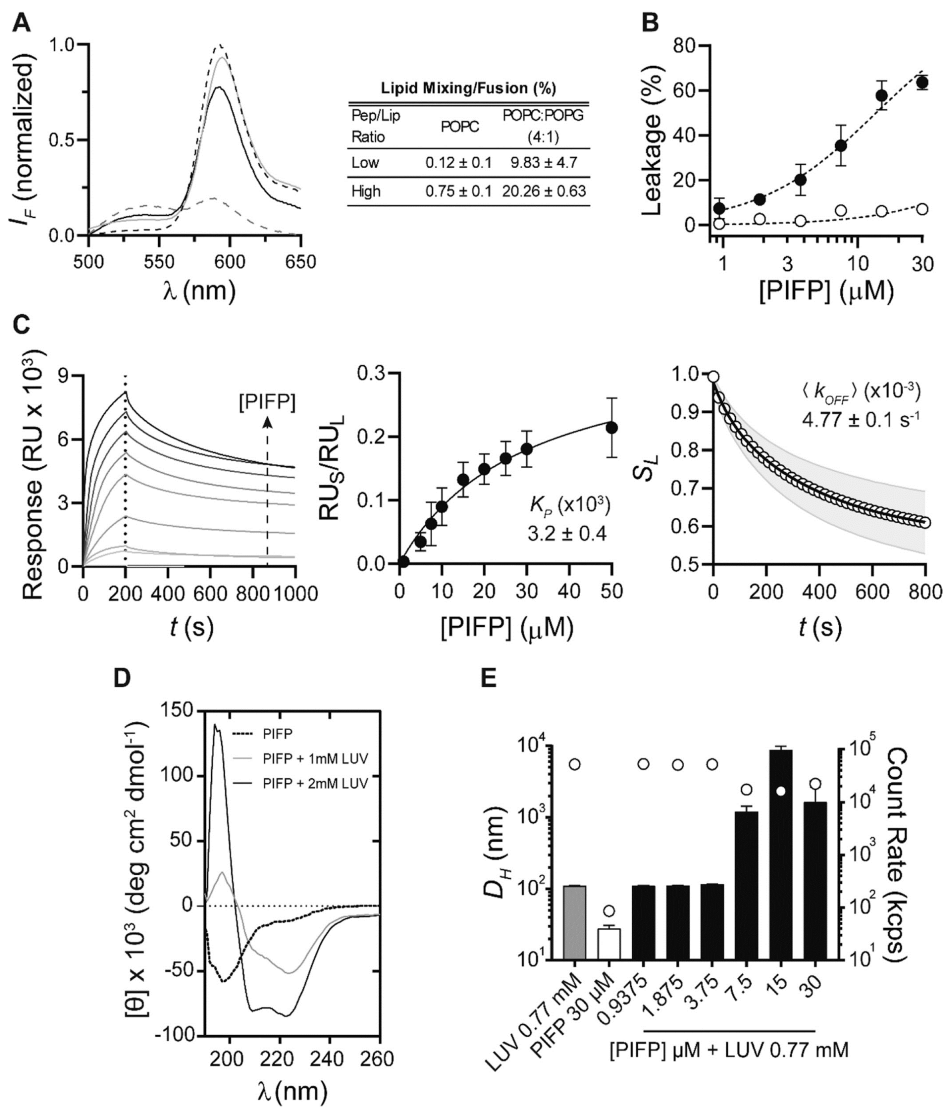


Figure 4.2: Characterization of PIFP–lipid membrane interactions in vitro. (A) Representative FRET PIFP–lipid mixing induction spectra (left) in POPC:POPG (4:1) vesicles (black dashed line, control; gray dashed line, Triton 1% (v/v); gray line, PIFP low pep/lip ratio; black line, PIFP high pep/lip ratio). PIFP–lipid fusion induction (%) (right) was extrapolated from equation 2. (B) PIFP–membrane perturbation induction. Leakage (%) from 0.77 mM large unilamellar vesicles (LUVs) [white, POPC; black, POPC:POPG (4:1)] loaded with 50 mM 5,6-CF was obtained upon incubation with PIFP in concentrations ranging between 0.9375 and 30 μ M. (C) Surface plasmon resonance (SPR) PIFP–lipid (un)binding kinetics. (Left) SPR sensorgrams of increasing PIFP concentrations (gray scale, from lighter to darker, 1, 5, 7.5, 10, 15, 20, 25, 30, and 50 μ M of PIFP) interacting with POPC:POPG (4:1) small unilamellar vesicle (SUV). A vertical dashed line at 200 s separates association and dissociation phases. (Middle) PIFP partition extends toward POPC:POPG (4:1) SUV. Individual values of RUS and RUL were obtained from each PIFP sensorgram at 200 s at the different concentrations tested. Data was fitted with equation 4, from which the K_P value was obtained. (Right) PIFP dissociation kinetics. Membrane-associated PIFP (SL) was plotted as a function of the dissociation time. Standard deviation interval is represented in gray. The data obtained was fitted by equation 5. (D) PIFP secondary structure. PIFP CD spectra were obtained in aqueous solution and in the presence of 1 and 2 mM of POPC:POPG (4:1) LUV. (E) PIFP and PIFP–LUV aggregation profiles. Z-average DH (columns) and count rate (circles) of the particles formed upon titration of PIFP in a 0.77 mM POPC:POPG (4:1) LUV solution were obtained by dynamic light scattering (DLS). Particles formed in the presence of 0.77 mM LUV and PIFP 30 μ M sample solution were obtained as controls.

on a small unilamellar vesicle (SUV)-covered surface coupled with a flow system. Peptide-to-lipid response ratios were plotted as a function of the injected peptide concentration, allowing an estimation of the peptide partition coefficient (K_P), as described elsewhere [328]. PIFP exhibits high lipid binding responses, in a concentration-dependent fashion (Figure 4.2C, left panel), leading to a $K_P \approx 3.2 \times 10^3$ (Figure 4.2C, middle panel). Given the PIFP fusion capacity, a K_P of this magnitude-similar to values described for other fusion peptides [329]-was expected. The sensorgram analysis also shows that there are variations in the dissociation kinetics (200–1000 s after injection) with the increasing peptide concentration, especially comparing the 30 and 50 μ M PIFP sensorgrams. For this reason, PIFP–membrane dissociation kinetics were further evaluated. PIFP–membrane dissociation follows a second-order kinetics, most likely derived from the presence of two peptide populations (Figure 4.2C, right panel): one faster, with lower mass and decreased retention, most likely corresponding to the individual peptide molecules; and another slower, with higher mass and increased retention, most likely corresponding to clusters formed by peptide molecules. Accompanied by a significant variability within the concentration range, the dissociation kinetics analysis suggests the presence of peptide–peptide and/or peptide–lipid aggregation dependent on the PIFP concentration. The weight-averaged dissociation constant ($\langle k_{off} \rangle$) extrapolated is $\approx 4.77 \times 10^{-3} \text{ s}^{-1}$, which is in fact quite similar to the $\langle k_{off} \rangle$ described for an aggregation-prompt peptide observed by Figueira *et al.*[328]. The PIFP–lipid binding analysis reinforces the argument of an

oligomerization/aggregation process present on the PIFP fusion mechanism. However, the macromolecular organization of the peptide-vesicle aggregates was still inconclusive.

4.4.3 PIFP Helix Content Depends on the Membrane Saturation Level

Another relevant aspect to evaluate is the PIFP secondary structure, not only in solution but also upon interaction with lipids [327]. Similar to other fusion peptides, PIFP evolves from a mainly random-coil conformation in aqueous solution to a α -helical conformation upon interaction with anionic POPC:POPG (4:1) membranes (Figure 4.2D). Additionally, we also observe that the PIFP α -helix content in the presence of large unilamellar vesicles (LUVs) depends on the lipid concentration. This can be inferred from (i) the existence of more than one isosbestic point in the spectra of Figure 4.3 and (ii) from the obvious departure from an α -helical spectrum in the 205–225 nm range at 1 mM of LUVs, where the spectrum minima are distorted. Note that, at 1 mM, a mixture of aqueous- and membrane-bound peptide populations is expected (for a $K_P \approx 3.2 \times 10^3$ and 1 mM lipid, about 30% of the peptide is expected to be unbound [330]); still, the spectrum at 1 mM lipid in Figure 4.2D cannot be explained by a combination of those at 0 and 2 mM lipids (hence the multiple isosbestic points), indicating the existence of at least a third population with different structural characteristics. Importantly, it is at lower lipid concentrations that the PIFP reaches the highest local peptide/lipid ratios in the membrane, which seem to be the conditions for this third, less helical population to occur. The relevance of these different peptide populations is discussed below.

4.4.4 Vesicle Aggregation Triggering Is Dependent on the PIFP Concentration

Membrane-surface peptide aggregation/oligomerization is commonly shared between viral fusion peptides [331]. The increased local peptide concentration on the membrane's surface enhances the destabilization of lipid-lipid interactions and associated membrane tension, promoting the mixture of the viral and host cell membrane lipid content or, in other cases, the formation of a pore [332]. However, the fusion peptide aggregation/oligomerization process differs between viruses, with direct impact on its fusion mechanism. PIFP revealed the tendency to accumulate on the lipid vesicle surface in a concentration-dependent fashion, perturbing its membranes and promoting the fusion between anionic vesicles. However, the aggregation profile of both PIFP alone and PIFP-LUV structures was still unclear. Using dynamic light scattering (DLS), we characterized the macromolecular organization of these structures. Hydrodynamic diameter (DH) and particle count rate were obtained upon PIFP consecutive addition to a 0.77 mM lipid vesicle solution (Figure 4.2E). Three important observations are worth mentioning: (1) PIFP does not form aggregates in aqueous solution; (2) upon increasing the PIFP concentration from 3.75 to 7.5 μ M, the LUV particles grow, at least, 10-fold on its DH, with a high variability on their sizes (Supplementary Figure S4.1); and (3) in parallel with this particle size increase, the particle count suffers an approximately 5-fold reduction. These

results suggest that upon reaching a critical aggregation concentration (CAC), between 3.75 and 7.5 μM , PIFP triggers the aggregation of lipid vesicles, increasing the particle size detected in solution. Taking the upper-bound PIFP CAC as 7.5 μM yields an approximate maximum ratio of 1 peptide to 20 anionic lipids in the system; this represents an overall balanced charge of -15, which is still far for an electrostatic neutralization hypothesis. Considering all of the data presented so far, it is most likely that PIFP aggregation triggering and simultaneous fusion mechanisms are dependent on a membrane-surface-controlled oligomerization and consequent pore formation. However, this hypothesis is hard to prove *in vitro*. For this reason, an *in silico* analysis of the PIFP–membrane interactions were performed.

4.4.5 CG Simulations Reveal the PIFP Aggregates When Inside a POPC:POPG Membrane

To complement the data obtained experimentally and to better understand what occurs at a molecular level upon PIFP insertion in the membrane, we performed CG and AA MD simulations. First, CG simulations of PIFP in a membrane bilayer were performed to understand how the PIFP assembles and later convert it to atomistic detail to analyze the molecular interactions that stabilize it. Only a mixed bilayer composed of POPC:POPG (4:1) was studied since the data we obtained for the PIFP–lipid fusion induction showed that this peptide has virtually no activity in pure POPC membranes (Figure 4.2A). To study the influence of different peptide concentrations, we built four Martini 3 [187] CG systems with increasing peptide/lipid concentrations and simulated each one for 20 μs . The detailed composition of the systems is shown in those tested experimentally, and a snapshot of these systems can be seen in Supplementary Figure S4.2. When performing preliminary short simulations of these systems, we observed that, when inserted vertically in the membrane, the peptides would come to the top of the outer leaflet. Since previous NMR data shows that the PIFP adopts a membrane-spanning conformation when interacting with POPC:POPG (4:1) membranes [149], we applied a flat-bottomed harmonic well potential in all systems to ensure that the peptide would remain vertically inserted in the membrane. It is worth noting that, further on, when converting these CG systems to AA, the peptides were able to remain vertically inserted in the membrane without needing any restraints. Moreover, our objective was not to assess the peptide orientation in the membrane, which was already known, but rather to determine how the peptides assemble and interact with each other in the membrane. Thus, we used the experimental findings regarding the PIFP orientation in the membrane to guide the simulations and reduce the degrees of freedom of the system.

After 20 μs of CG simulation, we saw that, in all of the systems with multiple peptides, these interacted to form oligomeric structures (as shown in the molecular images in Figure 4.3). Visual inspection of the trajectories revealed that these contacts were fairly persistent; i.e., after the first contact between any two peptides, the interaction was

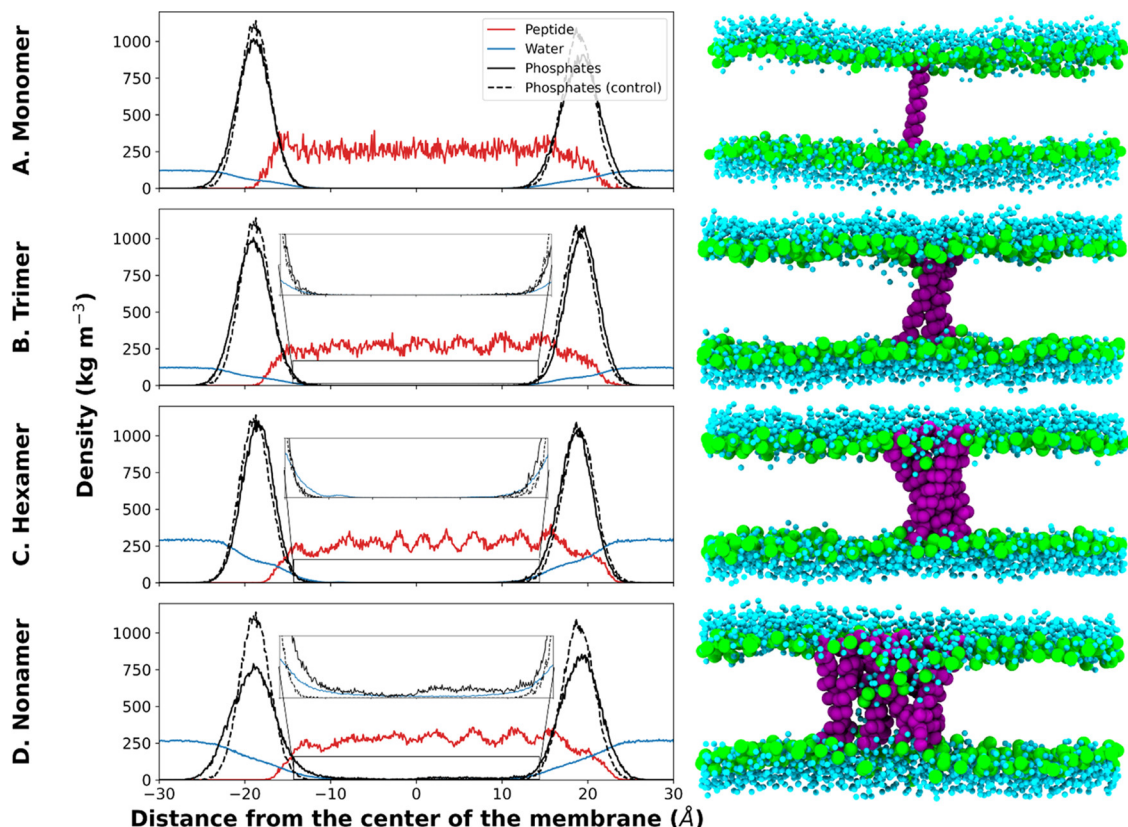


Figure 4.3: Density profiles obtained from CG simulations of the PIFP. The density profiles for the four systems with increasing peptide/lipid ratios (monomer < trimer < hexamer < nonamer) show the overall behavior of the system components during the simulation. The dashed lines represent the phosphate densities in systems without peptides; the solid lines are all relative to the systems with the PIFPs. The images on the right side are snapshots of the last frame of CG simulations, where the PIFPs are shown in purple, water molecules in blue, and the lipid phosphate groups in green.

maintained throughout the simulation. The concentration-dependent peptide–peptide and/or peptide–lipid aggregation observed when performing a PIFP–lipid binding analysis also reinforces this observation (Figure 4.3). The PIFPs stayed inside the membrane throughout the simulation, as seen in Figure 4.3, which indicates that the flat-bottom restraint applied was sufficient to recreate the experimentally observed orientation. In terms of peptide-induced membrane effects, we saw that, for the system with the highest peptide/lipid ratio (nine peptides), there is a tendency for the membrane leaflets to come close to each other. This is clear in Figure 4.3 D where there is a very slight decrease in the membrane thickness due to small membrane deformations better seen in the zoomed-in inset. Membrane hydration increases with increasing PIFP concentrations. For the two systems with the highest PIFP concentrations, waters can penetrate deeper than the lipid headgroups and reach the membrane core (Figure 4.3 C,D), which might be facilitated by the peptide aggregation mentioned above. A common effect of the addition of FPs to a membrane is the increase [306–308] or decrease in the lipid tail order [160, 164]. To measure

the membrane disturbance induced by these peptides on the membrane, we determined the lipid order parameters in all four CG simulations (Supplementary Figure S4.3). These results showed that lipid tail order decreases with increasing peptide concentrations. The behavior seen in the CG simulations is in agreement with what was seen experimentally, showing not only that these peptides tend to aggregate when inside a membrane [301] but also that these clusters facilitate the entry of water molecules to its core [168].

4.4.6 PIFP That Can Adopt Moderately Helical Structures in Water and Helicity in the Membrane Is Influenced by the Peptide/Lipid Ratio

During the fusion process, the F protein suffers conformational changes and, before inserting into the membrane, the PIFP becomes exposed to water. To characterize the peptide's structural properties in aqueous solution, we performed metadynamics simulations and, in line with the CD results, the PIFP is mainly a random coil in aqueous solution, but it has a heterogeneous energy landscape and can easily adopt moderately helical structures (Figure 4.4). This can facilitate its transition into the helical structure that it adopts in the membrane.

To analyze the PIFP secondary structure in the membrane with atomistic detail, we converted the final systems of the CG simulations to AA resolution. Triplicates of all four systems were simulated for 1 μ s without any restraints. The analysis of the peptide's secondary structure in a POPC:POPG membrane revealed that at low peptide/lipid ratio (1 peptide copy) the peptide maintained its helical conformation losing it only at the N- and C-terminal ends (Figure 4.5A). For the other three systems, there was also the loss of structure at both ends, with the N-terminus being the most impacted (Figure 4.5B-D). In the system with three peptides, the first two N-terminal residues lose entirely their helical structure, in all replicates. In the hexamer, the loss of helicity affects five N-terminal residues, and in the nonamer, six residues are affected. The loss of structure seen in these regions may be due to the passage of water molecules into the porelike structures and consequent increase in the number of PIFP–water contacts in this region (Supplementary Figure S4.4). This is in line with the results seen in metadynamics simulations and CD analysis, which shows that the peptide's structure is destabilized in the aqueous environment. Concurrently, in the six-copy system, there is additional loss of α -helical structure at residues T117, A118, A119, and Q120. This is the same region where a bend, with its vertex at T117, is found in the prefusion crystal structure of the parainfluenza fusion protein [165, 169], indicating that this region might be flexible.

When looking at the interpeptide contacts between residues (those within 4.5 Å; Figure 4.6), these are prevalent in the C-terminal region of the trimer, unlike the N-terminal region, where no interpeptide contacts are seen (Figure 4.6A). A different behavior is seen in the systems with higher PIFP concentrations, where both terminals have peptide-peptide contacts (Figure 4.6B,C). The prevalence of these contacts in the N-terminal region increases from the hexamer to the nonamer but they are always less frequent than in the

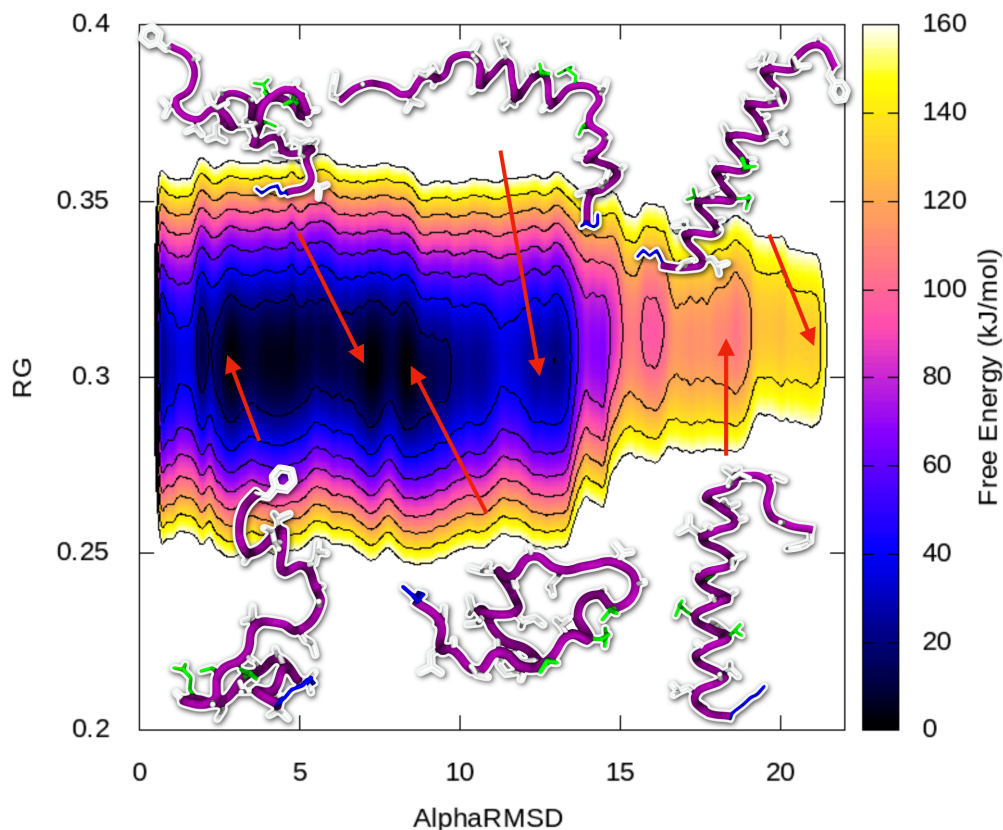


Figure 4.4: Free-energy surface of the PIFP in water obtained from Metadynamics calculations. A two-dimensional (2D) free-energy landscape was built as a function of the CVs AlphaRMSD and radius of gyration (RG). The free energy was reconstructed from the bias potential, and contour lines are used to represent regions with identical free-energy values. Structures corresponding to the most relevant basins are illustrated using molecular images built with VMD [227] using the same color scheme as in Figure 4.1C.

C-terminal region.

The regions of the PIFPs, where low percentages of interpeptide contacts were identified, also form stable interactions with the membrane lipids (Figure S4.5), guaranteeing the stability of the oligomeric structures in the membrane. When analyzing the interactions between the PIFPs and the PG lipid heads, the N-terminal region that is predicted to insert all the way through the membrane is the one with the most contacts (Figure S4.6). This is compatible with the fact that, in the plasma membrane, the negatively charged lipids are found in the inner layer of the membrane and, consequently, only the N-terminal region of the peptide would be able to interact with these lipid heads. Since peptide bending has been described as an important FP feature to induce membrane disturbance

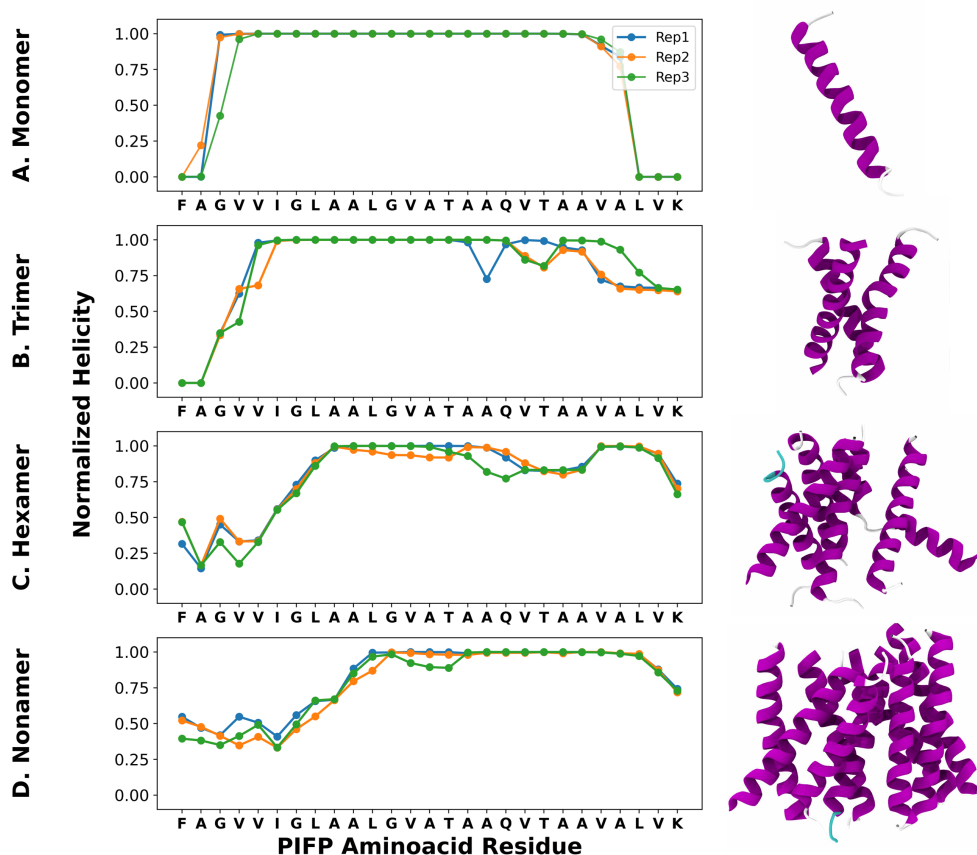


Figure 4.5: PIFP dynamic helicity during atomistic MD simulations. The plots show the average helicity values obtained for each residue in each of three replicate 1 μ s long MD simulations (data for each replicate is shown in a different color). The images to the right are snapshots of the simulation with the helix motifs highlighted in purple.

and consequent reduction of the energetic barrier for membrane fusion [157, 333], we analyzed if the PIFP forms the same bend inside this membrane bilayer (Supplementary Figure S4.7). The helix geometry analysis showed that in one replicate of the hexameric system, there is increased departure from a straight helix (represented in Figure S4.7 by helix bending angles farther from 180 degrees), supporting that the local loss of helicity around T117 is induced by an increase in peptide curvature. Overall, there seems to be a correlation between the number of copies of the PIFP and the loss of α -helical structure. This observation is in good agreement with the data we obtained by CD (Figure 4.2D) and previous NMR structural analysis [149]. Both the CD and NMR experiments [149] showed that at low peptide/lipid ratios, in a POPC:POPG membrane, the peptide has an α -helical structure, and at higher ratios, our CD data shows that the helical structure is less well defined (Figure 4.2D, gray line). Finally, we hypothesize that at low PIFP concentrations the peptide maintains its α -helical structure due to the hydrophobic environment inside the membrane, losing its structure only in the terminal regions (more exposed to water).

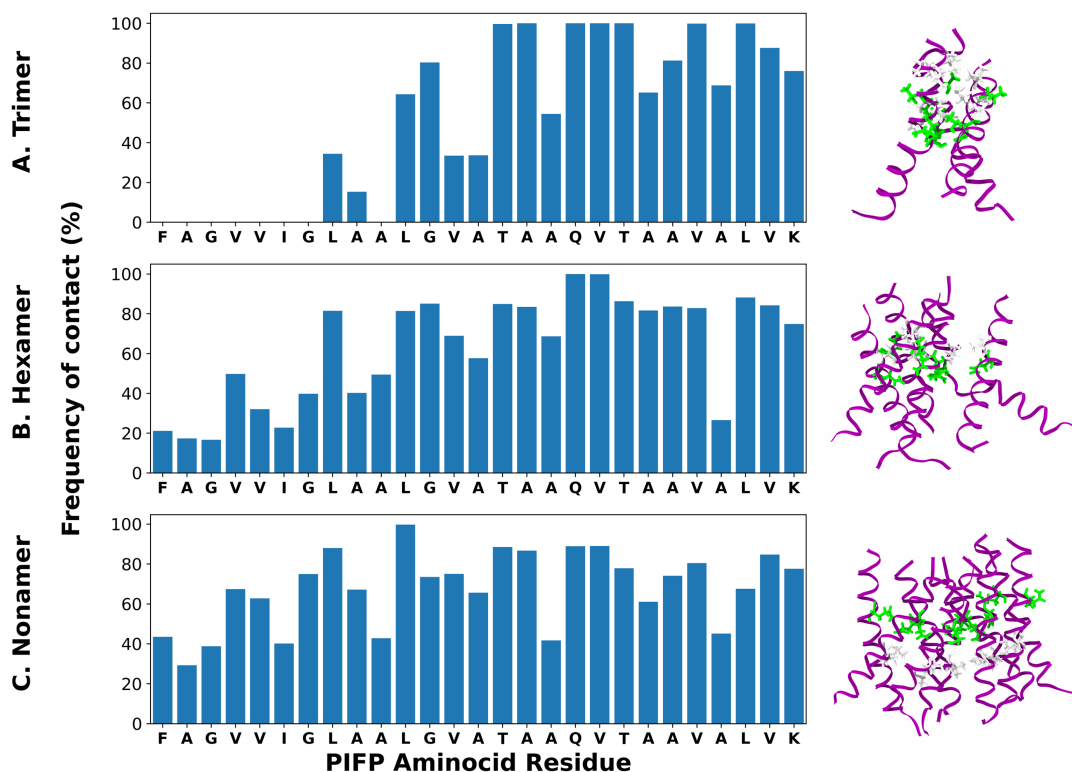


Figure 4.6: Prevalence of peptide-peptide contacts. The plots show the prevalence of contacts between different peptides, within 4.5 Å, established by each amino acid residue, and averaged over the simulation. This analysis was only performed for the systems with more than one peptide. Each plot has to its right a snapshot of the final frame of the corresponding simulation, where the PIFPs are shown in purple and the residues involved in the peptide-peptide contacts with a prevalence higher than 90% are highlighted in green (polar residues) and white (nonpolar residues).

With the increase in the number of copies of the PIFP, there is a loss of structure due to the formation of prevalent interpeptide interactions.

4.4.7 PIFP Peptide-Peptide Interactions Stabilize Porelike Structures

As mentioned earlier, during the membrane fusion process, after binding to the target cell, conformational changes occur in the F protein leading to its trimerization, and at least two parainfluenza F protein trimers are necessary for the viral fusion process to occur [179, 297, 298]. This means that at least six fusion peptides are involved in the fusion process. To determine if there is a specific contact region between the PIFPs, we analyzed the data obtained for the peptide-peptide contacts occurring between residues of different peptides (Figure 4.6) and determined the accessibility of water to the core of the membrane (Figure 4.7A–C).

Through the contact analysis, we saw that, in the systems with more than one peptide,

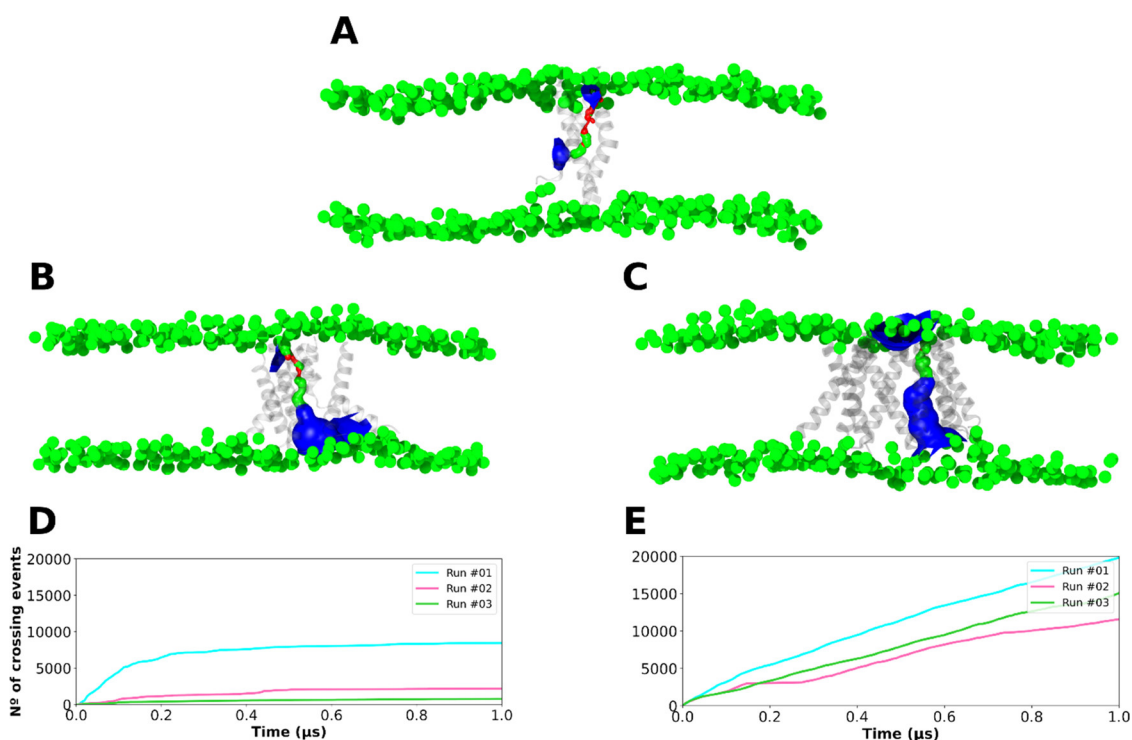


Figure 4.7: PIFP pore characterization. Panels (A–C) are figures generated by the HOLE2 analysis tool from MD analysis that show the accessibility of water to the membrane for the trimeric, hexameric, and nonameric systems, respectively. The parts that are inaccessible to water (pore radius < 1.15 Å) are shown in red, water-accessible parts (1.15 Å $<$ pore radius < 2.30 Å) are shown in green, and wide areas (pore radius > 2.30 Å) are displayed in blue. Panels (D) and (E) show the number of water-crossing events that were determined for the systems with six and nine PIFPs, respectively.

the residues forming more prevalent contacts were those with hydrophobic side chains (L110, L113, A118, V121, V125, and L127) and the polar residues (T117, Q120, and T122). Additionally, in the trimeric PIFP, there is 100% of prevalence of contacts involving residues T117, A118, Q120, V121, and T122, all located in the C-terminal region of the PIFP. These contacts create a specific surface of contact stabilizing this end of the structure (Figure 4.6A, molecular image). This is also the region where the HOLE2 tool found the space for the entry of water molecules (Figure 4.7A), but the inexistence of interpeptide interactions at the other end (the N-terminal region) prevents the formation of a viable pore. In the hexamer, contacts involving Q120 and V121 residues are maintained throughout the simulation, but, unlike the trimer, N-terminal residues form transient contacts (with a prevalence of around 20%) that stabilize the transmembrane structure enough to create just enough space for the passage of water through the membrane (Figure 4.7B). Finally, in the nonamer, even though interpeptide interactions are less prevalent, they are more vertically distributed. Contacts with L113 are seen all throughout the simulation, and the contacts at the N-terminal have an incidence close to 40%. Similarly to the trimer and hexamer, the N-terminal region has a relatively low interpeptide contact incidence,

which we found interesting given that several hyperfusogenic mutations occur in this region [334]. The effect of these point mutations on the peptide–peptide clustering, and consequently on the PIFP fusion, should be addressed in future studies. Nonetheless, the formation of several interpeptide contacts within the nonamers’ PIFPs forms a wide water-accessible region inside the membrane and, consequently, the ideal conformation for the formation of a porelike structure (Figure 4.7C). Looking at RMSD data of the nonamer (Supplementary Figure S4.8-4th row), there is a significant deviation in peptide 4 of replicate 3. Superimposing the last frames of the CG and AA simulations, considerable changes in conformation and bending of peptide 4 are clear (Supplementary Figure S4.9-peptide in red). However, the impact on the overall structure of the porelike structure and water penetration is identical when comparing AA and CG simulations (Supplementary Figure S4.4). These results show that the increase in the PIFP concentration facilitates the formation of peptide–peptide interactions between the PIFP’s hydrophobic and polar residues and that these stable interactions are of utmost importance for the assembly of porelike structures inside the membrane. Additionally, the formation of these structures in our simulations can explain the increased vesicle leakage seen experimentally at high peptide/lipid concentrations (Figure 4.2B).

4.4.8 Water Can Pass through the PIFP Porelike Structures

To determine if the previous porelike structures enable the passage of water, thus explaining the PIFP-induced leakage of 5,6-CF-loaded vesicles seen experimentally (Figure 4.2B), we analyzed the accessibility of water to the core of the membrane and the flux of water molecules through the pores. For the latter, we used the fluxer [335] script, which counts the number of crossing events across a finite thickness xy-plane membrane. The results showed that both systems form pores that are vehicles for the passage of water molecules through the membrane (Figure 4.7D,E). In the system with six peptides, the flux analysis showed that there is a positive flux of water until approximately 200 ns of simulation (in replicate 1). At this point, the flux is slowed/stopped (the slope of the line becomes close to 0). An explanation for the abrupt deceleration seen in the water flow can be the partial closing of the pore in this replicate. For replicates 2 and 3, the cumulative flux is lower but continuous; this suggests that in these replicates there is no initial open state (Figure 4.7D). For the nonamer, the flux of water inside the pore is continuous in all replicates. The water flux is initially identical in all three replicates, but different peptide rearrangements in each replicate then lead to some divergence. From 500 ns of simulation onward, the flux becomes again approximately the same for all three replicates (Figure 4.7E). A density analysis of the position of lipid heads and water molecules during the simulations was also performed (Figures S4.10 and S4.11, respectively), with results consistent with the passage of water through the core of the membrane. First, when looking at the lipid head density data for the two systems with the highest peptide concentration (from the top view), we see an opening on both sides

of the membrane (Supplementary Figure S4.10C,D). The absence of lipid heads in this area is an indication that the oligomeric structures can form a stable peptide arrangement that removes the lipids from this central area. In this same region, by analyzing the water density data, we can see that it is common to find water molecules (in blue) in the core of the membrane (in white) (Supplementary Figure S4.11C,D). Overall, we saw that at higher peptide concentrations there is a continuous flux of water through the membrane due to the formation of a stable oligomeric PIFP structure that, despite not being a requirement for membrane fusion, is a sign of strong PIFP-induced membrane disturbance. This flux of water through the membrane explains the experimental PIFP-induced leakage observed at higher PIFP concentrations (Figure 4.2B).

4.4.9 PIFP Induces Concentration-Dependent Lipid Tail Protrusion

It is known that to promote membrane fusion, the FPs must perturb the properties of fluid lipid bilayers. Effects induced by the FPs can be the induction of spontaneous curvature, hydration, changes in acyl chain order [336], or promotion of lipid tail protrusion. To measure the membrane disturbance induced by the increasing concentrations of peptide, we determined the order parameters, the number of membrane lipid tail protrusion events, and peptide–lipid contacts occurring during the simulations. The lipid order values were obtained by averaging the S-values for each acyl chain carbon of the lipid tails over time and were calculated for the lipids in a pure membrane (without peptides, as a control) and with the peptides (Supplementary Figure S4.12). Slightly higher S-values were obtained in the presence of the PIFPs, especially in the nonamer, whereas the monomer S-values are practically the same as the control's (Supplementary Figure S4.12A). The order behavior is the same when looking only at the lipids closest to the peptides (within 5 Å; Supplementary Figure S4.12B). Another mechanism that has been proposed to explain the FP-induced membrane fusion is the promotion of lipid tail protrusion events. These were first described based on MD data and correspond to the outward extension of lipid acyl chains, beyond the corresponding phosphate group, facilitating the fusion between opposite membranes [158–160, 175, 309] (Figure 4.8C,D). We also observed this when simulating the influenza FP in a membrane bilayer [159, 160, 175].

The lipid tail protrusion results showed that there is a gradual increase in the number of protruding lipids when the number of simulated peptides increases: monomer < trimer < hexamer < nonamer (Figure 4.8A). This increase in the peptide/lipid ratios leads to a linear increase in the number of protruding lipids, which precludes the occurrence of synergistic. We also analyzed which residues were closest to the protruding lipids (Figure 4.8B) and, as is the case for other FPs, the N-terminal residues [158–160, 172] seem to be those inducing the most lipid protrusion events. The N-terminal phenylalanine residue seems to play a key role in the induction of headgroup intrusions (the lipid headgroup penetrates deeper into the membrane) [158–160, 172] (Supplementary Figure S4.10), an event that facilitates the occurrence of lipid tail protrusion events. Peptide–lipid contact

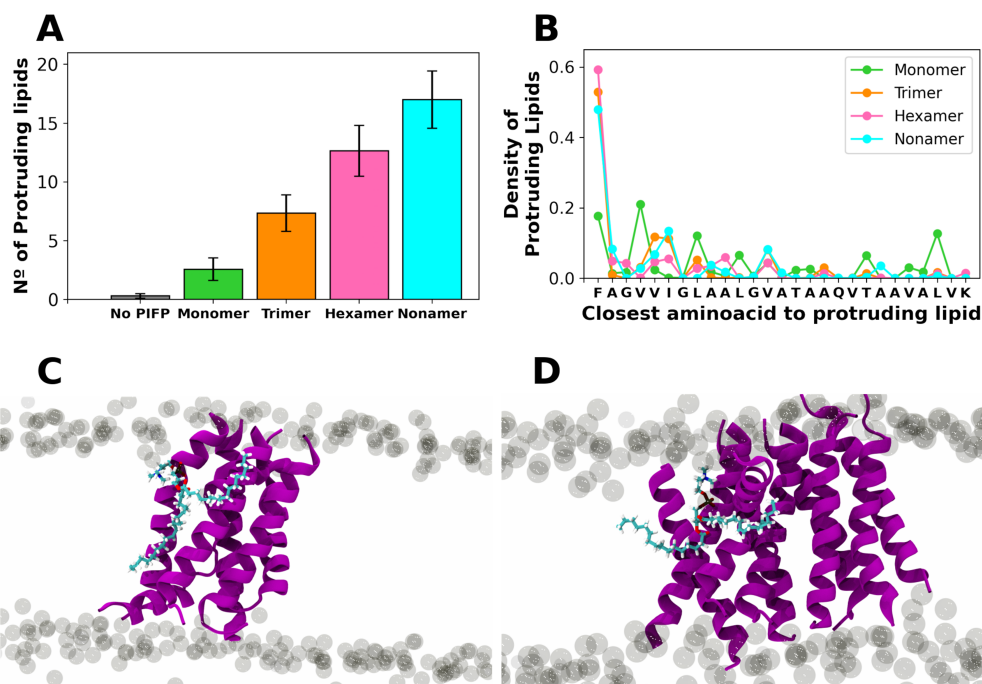


Figure 4.8: Lipid tail protrusion. (A) The plot shows the average number of protruding lipids at 4 Å from the peptides. A pure membrane (no PIFP) was used as a reference for the lipids that protrude without the influence of peptides. Error bars represent the 95% confidence intervals and were calculated with bootstrap resampling. Plot (B) shows the amino acid residues that are closest to a protruding lipid. Panels (C) and (D) are snapshots of the hexameric and nonameric simulations, respectively, where a lipid is highlighted in blue to evidence the lipid tail protrusion event.

analysis also shows that the N-terminal region of the trimer and hexamer interacts preferentially with POPG lipids (Supplementary Figure S4.5) and, on average, these oligomeric PIFPs contact with POPG lipids 1.5 more times than with POPC (Supplementary Figure S4.12). Interestingly, the results indicate that oligomerization facilitates the N-ter:POPG interactions since the interaction is much less frequent in the monomer than in the systems with multiple peptides. Lipid head density analysis shows that for the systems with higher peptide concentrations, it is common to find the lipid heads inside of the membrane (Supplementary Figure S4.10C,D). This is consistent with the notion that the N-terminal residues of the PIFP interact with the lipid heads, inducing lipid head intrusion that, in turn, facilitates the protrusion of the corresponding tails. These results show that the PIFP induces lipid tail protrusion that increases with the concentration of peptide. This is in very good agreement with the peptide-induced lipid mixing data, which show a similar trend: at higher peptide concentrations, there is more lipid mixing (Figure 4.1A, right panel).

4.5 Discussion

The aim of this work was to characterize the mechanisms of membrane fusion that occur during parainfluenza infection. With this in mind, we performed CG and atomistic molecular dynamics (MD) simulations, together with spectroscopic experiments of the PIFP in a POPC:POPG (with a ratio of 4:1) lipid bilayer. An important aspect that remained unclear is the effect of concentration/oligomerization of the PIFP. We addressed this question by performing a thorough computational and experimental characterization of systems with increasing peptide/lipid ratios. Experimental CD results showed that, upon interaction with membranes, the PIFP evolves from a random-coil conformation in solution to an α -helical conformation. Metadynamics simulations revealed that, although the peptide is mainly unstructured in water, it can adopt moderately helical structures, which might facilitate the acquisition of a helical structure upon membrane interaction. The concentration-dependent behavior of the secondary structure in the membrane was also seen in the AA MD simulations where the PIFP structure is α -helical at low PIFP concentrations, with some structure loss seen when increasing the peptide concentration. This is likely due to the formation of polar and hydrophobic contacts between the peptides, which lead to the formation of oligomeric structures. The formation of these oligomeric structures explains the 5,6-CF leakage increase at higher PIFP concentrations. In fact, when analyzing the peptide-membrane interactions, the simulation results show that with higher peptide concentrations the oligomeric structures form porelike structures that facilitate the flux of water through the membrane. Although the PIFP had been suggested to form a fusion pore [168], this is the first time that it is shown to spontaneously assemble into oligomeric structures that induce the formation of a water-permeable pore. Moreover, the computational analyses reveal that these peptides can induce lipid protrusion events without greatly altering the lipid order. This is in line with what has been observed for other FPs [159, 160, 164, 172, 175] and indicates that lipid tail protrusion plays an important role in viral membrane fusion (see Lousa and Soares [161] for a review that discusses this topic). The results obtained complement the previous experimental studies and strongly support the hypothesis that the FP promotes membrane fusion by destabilizing the host membrane, a role that is better fulfilled at higher peptide concentrations. Even though in our *in silico* studies both the hexameric and nonameric conformations play an essential role in the membrane destabilization, *in vitro* studies revealed that, when in phospholipid micelles, the PIFP assembles preferably into hexamers [168]. Interestingly, the effect of this peptide on the membrane, namely, the induction of lipid tail protrusion [159, 164, 172, 175] and the entry of waters to the core of the membrane when the peptides are transmembrane, [337] is similar to that observed for the influenza fusion peptide, supporting the hypothesis that fusion peptides from different viruses share common mechanisms. These findings provide a valuable contribution to the advance of the knowledge on viral fusion peptides, which, given their importance, can be very useful for the development of antiviral therapies.

4.6 Supplementary Information of Parainfluenza Fusion Peptide Promotes Membrane Fusion by Assembling into Oligomeric Porelike Structures

4.6.1 Molecular Dynamics Simulations Analysis Tools

Secondary structure analysis – The do_dssp program was used to obtain the secondary structure data of the PIFPs throughout the simulation. This tool reads the trajectory file and computes the secondary structure for each time frame by calling the dssp program [338].

Inter-peptide contact analysis – The prevalence of peptide-peptide contacts formed between different peptides was determined by counting the number of residues at less than 4.5 Å from another peptide's residue.

Pore formation analysis – The HOLE2 Analysis tool from MDAnalysis [233, 285] was used to analyze the propensity for the formation of pore structures in all systems with more than one peptide.

Water flux analysis – The fluxer.py script [335] developed to count the number of crossing events across a finite width *xy*-plane membrane, was used to quantify the flux of water across the membranes.

Lipid-tail protrusion events analysis – Lipid tail protrusion events were determined by counting the number of lipids, around 4Å of the peptides, for which, at least one of the carbon atoms in the tail protruded more than 0.1 nm above the phosphate headgroup. The number of events per frame was calculated for both membranes with and without peptides, with the latter serving as a baseline for the protrusion events that are not induced by the peptides.

Lipid tail order analysis – Lipid order values were obtained by averaging the S-values for each acyl chain carbon of the lipid tails over time. These values were calculated for the lipids in a pure membrane (without peptides - as a control) and in the systems with the peptides.

Helix geometry analysis – The HELANAL tool from MDAnalysis [339] was used to characterize the helical geometry of the PIFPs from its C- α atom coordinates. In systems with more than one peptide the average angle

Water and Lipid head density analysis – The water and lipid head density was determined throughout the simulation using DensityAnalysis from MDAnalysis [233, 285]. This program reads the trajectory file frame by frame, and the atoms in atomgroup are histogrammed on a 3D grid with spacing delta.

Root mean square deviation (RMSD) – Temporal evolution of the C α root mean square deviation in AA MD simulations was calculated with the MDAnalysis.analysis.rms module. The starting structure of each simulation was used as the reference structure and the initialization steps from the calculation were excluded. Only the C α atoms were used for fitting and for the RMSD calculations.

Root mean square fluctuation (RMSF) – The root mean square fluctuation (RMSF) was calculated with the GROMACS rmsf tool using the starting structure of each simulation

CHAPTER 4. PARAINFLUENZA FUSION PEPTIDE PROMOTES MEMBRANE FUSION BY ASSEMBLING INTO OLIGOMERIC PORELIKE STRUCTURES

as the reference structure and the initialization steps from the calculation were excluded. Only the last 500ns of simulation were considered for the RMSF calculations.

CG to AA conversion – The last frame from the CG simulations to atomistic detail was performed with the backward.py script [318].

PIFP Concentration	Number of PIFP	Pep/Lip Ratio
High	9	1/42
Medium	6	1/67
Low	3	1/154
Very low	1	1/492

Table S4.1: List with detailed composition of the systems to be studied

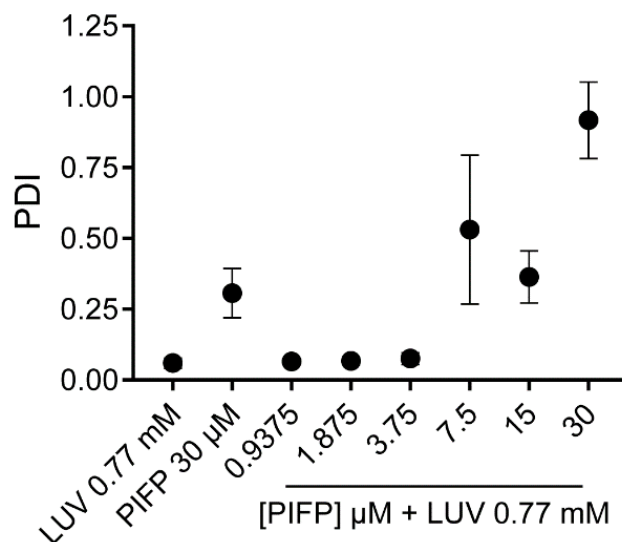


Figure S4.1: PIFP-induced LUV aggregates polydispersity index (PDI). PDI of the resultant particles formed upon titration of PIFP in a 0.77 mM POPC:POPG (4:1) LUV solution were obtained by DLS.

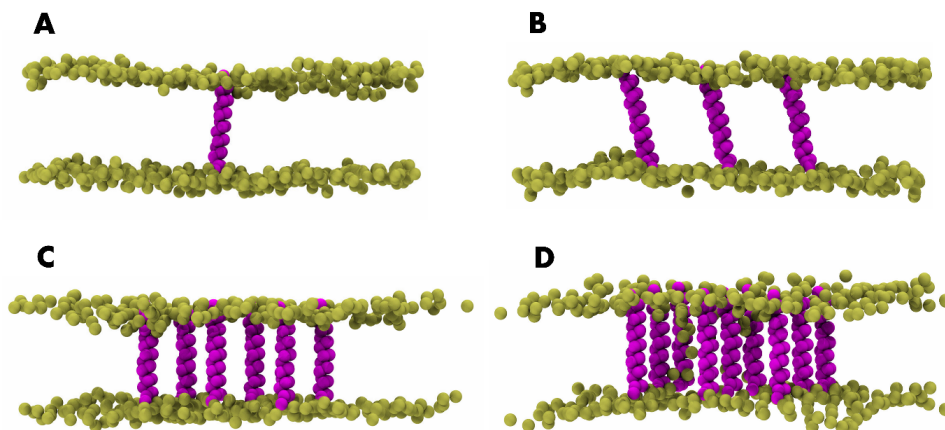


Figure S4.2: CG systems created to study the PIFP's behavior in a POPC:POPG membrane. Images are shown in a crescent peptide/lipid ratio: The monomeric system has the lowest peptide/lipid ratio (A - 1/492), followed by the trimeric (B - 1/154), the hexameric (C - 1/67) and finally the nonameric system (D - 1/42).

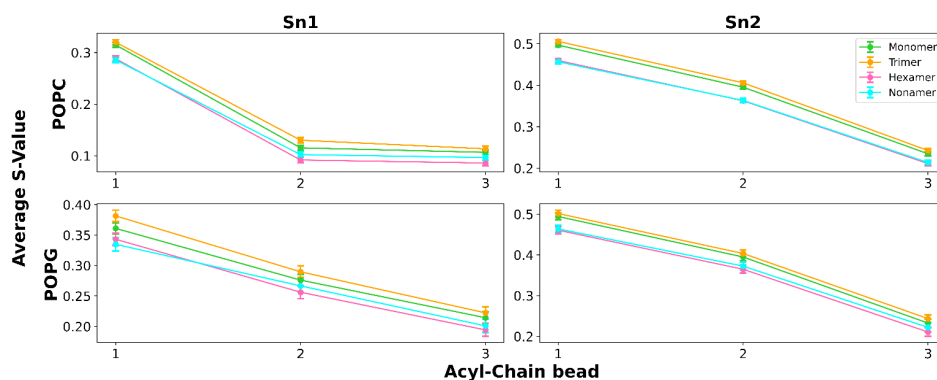


Figure S4.3: Lipid order parameters for the lipid tails in the CG systems. (A) Lipid order values calculated for all membrane lipids in all membrane systems. Values obtained for the POPC lipids (top) and POPG (bottom) are shown separately. Order values for saturated and unsaturated tails are shown in the 1st and 2nd rows, respectively. Error bars represent the 95 % confidence intervals and were calculated with bootstrap resampling.

CHAPTER 4. PARAINFLUENZA FUSION PEPTIDE PROMOTES MEMBRANE FUSION BY ASSEMBLING INTO OLIGOMERIC PORELIKE STRUCTURES

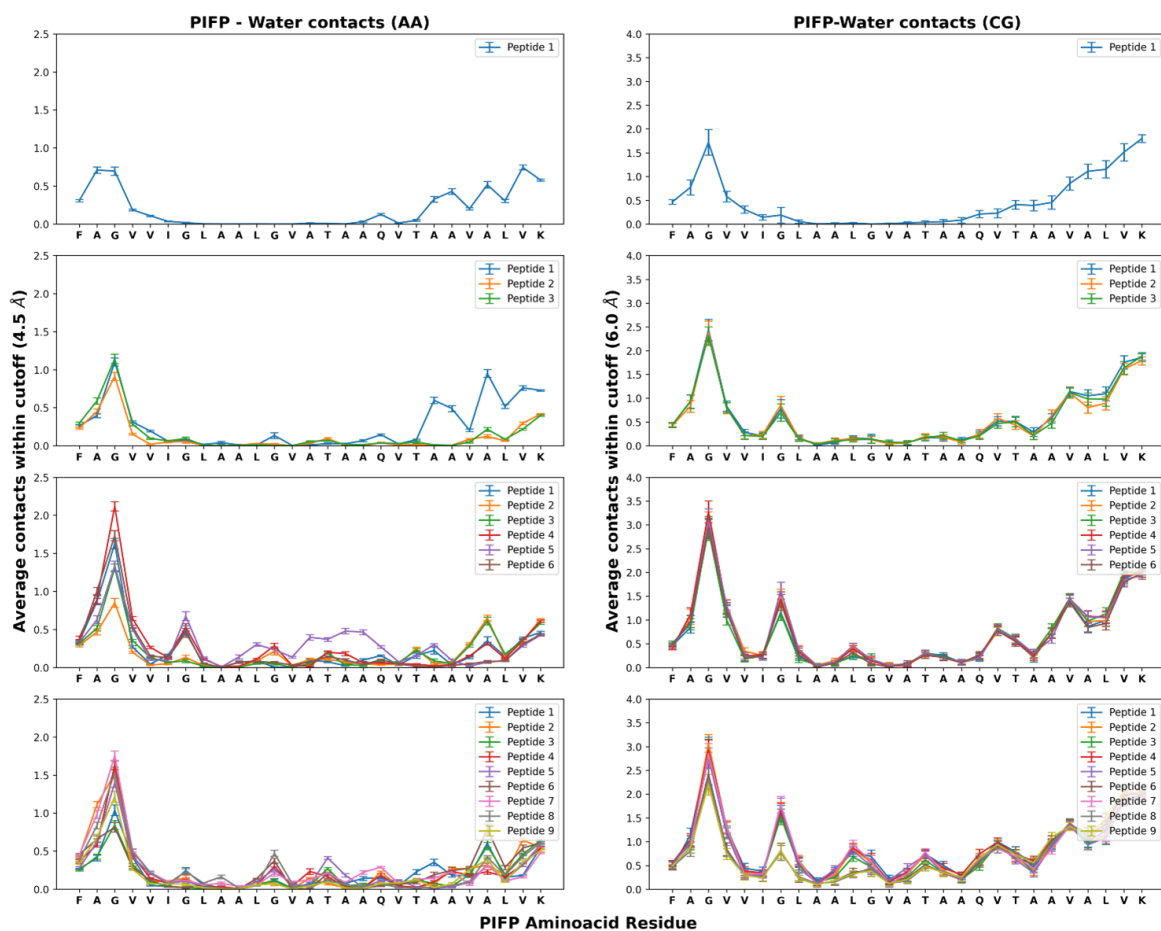


Figure S4.4: Average number of contacts formed by each residue of the PIFPs and the water molecules (left column – AA simulations; right column – CG simulations), within a 4.5 Å radius. Each row corresponds to a given system: Monomeric (1st row), Trimeric (2nd row), Hexameric (3rd row), and Nonameric (4th row).

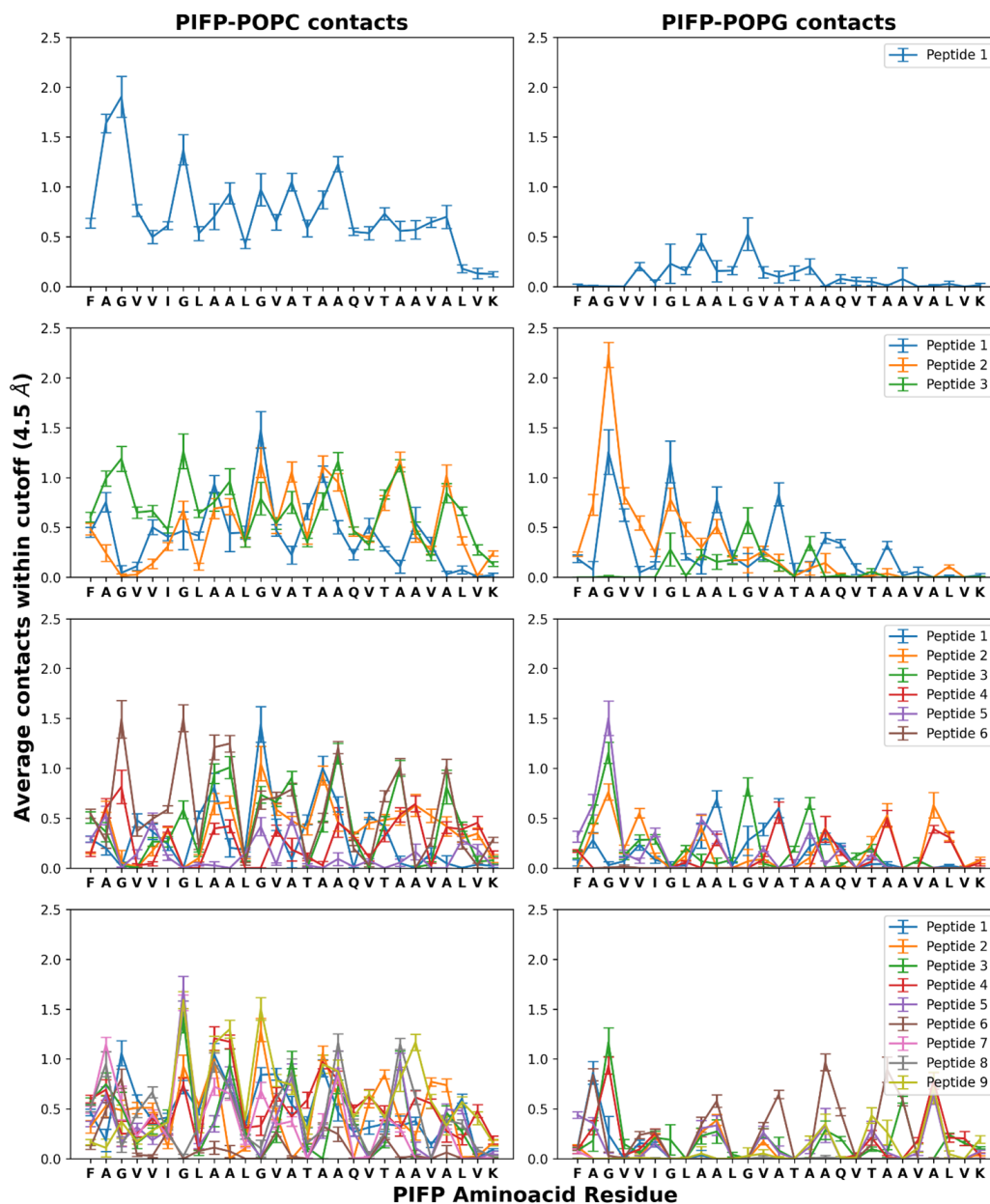


Figure S4.5: Average number of contacts formed by each residue of the PIFPs and the membrane lipids (left column – POPC; right column – POPG), within a 4.5 Å radius, in the AA simulations. Each row corresponds to a given system: Monomeric (1st row), Trimeric (2nd row), Hexameric (3rd row), and Nonameric (4th row).

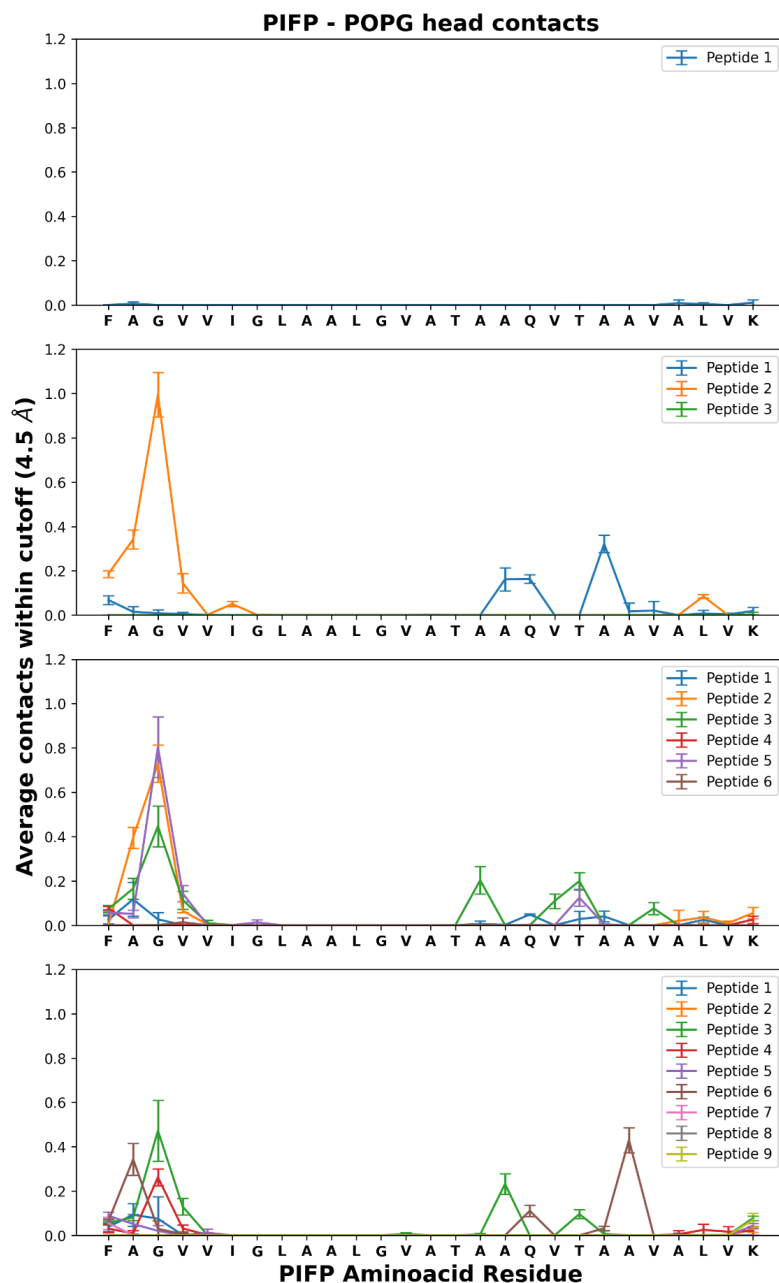


Figure S4.6: Average number of contacts formed by each residue of the PIFPs and the heads of the POPG lipids, within a 4.5 Å radius, in the AA simulations. Each row corresponds to a given system: Monomeric (1st row), Trimeric (2nd row), Hexameric (3rd row), and Nonameric (4th row).

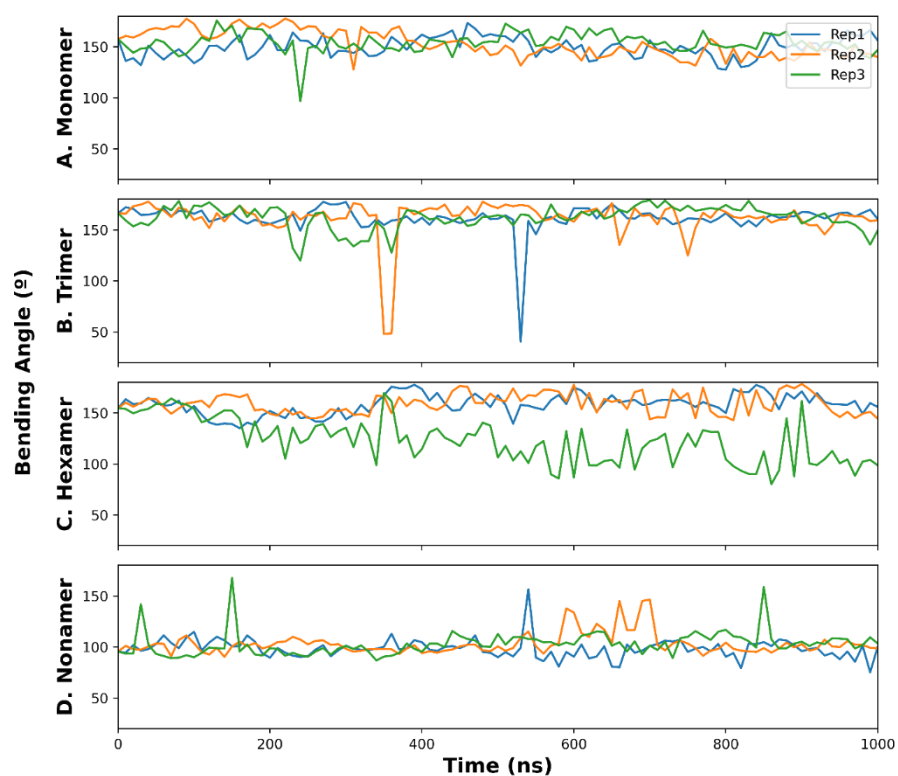


Figure S4.7: PIFP helix geometry in AA simulations. The HELANAL tool from MDAnalysis was used to characterize the helical geometry of the PIFP from its $C-\alpha$ atom coordinates. In systems with more than one peptide the plot shows the average angle. Each row corresponds to a given system: Monomeric (1st row), Trimeric (2nd row), Hexameric (3rd row), and Nonameric (4th row).

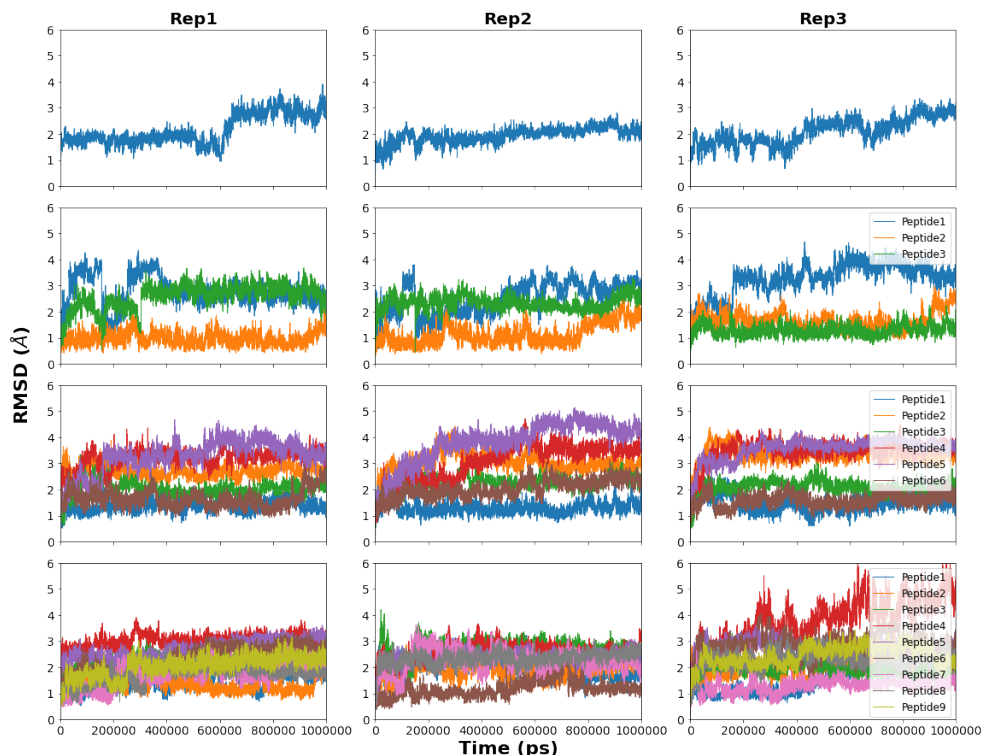


Figure S4.8: RMSD for the PIFP. Temporal evolution of the $C\alpha$ root mean square deviation in AA MD simulations. The root mean square deviation (RMSD) was calculated with the MDAnalysis.analysis.rms module from MDAnalysis. Each row corresponds to a given system: Monomeric (1st row), Trimeric (2nd row), Hexameric (3rd row), and Nonameric (4th row). Each column corresponds to an independent replicate simulation performed in the same conditions.

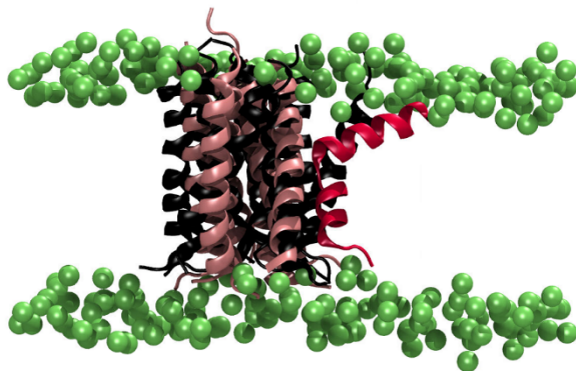


Figure S4.9: Superimposed snapshots of the nonameric AA (replicate 3) and CG simulations last frames. PIFPs in black correspond to the last frame of the CG simulation. PIFPs in pink correspond to the last frame of the replicate 3 AA simulation, with the peptide 4 in red.

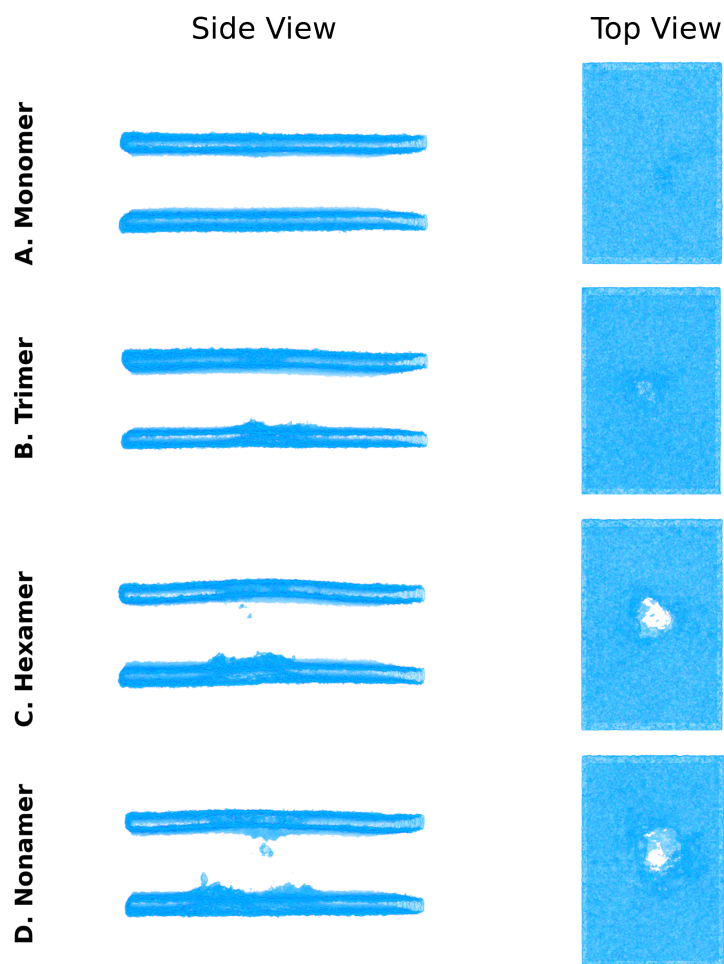


Figure S4.10: Lipid head position with probability higher than 40%. Lipid head positions were determined throughout AA simulations using DensityAnalysis from MDAnalysis. Each row corresponds to a given system: Monomeric (1st row), Trimeric (2nd row), Hexameric (3rd row), and Nonameric (4th row).

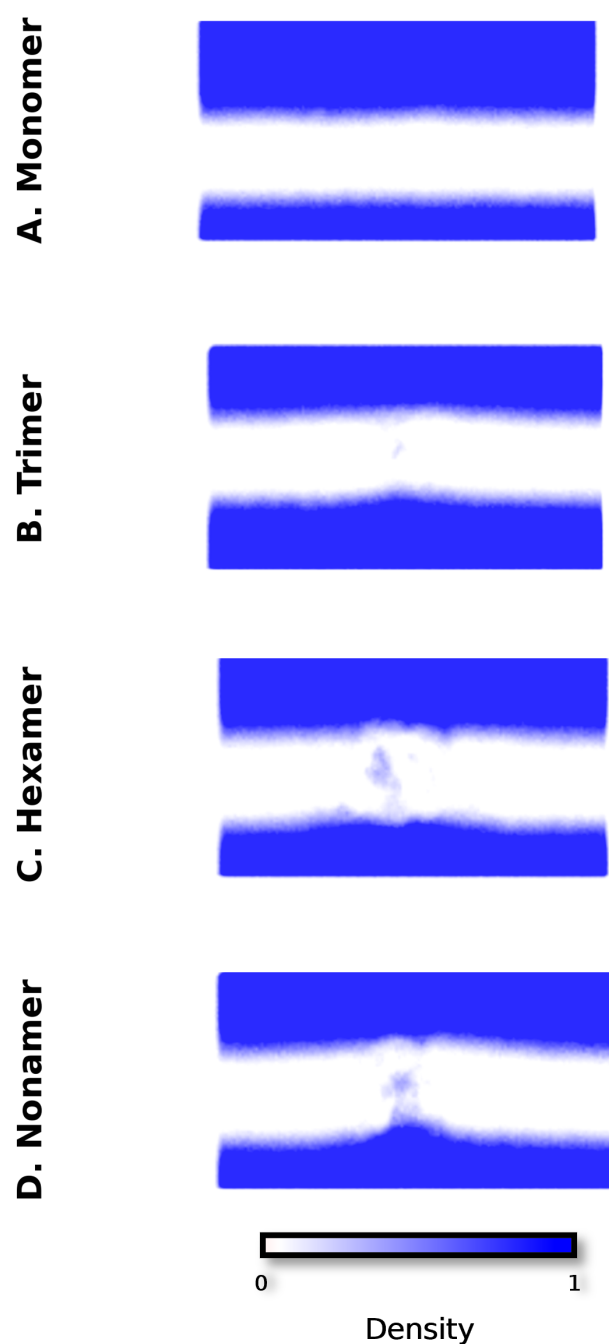


Figure S4.11: Water density in AA simulations. Water molecule location was determined throughout the simulation using DensityAnalysis from MDAnalysis. Each row corresponds to a given system: Monomeric (1st row), Trimeric (2nd row), Hexameric (3rd row), and Nonameric (4th row).

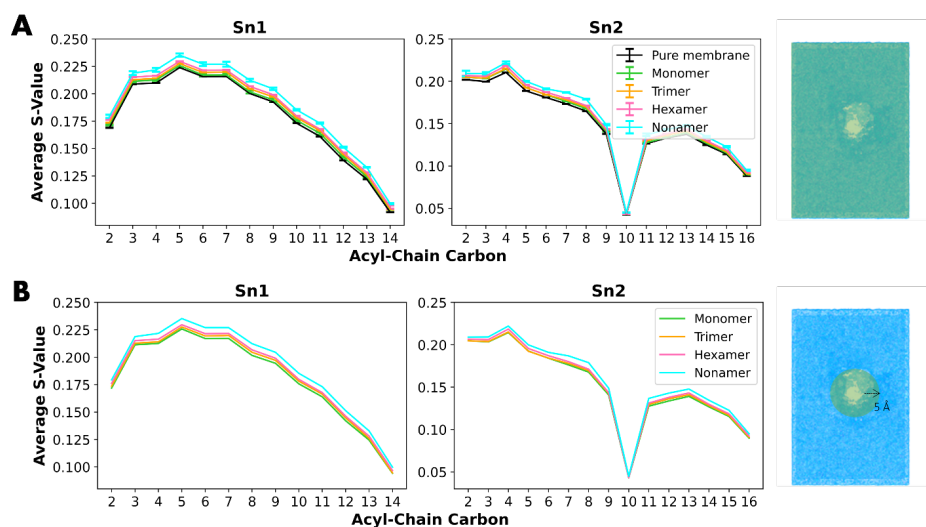


Figure S4.12: Lipid order parameters for the lipid tails in the AA systems. (A) Lipid order values calculated for all membrane lipids in all membrane systems. Error bars represent the 95 % confidence intervals and were calculated with bootstrap resampling. (B) Lipid order values for lipids with atoms within 5 Å of a peptide. Order values for saturated (on the left) and unsaturated (on the right) tails are shown separately. The images to the right depict the lipid regions considered in each case, as shaded top views of Figure II9.

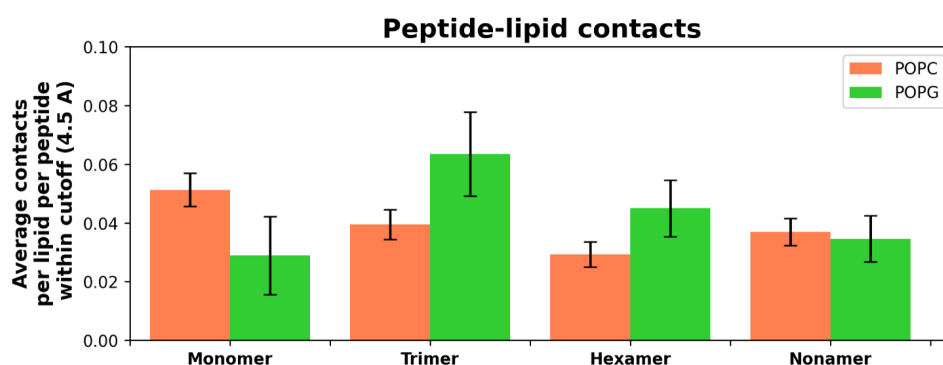


Figure S4.13: Bar plot of average number of contacts formed between the PIFP and the membrane lipids (POPC and POPG), in the simulations performed in POPC:POPG (4:1), normalized by the number of lipids and peptides. A contact was defined when the minimum peptide-lipid distance was lower or equal to 4.5 Å. Error bars represent the 95 % confidence intervals and were calculated with bootstrap resampling.

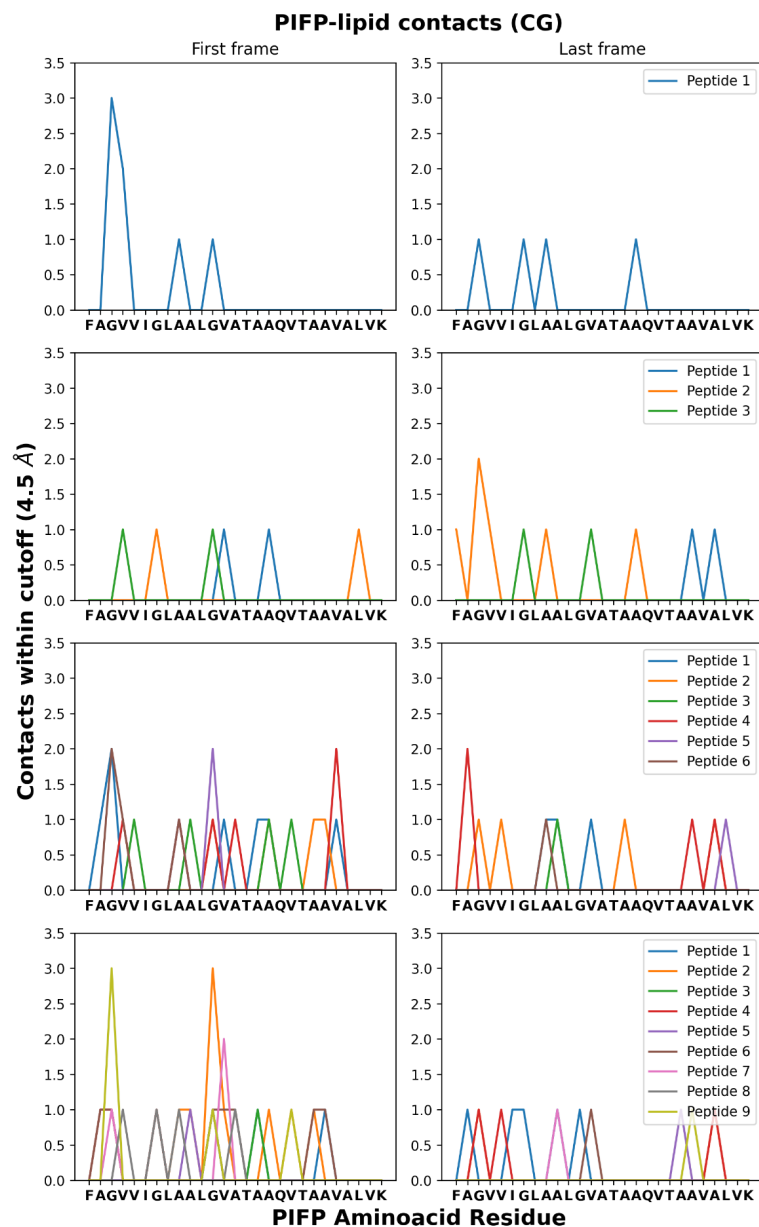


Figure S4.14: Average number of contacts formed by each residue of the PIFPs peptides and the membrane lipids, within a 6.0 Å radius, in the first and last frames of the CG simulations. Each row corresponds to a given system: Monomeric (1st row), Trimeric (2nd row), Hexameric (3rd row), and Nonameric (4th row).

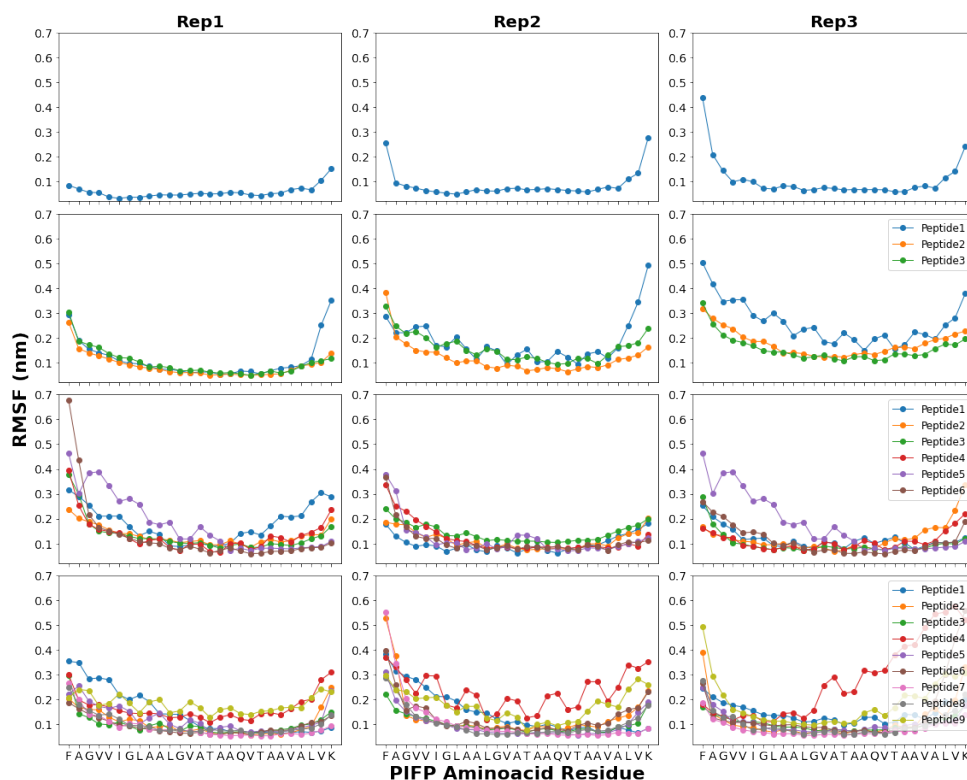


Figure S4.15: Residue root mean square fluctuation in AA MD simulations. The root mean square fluctuation (RMSF) was calculated with the GROMACS rmsf tool. Only the last 500ns of the simulation were considered for the calculation. Each row corresponds to a given system: Monomeric (1st row), Trimeric (2nd row), Hexameric (3rd row), and Nonameric (4th row). Each column corresponds to an independent replicate simulation performed in the same conditions.

SUBSTITUTING THE PARAINFLUENZA FUSION
PEPTIDE N-TERMINAL RESIDUE IMPACTS ITS
FUSOGENICITY

This chapter contains unpublished data

Authors: Mariana Valério, Carolina C. Buga, Miguel A.R.B. Castanho, Manuel N. Melo, Cláudio M. Soares, Ana Salomé Veiga, and Diana Lousa

Author contributions: M.V., M.N.M., C.M.S., and D.L. designed the computational systems and analysis and M.V. performed them. C.C.B., M.A.R.B.C., and A.S.V. designed the experimental assays and C.C.B. performed the experiments.

5.1 Abstract

Paramyxoviruses are enveloped viruses that carry a negative-sense RNA genome and require the host cell machinery to replicate. In the case of the parainfluenza virus, the process of cell entry begins with the recognition and attachment to target receptors, followed by proteolytic cleavage of the fusion glycoprotein F. This cleavage exposes the fusion peptide (FP) region, which is responsible for binding to the target membrane and initiating the fusion process.

Previously, molecular dynamics (MD) simulations performed by us have shown that the parainfluenza fusion peptides (PIFPs) spontaneously aggregate in a model membrane, in a concentration-dependent manner, to form a pore-like structure. With the aim of understanding which PIFP residues are most relevant for their insertion and aggregation in the membrane, we investigated the effect of specific substitutions on the properties of the peptide-induced pore-like structure. This analysis combined experimental biophysical data with coarse-grain (CG) and all-atom (AA) MD simulations. We selected the substitutions, Q120A and F103A, based on previous research findings highlighting their significance in maintaining peptide-peptide interactions within the membrane context [168] and in the PIFP capability to induce lipid tail protrusion events [203], respectively.

We observed a substantial impact of the phenylalanine substitution on peptide-induced lipid mixing and membrane leakage, when compared to the wild-type peptide. The Q120A substitution appears to have minimal impact on lipid mixing, while exhibiting a moderate reduction in membrane leakage. Our findings provide a comprehensive understanding of how these specific substitutions impact the properties and behavior of the PIFP: the N-terminal phenylalanine plays a central role in membrane insertion and fusion, and the glutamine residue is crucial for controlling peptide oligomerization and pore formation.

5.2 Introduction

Parainfluenza viruses (PIV) are members of the Paramyxoviridae family, which also encompasses other pathogens affecting both humans and animals, such as Hendra and Nipah viruses. PIV poses a significant global disease burden, especially for children under 5 years old [294]. These viruses cause numerous hospitalizations and deaths worldwide, yet specific treatments remain elusive.

The fusion peptide (FP) of parainfluenza virus 5 (PIV5) plays a pivotal role in the viral infection process, a characteristic shared with related viruses such as influenza, SARS-CoV-2 and Dengue. These enveloped viruses, which contain a negative-sense RNA genome, rely on the fusion between their viral envelope and the host plasma membrane during the early stages of infection to insert the viral genome into the host cell. This fusion process is facilitated by viral fusion proteins, and in the case of PIV, it involves the interaction between the receptor-binding glycoprotein (HN) and the fusion glycoprotein. Unlike the influenza virus, PIV fuse their membrane with the host membrane

at the cell surface. In paramyxoviruses, the fusion between the viral and the host cell is orchestrated by the coordinated action of the receptor binding protein hemagglutinin-neuraminidase and the fusion protein (F protein), which are also privileged therapeutic targets. The Paramyxovirus F protein, initially synthesized as a precursor (F_0), undergoes cleavage to form a pre-fusion F trimer, where the hydrophobic fusion peptide region is initially concealed. Upon contact with the host cell, the F protein undergoes structural changes, exposing the hydrophobic fusion peptide that inserts into the host membrane. Subsequently, the F protein refolds into a stable postfusion structure, driving viral and cell membrane fusion and releasing the viral genetic material into the target cell [27, 127, 165, 292, 340–343]. The concealed PIV fusion peptide (Figure 5.1A) extends its N-terminal helix into the membrane upon activation and insertion, contributing to the fusion process.

Previous work by our team has shown that when present at high concentrations in an anionic membrane, the PIFP can promote fusion and/or membrane leakage through the formation of a water-permeable porelike structure. This structure promotes lipid head intrusion and lipid tail protrusion, which facilitates membrane fusion [203].

Building upon these findings, the current study aims to determine the role of two substituted residues on the PIFP's oligomerization and, membrane perturbation ability. We based the choice of substitutions on previous data that describes the relevance of Q120 residue for the maintenance of peptide-peptide interactions when inside of the membrane [168], and also on our own previous work showing that the N-terminal residue of this peptide - F103 - induces lipid tail protrusion [203], a feature that has been shown to precede and facilitate lipid mixing events [172]. To test the importance of these residues we substituted them to alanine, resulting in the following substitutions: Q120A and F103A.

It is important to emphasize that, after cleavage of the F protein, residue F103 becomes the N-terminal residue of the PIFP (as seen in figure 5.1). In simpler terms, the numbering for F103 reflects its position in the F_0 protein sequence before cleavage, which subsequently transforms into its active F_1 form, in which F103 becomes the N-terminal residue.

In this study we succeeded in conducting a comprehensive analysis of the effect of these substitutions on the PIFP properties, by combining experimental biophysical data with results from coarse-grain and all-atom molecular dynamics (MD) simulations. Our research elucidated a notable influence of the N-terminal phenylalanine substitution on both lipid mixing and membrane leakage induced by the peptide. In contrast, the Q120A substitution had a more limited impact on leakage and no significant effect on lipid mixing but affected the number and type of interactions that are formed between peptide copies inside the membrane in MD simulations. Overall, this study allows for the understanding of the consequences of these substitutions on PIFP function in the context of the host membrane, shedding light into the specific roles of these residues in membrane fusion.

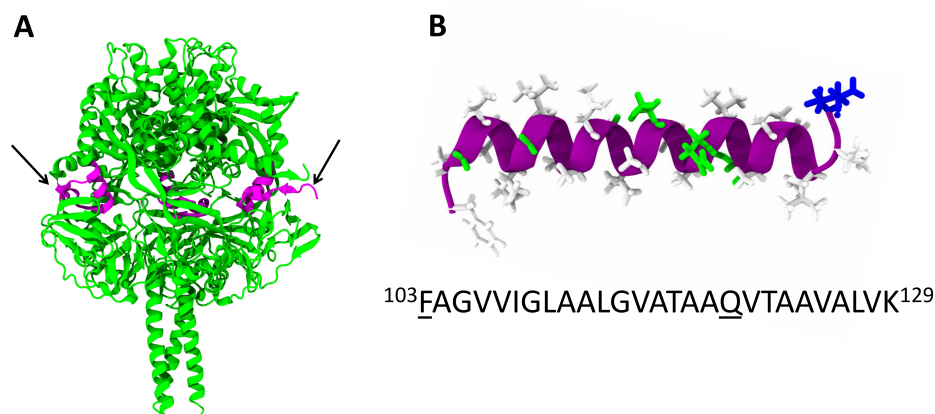


Figure 5.1: Parainfluenza fusion peptide in F protein context and after cleavage. (A) Cartoon representation of the crystal structure of the cleaved prefusion F protein trimer, 4GIP, viewed from the side, with the FP highlighted colored in pink. Arrows indicate the protease cleavage sites. (B) Molecular image of the PIFP with the backbone in purple and the residue side-chains colored according to their biophysical properties: nonpolar (white), basic (blue), and polar residues (green) and below the PIFP amino acid sequence with the F103 and Q120 substitution sites highlighted.

5.3 Methods

5.3.1 Instrumentation

Liposomes were extruded through 50-nm and 100-nm pore Nucleopore polycarbonate membranes (Whatman/GE Healthcare, UK) using a LiposoFast-Basic plus Stabilizer setup from Avestin (Mannheim, Germany) with Hamilton (Bonaduz, Switzerland) syringes. Dynamic light scattering (DLS) measurements were carried out in a Malvern Instruments ZetasizerNano ZS (Worcestershire, UK). Fluorescence measurements were performed in a Varian Cary Eclipse spectrofluorometer (Crawley, UK) and in a FLS920 series Edinburgh Instruments spectrofluorometer (Livingston, UK). For the surface plasmon resonance (SPR) experiments, a GEHealthcare Biacore X100 (Little Chalfont, UK) was used.

5.3.2 Lipid Vesicles Preparation

Small unilamellar vesicles (SUVs) and large unilamellar vesicles (LUVs) were used as membrane model systems. The proper amount of lipids were solubilized in chloroform in a round-bottom flask and a gentle nitrogen flow was used to evaporate the organic solvent. The newly formed lipidic film was kept under vacuum overnight. The rehydration of the lipidic film was performed with 10mM HEPES pH 7.4, 150mM NaCl buffer followed by ten freeze/thaw cycles forming a suspension of multilamellar vesicles. SUVs and LUVs

were achieved by extrusion of the multilamellar vesicles through polycarbonate filters with pore size of 50-nm and 100-nm, respectively.

5.3.3 PIFPs-induced lipid mixing

Lipid mixing induced by the peptides was evaluated by a FRET approach. POPC:POPS (4:1) LUVs labeled with both 1% N-NBD-PE (donor) and 1% N-Rh-PE (acceptor) and unlabeled LUVs were mixed in a 1:4 ratio in buffer 10 mM HEPES pH 7.4, 150 mM NaCl. Titrations of increasing concentrations from 2.5 up to 15 μ M of each peptide were performed in a 0.77 mM LUV sample. Fluorescence emission spectra were collected between 500 and 650 nm (λ_{ex} =470nm) after a 15 min incubation at 37°C. For each peptide, triplicates were performed and the lipid mixing efficiency (R) was determined:

$$R = \frac{I_{F530nm}}{I_{F588nm}} \quad (5.1)$$

Where I_F 530 nm is the fluorescence emission intensity recorded at 530 nm and I_F 588 nm is the fluorescence emission intensity recorded at 588 nm. Then, fusion (%) was determined using the following equation:

$$Lipidmixing(\%) = \frac{R_S - R_0}{R_{max} - R_0} \times 100 \quad (5.2)$$

Where R_S is the R of the sample, R_0 is the R of the negative control corresponding to the vesicles in the absence of peptide and R_{max} is the R for the positive control corresponding to the vesicles in the presence of 1% (v/v) Triton X-100.

5.3.4 PIFPs-induced Vesicle Leakage

The content leakage of POPC:POPS (4:1) vesicles induced by the peptides was determined by 5,6-carboxyfluorescein (5,6-CF) dequenching using LUVs loaded with 50 mM 5,6-CF as described elsewhere [312]. Briefly, peptide concentrations ranging from 0.9375 to 30 μ M were incubated with 0.77 mM LUVs in buffer 10 mM HEPES pH 7.4, 150 mM NaCl for 15 min at 37°C. Fluorescence emission intensity (λ_{ex} = 493 nm and λ_{emi} =520 nm) was monitored and the leakage percentage was determined as described previously [312].

5.3.5 Surface Plasmon Resonance

PIFP-Q120A and PIFP-F103A affinity toward POPC:POPS (4:1) vesicles was evaluated by SPR experiments. Each L1 sensor chip was rinsed with three injections of 20 mM CHAPS before starting an experiment. The lipid surface was prepared with 1 mM POPC:POPS (4:1) SUV sample injected at a 2 μ L/min flow rate for 2400 s. Then, a 36 s injection of 10 mM NaOH at 50 μ L/min was performed to remove loosely bound vesicles. Peptide samples were prepared in 10 mM HEPES pH 7.4, 150 mM NaCl in concentrations ranging from 2.5 to 50 μ M and injected over the lipid surface at a 5 μ L/min flow speed during 500

s (association time). For each sample a 1000 s dissociation time was allowed. After each run, the L1 sensor chip was regenerated with sequential injections of 20 mM CHAPS (5 $\mu\text{L}/\text{min}$ for 60 s), 0.5 % (v/v) SDS (5 $\mu\text{L}/\text{min}$ for 60 s), 10 mM NaOH with 20 % (v/v) methanol (50 $\mu\text{L}/\text{min}$ for 36 s) and 10 mM NaOH (50 $\mu\text{L}/\text{min}$ for 36 s). After correcting the response values for each peptide molecular weight, the partition constants (K_P) were calculated using the following equation:

$$\frac{RU_S}{RU_L} = \frac{\gamma_L K_P \frac{M_S}{M_L} [S]_W}{1 + \sigma \gamma_L K_P [S]_W} \quad (5.3)$$

where RU_S is the total sample response, RU_L corresponds to the total lipid deposition response, γ_L is the lipid molar volume, M_S represents the peptide molecular weight, M_L is the lipid vesicles molecular weight, $[S]_W$ is the sample concentration and σ corresponds to the lipid to solute ratio at the saturation point.

5.3.6 Dynamic Light Scattering

Lipid vesicle aggregation induced by the peptides was monitored by DLS. Titrations of increasing concentrations from 0.9375 up to 30 μM of each peptide were performed in a 0.77 mM POPC:POPS (4:1) LUV sample in buffer 10 mM HEPES pH 7.4, 150 mM NaCl. After each peptide addition, an incubation of 5 min at 37°C was performed. Measurements comprised three individual runs, each corresponding to an average autocorrelation curve obtained from at least ten replicate sample scans. The CONTIN-based method [313] was applied to determine the diffusion coefficient (D) values of each sample. D values were then used to calculate the particles Z-average hydrodynamic diameter and derived count rate through the Stokes-Einstein-Sutherland equation [314].

5.3.7 Coarse-Grain MD Simulations of the PIFP in a POPC:POPS membrane

Solid-state NMR studies of PIFP revealed that the peptide adopts a fully α -helical conformation inside POPC:POPG membranes [149]. Based on these data, a fully α -helical PIFP was created using the *remodeling* tool from Rosetta [315] and converted to a MARTINI coarse grained topology [201, 316] with the martinize 2.0 program [187]. This program was also used to generate the topology and structure files based on the atomistic pdb file.

Triplicate coarse-grained (CG) simulations of the wt, Q120A, and F103A PIFPs were performed in a POPC:POPS membrane (at 80:20 molar ratio). Aiming to examine their behavior at high concentrations, the systems had one peptide per 100 lipids, equivalent to $\approx 7.5 \mu\text{M}$ in the experimental setup. The N-terminal residue of the monomers was protonated and the C termini's amidated to reproduce the state of peptides in the context of the F protein and the experimental setup. The axes of the α -helices in each dimer were oriented parallel to each other and separated by at least 3 nm.

All simulations were performed with the GROMACS 2020.3 package [220], using the Martini 3 force field [187] and lipid topologies released with it. The system was built using

the INSANE script, in a 10 x 15 x 14 nm box. The total charge of the systems was first neutralized with either Na⁺ or Cl⁻ ions, and then an additional 140 mM NaCl was added.

NMR data on the position of the wt PIFP in a POPC:POPG membrane shows that they should stay vertically inside the membrane [149]. Preliminary tests showed that, contrary to the experimental data, when using the Martini 3 protein model the PIFP's were leaving the membrane. To prevent this from happening we scaled down the backbone-water interactions of the peptide as suggested in [344].

The systems were energy-minimized using the steepest-descent method for 5000 steps. In this step the peptide's backbone atoms were position-restrained using a force constant of 500 kJ mol⁻¹nm⁻². Before performing production runs, an equilibration step was carried out for 20 ns. The simulations were run in the isothermal-isobaric (NPT) ensemble, coupling the system to a temperature and pressure bath using the V-rescale and Parrinello-Rahman [224] schemes, respectively (the Berendsen barostat was employed during equilibration). The target temperature was set to 310 K and semi-isotropic pressure coupling also set to 1 atm using a compressibility of 3 x 10⁻⁴ bar⁻¹. Reaction-field electrostatics were employed, and Lennard-Jones and Coulombic interactions were both cut-off at 1.1 nm[317]. Each system was run for 30 μ s.

5.3.8 Atomistic MD Simulations of the PIFP in a POPC/POPS

Atomistic MD simulations of the wt, Q120A and F103A fusion peptides were performed using the last frame of the CG simulations as the starting structure. The systems were converted from coarse-grain to atomistic detail using the backward.py script [318]. All atomistic simulations were performed with the GROMACS 2020.3 [220] package and using the Amber14sb forcefield [185], alongside the TIP3P water model [345]. The minimization and equilibration protocols used default parameters generated by CHARMM-GUI for GROMACS [319–321]. In short, after 5000 steps of minimization using the steepest-descent algorithm, the system temperature was maintained at 310 K by coupling to a Nosé-Hoover thermostat [322] with a 1 ps coupling constant. The system was then coupled semi-isotropically to a Parrinello-Rahman barostat [224] and further equilibrated for 125 ps using 1 ps timestep, and 500 ps with a 1 ps timestep. During the equilibration, atomic positions restraints on the protein heavy atoms were gradually relaxed in six steps from 4000 to 2000 kJ nm⁻²mol⁻¹ and from 500 to 200 kJ nm⁻²mol⁻¹ the protein backbone and side-chains. The lipid's phosphate atoms were also gradually relaxed from 1000 to 40 kJ nm⁻²mol⁻¹. After the sixth step, the simulations were performed without any restraints. Three replicates were simulated for each system with all simulations being run for 1.5 μ s.

5.4 Results

5.4.1 F103A and Q120A substitutions do not affect the PIFP ability to induce vesicle aggregation in a concentration-dependent manner

This study builds on previous research that highlighted the importance of the Q120 residue in maintaining peptide-peptide interactions between the PIFPs when inside a membrane [168], and our work showing the ability of the N-terminal residue, F103, to induce lipid tail protrusion — an intermediate step of lipid mixing events [172, 203]. Despite the available knowledge, a comprehensive understanding of the roles of these residues in peptide-peptide and peptide-membrane interactions remains elusive. To bridge this knowledge gap we studied three PIFPs: the wt and two others with the following point-substitutions Q120A and F103A. These peptides comprehend residues 103-129 of the PIV5 fusion protein and on the C-terminus they are tethered to an 8-amino-3,6-dioxaoctanoic acid that bridges a polylysine tail.

The aggregation/oligomerization of viral fusion peptides on the cell membrane surface is a shared feature among diverse viruses, including PIV. In the PIFP case, a concentration-dependent accumulation of the peptide on the surface of lipid vesicles and a consequent triggering of vesicle aggregation was previously reported [203]. To investigate this effect in the presence of the Q120A and F103A substitutions, DLS was employed to characterize the macromolecular organisation of the structures formed by the peptides alone and peptide vesicle mixtures. The measurement of the hydrodynamic diameter (D_H) as well as the particle count rate upon consecutive addition of the peptides to a lipid vesicle solution allowed us to obtain insights regarding the aggregation process.

For the PIFP-Q120A, there are no observed aggregates in the aqueous solution (Figure 5.2A). However, when LUVs are introduced, an increase in Q120A concentration from 3.75 to 7.5 μM results in a significant 24-fold increase in the LUVs D_H and a 2-fold decrease in the particle count. Thus, the PIFP-Q120A presents a critical aggregation concentration (CAC) between 3.75 and 7.5 μM , similar to what was seen previously for the wt PIFP [203].

Regarding the F103A substitution, a similar pattern is observed, where the peptide alone does not form aggregates (Figure 5.2B). As the concentration of PIFP-F103A increases from 3.75 to 7.5 μM , there is an increase in the LUVs D_H along with a decrease in the particle count. However, the extent of the PIFP-F103A effect on the LUVs differs from the previous substitution. For this one only a 5-fold increase in LUVs D_H and 1.7-fold reduction of particle counts were observed.

These results suggest that, similarly to the wt PIFP, upon reaching a CAC, between 3.75 and 7.5 μM , the Q120A and F103A PIFPs trigger the aggregation of lipid vesicles, increasing the particle size detected in solution.

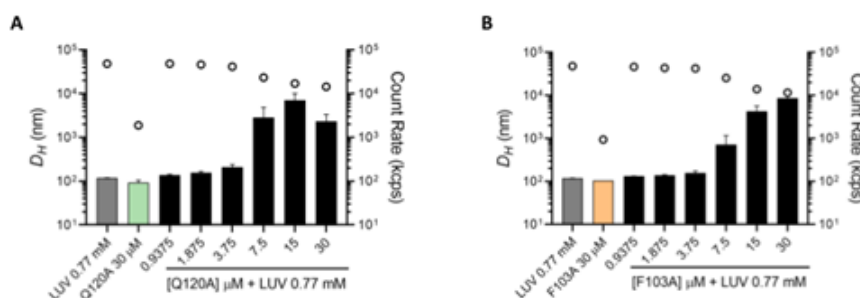


Figure 5.2: PIFPs and PIFPs-vesicles aggregation profiles. Monitorization with DLS of Z-average D_H (columns) and count rate (circles) of particles formed during (A) Q120A and (B) F103A titration in a 0.77 mM POPC:POPS (4:1) LUVs solution. Particles formed in solutions containing only 0.77mM LUVs or PIFPs 30 μ M were characterized as controls.

5.4.2 PIFP-Q120A has a higher partition coefficient than PIFP-F103A

Having established the ability of the Q120A and F103A to induce lipid vesicle aggregation, it became essential to characterize the lipid binding properties of these peptides. To assess peptide-lipid binding, we utilized surface plasmon resonance (SPR), a technique that offers real-time detection of bound peptides on a surface coated with small unilamellar vesicles. Each peptide partition coefficient was estimated by plotting peptide-to-lipid response ratios as a function of the injected peptide concentration, as described previously [328]. Although both peptides displayed an increase in lipid binding responses in a concentration-dependent manner, different partition coefficient (K_P) values were obtained (figure ??). PIFP-Q120A exhibited a K_P value of approximately 9.8×10^3 , which exceeds the K_P value reported for the wt PIFP ($K_P \approx 3.2 \times 10^3$) [203]. On the contrary, PIFP-F103A revealed a lower K_P value ($\approx 1.8 \times 10^3$) than the wt PIFP.

The analysis of peptide-lipid binding reveals that the Q120A substitution does not compromise the peptide's ability to bind to the membrane, and actually even increases the peptide's partition coefficient, possibly due to the higher hydrophobic nature of the peptide after the Q120A substitution. Conversely, the F103A substitution substantially reduces the peptide's affinity for the lipid membrane, probably due to the importance of the N-terminal phenylalanine for the insertion of the peptide into the membrane bilayer.

5.4.3 The N-terminal phenylalanine residue is critical for the PIFP's ability to induce lipid mixing and membrane leakage

To explore the effect of the selected substitutions on the PIFP fusogenic properties, we examined Q120A and F103A's ability to induce both lipid mixing and vesicle leakage. Lipid mixing was monitored by peptide-induced FRET efficiency variations in a mixture composed of NDB-RhB-labelled and unlabelled vesicles while vesicle leakage was evaluated by monitoring the peptide-induced leakage of 5,6-carboxyfluorescein (5,6-CF)-loaded vesicles.

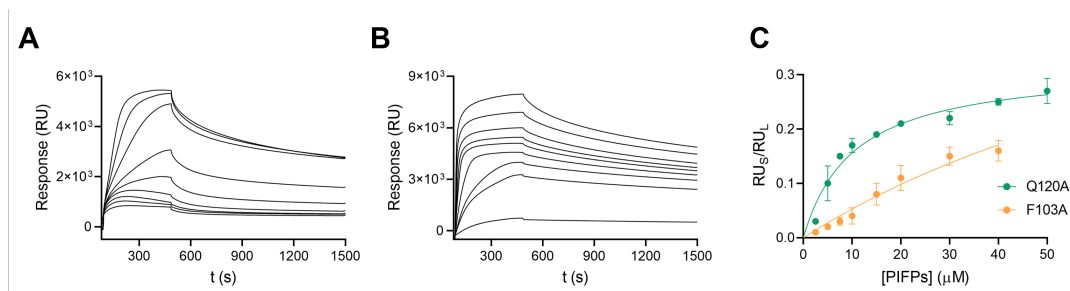


Figure 5.3: Characterization of PIFPs-lipid membrane (un)binding in vitro. Surface plasmon resonance (SPR) sensorgrams of increasing (A) Q120A and (B) F103A concentrations (2.5, 5, 7.5, 10, 15, 20, 30 and 40 μM) interacting with POPC:POPS (4:1) vesicles. (C) K_P values determined by representing the individual values of RU_S and RU_L recorded for each PIFP sensorgram at 500s as the function of the corresponding concentration. Data were fitted with equation 5.3.

Interestingly, Q120A and F103A substitutions revealed different effects on lipid mixing and vesicle leakage. On one hand, F103A substitution impaired the peptide-induced lipid mixing (Figure 5.4A) and nearly abolished the membrane leakage effect (Figure 5.4B). Meanwhile, the Q120A substitution resulted in a significant reduction in the FP-induced membrane leakage in all the peptide concentrations tested but no significant effect on the PIFP-induced lipid mixing (Figure 5.4A).

In summary, these findings suggest that substituting the N-terminal phenylalanine has a significant impact on the membrane perturbation properties of the PIFP, while the substitution of the glutamine residue only affects the ability of the peptide to induce membrane leakage. From these results, it seems clear that both lipid mixing and leakage ability depend on two key factors: the peptide concentration within the membrane and the influence of peptide-peptide interactions on membrane disruption. The lower partition coefficient of the F103A mutant contributes to the decrease in lipid mixing and leakage ability, suggesting that this residue is crucial for membrane fusion.

On the other hand, the Q120A substitution, which leads to a different profile of induced vesicle leakage, is indicative that this glutamine residue may be involved in regulating the network of interpeptide interactions and/or peptide-lipid interactions required for the flow of content through the pore-like structure, similarly to what was seen for the wt PIFP. In addition, the reduction in vesicle leakage can also be attributed to the substitution of the polar glutamine with a more hydrophobic residue that reduces the propensity of water molecules to pass through the membrane. However, at this stage, the validation of this hypothesis *in vitro* faces significant challenges. Therefore, we followed with an *in silico* approach to delve deeper into this matter.

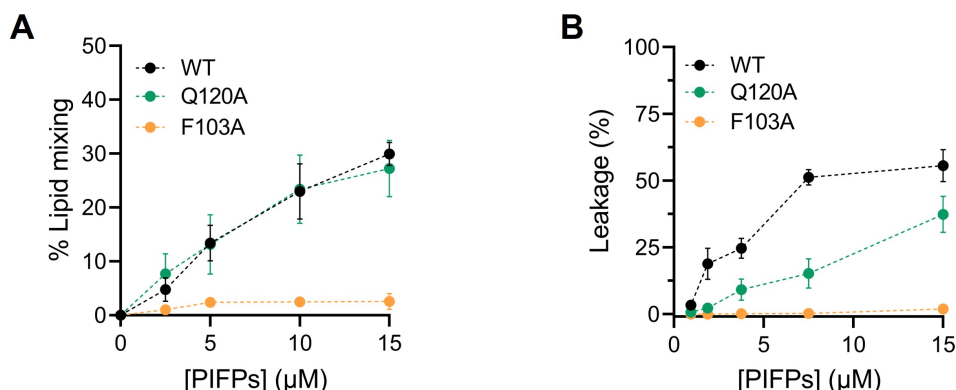


Figure 5.4: Characterization of PIFPs fusogenic properties *in vitro*. (A) Percentage of PIFPs-induced lipid mixing in POPC:POPS (4:1) vesicles, extrapolated from equation 5.1, at increasing peptide concentration. (Black dots, PIFP-wt; green dots, PIFP-Q120A; orange dots, PIFP-F103A). (B) Percentage of PIFPs-induced membrane leakage. Leakage of 5,6-CF from 0.77mM LUVs upon incubation with increasing concentrations up to 15 μM of (black) PIFP-wt (green) PIFP-Q120A and (orange) PIFP-F103A.

5.4.4 MD simulations reveal that F103A and Q120A assemble into oligomeric structures stabilized by distinct interaction networks

To compare and validate our experimental results, we conducted triplicate MD simulations of POPC:POPS systems with high concentrations of these PIFPs, with one peptide per 100 lipids, to a total of six peptides per simulation box. The N-terminal residue of the monomers was protonated and the C termini's amidated to reproduce the state of peptides in the context of the F protein and the experimental setup.

Simulations were started with the peptides already inserted in the membrane in a transmembrane configuration. The WT peptide has been confirmed to insert in such a way, via NMR [149]. Although the two mutations tested could impact the insertion of the peptide, we designed the simulations assuming that the peptides would be inserted in the membrane. Thus, all the results obtained are based on this assumption. Despite the decrease in partition, the F103A peptide continues to have affinity to the membrane and maintain its hydrophobic character. And so it is reasonable to assume that it can adopt a “membrane-spanning configuration”. The Q120A substitution should even facilitate transmembrane insertion, as a polar residue is being swapped by a hydrophobic one placed in the middle of the helix, which will reside close to the membrane core.

During the CG simulations, all peptides aggregate inside the membrane (as seen in Figures 5.5D, E and F) and this was the initial conformation used in the atomistic MD simulations (Figures 5.5G, H and I).

Upon visual inspection of the systems after 1.5 μs of simulation at atomistic detail, it is evident that all peptides remain embedded within the membrane, and the PIFP

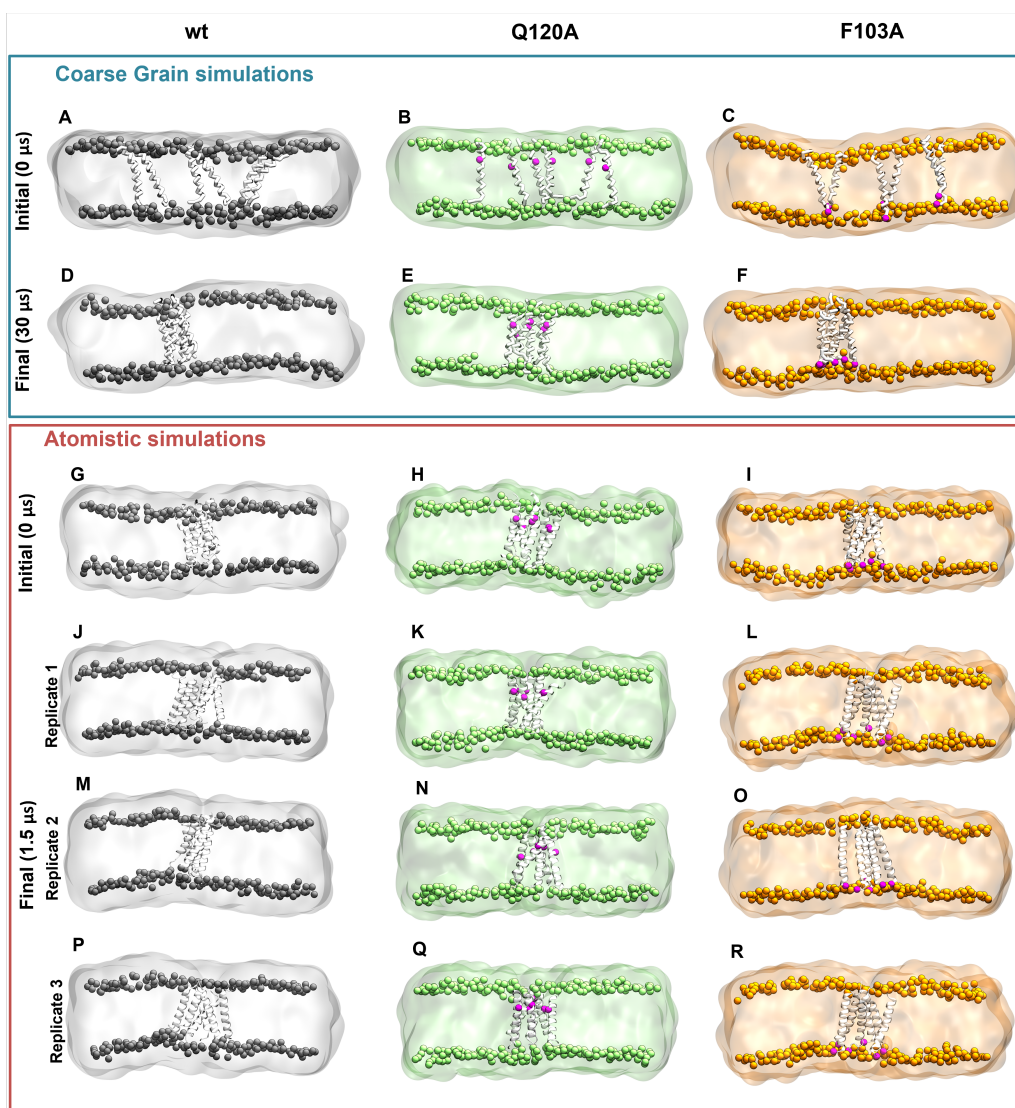


Figure 5.5: Snapshots of the initial and final frames of the CG and atomistic MD simulations. Initially, six PIFPs were inserted into a POPC:POPS (4:1) membrane (A, B and C). The three systems were then simulated in CG detail for 30 μ s, with peptides aggregating inside of the membrane (D, E and F). The last frame of the CG simulation was converted to atomistic detail recurring to the backward tool [318] (G, H and I). Three independent replicates of each system were then simulated for 1,5 μ s (wt: J, M and P; Q120A: K, N and Q; F103A: L, O and R). Lipids belonging to the wt, Q120A and F103A systems are shown in gray, green and orange, respectively. Substitution sites Q120A and F103A are shown in magenta.

aggregates persist, as seen in the lower section of figure 5.5, supporting the assumption that the peptide is stably inserted in the membrane. . However, this initial analysis, while informative, does not provide a comprehensive understanding of the true influence of the substitutions on the aggregated structures. To achieve this, we started by analyzing the impact of the single substitutions on the formation of interpeptide interactions. The

RIP-MD tool was used [238], to which 250 frames of each replicate were supplied. All hydrogen bonds and pi-stacking interactions were calculated, and the residue interaction networks (RINs) for the peptide-peptide interactions plotted using Cytoscape [239] (Figure 5.6). To facilitate the analysis of the RINs we assigned numbers from 1 to 6 to each of the six peptides in the simulation box. These numbers remained consistent throughout the simulation and between the wt, Q120A, and F103A systems.

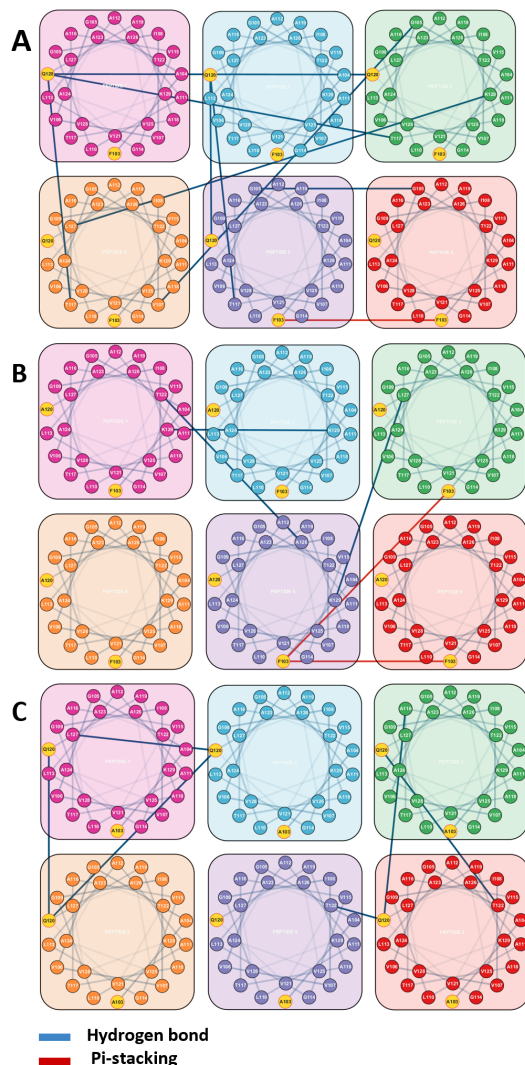


Figure 5.6: Residue interaction networks (RINs) for the peptide-peptide interactions of the (A) wt, (B) Q120A and (C) F103A systems. The RINs were determined using RIP-MD. Hydrogen bonds are shown in blue and pi-pi interactions are shown in red.

In the wt system, we observed the presence of eight persistent hydrogen bonds, found in more than 30% of the frames analyzed. The majority of these hydrogen bonds were mediated by Q120 residues, with the most frequent occurring between two glutamine residues in peptides 1 and 3; and 2 and 5. Additionally, there were interactions observed

between Q120 and T117, occurring in three separate instances, one between peptides 1 and 4, one between peptides 1 and 3, and another one between peptides 2 and 5. Apart from the interactions mediated by glutamine or threonine residues, we also detected some backbone interactions. These interactions involved G105 that interacts with itself or with valine 107. Besides the hydrogen bonds, we also observed a less persistent pi-stacking interaction, present in less than 10% of the simulation frames analyzed. This interaction involved the N-terminal phenylalanine residues of peptides 5 and 6 (Figure 5.6A).

When substituting the Q120 amino acid residue for alanine, a noticeable decrease in the overall number of interactions was observed, from eight to three. Interestingly, there was a complete abolition of interactions mediated by the substituted A120 residue, as depicted in Figure 5.6B. New hydrogen bonds emerged, involving residues K129, T122, A126, and L127, with a persistence higher than 30% of the analyzed frames. Specifically, we observed the formation of the following inter-peptide hydrogen bonds: T122-A126 between peptides 1 and 5, K129-K129 between peptides 1 and 2, and L127-K129 between peptides 3 and 5. Regarding the pi-stacking interactions, a new interaction between two F103 residues appeared between peptides 3 and 5, although its persistence remained low, at less than 5%.

Finally, in the system where the F103 amino acid residue was substituted with an alanine residue, the pi-stacking interactions previously mediated by this residue ceased to exist. Interestingly, the number of persistent inter-peptide hydrogen bond interactions increased compared to the Q120A system, going from three to six. However, this number decreased compared to the wt system, where we saw eight hydrogen bond interactions. In this system the Q120 amino acid residue played a crucial role in maintaining peptide-peptide interactions within the peptides, by mediating all interactions observed. These include the following pairs: Q120A-Q120A between peptides 1 and 2, and peptides 2 and 4; Q120A-T122 between peptides 3 and 6, and peptides 5 and 6; Q120A-L127 between peptides 1 and 2; and Q120A-A116 between peptides 3 and 6.

In conclusion, the introduction of both these substitutions had distinct impacts on the peptide's ability to aggregate inside the membrane. The Q120A substitution led to a significant reduction in overall interactions, including the complete abolishment of interactions mediated by the substituted A120 residue, and while new hydrogen bonds formed, they remained fewer than in the wt system, indicating the crucial role of Q120 in maintaining peptide-peptide interactions and oligomerization inside the membrane. On the other hand, the F103A substitution eliminated pi-stacking interactions but increased hydrogen bond interactions compared to the Q120A system. However, these remained fewer than in the wt system, with Q120 still playing a pivotal role in mediating interactions. These results underline the importance of specific amino acid residues, particularly Q120, in stabilizing and modulating the interactions of these peptides within the membrane, allowing for their aggregation.

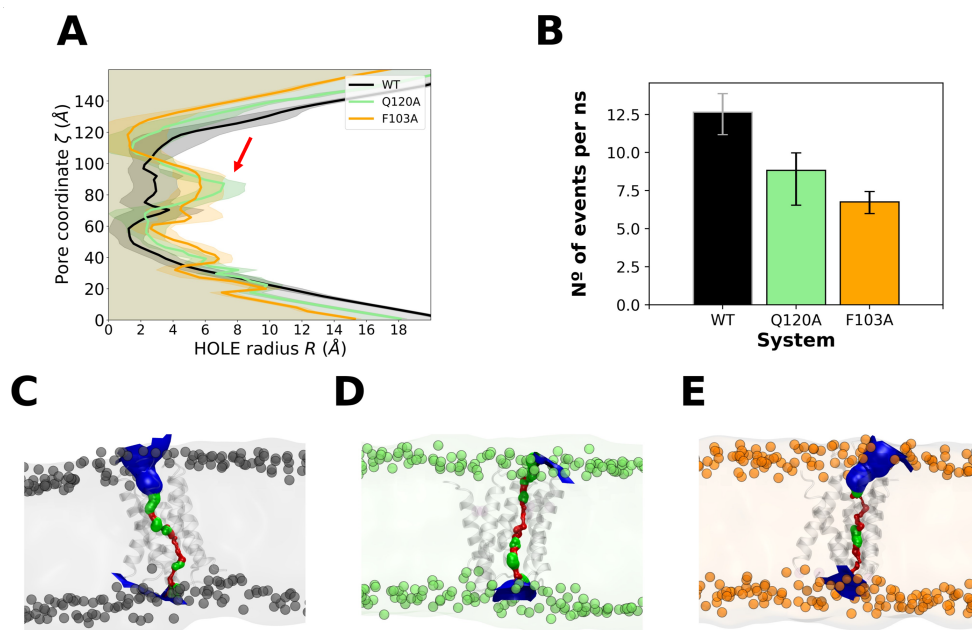


Figure 5.7: PIFP pore characterization. Graph (A) shows the radius of the pores for system wt (in black), Q120A (in green) and F103A (in orange). The pore radius was determined using the spherical probe of the HOLE algorithm [346, 347]. Below the curves, the average space available for water passage is highlighted. Panel (B) displays a snapshot from the wt, Q120A and F103A simulations, clearly illustrating the space available in the membrane for the passage of water.

5.4.5 Q120A and F103A substitution lead to decrease in water flux through the membrane

The aggregation behavior of the PIFPs in the membrane, seen here, was also shown by us in a previous study [203]. Our earlier observations strongly suggested that peptide aggregation inside the membrane played a pivotal role in promoting the formation of water-permeable pore-like structures. To determine whether these substituted peptides would also allow for the flow of water through them, we looked at the available space inside the membrane and calculated how many water molecules could pass through it per time unit (ns). To do this, we examined the last 1 μ s of simulation by using the HOLE method [346, 347].

Knowing that a radius of at least 1.15 Å is necessary to accommodate a water molecule, our results show that, on average, all the systems studied form channels sufficiently large to allow water molecules along their whole length (Figures 5.7A, C, D and E). This spacing is even wider in the region close to the substitution site Q120A, as shown in figure 5.7A by the red arrow. This observation is also in line with what we show in Figure 5.6B, that when substituting residue Q120 to alanine, a notable decrease in the overall number of interactions was observed, and no new interactions are formed by the substituted A120 residue.

Looking at the average number of waters crossing the membrane per time unit, we see that the flux (measured in molecules/ns) through the F103A pore is the highest, followed by the Q120A system, and finally the wt system ($\text{Flux}_{\text{wtpore}} = 12.65 \pm 2.94 > \text{Flux}_{\text{Q120pore}} = 8.83 \pm 4.58 > \text{Flux}_{\text{F103Apore}} = 6.75 \pm 1.53$).

Overall, the wt system exhibits the highest water flux through the pore, followed by the Q120A and F103A systems. The same trend was seen in the biophysical assays, where the introduction of the glutamine substitution (Q120A) results in a decrease of 5,6-CF leakage, which is even lower when substituting the N-terminal phenylalanine (Figure 5.4D). Nonetheless, in the leakage assays we saw that the substitution of the N-terminal phenylalanine led to almost abolition of leakage, and, in our computational results, we still see some flux through the pore-like structure. This discrepancy between *in silico* and *in vitro* experiments may be explained by a fundamental difference in the setup of the simulations compared to the biophysical assays. In the computational systems, the PIFPs were initially inserted into the membrane, as described by NMR structure data indicating that the peptide is transmembrane [149]. In contrast, in the experimental setup, the peptides are initially placed in solution and are incorporated into the membrane based on their propensity, i.e. the K_P . In fact, the experimental data show that the F103A peptide has a K_P for the membrane 3-fold and 5.4-fold lower than that of the wt or Q120A systems, respectively (Figure 5.3C). This may be the reason for the very low water leakage seen experimentally, i.e. since the peptides are less incorporated in the membrane, there is less leakage. Therefore, the comparison of the water flux of the wt and the substituted peptides observed in our simulations takes into account only what happens after the peptides insert into the membrane and does not account for the differences in water/lipid partition between peptides.

In sum, in both the experimental and *in silico* data the Q120A and F103A substitutions lead to a decrease in the flux of water through the membrane when compared with the wt.

5.4.6 Substitution of the N-terminal phenylalanine residue significantly affects lipid tail protrusion

The passage of water through a membrane represents an atypical event given the membrane's hydrophobic nature. Our observation that upon the addition of these peptides, there is the passage of water through them, indicates that they must significantly perturb the lipid bilayer. To evaluate this, we analysed one of the computational hallmarks of membrane perturbation - lipid tail protrusion. This phenomenon was initially identified through MD data and is characterized by the outward extension of lipid acyl chains beyond their corresponding phosphate headgroups (as depicted in figures 5.8C, D, and E).

To measure the membrane disturbance induced by the PIFPs we determined the number of membrane lipid tail protrusion events occurring in the last 500 ns of the atomistic simulations. Results show that both substitutions lead to a decrease in the

number of lipid acyl protrusion events (Figure 5.8A), with the F103A substitution leading to the lowest number of lipid tail protrusion events (Figure 5.8A). Upon analyzing the amino acid residues near the lipid protrusion sites, we observed that the A103 substitution significantly reduces the frequency of lipid-tail protrusions close to the N-terminal residue when compared with both the wt and Q120A systems (Figure 5.8B). For the Q120A system, a closer examination of the amino acid residues near the lipid protrusion sites revealed that the substitution of Q120 to an alanine leads to an abolishment of the lipid tail protrusion events induced by this residue when compared with wt (Figure 5.8B).

These results are commonly correlated with lipid mixing due to the known effect of lipid tail protrusion in facilitating fusion between opposing membranes [158–160, 174, 175, 348]. In this study, both the biophysical and *in silico* assays point to a lower lipid mixing induced by the F103 mutant, whereas in the case of the Q120A the integration of the computational and experimental results is not so straightforward. Computational results show a lower lipid tail protrusion of Q120A when compared with wt (Figure 5.7A), whereas experimental results show no significant differences in lipid mixing (Figure 5.4D). As noted before, our computational analysis disregards the partition of the peptide into the membrane and focuses only on the effect that the peptide has on the lipids once it is already inserted into the membrane. SPR data shows that the Q120A peptide has a K_P for the membrane 3-fold higher than that of the wt system. Consequently, when we take into account that the partition coefficient of the Q120A to the membrane is 3-fold higher than the one of the wt, the lipid tail protrusion and lipid mixing results become consistent.

5.5 Conclusion

Our study builds upon previous research where the PIFP is shown to induce membrane fusion by forming concentration-dependent oligomeric pore-like structures in the membrane. Herein, we extended this research by investigating the impact of specific substitutions on the PIFP ability to interact with and disturb the membrane. This was achieved by combining a range of experimental biophysical techniques with computational simulations to gain valuable insights into how specific substitutions affect PIFP properties.

We found that the Q120A substitution does not significantly impact the peptide's ability to induce vesicle fusion. This is supported by its similar lipid mixing and vesicle aggregation capabilities when compared to the wt peptide. *In silico* simulation results also align with this trend, especially when taking into account the K_P 's of the PIFPs. We hypothesize that the higher hydrophobicity of the alanine residue contributes to its enhanced membrane insertion capability. Simultaneously, this substitution significantly affects the formation of peptide-peptide interactions within the membrane, and leads to in reduced water flux through the membrane.

Conversely to the Q120A peptide, with the N-terminal phenylalanine substitution (F103A), we observe a significant reduction in the peptide's ability to induce vesicle fusion.

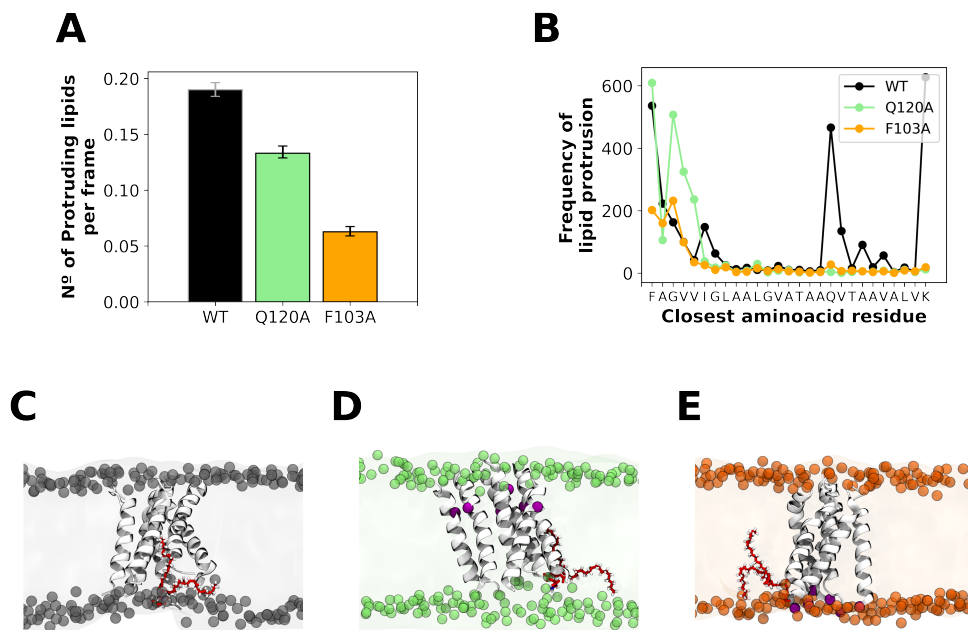


Figure 5.8: Lipid tail protrusion. (A) The plot shows the average number of protruding lipids at 4 Å from the peptides. Error bars represent the 95% confidence intervals and were calculated with bootstrap resampling. Plot (B) shows the amino acid residues that are closest to a protruding lipid. Panels (C), (D) and (E) are snapshots of the wt, Q120A and F103A simulations, respectively, where a lipid is shown in the membrane context as a representation of a lipid tail protrusion event. The substitution sites are highlighted in red.

Both lipid mixing and membrane leakage are nearly abolished in contrast to the wt or Q120A systems. This drastic decrease is partially due to the substantially lower partition coefficient of this peptide to the membrane, nearly two times lower than that of the wt peptide and five times lower than that of Q120A. This could be attributed to the fact that we are substituting an aromatic residue. These are commonly found in viral FP and have been shown to play a vital role in helping the peptide insert into membranes [145, 349]. Another important feature of aromatic residues is their ability to interact favorably with membrane phospholipids and, to partition at the lipid tail-lipid head interface, and this interaction can help destabilize the membrane and promote bilayer fusion [145]. Furthermore, the fact that F103A is an N-terminal residue in the fusion peptide may also play a role in its ability to induce membrane fusion. Previous studies have demonstrated that altering the N-terminal residue of the influenza FP to various amino acid residues has different outcomes: either retaining complete fusion capability, losing complete fusion capability, or exclusively promoting hemifusion [350, 351]. This highlights the pivotal role of the N-terminal residue in the fusion process. The results presented here clearly indicate that having a phenylalanine at the PIFP N-terminal position is critical for its fusogenic activity.

CHAPTER 5. SUBSTITUTING THE PARAINFLUENZA FUSION PEPTIDE N-TERMINAL RESIDUE IMPACTS ITS FUSOGENICITY

Overall, our findings shed light on the roles of the N-terminal phenylalanine and glutamine residues in the PIFP oligomerization, membrane perturbation, and leakage capabilities. This knowledge can be valuable for understanding viral fusion processes, as well as highlighting the significance of controlled fusion and leakage in the context of viral infection.

COMPUTATIONAL MICROSCOPY OF THE
VIRAL ENTRY PROCESS AND THE ROLE OF THE
FUSION MACHINERY

This chapter contains unpublished data

Authors: Mariana Valério, Cláudio M. Soares, Diana Lousa, Manuel N. Melo and Paulo C. T. Souza

Author contributions: M.V., C.M.S., D.L., M.N.M., and P.C.T.S designed the computational systems and analysis and M.V performed them.

6.1 Abstract

Enveloped viruses such as SARS-CoV-2 and parainfluenza, are coated by a lipid bilayer. During the early steps of infection, the fusion of this viral envelope with the host plasma membrane is a critical step for the delivery of the viral genome to the host cell. This mechanism is promoted and mediated by viral fusion proteins.

In this exploratory study, we investigated how fusion proteins from both SARS-CoV-2 and parainfluenza promote viral fusion processes by simulating the fusion process. We used coarse-grained molecular dynamics simulations to attempt fusion between simplified viral vesicles and cell membrane models, both in the presence and absence of fusion proteins. Our simulations revealed an increase in membrane fusion when fusion proteins were present in both the SARS-CoV-2 and parainfluenza systems. This attempt at fusion was characterized by the mixing of lipids from vesicles and cell membranes, accompanied by the formation of a large number and more well-defined hemifused structures when compared to simulations conducted without fusion proteins. Even though none of the systems simulated reached full fusion, these preliminary simulations provide clues into the mechanism of membrane fusion induced by fusion proteins and open important avenues to study this process.

6.2 Introduction

Membrane fusion is an important stage in the life cycle of enveloped viruses. To successfully infect a host cell, a virus must introduce its genetic material into the cell's cytoplasm. Initially, this genetic material is separated from the cell's cytoplasm by two distinct membranes: the viral envelope and the cellular membrane. Through the fusion of these membranes, the cell's cytoplasm becomes connected with the inner content of the virion, where the viral genetic material is found. Understanding the intricate details behind viral membrane fusion — a pivotal step in infection — has been a subject of intense investigation [159, 160, 172, 175, 203, 352]. This process is primarily orchestrated by viral fusion proteins that possess a common structural element known as the fusion peptide (FP). The FP is pivotal in facilitating membrane fusion by inserting into the host membrane, thereby anchoring the viral particle and inducing membrane perturbations. This process is illustrated in Figure 6.1C, which describes the viral fusion process for class I fusion proteins.

The membrane fusion process of enveloped viruses involves several intermediates, as detailed in Figure 6.2. It commences with the initial membrane apposition (Figure 6.2A), that upon the formation of hydrophobic defects (Figure 6.2B) progresses to a fusion stalk (Figure 6.2C), followed by hemifusion diaphragm expansion (Figure 6.2D), and ultimately the fusion pore opening (Figure 6.2E) [353–356]. While there is general consensus on these intermediates, uncertainties persist regarding the precise role of viral fusion proteins and any potential off-pathway states during fusion.

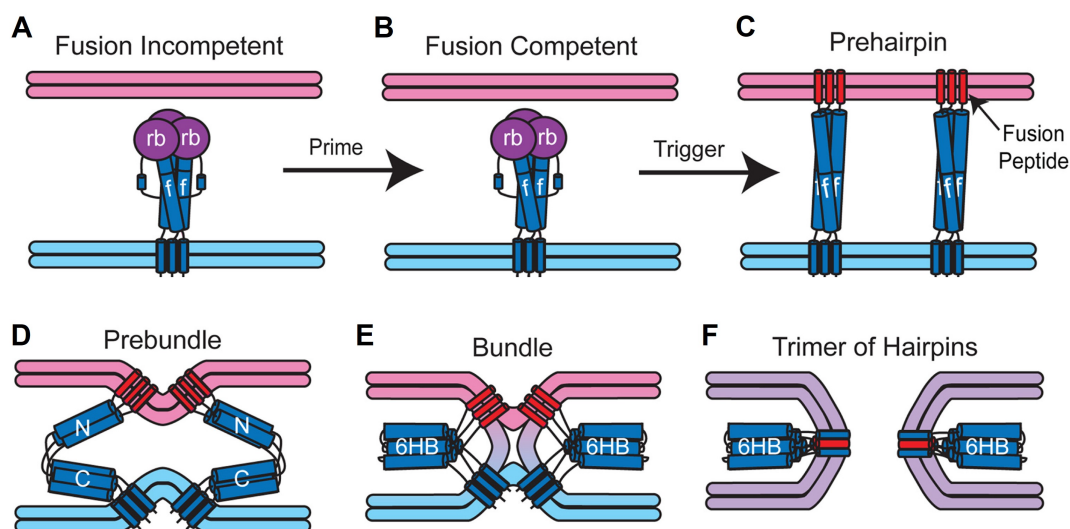


Figure 6.1: Representation of the classical viral fusion process. The model shown is for a class I fusion protein, but related models apply to class II and III fusion proteins. (A) Prior to triggering, the receptor-binding subunit (deep purple, rb) securely holds the fusion subunit (dark blue, f) in place. (B) Upon triggering the prehairpin protein embeds into the target membrane using the fusion peptide (red). (C) The prehairpin then folds back, causing the N- and C-terminal regions to assemble into a six-helix bundle (6HB). This progressive action pulls the target (pink) and viral (light blue) closer. (D) The membranes move through stages of close apposition, leading to (E) Hemifusion, where the membranes partially merge. (F) Ultimately, fusion pore formation occurs, fully merging the membranes. Reprinted with permission from [33]. Copyright © 2016 John Wiley & Sons A/S.

In this work, we focus on the fusion processes of two well-known class I enveloped viruses, SARS-CoV-2 and parainfluenza. Though distinct, both exhibit striking parallels with the influenza virus which is arguably one of the most extensively studied viruses in the realm of membrane fusion. Yet, despite the wealth of knowledge accumulated over the years [157, 160, 172, 175], establishing a unified mechanism for influenza fusion has proven challenging. The observed similarities in the fusion mechanisms of these diverse viruses provide a unique opportunity to gain insights into the broader principles governing viral membrane fusion. By leveraging these shared features and employing state-of-the-art simulation techniques, we aim to shed light on the fusion mechanisms of SARS-CoV-2 and parainfluenza, contributing to a deeper understanding of membrane fusion across enveloped viruses.

To investigate the fusion of viral lipid envelopes with cell membrane models, we performed exploratory coarse-grained (CG) molecular dynamics (MD) simulations. In these simulations, we studied the influence of viral fusion proteins on viral fusion processes, by simulating simplified viral vesicles and cell membrane models, both in the presence and absence of fusion proteins. Analogous setups have been shown to lead to the spontaneous fusion of vesicles containing the influenza fusion protein [172] or DNA-lipoplexes [357],

within the simulation time accessible to CG Martini simulations.

In our simulations, we observed a significant increase in lipid mixing in the presence of both the SARS-CoV-2 and parainfluenza fusion proteins. This increased fusion was characterized by the mixing of lipids from vesicles and cell membranes, resulting in the formation of more hemifusion intermediate structures, as compared to simulations conducted in the absence of fusion proteins. These preliminary simulations pave the way for a detailed understanding of membrane fusion and the role of fusion proteins.

6.3 Methods

To investigate the fusion behaviour of the SARS-CoV-2 and parainfluenza viral-like vesicles with cell membranes, we built several fusion systems with varying protein and lipid compositions (Figure 6.3). All simulations were run with the GROMACS 2020.3 package [220], using the Martini 3 forcefield [187] and the phospholipid topologies released with it. The Martini 3 cholesterol model from Borges-Araújo *et al.* was used [358].

6.3.1 Test simulations to define the system size, composition, and simulation conditions

As a preliminary test for our fusion simulations we used the TS2CG software [362] to construct a 70 nm viral vesicle with two SARS-CoV-2 fusion proteins on its surface and a 100% POPC lipid composition. For the fusion proteins, we generated a conformation where the transmembrane domain (TMD) and fusion peptide (FP) region are not yet bundled together. This separation allows for their individual insertion into the viral and cell membranes, as seen in figure 6.1E. Furthermore, selecting a conformation that is in proximity to the post-fusion state allows for a quicker transition to this final

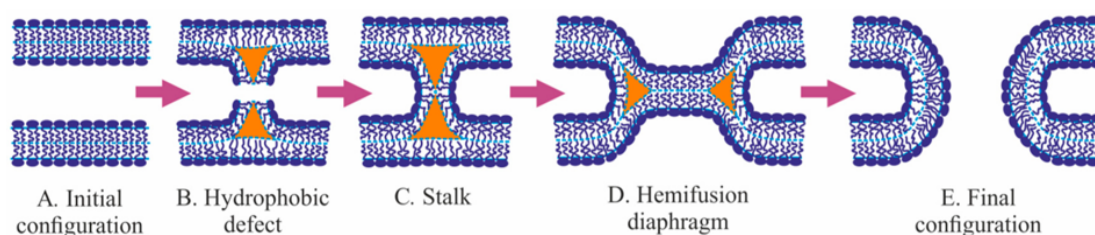


Figure 6.2: Representation of the stages of membrane fusion in a sequential manner. (A) Initial configuration, where two parallel membranes are depicted. (B) Formation of hydrophobic defects on top of two opposing bulges on the membranes. (C) Merger of contact monolayers results in the formation of the stalk. (D) Expansion of the stalk leads to the formation of the hemifusion diaphragm. (E) Pore formation on the central bilayer of the hemifusion diaphragm leads to the final configuration of the fusion pore. Orange areas represent voids between contact and distal monolayers. Reprinted with permission from [356]. Copyright © 2020 MDPI.

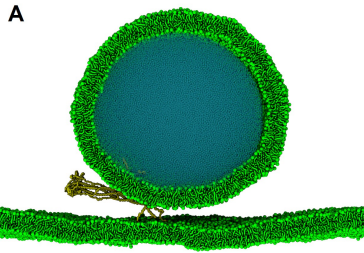
Preliminary Fusion System

Viral vesicle composition

POPC:100

Host membrane composition

POPC:100



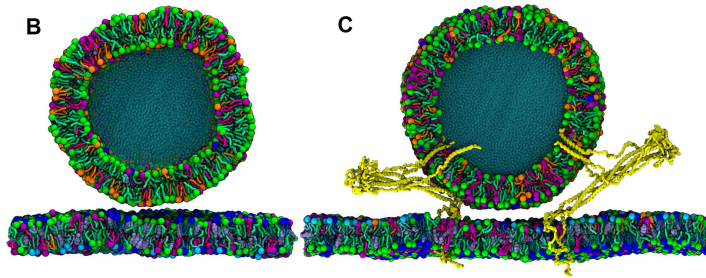
SARS-CoV-2 Fusion Systems

Viral vesicle composition

POPC:56 POPI:18
POPE:21 POPS:1
DPSM:2 CHOL:2

Host membrane composition

Outer leaflet	Inner leaflet
POPC:24	POPC:14
PAPC:12	PAPC:8
POPE:2	POPE:5
DIPE:6	DIPE:16
DPSM:24	DPSM:11
CHOL:31	PAPS:25
	SAP6:2
	CHOL:28



Parainfluenza Fusion Systems

Viral vesicle composition

POPC:16 DPSM:13
PAPC:8 SAP6:1
POPE:7 PAPS:10
DIPE:20 CHOL:25

Host membrane composition

Outer leaflet	Inner leaflet
POPC:24	POPC:14
PAPC:12	PAPC:8
POPE:2	POPE:5
DIPE:6	DIPE:16
DPSM:24	DPSM:11
CHOL:31	PAPS:25
	SAP6:2
	CHOL:28

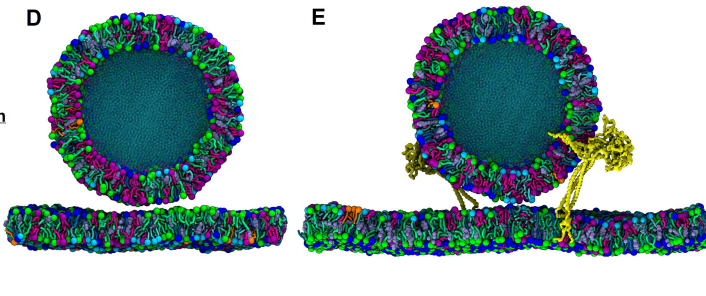


Figure 6.3: Lipid and Protein Composition of Fusion Systems. This figure provides an overview of the lipid and protein compositions in our fusion systems. We created five distinct systems: an initial system with 100% POPC lipids and two SARS-CoV-2 fusion proteins (A); two SARS-CoV-2 fusion systems, one with fusion proteins (C) and one without (B); two parainfluenza fusion systems, one with fusion proteins (E) and one without (D). The SARS-CoV-2 viral vesicle lipid composition was based on the membrane composition of Vero and A549-derived viruses [359], while the viral vesicle lipid composition was obtained from data on virions cultivated monkey kidney, baby hamster kidney, and adult hamster kidney cells [360]. The host cell membrane composition was obtained from Ingólfsson *et al.*'s eight-component average plasma membrane model [361].

state, expediting the fusion of both membranes. This bundle conformation (Figure 6.1E) is not readily accessible in the Protein Data Bank (PDB), so we generated it from the coordinates of the SARS-CoV-2 pre-fusion (PDB ID: 6XR8 [63]) and post-fusion S proteins (PDB ID: 6M3W [363]), respectively. Both structures were superimposed, as shown in supplementary figure S6.1A, to extract the bundle conformation.

In order to sample the conformational flexibility of the bundle structure we then converted the atomistic protein structure into a GōMartini representation [364–366], using the martinize2 tool[367]. The `-gows-include` flag was used during this conversion to generate a protein.itp file that includes the necessary statements for the Gō-like model. Subsequently, a contact map defining Gō-like contacts was established. This map was obtained through a web service (<http://pomalab.ippt.pan.pl/GoContactMap/>). Afterward, the contact map obtained was used, together with the CG protein structure, to generate the final parameters required for the GōMartini model. To ensure that the transmembrane regions of the fusion protein stayed inside the membrane protein-water interactions were scaled as suggested in [344].

With the vesicle already assembled (Figure 6.4B) we started by hydrating the system, adding 150 mM NaCl to mimic biological salinity and then neutralized with Na⁺ ions. Following this step, the system underwent energy minimization using both soft-core and regular methods, as well as a six-step equilibration with pores on the vesicle to balance the water molecules inside and outside the vesicle.

To construct the 100% POPC host cell membrane model containing the two fusion proteins, we utilized insane.py [318] (Figure 6.4C). Similarly to the viral vesicle system, we added water and ions, followed by energy minimization, and equilibration. The pressure coupling was set to semi-isotropic in both the xy and z dimensions, and maintained from this point onward.

After equilibrating, both the vesicle and host membrane systems, we superimposed the fusion proteins from both systems using the FP region as a reference (Figure 6.4D). The duplicated proteins were removed, as well as all the waters and ions. The simulation box dimensions were set to 100 nm x 100 nm x 80 nm in the xyz axes, accommodating a total of approximately 3 000 000 beads. The system was newly solvated, neutralized, and 150 mM of NaCl was added. It was then energy minimized, and equilibrated in six steps. Subsequently, we ran three independent fusion simulations, each spanning 10 μ s.

6.3.2 Simulations of viral fusion without fusion proteins

With the insights gained from the preliminary tests, we went on to build fusion systems with realistic membrane compositions, but without any viral protein on their surface. The cell model membrane with an inner leaflet composition of 14% POPC, 8% PAPC, 5% POPE, 16% DIPE, 11% DPSM, 28% CHOL, 25% PAPS, 2% SAP6 and an outer leaflet composition of 24% POPC, 12% PAPC, 2% POPE, 6% DIPE, 24% DPSM and 31% CHOL. The viral vesicle of the SARS-CoV-2 a membrane composition of 56% POPC, 21% POPE, 2% DPSM, 18% POPI, 1% POPS and 2% CHOL. And the parainfluenza membrane 16% POPC, 8% PAPC, 7% POPE, 20% DIPE, 13% DPSM, 1% SAP6, 10% PAPS and 25% CHOL.

As before, the viral vesicle with 30 nm was built using the TS2CG software [362], followed by the addition of water and ions (150 mM NaCl). Charge neutrality was maintained by adding counter ions Na⁺. Subsequently, the system underwent energy

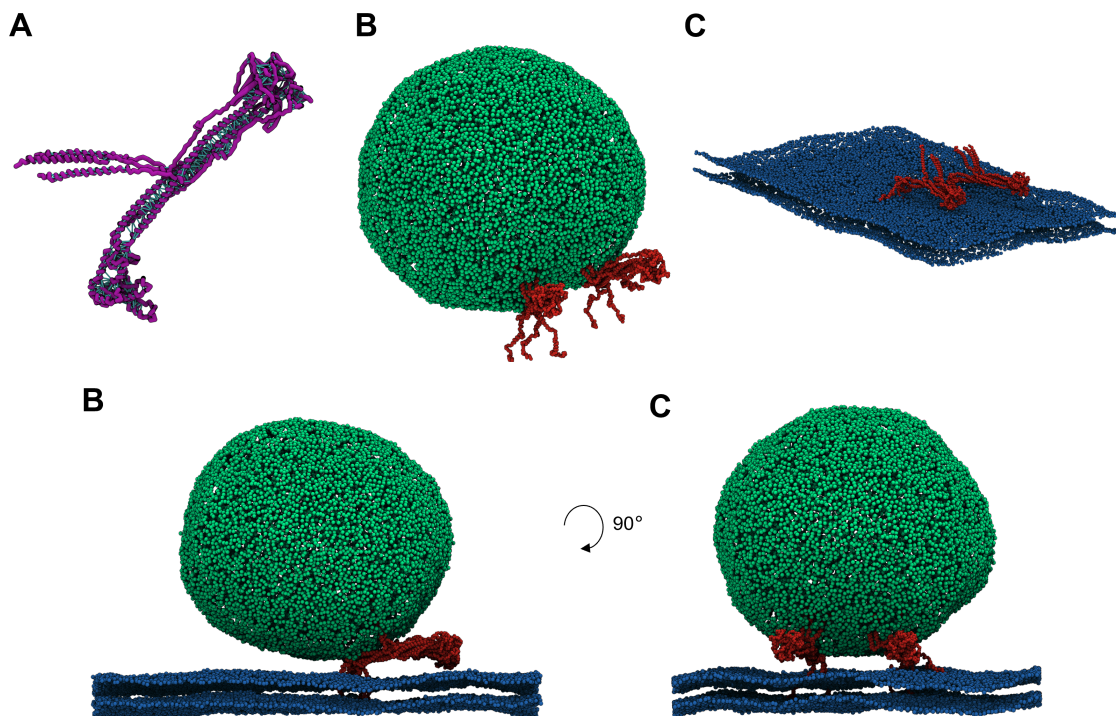


Figure 6.4: Snapshots of the preliminary fusion system components. (A) SARS-CoV-2 fusion glycoprotein in coarse-grain detail (in purple) and Gō-model elastics (in blue). (B) Simplified viral vesicle with 100% POPC lipid composition and two fusion glycoproteins (in red). (C) Host cell membrane (100% POPC) with two fusion glycoproteins (in red). (D) Fusion system, composed of a vesicle (in green) and host membrane (in blue) assembled into a near-fusion state.

minimization (soft-core and regular) and a six-step equilibration, as described before, over a duration of 100 ns. The host cell membrane model that was used to simulate the host cell membrane was generated with the *insane.py* script [318] and subsequently solvated, neutralized, and 150 mM of NaCl was added. It was then energy minimized and equilibrated under semi-isotropic pressure coupling.

Following equilibration, all the water molecules and ions were removed, and the viral vesicle was positioned approximately 2 nm away from the host cell membrane. Simulation box dimensions were set at 40 nm × 40 nm × 45 nm in the xyz axes, accommodating a total of $\approx 600,000$ beads. The system underwent solvation, salinization, energy minimization, and six additional equilibration steps. Semi-isotropic pressure coupling was maintained from this phase onwards.

After system equilibration, GROMACS was used to introduce a pulling potential between 10 lipids from the outer leaflet of the vesicle and the 10 closest lipids from host cell membrane. As described in Bruininks *et al.* [357], a brief pulling phase, lasting less than five ns, was implemented using a harmonic biasing force with a constant velocity

of -1 m/s. Once an initial stalk structure was formed, the biasing potential was removed. Subsequently, the system was re-equilibrated, and three independent fusion simulations were conducted, each spanning 10 μ s.

6.3.2.1 Simulations of viral fusion in the presence of fusion proteins

To assess the influence of incorporating fusion proteins while preserving enhanced performance, we built a new fusion system composed of a 30nm viral vesicle with two fusion proteins and the host cell model membrane. The SARS-CoV-2 fusion protein model described in the first section of the methods was also used here. For the parainfluenza, the pre-fusion and post-fusion states of the of F protein were obtained from the combination of PDB structures 2B9B [340] and 1ZTM [165], respectively (Figure S6.1B). Gō-models were also employed here.

The viral vesicles with realistic lipid compositions, as described in the previous section, and two viral proteins were assembled using the TS2CG software [362]. Only the transmembrane domain of the fusion protein was inserted in the vesicle, as shown in figure 6.4B. The system was then hydrated, and the biological salinity of 150 mM NaCl was added, together with neutralizing Na⁺ ions. Subsequently, the system underwent energy minimization (soft-core and regular) and a six-step equilibration over a duration of 100 ns.

The host cell membrane with the two fusion proteins was generated using insane.py [318]. As described in the previous section, the membrane was solvated, neutralized, and 150 mM of NaCl was added. It was then energy minimized, and equilibrated. Pressure coupling was configured to be semi-isotropic in both the xy and z dimensions from here onwards.

After equilibration, all waters and ions were removed and the proteins in the vesicle and target membrane were aligned, using the fusion peptide (FP) region as a reference point. Following this, the system underwent solvation, salinization, energy minimization, and an additional six equilibration steps.

After the final system equilibration, GROMACS was used to introduce a pulling potential as described in the previous section. Subsequently, the system was equilibrated again, and three independent fusion simulations were conducted, each spanning 10 μ s. The simulation box dimensions were set at 60 nm x 40 nm x 40 nm in the xyz axes, accommodating a total of $\approx$ 900 000 beads.

6.4 Results

6.4.1 Test simulation to define the system size, composition and simulation conditions

In our initial attempt to simulate viral fusion using Martini 3 CG models [187], we constructed a 70 nm viral vesicle with two SARS-CoV-2 fusion proteins. Our choice for

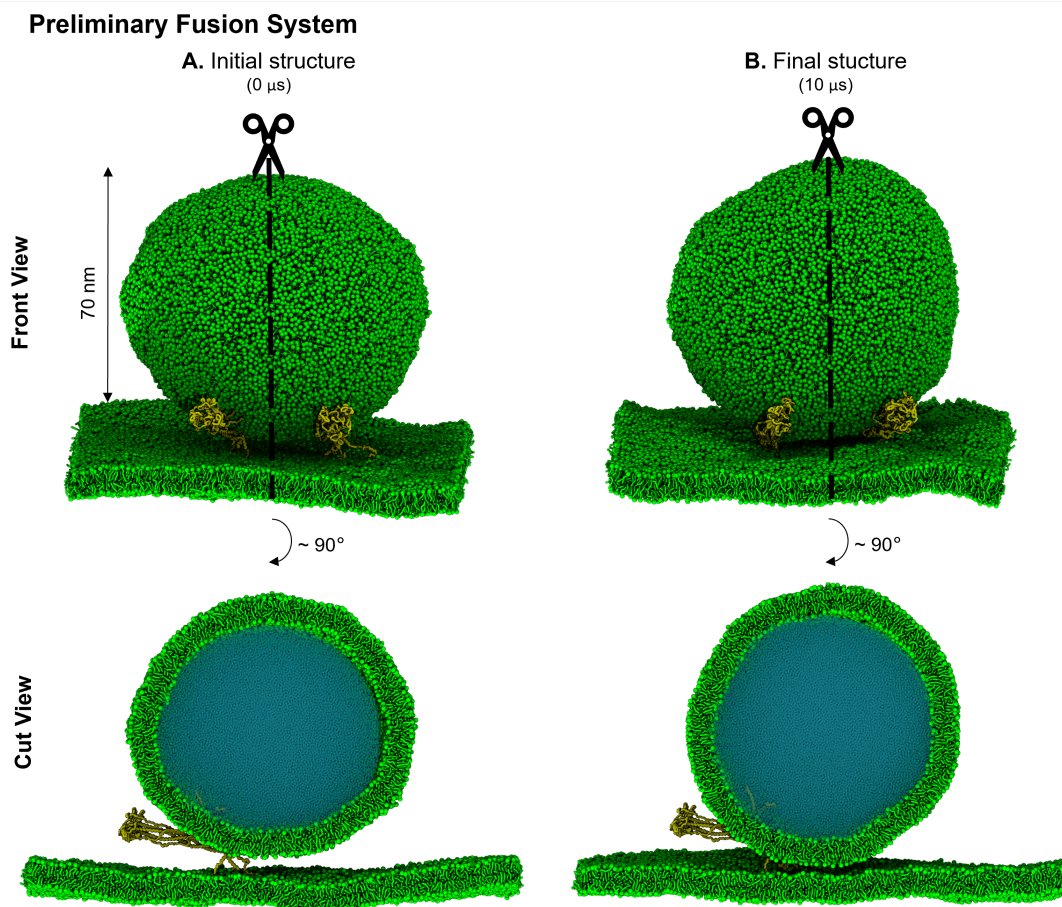


Figure 6.5: Initial and final frames of the preliminary viral fusion simulation. A 70 nm viral vesicle initially sits on top of the host cell membrane, connected by two SARS-CoV-2 fusion proteins (A). After a 10 μ s long simulation there is no mixing between the viral and host membranes (B). The top panel displays the "Front view" of the system, along with the designated cut plane for the "Cut view" image. In the bottom panel, the system is depicted following a vertical cut and an approximate 90-degree rotation. The POPC lipids are shown in green and the protein in yellow.

vesicle diameter and number of fusion proteins was based on prior research indicating that the average diameter of a SARS-CoV-2 particle is approximately 70 nm (see [206, 368]) and that the critical number of fusion proteins necessary for fusion to occur in Class I viruses, such as West Nile and Kunjin, is two [44]. The vesicle was placed 5 nm away from the host cell membrane and simulated for 10 μ s. The lipid composition of both the vesicle and target membranes was exclusively POPC lipids.

During the simulation, the vesicle and host membrane repeatedly approached and then moved apart without ever coming close enough to create a hydrophobic defect (Figure 6.5). As a consequence of the lack of interaction between the lipids of both membranes, none of the stages of fusion described in figure 6.2 was observed. We hypothesize that this outcome is due to the significant barrier created by the presence of water between the

vesicle and the target membrane. At close distances, the polar heads of the lipids orient water molecules in such a way that two polar surfaces in water repel at short distances [369]. This phenomenon is called hydration repulsion [370, 371]. Additionally, the sheer scale of the system, consisting of approximately 3,000,000 beads, makes it extremely costly to go beyond 10 μ s and this time scale is not enough to overcome the large energy barrier to fusion. These test simulations indicated that it would not be possible to observe simulation events with this setup. Therefore, we adapted our setup as described in the following section.

6.4.2 Simulations of viral fusion without fusion proteins

To enhance the performance of our fusion simulations, we reduced the size of the viral particle from 70 to 30 nm. We also benefited from insights on existing research describing the use of a pulling potential to speed up the initial contact between the hydrophobic tails of the vesicle and target cell membranes [357].

Aiming for greater biological accuracy on the membrane composition of both the viral and host cell membranes, we based the selection of their lipid compositions on previous studies. For SARS-CoV-2, we used the composition determined for Vero and A549-derived viruses [359], while for parainfluenza, we relied on data from virions cultivated in various cell lines, including monkey kidney, baby hamster kidney, and adult hamster kidney cells [360]. The host cell lipid composition was taken from Ingólfsson *et al.*, 2020, where the authors determined how much complexity could be reduced while maintaining an accurate representation of the "Average" and "Brain" plasma membrane properties [361]. We used the eight-component (A-8) average plasma membrane composition. All the detailed membrane compositions can be found in figure 6.3.

Initially, three replicates of both the SARS-CoV-2 and parainfluenza fusion systems without fusion proteins were run for 10 μ s. Interestingly, unlike what we observed in the test system, we saw progress toward the hemifused state for both viral systems. This was likely due to the crucial role of the initial lipid-pulling step in overcoming the repulsive barrier created by the water molecules. An initial hydrophobic defect (Figure 6.2B) is formed and, at the point of interbilayer contact we see the emergence of a stalk structure (Figure 6.6C and G). In this intermediate state, lipids from the outer leaflet of the vesicle mix with the top leaflet of the cell membrane model. The distant monolayers then merge into a single bilayer structure called a hemifusion intermediate structure, as seen in Figure 6.6D and H. Ideally, this process would lead to the creation of a fusion pore, connecting the internal aqueous compartments initially enclosed by the separate membranes. However, in our simulations, none of the replicates reached a fully fused state during the simulation period.

Comparing both the protein-less SARS-CoV-2 and parainfluenza fusion systems indicates that the progression of membrane fusion is identical between the three replicates of both systems. In the last frame of all replicates, we see extensive mixing of vesicle and

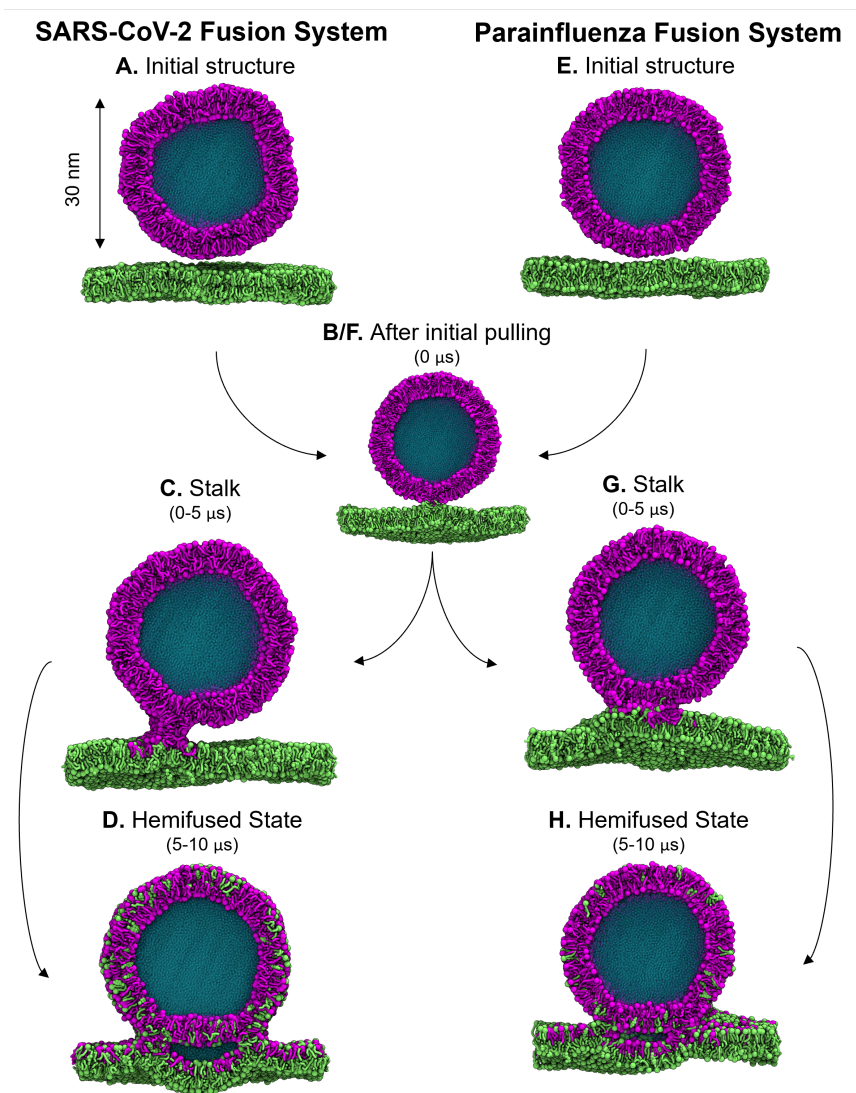


Figure 6.6: Schematic view of SARS-CoV-2 and parainfluenza viral fusion simulations without fusion proteins. Initially, the 30nm viral vesicle sits on top of the host cell membrane (A and B). After applying an initial lipid-pulling (B and F), the stalk structure is formed (G and C), which eventually leads to the formation of a wide hemifusion intermediate (D and H). The pathways indicated are based on 3 independent fusion experiments, each 10 μ s long. None of the replicates leads to full fusion of the membranes. The lipids are colored in accordance with the membrane they were a part of in the initial structure. Lipids that were part of the vesicle are shown in magenta and the ones belonging to the host cell are green. All trajectories were treated with MDVWhole [372] to prevent molecules and assemblies from splitting over the boundaries.

host cell lipids, and the formation of one hemifusion intermediate (Supplementary figures S6.2 and S6.3). This suggests that the differences in lipid composition between the two viruses had only a minor impact on the fusion process. The only behavioral difference was found in replicates 2 and 3 of the parainfluenza fusion system. Here, the simulation box shrunk in the x and y dimensions due to significant buckling of the model cell membrane

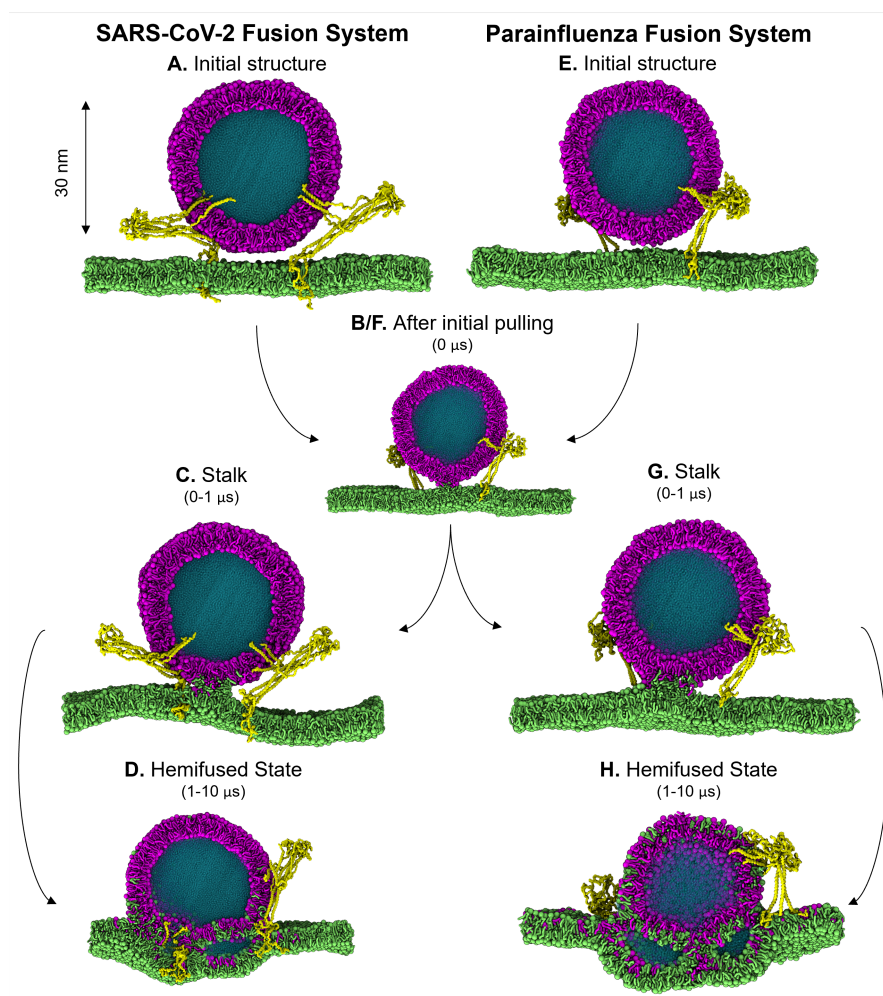


Figure 6.7: Schematic view of SARS-CoV-2 and parainfluenza viral fusion simulations with fusion proteins. The 30nm viral vesicle initially sits on top of the host cell membrane, connected by two fusion proteins (A and E). After applying an initial lipid-pulling (B and F), the stalk structure is formed (C and G), which eventually leads to the formation of a wide hemifusion intermediate (D and H). The pathways indicated are based on 3 independent fusion experiments, each 10 μ s long. None of the replicates leads to full fusion of the membranes. Lipids that were initially part of the vesicle are shown in magenta and the ones belonging to the host cell are green. Proteins are shown in yellow and waters in blue. All trajectories were treated with MDVWhole [372].

(Supplementary figure S6.3, in the second and third columns). This caused the viral vesicle to interact with its own periodic images. To address this undesired stimulation artifact, several strategies were discussed in the "Conclusions and futures perspectives" section.

6.4.3 Simulations of viral fusion promoted by fusion proteins

Having gained insights into viral-host membrane fusion in the absence proteins, we proceeded to investigate the impact of fusion proteins on the viral membrane fusion process. To investigate this we constructed SARS-CoV-2 and parainfluenza viral vesicles,

each containing two fusion protein copies, interacting with the host cell membrane. Similarly to the systems without fusion protein, we conducted triplicate simulations of both the SARS-CoV-2 and parainfluenza systems, lasting for 10 μ s each. Here the initial lipid-pulling step was also applied.

Despite the identical simulation duration, a quick visual inspection revealed a significantly higher degree of lipid mixing in the presence of fusion proteins. Here we observed the mixing of the inner bilayer of the host cell membrane (Supplementary figures S6.4 and S6.5), a phenomenon not seen in the simulations without fusion protein. Furthermore, the presence of these proteins resulted in the generation of a hemifusion diaphragm structure within the initial 1 μ s of simulation as well as, the formation of a greater number of these structures (as depicted in Figure 6.7D and H). In the simulations conducted without fusion proteins the initial stalk led to the formation of only one hemifusion state, while here, our results lead us to conclude that the perturbations caused by the fusion peptide (FP) region on the cell membrane significantly aid in accelerating the formation of fusion intermediates.

Comparing SARS-CoV-2 and parainfluenza, initial observations suggest that the degree of membrane fusion is comparable between the two systems. A visual evaluation of the impact of the different lipid compositions and the presence of fusion proteins on the fusion process is hindered by a recurring challenge in all replicates: the simulation box shrunk in the x and y dimensions due to significant fluctuations in the model cell membrane. This phenomenon is clearly depicted in supplementary figures S6.4 and S6.5. Unfortunately, this artifact prevents us from correctly analyzing the simulation, prompting the discussion of strategies to enhance the reliability of our results, in the next section.

6.5 Conclusions and future perspectives

In this exploratory study, we investigated the impact of SARS-CoV-2 and parainfluenza fusion proteins on viral membrane fusion. We utilized coarse-grained molecular dynamics simulations to explore the fusion process between simplified viral vesicles and host cell membrane models, both in the presence and absence of fusion proteins.

In our test simulations, we witnessed how the repulsive hydration and electrostatic forces are at play near the membrane surface, significantly hampering the initial contact between both the viral and host membranes.

To facilitate the process, we then simulated both SARS-CoV-2 and parainfluenza fusion systems with realistically complex membrane compositions, in the absence of proteins. To further enhance the performance of the simulation, we reduced significantly the size of the viral particle and to expedite the initial contact between the hydrophobic tails of the vesicle and target cell membranes we applied a pulling potential. This procedure mitigated the initial barrier, creating a localized hydrophobic defect (as shown in figure 6.2B) that promoted the interaction between the two lipid bilayers. In the precise location

where the interbilayer contact is established, we saw the formation of a stalk structure (Figure 6.6C and G).

According to the classical membrane fusion model, the stalk structure facilitates lipid intermixing between contacting layers, keeping them separate from the distal layers. The distal monolayers then merge into a hemifused state (Figure 6.2D), leading to the formation of a fusion pore connecting the initially enclosed aqueous compartments within separate membranes (illustrated in figure 6.2E) [353–356].

In our simulations of the fusion systems containing fusion proteins, we observed extensive lipid mixing between the viral vesicle and host cell membrane, accompanied by the formation of several hemifusion diaphragms (Figure 6.7D and H). However, no complete fusion events were observed. To a lesser extent, this phenomenon was also observed in the systems without the viral fusion protein, resulting in the formation of a single hemifused intermediate (Figure 6.6D and H) that, similarly, did not progress to the formation of a fusion pore.

Hemifusion structures are also observable through experimental confocal fluorescence microscopy. These structures can have lifespans ranging from milliseconds to seconds, and in some cases, even longer durations. Their stability and duration can be affected by various factors, including lipid composition, the presence of fusion proteins, and the presence of divalent cations [373, 374].

The limited duration of our simulations, which currently reaches up to the microsecond time scale, offers a possible explanation for the absence of full fusion events. Extending our simulations to the second timescale would demand extensive computational resources and time. An interesting alternative could be to investigate the impact of incorporating divalent ions into our fusion simulations. Notably, calcium (Ca^{2+}) and magnesium (Mg^{2+}) are present in high concentrations in the fusion environment and are recognized as significant facilitators of vesicle fusion, as they effectively reduce the duration of the HD intermediate [374].

Experimental findings also highlight the potential influence of lipid composition on the duration of the HD intermediate [373, 374]. This opens the door to another interesting discussion. When we transitioned to more realistically complex membranes, we not only reduced the size of the viral particle but also introduced an initial pulling force. These adjustments significantly mitigated the initial hydrophilic barrier, thereby facilitating the membrane fusion process. However, it remained unclear whether the initial hydrophilic repulsion was due to the 100% POPC membrane, since we performed the simultaneous alteration of multiple variables. Future work can concentrate on identifying which lipid factors might be aiding in the fusion process.

Even though a complete fusion event was not observed in our simulations, we have made significant strides in understanding the pivotal role of fusion proteins in this process. In fact, in the systems containing a realistically complex membrane without fusion protein, we see no mixing of the lipids in the inner bilayer of the cell membrane. Remarkably, the introduction of viral fusion proteins alone was sufficient to trigger mixing between the

vesicle and the inner cell membrane bilayer. This could be further improved by tweaking the protein model used. In our systems, the protein is held in a bundle conformation, however, it is conceivable that conformational rearrangements during the process could facilitate fusion and aid in surpassing the hemifusion state. While implementing this within the Martini framework may not be straightforward, one potential approach is to explore a two-state Gō-model to replicate such conformational changes.

Besides the inability to see full fusion of the viral and host membranes, we encountered another persistent issue across most of the fusion systems: the simulation box experienced periodic shrinkage in the x and y dimensions, primarily caused by notable deformations in the model cell membrane. This unintended artifact is illustrated in figures S6.4 and S6.5. To overcome this challenge, we plan to implement several strategies to enhance the reliability of our results:

- **Increase the simulation box size:** Enlarging the simulation box will help prevent interactions between the primary system and its periodic images.

- **Modify neighbor list updates:** We will investigate the possibility of increasing the frequency of neighbor list updates, as the membrane deformation problem has been attributed to the combination of short outer cutoffs and infrequent neighbor list updates, leading to the miss of long-range attractive Lennard-Jones interactions [375].

- **Consider bilayer re-scaling:** An alternative approach involves re-scaling the outer vesicle and top cell bilayers to match the number of lipids in the inner vesicle and bottom cell bilayers. This adjustment may enhance the likelihood of achieving full membrane fusion, as suggested in previous studies [357].

Once these strategies are implemented, and the simulation runs smoothly, we will proceed with a detailed analysis of the simulation results. We will start by assessing whether complete fusion is achieved and, if yes, in which simulation systems. Secondly, we will quantify the lipid mixing between viral and host membranes, examining whether specific lipids tend to localize in particular regions of the stalk and hemifusion structures. Next, we will focus on the precise location of the FP region within the stalk, hemifusion diaphragm and fusion pore structures. This will be followed by the analysis of which types of lipids interact most prominently with the FPs during the membrane fusion process. Lastly, to validate the accuracy of our simulations and provide additional context for our findings, we will compare our simulation results with experimental observations, focusing on the behavior of the fusion proteins.

While this is still preliminary work, we find this system will be a promising tool for unveiling the secrets of viral membrane fusion.

**6.6 Supplementary Information of Computational microscope
of the Viral Entry Process and the role of the fusion
machinery**

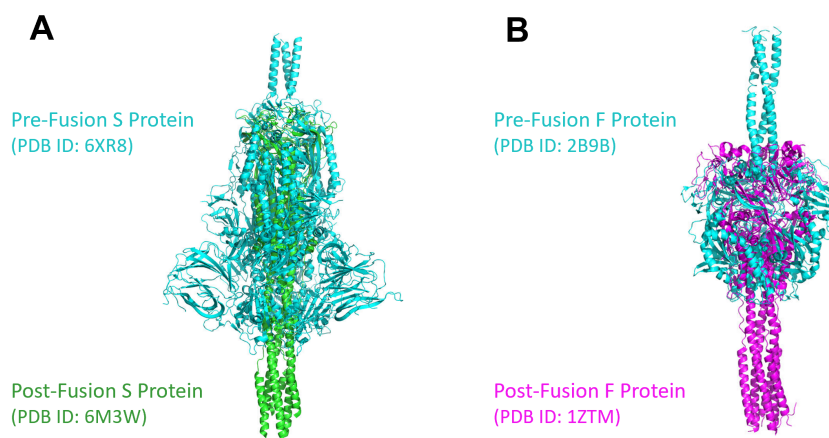


Figure S6.1: Building the pre-bundle state of the SARS-CoV-2 and PIV fusion proteins. (A) Superposition of both the pre- and post-fusion of the SARS-CoV-2 S protein. (B) Superposition of both the pre- and post-fusion of the Parainfluenza F protein.

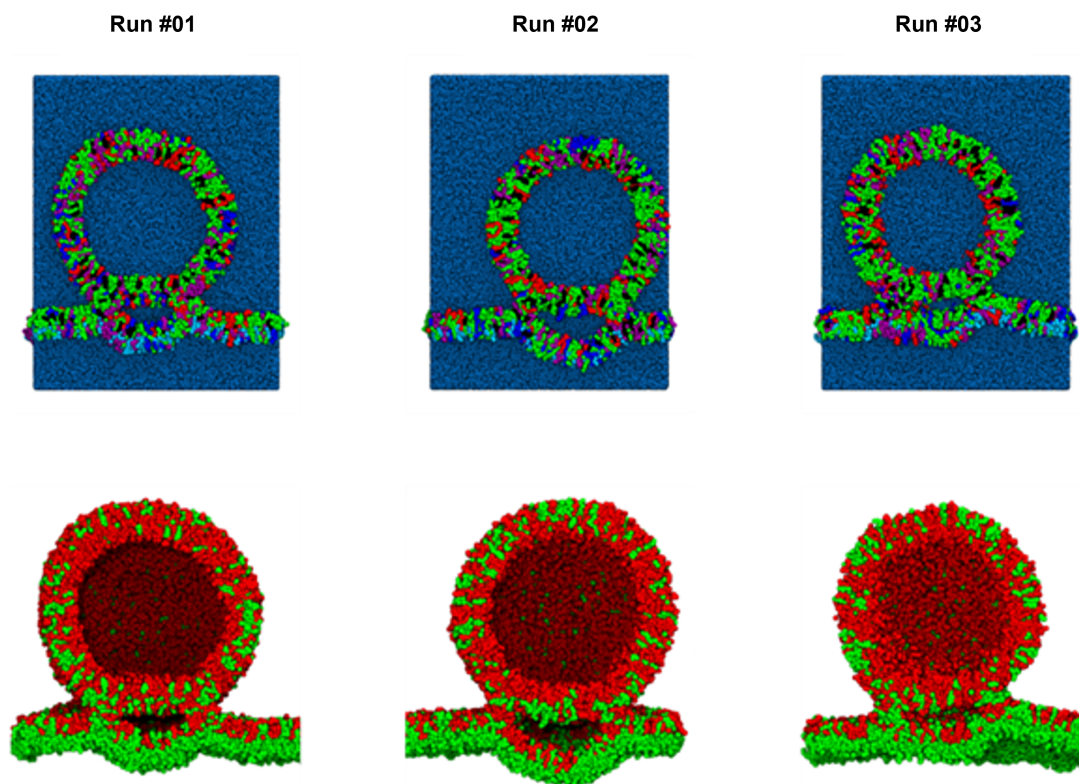


Figure S6.2: Snapshots of the SARS-CoV-2 fusion simulations without fusion proteins. All images are snapshots of the last frame. On the first row lipids are colored by head type (PC:green, PE:purple, PS:cyan, PI and PI(4,5)P₂:orange and SM:blue) and the cholesterol molecule is shown in mauve. On the second row the lipids that were part of the vesicle in frame 1 are shown in magenta and the host cell lipids in green. All trajectories were treated with MDVWhole [372] to prevent molecules and assemblies from splitting over the boundaries.

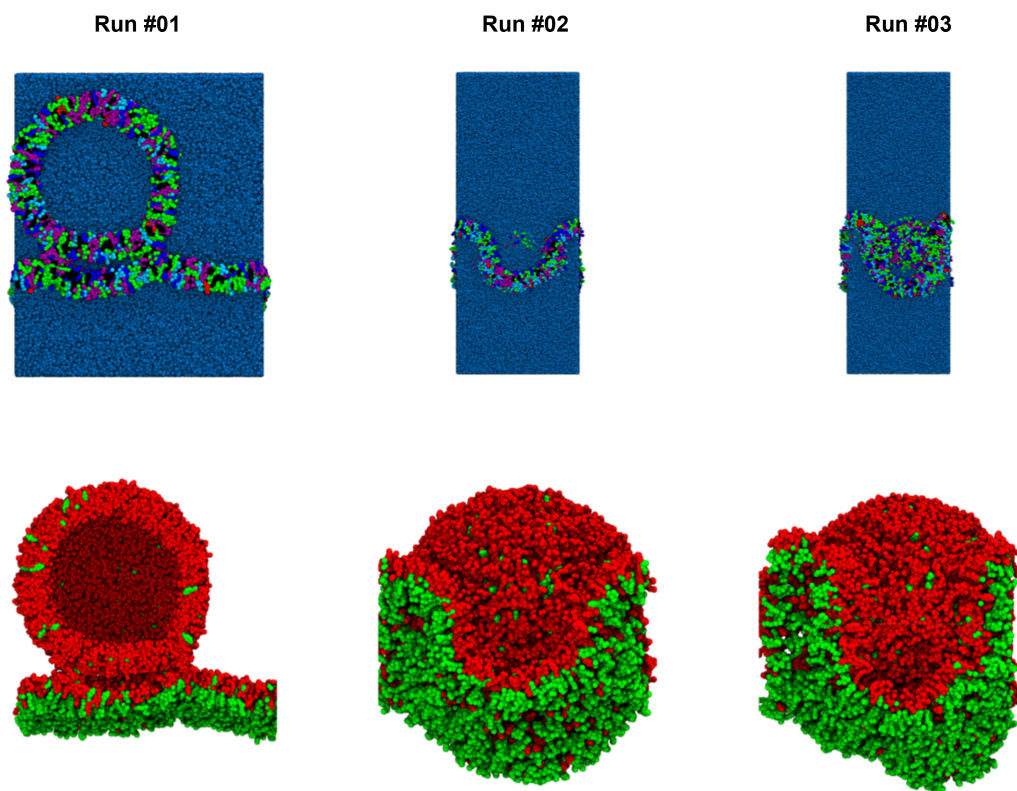


Figure S6.3: Snapshots of the parainfluenza fusion simulations without fusion proteins. All images are snapshots of the last frame. On the second row the lipids that were part of the vesicle in frame 1 are shown in magenta and the host cell lipids in green. On the last row lipids are colored by head type (PC:green, PE:purple, PS:cyan, PI and PI(4,5)P₂:orange and SM:blue) and the cholesterol molecule is shown in mauve. All trajectories were treated with MDVWhole [372] to prevent molecules and assemblies from splitting over the boundaries.

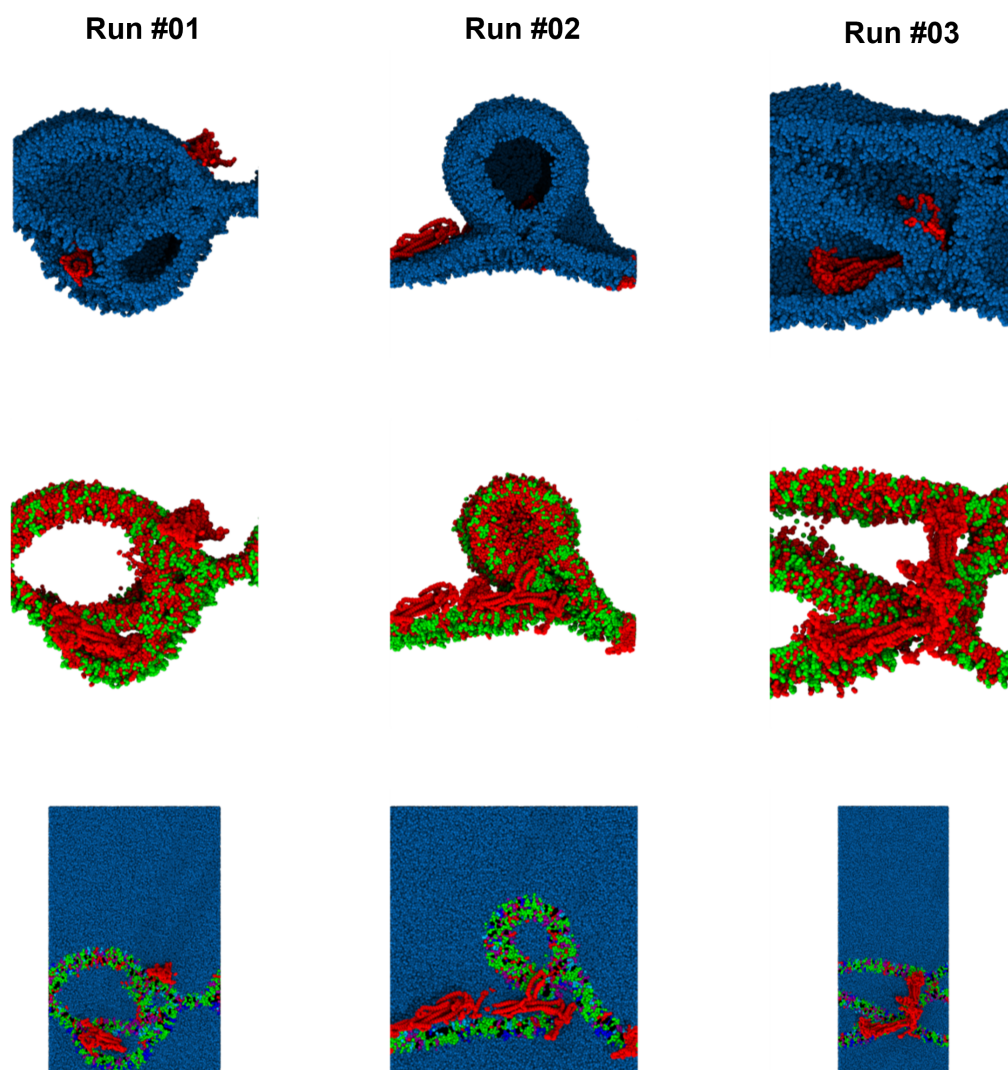


Figure S6.4: Snapshots of the SARS-CoV-2 fusion simulations with fusion proteins. All images are snapshots of the last frame. On the first row lipids are colored by head type (PC:green, PE:purple, PS:cyan, PI and PI(4,5)P₂:orange and SM:blue) and the cholesterol molecule is shown in mauve. On the second row the lipids that were part of the vesicle in frame 1 are shown in magenta and the host cell lipids in green. Protein backbone beads are shown in red. All trajectories were treated with MDVWhole [372] to prevent molecules and assemblies from splitting over the boundaries.

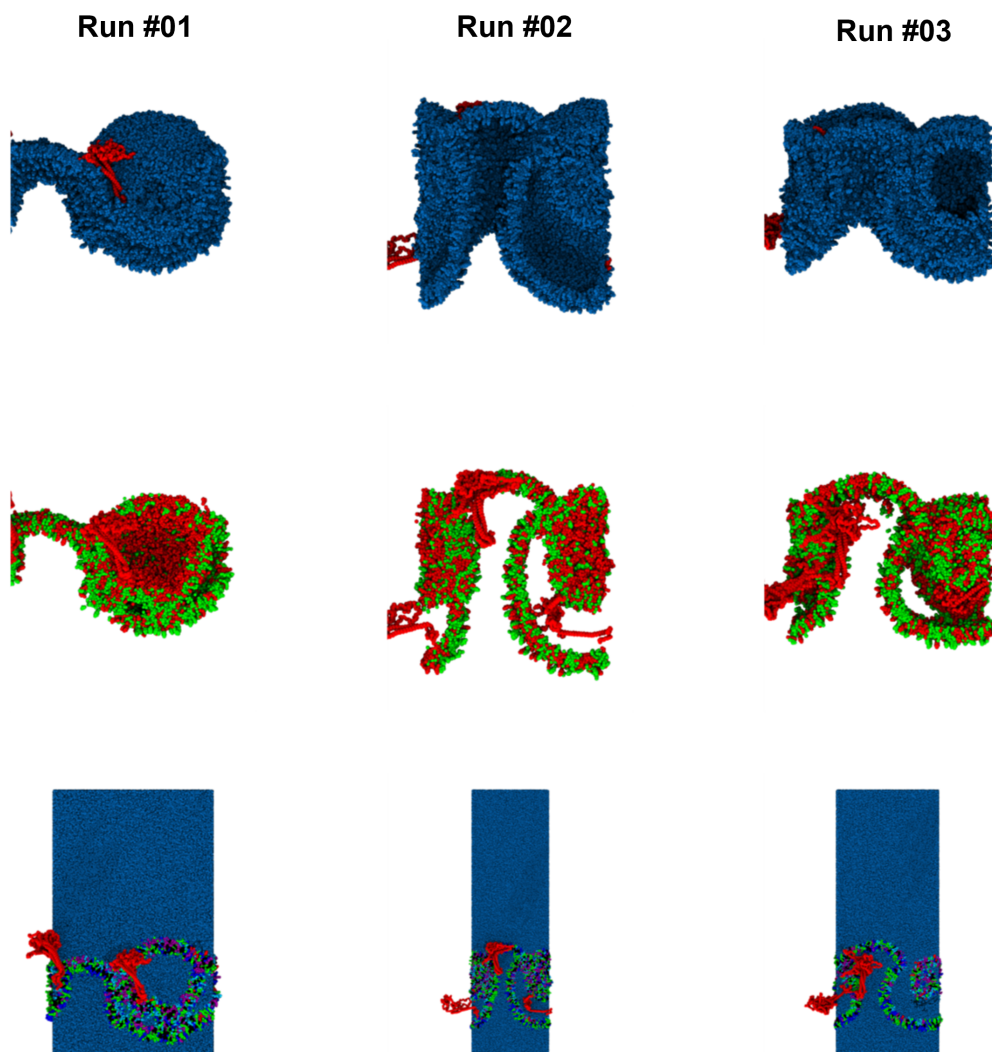


Figure S6.5: Snapshots of the parainfluenza fusion simulations with fusion proteins. All images are snapshots of the last frame. On the second row the lipids that were part of the vesicle in frame 1 are shown in magenta and the host cell lipids in green. On the last row lipids are colored by head type (PC:green, PE:purple, PS:cyan, PI and PI(4,5)P₂:orange and SM:blue) and the cholesterol molecule is shown in mauve. Protein backbone beads are shown in red. All trajectories were treated with MDVWhole [372] to prevent molecules and assemblies from splitting over the boundaries.

GENERAL CONCLUSIONS

In this dissertation we investigated the role of fusion proteins in the viral fusion process, focusing on the target cell receptor recognition and membrane fusion steps. For this, we focused on two viruses - the SARS-CoV-2 and parainfluenza -, leading to the following studies:

1. Exploring the impact of mutations on the SARS-CoV-2 RBD conformational dynamics and antibody neutralization;
2. Parainfluenza Fusion Peptide Promotes Membrane Fusion by Assembling into Oligomeric Porelike Structure;
3. Substituting the Parainfluenza Fusion Peptide N-terminal residue impacts its fusogenicity;
4. Computational microscopy of the viral entry process and the role of the fusion machinery.

We based our first study on the prior knowledge that, during the COVID-19 pandemic, many naturally occurring mutations emerged in the SARS-CoV-2 spike protein, several of them in the RBD region. Some of these mutations led to the emergence of dominant viral variants in specific regions due to their increased fitness advantage [83, 176]. To investigate if these mutations induced significant changes in the conformational dynamics of the RBD, when compared with the wild-type strain, we conducted long atomistic molecular dynamics simulations. The wild-type RBD and four VOCs (Alpha, Beta, Delta, and Omicron) were studied. The simulation results revealed two distinct conformations of the RBD - an "open" conformation facilitating ACE2 binding and a "closed" conformation blocking the binding surface. The Alpha and Beta variants shifted the equilibrium towards the open conformation, potentially enhancing ACE2 binding affinity, while, the Delta and Omicron variants exhibited significant deviations, rarely adopting the closed conformation and displaying altered orientations. These findings suggest that VOCs, particularly the Omicron and Delta variants, modulate the conformational dynamics of the RBD to facilitate ACE2 binding and possibly aid in antibody escape. We also contributed to a study to a study aimed at identifying the effects of single and synergistic amino acid substitutions in promoting SARS-CoV-2 immune escape. For this we analyzed 57 PDB structures of the S

protein bound to neutralizing antibodies to find which residues interact more commonly with antibodies, and compared this data with RBD-ACE2 simulations. Results showed that E484 and S494 mutants may be cause for concern due to their ability to escape the immune system.

Another critical step in membrane fusion is the perturbation induced by the insertion of the fusion peptide into the host membrane. In this thesis, this process was studied by focusing on the influenza fusion peptide, for which little was known regarding its mechanism of action. In this virus, fusion is mediated by the F protein, which after binding to the host cell, undergoes major conformational rearrangements and is proteolytically cleaved, exposing the FP region [177, 178]. After being exposed, this region is inserted into the target membrane, [179] facilitating membrane fusion. To investigate how the PIFP interacts with and perturbs the membrane, we employed molecular dynamics simulations and compared our results with biophysical assays performed by our collaborators at IMM, Lisbon. At high concentrations, the parainfluenza fusion peptide (PIFP) induced the formation of a water-permeable pore-like structure, which facilitated lipid head intrusion and lipid tail protrusion, ultimately promoting membrane fusion. Experimental results supported these findings, demonstrating that the PIFP could induce fusion and/or membrane leakage depending on the peptide/lipid ratio. With the MD simulations, we were also able to identify key amino acid residues of the PIFP involved in membrane fusion.

Building upon these findings, we then introduced alanine mutations to key amino acids within the PIFP that were identified as critical for membrane fusion, specifically Q120 and F103. Employing a combination of molecular dynamics simulations and biophysical assays, we explored the impact of these mutations. The results of the biophysical assays indicated that the Q120A mutation did not impede the peptide's ability to induce vesicle fusion when compared to the wild-type (wt) peptide. This observation was corroborated by *in-silico* data. Conversely, the mutation of the phenylalanine residue significantly impaired the peptide's capacity to induce vesicle fusion and membrane leakage. This effect could be attributed to a markedly lower partition coefficient for this peptide within the membrane, which was nearly twice as low as that of the wt peptide and five times lower than Q120A.

In the last chapter, we delve into exploratory research on the role of SARS-CoV-2 and parainfluenza fusion proteins in viral fusion. We employed coarse-grained molecular dynamics simulations to investigate fusion processes between simplified viral vesicles and cell membrane models, both with and without fusion proteins. Intriguingly, within systems featuring a fusion protein, we observed a significant enhancement toward membrane fusion for both SARS-CoV-2 and parainfluenza. These simulations showcased extensive mixing of lipids from the vesicles and cell membranes, and the formation of a greater number of hemifusion structures compared to simulations conducted without fusion proteins. While achieving full membrane fusion was not possible due to certain limitations, we plan to explore the avenues detailed in Chapter 6 in hopes of achieving it.

Collectively, this dissertation journey integrated research on both SARS-CoV-2 and parainfluenza, providing a comprehensive understanding of fusion protein mode of action during viral fusion at a molecular level. These findings allow for the expansion of knowledge of viral entry and lead the way to potential interventions aimed at blocking and combating infectious diseases, namely at the cell entry stage.

BIBLIOGRAPHY

- [1] J. M. Lourenço. *The NOVAthesis L^AT_EX Template User's Manual*. NOVA University Lisbon. 2021. URL: <https://github.com/joaomlourenco/novathesis/raw/main/template.pdf> (cit. on p. i).
- [2] S.-Y. Zhang et al. "Human Toll-like receptor-dependent induction of interferons in protective immunity to viruses". In: *Immunological reviews* 220.1 (2007), pp. 225–236 (cit. on p. 1).
- [3] D. M. Morens and e. al et. "The COVID-19 pandemic: what can we learn from it?" In: *Journal of Clinical Investigation* 130.5 (2020), pp. 2140–2145 (cit. on p. 1).
- [4] World Health Organization. *Measles*. Published: [<https://www.who.int/news-room/fact-sheets/detail/measles>]. 2018 (cit. on p. 1).
- [5] E. Simons and e. al et. "Assessment of the 2010 global measles mortality reduction goal: results from a model of surveillance data". In: *The Lancet* 379.9832 (2019), pp. 2173–2178 (cit. on p. 1).
- [6] B. N. Fields and e. al et. *Fields virology*. Wolters Kluwer Health/Lippincott Williams & Wilkins, 2013 (cit. on p. 1).
- [7] M. M. Patel and e. al et. "Global epidemiology of hepatitis A virus infection: new estimates of age-specific antibody to HAV seroprevalence". In: *Hepatology* 48.4 (2008), pp. 1333–1341 (cit. on p. 1).
- [8] T. Solomon and e. al et. "Neurological manifestations of dengue infection". In: *The Lancet* 369.9572 (2007), pp. 2052–2059 (cit. on p. 1).
- [9] A. Clark and e. al et. "Global, regional, and national estimates of the population at increased risk of severe COVID-19 due to underlying health conditions in 2020: a modelling study". In: *The Lancet Global Health* 8.8 (2017), e1003–e1017 (cit. on p. 1).
- [10] UNICEF. *Progress on household drinking water, sanitation, and hygiene 2000–2017: Special focus on inequalities*. Published: [<https://www.unicef.org/reports/progress-on-drinking-water-sanitation-and-hygiene-2019>]. 2019 (cit. on p. 1).

- [11] P. Murcia, W. Donachie, and M. Palmarini. "Viral pathogens of domestic animals and their impact on biology, medicine and agriculture". In: *Encyclopedia of Microbiology* (2009), p. 805 (cit. on p. 1).
- [12] P. Zhou et al. "A pneumonia outbreak associated with a new coronavirus of probable bat origin". In: *nature* 579.7798 (2020), pp. 270–273 (cit. on p. 1).
- [13] S. K. Lam and K. B. Chua. "Nipah virus encephalitis outbreak in Malaysia". In: *Clinical Infectious Diseases* 34.Supplement_2 (2002), S48–S51 (cit. on p. 2).
- [14] M. S. Islam et al. "Nipah virus transmission from bats to humans associated with drinking traditional liquor made from date palm sap, Bangladesh, 2011–2014". In: *Emerging infectious diseases* 22.4 (2016), p. 664 (cit. on p. 2).
- [15] G. Conroy. "Nipah virus outbreak: what scientists know so far." In: *Nature* (2023) (cit. on p. 2).
- [16] W. A. Orenstein and e. al et. "Vaccines". In: *Plotkin's Vaccines*. Elsevier, 2004, pp. 53–92 (cit. on p. 2).
- [17] World Health Organization. *Vaccines and immunization: diseases and vaccines*. Published: [link to article]. 2020 (cit. on p. 2).
- [18] D. S. Dimitrov. "Virus entry: molecular mechanisms and biomedical applications". In: *Nature Reviews Microbiology* 2.2 (2004), pp. 109–122 (cit. on pp. 2, 4).
- [19] J. E. Jones, V. Le Sage, and S. S. Lakdawala. "Viral and host heterogeneity and their effects on the viral life cycle". In: *Nature Reviews Microbiology* 19.4 (2021), pp. 272–282 (cit. on p. 2).
- [20] M. S. Maginnis. "Virus–receptor interactions: the key to cellular invasion". In: *Journal of molecular biology* 430.17 (2018), pp. 2590–2611 (cit. on pp. 2, 7).
- [21] V. Más and J. A. Melero. "Entry of enveloped viruses into host cells: membrane fusion". In: *Structure and Physics of Viruses: An Integrated Textbook* (2013), pp. 467–487 (cit. on p. 3).
- [22] M. G. Mateu. "Introduction: The structural basis of virus function". In: *Structure and physics of viruses: an integrated textbook* (2013), pp. 3–51 (cit. on p. 3).
- [23] J. Mercer and A. Helenius. "Virus entry by endocytosis". In: *Annual Review of Biochemistry* 79 (2010), pp. 803–833 (cit. on p. 3).
- [24] S. B. Sieczkarski and G. R. Whittaker. "Influenza virus can enter and infect cells in the absence of clathrin-mediated endocytosis". In: *Journal of Virology* 76.20 (2002), pp. 10455–10464 (cit. on p. 3).
- [25] E. G. Acosta and R. Bartenschlager. "The multifaceted role of clathrin-mediated endocytosis during viral infection". In: *Trends in Microbiology* 17.12 (2009), pp. 590–597 (cit. on p. 3).

- [26] G. J. Doherty and H. T. McMahon. "Mechanisms of endocytosis". In: *Annual review of biochemistry* 78 (2009). Publisher: Annual Reviews, pp. 857–902 (cit. on p. 3).
- [27] J. M. White et al. "Membrane fusion". In: *Cell* 135.3 (2008). Publisher: Elsevier, pp. 603–613 (cit. on pp. 3, 116).
- [28] S. C. Harrison. "Viral membrane fusion". en. In: *Virology* 479-480 (2015-05), pp. 498–507. ISSN: 00426822. DOI: 10.1016/j.virol.2015.03.043. URL: <https://linkinghub.elsevier.com/retrieve/pii/S004268221500183X> (visited on 2022-01-28) (cit. on pp. 3, 7).
- [29] S. Belouzard, V. C. Chu, and G. R. Whittaker. "Entry route of the enveloped virus hepatitis C virus depends on clathrin-mediated endocytosis". In: *Journal of Virology* 83.1 (2009), pp. 398–411 (cit. on p. 3).
- [30] J. Shang et al. "Structural basis of receptor recognition by SARS-CoV-2". en. In: *Nature* 581.7807 (2020-05), pp. 221–224. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-020-2179-y. URL: <http://www.nature.com/articles/s41586-020-2179-y> (visited on 2021-03-03) (cit. on pp. 3, 11, 34, 46).
- [31] C. B. Jackson et al. "Mechanisms of SARS-CoV-2 entry into cells". In: *Nature Reviews Molecular Cell Biology* (2021-10). Publisher: Nature Publishing Group, pp. 1–18. ISSN: 14710080. DOI: 10.1038/s41580-021-00418-x (cit. on pp. 3, 5, 32, 33).
- [32] K. Miyauchi et al. "HIV enters cells via endocytosis and dynamin-dependent fusion with endosomes". In: *Cell* 137.3 (2009), pp. 433–444 (cit. on p. 3).
- [33] J. M. White and G. R. Whittaker. "Fusion of enveloped viruses in endosomes". In: *Traffic* 17.6 (2016). Publisher: Wiley Online Library, pp. 593–614 (cit. on pp. 5, 12, 136).
- [34] R. A. Lamb, T. S. Jardetzky, and T. Liu. "Molecular basis of viral membrane fusion". In: *Nature* 7.1 (2007). Publisher: Nature Publishing Group, pp. 89–100 (cit. on p. 5).
- [35] C. Lozada et al. "Identification and characteristics of fusion peptides derived from enveloped viruses". In: *Frontiers in Chemistry* 9 (2021), p. 689006 (cit. on pp. 6, 13).
- [36] M. Kielian and F. A. Rey. "Virus membrane-fusion proteins: More than one way to make a hairpin". In: *Nature Reviews Microbiology* 4.1 (2006). Publisher: Nature Publishing Group, pp. 67–76 (cit. on p. 6).
- [37] D. C. Chan et al. "Core structure of gp41 from the HIV envelope glycoprotein". In: *Cell* 89.2 (1997). Publisher: Elsevier, pp. 263–273 (cit. on p. 7).
- [38] J. M. Kilby et al. "Potent suppression of HIV-1 replication in humans by T-20, a peptide inhibitor of gp41-mediated virus entry". In: *Nature Medicine* 4.11 (1998). Publisher: Nature Publishing Group, pp. 1302–1307 (cit. on p. 7).

- [39] Y. Zhu et al. "Design of potent membrane fusion inhibitors against SARS-CoV-2, an emerging coronavirus with high fusogenic activity". In: *Journal of virology* 94.14 (2020), e00635–20 (cit. on p. 7).
- [40] M. I. Barton et al. "Effects of common mutations in the SARS-CoV-2 Spike RBD and its ligand, the human ACE2 receptor on binding affinity and kinetics". In: *Elife* 10 (2021), e70658 (cit. on p. 7).
- [41] V. Kumar et al. "Possible link between higher transmissibility of alpha, kappa and delta variants of SARS-CoV-2 and increased structural stability of its spike protein and hACE2 affinity". In: *International journal of molecular sciences* 22.17 (2021), p. 9131 (cit. on p. 7).
- [42] J. Grove and M. Marsh. "The cell biology of receptor-mediated virus entry". In: *Journal of Cell Biology* 195.7 (2011), pp. 1071–1082 (cit. on p. 7).
- [43] T. Ivanovic et al. "Influenza-virus membrane fusion by cooperative fold-back of stochastically induced hemagglutinin intermediates". In: *Elife* 2 (2013), e00333 (cit. on p. 7).
- [44] L. H. Chao et al. "Sequential conformational rearrangements in flavivirus membrane fusion". In: *Elife* 3 (2014), e04389 (cit. on pp. 7, 142).
- [45] Y. Cai et al. "Distinct conformational states of SARS-CoV-2 spike protein". In: *Science* 369.6511 (2020), pp. 1586–1592 (cit. on p. 7).
- [46] A. C. Walls et al. "Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein". In: *Cell* 181.2 (2020-04). Publisher: Cell Press, 281–292.e6. ISSN: 10974172. DOI: 10.1016/j.cell.2020.02.058 (cit. on pp. 7, 9, 46).
- [47] J. M. White et al. "Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme". In: *Critical reviews in biochemistry and molecular biology* 43.3 (2008). Publisher: Taylor & Francis, pp. 189–219 (cit. on p. 7).
- [48] R. M. DuBois et al. "The structural basis for receptor recognition by New World hemorrhagic fever arenaviruses". In: *Nature* 534.7609 (2016). Publisher: Nature Publishing Group, pp. 1–11 (cit. on p. 7).
- [49] M. de Graaf and R. A. Fouchier. "Role of receptor binding specificity in influenza A virus transmission and pathogenesis". In: *The EMBO journal* 33.8 (2014), pp. 823–841 (cit. on p. 8).
- [50] Y. Shi et al. "Enabling the 'host jump': structural determinants of receptor-binding specificity in influenza A viruses". In: *Nature Reviews Microbiology* 12.12 (2014), pp. 822–831 (cit. on p. 8).
- [51] L. Gillespie et al. "Endocytic function is critical for influenza A virus infection via DC-SIGN and L-SIGN". In: *Scientific reports* 6.1 (2016), p. 19428 (cit. on p. 8).

-
- [52] S. L. Londrigan et al. "N-linked glycosylation facilitates sialic acid-independent attachment and entry of influenza A viruses into cells expressing DC-SIGN or L-SIGN". In: *Journal of virology* 85.6 (2011), pp. 2990–3000 (cit. on p. 8).
 - [53] I. V. Alymova et al. "Receptor-binding specificity of the human parainfluenza virus type 1 hemagglutinin–neuraminidase glycoprotein". In: *Glycobiology* 22.2 (2012), pp. 174–180 (cit. on p. 8).
 - [54] T. Takimoto et al. "Role of the hemagglutinin-neuraminidase protein in the mechanism of paramyxovirus-cell membrane fusion". In: *Journal of virology* 76.24 (2002), pp. 13028–13033 (cit. on p. 8).
 - [55] R. A. Lamb, R. G. Paterson, and T. S. Jardetzky. "Paramyxovirus membrane fusion: lessons from the F and HN atomic structures". In: *Virology* 344.1 (2006), pp. 30–37 (cit. on p. 8).
 - [56] M. Hoffmann et al. "SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor". In: *Cell* 181.2 (2020-04). Publisher: Cell Press, 271–280.e8. ISSN: 10974172. DOI: 10.1016/j.cell.2020.02.052 (cit. on pp. 8, 9, 33).
 - [57] C. B. Jackson et al. "Mechanisms of SARS-CoV-2 entry into cells". In: *Nature Reviews Molecular Cell Biology* 23.1 (2021-10), pp. 3–20. DOI: 10.1038/s41580-021-00418-x. URL: <https://doi.org/10.1038/s41580-021-00418-x> (cit. on p. 8).
 - [58] Z.-Y. Yang et al. "pH-dependent entry of severe acute respiratory syndrome coronavirus is mediated by the spike glycoprotein and enhanced by dendritic cell transfer through DC-SIGN". In: *Journal of virology* 78.11 (2004), pp. 5642–5650 (cit. on p. 8).
 - [59] R. Amraei et al. "CD209L/L-SIGN and CD209/DC-SIGN act as receptors for SARS-CoV-2". In: *ACS Central Science* 7.7 (2021), pp. 1156–1165 (cit. on p. 8).
 - [60] S. A. Jeffers et al. "CD209L (L-SIGN) is a receptor for severe acute respiratory syndrome coronavirus". In: *Proceedings of the National Academy of Sciences* 101.44 (2004), pp. 15748–15753 (cit. on p. 8).
 - [61] J. Liu et al. "Molecular architecture of native HIV-1 gp120 trimers". In: *Nature* 455.7209 (2008), pp. 109–113 (cit. on p. 8).
 - [62] D. Wrapp et al. "Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation". In: *Science* 367.6483 (2020). Publisher: American Association for the Advancement of Science, pp. 1260–1263 (cit. on pp. 9, 33, 34, 46).
 - [63] Y. Cai et al. "Distinct conformational states of SARS-CoV-2 spike protein". en. In: *Science* (2020-07), eabd4251. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.abd4251. URL: <https://www.sciencemag.org/lookup/doi/10.1126/science.abd4251> (visited on 2021-03-22) (cit. on pp. 9, 138).

- [64] Y. Huang et al. "Structural and functional properties of SARS-CoV-2 spike protein: potential antiviral drug development for COVID-19". In: *Acta Pharmacologica Sinica* 41.9 (2020-09). Publisher: Springer Nature, pp. 1141–1149. ISSN: 17457254. DOI: 10.1038/s41401-020-0485-4 (cit. on pp. 9, 33).
- [65] D. J. Benton et al. "Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion". In: *Nature* 588.7837 (2020-12). Publisher: Nature, pp. 327–330. ISSN: 1476-4687. DOI: 10.1038/S41586-020-2772-0 (cit. on pp. 9, 33).
- [66] B. J. Bosch et al. "The Coronavirus Spike Protein Is a Class I Virus Fusion Protein: Structural and Functional Characterization of the Fusion Core Complex". In: *Journal of Virology* 77.16 (2003-08). Publisher: American Society for Microbiology, pp. 8801–8811. ISSN: 0022-538X. DOI: 10.1128/JVI.77.16.8801-8811.2003 (cit. on pp. 9, 33).
- [67] Q. Wang et al. "Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2". In: *Cell* 181.4 (2020-05). Publisher: Cell Press, 894–904.e9. ISSN: 10974172. DOI: 10.1016/j.cell.2020.03.045 (cit. on pp. 9, 11, 33, 34, 46).
- [68] Y. Yuan et al. "Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains". en. In: *Nature Communications* 8.1 (2017-04), p. 15092. ISSN: 2041-1723. DOI: 10.1038/ncomms15092. URL: <http://www.nature.com/articles/ncomms15092> (visited on 2022-09-12) (cit. on p. 9).
- [69] M. Gui et al. "Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding". en. In: *Cell Research* 27.1 (2017-01), pp. 119–129. ISSN: 1001-0602, 1748-7838. DOI: 10.1038/cr.2016.152. URL: <http://www.nature.com/articles/cr2016152> (visited on 2022-09-12) (cit. on p. 9).
- [70] M. Gur et al. "Conformational transition of SARS-CoV-2 spike glycoprotein between its closed and open states". In: *The Journal of chemical physics* 153.7 (2020) (cit. on p. 9).
- [71] Y. Gong et al. "The glycosylation in SARS-CoV-2 and its receptor ACE2". In: *Signal Transduction and Targeted Therapy* 6.1 (2021), p. 396 (cit. on p. 9).
- [72] Y. Watanabe et al. "Exploitation of glycosylation in enveloped virus pathobiology". In: *Biochimica et Biophysica Acta (BBA)-General Subjects* 1863.10 (2019), pp. 1480–1497 (cit. on p. 9).
- [73] Y. Watanabe et al. "Vulnerabilities in coronavirus glycan shields despite extensive glycosylation". In: *Nature communications* 11.1 (2020), p. 2688 (cit. on p. 9).
- [74] R. Raman et al. "Glycan–protein interactions in viral pathogenesis". In: *Current opinion in structural biology* 40 (2016), pp. 153–162 (cit. on p. 9).

- [75] M. Yang et al. "Conformational heterogeneity of the HIV envelope glycan shield". In: *Scientific reports* 7.1 (2017), p. 4435 (cit. on p. 9).
- [76] M. Crispin, A. B. Ward, and I. A. Wilson. "Structure and immune recognition of the HIV glycan shield". In: *Annual review of biophysics* 47 (2018), pp. 499–523 (cit. on p. 9).
- [77] Y. Watanabe et al. "Site-specific glycan analysis of the SARS-CoV-2 spike". In: *Science* 369.6501 (2020), pp. 330–333 (cit. on p. 9).
- [78] L. Casalino et al. "Beyond shielding: The roles of glycans in the SARS-CoV-2 spike protein". In: *ACS Central Science* 6.10 (2020-10). Publisher: American Chemical Society, pp. 1722–1734. ISSN: 23747951. DOI: 10.1021/ACSCENTSCI.0C01056/SUPPL_FILE/0C0C01056_SI_006.ZIP (cit. on pp. 9, 12, 33–35).
- [79] M. Valério et al. "SARS-CoV-2 variants impact RBD conformational dynamics and ACE2 accessibility". In: *Frontiers in Medical Technology* 4 (2022-10), p. 1009451. ISSN: 2673-3129. DOI: 10.3389/fmedt.2022.1009451. URL: <https://www.frontiersin.org/articles/10.3389/fmedt.2022.1009451/full> (visited on 2022-10-20) (cit. on p. 10).
- [80] *Science Brief: Emerging SARS-CoV-2 Variants* \textbar CDC. 0000. URL: <https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/scientific-brief-emerging-variants.html> (cit. on pp. 10, 34, 41, 48).
- [81] H. Tegally et al. *Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa*. en. preprint. *Epidemiology*, 2020-12. DOI: 10.1101/2020.12.21.20248640. URL: <http://medrxiv.org/lookup/doi/10.1101/2020.12.21.20248640> (visited on 2021-03-03) (cit. on pp. 10, 34, 44, 48, 51).
- [82] T. Fujino et al. "Novel SARS-CoV-2 variant in travelers from Brazil to Japan". In: *Emerging infectious diseases* 27.4 (2021), p. 1243 (cit. on pp. 10, 48).
- [83] S. Cherian et al. "Sars-cov-2 spike mutations, l452r, t478k, e484q and p681r, in the second wave of covid-19 in Maharashtra, India". In: *Microorganisms* 9.7 (2021-07). Publisher: Multidisciplinary Digital Publishing Institute, p. 1542. ISSN: 20762607. DOI: 10.3390/microorganisms9071542 (cit. on pp. 10, 17, 34, 42, 48, 155).
- [84] L. Wang and G. Cheng. "Sequence analysis of the emerging SARS-CoV-2 variant Omicron in South Africa". In: *Journal of Medical Virology* 94.4 (2022-04). Publisher: John Wiley & Sons, Ltd, pp. 1728–1733. ISSN: 1096-9071. DOI: 10.1002/JMV.27516 (cit. on pp. 10, 34, 44).
- [85] R. Viana et al. "Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa". In: *Nature* 2022 (2022-01). Publisher: Nature Publishing Group, pp. 1–8. ISSN: 1476-4687. DOI: 10.1038/s41586-022-04411-y (cit. on pp. 10, 34).

- [86] D. V. Parums. "World Health Organization (WHO) Variants of concern lineages under monitoring (VOC-LUM) in response to the global spread of lineages and sublineages of omicron, or B. 1.1. 529, SARS-CoV-2". In: *Medical science monitor: international medical journal of experimental and clinical research* 28 (2022), e937676–1 (cit. on p. 10).
- [87] D. Wang et al. "Analysis of the N-glycosylation profiles of the spike proteins from the Alpha, Beta, Gamma, and Delta variants of SARS-CoV-2". In: *Analytical and Bioanalytical Chemistry* (2023), pp. 1–15 (cit. on p. 10).
- [88] A. Shajahan et al. "Site Specific N-and O-glycosylation mapping of the Spike Proteins of SARS-CoV-2 Variants of Concern". In: *Scientific Reports* 13.1 (2023), p. 10053 (cit. on p. 10).
- [89] M. L. Newby et al. "Variations within the glycan shield of SARS-CoV-2 impact viral spike dynamics". In: *Journal of Molecular Biology* 435.4 (2023), p. 167928 (cit. on p. 10).
- [90] J. Lan et al. "Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor". en. In: *Nature* 581.7807 (2020-05), pp. 215–220. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-020-2180-5. URL: <http://www.nature.com/articles/s41586-020-2180-5> (visited on 2021-03-21) (cit. on pp. 10, 11, 33–35, 38, 46, 50).
- [91] R. Yan et al. "Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2". en. In: *Science* 367.6485 (2020-03), pp. 1444–1448. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.abb2762. URL: <https://www.sciencemag.org/lookup/doi/10.1126/science.abb2762> (visited on 2021-03-03) (cit. on pp. 10, 51).
- [92] T. Xiao et al. "A trimeric human angiotensin-converting enzyme 2 as an anti-SARS-CoV-2 agent". In: *Nature Structural & Molecular Biology* 28.2 (2021), pp. 202–209 (cit. on p. 10).
- [93] M. Alenquer et al. "Signatures in SARS-CoV-2 spike protein conferring escape to neutralizing antibodies". In: *PLoS Pathogens* 17.8 (2021-08). Publisher: Public Library of Science, e1009772. ISSN: 15537374. DOI: 10.1371/journal.ppat.1009772 (cit. on pp. 10, 33, 43, 51).
- [94] C. S. Lupala et al. "Computational simulations reveal the binding dynamics between human ACE2 and the receptor binding domain of SARS-CoV-2 spike protein". In: *Quantitative Biology* 9.1 (2021-03). Publisher: Higher Education Press, p. 61. ISSN: 2095-4689. DOI: 10.15302/j-qb-020-0231 (cit. on pp. 10, 33).
- [95] F. F. Yan and F. Gao. "Comparison of the binding characteristics of SARS-CoV and SARS-CoV-2 RBDs to ACE2 at different temperatures by MD simulations". In: *Briefings in Bioinformatics* 22.2 (2021-03). Publisher: Oxford Academic, pp. 1122–1136. ISSN: 14774054. DOI: 10.1093/bib/bbab044 (cit. on pp. 10, 33).

- [96] C. Xu et al. "Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM". In: *Science Advances* 7.1 (2021-01). Publisher: American Association for the Advancement of Science. ISSN: 23752548. DOI: 10.1126/sciadv.abe5575 (cit. on pp. 10, 33, 46).
- [97] L. Cao et al. "De novo design of picomolar SARS-CoV-2 miniprotein inhibitors". In: *Science* 370.6515 (2020-10). Publisher: American Association for the Advancement of Science. ISSN: 10959203. DOI: 10.1126/science.abd9909 (cit. on pp. 10, 33).
- [98] R. Alexpandi et al. "Quinolines-Based SARS-CoV-2 3CLpro and RdRp Inhibitors and Spike-RBD-ACE2 Inhibitor for Drug-Repurposing Against COVID-19: An in silico Analysis". In: *Frontiers in Microbiology* 11 (2020-07). Publisher: Frontiers, p. 1796. ISSN: 1664-302X. DOI: 10.3389/fmicb.2020.01796 (cit. on pp. 10, 33).
- [99] I. E. Awad et al. "High-throughput virtual screening of drug databanks for potential inhibitors of SARS-CoV-2 spike glycoprotein". In: *Journal of Biomolecular Structure and Dynamics* (2020). Publisher: Taylor & Francis. ISSN: 15380254. DOI: 10.1080/07391102.2020.1835721 (cit. on pp. 10, 33).
- [100] A. K. Padhi et al. "Unraveling the mechanism of arbidol binding and inhibition of SARS-CoV-2: Insights from atomistic simulations". In: *European Journal of Pharmacology* 894 (2021-03). Publisher: Elsevier, p. 173836. ISSN: 18790712. DOI: 10.1016/j.ejphar.2020.173836 (cit. on pp. 10, 33).
- [101] V. Kumar, H. Liu, and C. Wu. "Drug repurposing against SARS-CoV-2 receptor binding domain using ensemble-based virtual screening and molecular dynamics simulations". In: *Computers in Biology and Medicine* 135 (2021-08). Publisher: Pergamon, p. 104634. ISSN: 18790534. DOI: 10.1016/j.compbiomed.2021.104634 (cit. on pp. 10, 11, 33).
- [102] C. N. Patel et al. "Pinpointing the potential hits for hindering interaction of SARS-CoV-2 S-protein with ACE2 from the pool of antiviral phytochemicals utilizing molecular docking and molecular dynamics (MD) simulations". In: *Journal of Molecular Graphics and Modelling* 105 (2021-06). Publisher: Elsevier, p. 107874. ISSN: 18734243. DOI: 10.1016/j.jmgm.2021.107874 (cit. on pp. 10, 33).
- [103] K. Wu et al. "Mechanisms of host receptor adaptation by severe acute respiratory syndrome coronavirus". In: *Journal of Biological Chemistry* 287.12 (2012), pp. 8904–8911 (cit. on p. 10).
- [104] J. Shang et al. "Structural basis of receptor recognition by SARS-CoV-2". In: *Nature* 581.7807 (2020), pp. 221–224 (cit. on p. 10).
- [105] B. Jawad et al. "Key interacting residues between RBD of SARS-CoV-2 and ACE2 receptor: combination of molecular dynamics simulation and density functional calculation". In: *Journal of chemical information and modeling* 61.9 (2021), pp. 4425–4441 (cit. on pp. 10, 11).

- [106] J. Verma and N. Subbarao. "Insilico study on the effect of SARS-CoV-2 RBD hotspot mutants' interaction with ACE2 to understand the binding affinity and stability". In: *Virology* 561 (2021), pp. 107–116 (cit. on pp. 10, 11).
- [107] M. Yuan et al. "Structural and functional ramifications of antigenic drift in recent SARS-CoV-2 variants". In: *Science* 373.6556 (2021), pp. 818–823 (cit. on p. 11).
- [108] A. Khan et al. "Higher infectivity of the SARS-CoV-2 new variants is associated with K417N/T, E484K, and N501Y mutants: an insight from structural data". In: *Journal of cellular physiology* 236.10 (2021), pp. 7045–7057 (cit. on p. 11).
- [109] W. B. Wang et al. "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". In: *Journal of Molecular Graphics and Modelling* 109 (2021), p. 108035 (cit. on p. 11).
- [110] A. Hagen. *How ominous is the Omicron variant (b.1.1.529)?* 2021-12. URL: <https://asm.org/Articles/2021/December/How-Ominous-is-the-Omicron-Variant-B-1-1-529> (cit. on p. 11).
- [111] C. Chen et al. "Computational prediction of the effect of amino acid changes on the binding affinity between SARS-CoV-2 spike RBD and human ACE2". In: *Proceedings of the National Academy of Sciences* 118.42 (2021), e2106480118 (cit. on p. 11).
- [112] F. Li et al. "Structural biology: Structure of SARS coronavirus spike receptor-binding domain complexed with receptor". In: *Science* 309.5742 (2005-09), pp. 1864–1868. ISSN: 00368075. DOI: 10.1126/science.1116480 (cit. on pp. 11, 34, 46).
- [113] T. Zhou et al. "Cryo-EM Structures of SARS-CoV-2 Spike without and with ACE2 Reveal a pH-Dependent Switch to Mediate Endosomal Positioning of Receptor-Binding Domains". In: *Cell Host and Microbe* 28.6 (2020-12). Publisher: Elsevier, 867–879.e5. ISSN: 19346069. DOI: 10.1016/j.chom.2020.11.004 (cit. on pp. 11, 34, 46).
- [114] M. Rapp et al. "Modular basis for potent SARS-CoV-2 neutralization by a prevalent VH1-2-derived antibody class". In: *Cell Reports* 35.1 (2021-04). Publisher: Elsevier. ISSN: 22111247. DOI: 10.1016/j.celrep.2021.108950 (cit. on pp. 11, 34, 46).
- [115] M. A. Tortorici et al. "Ultrapotent human antibodies protect against SARS-CoV-2 challenge via multiple mechanisms". In: *Science* 370.6519 (2020-11). Publisher: American Association for the Advancement of Science, pp. 950–957. ISSN: 10959203. DOI: 10.1126/science.abe3354 (cit. on pp. 11, 34, 46).
- [116] Y. Wu et al. "A noncompeting pair of human neutralizing antibodies block COVID-19 virus binding to its receptor ACE2". In: *Science* 368.6496 (2020-06). Publisher: American Association for the Advancement of Science, pp. 1274–1278. ISSN: 10959203. DOI: 10.1126/science.abc2241 (cit. on pp. 11, 34, 46).

-
- [117] F. Bertoglio et al. "A SARS-CoV-2 neutralizing antibody selected from COVID-19 patients binds to the ACE2-RBD interface and is tolerant to most known RBD mutations". In: *Cell Reports* 36.4 (2021-07). Publisher: Elsevier. ISSN: 22111247. DOI: 10.1016/j.celrep.2021.109433 (cit. on pp. 11, 34, 46).
- [118] J. Kreye et al. "A Therapeutic Non-self-reactive SARS-CoV-2 Antibody Protects from Lung Pathology in a COVID-19 Hamster Model". In: *Cell* 183.4 (2020-11). Publisher: Elsevier, 1058–1069.e19. ISSN: 10974172. DOI: 10.1016/j.cell.2020.09.049 (cit. on pp. 11, 34, 46).
- [119] J. Hansen et al. "Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail". In: *Science* 369.6506 (2020-08). Publisher: American Association for the Advancement of Science, pp. 1010–1014. ISSN: 10959203. DOI: 10.1126/science.abd0827 (cit. on pp. 11, 34, 46).
- [120] M. Yuan et al. "Structural basis of a shared antibody response to SARS-CoV-2". In: *Science* 369.6507 (2020-08). Publisher: American Association for the Advancement of Science, pp. 1119–1123. ISSN: 10959203. DOI: 10.1126/science.abd2321 (cit. on pp. 11, 34, 46).
- [121] P. Baral et al. "Mutation-induced changes in the receptor-binding interface of the SARS-CoV-2 Delta variant B.1.617.2 and implications for immune evasion". In: *Biochemical and Biophysical Research Communications* 574 (2021-10). Publisher: Academic Press, pp. 14–19. ISSN: 10902104. DOI: 10.1016/j.bbrc.2021.08.036 (cit. on pp. 12, 34, 40).
- [122] N. Bhattarai et al. "Structural and Dynamical Differences in the Spike Protein RBD in the SARS-CoV-2 Variants B.1.1.7 and B.1.351". In: *Journal of Physical Chemistry B* 125.26 (2021-07). Publisher: American Chemical Society, pp. 7101–7107. ISSN: 15205207. DOI: 10.1021/acs.jpcc.1c01626 (cit. on pp. 12, 34, 40).
- [123] G. Wirnsberger et al. "Clinical grade ACE2 as a universal agent to block SARS-CoV-2 variants". In: *bioRxiv* (2021-09). Publisher: Cold Spring Harbor Laboratory, p. 2021.09.10.459744. DOI: 10.1101/2021.09.10.459744 (cit. on pp. 12, 34, 45).
- [124] G. Nelson et al. "Millisecond-scale molecular dynamics simulation of spike RBD structure reveals evolutionary adaption of SARS-CoV-2 to stably bind ACE2". In: *bioRxiv* (2020-12). ISBN: 10.1101/2020.12.1 Publisher: Cold Spring Harbor Laboratory, p. 2020.12.11.422055. DOI: 10.1101/2020.12.11.422055 (cit. on pp. 12, 34).
- [125] B. Turoňová et al. "In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges". In: *Science* 370.6513 (2020-10). Publisher: American Association for the Advancement of Science, pp. 203–208. ISSN: 10959203. DOI: 10.1126/SCIENCE.ABD5223/SUPPL_FILE/ABD5223S1.MOV (cit. on pp. 12, 34).

- [126] H. Woo et al. "Developing a fully glycosylated full-length SARS-COV-2 spike protein model in a viral membrane". In: *Journal of Physical Chemistry B* 124.33 (2020-08). Publisher: American Chemical Society, pp. 7128–7137. ISSN: 15205207. DOI: 10.1021/ACS.JPCB.0C04553/SUPPL_FILE/JP0C04553_SI_001.MP4 (cit. on pp. 12, 34).
- [127] S. C. Harrison. "Viral membrane fusion". In: *Nature Structural & Molecular Biology* 15.7 (2008). Publisher: Nature Publishing Group, pp. 690–698 (cit. on pp. 12, 116).
- [128] G. B. Melikyan. "Common principles and intermediates of viral protein-mediated fusion: the HIV-1 paradigm". In: *Retrovirology* 5.1 (2008). Publisher: BioMed Central, p. 111 (cit. on p. 12).
- [129] A. Lorin et al. "Implication des peptides de fusion des glycoprotéines de fusion virales de classe I dans la fusion membranaire". In: *BASE* (2007) (cit. on p. 13).
- [130] S. Galdiero et al. "Intracellular delivery: exploiting viral membranotropic peptides". In: *Current drug metabolism* 13.1 (2012), pp. 93–104 (cit. on p. 13).
- [131] B. Apellániz et al. "The three lives of viral fusion peptides". en. In: *Chemistry and Physics of Lipids* 181 (2014-07), pp. 40–55. ISSN: 00093084. DOI: 10.1016/j.chemphyslip.2014.03.003. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0009308414000413> (visited on 2021-01-06) (cit. on p. 13).
- [132] A. Falanga et al. "Membranotropic peptides mediating viral entry". In: *Peptide Science* 110.5 (2018), e24040 (cit. on p. 13).
- [133] J.-M. Crowet et al. "Multi-scale simulation of the simian immunodeficiency virus fusion peptide". In: *The Journal of Physical Chemistry B* 116.46 (2012), pp. 13713–13721 (cit. on p. 13).
- [134] M. S. Freitas et al. "Structure of the Ebola fusion peptide in a membrane-mimetic environment and the interaction with lipid rafts". In: *Journal of Biological Chemistry* 282.37 (2007), pp. 27306–27314 (cit. on p. 13).
- [135] H.-F. Li et al. "Mutagenesis of the fusion peptide-like domain of hepatitis C virus E1 glycoprotein: involvement in cell fusion and virus entry". In: *Journal of biomedical science* 16 (2009), pp. 1–18 (cit. on p. 13).
- [136] L. G. M. Basso et al. "Membranotropic and biological activities of the membrane fusion peptides from SARS-CoV spike glycoprotein: The importance of the complete internal fusion peptide domain". In: *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1863.11 (2021), p. 183697 (cit. on p. 13).
- [137] A. L. Lai et al. "The SARS-CoV fusion peptide forms an extended bipartite fusion platform that perturbs membrane order in a calcium-dependent manner". In: *Journal of molecular biology* 429.24 (2017), pp. 3875–3892 (cit. on p. 13).

- [138] A. Santamaria et al. "Strikingly different roles of SARS-CoV-2 fusion peptides uncovered by neutron scattering". In: *Journal of the American Chemical Society* 144.7 (2022), pp. 2968–2979 (cit. on p. 13).
- [139] S. L. Schaefer, H. Jung, and G. Hummer. "Binding of SARS-CoV-2 fusion peptide to host endosome and plasma membrane". In: *The Journal of Physical Chemistry B* 125.28 (2021), pp. 7732–7741 (cit. on p. 13).
- [140] R. K. Koppiseti, Y. G. Fulcher, and S. R. Van Doren. "Fusion peptide of SARS-CoV-2 spike rearranges into a wedge inserted in bilayered micelles". In: *Journal of the American Chemical Society* 143.33 (2021), pp. 13205–13211 (cit. on p. 13).
- [141] W. Shi et al. "Cryo-EM structure of SARS-CoV-2 postfusion spike in membrane". In: *Nature* (2023), pp. 1–2 (cit. on p. 13).
- [142] R. F. Garry and S. Dash. "Proteomics computational analyses suggest that hepatitis C virus E1 and pestivirus E2 envelope glycoproteins are truncated class II fusion proteins". In: *Virology* 307.2 (2003), pp. 255–265 (cit. on p. 13).
- [143] J. M. White. "Membrane fusion". In: *Science* 258.5084 (1992), pp. 917–924 (cit. on p. 12).
- [144] T. Suárez et al. "Calcium-dependent conformational changes of membrane-bound Ebola fusion peptide drive vesicle fusion". In: *FEBS letters* 535.1-3 (2003), pp. 23–28 (cit. on p. 12).
- [145] S. H. White and W. C. Wimley. "Membrane protein folding and stability: physical principles". In: *Annual review of biophysics and biomolecular structure* 28.1 (1999), pp. 319–365 (cit. on pp. 12, 131).
- [146] W. C. Wimley and S. H. White. "Experimentally determined hydrophobicity scale for proteins at membrane interfaces". In: *Nature structural biology* 3.10 (1996), pp. 842–848 (cit. on p. 13).
- [147] X. Han and L. K. Tamm. "A host–guest system to study structure–function relationships of membrane fusion peptides". In: *Proceedings of the National Academy of Sciences* 97.24 (2000), pp. 13097–13102 (cit. on p. 13).
- [148] X. Han and L. K. Tamm. "pH-dependent self-association of influenza hemagglutinin fusion peptides in lipid bilayers". In: *Journal of molecular biology* 304.5 (2000), pp. 953–965 (cit. on p. 13).
- [149] H. Yao and M. Hong. "Conformation and Lipid Interaction of the Fusion Peptide of the Paramyxovirus PIV5 in Anionic and Negative-Curvature Membranes from Solid-State NMR". en. In: *Journal of the American Chemical Society* 136.6 (2014-02), pp. 2611–2624. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/ja4121956. URL: <https://pubs.acs.org/doi/10.1021/ja4121956> (visited on 2020-11-30) (cit. on pp. 13, 15, 16, 77, 80, 81, 87, 91, 119, 120, 124, 129).

- [150] H. Yao and M. Hong. "Membrane-dependent conformation, dynamics, and lipid interactions of the fusion peptide of the paramyxovirus PIV5 from solid-state NMR". In: *Journal of molecular biology* 425.3 (2013), pp. 563–576 (cit. on p. 13).
- [151] A. L. Lai et al. "Fusion activity of HIV gp41 fusion domain is related to its secondary structure and depth of membrane insertion in a cholesterol-dependent fashion". In: *Journal of molecular biology* 418.1-2 (2012), pp. 3–15 (cit. on p. 13).
- [152] X. Han et al. "Membrane structure and fusion-triggering conformational change of the fusion domain from influenza hemagglutinin". In: *Nature structural biology* 8.8 (2001), pp. 715–720 (cit. on p. 13).
- [153] J. L. Lorieau, J. M. Louis, and A. Bax. "The complete influenza hemagglutinin fusion domain adopts a tight helical hairpin arrangement at the lipid: water interface". In: *Proceedings of the National Academy of Sciences* 107.25 (2010), pp. 11341–11346 (cit. on p. 14).
- [154] K. J. Cross et al. "Composition and functions of the influenza fusion peptide". In: *Protein and peptide letters* 16.7 (2009), pp. 766–778 (cit. on p. 14).
- [155] Q. Huang, C.-L. Chen, and A. Herrmann. "Bilayer conformation of fusion peptide of influenza virus hemagglutinin: a molecular dynamics simulation study". In: *Biophysical Journal* 87.1 (2004), pp. 14–22 (cit. on p. 14).
- [156] J. Li, P. Das, and R. Zhou. "Single mutation effects on conformational change and membrane deformation of influenza hemagglutinin fusion peptides". In: *The Journal of Physical Chemistry B* 114.26 (2010), pp. 8799–8806 (cit. on p. 14).
- [157] S. Légaré and P. Lagüe. "The Influenza Fusion Peptide Adopts a Flexible Flat V Conformation in Membranes". en. In: *Biophysical Journal* 102.10 (2012-05), pp. 2270–2278. ISSN: 00063495. DOI: 10.1016/j.bpj.2012.04.003. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0006349512004158> (visited on 2021-01-18) (cit. on pp. 14, 77, 91, 136).
- [158] S. Légaré and P. Lagüe. "The influenza fusion peptide promotes lipid polar head intrusion through hydrogen bonding with phosphates and N-terminal membrane insertion depth". In: *PROTEINS: Structure, Function, and Bioinformatics* 82.9 (2014). Publisher: Wiley Online Library, pp. 2118–2127 (cit. on pp. 14, 15, 77, 95, 130).
- [159] D. Lousa et al. "Fusing simulation and experiment: The effect of mutations on the structure and activity of the influenza fusion peptide". In: *Scientific reports* 6.1 (2016). Publisher: Nature Publishing Group, pp. 1–14 (cit. on pp. 14, 15, 77, 95, 97, 130, 135).
- [160] B. L. Victor et al. "Self-assembly molecular dynamics simulations shed light into the interaction of the influenza fusion peptide with a membrane bilayer". In: *Journal of chemical information and modeling* 55.4 (2015). Publisher: ACS Publications, pp. 795–805 (cit. on pp. 14, 15, 77, 88, 95, 97, 130, 135, 136).

- [161] D. Lousa and C. M. Soares. "Molecular mechanisms of the influenza fusion peptide: insights from experimental and simulation studies". en. In: *FEBS Open Bio* (2021-11), pp. 2211–5463.13323. ISSN: 2211-5463, 2211-5463. DOI: 10.1002/2211-5463.13323. URL: <https://onlinelibrary.wiley.com/doi/10.1002/2211-5463.13323> (visited on 2021-12-12) (cit. on pp. 14, 97).
- [162] R. Worch et al. "Three conserved residues of influenza fusion peptide alter its behavior at the membrane interface". In: *EUROPEAN BIOPHYSICS JOURNAL WITH BIOPHYSICS LETTERS*. Vol. 46. SPRINGER 233 SPRING ST, NEW YORK, NY 10013 USA. 2017, S392–S392 (cit. on p. 14).
- [163] R. Worch et al. "Charged N-terminus of influenza fusion peptide facilitates membrane fusion". In: *International journal of molecular sciences* 19.2 (2018), p. 578 (cit. on p. 14).
- [164] P. Larsson and P. M. Kasson. "Lipid Tail Protrusion in Simulations Predicts Fusogenic Activity of Influenza Fusion Peptide Mutants and Conformational Models". en. In: *PLoS Computational Biology* 9.3 (2013-03). Ed. by E. Tajkhorshid, e1002950. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1002950. URL: <https://dx.plos.org/10.1371/journal.pcbi.1002950> (visited on 2021-01-06) (cit. on pp. 15, 77, 88, 97).
- [165] H.-S. Yin et al. "Structure of the parainfluenza virus 5 F protein in its metastable, prefusion conformation". en. In: *Nature* 439.7072 (2006-01), pp. 38–44. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature04322. URL: <http://www.nature.com/articles/nature04322> (visited on 2021-04-23) (cit. on pp. 15, 16, 89, 116, 141).
- [166] C. M. Horvath and R. A. Lamb. "Studies on the fusion peptide of a paramyxovirus fusion glycoprotein: roles of conserved residues in cell fusion." In: *Journal of virology* 66.4 (1992). Publisher: Am Soc Microbiol, pp. 2443–2455 (cit. on pp. 15, 76).
- [167] S. G. Peisajovich et al. "Sendai virus N-terminal fusion peptide consists of two similar repeats, both of which contribute to membrane fusion". In: *European journal of biochemistry* 269.17 (2002), pp. 4342–4350 (cit. on p. 15).
- [168] J. E. Donald et al. "Transmembrane orientation and possible role of the fusogenic peptide from parainfluenza virus 5 (PIV5) in promoting fusion". en. In: *Proceedings of the National Academy of Sciences* 108.10 (2011-03), pp. 3958–3963. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1019668108. URL: <http://www.pnas.org/cgi/doi/10.1073/pnas.1019668108> (visited on 2020-11-30) (cit. on pp. 15, 16, 77, 89, 97, 115, 116, 121).
- [169] B. D. Welch et al. "Structure of the cleavage-activated prefusion form of the parainfluenza virus 5 fusion protein". en. In: *Proceedings of the National Academy of Sciences* 109.41 (2012-10), pp. 16672–16677. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1213802109. URL: <http://www.pnas.org/cgi/doi/10.1073/pnas.1213802109> (visited on 2021-04-23) (cit. on pp. 16, 89).

- [170] F. Martín-García et al. "Molecular dynamics analysis of conformational change of paramyxovirus F protein during the initial steps of membrane fusion". In: *Biochemical and biophysical research communications* 420.1 (2012), pp. 42–47 (cit. on pp. 15, 16).
- [171] R. Su, J. Zeng, and B. O'Shaughnessy. "Host cell membrane capture by the SARS-CoV-2 spike protein fusion intermediate". In: *Biophysical Journal* 121.3 (2022), 459a (cit. on pp. 15, 16).
- [172] A. Pabis, R. J. Rawle, and P. M. Kasson. "Influenza hemagglutinin drives viral entry via two sequential intramembrane mechanisms". In: *Proceedings of the National Academy of Sciences* 117.13 (2020). Publisher: National Acad Sciences, pp. 7200–7207 (cit. on pp. 15, 27, 77, 95, 97, 116, 121, 135, 136).
- [173] P. M. Kasson and V. S. Pande. "Control of membrane fusion mechanism by lipid composition: predictions from ensemble molecular dynamics". In: *PLoS computational biology* 3.11 (2007), e220 (cit. on p. 15).
- [174] P. M. Kasson and V. S. Pande. "Structural Basis for Influence of Viral Glycans on Ligand Binding by Influenza Hemagglutinin". en. In: *Biophysical Journal* 95.7 (2008-10), pp. L48–L50. ISSN: 00063495. DOI: 10.1529/biophysj.108.141507. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0006349508784522> (visited on 2022-01-28) (cit. on pp. 15, 130).
- [175] D. Lousa et al. "Effect of pH on the influenza fusion peptide properties unveiled by constant-pH molecular dynamics simulations combined with experiment". en. In: *Scientific Reports* 10.1 (2020-12), p. 20082. ISSN: 2045-2322. DOI: 10.1038/s41598-020-77040-y. URL: <http://www.nature.com/articles/s41598-020-77040-y> (visited on 2021-12-12) (cit. on pp. 15, 95, 97, 130, 135, 136).
- [176] H. Tegally et al. "Detection of a SARS-CoV-2 variant of concern in South Africa". en. In: *Nature* 592.7854 (2021-04), pp. 438–443. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-021-03402-9. URL: <http://www.nature.com/articles/s41586-021-03402-9> (visited on 2022-09-25) (cit. on pp. 17, 155).
- [177] R. A. Lamb and G. D. Parks. "Paramyxoviridae". In: *Fields Virology: Sixth Edition*. Wolters Kluwer Health Adis (ESP), 2013 (cit. on pp. 17, 75, 156).
- [178] C. J. Russell, T. S. Jardetzky, and R. A. Lamb. "Membrane fusion machines of paramyxoviruses: capture of intermediates of fusion". In: *The EMBO Journal* 20.15 (2001). Publisher: John Wiley & Sons, Ltd, pp. 4024–4034 (cit. on pp. 17, 76, 156).
- [179] L. Hernandez et al. "Virus-cell and cell-cell fusion". In: *Annual review of cell and developmental biology* 12.1 (1996). Publisher: Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, pp. 627–661 (cit. on pp. 17, 76, 92, 156).

-
- [180] N. L. Doltsinis. *Molecular Dynamics Beyond the Born-Oppenheimer Approximation: Mixed Quantum–Classical Approaches*. NIC Series 31. 2006 (cit. on p. 19).
 - [181] J. R. Perilla et al. “Molecular dynamics simulations of large macromolecular complexes”. In: *Current opinion in structural biology* 31 (2015), pp. 64–74 (cit. on p. 20).
 - [182] W. Van Gunsteren et al. “The GROMOS96 manual and user guide”. In: *Biomos, Zurich, Switzerland* (1996) (cit. on p. 20).
 - [183] H. J. Berendsen, D. van der Spoel, and R. van Drunen. “GROMACS: A message-passing parallel molecular dynamics implementation”. In: *Computer physics communications* 91.1-3 (1995), pp. 43–56 (cit. on p. 20).
 - [184] O. M. Becker et al. *Computational biochemistry and biophysics*. Crc Press, 2001 (cit. on pp. 20, 24).
 - [185] J. A. Maier et al. “ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB”. en. In: *Journal of Chemical Theory and Computation* 11.8 (2015-08), pp. 3696–3713. ISSN: 1549-9618, 1549-9626. DOI: 10.1021/acs.jctc.5b00255. URL: <https://pubs.acs.org/doi/10.1021/acs.jctc.5b00255> (visited on 2021-03-16) (cit. on pp. 21, 35, 50, 120).
 - [186] J. Huang et al. “CHARMM36m: an improved force field for folded and intrinsically disordered proteins”. In: *Nature Methods* 14.1 (2017-01), pp. 71–73. ISSN: 1548-7105. DOI: 10.1038/nmeth.4067. URL: <https://doi.org/10.1038/nmeth.4067> (cit. on pp. 21, 82).
 - [187] P. C. T. Souza et al. “Martini 3: a general purpose force field for coarse-grained molecular dynamics”. en. In: *Nature Methods* 18.4 (2021-04), pp. 382–388. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/s41592-021-01098-3. URL: <http://www.nature.com/articles/s41592-021-01098-3> (visited on 2021-04-21) (cit. on pp. 21, 27, 81, 87, 119, 137, 141).
 - [188] M. Abraham et al. “GROMACS 2023.3 Manual”. In: (2023). DOI: 10.5281/ZENODO.10017699. URL: <https://zenodo.org/doi/10.5281/zenodo.10017699> (cit. on p. 21).
 - [189] T. Darden, D. York, and L. Pedersen. “Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems”. en. In: *The Journal of Chemical Physics* 98.12 (1993-06), pp. 10089–10092. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.464397. URL: <http://aip.scitation.org/doi/10.1063/1.464397> (visited on 2021-03-16) (cit. on pp. 22, 36, 50).

- [190] I. G. Tironi et al. "A generalized reaction field method for molecular dynamics simulations". en. In: *The Journal of Chemical Physics* 102.13 (1995-04), pp. 5451–5459. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.469273. URL: <https://pubs.aip.org/aip/jcp/article/102/13/5451-5459/481977> (visited on 2023-06-13) (cit. on p. 23).
- [191] A. R. Leach. *Molecular modelling: principles and applications*. Pearson education, 2001 (cit. on pp. 23–25).
- [192] A. Hinchliffe. *Molecular modelling for beginners*. John Wiley & Sons, 2003 (cit. on p. 23).
- [193] W. F. Van Gunsteren and H. J. Berendsen. "Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry". In: *Angewandte Chemie International Edition in English* 29.9 (1990), pp. 992–1023 (cit. on p. 24).
- [194] T. L. Hill. *An introduction to statistical thermodynamics*. Courier Corporation, 1986 (cit. on p. 24).
- [195] H. J. Berendsen et al. "Molecular dynamics with coupling to an external bath". In: *The Journal of chemical physics* 81.8 (1984). Publisher: American Institute of Physics, pp. 3684–3690 (cit. on pp. 25, 35, 50).
- [196] S. J. Marrink and D. P. Tieleman. "Perspective on the Martini model". In: *Chemical Society Reviews* 42.16 (2013), pp. 6801–6822 (cit. on p. 26).
- [197] Y. Duan and P. A. Kollman. "Computational protein folding: From lattice to all-atom". In: *IBM Systems Journal* 40 (2001), pp. 297–309 (cit. on p. 26).
- [198] S. Fraga and J. M. R. Parker. "Modelling of peptide and protein structures". In: *Amino Acids* 7 (1994), pp. 175–202 (cit. on p. 26).
- [199] Y. Sugita and Y. Okamoto. "Replica-exchange molecular dynamics method for protein folding". In: *Chemical Physics Letters* 314 (1999), pp. 141–151 (cit. on p. 26).
- [200] A. Laio and M. Parrinello. "Escaping free-energy minima". en. In: *Proceedings of the National Academy of Sciences* 99.20 (2002-10), pp. 12562–12566. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.202427399. URL: <http://www.pnas.org/cgi/doi/10.1073/pnas.202427399> (visited on 2021-07-16) (cit. on pp. 26, 82).
- [201] S. J. Marrink et al. "The MARTINI force field: coarse grained model for biomolecular simulations". In: *The journal of physical chemistry B* 111.27 (2007). Publisher: ACS Publications, pp. 7812–7824 (cit. on pp. 27, 81, 119).
- [202] F. Di Natale et al. "A massively parallel infrastructure for adaptive multiscale simulations: modeling RAS initiation pathway for cancer". In: *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*. 2019, pp. 1–16 (cit. on p. 27).

- [203] M. Valério et al. "Parainfluenza Fusion Peptide Promotes Membrane Fusion by Assembling into Oligomeric Porelike Structures". en. In: *ACS Chemical Biology* 17.7 (2022-07), pp. 1831–1843. ISSN: 1554-8929, 1554-8937. DOI: 10.1021/acscchembio.2c00208. URL: <https://pubs.acs.org/doi/10.1021/acscchembio.2c00208> (visited on 2022-10-20) (cit. on pp. 27, 115, 116, 121, 122, 128, 135).
- [204] K. G. Andersen et al. "The proximal origin of SARS-CoV-2". In: *Nature medicine* 26.4 (2020). Publisher: Nature Publishing Group, pp. 450–452 (cit. on p. 32).
- [205] F. Wu et al. "A new coronavirus associated with human respiratory disease in China". In: *Nature* 579.7798 (2020-03). Publisher: Nature Research, pp. 265–269. ISSN: 14764687. DOI: 10.1038/s41586-020-2008-3 (cit. on p. 32).
- [206] N. Zhu et al. "A novel coronavirus from patients with pneumonia in China, 2019". In: *New England Journal of Medicine* (2020). Publisher: Mass Medical Soc (cit. on pp. 32, 48, 142).
- [207] F. Zhou et al. "Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study". In: *The Lancet* 395.10229 (2020-03). Publisher: Lancet Publishing Group, pp. 1054–1062. ISSN: 1474547X. DOI: 10.1016/S0140-6736(20)30566-3 (cit. on p. 32).
- [208] W. H. Organization. *WHO Coronavirus (Covid19)*. Pages: 21–49 Publication Title: World Health Organization. 2020. URL: <https://covid19.who.int/> (cit. on p. 32).
- [209] F. Li. "Structure, function, and evolution of coronavirus spike proteins". In: *Annual review of virology* 3 (2016). Publisher: Annual Reviews, pp. 237–261 (cit. on p. 33).
- [210] Y. Wan et al. "Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus". In: *Journal of Virology* 94.7 (2020-01). Publisher: American Society for Microbiology. ISSN: 0022-538X. DOI: 10.1128/jvi.00127-20 (cit. on p. 33).
- [211] A. C. Walls et al. "Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion". In: *Proceedings of the National Academy of Sciences of the United States of America* 114.42 (2017-10). Publisher: National Academy of Sciences, pp. 11157–11162. ISSN: 10916490. DOI: 10.1073/pnas.1708727114 (cit. on p. 33).
- [212] R. D. de Vries et al. "Intranasal fusion inhibitory lipopeptide prevents direct-contact SARS-CoV-2 transmission in ferrets". In: *Science* 371.6536 (2021-03). Publisher: American Association for the Advancement of Science (AAAS), pp. 1379–1382. ISSN: 0036-8075. DOI: 10.1126/science.abf4896 (cit. on p. 33).
- [213] M. Yuan et al. "A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV". In: *Science (New York, N.Y.)* 368.6491 (2020-05). Publisher: Science, pp. 630–633. ISSN: 1095-9203. DOI: 10.1126/SCIENCE.ABB7269 (cit. on p. 33).

- [214] M. Hussain et al. "Structural variations in human ACE2 may influence its binding with SARS-CoV-2 spike protein". In: *Journal of Medical Virology* 92.9 (2020-09). Publisher: John Wiley & Sons, Ltd, pp. 1580–1586. ISSN: 10969071. DOI: 10.1002/jmv.25832 (cit. on p. 33).
- [215] F. Ali et al. "ACE2 coding variants in different populations and their potential impact on SARS-CoV-2 binding affinity". In: *Biochemistry and Biophysics Reports* 24 (2020-12). Publisher: Elsevier, p. 100798. ISSN: 24055808. DOI: 10.1016/j.bbrep.2020.100798 (cit. on p. 33).
- [216] Z. T. Muhseen et al. "Promising terpenes as SARS-CoV-2 spike receptor-binding domain (RBD) attachment inhibitors to the human ACE2 receptor: Integrated computational approach". In: *Journal of Molecular Liquids* 320 (2020-12). Publisher: Elsevier, p. 114493. ISSN: 01677322. DOI: 10.1016/j.molliq.2020.114493 (cit. on p. 33).
- [217] S. S. Abdool Karim and T. de Oliveira. "New SARS-CoV-2 Variants — Clinical, Public Health, and Vaccine Implications". In: *New England Journal of Medicine* 384.19 (2021-05). Publisher: Massachusetts Medical Society, pp. 1866–1868. ISSN: 0028-4793. DOI: 10.1056/nejmc2100362 (cit. on p. 34).
- [218] M. Shah et al. "Mutations in the SARS-CoV-2 spike RBD are responsible for stronger ACE2 binding and poor anti-SARS-CoV mAbs cross-neutralization". In: *Computational and Structural Biotechnology Journal* 18 (2020-01). Publisher: Elsevier B.V., pp. 3402–3414. ISSN: 20010370. DOI: 10.1016/j.csbj.2020.11.002 (cit. on p. 34).
- [219] M. J. Abraham et al. "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers". en. In: *SoftwareX* 1-2 (2015-09), pp. 19–25. ISSN: 23527110. DOI: 10.1016/j.softx.2015.06.001. URL: <https://linkinghub.elsevier.com/retrieve/pii/S2352711015000059> (visited on 2022-10-19) (cit. on p. 35).
- [220] Lindahl et al. *GROMACS 2020.3 Source code*. 2020-07. DOI: 10.5281/ZENODO.3923645. URL: <https://zenodo.org/record/3923645> (visited on 2021-03-16) (cit. on pp. 35, 50, 81, 82, 119, 120, 137).
- [221] P. Mark and L. Nilsson. "Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K". en. In: *The Journal of Physical Chemistry A* 105.43 (2001-11), pp. 9954–9960. ISSN: 1089-5639, 1520-5215. DOI: 10.1021/jp003020w. URL: <https://pubs.acs.org/doi/10.1021/jp003020w> (visited on 2021-03-16) (cit. on pp. 35, 50).
- [222] L. Schrodinger. "The PyMOL molecular graphics system". In: *Version 1* (2015), p. 8 (cit. on pp. 35, 36).

- [223] G. Bussi, D. Donadio, and M. Parrinello. "Canonical sampling through velocity rescaling". en. In: *The Journal of Chemical Physics* 126.1 (2007-01), p. 014101. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.2408420. URL: <http://aip.scitation.org/doi/10.1063/1.2408420> (visited on 2021-03-16) (cit. on pp. 35, 50).
- [224] M. Parrinello and A. Rahman. "Polymorphic transitions in single crystals: A new molecular dynamics method". In: *Journal of Applied physics* 52.12 (1981). Publisher: American Institute of Physics, pp. 7182–7190 (cit. on pp. 36, 50, 81, 82, 120).
- [225] U. Essmann et al. "A smooth particle mesh Ewald method". en. In: *The Journal of Chemical Physics* 103.19 (1995-11), pp. 8577–8593. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.470117. URL: <http://aip.scitation.org/doi/10.1063/1.470117> (visited on 2021-03-16) (cit. on pp. 36, 50).
- [226] B. Hess et al. "LINCS: A Linear Constraint Solver for molecular simulations". In: *Journal of Computational Chemistry* 18.12 (1997-09). Publisher: John Wiley and Sons Inc., pp. 1463–1472. ISSN: 01928651. DOI: 10.1002/(SICI)1096-987X(199709)18:12<textless1463::AID-JCC4>textgreater3.0.CO;2-H (cit. on pp. 36, 50).
- [227] W. Humphrey, A. Dalke, and K. Schulten. "VMD: Visual molecular dynamics". en. In: *Journal of Molecular Graphics* 14.1 (1996-02), pp. 33–38. ISSN: 02637855. DOI: 10.1016/0263-7855(96)00018-5. URL: <https://linkinghub.elsevier.com/retrieve/pii/0263785596000185> (visited on 2021-03-03) (cit. on pp. 36, 52, 90).
- [228] E. F. Pettersen et al. "UCSF Chimera - A visualization system for exploratory research and analysis". In: *Journal of Computational Chemistry* 25.13 (2004-10), pp. 1605–1612. ISSN: 01928651. DOI: 10.1002/jcc.20084 (cit. on p. 36).
- [229] I. T. Jolliffe and J. Cadima. "Principal component analysis: A review and recent developments". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2065 (2016-04). Publisher: The Royal Society Publishing. ISSN: 1364503X. DOI: 10.1098/rsta.2015.0202 (cit. on p. 36).
- [230] I. T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. New York: Springer-Verlag, 2002. ISBN: 0-387-95442-2. DOI: 10.1007/b98835 (cit. on p. 36).
- [231] S. R. Campos and A. M. Baptista. "Conformational analysis in a multidimensional energy landscape: Study of an arginylglutamate repeat". In: *Journal of Physical Chemistry B* 113.49 (2009-12). Publisher: American Chemical Society, pp. 15989–16001. ISSN: 15206106. DOI: 10.1021/jp902991u (cit. on pp. 36, 37).
- [232] R. Mickey and A. C. Rencher. "Methods of Multivariate Analysis". In: *Journal of the American Statistical Association* 93.443 (1998). ISBN: 978-0-471-46172-2 Publisher: J. Wiley, p. 1239. ISSN: 01621459. DOI: 10.2307/2669873 (cit. on p. 36).
- [233] N. Michaud-Agrawal et al. "MDAnalysis: a toolkit for the analysis of molecular dynamics simulations". In: *Journal of computational chemistry* 32.10 (2011). Publisher: Wiley Online Library, pp. 2319–2327 (cit. on pp. 36, 49, 50, 99).

- [234] B. Silverman. *Density Estimation for Statistics and Data Analysis*. Routledge, 2018-02. ISBN: 978-1-315-14091-9. DOI: 10.1201/9781315140919 (cit. on p. 36).
- [235] S. R. Campos and A. M. Baptista. *Molecular Simulation Lab — In-House Software*. 0000. URL: <https://www.itqb.unl.pt/labs/molecular-simulation/in-house-software> (cit. on p. 36).
- [236] O. M. Becker and M. Karplus. “The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics”. In: *Journal of Chemical Physics* 106.4 (1997-06). Publisher: American Institute of PhysicsAIP, pp. 1495–1517. ISSN: 00219606. DOI: 10.1063/1.473299 (cit. on p. 37).
- [237] F. H. Stillinger and T. A. Weber. “Packing structures and transitions in liquids and solids”. In: *Science* 225.4666 (1984-09). Publisher: American Association for the Advancement of Science, pp. 983–989. ISSN: 00368075. DOI: 10.1126/science.225.4666.983 (cit. on p. 37).
- [238] S. Contreras-Riquelme et al. “RIP-MD: a tool to study residue interaction networks in protein molecular dynamics”. en. In: *PeerJ* 6 (2018-12), e5998. ISSN: 2167-8359. DOI: 10.7717/peerj.5998. URL: <https://peerj.com/articles/5998> (visited on 2022-10-20) (cit. on pp. 37, 126).
- [239] P. Shannon et al. “Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks”. en. In: *Genome Research* 13.11 (2003-11), pp. 2498–2504. ISSN: 1088-9051. DOI: 10.1101/gr.1239303. URL: <http://genome.cshlp.org/lookup/doi/10.1101/gr.1239303> (visited on 2022-10-20) (cit. on pp. 37, 126).
- [240] C. A. Banho et al. “Impact of SARS-CoV-2 Gamma lineage introduction and COVID-19 vaccination on the epidemiological landscape of a Brazilian city”. en. In: *Communications Medicine* 2.1 (2022-12), p. 41. ISSN: 2730-664X. DOI: 10.1038/s43856-022-00108-5. URL: <https://www.nature.com/articles/s43856-022-00108-5> (visited on 2022-09-12) (cit. on p. 37).
- [241] P. D. Yadav et al. “Isolation and characterization of SARS-CoV-2 Beta variant from UAE travelers”. en. In: *Journal of Infection and Public Health* 15.2 (2022-02), pp. 182–186. ISSN: 18760341. DOI: 10.1016/j.jiph.2021.12.011. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1876034121004299> (visited on 2022-09-12) (cit. on p. 37).
- [242] S. L. Rath and K. Kumar. “Investigation of the Effect of Temperature on the Structure of SARS-CoV-2 Spike Protein by Molecular Dynamics Simulations”. In: *Frontiers in Molecular Biosciences* 7 (2020-10), p. 583523. ISSN: 2296-889X. DOI: 10.3389/fmolb.2020.583523. URL: <https://www.frontiersin.org/article/10.3389/fmolb.2020.583523/full> (visited on 2022-09-12) (cit. on p. 40).

- [243] N. C. Wu et al. "An Alternative Binding Mode of IGHV3-53 Antibodies to the SARS-CoV-2 Receptor Binding Domain". In: *Cell Reports* 33.3 (2020-10). Publisher: Elsevier. ISSN: 22111247. DOI: 10.1016/j.celrep.2020.108274 (cit. on pp. 43, 46).
- [244] CoVariants. 0000. URL: <https://covariants.org/variants/21K.Omicron> (cit. on p. 43).
- [245] E. Simon-Loriere and O. Schwartz. "Towards SARS-CoV-2 serotypes?" In: *Nature Reviews Microbiology* 2022 20:4 20.4 (2022-02). ISBN: 0123456789 Publisher: Nature Publishing Group, pp. 187–188. ISSN: 1740-1534. DOI: 10.1038/s41579-022-00708-x (cit. on p. 43).
- [246] P. A. Desingu, K. Nagarajan, and K. Dhama. "Emergence of Omicron third lineage BA.3 and its importance". In: *Journal of Medical Virology* 94.5 (2022-05). ISBN: 2198/258129 Publisher: John Wiley & Sons, Ltd, pp. 1808–1810. ISSN: 1096-9071. DOI: 10.1002/JMV.27601 (cit. on p. 44).
- [247] M. McCallum et al. "Molecular basis of immune evasion by the Delta and Kappa SARS-CoV-2 variants." In: *Science (New York, N.Y.)* (2021-08). Publisher: Cold Spring Harbor Laboratory, eabl8506. ISSN: 1095-9203. DOI: 10.1126/science.abl8506 (cit. on p. 45).
- [248] T.-J. Yang et al. "Structure-activity relationships of B.1.617 and other SARS-CoV-2 spike variants". In: *bioRxiv* (2021-09). Publisher: Cold Spring Harbor Laboratory, p. 2021.09.12.459978. DOI: 10.1101/2021.09.12.459978 (cit. on p. 45).
- [249] J. W. Saville et al. "Structural and Biochemical Rationale for Enhanced Spike Protein Fitness in Delta and Kappa SARS-CoV-2 Variants". In: *bioRxiv* (2021-09). Publisher: Cold Spring Harbor Laboratory, p. 2021.09.02.458774. DOI: 10.1101/2021.09.02.458774 (cit. on p. 45).
- [250] F. Tian et al. "N501y mutation of spike protein in sars-cov-2 strengthens its binding to receptor ace2". In: *eLife* 10 (2021-08). Publisher: eLife Sciences Publications, Ltd. ISSN: 2050084X. DOI: 10.7554/eLife.69091 (cit. on p. 45).
- [251] C. Laffeber et al. "Experimental Evidence for Enhanced Receptor Binding by Rapidly Spreading SARS-CoV-2 Variants". In: *Journal of Molecular Biology* 433.15 (2021-07). Publisher: Academic Press, p. 167058. ISSN: 10898638. DOI: 10.1016/j.jmb.2021.167058 (cit. on p. 45).
- [252] P. Supasa et al. "Reduced neutralization of SARS-CoV-2 B.1.1.7 variant by convalescent and vaccine sera". In: *Cell* 184.8 (2021-04). Publisher: Elsevier, 2201–2211.e7. ISSN: 10974172. DOI: 10.1016/j.cell.2021.02.033 (cit. on p. 45).

- [253] d. A. S. Souza et al. "Molecular dynamics analysis of fast-spreading severe acute respiratory syndrome coronavirus 2 variants and their effects in the interaction with human angiotensin-converting enzyme 2". In: *bioRxiv* (2021-06). Publisher: Cold Spring Harbor Laboratory, p. 2021.06.14.448436. DOI: 10.1101/2021.06.14.448436 (cit. on p. 45).
- [254] J. Lan et al. "Structural insights into the SARS-CoV-2 Omicron RBD-ACE2 interaction". en. In: *Cell Research* 32.6 (2022-06), pp. 593–595. ISSN: 1748-7838. DOI: 10.1038/s41422-022-00644-8. URL: <https://www.nature.com/articles/s41422-022-00644-8> (visited on 2022-09-25) (cit. on p. 45).
- [255] W. Dejnirattisai et al. "SARS-CoV-2 Omicron-B.1.1.529 leads to widespread escape from neutralizing antibody responses". In: *Cell* 185.3 (2022-02). Publisher: Cell Press, 467–484.e15. ISSN: 0092-8674. DOI: 10.1016/j.cell.2021.12.046 (cit. on p. 45).
- [256] B. Meng et al. "Altered TMPRSS2 usage by SARS-CoV-2 Omicron impacts infectivity and fusogenicity". In: *Nature* 2022 (2022-02). Publisher: Nature Publishing Group, pp. 1–9. ISSN: 1476-4687. DOI: 10.1038/s41586-022-04474-x (cit. on p. 45).
- [257] E. Socher et al. "Computational decomposition reveals reshaping of the SARS-CoV-2-ACE2 interface among viral variants expressing the N501Y mutation". In: *Journal of Cellular Biochemistry* 122.12 (2021). Publisher: John Wiley & Sons, Ltd, pp. 1863–1872. ISSN: 10974644. DOI: 10.1002/jcb.30142 (cit. on p. 45).
- [258] F. Ali, A. Kasry, and M. Amin. "The new SARS-CoV-2 strain shows a stronger binding affinity to ACE2 due to N501Y mutant". In: *Medicine in Drug Discovery* 10 (2021-06). Publisher: Elsevier, p. 100086. ISSN: 25900986. DOI: 10.1016/j.medidd.2021.100086 (cit. on p. 45).
- [259] B. Luan, H. Wang, and T. Huynh. "Enhanced binding of the N501Y-mutated SARS-CoV-2 spike protein to the human ACE2 receptor: insights from molecular dynamics simulations". In: *FEBS Letters* 595.10 (2021-05). Publisher: John Wiley & Sons, Ltd, pp. 1454–1461. ISSN: 18733468. DOI: 10.1002/1873-3468.14076 (cit. on p. 45).
- [260] W. S. Ahmed, A. M. Philip, and K. H. Biswas. "Decreased Interfacial Dynamics Caused by the N501Y Mutation in the SARS-CoV-2 S1 Spike:ACE2 Complex". In: *Frontiers in Molecular Biosciences* 9 (2022-07), p. 846996. ISSN: 2296-889X. DOI: 10.3389/fmolb.2022.846996. URL: <https://www.frontiersin.org/articles/10.3389/fmolb.2022.846996/full> (visited on 2022-09-12) (cit. on p. 45).
- [261] G. Nelson et al. "Molecular dynamic simulation reveals E484K mutation enhances spike RBD-ACE2 affinity and the 1 combination of E484K, K417N and N501Y mutations (501Y.V2 variant) induces conformational change greater than N501Y mutant alone, potentially resulting in an esc". In: *bioRxiv* (2021-01). Publisher:

- Cold Spring Harbor Laboratory, p. 2021.01.13.426558. DOI: 10.1101/2021.01.13.426558 (cit. on p. 45).
- [262] N. Mandal, A. K. Padhi, and S. L. Rath. "Molecular insights into the differential dynamics of SARS-CoV-2 variants of concern". en. In: *Journal of Molecular Graphics and Modelling* 114 (2022-07), p. 108194. ISSN: 10933263. DOI: 10.1016/j.jmgm.2022.108194. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1093326322000730> (visited on 2022-09-12) (cit. on p. 45).
- [263] A. Norman et al. "Discovery of Cyclic Peptide Ligands to the SARS-CoV-2 Spike Protein Using mRNA Display". In: *ACS Central Science* 7.6 (2021-06). Publisher: American Chemical Society, pp. 1001–1008. ISSN: 2374-7943. DOI: 10.1021/acscentsci.0c01708 (cit. on p. 46).
- [264] W. Jiang et al. "Characterization of MW06, a human monoclonal antibody with cross-neutralization activity against both SARS-CoV-2 and SARS-CoV". In: *mAbs* 13.1 (2021-01). Publisher: Taylor & Francis, p. 1953683. ISSN: 19420870. DOI: 10.1080/19420862.2021.1953683 (cit. on p. 46).
- [265] H. Yao et al. "Rational development of a human antibody cocktail that deploys multiple functions to confer Pan-SARS-CoVs protection". In: *Cell Research* 31.1 (2021-12). Publisher: Nature Publishing Group, pp. 25–36. ISSN: 17487838. DOI: 10.1038/s41422-020-00444-y (cit. on p. 46).
- [266] D. Fu et al. "Structural basis for SARS-CoV-2 neutralizing antibodies with novel binding epitopes". In: *PLoS Biology* 19.5 (2021-05). Publisher: Public Library of Science, e3001209. ISSN: 15457885. DOI: 10.1371/journal.pbio.3001209 (cit. on p. 46).
- [267] Q. Li et al. "The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity". en. In: *Cell* 182.5 (2020-09), 1284–1294.e9. ISSN: 00928674. DOI: 10.1016/j.cell.2020.07.012. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0092867420308771> (visited on 2020-11-30) (cit. on p. 46).
- [268] F. Quaglia et al. "SARS-CoV-2 variants preferentially emerge at intrinsically disordered protein sites helping immune evasion". In: *The FEBS journal* 289.14 (2022). Publisher: Wiley Online Library, pp. 4240–4250 (cit. on p. 46).
- [269] A. J. Greaney et al. *Comprehensive mapping of mutations to the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human serum antibodies*. en. preprint. Microbiology, 2021-01. DOI: 10.1101/2020.12.31.425021. URL: <http://biorxiv.org/lookup/doi/10.1101/2020.12.31.425021> (visited on 2021-03-03) (cit. on pp. 48, 51, 53, 54).

- [270] T. N. Starr et al. "Prospective mapping of viral mutations that escape antibodies used to treat COVID-19". en. In: *Science* 371.6531 (2021-02), pp. 850–854. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.abf9302. URL: <https://www.sciencemag.org/lookup/doi/10.1126/science.abf9302> (visited on 2021-03-04) (cit. on p. 48).
- [271] T. N. Starr et al. "Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding". en. In: *Cell* 182.5 (2020-09), 1295–1310.e20. ISSN: 00928674. DOI: 10.1016/j.cell.2020.08.012. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0092867420310035> (visited on 2021-03-10) (cit. on pp. 48, 51, 53, 54).
- [272] A. S. Luring and E. B. Hodcroft. "Genetic variants of SARS-CoV-2—what do they mean?" In: *Jama* 325.6 (2021), pp. 529–531 (cit. on p. 48).
- [273] M. E. Woolhouse and S. Gowtage-Sequeria. "Host range and emerging and reemerging pathogens". In: *Emerging infectious diseases* 11.12 (2005), p. 1842 (cit. on p. 48).
- [274] B. Korber et al. "Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus". In: *Cell* 182.4 (2020), pp. 812–827 (cit. on p. 48).
- [275] N. G. Davies et al. "Estimated transmissibility and impact of SARS-CoV-2 lineage B. 1.1. 7 in England". In: *Science* 372.6538 (2021), eabg3055 (cit. on p. 49).
- [276] N. G. Davies et al. "Increased mortality in community-tested cases of SARS-CoV-2 lineage B. 1.1. 7". In: *Nature* 593.7858 (2021), pp. 270–274 (cit. on p. 49).
- [277] L. van Dorp et al. "Recurrent mutations in SARS-CoV-2 genomes isolated from mink point to rapid host-adaptation". In: *BioRxiv* (2020), pp. 2020–11 (cit. on p. 49).
- [278] E. C. Thomson et al. "Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity". In: *Cell* 184.5 (2021), pp. 1171–1187 (cit. on p. 49).
- [279] R. E. Chen et al. "Resistance of SARS-CoV-2 variants to neutralization by monoclonal and serum-derived polyclonal antibodies". In: *Nature medicine* 27.4 (2021), pp. 717–726 (cit. on p. 49).
- [280] Y. Liu et al. "Neutralizing activity of BNT162b2-elicited serum". In: *New England Journal of Medicine* 384.15 (2021), pp. 1466–1468 (cit. on p. 49).
- [281] X. Xie et al. "Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera". In: *Nature medicine* 27.4 (2021), pp. 620–621 (cit. on p. 49).
- [282] V. V. Edara et al. "Neutralizing antibodies against SARS-CoV-2 variants after infection and vaccination". In: *Jama* 325.18 (2021), pp. 1896–1898 (cit. on p. 49).

- [283] A. Kuzmina et al. "SARS-CoV-2 spike variants exhibit differential infectivity and neutralization resistance to convalescent or post-vaccination sera". In: *Cell host & microbe* 29.4 (2021), pp. 522–528 (cit. on p. 49).
- [284] H. M. Berman. "The Protein Data Bank". In: *Nucleic Acids Research* 28.1 (2000-01), pp. 235–242. ISSN: 13624962. DOI: 10.1093/nar/28.1.235. URL: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/28.1.235> (visited on 2021-03-10) (cit. on pp. 49, 62).
- [285] R. J. Gowers et al. *MDAnalysis: a Python package for the rapid analysis of molecular dynamics simulations*. Tech. rep. Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2019 (cit. on pp. 49, 50, 99).
- [286] C. R. Harris et al. "Array programming with NumPy". en. In: *Nature* 585.7825 (2020-09), pp. 357–362. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-020-2649-2. URL: <http://www.nature.com/articles/s41586-020-2649-2> (visited on 2021-03-03) (cit. on p. 49).
- [287] J. Shang et al. "Cell entry mechanisms of SARS-CoV-2". en. In: *Proceedings of the National Academy of Sciences* 117.21 (2020-05), pp. 11727–11734. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.2003138117. URL: <http://www.pnas.org/lookup/doi/10.1073/pnas.2003138117> (visited on 2021-03-03) (cit. on p. 51).
- [288] W. Li et al. "Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus". en. In: *Nature* 426.6965 (2003-11), pp. 450–454. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature02145. URL: <http://www.nature.com/articles/nature02145> (visited on 2021-03-03) (cit. on p. 51).
- [289] A. J. Greaney et al. *Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition*. en. preprint. Microbiology, 2020-09. DOI: 10.1101/2020.09.10.292078. URL: <http://biorxiv.org/lookup/doi/10.1101/2020.09.10.292078> (visited on 2021-03-03) (cit. on p. 51).
- [290] S. J. Zost et al. "Potently neutralizing and protective human antibodies against SARS-CoV-2". en. In: *Nature* 584.7821 (2020-08), pp. 443–449. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-020-2548-6. URL: <http://www.nature.com/articles/s41586-020-2548-6> (visited on 2021-03-04) (cit. on p. 51).
- [291] C. M. Voloch et al. *Genomic characterization of a novel SARS-CoV-2 lineage from Rio de Janeiro, Brazil*. en. preprint. Genetic and Genomic Medicine, 2020-12. DOI: 10.1101/2020.12.23.20248598. URL: <http://medrxiv.org/lookup/doi/10.1101/2020.12.23.20248598> (visited on 2021-03-03) (cit. on p. 51).
- [292] A. Chang and R. E. Dutch. "Paramyxovirus fusion and entry: multiple paths to a common end". In: *Viruses* 4.4 (2012). Publisher: Molecular Diversity Preservation International, pp. 613–636 (cit. on pp. 75, 116).

- [293] R. E. Dutch. "Entry and fusion of emerging paramyxoviruses". In: *PLoS Pathog* 6.6 (2010). Publisher: Public Library of Science, e1000881 (cit. on p. 75).
- [294] G. R. Abedi et al. "Estimates of Parainfluenza Virus-Associated Hospitalizations and Cost Among Children Aged Less Than 5 Years in the United States, 1998–2010". en. In: *Journal of the Pediatric Infectious Diseases Society* 5.1 (2016-03), pp. 7–13. ISSN: 2048-7193, 2048-7207. DOI: 10.1093/jpids/piu047. URL: <https://academic.oup.com/jpids/article-lookup/doi/10.1093/jpids/piu047> (visited on 2021-12-30) (cit. on pp. 75, 115).
- [295] J. J. Skehel and D. C. Wiley. "Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin". In: *Annual review of biochemistry* 69.1 (2000). Publisher: Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, pp. 531–569 (cit. on p. 75).
- [296] R. A. Lamb. "Paramyxovirus fusion: a hypothesis for changes". In: *Virology* 197.1 (1993). Publisher: Elsevier, pp. 1–11 (cit. on pp. 75, 76).
- [297] K. A. Baker et al. "Structural basis for paramyxovirus-mediated membrane fusion". In: *Molecular cell* 3.3 (1999). Publisher: Elsevier, pp. 309–319 (cit. on pp. 76, 92).
- [298] D. M. Eckert and P. S. Kim. "Mechanisms of viral membrane fusion and its inhibition". In: *Annual review of biochemistry* 70.1 (2001). Publisher: Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, pp. 777–810 (cit. on pp. 76, 92).
- [299] W. P. Russ and D. M. Engelman. "The GxxxG motif: a framework for transmembrane helix-helix association". In: *Journal of molecular biology* 296.3 (2000). Publisher: Elsevier, pp. 911–919 (cit. on p. 76).
- [300] G. Kleiger et al. "GXXXG and AXXXA: Common alpha-Helical Interaction Motifs in Proteins, Particularly in Extremophiles[†]". en. In: *Biochemistry* 41.19 (2002-05), pp. 5990–5997. ISSN: 0006-2960, 1520-4995. DOI: 10.1021/bi0200763. URL: <https://pubs.acs.org/doi/10.1021/bi0200763> (visited on 2021-04-23) (cit. on p. 76).
- [301] S. Bagai and R. A. Lamb. "A Glycine to Alanine Substitution in the Paramyxovirus SV5 Fusion Peptide Increases the Initial Rate of Fusion". en. In: *Virology* 238.2 (1997-11), pp. 283–290. ISSN: 00426822. DOI: 10.1006/viro.1997.8858. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0042682297988589> (visited on 2021-04-25) (cit. on pp. 76, 89).
- [302] G. P. Pattnaik, G. Meher, and H. Chakraborty. "Exploring the mechanism of viral peptide-induced membrane fusion". In: *Biochemical and Biophysical Roles of Cell Surface Molecules* (2018). Publisher: Springer, pp. 69–78 (cit. on p. 77).

- [303] P. Lagüe, B. Roux, and R. W. Pastor. “Molecular dynamics simulations of the influenza hemagglutinin fusion peptide in micelles and bilayers: conformational analysis of peptide and lipids”. In: *Journal of molecular biology* 354.5 (2005). Publisher: Elsevier, pp. 1129–1141 (cit. on p. 77).
- [304] F. Collu et al. “Assembly of influenza hemagglutinin fusion peptides in a phospholipid bilayer by coarse-grained computer simulations”. In: *Frontiers in molecular biosciences* 2 (2015). Publisher: Frontiers, p. 66 (cit. on p. 77).
- [305] M. Fuhrmans and S. J. Marrink. “Molecular view of the role of fusion peptides in promoting positive membrane curvature”. In: *Journal of the American Chemical Society* 134.3 (2012). Publisher: ACS Publications, pp. 1543–1552 (cit. on p. 77).
- [306] A. L. Lai and J. H. Freed. “HIV gp41 fusion peptide increases membrane ordering in a cholesterol-dependent fashion”. In: *Biophysical journal* 106.1 (2014). Publisher: Elsevier, pp. 172–181 (cit. on pp. 77, 88).
- [307] G. Meher and H. Chakraborty. “Membrane composition modulates fusion by altering membrane properties and fusion peptide structure”. In: *The Journal of membrane biology* 252.4 (2019). Publisher: Springer, pp. 261–272 (cit. on pp. 77, 88).
- [308] H. Chakraborty et al. “Depth-dependent membrane ordering by hemagglutinin fusion peptide promotes fusion”. In: *The Journal of Physical Chemistry B* 121.7 (2017). Publisher: ACS Publications, pp. 1640–1648 (cit. on pp. 77, 88).
- [309] P. M. Kasson, E. Lindahl, and V. S. Pande. “Atomic-resolution simulations predict a transition state for vesicle fusion defined by contact of a few lipid tails”. In: *PLoS Comput Biol* 6.6 (2010). Publisher: Public Library of Science, e1000829 (cit. on pp. 77, 95).
- [310] H. J. Risselada et al. “Line-tension controlled mechanism for influenza fusion”. In: *PLoS One* 7.6 (2012). Publisher: Public Library of Science San Francisco, USA, e38302 (cit. on p. 77).
- [311] A. Holt and J. A. Killian. “Orientation and dynamics of transmembrane peptides: the power of simple models”. In: *European Biophysics Journal* 39.4 (2010). Publisher: Springer, pp. 609–621 (cit. on p. 77).
- [312] Y.-H. Huang et al. “The Biological Activity of the Prototypic Cyclotide Kalata B1 Is Modulated by the Formation of Multimeric Pores”. en. In: *Journal of Biological Chemistry* 284.31 (2009-07), pp. 20699–20707. ISSN: 00219258. DOI: 10.1074/jbc.M109.003384. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0021925817483931> (visited on 2021-11-23) (cit. on pp. 79, 118).
- [313] S. W. Provencher. “CONTIN: a general purpose constrained regularization program for inverting noisy linear algebraic and integral equations”. In: *Computer Physics Communications* 27.3 (1982). Publisher: Elsevier Science, pp. 229–242 (cit. on pp. 79, 119).

- [314] B. J. Berne and R. Pecora. *Dynamic light scattering: with applications to chemistry, biology, and physics*. Dover ed. Mineola, N.Y: Dover Publications, 2000. ISBN: 978-0-486-41155-2 (cit. on pp. 79, 119).
- [315] R. F. Alford et al. "The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design". en. In: *Journal of Chemical Theory and Computation* 13.6 (2017-06), pp. 3031–3048. ISSN: 1549-9618, 1549-9626. DOI: 10.1021/acs.jctc.7b00125. URL: <https://pubs.acs.org/doi/10.1021/acs.jctc.7b00125> (visited on 2021-07-16) (cit. on pp. 80, 119).
- [316] L. Monticelli et al. "The MARTINI coarse-grained force field: extension to proteins". In: *Journal of chemical theory and computation* 4.5 (2008). Publisher: ACS Publications, pp. 819–834 (cit. on pp. 81, 119).
- [317] F. Benedetti and C. Loison. "Comment to: "Martini straight: Boosting performance using a shorter cutoff and GPUs" by D.H. de Jong, S. Baoukina, H.I. Ingólfsson, and S.J. Marrink". en. In: *Computer Physics Communications* 228 (2018-07), pp. 146–151. ISSN: 00104655. DOI: 10.1016/j.cpc.2018.02.003. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0010465518300316> (visited on 2021-10-19) (cit. on pp. 81, 120).
- [318] T. A. Wassenaar et al. "Computational lipidomics with insane: a versatile tool for generating custom membranes for molecular simulations". In: *Journal of chemical theory and computation* 11.5 (2015). Publisher: ACS Publications, pp. 2144–2155 (cit. on pp. 82, 100, 120, 125, 139–141).
- [319] S. Jo, T. Kim, and W. Im. "Automated builder and database of protein/membrane complexes for molecular dynamics simulations". In: *PloS one* 2.9 (2007). Publisher: Public Library of Science, e880 (cit. on pp. 82, 120).
- [320] S. Jo et al. "CHARMM-GUI: a web-based graphical user interface for CHARMM". In: *Journal of computational chemistry* 29.11 (2008). Publisher: Wiley Online Library, pp. 1859–1865 (cit. on pp. 82, 120).
- [321] S. Jo et al. "CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes". In: *Biophysical journal* 97.1 (2009). Publisher: Elsevier, pp. 50–58 (cit. on pp. 82, 120).
- [322] D. J. Evans and B. L. Holian. "The Nose–Hoover thermostat". en. In: *The Journal of Chemical Physics* 83.8 (1985-10), pp. 4069–4074. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.449071. URL: <http://aip.scitation.org/doi/10.1063/1.449071> (visited on 2021-04-23) (cit. on pp. 82, 120).
- [323] G. A. Tribello et al. "PLUMED 2: New feathers for an old bird". en. In: *Computer Physics Communications* 185.2 (2014-02), pp. 604–613. ISSN: 00104655. DOI: 10.1016/j.cpc.2013.09.018. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0010465513003196> (visited on 2021-07-16) (cit. on p. 82).

- [324] The PLUMED consortium. "Promoting transparency and reproducibility in enhanced molecular simulations". en. In: *Nature Methods* 16.8 (2019-08), pp. 670–673. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/s41592-019-0506-8. URL: <http://www.nature.com/articles/s41592-019-0506-8> (visited on 2021-07-16) (cit. on p. 82).
- [325] O. Maier, V. Oberle, and D. Hoekstra. "Fluorescent lipid probes: some properties and applications (a review)". en. In: *Chemistry and Physics of Lipids* 116.1-2 (2002-06), pp. 3–18. ISSN: 00093084. DOI: 10.1016/S0009-3084(02)00017-8. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0009308402000178> (visited on 2021-04-21) (cit. on p. 83).
- [326] L. G. M. Basso et al. "SARS-CoV fusion peptides induce membrane surface ordering and curvature". en. In: *Scientific Reports* 6.1 (2016-12), p. 37131. ISSN: 2045-2322. DOI: 10.1038/srep37131. URL: <http://www.nature.com/articles/srep37131> (visited on 2021-04-21) (cit. on p. 83).
- [327] R. M. Epand. "Fusion peptides and the mechanism of viral fusion". en. In: *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1614.1 (2003-07), pp. 116–121. ISSN: 00052736. DOI: 10.1016/S0005-2736(03)00169-X. URL: <https://linkinghub.elsevier.com/retrieve/pii/S000527360300169X> (visited on 2021-04-21) (cit. on pp. 83, 86).
- [328] T. N. Figueira et al. "Quantitative analysis of molecular partition towards lipid membranes using surface plasmon resonance". en. In: *Scientific Reports* 7.1 (2017-05), p. 45647. ISSN: 2045-2322. DOI: 10.1038/srep45647. URL: <http://www.nature.com/articles/srep45647> (visited on 2021-04-21) (cit. on pp. 85, 122).
- [329] N. C. Santos, M. Prieto, and M. A. R. B. Castanho. "Interaction of the Major Epitope Region of HIV Protein gp41 with Membrane Model Systems. A Fluorescence Spectroscopy Study[†]". en. In: *Biochemistry* 37.24 (1998-06), pp. 8674–8682. ISSN: 0006-2960, 1520-4995. DOI: 10.1021/bi9803933. URL: <https://pubs.acs.org/doi/10.1021/bi9803933> (visited on 2021-04-21) (cit. on p. 85).
- [330] N. C. Santos, M. Prieto, and M. A. Castanho. "Quantifying molecular partition into model systems of biomembranes: an emphasis on optical spectroscopic methods". en. In: *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1612.2 (2003-06), pp. 123–135. ISSN: 00052736. DOI: 10.1016/S0005-2736(03)00112-3. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0005273603001123> (visited on 2021-10-19) (cit. on p. 86).
- [331] Y. Xia, J. Sun, and D. Liang. "Aggregation, Fusion, and Leakage of Liposomes Induced by Peptides". en. In: *Langmuir* 30.25 (2014-07), pp. 7334–7342. ISSN: 0743-7463, 1520-5827. DOI: 10.1021/la501618f. URL: <https://pubs.acs.org/doi/10.1021/la501618f> (visited on 2021-04-21) (cit. on p. 86).

- [332] S. Nir, F. Nicol, and F. C. Szoka. "Surface aggregation and membrane penetration by peptides: relation to pore formation and fusion". en. In: *Molecular Membrane Biology* 16.1 (1999-01), pp. 95–101. ISSN: 0968-7688, 1464-5203. DOI: 10.1080/096876899294814. URL: <http://www.tandfonline.com/doi/full/10.1080/096876899294814> (visited on 2021-04-21) (cit. on p. 86).
- [333] A. L. Lai and L. K. Tamm. "Locking the Kink in the Influenza Hemagglutinin Fusion Domain Structure*". en. In: *Journal of Biological Chemistry* 282.33 (2007-08), pp. 23946–23956. ISSN: 00219258. DOI: 10.1074/jbc.M704008200. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0021925820543792> (visited on 2021-01-18) (cit. on p. 91).
- [334] J. B. Johnson, A. P. Schmitt, and G. D. Parks. "Point Mutations in the Paramyxovirus F Protein That Enhance Fusion Activity Shift the Mechanism of Complement-Mediated Virus Neutralization". en. In: *Journal of Virology* 87.16 (2013-08), pp. 9250–9259. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/JVI.01111-13. URL: <https://journals.asm.org/doi/10.1128/JVI.01111-13> (visited on 2022-02-18) (cit. on p. 94).
- [335] M. N. Melo et al. "High-Throughput Simulations Reveal Membrane-Mediated Effects of Alcohols on MscL Gating". en. In: *Journal of the American Chemical Society* 139.7 (2017-02), pp. 2664–2671. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/jacs.6b11091. URL: <https://pubs.acs.org/doi/10.1021/jacs.6b11091> (visited on 2021-04-21) (cit. on pp. 94, 99).
- [336] L. K. Tamm and X. Han. "Viral fusion peptides: a tool set to disrupt and connect biological membranes". In: *Bioscience reports* 20.6 (2000). Publisher: Portland Press Ltd., pp. 501–518 (cit. on p. 95).
- [337] M. Michalski and P. Setny. "Membrane-Bound Configuration and Lipid Perturbing Effects of Hemagglutinin Subunit 2 N-Terminus Investigated by Computer Simulations". In: *Frontiers in Molecular Biosciences* 9 (2022-01), p. 826366. ISSN: 2296-889X. DOI: 10.3389/fmolb.2022.826366. URL: <https://www.frontiersin.org/articles/10.3389/fmolb.2022.826366/full> (visited on 2022-02-18) (cit. on p. 97).
- [338] W. Kabsch and C. Sander. "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features". In: *Biopolymers: Original Research on Biomolecules* 22.12 (1983). Publisher: Wiley Online Library, pp. 2577–2637 (cit. on p. 99).
- [339] M. Bansal, S. Kumart, and R. Velavan. "HELANAL: A Program to Characterize Helix Geometry in Proteins". en. In: *Journal of Biomolecular Structure and Dynamics* 17.5 (2000-04), pp. 811–819. ISSN: 0739-1102, 1538-0254. DOI: 10.1080/07391102.2000.10506570. URL: <http://www.tandfonline.com/doi/abs/10.1080/07391102.2000.10506570> (visited on 2021-04-21) (cit. on p. 99).

- [340] H.-S. Yin et al. "Structure of the uncleaved ectodomain of the paramyxovirus (hPIV3) fusion protein". In: *Proceedings of the National Academy of Sciences* 102.26 (2005), pp. 9288–9293 (cit. on pp. 116, 141).
- [341] A. Moscona et al. "Entry of parainfluenza virus into cells as a target for interrupting childhood respiratory disease". In: *The Journal of clinical investigation* 115.7 (2005), pp. 1688–1698 (cit. on p. 116).
- [342] R. K. Plemper. "Cell entry of enveloped viruses". In: *Current opinion in virology* 1.2 (2011), pp. 92–100 (cit. on p. 116).
- [343] E. C. Smith et al. "Viral entry mechanisms: the increasing diversity of paramyxovirus entry". In: *The FEBS journal* 276.24 (2009), pp. 7217–7227 (cit. on p. 116).
- [344] J. K. Spinti, F. N. Nunes, and M. N. Melo. "Room for improvement in the initial martini 3 parameterization of peptide interactions". In: *Chemical Physics Letters* 819 (2023), p. 140436 (cit. on pp. 120, 139).
- [345] P. Mark and L. Nilsson. "Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K". In: *The Journal of Physical Chemistry A* 105.43 (2001), pp. 9954–9960 (cit. on p. 120).
- [346] O. S. Smart et al. "HOLE: A program for the analysis of the pore dimensions of ion channel structural models". en. In: *Journal of Molecular Graphics* 14.6 (1996-12), pp. 354–360. ISSN: 02637855. DOI: 10.1016/S0263-7855(97)00009-X. URL: <https://linkinghub.elsevier.com/retrieve/pii/S026378559700009X> (visited on 2023-05-10) (cit. on p. 128).
- [347] O. Smart, J. Goodfellow, and B. Wallace. "The pore dimensions of gramicidin A". en. In: *Biophysical Journal* 65.6 (1993-12), pp. 2455–2460. ISSN: 00063495. DOI: 10.1016/S0006-3495(93)81293-1. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0006349593812931> (visited on 2023-05-10) (cit. on p. 128).
- [348] Y. Zhang et al. "Lipid acyl chain protrusion induced by the influenza virus hemagglutinin fusion peptide detected by NMR paramagnetic relaxation enhancement". In: *Biophysical Chemistry* 299 (2023), p. 107028 (cit. on p. 130).
- [349] J. L. Nieva and T. Suárez. "Hydrophobic-at-interface regions in viral fusion protein ectodomains". In: *Bioscience Reports* 20 (2000), pp. 519–533 (cit. on p. 131).
- [350] H. Qiao et al. "A specific point mutant at position 1 of the influenza hemagglutinin fusion peptide displays a hemifusion phenotype". In: *Molecular biology of the cell* 10.8 (1999), pp. 2759–2769 (cit. on p. 131).
- [351] A. L. Lai et al. "Fusion peptide of influenza hemagglutinin requires a fixed angle boomerang structure for activity". In: *Journal of Biological Chemistry* 281.9 (2006), pp. 5760–5770 (cit. on p. 131).

- [352] R. Worch et al. "Charged N-terminus of influenza fusion peptide facilitates membrane fusion". In: *International journal of molecular sciences* 19.2 (2018). Publisher: Multidisciplinary Digital Publishing Institute, p. 578 (cit. on p. 135).
- [353] A. Joardar, G. P. Pattnaik, and H. Chakraborty. "Mechanism of membrane fusion: Interplay of lipid and peptide". In: *The Journal of Membrane Biology* 255.2-3 (2022), pp. 211–224 (cit. on pp. 135, 147).
- [354] V. Markin, M. Kozlov, and V. Borovjagin. "On the theory of membrane fusion. The stalk mechanism". In: *Gen. Physiol. Biophys* 3.5 (1984), pp. 361–377 (cit. on pp. 135, 147).
- [355] H. J. Risselada, G. Bubnis, and H. Grubmüller. "Expansion of the fusion stalk and its implication for biological membrane fusion". In: *Proceedings of the National Academy of Sciences* 111.30 (2014), pp. 11043–11048 (cit. on pp. 135, 147).
- [356] S. A. Akimov et al. "Continuum models of membrane fusion: Evolution of the theory". In: *International journal of molecular sciences* 21.11 (2020), p. 3875 (cit. on pp. 135, 137, 147).
- [357] B. M. Bruininks et al. "A molecular view on the escape of lipoplexed DNA from the endosome". en. In: *eLife* 9 (2020-04), e52012. ISSN: 2050-084X. DOI: 10.7554/eLife.52012. URL: <https://elifesciences.org/articles/52012> (visited on 2022-01-29) (cit. on pp. 136, 140, 143, 148).
- [358] L. Borges-Araújo et al. "Martini 3 Coarse-Grained Force Field for cholesterol". In: *Journal of Chemical Theory and Computation* (2023) (cit. on p. 137).
- [359] Z. Saud et al. "The SARS-CoV2 envelope differs from host cells, exposes procoagulant lipids, and is disrupted in vivo by oral rinses". In: *Journal of Lipid Research* 63.6 (2022) (cit. on pp. 138, 143).
- [360] H.-D. Klenk and P. W. Choppin. "Lipids of plasma membranes of monkey and hamster kidney cells and of parainfluenza virions grown in these cells". In: *Virology* 38.2 (1969), pp. 255–268 (cit. on pp. 138, 143).
- [361] H. I. Ingólfsson et al. "Capturing biologically complex tissue-specific membranes at different levels of compositional complexity". In: *The Journal of Physical Chemistry B* 124.36 (2020), pp. 7819–7829 (cit. on pp. 138, 143).
- [362] W. Pezeshkian et al. "Backmapping triangulated surfaces to coarse-grained membrane models". In: *Nature communications* 11.1 (2020), p. 2296 (cit. on pp. 137, 139, 141).
- [363] X. Fan et al. *Cryo-EM analysis of the post-fusion structure of the SARS-CoV spike glycoprotein*. *Nat Commun* 11: 3618. 2020 (cit. on p. 138).

- [364] A. B. Poma, M. Cieplak, and P. E. Theodorakis. "Combining the MARTINI and structure-based coarse-grained approaches for the molecular dynamics studies of conformational transitions in proteins". In: *Journal of Chemical Theory and Computation* 13.3 (2017), pp. 1366–1374 (cit. on p. 139).
- [365] M. I. Mahmood, A. B. Poma, and K.-i. Okazaki. "Optimizing Gō-MARTINI coarse-grained model for F-BAR protein on lipid membrane". In: *Frontiers in molecular biosciences* 8 (2021), p. 619381 (cit. on p. 139).
- [366] Z. Liu et al. "Mapping mechanostable pulling geometries of a therapeutic anticalin/CTLA-4 protein complex". In: *Nano letters* 22.1 (2021), pp. 179–187 (cit. on p. 139).
- [367] P. C. Kroon et al. "Martinize2 and Vermouth: Unified Framework for Topology Generation". In: *arXiv preprint arXiv:2212.01191* (2022) (cit. on p. 139).
- [368] D. F. Barreto-Vieira et al. "SARS-CoV-2: ultrastructural characterization of morphogenesis in an in vitro system". In: *Viruses* 14.2 (2022), p. 201 (cit. on p. 142).
- [369] R. Rand and V. Parsegian. "Hydration forces between phospholipid bilayers". In: *Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes* 988.3 (1989), pp. 351–376 (cit. on p. 143).
- [370] S. Leikin et al. "Hydration forces". In: *Annual review of physical chemistry* 44.1 (1993), pp. 369–395 (cit. on p. 143).
- [371] J. N. Israelachvili and R. M. Pashley. "Molecular layering of water at surfaces and origin of repulsive hydration forces". In: *Nature* 306.5940 (1983), pp. 249–250 (cit. on p. 143).
- [372] B. M. H. Bruininks, T. A. Wassenaar, and I. Vattulainen. "Unbreaking Assemblies in Molecular Simulations with Periodic Boundaries". In: *Journal of Chemical Information and Modeling* (2023-05). DOI: 10.1021/acs.jcim.2c01574. URL: <https://doi.org/10.1021/acs.jcim.2c01574> (cit. on pp. 144, 145, 151–154).
- [373] J. Nikolaus et al. "Direct visualization of large and protein-free hemifusion diaphragms". In: *Biophysical journal* 98.7 (2010), pp. 1192–1199 (cit. on p. 147).
- [374] J. M. Warner and B. O'Shaughnessy. "Evolution of the hemifused intermediate on the pathway to membrane fusion". In: *Biophysical journal* 103.4 (2012), pp. 689–701 (cit. on p. 147).
- [375] H. Kim, B. Fábián, and G. Hummer. "Neighbor List Artifacts in Molecular Dynamics Simulations". In: (2023) (cit. on p. 148).



Oeiras, December, 2023

Unlocking Viral Entry: The Key Role of Fusion Proteins

Mariana Valério



IBRAG
instituto de biologia
roberto alcântara gonçalves