



Ana Carolina Carvalho Buga
Bachelor's degree in Biochemistry

Spike glycoprotein: molecular characterization and inactivation of a key player in SARS-CoV-2 infection

Dissertation to obtain the master's degree in Biochemistry for Health

Supervisor: Dr. Diana Lousa, Researcher, ITQB-NOVA
Co-supervisor: Prof. Dr. Ana Salomé Veiga, Assistant Professor, IMM-FMUL

[Fevereiro 2022]



Ana Carolina Carvalho Buga
Bachelor's degree in Biochemistry

Spike glycoprotein: molecular characterization and inactivation of a key player in SARS-CoV-2 infection

Dissertation to obtain the master's degree in Biochemistry for Health

Supervisor: Dr. Diana Lousa, Researcher, ITQB-UNL
Co-supervisor: Prof. Dr. Ana Salomé Veiga, Assistant Professor, IMM-FMUL

This work was supported by Fundação para a Ciência e Tecnologia (FCT) through the grant PTDC/CCI-BIO/28200/2017 and by Project LISBOA-01-0145-FEDER-007660 (Microbiologia Molecular, Estrutural e Celular) funded by FEDER and FCT



Júri:

Presidente: Prof. Doutor(a) [Nome Completo]
Arguente (s): Doutor Manuel Nuno de Sousa Pereira Simões de Melo
Vogal (ais): Prof. Doutor(a) [Nome Completo]

[Local de realização das provas públicas de defesa da dissertação]

[Mês e 2022]

Spike glycoprotein: molecular characterization and inactivation of a key player in SARS-CoV-2 infection

Copyright © Ana Carolina Carvalho Buga, Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa.

O Instituto de Tecnologia Química e Biológica António Xavier e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

Acknowledgements

At this point of my master thesis, I would like to acknowledge the people who contributed crucially for the success of this journey.

First, I would like to thank my supervisor Dr. Diana Lousa for the amazing opportunity to develop the current project and for giving me the flexibility needed to pursue the subjects I am interested in. I specially acknowledge her devotion, guidance, the valuable expertise and the time spent addressing my doubts and ideas. Moreover, I would like to thank my co-supervisor Professor Dr. Ana Salomé Veiga for the patience, insightful advices and essential knowledge which contributed to the improvement not only of the present project but also to my scientific journey.

I would also like to acknowledge Dr. João Vicente for the unexpected chance of work alongside him at the Macromolecular Crystallography Unit. I am truly grateful for the support, attention, problem-solving skills taught, the endless brainstorming and of course for the usual healthy scares.

Furthermore, I want to show my gratitude to both Prof. Dr. Cláudio Soares and Prof. Dr. Miguel Castanho for making this step possible by accepting me in their labs and contributing with valuable insights.

Additionally, I want to specially thank a few more people for their precious contributions and shared knowledge, namely Mariana Valério, Diogo A. Mendonça, Marta Alenquer and Alexandra Balola.

To my colleagues and friends of the Cláudio M. Soares (ITQB NOVA) and Miguel Castanho's (iMM) Labs and Macromolecular Crystallography Unit, for the positivity, patience and availability which transforms the work environment completely allowing to face and solve problems more easily.

Finally, I would like to thank my family and my boyfriend for the unconditional love and support given during my academic journey as well as in my personal life, providing me everything I need to accomplish my goals and to become a better human being.

Abstract

The coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), emerged in late 2019 and quickly spread worldwide. So far, over 370 million infections and 5.6 million deaths were reported by the World Health Organization. Although a great effort has been made by the scientific community to study SARS-CoV-2, more research is needed to advance the knowledge on this virus.

SARS-CoV-2 is an enveloped virus and its entry into host cells is mediated by a viral surface protein, the spike glycoprotein (S-protein). This protein is composed of two subunits (S1 and S2) that contain essential domains for the viral entry mechanism, namely the receptor-binding domain (RBD), known to bind to angiotensin-converting enzyme 2 (ACE2) present at the host cells surface, and the fusion peptide (FP), which promotes the fusion between the viral and host cell membranes. However, despite the relevance of the FP for viral entry, its location on the S-protein and amino acid sequence are still unclear.

In this project, the main goals were: 1) to inhibit SARS-CoV-2 entry into host cells by targeting the S-protein and 2) to unravel the FP of this virus.

Two distinct approaches were established to fulfill the first goal: one relied on the development of RBD-interacting chimeric (RIChs) mini-proteins to prevent the RBD-ACE2 interaction, while the other aimed to develop a broad-spectrum vaccine against coronaviruses. Five RIChs were computationally designed using the Rosetta software, by grafting a segment of the ACE2 helix involved in the RBD binding into stable protein scaffolds. RIChs were successfully expressed in *Escherichia coli* (*E. coli*) and purified by chromatographic methods. Surface plasmon resonance (SPR) experiments revealed that three out of five RIChs bound to the RBD, and the best candidate's affinity is on the same order of magnitude as ACE2 affinity. However, none of these proteins exhibited neutralizing activity. An immunogenic construct was also computationally designed using Rosetta software, by grafting the conserved residues within the S2' cleavage site and the FP region into the extremely stable unnatural protein. After expression in *E. coli* and purification, the immunogenic construct was used for immunizing rabbits. The antibodies produced bound to the S-protein with high affinity, however these did not prevent the S2 cleavage.

To fulfill the second goal, four putative FPs are proposed according to previous knowledge combined with computational tools. The experimental and computational characterization of these peptides aimed to identify which one of the putative FPs is required to trigger the membrane fusion process. The experimental assays in membrane model systems revealed that the putative FPs-induced fusion and perturbation differs between the peptides. Molecular dynamics simulations allowed to observe different insertion modes into the membrane and the establishment of specific peptide-membrane interactions. The results obtained suggest the requirement of a fusion module to trigger membrane fusion.

Overall, this work paves the way for the development of antiviral strategies targeting SARS-CoV-2 spike protein.

Keywords: SARS-CoV-2, spike-glycoprotein, fusion peptide (FP), Receptor binding domain (RBD), biophysical characterization.

Resumo

A doença causada por coronavírus 2019 (COVID-19) associada ao coronavírus da síndrome respiratória aguda grave 2 (SARS-CoV-2) surgiu no fim de 2019 e rapidamente se disseminou mundialmente. Até janeiro de 2022, mais de 370 milhões casos positivos e 5.6 milhões de óbitos foram registados de acordo com a Organização Mundial da Saúde. Apesar do grande empenho científico para estudar o SARS-CoV-2, mais investigação científica é necessária para expandir o conhecimento sobre este vírus.

O SARS-CoV-2 é um vírus com invólucro cuja entrada na célula hospedeira é mediada por uma proteína na superfície viral, a *spike glycoprotein* (proteína-S). Esta proteína é constituída pelas subunidades (S1 e S2) que contêm domínios essenciais para o mecanismo de entrada viral, nomeadamente: o domínio de ligação ao receptor (RBD), que se liga à enzima ACE2 presente na superfície da célula hospedeira; e o péptido de fusão (FP), que promove a fusão entre a membrana viral e a membrana da célula hospedeira. Apesar da importância do FP na entrada viral, a sua localização na proteína-S e a sua sequência de aminoácidos permanecem desconhecidas.

Os objetivos principais deste projeto foram: 1) impedir a entrada do SARS-CoV-2 nas células hospedeiras através da inibição da proteína-S e 2) identificar o FP deste vírus.

Para cumprir o primeiro objetivo, duas abordagens foram estabelecidas: numa *RBD-interacting chimeric* (RIChs) *mini-proteins* foram desenvolvidas para bloquearem a interação entre o RBD e a ACE2, na outra uma vacina de largo espectro contra coronavírus foi criada. Cinco RIChs foram desenhadas computacionalmente recorrendo ao software Rosetta, com qual um segmento da hélice da ACE2 envolvido na ligação com o RBD foi enxertado em proteínas estáveis. As RIChs foram expressas em *Escherichia coli* (*E. coli*) e purificadas através de métodos cromatográficos. Ensaios de *Surface plasmon resonance* (SPR) demonstraram que três RIChs se ligam ao RBD, sendo que a melhor candidata apresenta uma afinidade na mesma ordem de magnitude que a afinidade da ACE2. Contudo, nenhuma destas proteínas revelou atividade neutralizante. Uma proteína imunogénica também foi desenhada computacionalmente com o software Rosetta, no qual uma sequência de aminoácidos conservada, correspondente ao sítio de clivagem S2' e da região do FP, foi enxertada na proteína artificial. Após a expressão em *E. coli* e purificação, a proteína imunogénica foi usada na imunização de coelhos. Os anticorpos produzidos demonstraram capacidade de se ligarem à proteína-S com alta afinidade. No entanto, estes não inibiram a clivagem da subunidade S2.

Para atingir o segundo objetivo, quatro FPs putativos são propostos de acordo com conhecimento anterior e com uso de ferramentas computacionais. A caracterização experimental e computacional destes FPs putativos pretendia identificar qual destes é necessário para desencadear o processo de fusão membranar. Os ensaios experimentais com sistemas modelo de membranas revelaram que a fusão e perturbação induzida pelos FPs putativos difere. Simulações de dinâmica molecular permitiram observar diferentes modos de inserção e contactos péptido-membrana. Desta forma, os resultados obtidos indicam que um módulo de fusão é necessário para desencadear o processo de fusão.

Em suma, este trabalho abre caminho para o desenvolvimento de estratégias antivirais tendo como alvo a proteína-S.

Keywords: SARS-CoV-2, spike-glycoprotein, péptido de fusão (FP), domínio de ligação ao recetor (RBD), caracterização biofísica.

Table of contents

Abstract	III
Resumo.....	V
1. Introduction	1
1.1. COVID-19.....	1
1.2. SARS-CoV-2.....	1
1.3. S-protein.....	2
1.3.1. Glycans.....	3
1.3.2. Cleavage sites	3
1.3.3. Receptor binding domain	4
1.3.4. Fusion peptide	6
1.4. SARS-CoV-2 entry mechanism	8
1.5. COVID-19: current treatments and promising therapeutic strategies.....	9
1.5.1. Current treatments and prevention	9
1.5.2. Promising therapeutic targets	10
1.6. Importance of fundamental and applied research	11
2. Objectives	13
3. Materials and methods.....	15
3.1. <i>In silico</i> design.....	15
3.1.1. RBD-interacting chimeric proteins.....	15
3.1.2. Immunogenic construct design	16
3.2. Identification of SARS-CoV-2 putative fusion peptides.....	17
3.3. <i>In silico</i> characterization	17
3.3.1. Molecular dynamics (MD) simulations	17
3.3.2. SARS-CoV-2 FPs MD simulations setup.....	20
3.4. Protein production.....	21
3.4.1. Plasmids and competent cells	21
3.4.2. Expression.....	22
3.5. Protein purification	24
3.5.1. RBD-interacting chimeric mini-proteins	24
3.5.2. Immunogenic construct.....	25
3.5.3. SARS-CoV-2 Fusion Module	25
3.5.4. Protein quantification	26
3.6. Chemical Synthesis of the FP1, FP2 and FP3	26
3.7. Experimental characterization.....	26
3.7.1. Structural characterization	26

Far-UV circular dichroism	26
Dynamic light scattering.....	28
3.7.2. Functional characterization	29
Surface plasmon resonance	29
Dye-free Differential Scanning Fluorimetry	31
Immunization program.....	31
Assessment of the TSSFP-induced antibodies.....	32
Neutralization assays.....	33
Spike glycoprotein cleavage assay.....	33
Liposome preparation and characterization	34
Fluorescence spectroscopy	34
4. Results and discussion	38
4.1. RBD-interacting chimeric proteins	38
4.1.1. <i>In silico</i> design.....	38
RICH 1.....	39
RICH 2.....	40
RICH 3.....	40
RICH 4.....	41
RICH 5.....	41
4.1.2. Expression and purification	42
4.1.3. Structural characterization	49
4.1.4. Functional characterization	52
RICHs affinity for the SARS-CoV-2 RBD	52
RICH 1 and RBD structural stability changes induced by complex formation ...	53
Neutralization ability	55
4.2. Immunogenic construct.....	57
4.2.1. <i>In silico</i> design.....	57
4.2.2. Expression and purification	57
4.2.3. Structural characterization	60
Secondary structure characterization	60
Thermal stability	60
Storage stability.....	61
4.2.4. Functional characterization	62
Evaluation of the antibodies generated	62
Neutralization ability of the antibodies generated	65
Effect of the antibodies in fusion mechanism steps	67

4.3. SARS-CoV-2 putative fusion peptides	69
4.3.1. Identification of SARS-CoV-2 putative fusion peptides	69
4.3.2. Expression and purification of the FM.....	70
4.3.3. FPs aggregation in aqueous solution	71
4.3.4. SARS-CoV-2 putative FPs effect on membrane model systems	73
FPs-induced fusion and perturbation on lipid vesicles.....	73
FPs-induced vesicle aggregation	76
4.3.5. <i>In silico</i> characterization of SARS-CoV-2 putative FPs	78
RMSD	79
Secondary structure	80
Insertion of the FPs in the membrane bilayer.....	85
FPs-membrane bilayer interactions	87
FPs-induced changes in the membrane proprieties.....	91
4.3.6. Comparison with previous studies of SARS-CoV-2 putative FPs	94
5. Conclusions and future perspectives	95
6. References	96
7. Annex.....	102
7.1. Plasmids	102
7.2. FM expression vector produced by FastCloning method	103
7.2.1. PCR amplifications	104
7.3. RICH 1 expression tests.....	105
7.4. Calibration curves	106
7.5. Structural characterization.....	107
7.6. Functional characterization.....	108

Figures index

1. Introduction

Figure 1.1 Structural illustration of SARS-CoV-2 virion.	2
Figure 1.2 SARS-CoV-2 spike-glycoprotein.	3
Figure 1.3 SARS-CoV-2 receptor binding domain (RBD) motifs (PDB ID:6VW1).	4
Figure 1.4 ACE2 complex with SARS-CoV and SARS-CoV-2 RBD.	6
Figure 1.5 Schematic representation of the membrane fusion mediated by viral fusion proteins schematized.	7
Figure 1.6 SARS-CoV-2 FPs proposed in the literature.	8
Figure 1.7 SARS-CoV-2 entry mechanism.	9

2. Objectives

Figure 2.1 Main goals of this project.	13
---	----

3. Materials and methods

Figure 3.1 Immunogenic fragment identification (highlighted in orange) and its conservation among coronaviruses able to infect humans including HCoV-OC43, HCoV-HKU1, MERS-CoV, SARS-CoV-2 and SARS-CoV.	15
Figure 3.2 General representation of a potential energy surface which maps the potential energy as a function of the system configuration.	18
Figure 3.3 Schematized representation of the periodic boundary conditions.	18
Figure 3.4 Origin of the CD effect.	26
Figure 3.5 Schematic representation of the SPR principle and the sensorgram generated.	29
Figure 3.6 Immunization program schedule with the boosts and bleeds days.	30
Figure 3.7 Jablonski diagram illustrates the stages responsible for fluorescence generation.	33

4. Results and discussion

Figure 4.1 RBD-interacting chimeric proteins (RIChs) structure in the context of the SARS-CoV-2 RBD.	37
Figure 4.2 RICH 1-RBD interface.	38
Figure 4.3 Binding interface of RICH 2 and RBD.	39
Figure 4.4 RICH 3 and RBD interface.	40
Figure 4.5 Binding interface of RICH 4 and RBD.	40
Figure 4.6 Binding interface of RICH 5 and RBD.	41
Figure 4.7 Chromatograms and respective SDS-PAGE gels and WB of RICH 1 purification steps.	42
Figure 4.8 Chromatograms of the RICH 2 three-step purification and the final SDS-PAGE gel.	43
Figure 4.9 Chromatograms and respective SDS-PAGE gels of RICH 3 purification steps.	44
Figure 4.10 Chromatograms of the RICH 4 purification and the final SDS-PAGE gel.	45
Figure 4.11 Chromatograms and respective SDS-PAGE gels of RICH 5 purification steps.	46
Figure 4.12 Chromatograms and corresponding SDS-PAGE gels of scaffold 1 purification.	47

Figure 4.13 Chromatograms and respective SDS-PAGE gels and WB of 5XEF purification steps.	48
Figure 4.14 CD spectra of RICHs.	49
Figure 4.15 Content of each secondary structure element predicted by BeStSel web server-based tool for each RICH establish from the CD spectra collected.	50
Figure 4.16 CD spectra collected at 20 °C, 90 °C and after recooling to 20 °C of each RICH.....	51
Figure 4.17 Sensorgrams for the binding of RICHs, BSA and hACE2 to the SARS-CoV-2 RBD obtained by surface plasmon resonance (SPR).	52
Figure 4.18 Thermal stability of RICH 1, RBD and the RICH1-RBD complex monitored using dye-free DSF following the intrinsic fluorescence changes at 330 nm, 350 nm and for the 350/330 ratio.	53
Figure 4.19 First derivative of the fluorescence data collected at 330 nm during the thermal denaturation assay.	54
Figure 4.20 Effect of RICH 1, RICH 3 and RICH 5 in the infection of 293T-ACE2 cells by spike-pseudotyped lentiviruses.	55
Figure 4.21 Designed α -helix of the TSSFP or immunogenic construct structure overlayed with its native scaffold.	56
Figure 4.22 TSSFP expression profile evaluated by WB.	57
Figure 4.23 Chromatograms and respective SDS-PAGE gels of TSSFP purification steps.....	58
Figure 4.24 Secondary structure evaluation of TSSFP.	59
Figure 4.25 TSSFP thermostability evaluation by CD.	60
Figure 4.26 TSSFP storage stability evaluated by CD.	60
Figure 4.27 Variation of the mean residue ellipticity at 222 nm as a function of the storage time.	61
Figure 4.28 Evaluation of the antibodies generated in the immunization program by western blot.	62
Figure 4.29 ELISA result showing the ability of the ABs in the samples collected during the immunization program to bind to the SARS-CoV-2 S-protein.	63
Figure 4.30 Neutralization curves of the samples collected during the immunization program with TSSFP of rabbit 1 and rabbit 2.	65
Figure 4.31 Neutralization curves of the control samples.	66
Figure 4.32 Evaluation of the purified antibodies to prevent S2' cleavage.	66
Figure 4.33 SARS-CoV-2 putative FPs identified.	68
Figure 4.34 FM expression profile evaluated by WB.	69
Figure 4.35 SDS-PAGE gels and chromatogram of FM purification steps.	70
Figure 4.36 DLS profile of samples containing increasing concentrations of FP1, FP2 or FP3 in 10 mM HEPES 150 mM NaCl pH 7.4.	71
Figure 4.37 Evaluation of the FP1, FP2 and FP3 aggregation using ANS probe.	72
Figure 4.38 Evaluation of the FP1-induced fusion and perturbation in LUVs composed of POPC, POPC:POPG (4:1) and POPC:POPG:Chol (3:1:1).	73
Figure 4.39 Evaluation of the FP2-induced fusion and perturbation in LUVs composed of POPC, POPC:POPG (4:1) and POPC:POPG:Chol (3:1:1).	74

Figure 4.40 Evaluation of the FP3-induced fusion and perturbation in LUVs composed of POPC, POPC:POPG (4:1) and POPC:POPG:Chol (3:1:1).	74
Figure 4.41 Evaluation of the FM-induced fusion and perturbation in LUVs composed of POPC, POPC:POPG (4:1) and POPC:POPG:Chol (3:1:1).	75
Figure 4.42 Comparison of the percentage of fusion and leakage induced by each FP at a ratio of lip/pep of 20 on vesicles.	75
Figure 4.43 Representative DLS profile of a mixture containing FP1 or FP2 and LUVs.	76
Figure 4.44 DLS profiles of samples containing either POPC or POPC:POPG (4:1) LUVs with increasing concentration of FP3 and the FP3 profile.	77
Figure 4.45 SARS-CoV-2 spike-protein (PDB ID: 6ZGE, blue) with the four putative FPs highlighted (pink).	78
Figure 4.46 Temporal evolution of the RMSD for all the replicas of each putative FPs.	79
Figure 4.47 Evaluation of the secondary structure of FP1.	80
Figure 4.48 Evaluation of the secondary structure of FP2.	81
Figure 4.49 Evaluation of the secondary structure of FP3.	82
Figure 4.50 Evaluation of the secondary structure of FM.	83
Figure 4.51 Density distribution along the Z-axis of the putative FPs, the phosphate groups and water molecules.	84
Figure 4.52 Average insertion of the FPs per residue in the membrane bilayer considering the phosphate groups of the lower leaflet as the reference.	85
Figure 4.53 POPC and POPS structural division into “Heads” and “Tails”.	86
Figure 4.54 Interactions established between FP1 and the membrane bilayer.	87
Figure 4.55 FP2-membrane bilayer interactions.	88
Figure 4.56 FP3-membrane bilayer interactions.	89
Figure 4.57 Interactions established between FM and the membrane bilayer.	90
Figure 4.58 Order parameters determined for the POPC and POPS tails.	91
Figure 4.59 Average protruding lipids per frame induced by each of the FPs.	92
Figure 4.60 Global membrane thickness determined for each system.	92
Figure 4.61 Snapshots of the FPs-induced perturbation in the membrane bilayer.	92

7. Annex

Figure A1 Maps of the expression plasmid.	101
Figure A2 FastCloning procedure.	102
Figure A3 PCR products analyzed with 1% agarose gel electrophoresis.	103
Figure A4 RICH 1 expression evaluated by WB.	105
Figure A5 Calibration curve of the S75 10/300 GL column and the correspondent equation.	105
Figure A6 Typical CD spectra of each secondary structure element.	106
Figure A7 CD spectra of RICHs and CAPITO fit.	106
Figure A9 CD spectra of TSSFP and CAPITO fit.	107
Figure A8 Sensorgrams for the binding of scaffold 1 and scaffold 5 to the SARS-CoV-2 RBD obtained by surface plasmon resonance.	107

Tables index

3. Materials and methods

Table 3.1 Detailed information on each equilibration step.	20
Table 3.2 Conditions tested for the expression of TSSFP using <i>E.coli</i> BL21(DE3) and an induction OD _{600nm} of 0.6.	21
Table 3.3 Conditions of the FM expression tests.	21
Table 3.4 Expression scale-up conditions.	22

4. Results and discussion

Table 4.1 Overall characteristics of each protein purification procedure, including the oligomeric state.	41
Table 4.2 Binding constants determined by the Biacore Insight Evaluation Software for the proteins which bound to the SARS-CoV-2 RBD.	52
Table 4.3 SARS-CoV-2 putative FPs main characteristics.	68

7. Annex

Table A1 Primers used in the PCR amplification.	103
Table A2 Primers used in the colony PCR amplification.	104
Table A3 Conditions of the RICH 1 expression tests where the induction started at an OD _{600nm} of 0.6 and was conducted at 18°C for 18h.	104
Table A4 Protein standards used to determine the calibration curve.	105

List of abbreviations

Coronavirus disease 2019 (COVID-19)
Severe acute respiratory syndrome coronaviruses 2 (SARS-CoV-2)
RBD-interacting chimeric (RIChs)
Escherichia coli (*E. coli*)
Surface plasmon resonance (SPR)
Spike glycoprotein (S-protein)
Receptor-binding domain (RBD)
Angiotensin-converting enzyme 2 (ACE2)
Fusion peptide (FP)
Acute respiratory distress syndrome (ARDS)
Coronaviruses (CoVs)
Severe acute respiratory syndrome coronavirus (SARS-CoV)
Middle East respiratory syndrome coronavirus (MERS-CoV)
Nonstructural proteins (nsps)
Nucleocapsid protein (N-protein)
Membrane glycoprotein (M-protein)
Envelope protein (E-protein)
N-terminal domain (NTD)
Receptor binding motif (RBM)
Heptad-repeat domains (HR1/HR2)
Transmembrane domain (TM)
Cytoplasmic domain (CT)
Glycosaminoglycans (GAGs)
Transmembrane serine protease 2 (TMPRSS2)
Fusion peptide 1 (FP1)
Fusion peptide 2 (FP2)
Fusion peptide 3 (FP3)
Internal fusion peptide (IFP)
N-hydroxycytidine (NHC)
Interleukin-6 (IL-6)
Shape complementary (Sc)
Binding energy (dG)
Human Coronavirus OC43 (HCoV-OC43)
Human Coronavirus HKU1 (HCoV-HKU1)
Molecular dynamics (MD)
Periodic boundary conditions (PBC)
Positioning of Proteins in Membranes (PPM)
Room temperature (RT)
Isopropyl- β -D-thiogalactoside (IPTG)
Decylmaltoside (DM)
Circular dichroism (CD)
Digital integration time (D.I.T.)
Dye-free differential scanning fluorimetry (dye-freeDSF)

Enzyme linked immunosorbent assays (ELISA)
Neutralization titer (NT)
1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC)
1-Palmitoyl-2-oleoyl-sn-glycero-3-(phospho-rac-(1-glycerol)) (POPG)
Cholesterol (Chol)
Multilamellar vesicles (MLVs)
Fluorescence resonance energy transfer (FRET)
1-anilinonaphthalene-8-sulfonate (ANS)
5,6-carboxyfluorescein (5,6-CF)
Western blot (WB)
Antibodies (ABs)
Root mean square deviation (RMSD)
Cryo-electron microscopy (cryo-EM)

1. Introduction

1.1. COVID-19

The coronavirus disease 2019 (COVID-19) emerged in December 2019 in the Wuhan region in China, where the first patients were identified and epidemiologically associated with the Wuhan wet animal market. Despite the suggested zoonotic origin of COVID-19, it was the human-to-human transmission that allowed the quick spread of the disease worldwide leading to the current pandemic state.¹ On January 30, 2022, over 370 million cases of infections and 5.6 million deaths were reported by the World Health Organization.² COVID-19 originated not only a health crisis but also an economic one³

COVID-19 is an infectious disease, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which can be divided into three main phases: I - asymptomatic incubation period; II - non-severe symptomatic phase; III - severe respiratory symptomatic phase. During the first two phases, an adaptive immune response is necessary to eradicate the virus and prevent the disease progression to more severe stages. Thus, strategies aiming to improve the immune response are essential. When the immune response is compromised, the virus propagates and destroys cells.⁴ The innate inflammation induced by the damaged cells in the lungs originates life-threatening respiratory disorders characteristic of the severe stage, such as acute respiratory distress syndrome (ARDS).¹

Coronaviruses (CoVs) are a group of viruses able to infect different species and known to cause mild to severe respiratory infections in humans. In the twenty-first century, three highly pathogenic CoVs have emerged, becoming public health concerns: severe acute respiratory syndrome coronavirus (SARS-CoV, 2002), Middle East respiratory syndrome coronavirus (MERS-CoV, 2012) and SARS-CoV-2 (2019).⁵

Given the major impact of COVID-19 and of outbreaks cause by CoVs, it is of utmost importance to understand the mechanism of infection of these viruses at the molecular level to prevent it and leverage this knowledge in the development of efficient treatments.

1.2. SARS-CoV-2

SARS-CoV-2 is an enveloped, positive-sense single-stranded RNA virus, with a genome organization similar to other CoVs. Two thirds of the virus genome encode for 16 nonstructural proteins (nsps), which have different roles, such as modulating the host immune system, contributing to evade the immune system and assisting in the viral replication. The remaining one third encodes for accessory proteins with immune modulatory roles and 4 structural proteins: nucleocapsid protein (N-protein), membrane glycoprotein (M-protein), envelope protein (E-protein) and spike glycoprotein (S-protein).⁶ These are assembled as illustrated in Figure 1.1 and play specific roles.

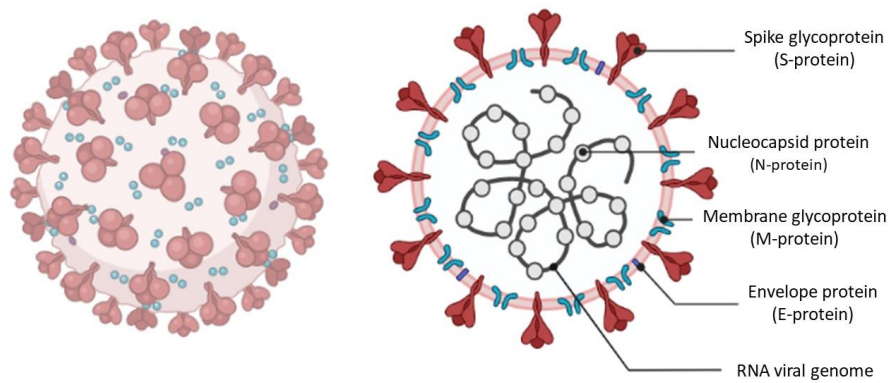


Figure 1.1 | Structural illustration of SARS-CoV-2 virion. Created with BioRender.com⁷.

N-protein's main function is to protect the virus RNA by compacting it. Additionally, the N-protein is also able to manipulate the host cellular machinery and deregulate the cell cycle. M-protein ensures the viruses assemble through protein-protein interactions with the N-protein (M-N), with itself (M-M) and with the S-protein (M-S). The E-protein is responsible for forming the virus outer capsid and has an important impact in the virus virulence factor. The S-protein has two essential roles: facilitate the viral particle and host cells attachment as well as mediate fusion between the viral and host membranes, enabling infection to occur.⁸

1.3. S-protein

CoVs S-protein forms homotrimers protruding from the viral surface giving the "crown"-like appearance from which its designation is derived. This protein is the key for the infection mechanism and therefore plays a pivotal role in the viral pathogenesis.^{8 9}

SARS-CoV-2 S-protein contains 1273 amino acid residues, and is composed by two subunits: S1, responsible for recognition and binding to the host cell receptor, and S2, comprising the machinery responsible for the fusion between host and viral membranes. At the amino acid sequence level, the S2 subunit of SARS-CoV-2 and SARS-CoV share 90% identity, while S1 subunits have only 63% identity. The differences in S1 subunit are believed to be the cause of the higher pathogenesis of SARS-CoV-2.⁹

Each subunit can be divided in functional domains as shown in Figure 1.2. The S1 subunit has 3 domains with specific functions: the N-terminal domain (NTD), which determines the host range; the receptor binding domain (RBD), which recognizes the receptor on the host cell surface; and the receptor binding motif (RBM), which corresponds to the region of the RBD that binds and interacts with the host receptor. For the S2 subunit, 4 domains are described: the fusion peptide (FP), which inserts into the host cell membrane and plays a key role in the fusion between the host cell and the viral membranes; two heptad-repeat domains (HR1/HR2), which form a 6-helix bundle that assists in mediating the membrane fusion; the transmembrane domain (TM), which is essential for the S-protein trimerization and membrane fusion; and the cytoplasmic domain (CT), which helps to anchor the trimer to the viral membrane and is also involved in the cell-cell fusion process.^{8 10}

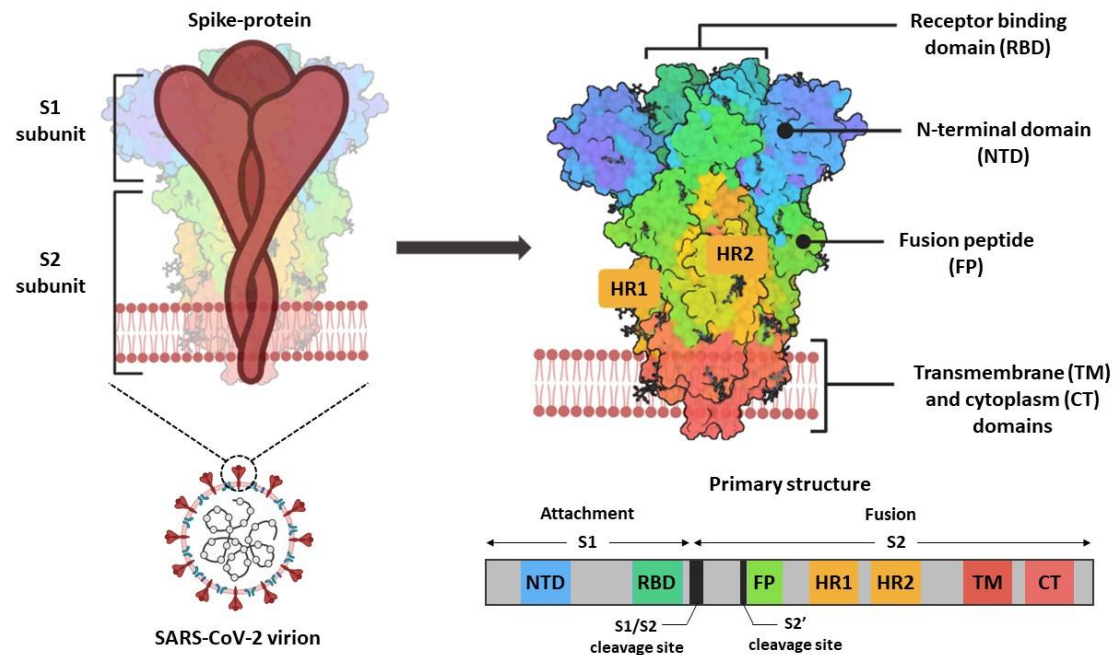


Figure 1.2 | SARS-CoV-2 spike-glycoprotein. (Left) Diagram of the S-protein composed by subunit S1 and S2. (Top right) S-protein structure determined by X-ray crystallography (PDB ID: 6VXX) with functional domains identified. (Bottom right) S-protein primary structure. Created with BioRender.com¹¹.

1.3.1. Glycans

The S-protein is extensively glycosylated by the host glycosylation apparatus. Thus, the glycans role in the viral infection should be considered during the design of biopharmaceuticals tarding the S-protein.

The S-protein glycosylation improves the viral attachment to the host cells¹² in two stages: prior to RBD-ACE2 binding, in which the mannose moieties present in the glycans mediate the interaction with cell surface attachment factors, such as the sialic acid-containing oligosaccharides and glycosaminoglycans (GAGs); and during the RBD-ACE2 interaction, in which the glycans are present in the complex interface.¹³ Furthermore, the S-protein glycosylation provides an additional strategy to evade the immune system¹², since the protein is shielded by the cell “self”-glycans masking the polypeptide epitopes from possible neutralizing antibodies recognition.¹³ This might diminish the efficacy of vaccines targeting the S-protein.

1.3.2. Cleavage sites

During the viral entry, the S-protein undergoes conformational changes induced by two proteolytic cleavages. For some CoVs, a priming cleavage occurs at the S1 and S2 interface (S1/S2 cleavage site), which usually enables the S-protein not only to better bind to the receptor but also to expose inaccessible cleavage sites. Then, a cleavage within the S2 region (S2' cleavage site) occurs, releasing the FP and thus triggering the membrane fusion.¹⁴

Unlike most related CoVs, SARS-CoV-2 has a multibasic motif P-R-R-A-R in the S1/S2 interface, forming a furin cleavage site.⁶ This transmembrane protease is expressed in eukaryotic cells and is able to cleave a broad range of precursors, including growth factors, hormones and cell surface receptors. Furthermore, it is known that furin is essential for the

activation of other viruses, such as HIV and Ebola.¹⁵ The furin cleavage site appears to expand SARS-CoV-2 tropism and transmissibility, since the expression of furin-like proteases is common in several animals.

The S2' cleavage site, located immediately upstream of the FP, is a dibasic site with K-R-S amino acids residues. The host transmembrane serine protease 2 (TMPRSS2) is responsible for the S2' cleavage. Although TMPRSS2 physiological role is yet unknown, this is a widely expressed protein in epithelial cells of the respiratory, urogenital and gastrointestinal tract. Viruses including parainfluenza, SARS-CoV and MERS-CoV are also proteolytically activated by this protease.¹⁵ In TMPRSS2 negative cells, there is evidence that CoVs may use endosomal cathepsin L to infect them. In these cases, the late pathway or viral entry through the endosomal membrane is preferable to the early pathway or fusion with the cell plasma membrane.^{16 17}

1.3.3. Receptor binding domain

Similarly, to SARS-CoV and MERS, SARS-CoV-2 uses the human angiotensin-converting enzyme 2 (hACE2) as the host cell receptor. ACE2 is a transmembrane enzyme highly expressed in capillary rich organs such as lungs, gut, kidneys and brain. This enzyme is crucial for blood pressure regulation since it counterbalances the ACE activity. The ACE cleaves angiotensin I into angiotensin II, which is a vasoconstrictor, while ACE2 cleaves the angiotensin II C-terminal and hydrolyses it into angiotensin, a vasodilator. In addition, ACE2 can cleave other peptides, such as neurotensin and is also involved in the regulation of membrane trafficking of neutral amino acids.¹⁸

The RBD is a crucial domain of the S1 subunit of the S-protein, since it recognizes the ACE2 receptor at the host cell surface. Two motifs build the RBD, as illustrated in Figure 1.3: the core formed by five anti-parallel β -sheets (β 1, β 2, β 3, β 4 and β 7) and three short α -helices (α 1, α 2 and α 3); and the receptor binding motif (RBM), which is composed by two β -sheets (β 5 and β 6), an extended loop and comprises the most important residues for ACE2 binding.¹⁹

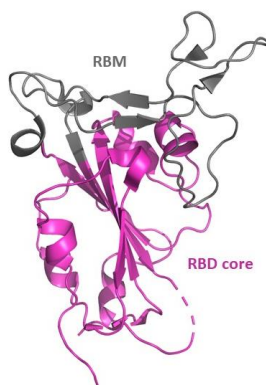


Figure 1.3 | SARS-CoV-2 receptor binding domain (RBD) motifs (PDB ID:6VW1). (Pink) RBD core. (Grey) Receptor binding motif (RBM).

Despite the high similarity between SARS-CoV and SARS-CoV-2, RBD-ACE2 binding affinity of SARS-CoV-2 can be 10 to 20 fold higher relatively to SARS-CoV. Slight conformational changes due to some mutations (Figure 1.4A), as well as a higher affinity of the RBM for the ACE2 receptor, may explain the differences.

SARS-CoV RBD has a Pro-Pro-Ala motif (residues 469-471) in one of the ridge loops of the RBM, forcing it to adopt a sharp turn. In contrast, SARS-CoV-2 contains a Gly-Val-Glu-Gly motif (residues 482-485) in the same region. This four-residue motif composed by two bulky and two flexible residues allows the loop to adopt new conformations. In addition, the main-chain hydrogen bond formed between Asn487 and Ala475 contributes to a more compact conformation of the SARS-CoV-2 RBM ridge moving the Ala475 loop closer to ACE2. As a result of these differences, the SARS-CoV-2 RBM is able to establish more contacts with ACE2 N-terminal helix, including a new hydrogen bond between the RBM Ala475 main chain carboxyl and the ACE2 Ser19 side chain hydroxyl (Figure 1.4B). Moreover, the Phe486 in SARS-CoV-2 RBM inserts into an ACE2 hydrophobic pocket formed by Met82, Leu79 and Tyr83 while in SARS-CoV RBM the corresponding residue is the Leu472 which forms a weaker interaction.²⁰

Two virus-binding hotspots in the ACE2 interface have been previously identified: the Lys31 or hotspot 31 and Lys353 or hotspot 353; both form a salt bridge between ACE2 and the viral residues.²¹ The SARS-CoV hotspot 31 involves Tyr442 support while in the SARS-CoV-2 the mutation to a Leu(455) provides less support due to its less bulky side chain, resulting in the hotspot 31 structural rearrangement. Consequently, the salt bridge between ACE2 Lys31 and Glu35 breaks driving both residues to form hydrogen bonds with the SARS-CoV-2 RBM Gln493. In a similar way, hotspot 353 at the SARS-CoV RBM-ACE2 interface also requires support from Thr487 while the Asn501 at the SARS-CoV-2 RBM-ACE2 interface provides less support. Therefore, ACE2 Lys353 undergoes slight conformational changes, forming a hydrogen bond with the SARS-CoV-2 RBM main chain. Thus, in the SARS-CoV-2 RBM-ACE2 interface, both hotspots suffered an adjustment reducing the need for support given by nearby RBD residues, maintaining the stabilization.

SARS-CoV-2 RBD has six crucial residues to bind to ACE2: Leu455, Phe486, Gln493, Ser494, Asn501 and Tyr505. Five out of these six differ in SARS-CoV RBD corresponding to Tyr442, Leu472, Asn479, Asp480, Thr487 and Tyr491. These differences enhance SARS-CoV-2 recognition and interaction with ACE2, resulting in an infectivity increase (Figure 1.4C).²²

Unexpectedly, despite the improved binding affinity for ACE2, the binding ability of SARS-CoV-2 is not much higher than the SARS-CoV. Several cryoelectron microscopy (cryo-EM) studies revealed that the SARS-CoV-2 RBD adopts diverse conformations of receptor-accessible ("open") and receptor-inaccessible ("closed") states.²³ In the closed state, the three RBDs present a "down" conformation, which makes them inaccessible to ACE2 and hides the S2 core, while in the open state, the RBDs acquire an "up" form becoming accessible to ACE2. The latter is an asymmetric conformation where it is possible to find either one, two or three "up" RBDs. Unlike SARS-CoV RBD, which is mostly in the "up" form, SARS-CoV-2 RBD is more frequently in the "down" form decreasing its accessibility to ACE2. However, this characteristic may have some advantages to evade the immune system by decreasing the RBD exposure.^{24 25 26}

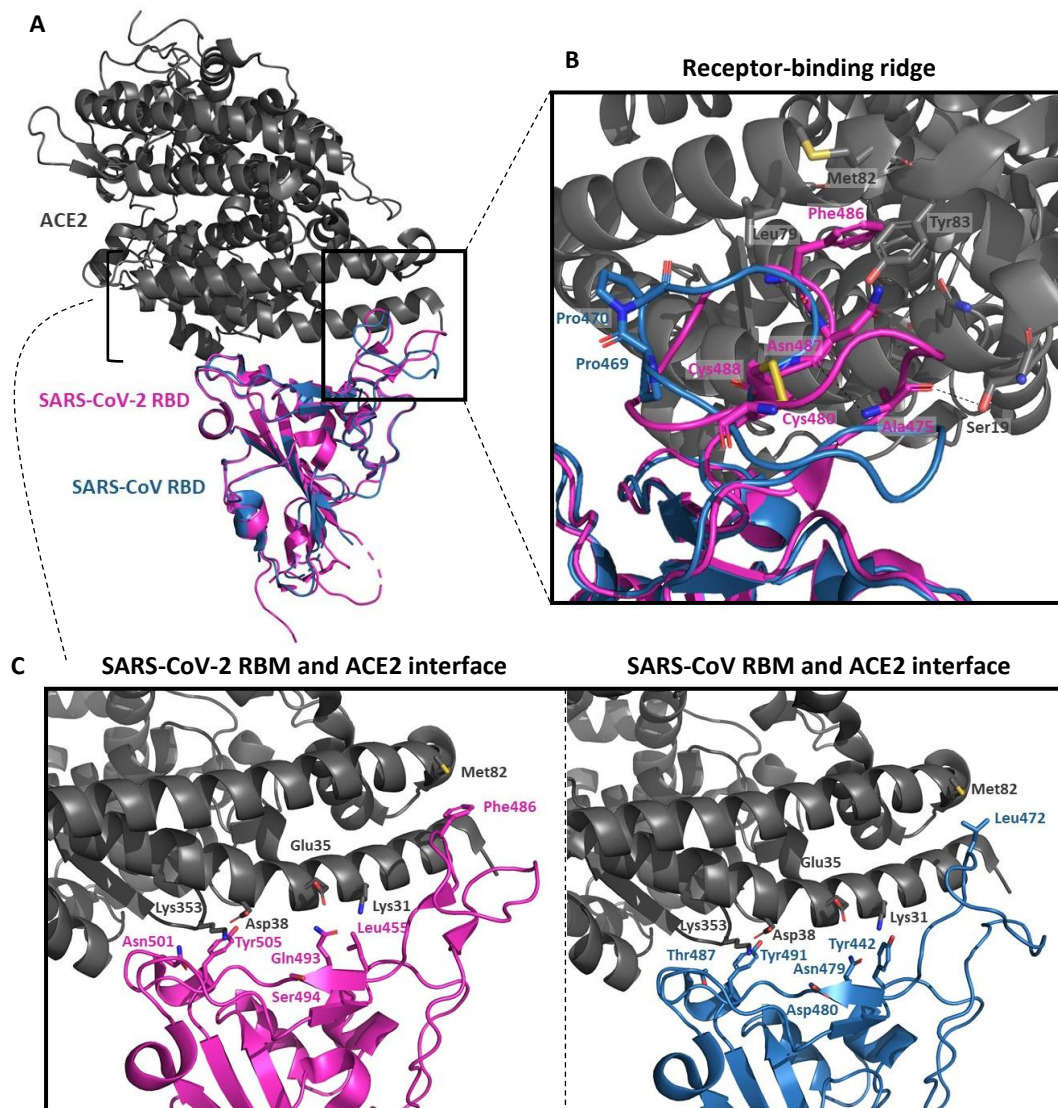


Figure 1.4 | ACE2 complex with SARS-CoV and SARS-CoV-2 RBD. (A) (blue) Structure of SARS-CoV RBD (PDB ID:2AJF) and (pink) SARS-CoV-2 RBD (PDB ID:6VW1) complexed with ACE2. (B) Differences in the receptor-binding ridge of (blue) SARS-CoV and (pink) SARS-CoV-2. (C) Binding interface of (blue) SARS-CoV RBM or (pink) SARS-CoV-2 RBM and (grey) ACE2.

1.3.4. Fusion peptide

Membrane fusion mediated by viral fusion proteins is a complex process still not fully understood. However, a general sequence of events is expected to occur as shown in Figure 1.5. First, the fusion protein is in the pre-fusion conformation with its FP hidden. Exposure to an external stimulus triggers a conformational change that results in FP insertion in the host membrane and the formation of an extended intermediate of the fusion protein that forms a bridge between both membranes. Next, the intermediate conformation collapses, bringing both membranes closer to each other and promoting the hemifusion state where the viral and host outer leaflets merge into a hemifusion stalk. Finally, the post-fusion conformation is followed by the formation of a fusion pore.²⁷

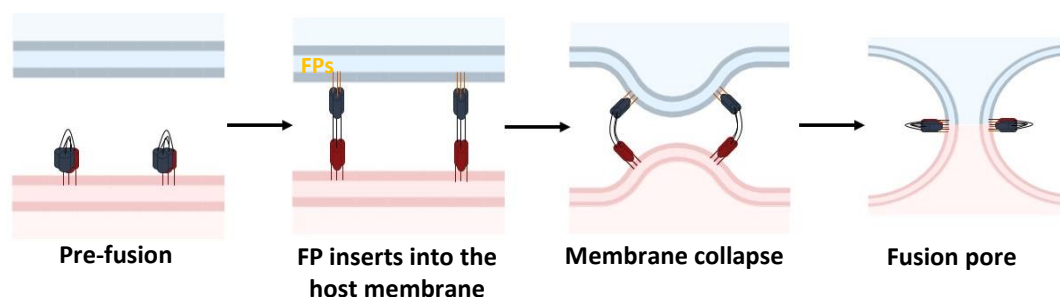


Figure 1.5 | Schematic representation of the membrane fusion mediated by viral fusion proteins schematized. Created with BioRender.com.

The term fusion peptide is widely used in the literature; however, a precise definition is lacking as well as absolute criteria to define it. Broadly speaking, the FP corresponds to the region of the fusion protein that inserts into and destabilizes the host membrane. Nonetheless, a combination of general characteristics may be applied to identifying FPs. FPs tend to be hydrophobic segments, rich in Gly residues and are usually located at the N-terminus of the fusion protein, although some internal FPs have been also reported.^{28 29 30} In addition, since FPs are key pieces for the fusogenic activity of fusion proteins, another way to identify them may lie on the existence of mutations affecting this region that lead to loss of activity. For this reason, FPs amino acid sequence tends to be highly conserved within viral families.²⁷ The mechanism by which the FPs promote membrane fusion might include acting as an anchor to join viral and host membranes, induction of a negative membrane curvature or lowering the lipid monolayer rupture tension.³¹

The study of the properties and function of FPs is possible using small synthetic peptide models harboring the same amino acid sequence as the original FP. Although these peptides are usually unable to induce the complete fusion between lipid vesicles, because the process depends on the action of several components of the fusion protein, they are able to destabilize membrane bilayers and promote lipid mixing. Additionally, the transition between aqueous to membrane-like environment is usually associated with a change in the FPs secondary structure from a random coil to a α -helix or β -strand conformation.³²

SARS-CoV-2 FP is still a matter of debate among the scientific community since there is no consensus about its amino acid sequence and exact position in the S-protein. Several experimental and theoretical studies have been conducted to clarify this matter. The most widely used approach is the comparison to other CoVs, since FPs are usually conserved regions. Three potential FPs have been proposed following this strategy: fusion peptide 1 (FP1), fusion peptide 2 (FP2) and internal fusion peptide (IFP) (Figure 1.6).^{33 34 35 36}

Recently, the FP1-FP2 (residues 816-855) structure in bilayered micelles was determined by NMR, revealing a globular fold composed of three α -helices. The first helix from Phe817 to Val826 is adjacent to the second helix formed between Ala831 and Tyr837. The third helix, which spans from Ala845 to Gln853 appears to be less stable and uniform. α -helices 1 and 2 seem to be inserted to about half the acyl thickness of a leaflet while α -helix 3 protrudes from the bilayer becoming semi-accessible to the solvent. Data collected using circular dichroism and NMR in another work aiming to identify the structural features of the SARS-CoV-2 fusion domain also indicate that the Ser816-Gly838 segment is more deeply inserted in the bilayer membrane while the Asp839-Phe855 region has a shallower interaction.³³

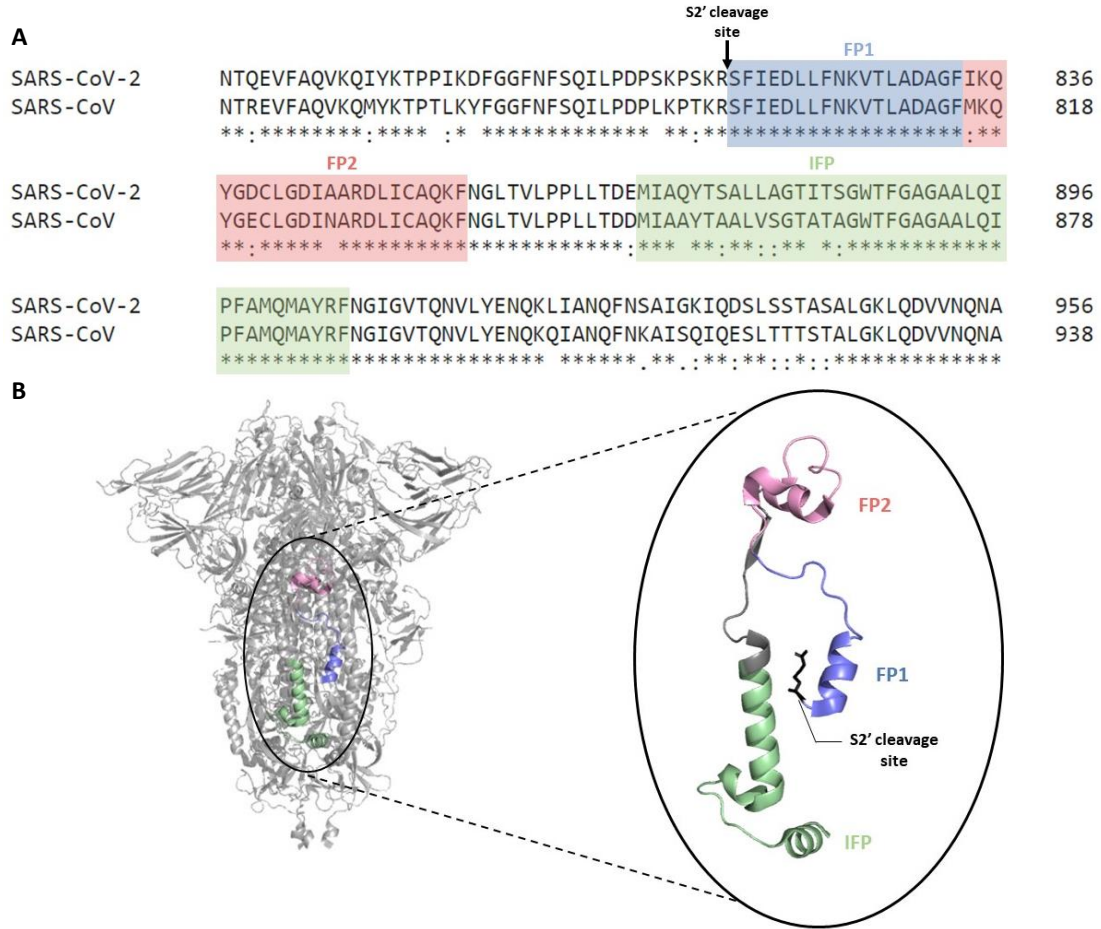


Figure 1.6 | SARS-CoV-2 FPs proposed in the literature. (A) Sequence alignment of SARS-CoV-2 and SARS S-protein with the FP1 (blue), FP2 (pink) and IFP (green) highlighted. (B) FPs localization and structure within the S-protein (PDB ID: 6ZGE).

The activity of the putative SARS-CoV-2 FPs has been evaluated using membrane model systems composed of different lipid mixtures, in the presence of calcium and at distinct pH values. The overall findings revealed that the IFP is more efficient in inducing leakage and lipid mixing of vesicles when compared with both N-terminal FPs. The activity of the three FPs seems to be enhanced in the presence of anionic phospholipids suggesting an electrostatic interaction between the lipids headgroup and the FPs.^{34 35} Moreover, the FP1-FP2 segment fusogenicity, as well as its propensity to form helical structures seem to be boosted by calcium ions. Possible calcium ions engagement spots have been predicted by molecular dynamics simulations pointing out to Asp830/Asp839 and Glu819/Asp820.³⁶

1.4. SARS-CoV-2 entry mechanism

The SARS-CoV-2 entry mechanism relies on the S-protein domains which play specific and pivotal roles, as shown in Figure 1.7. Unlike SARS-CoV, SARS-CoV-2 S-protein is preactivated by furin at the S1/S2 cleavage site - due to its multibasic motif – separating both subunits which yet remain noncovalently bound. This proteolytic cleavage leads to structural changes in the S-protein contributing to the increase of RBD accessibility. Once the RBD in the “up” form binds

to the host receptor ACE2, the TMPRSS2 protease on the host cell surface cleaves the S-protein in the S2' cleavage site. The structural changes associated with the S2 domain cleavage enable the FP release and its insertion into the host cell membrane.³⁷ The FP bridges both membranes diminishing the distance between them and, thus, promoting the membrane fusion.

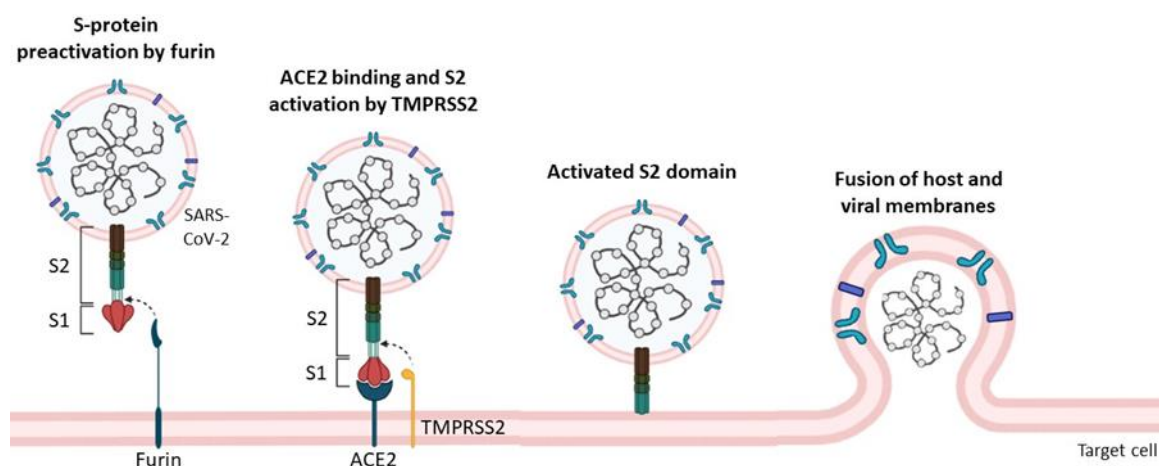


Figure 1.7 | SARS-CoV-2 entry mechanism. The SARS-CoV-2 S-protein is preactivated by furin at the S1/S2 cleavage site, followed by the high affinity binding of RBD to the ACE2 host cell receptor. Subsequently, the S2 domain is activated by proteolytic cleavage at S2' site by TMPRSS2. The FP is released and inserts into the host cell membrane triggering membrane fusion. Created with BioRender.com.

1.5. COVID-19: current treatments and promising therapeutic strategies

1.5.1. Current treatments and prevention

COVID-19 quickly spread worldwide causing an unexpected pandemic. Thus, therapeutics based on former knowledge acquired with diseases caused by similar viruses such as SARS, MERS and influenza, were applied. These include general treatments, anti-inflammatory compounds, anti-viral drugs and immunotherapy.³ General treatments are based on monitoring the symptomatic patients, providing supportive cares such as oxygenation, ventilation and fluid management when needed.

Glucocorticoids and tocilizumab have been used as anti-inflammatory agents in COVID-19 patients. When systemic or locally administered, glucocorticoids may modulate inflammatory-mediated lung injury and consequently reduce the disease progression to respiratory failure. Tocilizumab is a recombinant-humanized monoclonal antibody able to bind the interleukin-6 (IL-6) receptor which is increased in COVID-19 patients' serum and contributes to the development of severe states. However, the use of anti-inflammatory agents is still a matter of debate since its potential benefits and harms are not yet fully understood.³⁸

Remdesivir, a broad spectrum anti-viral, is a nucleoside analogue proposed to inhibit SARS-CoV-2 non-structural proteins affecting the viral RNA synthesis, therefore decreasing viral infection. However, some studies revealed that Remdesivir is only suitable for early disease stages because its activity is irrelevant once the viral replication peak is reached.^{38 39}

In addition, the convalescent plasma therapy consists of the administration of plasma collected from recovered patients which contains neutralizing antibodies able to help the infected patient to fight the infection.

Regarding COVID-19 specific treatments, two oral antiviral drugs were recently approved, namely Molnupiravir and Paxlovid developed by Merck and Pfizer, respectively. On one hand, Molnupiravir is a ribonucleoside prodrug of N-hydroxycytidine (NHC), which upon phosphorylation inside the cells is used by the viral RNA polymerase. The NHC triphosphate misdirects the polymerase to use either adenosine or guanine during the viral replication leading to deleterious errors. The accumulation of these errors prevents the proper viral replication.⁴⁰ On the other hand, Paxlovid is a combination of two protease inhibitors, ritonavir and PF-07321332, which target the SARS-CoV-2 3CL protease required for viral replication preventing the increase in viral load.⁴¹

The development and production of vaccines to prevent severe COVID-19 were accomplished in record time. In Portugal, four different vaccines were used to immunize the population developed by: Moderna, Pfizer/BioNTech, AstraZeneca and Janssen Pharmaceuticals. Moderna (mRNA-1273) and Pfizer/BioNTech (BNT162b2) are both mRNA vaccines while AstraZeneca (AZD1222) and Janssen Pharmaceuticals (JNJ-78436735/Ad26.COVS.2.S) are replication-defective viral vector vaccines.

The mRNA molecule in vaccines is modified to contain the necessary information for the synthesis of the SARS-CoV-2 S-protein in its pre-fusion conformation. The molecule is administered encapsulated in a lipid nanoparticle vector to promote the uptake by the target cells and is then transcribed and translated by the host cell machinery. Once the viral antigen is produced, it is presented to the immune system triggering an adaptive immune response targeting the SARS-CoV-2 S-protein.⁴² The AstraZeneca vaccine employs the debilitated chimpanzee adenovirus (ChAdOx1) while Janssen Pharmaceuticals uses the replicating-defective adenovirus 26 to deliver the genetic information required to induce the expression of SARS-CoV-2 S-protein in its pre-fusion form as well.⁴²

Despite the outstanding scientific efforts in creating treatments and vaccines for COVID-19, it is still necessary to keep investing in the development of therapies more efficient than the current ones.

1.5.2. Promising therapeutic targets

Knowledge acquired about SARS-CoV and SARS-CoV-2 affords new possibilities to prevent the infection. The most promising candidate is the S-protein due to its crucial role for viral entry. There are several possible strategies under investigation to target the S-protein, including neutralizing antibodies, inhibiting the S-protein cleavage and blocking the RBD-ACE2 interaction.³

SARS-CoV-2 S-protein is still evolving, resulting in new mutations either associated with an increased affinity to the ACE2 or better capacity to evade the immune system. Nevertheless, it has been reported that neutralizing antibodies produced by some patients can reduce the disease severity. Therefore, presenting the S-protein as an immunogenic agent to the human

immune system remains a valid approach. This can be done by using the full-length S-protein or by using key immunogenic S-protein domains.^{43 44}

The approach of blocking the S-protein cleavage by inhibiting the host cell proteases has also been validated by recent studies. In this strategy, the serine protease inhibitor camostat mesylate, an approved drug to inhibit the TMPRSS2 protease, reduced the SARS-CoV-2 infection.⁴⁵

The RBD-ACE2 binding can be prevented by different strategies, including blocking the host ACE2 or blocking the RBD at the virus surface. The approach targeting the ACE2 can be achieved by administering RBDs in the soluble form or anti-ACE2 antibodies, however disadvantages might be associated with this strategy due to ACE2 important role in the blood pressure regulation.⁴⁶ The RBD blockage could be accomplished by administering ACE2, either in the soluble form or fused to immunoglobulin Fc domain as demonstrated in the literature.^{47 48} However, these methods face similar worries since effects of inhibiting or increasing the hACE2 in a systemic manner are not yet fully characterized or understood. To overcome these concerns, another approach has been developed in several studies, where small peptide inhibitors are designed to bind to the RBD and, thus, prevent the interaction with the ACE2 receptor.^{49 50 51} Some of the RBD inhibitors developed presented higher affinity *in vitro* than ACE2, reinforcing the potential of this strategy.

1.6. Importance of fundamental and applied research

Fundamental or basic research is a curiosity-driven investigation aiming to fill a gap in knowledge about some specific matter. On the other hand, the goal of applied research is to fulfill a need or provide solutions to a specific problem with the tools given by current knowledge. Although fundamental research discoveries may not have an immediate impact on society and their contributions may go unnoticed for years, they have been paving the way to impactful life-changing advances. Therefore, fundamental and applied research go hand-in-hand as complementary partners, since without one the other has either no foundations or no applicability.

COVID-19 is a suitable example demonstrating that our preparedness to respond to an outbreak is tightly associated with fundamental research advances. For instance, to control the viral transmission it was essential to know how this type of viruses infect the host, which had already been addressed with previous CoVs-related research. Another relevant example is the mRNA vaccines technology based on mRNA knowledge acquired throughout decades.⁵² Several more examples could be given to reinforce the point: a national or international preparedness for novel threats is only as good as its fundamental research and the investment in it.

2. Objectives

The scientific community efforts to fight the COVID-19 pandemic have been outstanding. However, the development of new therapeutics based on potential targets is not fully explored. Moreover, essential information on key steps of the SARS-CoV-2 infection mechanism is lacking. Therefore, this project had three main goals, as illustrated in Figure 2.1, and is an example of the importance of combining fundamental and applied research. The first two goals were to develop strategies to prevent the SARS-CoV-2 infection based on different targets. The third goal was to shed light into the fusion mechanism through the identification and characterization of the SARS-CoV-2 FP. To achieve these goals, we used a combination of well-established computational design and simulation methods as well as experimental techniques.

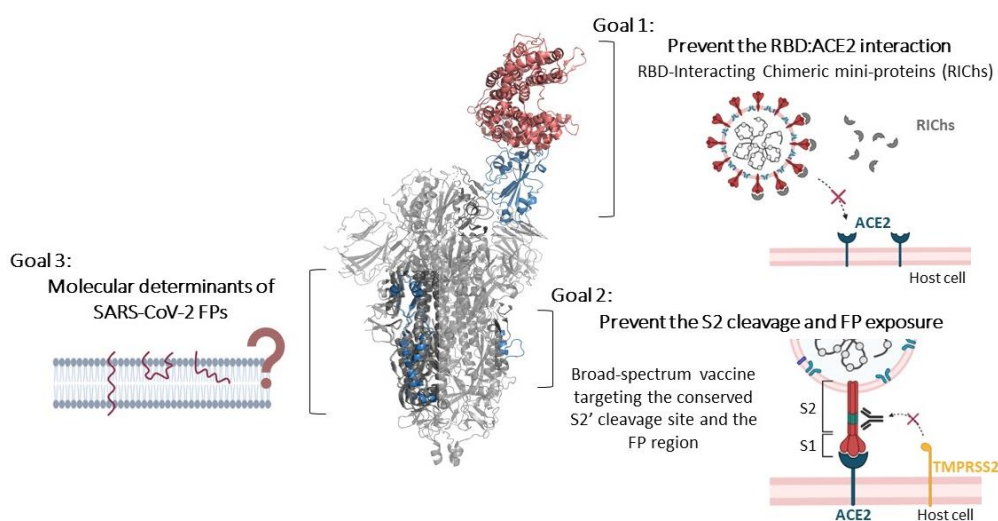


Figure 2.1 | Main goals of this project.

Regarding the first two goals, two approaches were explored to prevent viral infection. The first one intended to prevent SARS-CoV-2 RBD binding to ACE2 at the host cells surface, using RBD-interacting chimeric (RICHs) mini-proteins as decoys. These were designed with a top-down method where the ACE2 motif responsible to bind the RBD was grafted into protein scaffolds. The second one was based on the prevention of the proteolytic cleavage within the S2' cleavage site. An immunogenic construct was designed using the same top-down method, where the amino acid residues of the S2' cleavage site and a part of the FP1 region were grafted into a well-known and thermoresistant protein.

Regarding the third goal, putative FPs are proposed according to previous knowledge and computational tools. These FPs effects in membrane model systems were evaluated by experimental and *in silico* techniques aiming to identify which one of the putative FPs is required to trigger the membrane fusion process.

3. Materials and methods

3.1. *In silico* design

3.1.1. RBD-interacting chimeric proteins

RBD-interacting chimeric (RIChs) proteins were created using a motif-driven protein-protein interaction design protocol performed by Dr. Carlos Cruz, a member of the Protein Modelling Lab at ITQB in late 2019, so the methods used will be only briefly addressed here.

In motif-driven protein-protein interaction design, a motif known to bind to the desired target and with resolved structure is grafted onto a scaffold protein. Then, an improvement of the surrounding interaction surface is performed to accomplish optimal interactions and packing, resulting in higher binding affinities and complex stability improvement. This strategy aims to solve two major challenges: by using a known interacting motif, the design already begins with favorable interactions; and the grafted motif guides the designed protein towards a suitable orientation against the target.⁵³

The used protocol comprises the following three main steps performed with Rosetta software package⁵⁴: 1) selecting the binding motif(s) and determining the hotspots, 2) motif grafting and interface design, and 3) interface improvement.

Binding motif and hotspot determination

At the time RIChs were design, the knowledge about ACE2 and SARS-CoV-2 interaction was limited and achieved by some structural characterization studies and computational analysis. It was known that the binding was mediated by two ACE2 helices (H1 and H2) and one β -hairpin, as previously shown in Figure 1.4. Since the H1 helix (residues 19-52) was the main responsible for this interaction, it was chosen to be the binding motif. The hotspots were predicted via an interface alanine scanning where single substitutions to alanine are done in the interface residues and their impact on the binding energy is evaluated using the AlaScan filter in Rosetta scripts. A hotspot was selected if the residue replacement had a negative impact on the binding energy (ddG) using a threshold of 2.0 Rosetta energy units (r.e.u).

Motif grafting and interface design

The binding motif was grafted into the scaffold proteins within the RBD context. The scaffold proteins library consisted of 96 wild-type proteins with specific properties, including a solved structure at high resolution (< 0.25 nm), 90-300 residues length, high helical secondary structure content, monomeric assembly, without ligands and expressed in *E.coli*.

To improve the grafting success probability, the binding motif size was set to be flexible as long has it included the hotspot residues. The graftings were selected for interface design if the RMSD and atomic clashes remained below 0.1 nm and 5, respectively.

The interface design began with the replacement of the interface residues to alanine and side chain prepackings. Then, PackRotamerMover was employed to design core, boundary and interface residues within 0.15, 0.12 and 0.8 nm from the interface, respectively. The shape complementary (Sc) and binding energy (dG) were calculated for the minimized structures and used to rank the RIChs candidates.

Interface improvement

The candidates with the best Sc and dG proceeded to the interface improvement step. These types of assays aim to overcome a method limitation since the prior design was performed on static structures. Therefore, the residues near the RBD were the target of a new interface design. Here more than 50,000 conformational samples of RIChs candidates, generated after 6,000 backbone perturbation interactive Monte Carlo cycles and also side chain energy minimizations were evaluated in terms of Sc and dG. The best candidates were selected for experimental characterization.

3.1.2 Immunogenic construct design

Dr. Carlos Cruz, who created RIChs, also designed the immunogenic construct characterized in this project, using similar grafting methods.

According to the literature available at the time, a small S-protein fragment corresponding to the S2' cleavage site and a segment of the FP1 region (residues 807-824) revealed to be immunogenic inducing the generation of neutralizing antibodies in COVID-19 patients.⁵⁵ Another important reason for this fragment selection lies on its high conservation among diverse CoVs including SARS-CoV, MERS-CoV, Human Coronavirus OC43 (HCoV-OC43) and Human Coronavirus HKU1 (HCoV-HKU1) enabling the possible development of a broad-spectrum vaccine (Figure 3.1).

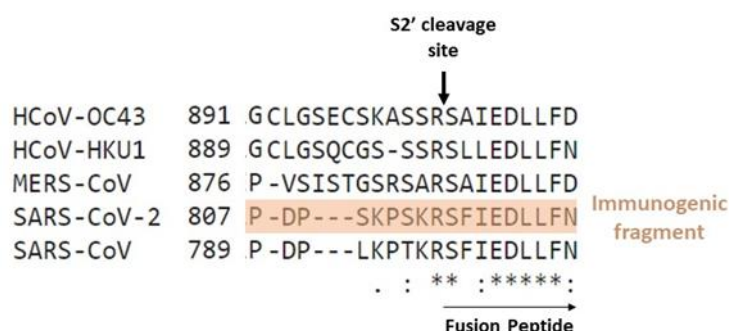


Figure 3.1 | Immunogenic fragment identification (highlighted in orange) and its conservation among coronaviruses able to infect humans including HCoV-OC43, HCoV-HKU1, MERS-CoV, SARS-CoV-2 and SARS-CoV.

The antibodies targeting this region could prevent the infection if they are able to block the S2 subunit activation by preventing the proteolytic cleavage. Furthermore, stereochemical effects of the antibody binding to the S-protein could also prevent the interaction of the FP with the host cell membrane.

Once the immunogenic fragment or epitope was selected, the search for a suitable protein scaffold began. The thermo-stable scaffold (TSS) revealed to be the best candidate for the following reasons: first, this is a small unnatural protein ensuring that polyclonal antibodies possible generated targeting the scaffold would not be harmful or interfere in any known biological pathway; second, TSS is highly thermo-stable which is a valuable feature for a therapeutic protein; third, its secondary structure has five β -strands and two α -helixes connected by short loops enabling to graft the epitope according to its secondary structure on the S-protein (small α -helix and random coil).

The design protocol was implemented with Rosetta software package in two main steps: motif grafting and construct improvement to better accommodate the motif.

Motif grafting

Unlike the protocol used for RICH design, the motif grafting in this pipeline is not automated. The region of the scaffold protein chosen for motif grafting was selected based on its compatible secondary structure in relation to the epitope. The motif alignment with the termini of the residues flanking the selected site on the scaffold protein was accomplished with the CCDEndsGraftMover tool. In addition, the residues located within a cut-off radius of 8 Å from the grafted motif were modified to stabilize the protein. Then, structural characterization was performed using the AbinitioRelax tool from Rosetta and short molecular dynamics simulations to evaluate the protein folding and the epitope flexibility, respectively.

Construct improvement

To select the best candidate, backbone perturbations and optimization of the interactions surrounding the epitope were performed on the residues within the 8 Å radius cut-off. These were iteratively mutated and the impact of these mutations on the complementarity between the motif residues and their neighbors were considered. This procedure was repeated dozens of times employing the Monte Carlo method and the PackRotamerMover implemented in Rosetta. The most promising candidates were ranked by the complementary interactions between the motif residues and their neighbors on the scaffold in order to select the best construct. The selected immunogenic construct was named TSSFP.

3.2. Identification of SARS-CoV-2 putative fusion peptides

The SARS-CoV-2 putative FPs propose in this project are the result of the combination between previous knowledge and the analysis of sequence alignments. According to the literature, FPs of many class I viral fusion proteins share similar characteristics such as: (I) location at the S2 subunit N-terminus, near or immediately adjacent to the proteolytic site within the S2 subunit; (II) conserved hydrophobic domains associated with membrane-interacting features; (III) potential amphipathic secondary structure and; (IV) abundance of glycine and alanine residues.^{27 56 57} Besides these characteristics above, the S2 subunit conservation among CoVs was also considered during the putative FPs identification. Additionally, recent works regarding this subject had already suggested possible FPs, in the following S-protein regions: 816-837, 816-855 and 869-905.

3.3. *In silico* characterization

Molecular dynamics (MD) simulations were used for the *in silico* characterization of the putative FPs, allowing us to examine how these peptides insert and interact with the membrane bilayer.

3.3.1. Molecular dynamics (MD) simulations

MD simulations are a well proven tool for understanding and characterizing proteins structural properties, such as folding and conformational changes, as well as understanding

microscopic interactions, including protein-protein interaction, drug binding and protein-membrane interactions. They can also be used to characterize the mechanisms of action of proteins or peptides at the molecular level. Computer simulations serve as a bridge between the microscopic or atomic level detail and the macroscopic world of the experimental assays. Moreover, MD simulations provide information on a temporal scale difficult or impossible to capture in the laboratory.⁵⁸

In MD simulations, atoms are treated as spheres, with point charges, connected by springs. The electronic degrees of freedom are not accounted for and bonds are not broken or created. These are obviously approximations, since molecular systems can only be precisely described using quantum mechanics laws. However, they are reasonable, since the goal of MD simulations is not to analyze phenomena in which the electronic degrees of freedom play a role and due to the fact that they rely on the Born-Oppenheimer approximation (which states that the nuclear and electronic degrees of freedom can be uncoupled). Moreover, we assume that nuclei can be treated as classical particles. Thus, the evolution of the positions and momenta of the particles of a system over time can be obtained by integrating Newton's second law of motion ($F = ma$)(Eq. 3.1):

$$F_i = m_i \frac{d^2 r_i}{dt^2}, \quad i = 1 \dots N \quad \text{Equation 3.1}$$

Equation 3.1 describes a system of N interacting particles, where each particle (i) is experiencing a force (F_i) that is proportional to its mass (m_i) times acceleration. Considering the energy of an atom represented as $U(i)$ and the three-dimensional features of the system, the force acting on each atom will be a three-dimensional vector which is the result of the combined partial derivatives of each principal axis (Eq. 3.2).

$$\frac{-dU(i)}{r_i} = -\left(\frac{\partial U}{\partial r_i} x + \frac{\partial U}{\partial r_i} y + \frac{\partial U}{\partial r_i} z\right) \quad \text{Equation 3.2}$$

The energy of the atoms is calculated using a potential energy function, often called a force field and generally represented as Eq.3.3:

$$U_{total} = U_{bonded} + U_{non-bonded} \quad \text{Equation 3.3}$$

where the bonded energy (U_{bonded}) comprises the energy associated with atomic bonds, angles, dihedrals and improper dihedrals while the non-bonded energy ($U_{non-bonded}$) is composed of the energy associated with the van der Waals and electrostatic interactions. The force field is a semi-empirical energy function that uses experimental data or quantum mechanics calculations to derive a set of defined parameters (force field parameters).^{59 60}

The potential energy of the simulation system varies with the atoms coordinates originating the potential energy surface represented in Figure 3.2⁶¹. The most important regions of this surface are the ones which represent stable states of the system also named energy minima points. Whereas more than one minimum may exist, there is usually only one global energy minimum. Minimization algorithms are used to identify the minimum points on the potential energy surface.⁵⁹

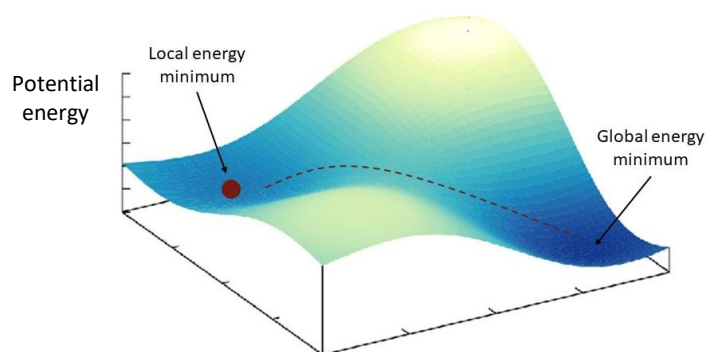


Figure 3.2 | General representation of a potential energy surface which maps the potential energy as a function of the system configuration. The local minima correspond to stable states of the system while the global energy minimum corresponds to the most stable state. A molecule may transit from a local minimum to the global one if it overcomes the energy barrier. Adapted from *Keith et al. (2021)*⁶¹

Biomolecules do not exist in vacuum. Therefore, a suitable representation of the solvent is important. The solvent can be simulated using implicit or explicit models. In the implicit model, the solvent is simulated as a continuum medium. In contrast, in the explicit model the molecules of the solvent are part of the system and interact with other molecules. Although the explicit model is more computationally expensive, it leads to more accurate results.

Another important matter to consider is the system boundary conditions which influence the molecules behavior when they approach the simulation box's limits or boundaries. The most common method to deal with this issue relies on the use of periodic boundary conditions (PBC) schematized in Figure 3.3. In PBC, the simulation unit cell is repeated in all directions and the box walls are removed, meaning that an atom may move between adjacent unit cells. Moreover, the atoms in a unit cell are under the influence of other atoms in the surrounding cells.⁵⁹

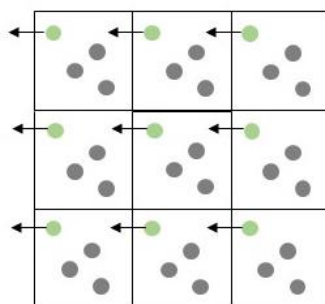


Figure 3.3 | Schematized representation of the periodic boundary conditions (PBC). Several copies of the same unit cell are repeated and if the atom in green leaves a certain box according to the displacement vector in black it will appear in the adjacent box.

MD simulation protocols can be designed to maintain some bulk properties of the system constant, including the number of particles (N), the temperature (T), the volume (V) or the pressure (P). Depending on the properties that are kept constant in MD simulations, different ensembles can be sampled. In this project the isothermal-isobaric ensemble where the

N, T and P are kept constant was used because it considers the system in contact with a thermostat and a barostat.⁶²

3.3.2 SARS-CoV-2 FPs MD simulations setup

In this thesis four systems were studied, each corresponding to one of the SARS-CoV-2 putative FPs identified. The FPs structures were obtained by truncating the SARS-CoV-2 S-protein (PDB ID: 6ZGE) as follow: FP1 (residues 816-837), FP2 (residues 816-855), FP3 (residues 869-905) and FM (residues 816-905).

The CHARMM-GUI Membrane Builder tool was used to build a POPC:POPS (4:1) membrane bilayer system with each putative FP. This is a well-established web-based user interface to build all-atom protein-membrane systems for MD simulations.⁶³ The overall process may be generalized in six steps: read the protein structure, orientate the protein, determine the system size, components generation, assembly and equilibration.⁶⁴

First, the PDB files of each FP generated after truncating the S-protein were used to generate a structure topology file. For FP1 and FM a disulfide bond between Cys840 and Cys841 was added. Next, the FPs were positioned in the membrane bilayer with the assistance of the Positioning of Proteins in Membranes (PPM) method implemented in CHARMM-GUI based on the free energy optimization associated with the protein transfer from an aqueous to a lipidic environment.⁶⁵

In the third step, the system box was set to be rectangular with specific dimensions of the xy plane and z-axis. The xy plane length was determined based on the lipids number and ratio comprising a total of 250 lipid molecules (200 POPC and 50 POPS). The z-axis length was chosen to guarantee that there would be no significant interactions between lipids and peptides in periodic images in the zz direction, resulting in a minimum water layer with a height of 22.5 Å below and above the membrane. In the subsequent step, the system was built using the replacement method and individual components were generated to solvate the peptide, including the lipid bilayer, water molecules and the suitable number of Na⁺ and Cl⁻ ions to neutralize the system achieving a concentration of 150 mM. Then, the protein/membrane complex and the remaining system components were assembled.

The last step was the equilibration aiming to relax the system. This is a progressive process where restraints are gradually diminished as the equilibration occurs. Different restraints were applied to the peptide, ions, lipid and water molecules through six cycles of 100 ps each: harmonic restraints on heavy atoms of the peptide and ions; repulsive planar restraints to avoid water molecules from penetrating into the membrane hydrophobic area; and, planar restraints to ensure the position of the lipids head groups along the z-axis. Table 3.1 specifies the restraints applied in each equilibration step. To ensure the equilibration success, in the two first steps, the NVT ensemble (constant volume and temperature) is used, while the NPT ensemble (constant pressure and temperature) is used in the remaining steps at 310 K. This temperature was chosen since it corresponds to the temperature of the human body (SARS-CoV-2 host) and also ensures that the membrane is in the biologically relevant liquid crystal phase.

Table 3.1 | Detailed information on each equilibration step. Force constants in kcal/(mol.Å²) for the diverse components of the system.

Step	Force Constants for Harmonic Restraint (kcal/(mol.Å ²))				
	Peptide Backbone	Peptide Sidechain	Water	Lipid	Ion
1	10.0	5.0	2.5	2.5	10.0
2	5.0	2.5	2.5	2.5	0.0
3	2.5	1.0	1.0	1.0	0.0
4	1.0	0.5	0.5	0.5	0.0
5	0.5	0.1	0.1	0.1	0.0
6	0.1	0.0	0.0	0.0	0.0

The simulations of each putative FP in a membrane bilayer were conducted using the GROMACS package⁶⁶, version 2020.3 and the CHARMM36m force field was used to model the peptide and lipid atoms⁶⁷. The TIP3P Water Model was also used.⁶⁸ The pressure was maintained at 1 bar employing the Parrinello-Rahman algorithm⁶⁹ with a semi-isotropic pressure coupling and a time coupling constant of 5.0 ps. The temperature was maintained at 310 K using the Nosé-Hoover algorithm^{70 71} with a time constant of 1.0 ps. PME method⁷² was applied to treat the long-range electrostatic interactions using the Verlet cut-off scheme with a 1 Å radius. The LINCS⁷³ algorithm was used to constrain all hydrogen-containing bonds. Three replicates of each system were simulated for 0.3 µs for FP1 and 2.5 µs for the remaining FPs.

3.4. Protein production

3.4.1. Plasmids and competent cells

The plasmids used to express the RICH, TSSFP and FM constructs were produced by GeneCust. The synthetic genes codon optimized for expression in *E.coli* were cloned into the pET28a(+) vector with kanamycin resistance, as schematized in Figure A1 presented in the annex section, except for FM. FM was built by the FastCloning⁷⁴ method to insert the FM sequence in the pETDuet vector with ampicillin resistance, as described in the annex section “FM expression vector produced by the FastCloning method”.

The following *E.coli* strains were used as competent cells to test the proteins expression and, subsequently for the expression scale-up: *E.coli* BL21(DE3), *E.coli* BL21(DE3) pLysS and *E.coli* BL21(DE3) star.

Transformations were performed with approximately 100 ng of expression plasmid and 50 µL of competent cells. After a 15 min incubation on ice, cells suffered a heat shock at 42 °C for 45 sec and were resuspended in 600 µL LB. Next, an incubation at 37 °C and 500 rpm (Eppendorf® Thermomixer Compact) for 1h was performed, followed by a centrifugation at room temperature (RT) and 3600 rpm for 2.5 min (Eppendorf® Microcentrifuge 5415 D). 500 µL of the supernatant was discarded and the cells pellet was resuspended in the remaining 150 µL to be plated in LB agar plates supplemented with 50 µg·mL⁻¹ kanamycin (Roth) or 100 µg·mL⁻¹ ampicillin (Sigma-Aldrich). Plates were incubated at 37 °C for 16h (Mettler INB 500 incubator).

3.4.2. Expression

Expression tests

The expression of some of the proteins used in this work needed to be either tested or optimized due to the different reasons such as: lack of information in the literature, low protein yield using the conditions described in the literature or being novel proteins. Therefore, TSSFP and FM expression tests were performed according to Tables 3.2 and 3.3, respectively.

RICH 1 expression was also a target of optimization rounds. However, none of the tested conditions were further pursued. Thus, the information related to this issue is presented in the annex section “RICH 1 expression tests”.

Table 3.2 | Conditions tested for the expression of TSSFP using *E.coli* BL21(DE3) and an induction OD_{600nm} of 0.6.

Culture medium	[IPTG ^a] (mM)	t _{expression} & T _{expression}
LB	0.1	5h 37°C
		18h 18°C
	0.5	5h 37°C
		18h 18°C
LB	0.1	5h 37°C
		18h 18°C
	0.5	5h 37°C
		18h 18°C

a - Isopropyl-β-D-thiogalactoside

Table 3.3 | Conditions of the FM expression tests. The induction began at OD_{600nm} 0.6 with 1 mM IPTG and was conducted at 18°C. Aliquots were collected at 2h, 4h and 18h.

<i>E. coli</i> strain	Culture medium
BL21(DE3)	LB
	M9
BL21(DE3)pLysS	LB
	M9
BL21(DE3)star	LB
	M9

The conditions tested for each protein were different, however the following procedures were common.

Pre-inoculum: Single colonies were used to inoculate 3 mL of LB medium supplemented with 50 µg·mL⁻¹ kanamycin or 100 µg·mL⁻¹ ampicillin, and incubated for 16h, at 37°C and 130 rpm (Aralab Agitorb 200).

Cell culture: 250 mL culture Erlenmeyers containing 50 mL culture medium supplemented with 50 µg·mL⁻¹ kanamycin or 100 µg·mL⁻¹ ampicillin were inoculated with the suitable volume of pre-inoculum to start the cell growth with an optical density at 600 nm (OD_{600nm}) of 0.05. Each Erlenmeyer was incubated at 37 °C and 120 rpm (Eppendorf® Innova® 44 shaker). The culture's OD_{600nm} was monitored along time.

Induction: When the OD_{600nm} reached 0.6, protein expression was induced with IPTG and incubated at the appropriate temperature and for the expression time defined in the test conditions.

Samples: Before the protein expression induction and at the end of the expression time 1 mL aliquots were collected and treated in order to separate the soluble from the insoluble fraction.

After a 10 min centrifugation (Eppendorf® Microcentrifuge 5415 R) at 12,000 rpm and 4 °C, the pellet was resuspended in 150 µL BugBuster® Protein Extraction Reagent (Milipore) lysis buffer supplemented with 0.3% (v/v) DNase (Sigma-Aldrich) and 1% (v/v) lysozyme from chicken egg white (Sigma-Aldrich). A 30 min incubation on ice was followed by a 30 min centrifugation at 13,000 rpm and 4 °C. The resulting supernatant and pellet were separated as soluble and insoluble fractions, respectively. 50 µL of SDS-PAGE sample loading buffer 4x were added to the soluble fractions while 200 µL of SDS-PAGE sample loading buffer 1x were added to the insoluble fractions. The latter were incubated at 95 °C for 10 min and vortexed.

Denatured samples were loaded into a NuPAGE™ 10% Bis-Tris gel (Invitrogen) and an electrophoresis was performed for 30 min at 200 V (Amersham Pharmacia Biotech EPS 601). Next, the proteins in the gel were transferred to polyvinylidene difluoride (PVDF) membranes using the iBlot™ 2 PVDF Regular Stacks and the Dry Blotting System (Thermo Fisher Scientific). The membranes were then blocked with 5% (w/v) skim milk (Sigma-Aldrich) in Tris buffer saline buffer containing 0.05% Tween 20 (TBS-T) for 30 min, at 37 °C and under gentle stirring (Heidolph Rotamax 120). An incubation with a mixture of the primary antibody 1:6,250 Monoclonal Anti-polyHistidine antibody produced in mouse (H1029 Sigma-Aldrich) and the secondary antibody 1:4,000 Anti-mouse IgG (Fc specific)-alkaline phosphatase antibody produced in goat (A2429 Sigma-Aldrich) in 5% (w/v) skim milk in TBS-T buffer for 1h, at 37 °C and under smooth stirring was performed. After three 10-min wash steps with TBS-T buffer, the WB was developed with the 1-Step™ NBT/BCIP Substrate Solution (Thermo Fisher Scientific).

Expression scale-up

The expression scale-up conditions for each protein are described in Table 3.4. The starter culture was performed as mentioned in the section “Expression tests” above. 2L culture flasks containing 1L of culture medium supplemented with 50 µg·mL⁻¹ kanamycin or 100 µg·mL⁻¹ ampicillin were inoculated with the appropriate volume of pre-inoculum to start cell growth with an OD_{600nm} of 0.05 and incubated at 37 °C and 120 rpm (Eppendorf® Innova® 44 shaker). When the OD_{600nm} reached 0.6, protein expression was induced with IPTG at the temperature and for the time presented in Table 3.5. Cells were recovered by centrifugation at 7500 rpm, 4 °C for 10 min (Beckman Coulter® Avanti® J-26 XPI, rotor JA-10) and then stored at -20 °C.

Table 3.4 | Expression scale-up conditions

Protein	<i>E. coli</i> strain	Culture medium	[IPTG] (mM)	t _{expression} & T _{expression}
RICH 1, 2, scaffold 1	BL21(DE3)pLysS	TB	1	18h 18°C
RICH 3	BL21(DE3)pLysS	LB		5h 37°C
RICH 4, 5, scaffold 5	BL21(DE3)	TB		18h 18°C
TSSFP	BL21(DE3)	2xYT	0.1	18h 18°C
FM	BL21(DE3)pLysS	LB	1	18h 18°C

3.5. Protein purification

Cells collected after each expression scale-up were lysed according to the following protocol. First, the pellet was resuspended in lysis buffer (3:1 ratio of volume of buffer, in mL, to cell mass, in g) composed of the selected buffer used for the first purification step supplemented with 1 mg·mL⁻¹ lysozyme (Sigma-Aldrich), 0.005 U benzonase (ChemCruz) and cOmplete™ EDTA-free Protease Inhibitor Cocktail tablets (Roche). Then, cellular disruption was accomplished by 8 sonication cycles with UP200S Ultrasonic Processor (Hielscher) with 50% amplitude and 0.6 cycle for 30 sec of continuous sonication and 30 sec of incubation on ice. To separate the soluble from the insoluble fraction, the lysate was centrifuged at 18,000 rpm, 4 °C for 30 min (Beckman Coulter® Avanti® J-26 XPI, rotor JA-25.50).

3.5.1 RBD-interacting chimeric mini-proteins

RICH 1 purification began with an affinity chromatography using a HiTrap™ Fast Flow Crude column (Cytiva) coupled to an ÄKTA FPLC system (Amersham Biosciences). A washing step with 20 mM Tris-HCl pH 7.5, 300 mM NaCl, 200 µM TCEP, 0.0005 % Decylmaltoside (DM), 10 mM Imidazole buffer was followed by an elution gradient of 12 column volumes (CVs) at 3 mL·min⁻¹ to 25 % of elution buffer containing 500 mM imidazole. Additionally, two elution steps were performed: the first with 50 % and the second with 100 % of elution buffer. The RICH 1-containing fractions collected during the chromatographic step were concentrated with an Amicon™ Ultra Centrifugal Filters Units (Merck) and loaded into a Superdex® 75 10/100 GL (Sigma-Aldrich). In this size-exclusion chromatography, the elution was done at 1 mL·min⁻¹ with 20 mM Tris-HCl pH 7.5, 300 mM NaCl, 100 µM TCEP, 0.0005 % DM buffer.

RICH 2 and 3 first purification step consisted in an affinity chromatography with a HiTrap TALON® Crude column (Cytiva) coupled to an ÄKTA FPLC system. After the washing step with 20 mM Tris-HCl pH 8.0, 300 mM NaCl, 10 % glycerol, 200 µM TCEP, 10 mM Imidazole buffer, the elution gradient of 15 or 11 CVs at 3 mL·min⁻¹ to 100 or 75 % of elution buffer containing 500 mM imidazole was performed for RICH 2 and RICH 3, respectively. For both proteins an elution step with 100 % elution buffer was done. The suitable resulting fractions were concentrated and loaded into a Superdex™ 75 prep grade column where the elution was accomplished with 20 mM Tris-HCl pH 8.0, 300 mM NaCl, 10 % glycerol, 200 µM TCEP buffer at 1 mL·min⁻¹. An additional size-exclusion chromatography was performed with a Superdex 75 10/300 GL (Sigma-Aldrich) column, from which both proteins were eluted at 0.5 mL·min⁻¹ with 20 mM Tris-HCl pH 7.4, 300 mM NaCl, 100 µM TCEP buffer.

RICH 4 and 5 were purified using a similar protocol to the RICH 2 and 3 purifications, with some modifications. In the affinity chromatography, the washing and elution buffer were composed of PBS pH 7.4, 200 µM TCEP, 10 mM or 500 mM Imidazole, respectively. The elution gradients were done for 12 and 15 CVs to 100 % elution buffer, for RICH 4 and 5 respectively. In the first size-exclusion chromatography, the used buffer was PBS pH 7.4, 200 µM TCEP.

As controls for the RICH proteins characterization, the respective scaffolds were also produced and purified. Scaffold 1 was purified with a four-step protocol. First, an affinity chromatography using a HiTrap™ Fast Flow Crude column coupled to an ÄKTA FPLC system was performed, where the washing step with 20 mM Tris-HCl pH 7.5, 600 mM NaCl, 10 % glycerol,

100 μ M TCEP, 10 mM Imidazole buffer was followed by an elution gradient of 18 CVs at 3 mL $\cdot$ min⁻¹ to 100 % of elution buffer containing 500 mM imidazole. An elution step with 100 % of elution buffer was also done. Next, the appropriate fractions were concentrated and loaded into a Superdex™ 75 prep grade column where the protein was eluted with 20 mM Tris-HCl pH 7.4, 600 mM NaCl, 10 % glycerol, 100 μ M TCEP buffer at 0.8 mL $\cdot$ min⁻¹. The third step consisted in an ionic-exchange chromatography using a HiTrap® Q Fast Flow column (Cytiva) where the washing step with 20 mM Tris-HCl pH 7.5, 10% glycerol, 100 μ M TCEP buffer was followed by the elution gradient of 30 CVs at 2 mL $\cdot$ min⁻¹ to 100 % of elution buffer containing 1 M NaCl. Finally, the scaffold 1-containing fractions underwent a hydrophobic chromatography using an HiPrep™ Octyl Fast Flow column (Cytiva). After the washing step with 50 mM Tris-HCl pH 7.5, 3 M NaCl buffer the protein was eluted using a gradient of 10 CVs at 2 mL $\cdot$ min⁻¹ to 100 % elution buffer without NaCl. At last, an elution step with 100 % elution buffer was performed.

The Scaffold 5 purification was similar to those of RICH 4 and 5, yet the elution gradient of the affinity chromatography was done for 20 CVs and the second size-exclusion chromatography was unnecessary.

After each purification step, the fractions obtained were examined by SDS-PAGE and WB when necessary.

3.5.2. Immunogenic construct

The TSFP purification started with an affinity chromatography using a HiTrap TALON® crude column coupled to an ÄKTAFPLC system, where the soluble fraction was loaded. After a washing step with 20 mM Tris-HCl pH 7.5, 300 mM NaCl, 10 % glycerol, 100 μ M TCEP and 5 mM imidazole to eliminate the contaminants with low affinity to the column resin, protein elution was performed with a 15 CVs gradient at 3 mL $\cdot$ min⁻¹ to 50% of elution buffer containing 20 mM Tris-HCl pH 7.5, 300 mM NaCl, 10 % glycerol, 100 μ M TCEP and 500 mM imidazole. A final elution step with 100% elution buffer containing 500 mM imidazole was made. Fractions of 1.6 mL were collected throughout the purification and evaluated by SDS-PAGE (12.5 %). TSFP-containing fractions were concentrated using an Amicon™ Ultra Centrifugal Filter before the size-exclusion chromatography step with Superose® 6 Increase 10/300 GL column (Sigma-Aldrich) in which proteins were eluted with 20 mM Tris-HCl pH 7.5, 300 mM NaCl, 10 % glycerol and 100 μ M TCEP. Fractions collected in this step were also evaluated by SDS-PAGE (12.5 %), pooled and loaded into a PD-10 Desalting column (Cytiva) eluted with 20 mM Tris-HCl pH 7.5, 10% glycerol and 100 μ M TCEP. The third purification step consisted in an ionic-exchange chromatography eluted with a 15 CV gradient at 2.5 mL $\cdot$ min⁻¹ to 50% of elution buffer containing 20 mM Tris-HCl pH 7.5, 300 mM NaCl, 10 % glycerol, 100 μ M TCEP and 1 M NaCl, followed by an elution step with 100% of the same buffer. The appropriate fractions were selected after SDS-PAGE analysis and loaded into a PD-10 Desalting column to exchange the buffer to PBS pH 7.4, 300 mM NaCl and 10 % glycerol.

3.5.3. SARS-CoV-2 Fusion Module

Despite FM being expressed with a His-Tag this tag revealed to be inaccessible and did not enable the interaction with either the HiTrap™ Fast Flow Crude or the HiTrap TALON® crude

column. The flow-through from the affinity chromatography column was loaded into an Amicon™ Ultra Centrifugal Filter with a molecular weight cut-off of 30 kDa aiming to retain the high molecular weight contaminants in the top of the concentrator. The FM was recovered in the flow-through and concentrated in an Amicon™ Ultra Centrifugal Filter with a 3 kDa cut-off. In the following step, the sample was injected into a Superdex™ 75 prep grade column coupled to an ÅKTA FPLC system (Amersham Biosciences), from which the protein was eluted with PBS pH 7.4, 200 mM NaCl, 10 % glycerol, 100 µM TCEP buffer at 1 mL·min⁻¹. The collected fractions were evaluated by SDS-PAGE. Finally, a WB targeting the His-Tag was performed to ensure the FM identity.

3.5.4. Protein quantification

After each purification, proteins concentration was determined using the Bradford method ⁷⁵adapted to be conducted in a 96-well plate.

3.6. Chemical Synthesis of the FP1, FP2 and FP3

FP1, FP2 and FP3 were synthesized by Bachem AG (Bubendorf, Switzerland) using solid phase peptide synthesis. The FPs amino acid sequences and molecular weights were confirmed by matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) and the purity of each peptide was determined by HPLC (Table 3.5).

Table 3.5 | Characteristics of the FPs chemically synthesised. FPs sequence, molecular weight (Da) assessed by MALDI-MS and purity (in percentage) determined by HPLC.

Peptide	Sequence	Molecular weight (Da)	Purity (%)
FP1	H-Ser-Phe-Ile-Glu-Asp-Leu-Leu-Phe-Asn-Lys-Val-Thr-Leu-AlaAsp-Ala-Gly-Phe-Ile-Lys-Gln-Tyr-OH trifluoroacetate salt	2532.9	94.6
FP2	H-Ser-Phe-Ile-Glu-Asp-Leu-Leu-Phe-Asn-Lys-Val-Thr-Leu-AlaAsp-Ala-Gly-Phe-Ile-Lys-Gln-Tyr-Gly-Asp-Cys-Leu-Gly-Asp-IleAla-Ala-Arg-Asp-Leu-Ile-Cys-Ala-Gln-Lys-Phe-OH trifluoroacetate salt	4424.1	88.7
FP3	H-Met-Ile-Ala-Gln-Tyr-Thr-Ser-Ala-Leu-Leu-Ala-Gly-Thr-Ile-ThrSer-Gly-Trp-Thr-Phe-Gly-Ala-Gly-Ala-Ala-Leu-Gln-Ile-Pro-Phe-AlaMet-Gln-Met-Ala-Tyr-Arg-OH trifluoroacetate salt	3910.6	78.3

3.7. Experimental characterization

3.7.1. Structural characterization

Far-UV circular dichroism

Far-UV circular dichroism (CD) spectropolarimetry allows to study the secondary and folding of a peptide or protein, as well as structural changes induced by variations in the protein environment factors such as temperature, pH and the presence of other proteins, lipids, or small molecules.⁷⁶

CD is the result of the differential absorption of circularly polarized light. A sample is irradiated by a polarized light beam with two circular components: clockwise (right, R) and counterclockwise (left, L). The result of this phenomena can be schematized as in Figure 3.4.

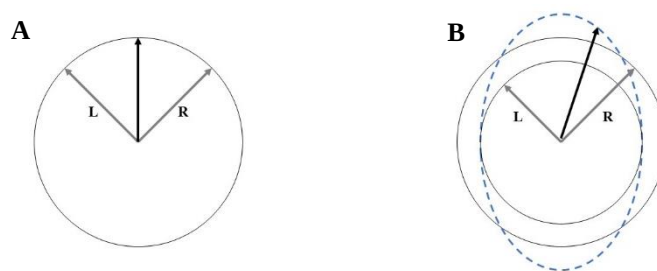


Figure 3.4 | Origin of the CD effect. (A) The absorption of both components, R and L, is similar in amplitude, resulting in a plane polarized radiation. (B) The magnitude of R and L absorption is different, so when combined they generate an elliptically polarized radiation and consequently a CD signal.

In situation A of Figure 3.4, both R and L components are equally affected thus no CD signal is generated. On the other hand, in situation B, there is a distinct absorption of each component resulting in a CD signal which is represented by the difference between L and R absorptions ($\Delta A = A_L - A_R$). Generally, this is also reported in terms of ellipticity (θ) in degrees through the following numerical relationship between ΔA and θ : $\theta = 32.98 \Delta A$. A CD signal is generated by a chiral molecule (optically active) either because it is intrinsically chiral, it is covalently bound to a chiral center, or it is in an asymmetric environment as the result of the three-dimensional structure.⁷⁷

In proteins, the absorption of optically active targets occurs in specific wavelength ranges. For example, the peptide bond absorption is observed below 240 nm; the aromatic side chains affect the 260 to 320 nm region; and the disulfide bonds are usually detected around 260 nm.

Secondary structure characterization

CD spectra in the peptide bond region are extremely useful to characterize protein secondary structure since each type of secondary structure element produces a specific CD profile, as shown in Figure A6 in the annex section. Besides comparing the obtained spectra with standard ones, two different algorithms were also employed, BeStSel and CAPITO,^{78 79} aiming to predict the secondary structure elements contribution.

For secondary structure characterization samples were prepared by dilution in 20 mM Tris-HCl pH 7.4 buffer to 0.15 mg·mL⁻¹ (RIChs) or in PBS pH 7.4, 300 mM NaCl and 10 % glycerol to 0.2 mg·mL⁻¹ (TSSFP).

CD spectra were collected at 20 °C in a Jasco J-815 CD Spectropolarimeter with the following acquisition parameters: measure range of 200-260 nm, data pitch of 0.1 nm, digital integration time (D.I.T.) of 8 sec, bandwidth of 2 nm and scanning speed of 50 nm·min⁻¹. A CD spectrum of the suitable buffer was subtracted to each sample spectrum before further data analysis.

CD data is presented as the mean residue ellipticity at wavelength λ , given by the following equation⁷⁷ (Eq. 3.4):

$$[\theta]_{mrw,\lambda} = \frac{MRW \times \theta_{\lambda}}{10 \times d \times c} \quad \text{Equation 3.4}$$

Where θ_{λ} is the observed ellipticity (degrees) at wavelength λ , d is the pathlength (cm) and c is the protein concentration ($\text{g}\cdot\text{mL}^{-1}$). MRW stands for Mean Residue Weight which is calculated from $M/(N-1)$ where M is the protein molecular mass (Da) and N is the number of amino acid residues.

Thermal stability

To assess the proteins thermal stability, CD spectra of the samples above were collected with the same acquisition parameters: at 20°C; after a thermal denaturation increasing 1 °C·min⁻¹ until 90 °C; and then after a re-cooling step to 20 °C.

Storage stability

Since TSSFP was designed to be used as a possible biopharmaceutical it is important to understand not only its thermostability but also its storage stability. To address this question, the secondary structure of two samples containing 0.2 mg·mL⁻¹ TSSFP in PBS pH 7.4, 300 mM NaCl and 10 % glycerol, which were stored at 4 °C or RT (≈ 24 °C), were monitored by CD at the following storage days: 1, 3, 5, 8, 15, 22 and 29. Spectra were collected and represented as mentioned in the section “Secondary structure characterization”.

Dynamic light scattering

Dynamic Light Scattering (DLS) is a rapid, non-invasive, and extremely versatile technique used in several applications such as detecting protein aggregation, protein-protein interaction and complex formation.⁸⁰ DLS measures the Brownian motion of particles in a dispersion enabling to determine their hydrodynamic size which is defined as the size of a hypothetical sphere that diffuses at the same rate as the particle under study.

Brownian motion is the random movement of a particle as a result of collisions with its surrounding molecules. The rate of Brownian motion is quantified as the translational diffusion coefficient (D) and it is related to the hydrodynamic size of a particle by the Stokes-Einstein Equation (Eq. 3.6).

$$D = \frac{k_B T}{6\pi\eta R_H} \quad \text{Equation 3.6}$$

where k_B is the Boltzmann constant ($1.38064852 \times 10^{-23} \text{ J}\cdot\text{K}^{-1}$), T is the temperature (K), η is the absolute viscosity (Pa·s) and R_H is the hydrodynamic radius (m).⁸¹

In DLS instruments, the sample is irradiated with a laser beam and the fluctuations of the scattered light are detected at a known scattering angle by a photon detector. The fluctuations result of the interference of light scattered by each particle; therefore, the intensity will change over time as the particles diffuse. The particles diffusion rate affects the speed of the intensity fluctuations, since smaller particles diffuse quickly which translates into more rapid fluctuations in the scattered light and the opposite occurs for bigger particles. The autocorrelation function represents how fast the intensity fluctuates over time and allows to extract the translational diffusion coefficient values and, thus the hydrodynamic size.⁸²

Besides the translational diffusion coefficient, other variables such as temperature and solvent viscosity are known to affect the particle diffusion rate, as shown in equation 3.6.

The size distribution measured is presented as an intensity distribution, which shows the intensity of scattered light from each size population present in the sample. These can also be converted into a number size distribution to evaluate the population prevalence, for example.

Peptide aggregation

The aggregation of the SARS-CoV-2 putative FPs FP1, FP2 and FP3 in aqueous solution, was monitored by changes in the DLS profile. Samples with different peptide concentrations ranging from 5 to 100 μM were prepared by diluting each peptide stock solution (in DMSO) with 10 mM HEPES pH 7.4, 150 mM NaCl. DLS measurements were performed in the Zetasizer Nano ZS90 (Malvern Panalytical), at 37 °C in quartz Ultra-Micro Cell for Nano series (Malvern Panalytical) cuvettes. Measurements consisted of at least 5 individual runs, each corresponding to an averaged autocorrelation curve obtained from at least 10 repeated sample scans. D values were calculated from autocorrelation curves using a CONTIN-based method⁸³. D values were used to determine intensity-averaged and number-averaged R_H distributions through the Stokes-Einstein equation (Equation 3.6).

3.7.2. Functional characterization

Surface plasmon resonance

Surface plasmon resonance (SPR) is a technique to monitor macromolecules interactions in real time as well as to determine specificity, affinity, and kinetic parameters. The SPR phenomenon occurs when a photon of incident light hits a metal surface, such as gold, exciting the electrons on the metal surface layer triggering electron movements, which are known as plasmon. As the result of plasmon oscillations, an electric field is generated between the metal surface and the sample solution.

In a typical SPR experiment (Figure 3.5), one interacting partner – the ligand - is immobilized on to the surface of the sensor chip. Then, a solution containing the other interaction partner – the analyte - is flown over the surface enabling the binding. The ligand-analyte binding induces a change in the refractive index at the SPR sensor surface which is expressed in resonance or response units (RU). The RU value at the start of the assay corresponds to the starting critical angle and a shift in this angle of 10^{-4} degree is equivalent to 1 RU.

The response is plotted as a function of time, creating a sensorgram through which the interaction progress can be characterized. Additionally, the detected response is directly proportional to the macromolecule's concentration on the sensor surface.

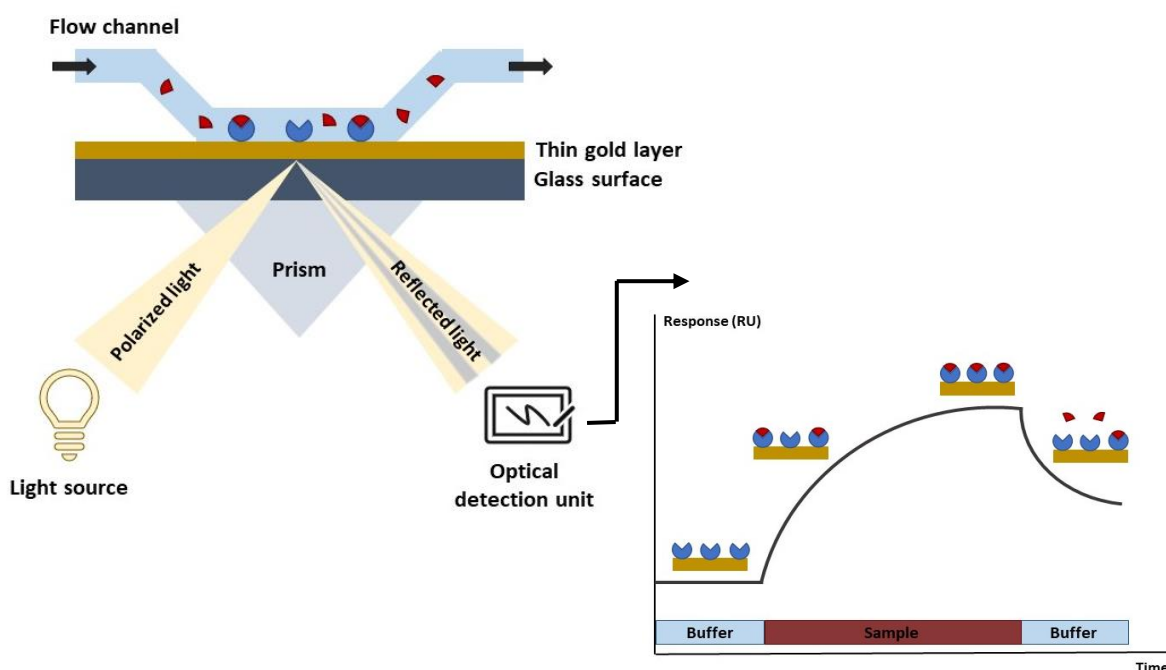


Figure 3.5 | Schematic representation of the SPR principle and the sensorgram generated.

Affinity of RIChs for RBD

The SPR assays reported here were performed by Dr. Pedro M. Sousa from the Merck Satellite Lab and Structure Biology for Drug Discovery Unit at iBET.

The affinity of RIChs 1-5 and scaffolds 1 and 2 for the RBD (kindly provided by Dr. Rute Castro and Dr. Paula Alves, iBET) was experimentally evaluated by SPR in a Biacore 4000 (GE Healthcare) equipment. The surface of the carboxymethylated dextran (CM5) chip was previously activated with a 1:1 ratio of 400 mM 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide and 100 mM N-hydroxysuccinimide for 10 min. RBD was diluted to $2 \mu\text{g}\cdot\text{mL}^{-1}$ in the selected buffer, based on literature data⁸⁴ (10 mM Bis-Tris pH 6.5) and immobilized onto the CM5 sensor chip according to standard amine coupling procedure, followed by a blockage step of the remaining carboxymethylated groups activated with a 7 min injection of 1M ethanolamine pH 8.5.

The RIChs 1-5, scaffolds 1 and 5 and controls were diluted in running buffer 10 mM Tris-HCl, 300 mM NaCl, 5 mM MgCl_2 , 0.1 mM EDTA, 0.05% (v/v) tween-20, 1 mM DTT, pH 7.5 and injected at eleven different concentrations using two-fold serial dilutions beginning with: 20 μM for RIChs 1,3 and 5, scaffolds 1 and 5 and BSA; 5.8 μM for RICh 2; 4.5 μM for RICh 4; and 0.5 μM for hACE2. BSA and hACE2 were used as negative and positive control, respectively. The low starting concentration of RIChs 2 and 4 was the result of the reduced purification yield. All assays were conducted at 25 °C.

The obtained sensorgrams were processed by subtracting the buffer blank from the binding responses recorded for each protein. The Biacore 4000 evaluation software was used to determine the binding constants from the resulting plots of steady-state SPR response against proteins concentration.

Dye-free Differential Scanning Fluorimetry

The dye-free differential scanning fluorimetry (dye-freeDSF) methodology monitors the intrinsic tryptophan fluorescence, reflecting changes in the protein microenvironment. Usually, these can be the result of diverse events, such as: folding or unfolding as a function of temperature; structure stabilization by cofactors or ligands; conformational changes induced by protein-protein interactions⁸⁵ and other factors affecting the protein three-dimensional structure.

In this technique, the tryptophan fluorescence is excited at 275 nm and emission is recorded at 330 nm and 350 nm. In an apolar environment, the tryptophan fluorescence peak is center around 330 nm. When located in a polar environment, the emission peak tends to decrease in intensity and shifts to 350 nm due to quenching by solvent molecules.⁸⁶ More details regarding fluorescence spectroscopy are given ahead in the section “fluorescence spectroscopy”.

Changes in the structural stability induced by the RBD interaction

In this functional characterization section, dye-freeDSF was used as a complementary technique to evaluate the interaction between RICH 1 and RBD by comparing the RICH 1 thermal denaturation profile in the presence and absence of RBD. Three different samples were prepared in triplicates in 20 mM Tris-HCl pH 7.5, 300 mM NaCl, 100 μ M TCEP: one containing 75 μ g·mL⁻¹ RBD; another with 740 μ g·mL⁻¹ RICH 1; and another including 490 μ g·mL⁻¹ RICH 1 and 75 μ g·mL⁻¹ RBD (10:1 molar ratio). High sensitivity capillaries were filled with 10 μ L of each sample. The temperature gradient was performed at 1 °C·min⁻¹ from 20 to 90°C and the intrinsic tryptophan fluorescence emission recorded. The thermal denaturation curves obtained at 330 nm, 350 nm and the fluorescence ratio 350/330 as well as the 330 nm first derivative were used for data analysis.

Immunization program

The immunogenic construct, TSSFP, was used for the immunization of two rabbits in order to test its ability to induce an immune response where antibodies against the epitope were produced. The immunization program was performed by Eurogentec according to the schedule in Figure 3.6.

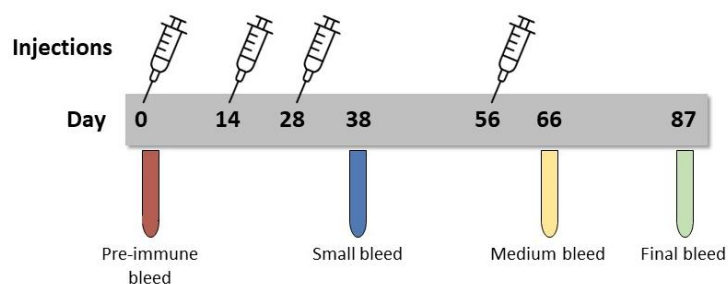


Figure 2.6 | Immunization program schedule with the boosts and bleeds days.

In each of the four injections, 256 µg of TSSFP were administrated with Freund's adjuvant at days: 0, 14, 28 and 56. Sera samples were collected as follows: pre-immune bleed at day 0; small bleed at day 38; medium bleed at day 66; and final bleed at day 87. At the end of the immunization program, part of the final bleed samples was purified by the company using affinity chromatography with the TSSFP, resulting in the purified antibodies.

Assessment of the TSSFP-induced antibodies

Sera samples from the immunization program described above were evaluated for the presence of antibodies capable of recognizing the spike-glycoprotein with the following techniques.

Western blot

The presence of antibodies able to bind to TSSFP and SARS-CoV-2 spike-glycoprotein was first assessed by western blot. For each serum sample, 200 ng of TSSFP and Spike-glycoprotein (kindly provided by Dr. Rute Castro and Dr. Paula Alves, iBET) were loaded into a NuPAGE™ 10% Bis-Tris gel (invitrogen) and after electrophoresis at 200V for 35 min (Amersham Pharmacia Biotech EPS 601), proteins were transferred to a PVDF membrane using the iBlot™ 2 PVDF Regular Stacks and the Dry Blotting System (Thermo Fisher Scientific). A blocking step with 5% (w/v) skim milk (Sigma-Aldrich) in TBS-T, for 1h at room temperature under gentle stirring was followed by an overnight (approximately 16h) incubation at 4 °C with 1:1,000 serum sample diluted with 5% skim milk powder in TBS-T. After three wash steps with TBS-T for 10 min, at room temperature and under gentle stirring, an incubation with 1:20,000 secondary antibody Anti-Rabbit IgG (whole molecule)-Peroxidase antibody produced in goat (Sigma-Aldrich) in 5% (w/v) skim milk in TBS-T for 1h, at room temperature and under gentle stirring was performed. Three washing steps were done before developing the membrane with SuperSignal™ West Pico Chemiluminescent Substrate (Thermo Fisher Scientific) in the iBright™ FL1500 (Thermo Fisher Scientific) imaging system.

Enzyme linked immunosorbent assay

After the WB, the antibodies ability to bind to SARS-CoV-2 spike-glycoprotein was also assessed by enzyme linked immunosorbent assays (ELISA) in collaboration with the Lymphocyte Physiology Lab at IGC.

In this assay, each well of 96-well plates was coated with 50 µL spike protein at 0.5 µg·mL⁻¹ in PBS and incubated overnight at 4 °C. After washing with PBS, the plates were blocked with 150 µL of 2 % BSA in PBS for 1h at room temperature. The sera samples from the immunization program – pre-immune, small, medium and final bleed - and the purified antibodies were subjected to a three-fold dilution series from 1:100 to 1:218,700 in 1 % BSA-PBS and incubated for 2h at room temperature. Another washing step was followed by the incubation with the secondary antibody goat anti-rabbit IgG-HRP (Santa Cruz Biotechnology, sc-2004) at 1:4,000 for 2h at room temperature. Plates were washed once again and incubated 10 min with 50 µL/well of TMB substrate (BioLegend). The reaction was stopped with 50 µL/well sulfuric acid and read at 450 nm on a plate reader. The endpoint titer was defined as the reciprocal of the highest analyte dilution that has an absorbance at 450 nm higher than 0.08.

Neutralization assays

The neutralization assays reported here, were accomplished thanks to a collaboration with the Cell Biology of Viral Infection Lab at IGC. In these assays, the 293T cell line engineered to express the human ACE receptor was used as the target host of the Spike-pseudotyped mutant lentiviruses described by M. Alenquer et al.⁸⁷

The main goal in these assays was to evaluate the RIChs and the antibodies produced against the immunogenic construct ability to prevent infection either by blocking the interaction between the RBD and ACE2 receptor or by shielding the S2' cleavage site and the FP region, respectively.

RICh 1, 3 and 5 samples were three-fold serially diluted over six concentrations, starting with 500 $\mu\text{g}\cdot\text{mL}^{-1}$. The distinct samples from the immunization program – pre-immune, small, large and final bleed and the purified antibodies – were four-fold serially diluted over six dilutions, beginning with a 1:10 dilution. Each dilution was incubated with spike-pseudotyped lentiviral particles for 1h at 37 °C. Next, the mixtures were added to a pre-seeded 96-well plate of 293T-ACE2 cells, with a final MOI of 0.2. As infection positive control, cells were infected in the absence of any inhibitor and, for background control, wells containing only cells were used. The fluorescent signal was determined after 48h of post-transduction using the GloMax Explorer System (Promega). All samples were analyzed in triplicates.

In the data analysis, the fluorescence units of the background control with cells only were subtracted to each sample fluorescence and then, these were normalized to the positive control. The half-maximal or 50 % neutralization titer (NT50), defined as the dilution reciprocal corresponding to a 50 % decrease in infection, was determined applying the four-parameter nonlinear regression (least squares regression without weighting; constrains: bottom = 0) (GraphPad Prism 5).

Spike glycoprotein cleavage assay

Spike-glycoprotein cleavage assays were performed to better understand the binding between this viral protein and the antibodies generated in response of the TSSFP. The antibodies are expected to bind to the cleavage site within spike S2 subunit and prevent proteolytic cleavage. To address this question, 300 ng of soluble SARS-CoV-2 spike-glycoprotein, either non-incubated or previously incubated with 600 ng purified antibodies, was added to a 96-well plate seeded with 293T-Ace2 cells. The cells and the extracellular medium were collected after 1h and 6h incubation.

Samples were loaded into a NuPAGE™ 10% Bis-Tris gel (Invitrogen) and after separation by electrophoresis at 200V for 35 min, the proteins were transferred to a PVDF membrane using the iBlot™ 2 PVDF Regular Stacks and the Dry Blotting System (Thermo Fisher Scientific). Next, a blocking step with 5% (w/v) skim milk (Sigma-Aldrich) in TBS-T was performed for 1h at room temperature under gentle stirring. The primary antibody SARS-CoV-2 Spike Protein S2 polyclonal antibody (PA5-114534 invitrogen) was diluted 1: 1000 in 5% (w/v) skim milk in TBS-T and incubated with the membrane for 1h, at room temperature and under gentle stirring, followed

by three wash steps with TBS-T for 7 min each. The secondary antibody Anti-Rabbit IgG (whole molecule)-Peroxidase (Sigma-Aldrich) was diluted and incubated as the primary antibody. After three wash steps, the WB was developed with SuperSignal™ West Pico Chemiluminescent Substrate (Thermo Fisher Scientific) in the iBright™ FL1500 (Thermo Fisher Scientific) imaging system.

Liposome preparation and characterization

Lipid films were prepared with the appropriate amount of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC, Avanti® Polar Lipids), 1-Palmitoyl-2-oleoyl-sn-glycero-3-(phospho-rac-(1-glycerol)) (POPG, Avanti® Polar Lipids) and cholesterol (Chol, Avanti® Polar Lipids) solubilized in chloroform according to the following ratios: POPC, POPC:POPG (4:1) and POPC:POPG:Chol (3:1:1). Stock solutions were dried under a N₂ stream and kept under vacuum overnight to eliminate the organic solvent. The lipid films were hydrated with 10 mM HEPES 150 mM NaCl pH 7.4, followed by ten freeze-thaw cycles. The resulting multilamellar vesicles (MLVs) were extruded using polycarbonate filters with a pore size of 100 nm to obtain large unilamellar vesicles (LUVs).

DLS was employed to monitor the LUVs hydrodynamic size in the presence of increasing FPs concentration. Samples containing 0.1 mM of POPC or POPC:POPG (4:1) LUVs and different lipid-to-peptide (lip/pep) molar ratios from 2 to 40 were prepared and analyzed as stated in the *peptide aggregation* section.

Fluorescence spectroscopy

Fluorescence is the result of a three-stage process that certain molecules, named fluorophores, undergo. This process is illustrated by the Jablonski diagram in Figure 3.7.

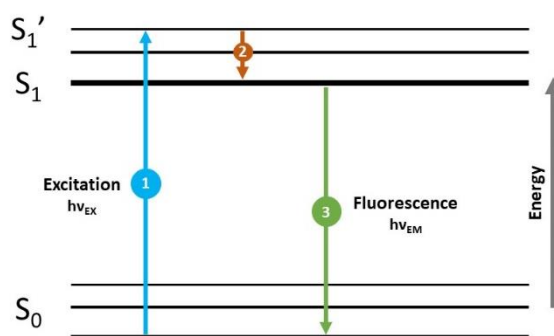


Figure 2.7 | Jablonski diagram illustrates the stages responsible for fluorescence generation. (1) Excitation (2) Excited-state lifetime (3) Fluorescence emission.

A fluorophore in the singlet ground state (S_0) reaches the excited electronic single state (S_1') when light with a certain energy ($h\nu_{EX}$) is applied (1). After light absorption, several processes affect the excited-state lifetime (2) such as conformational changes, interactions with the fluorophore molecular environment or energy transfer resulting in the dissipation of energy

and, consequently creating a relaxed singlet excited state (S_1) from which fluorescence emission occurs. A photon of specific energy ($h\nu_{EM}$) is emitted when the fluorophore returns to its ground state (3).

As a result of the energy dissipation during the excited-state lifetime, the photon emitted has lower energy than the excitation one, thus the excitation wavelength is higher than the emission wavelength. This energy or wavelength difference ($h\nu_{EX}-h\nu_{EM}$) is called the Stokes shift.

Generally, the emission spectrum displays the same vibrational energy spacing as the absorption spectrum resulting in symmetric spectra or a mirror image of the absorption bands. This can be explained by the Franck-Condon principle which states that all electronic transitions are vertical, so they occur without change in the position of the nuclei⁸⁸

One of the most important characteristics of a fluorophore is its quantum yield (Φ). Quantum yield is the ratio between the number of emitted and absorbed photons. Equation 3.7 relates the quantum yield and both radiative (k_r) and nonradiative (k_{nr}) pathways associated with the transition from the excited state to the ground state.

$$\Phi = \frac{k_r}{k_r + \sum k_{nr}} \quad \text{Equation 3.7}$$

Nonradiative transitions can be divided into diverse categories including vibrational relaxation, intersystem crossing and internal conversion. Vibrational relaxation occurs when the energy is dissipated from the molecule to its surroundings. Intersystem crossing corresponds to a transition from a singlet to a triplet state which can be followed by phosphorescence while internal conversion happens when a vibrational state of an excited state is coupled to a lower electronic transition one.

Another nonradiative process is the fluorescence resonance energy transfer (FRET) which may occur when the emission spectrum of a fluorophore (the donor) overlaps with the absorption spectrum of a nearby molecule (the acceptor). The FRET efficiency (E) of a given donor-acceptor pair is determined by the Equation 3.8.

$$E = \frac{R_0^6}{R_0^6 + r^6} \quad \text{Equation 3.8}$$

E is dependent of the Förster distance (R_0) which represents the molecular separation at which energy transfer achieves 50% efficiency and of the distance between both molecules (r).

Fluorescence spectroscopy is a non-destructive and highly versatile technique which can be applied to characterize and evaluate diverse phenomena. In this project, fluorescence spectroscopy was employed to examine the RBD-RICH 1 complex, the aggregation of SARS-CoV-2 putative FPs in aqueous solution and the FPs-induced lipid mixing or fusion and perturbation of LUVs.

Peptide aggregation

The aggregation of the peptides was monitored using the 1-anilinonaphthalene-8-sulfonate (ANS), a wide-used probe to study the local polarity of biological microenvironments. This is a non-covalent probe with a reduced fluorescence intensity in aqueous solutions which is substantially enhanced with the decrease of polarity of its environment. Additionally, a blue shift

(λ_{em} = 520 nm) is observed when ANS is in hydrophobic environments or in more viscous solutions.⁸⁹

Fluorescence measurements were performed in an FLS920 Fluorescence Spectrometer (Edinburgh Instruments) using Hellma® fluorescence quartz cuvettes (Merck), at 37 °C. The excitation wavelength was 365 nm and the emission wavelength was set from 400 to 600 nm. FP1, FP2 and FP3 were added into a 500 μ L solution containing 12.8 μ M ANS, through consecutive additions until achieving a concentration of 50 μ M for FP1 and 25 μ M for FP2 and FP3. Each addition was followed by a 15 min incubation before measurement.

Lipid mixing induced by FPs measured by FRET

In lipid mixing assays, unlabeled LUVs were mixed with labeled LUVs containing 1 mol% of fluorophores NBD-PE (Avanti® Polar Lipids) and RhB-PE (Avanti® Polar Lipids) at a 4:1 unlabeled:labeled molar ratio with 100 μ M total lipid. The lipid compositions used were: POPC, POPC:POPG (4:1), POPC:POPG:Chol (3:1:1). Lipid mixing between probe-containing vesicles and unlabeled ones results in a decrease in the FRET. Thus, the increase in the emission intensity of the donor, NBD-PE (530 nm), or the decrease of the emission intensity of the acceptor RhB-PE (588 nm) may be used to determine the extent of membrane fusion. The first option was used to calculate the percentage of lipid mixing or fusion according to Equation 3.9:

$$\%fusion = \frac{I_{530nm} - I_{530nm_0}}{I_{530nm_{100}} - I_{530nm_0}} \times 100 \quad \text{Equation 3.9}$$

where I_{530nm} is the fluorescence intensity of the sample measured at 530 nm, I_{530nm_0} and $I_{530nm_{100}}$ corresponds to the fluorescence intensity at 530 nm of the probe-labeled LUVs before peptide addition and with 0.2% (v/v) triton X-100, respectively.

Fluorescence measurements were performed in an FLS920 Fluorescence Spectrometer (Edinburgh Instruments) using Hellma® fluorescence quartz cuvettes (Merck), at 37 °C. The excitation wavelength was 470 nm and the emission wavelength was set from 500 to 650 nm. Peptides were added to 350 μ L of 100 μ M LUVs in 10 mM HEPES 150 mM NaCl pH 7.4, through consecutive additions until achieving a maximum lip/pep ratio of 4 for FP1, FP2 and FP3, and 20 for FM. Each addition was followed by a 15 min incubation. All samples except for FM were tested in duplicates.

Leakage assay

For this assay, the lipid films were hydrated with 10 mM HEPES 150 mM NaCl pH 7.4 containing 50 mM of 5,6-carboxyfluorescein (5,6-CF) probe. Permeabilization of the liposome's membrane induced by the peptides was evaluated by monitoring the release of the LUVs internal content. The 5,6-CF inside the LUVs is in a concentration that induces self-quenching and the measured fluorescence emission intensity is low. When the molecules are released from the vesicles the fluorescence emission intensity increases.

The intensity of the fluorescence emission was monitored at 37 °C using a Varian Cary Eclipse Fluorescence Spectrometer (Agilent) with the excitation and emission wavelengths at 490 and 520 nm, respectively. First, the stability of each LUV composition used was examined with samples containing 10 μ M LUVs under stirring for 20 min. Then 0.5 % (v/v) triton X-100 was added, and the fluorescence emission intensity was recorded for 10 min. Distinct LUVs

compositions were tested in these assays including POPC, POPC:POPG (4:1), POPC:POPG:Chol (3:1:1). Various lip/pep ratios were analyzed: 2 and 20 for FP1; 2, 20 and 80 for FP2 and FP3; 20 and 80 for FM. After 5 min LUVs stabilization, the peptides were added and the fluorescence intensity was followed for 10 min. Next, triton X-100 was added and the last 5 min of the experiment were conducted. FP1, FP2 and FP3 samples were analyzed in triplicates while FM samples were examined in duplicates due to the limited amount of available peptide.

The percentage of leakage was determined with Equation 3.10:

$$\% \text{ leakage} = \frac{I_{\text{peptide}} - I_0}{I_{100} - I_0} \times 100 \quad \text{Equation 3.10}$$

where I_{peptide} is the fluorescence emission intensity measured after peptide addition to the LUVs, I_0 and I_{100} corresponds to the fluorescence intensity of LUVs before peptide addition and with 0.5 % (v/v) triton X-100, respectively.

4. Results and discussion

4.1. RBD-interacting chimeric proteins

4.1.1. *In silico* design

The RBD-interacting chimeric (RICHs) proteins aiming to block the RBD-ACE2 binding were designed using the Rosetta software package. As explained in the *Materials and Methods* chapter, the used strategy consisted in grafting the ACE2 helical motif that interacts with the RBD onto stable scaffolds. In the protocol, four different scaffolds revealed to comprise suitable features for design, their crystallographic structures having the following PDB IDs: 1VXJ, 6ANR, 5XEF and 1QYS. Apart from the last, all the scaffolds are natural helix bundle proteins composed of at least 4 helices. 1QYS is the computationally designed protein which has a secondary structure composed of two α -helices and five β -strands.

An additional candidate (RICH 5) was designed by grafting a segment of the ACE2 H1 into scaffold 5XEF without mutations affecting the interface region. The structure of each of the five most promising candidates in the context of the RBD is presented in Figure 4.1.

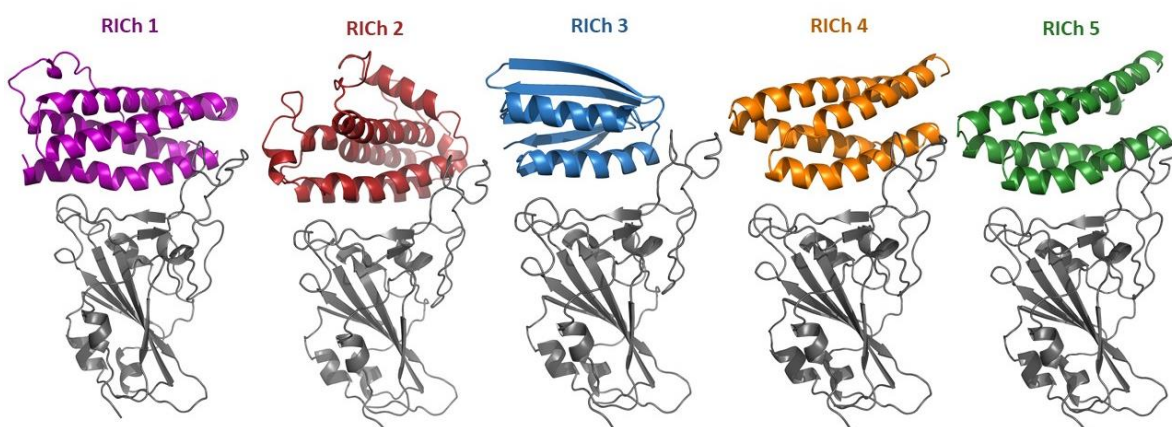


Figure 4.1 | RBD-interacting chimeric proteins (RICHs) structure in the context of the SARS-CoV-2 RBD (grey). (Purple) RICH 1 designed using the 1VXJ scaffold. (Red) RICH 2 based on scaffold 6ANR. (Blue) RICH 3 built with the rationally designed protein 1QYS. (Orange) RICH 4 designed using 5XEF as scaffold. (Green) RICH 5 composed of 5XEF scaffold and the conserved H1 segment of ACE2.

In all cases, the RICH's interface is composed of two parallel α -helices one being the ACE2 motif grafted while the other belongs to the scaffold native structure. The RBD-ACE2 interaction hotspots were maintained except for Asp38_{ACE2} which was substituted by a tryptophan in RICHs 2 and 4. The Asp38_{ACE2} originally forms a hydrogen bond with Tyr505_{RBD} and a salt bridge with Arg403_{RBD}. The substitution to a tryptophan breaks these interactions but establishes a π -stacking with Tyr505_{RBD}. Other native ACE2-RBD interactions were maintained in RICHs 1 and 5, including the hydrophobic contacts between Leu79_{ACE2}, Met82_{ACE2}, Tyr83_{ACE2} and Phe486_{RBD}. RICHs 1 and 5 were rated as the best candidates by the *in silico* evaluation.

RICH 1

Five residues of the RBD-ACE2 interface were preserved in the RICH 1 interface shown in Figure 4.2 (Asp127/Asp30_{ACE2}, Glu134/Glu37_{ACE2}, Asp135/Asp38_{ACE2}, Tyr138/Tyr41_{ACE2} and Leu52/Leu79_{ACE2}), two have similar properties (Leu56/Met82_{ACE2} and Gln132/Glu35_{ACE2}) and four were the result of interface improvement aiming to increase the binding affinity and effectiveness (Tyr120/Glu23_{ACE2}, Val124/Thr27_{ACE2}, Trp128/Lys31_{ACE2} and Ala131/His34_{ACE2}).

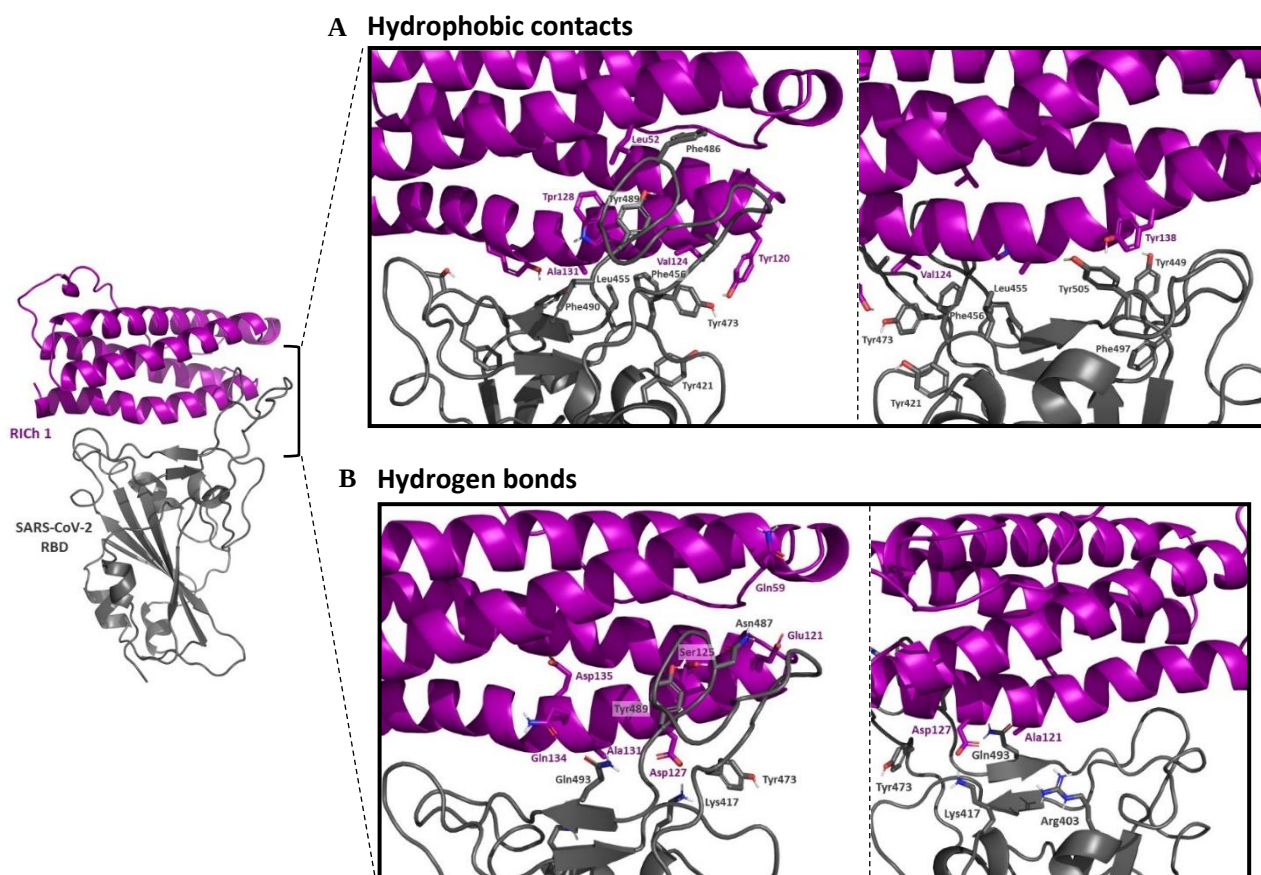


Figure 4.2 | RICH 1-RBD interface. (A) Hydrophobic contacts established between RICH 1 (purple) and SARS-CoV-2 RBD (grey). (B) Hydrogen bonds on the binding interface of RICH 1 (purple) and RBD (grey).

RICH 1-RBD interaction is dominated by hydrophobic contacts (Figure 4.2A) between three aromatic residues of RICH 1 and multiple residues of the RBD surface. The conserved Tyr138 is able to interact with Tyr449_{RBD}, Phe497_{RBD} and Tyr505_{RBD}. Tyr120 may form at least one π -stacking interaction with Phe456_{RBD}, Tyr421_{RBD} or Tyr473_{RBD} and it also establishes a hydrophobic contact with Ala475_{RBD}. Similarly, Trp128 may interact with Phe456_{RBD}, Tyr489_{RBD} or Phe490_{RBD}. The substitution of Thr27_{ACE2} with Val124 in a high hydrophobic region allows strengthening contacts with Phe456_{RBD}, Tyr473_{RBD}, Ala475_{RBD} and Tyr489_{RBD}. Moreover, the His131 substitution with Ala allows the contact with Leu455_{RBD}, while the Tyr substitution with Leu52 enables a persistent interaction with Phe486_{RBD}.

Aside from the extensive hydrophobic contacts, hydrogen bonds are also present in the RICH 1-RBD interface (Figure 4.2B). Gln59, Asp127, Gln134 and Asp135 ensure the native hydrogen bonds with Arg403_{RBD}, Lys417_{RBD}, Asn487_{RBD}, and Gln493_{RBD} while Glu121, Ser125 and Ala131 form additional hydrogen bonds with Tyr473_{RBD}, Tyr489_{RBD}, and Gln493_{RBD}.

RICH 2

Three residues of the RICH 2 interface (Figure 4.3) were maintained from ACE2 H1 helix: Asp30_{ACE2}, Lys31_{ACE2} and Leu45_{ACE2}. The remaining residues were either changed to residues with similar properties to mimic some native interactions or were mutated to residues with different properties to create new contacts.

RICH 2 and RBD binding occurs through diverse polar contacts: Lys10-Asp405_{RBD}, Asn13-Tyr505_{RBD}, Thr109-Asn487_{RBD}/Tyr489_{RBD}, Asp101-Lys417_{RBD}, Asp119-Tyr449_{RBD} and Tyr122-Thr500_{RBD}. Hydrophobic interactions established by the following residues are also present in the RICH 2-RBD interface: Met100-Phe486_{RBD}, Val108-Phe456_{RBD}/Ala475_{RBD} and Val115-Leu455_{RBD}. Additional contacts of Ile102, Tyr112 and Leu123 with Ser477_{RBD}, Gln493_{RBD} and Gln498 are observed in the complex formation. Tyr112 is also able to form a π -stacking contact with Tyr489_{RBD} contributing to a higher complex stability.

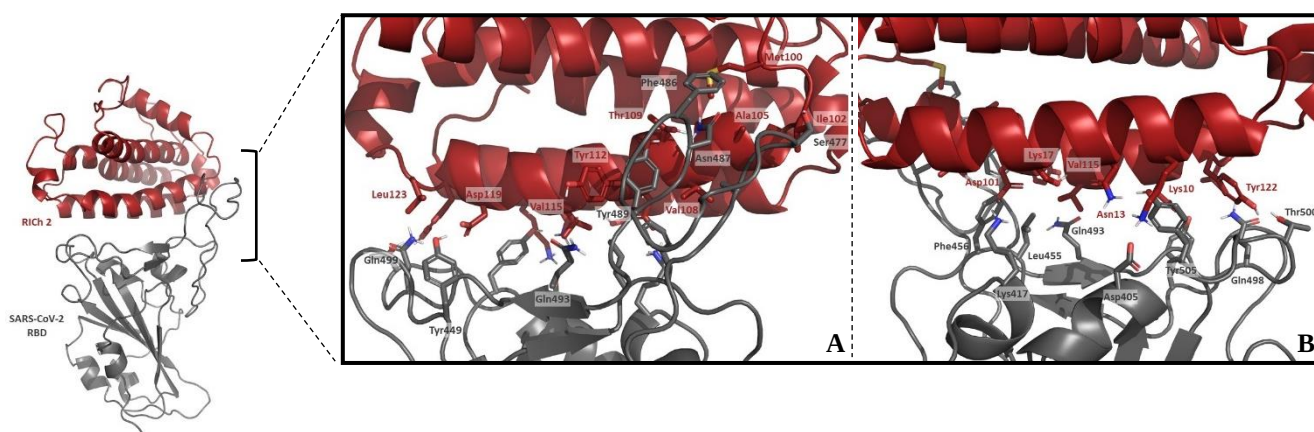


Figure 4.3 | Binding interface of RICH 2 (red) and RBD (grey). (A) Front view of the complex. (B) Back view of the complex.

RICH 3

RICH 3 interface comprises the residues highlighted in Figure 4.4 among which Ala25_{ACE2}, Lys26_{ACE2}, Asp30_{ACE2}, Asp38_{ACE2}, Tyr41_{ACE2} and Gln42_{ACE2} were conserved relatively to ACE2. The polar interactions on the RICH 2-RBD interface are established by His23-Tyr489_{RBD}, Lys24-Glu484_{RBD}, Asp58-Lys417_{ACE2}, Asp66-Tyr449_{RBD}/Gln498_{RBD}, Tyr69-Thr500_{RBD} and Gln70-Tyr449_{RBD}. A π -stacking formed by His23 and Phe486_{RBD} also contribute for the RICH 2-RBD binding.

The remaining contacts occur between Ala52, Ile55, Trp59, Val62, Ile63 and Tyr69 and Tyr503_{RBD}, Phe456_{RBD}, Tyr473_{RBD}, Ala475_{RBD}, Asn487_{RBD}, Gln493_{RBD} and Asn501_{RBD}.

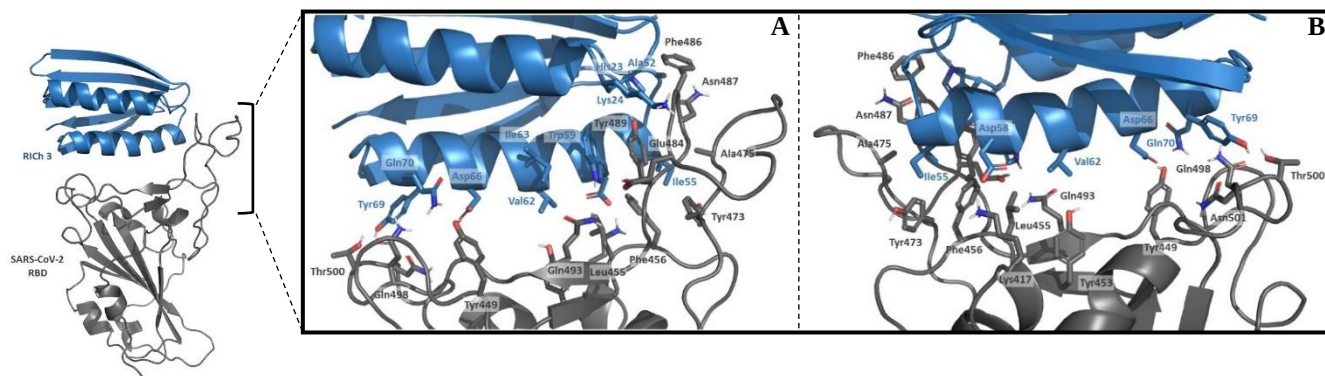


Figure 4.4 | RICH 3 (blue) and RBD (grey) interface. (A) Front view of the complex. (B) Back view of the complex.

RICH 4

RICH 4 interface (Figure 4.5) includes the ACE2 preserved residues Asp78 and Tyr81, while the Leu79_{ACE2}, Met82_{ACE2} and Tyr83_{ACE2} hydrophobic site is replaced by Val94 and Leu98. The conserved contacts comprise the hydrophobic interaction between Val94/Leu98-Phe486_{RBD}, the hydrogen bond formed by Tyr81-Thr500_{RBD} and the Asp78-Tyr449_{RBD} contact.

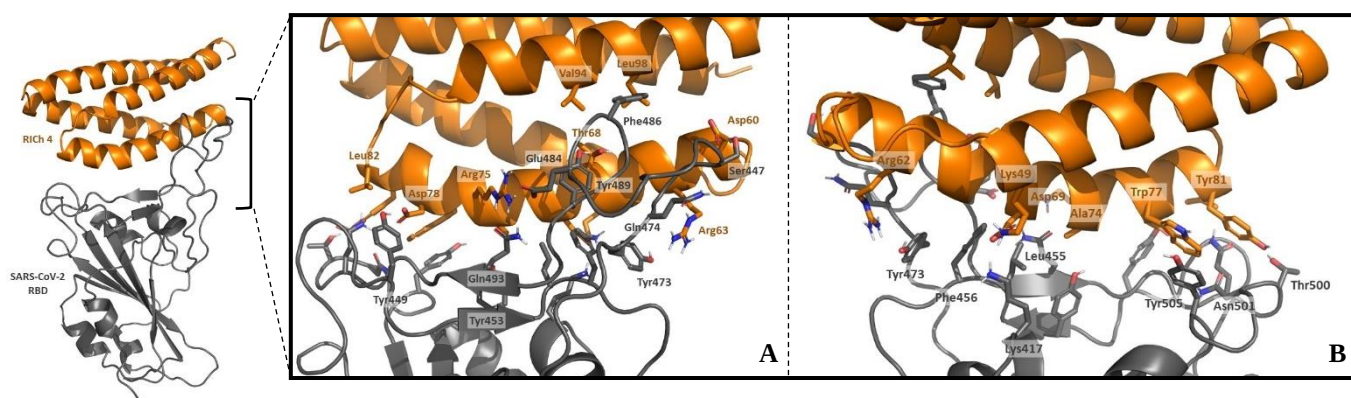


Figure 4.5 | Binding interface of RICH 4 (orange) and RBD (grey). (A) Front view of the complex. (B) Back view of the complex.

RICH 4-RBD binding is dominated by hydrogen and ionic bonds. Arg75 may establish interactions with two partners, namely Glu484_{RBD} and Gln493_{RBD}. An extensive network of hydrogen bonds in the interface is formed between multiple pairs of residues: Thr68-Tyr489_{RBD}, Asp69-Lys417_{RBD}, Asp60-Ser477_{RBD} and Arg63-Gln474_{RBD}. Aside from these, Trp77 and Tyr505_{RBD} share a π -stacking contact in addition to the hydrophobic interaction between Ala74-Leu455_{RBD}.

RICH 5

RICH 5 interface is presented in Figure 4.6. The residues Thr67, Phe68, Asp70, Lys71, His74, Asp78 and Tyr81 of the ACE2 were preserved while the Leu79_{ACE2}, Met82_{ACE2} and Tyr83_{ACE2} hydrophobic site was mimicked by the Val94 and Leu98 addition. Besides the native interactions,

new hydrogen bond contacts between Glu60-Tyr489_{RBD}/Ser477_{RBD}/Thr478_{RBD}, Asp78-Tyr489_{RBD} and Tyr81-Tyr448_{RBD} are possible.

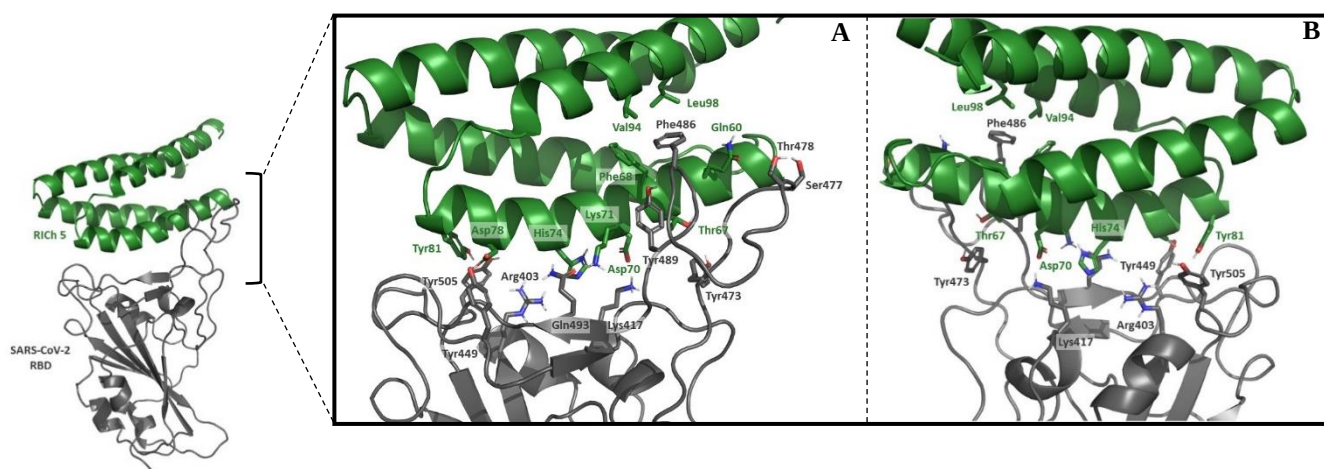


Figure 4.6 | Binding interface of RICH 5 (green) and RBD (grey). (A) Front view of the complex. (B) Back view of the complex.

4.1.2. Expression and purification

Scaffolds 1 and 5 are naturally existing proteins from *Thermotoga maritima* and *Bacillus cereus*. Therefore, it was possible to recombinantly express them in *E.coli* as already achieved in the literature.⁹⁰ On the other hand, RICHs 1-5 are novel engineered proteins and their stability was only predicted computationally. Still, their recombinant expression in *E.coli* was also successful.

RICH 1 purification was achieved in a two-step process which began by an affinity chromatography where the protein was eluted with two different imidazole concentrations: 100 mM and 250 mM, as shown in the chromatogram, SDS-PAGE gel and WB represented in Figure 4.7A, B and C. Next, the fractions within the green rectangle were concentrated and further purified by size-exclusion chromatography (Figure 4.7D, E and F). The elution volume (13.16 mL) corresponds to a molecular weight of 15.6 kDa, according to the calibration curve established for this column (Figure A5 in the annex section). This process had an overall yield of 0.5 mg of RICH 1 per culture medium liter.

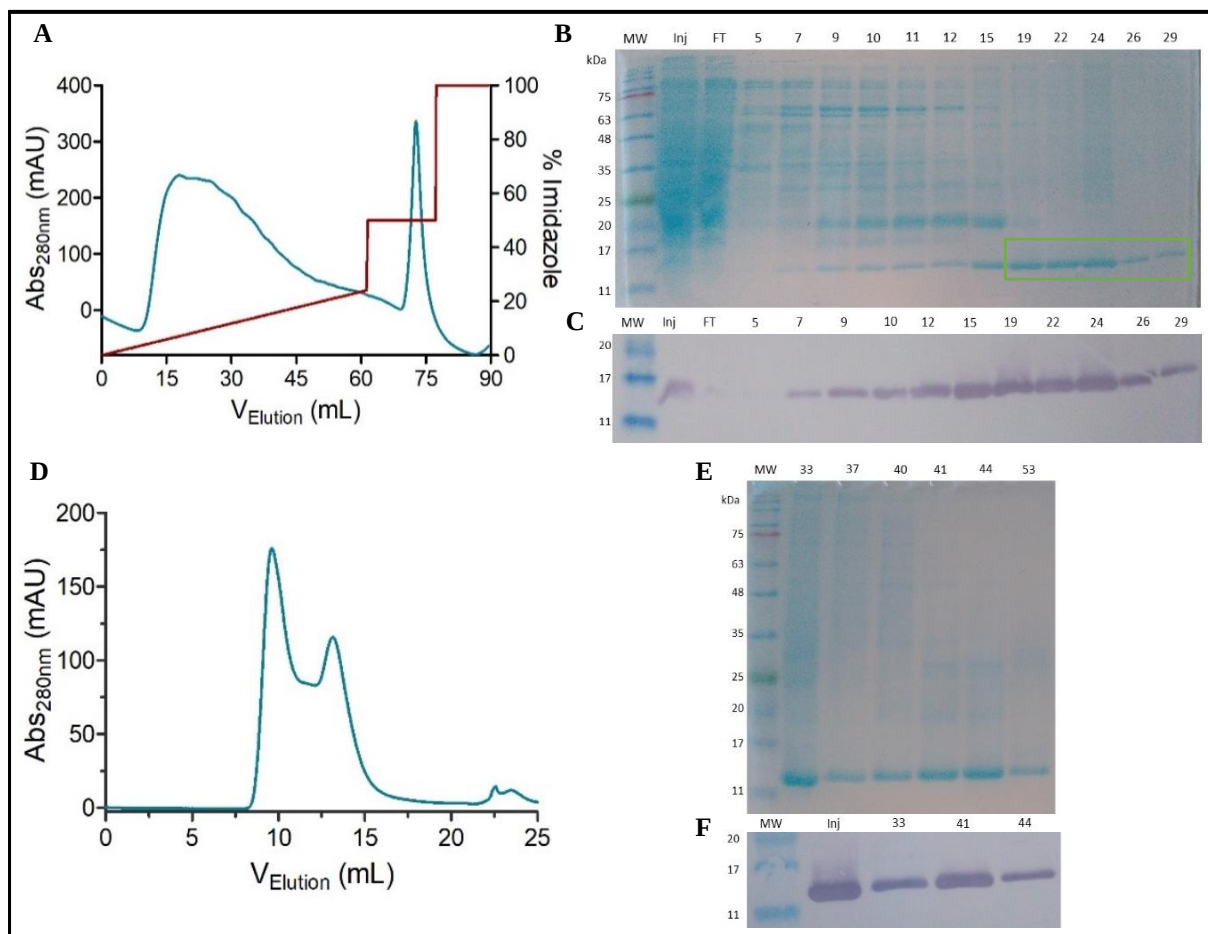


Figure 4.7 | Chromatograms and respective SDS-PAGE gels and WB of RICH 1 purification steps. (A) Elution profile of the affinity chromatography with HiTrap Fast Flow Crude column. (B) SDS-PAGE gel of the fractions collected during the affinity chromatography step. (C) WB of the affinity chromatography fractions. (D) Chromatogram obtained after the size-exclusion chromatography with Superdex 75 10/300 column. (E) SDS-PAGE gel where the fractions collected during the size-exclusion chromatography were examined. (F) WB of the size-exclusion chromatography fractions. Blue line – absorbance at 280 nm. Red line – elution gradient using buffer containing 500 mM imidazole. MW - NZYColour Protein Marker II. Inj – sample injected in each purification step. FT- flowthrough collected during the washing step. Numbers – number corresponding to the fractions collect during the chromatographies.

RICH 2 purification started with an affinity chromatography where the elution occurred with 191.5 mM imidazole, as shown in the chromatogram in Figure 4.8A. In the size-exclusion chromatographic steps, the RICH 2 elution volumes (respectively, 68.4 mL and 12.25 mL (Figure 4.8C and D) correspond to a molecular weight of 22.2 kDa. RICH 2 expression and purification process had an overall yield of 0.26 mg protein per culture medium liter.

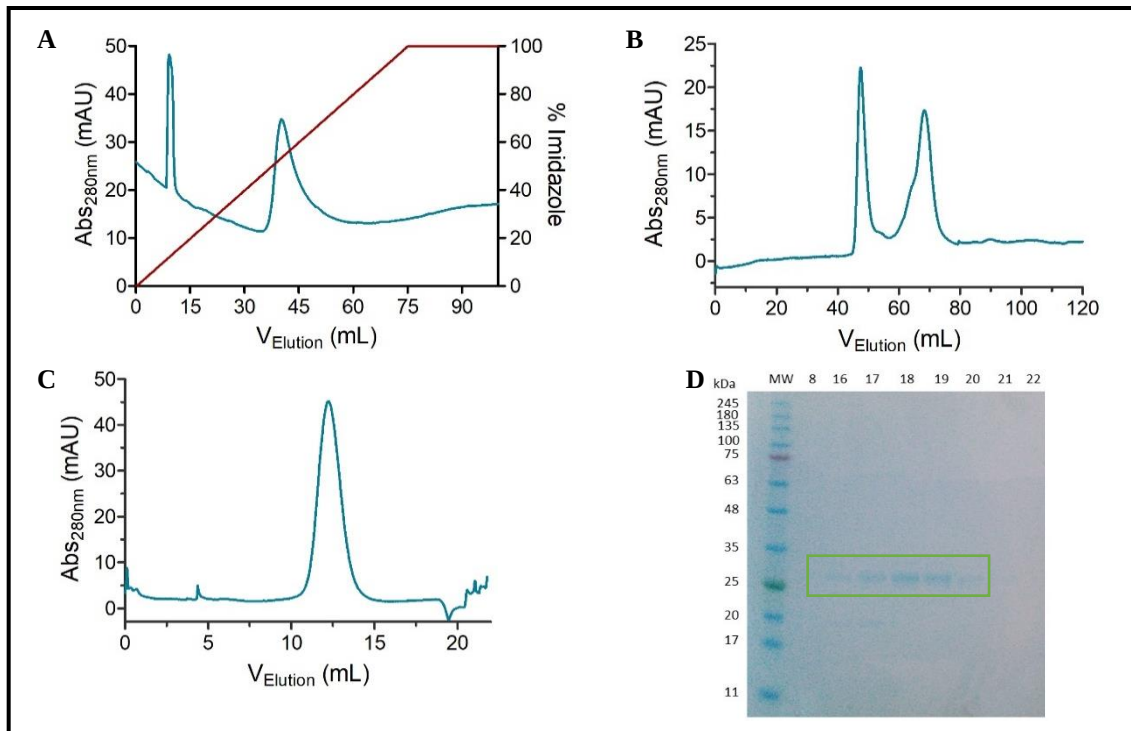


Figure 4.8 | Chromatograms of the RICH 2 three-step purification and the final SDS-PAGE gel. (A) Elution profile of the affinity chromatography with HiTrap TALON® Crude column. (B) Chromatogram acquired after the size-exclusion chromatography with Superdex 75 prep grade column. (C) Chromatogram of the last size-exclusion chromatography using a Superdex 10/300 GL column. (D) Analyzes of the fractions collected during the last purification step by SDS-PAGE. Blue line – absorbance at 280 nm. Red line – elution gradient using buffer containing 500 mM imidazole. MW - NZYColour Protein Marker II. Numbers – number corresponding to the fractions collect during the chromatography.

RICH 3 was first eluted with 109 mM imidazole from the affinity chromatography column, as indicated in the chromatogram and respective SDS-PAGE gel in Figure 4.9A and B. Next, RICH 3 was purified with a preparative size-exclusion chromatography column, as shown in Figure 4.9C and D. Lastly, in the final size-exclusion chromatography, the elution volume (12.74 mL) corresponds to a molecular weight of 18.4 kDa (Figure 4.9E and F). The yield of this process was 0.84 mg of RICH 3 per culture medium liter.

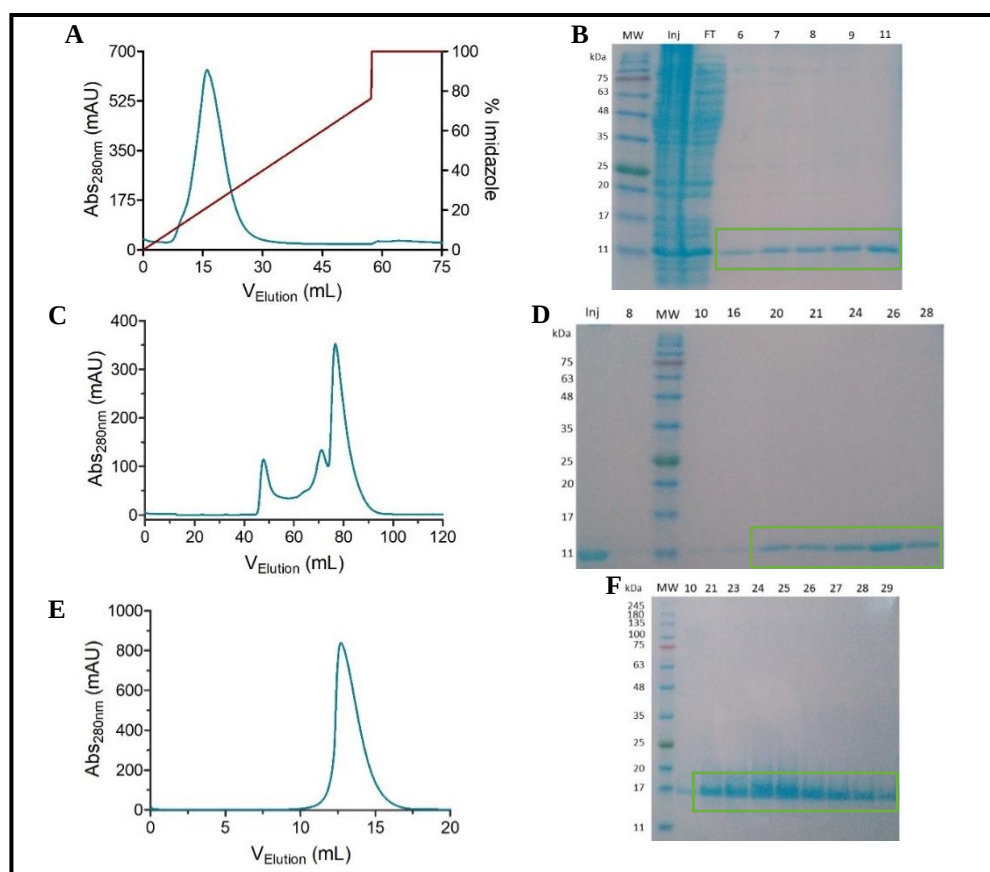


Figure 4.9 | Chromatograms and respective SDS-PAGE gels of RICH 3 purification steps. (A) Elution profile of the affinity chromatography with HiTrap Fast Flow Crude column. (B) SDS-PAGE gel of the fractions collected during the affinity chromatography step. (C) Chromatogram obtained after the size-exclusion chromatography with Superdex 75 prep grade column. (D) SDS-PAGE gel where the fractions collected during the size-exclusion chromatography were examined. (E) Chromatogram of the size-exclusion chromatography using a Superdex 10/300 GL column. (F) Evaluation of the fractions collected during the last purification step by SDS-PAGE gel. Blue line – absorbance at 280 nm. Red line – elution gradient using buffer containing 500 mM imidazole. MW - NZYColour Protein Marker II. Inj – sample injected in each purification step. FT- flowthrough collected during the washing step. Column numbers correspond to the fractions collected during the chromatographies.

RICH 4 purification was done in a two-step protocol starting with an affinity chromatography where the elution occurred with 238.5 mM imidazole, as shown in Figure 4.10A. In the size-exclusion chromatography (Figure 4.10B and C), the RICH 4 elution volume (10.91 mL) correspond to a molecular weight of 36.8 kDa. This process had an overall yield of 0.22 mg of RICH 4 per culture medium liter.

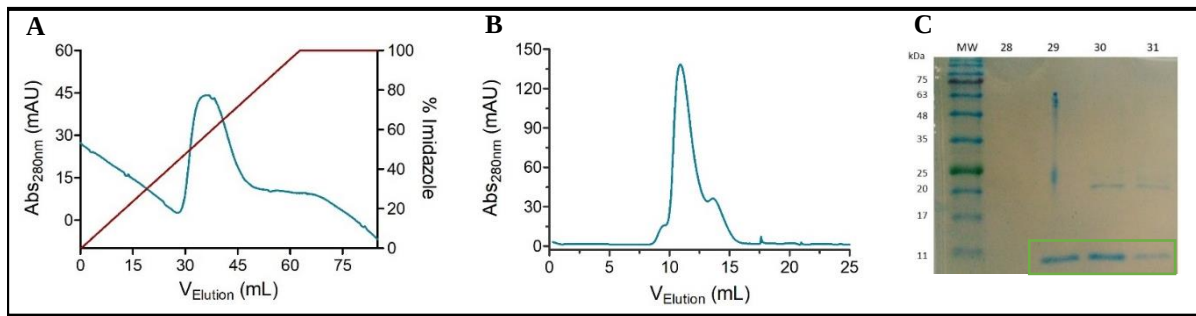


Figure 4.10 | Chromatograms of the RICH 4 purification and the final SDS-PAGE gel. (A) Elution profile of the affinity chromatography with HiTrap TALON® Crude column. (B) Chromatogram acquired after the size-exclusion chromatography using a Superdex 10/300 GL column. (C) Evaluation of the fractions collected during the last purification step by SDS-PAGE. Blue line – absorbance at 280 nm. Red line – elution gradient using buffer containing 500 mM imidazole. MW - NZYColour Protein Marker II. Numbers – number corresponding to the fractions collect during the chromatography.

RICH 5 was eluted with 95 mM imidazole from the affinity chromatography column, as shown in the chromatogram and respective SDS-PAGE gel in Figure 4.11A and B. In the size-exclusion chromatography using Superdex 75 prep grade column, RICH 5 was eluted with 59.0 mL and 62.4 mL (Figure 4.11C and D). Next, the RICH 5 elution volume (11.10 mL) from the final size-exclusion chromatography corresponded to a molecular weight of 34.2 kDa (Figure 4.11E and F). The yield of this process was 16.12 mg of RICH 5 per culture medium liter.

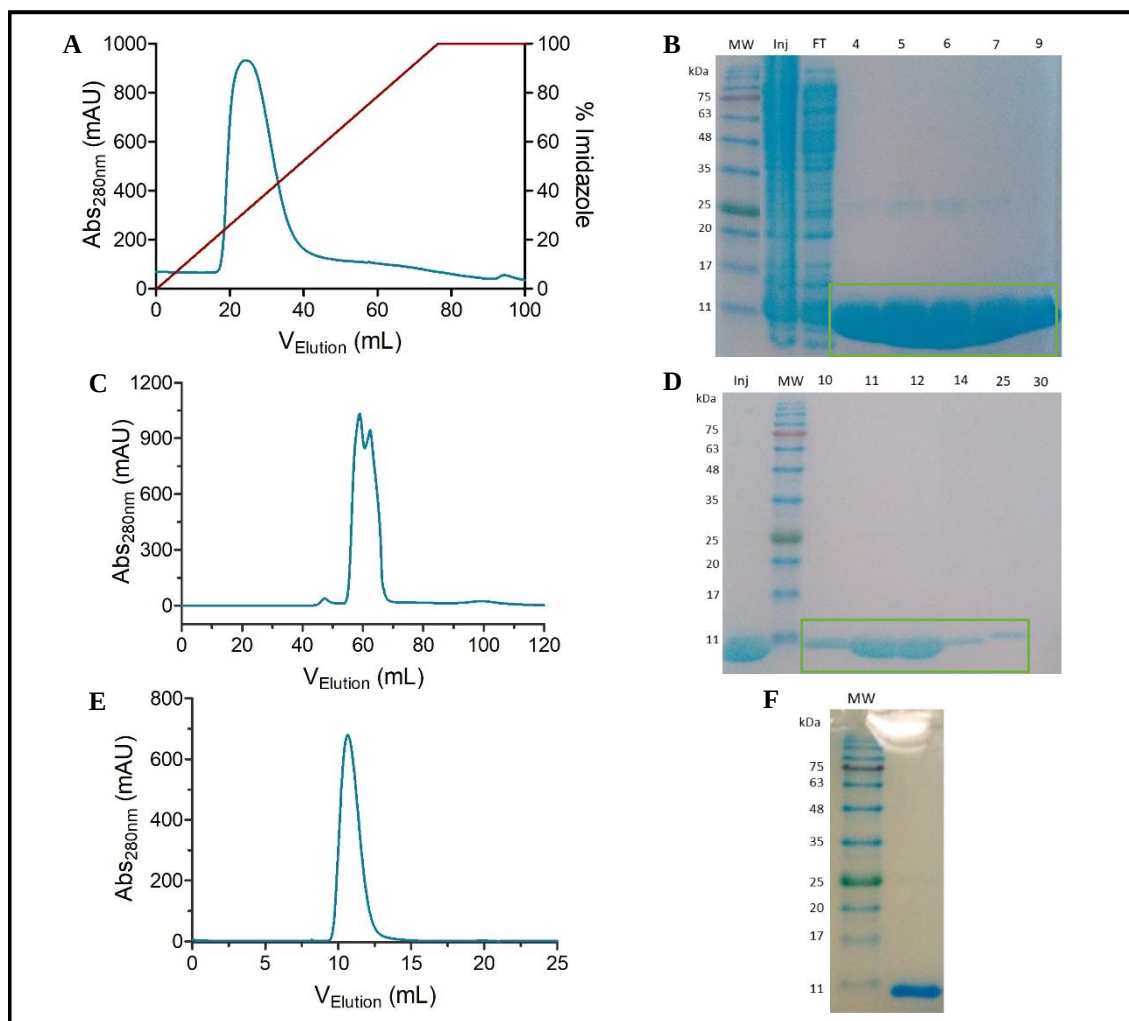


Figure 4.11 | Chromatograms and respective SDS-PAGE gels of RICH 5 purification steps. (A) Elution profile of the affinity chromatography with HiTrap Fast Flow Crude column. (B) SDS-PAGE gel of the fractions collected during the affinity chromatography step. (C) Chromatogram obtained after the size-exclusion chromatography with Superdex 75 prep grade column. (D) SDS-PAGE gel where the fractions collected during the size-exclusion chromatography were examined. (E) Chromatogram of the size-exclusion chromatography using a Superdex 10/300 GL column. (F) Evaluation of the fractions collected during the last purification step by SDS-PAGE gel. Blue line – absorbance at 280 nm. Red line – elution gradient using buffer containing 500 mM imidazole. MW - NZYColour Protein Marker II. Inj – sample injected in each purification step. FT- flowthrough collected during the washing step. Numbers – number corresponding to the fractions collect during the chromatographies.

Scaffold 1 purification was achieved through a four-step protocol starting with an affinity chromatography from which the elution was done with 135 mM imidazole (Figure 4.12A and B). In the size-exclusion chromatography (Figure 12C and D), the protein was eluted with 72.8 mL. The third step consisted in an ionic-exchange chromatography where scaffold 1 was eluted with 308 mM NaCl, as presented in Figure 4.12E and F. At the end, the protein was eluted with 50 mM Tris-HCl pH 7.5 from a hydrophobic chromatography column (Figure 4.12H and I). Scaffold 1 expression and purification protocol had a yield of 0.06 mg protein per culture medium liter.

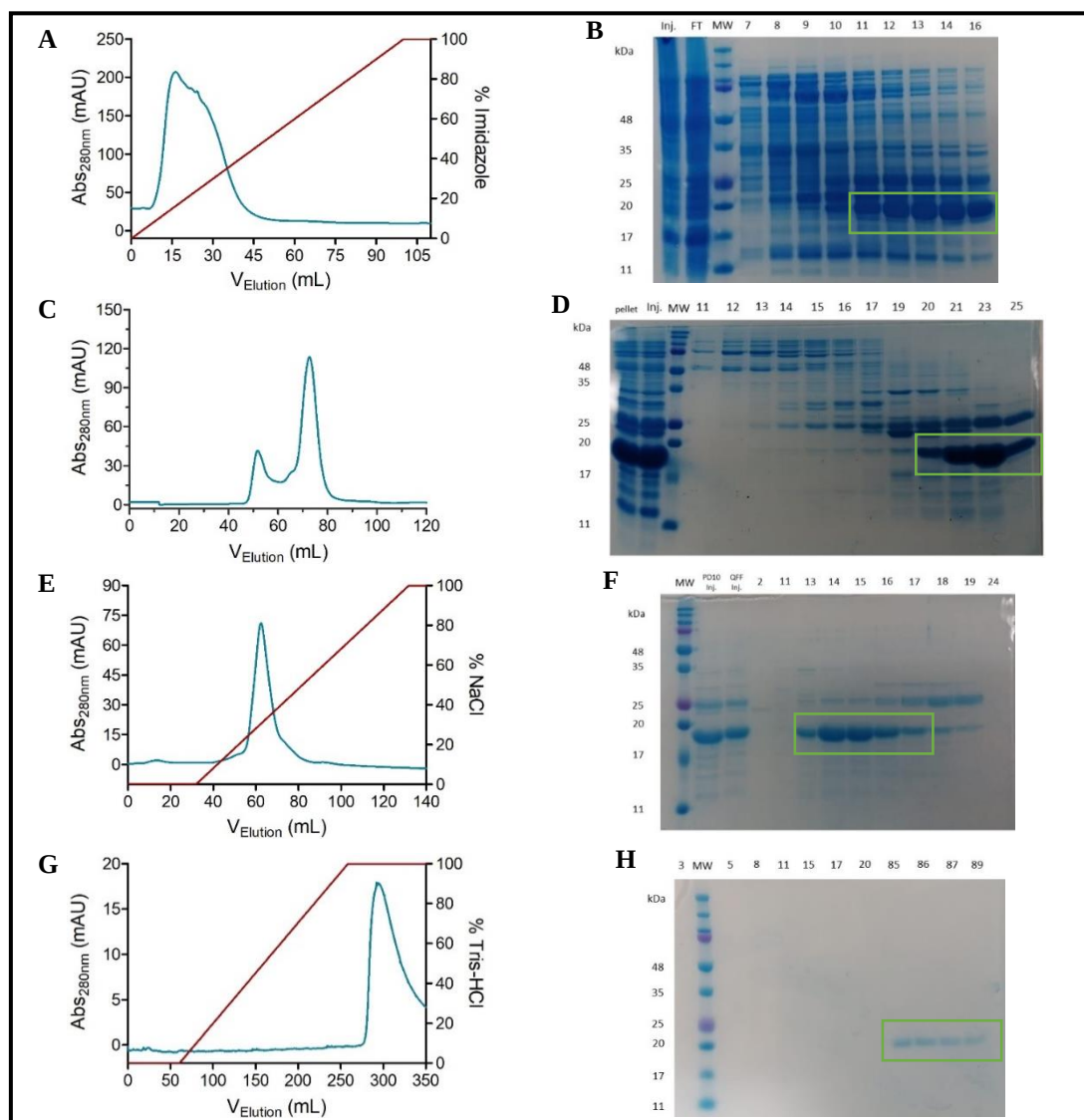


Figure 4.12 | Chromatograms and corresponding SDS-PAGE gels of scaffold 1 purification. (A) Elution profile of the affinity chromatography with HiTrap Fast Flow Crude column. (B) SDS-PAGE gel of the fractions collected. (C) Chromatogram obtained after the size-exclusion chromatography with Superdex 75 prep grade column. (D) SDS-PAGE gel (15%) where the fractions collected during the size-exclusion chromatography were examined. (E) Chromatogram of the ionic-exchange chromatography using a HiTrap Q Fast Flow column. (F) Evaluation of the fractions collected during the third purification step by SDS-PAGE gel. (G) Elution profile of the hydrophobic chromatography with Octyl Fast Flow Crude column. (H) SDS-PAGE gel of the fractions collected the last purification step. Blue line – absorbance at 280 nm. Red line – elution gradient using buffer containing either 500 mM imidazole, 1 M NaCl or 50 mM Tris-HCl pH 7.5.. MW - NZYColour Protein Marker II. Inj – sample injected in each purification step. FT- flowthrough collected during the washing step. Numbers – number corresponding to the fractions collect during the chromatographies.

Scaffold 5 used a two-step purification protocol which started with an affinity chromatography where the elution occurred with 54 mM imidazole, as shown in Figure 4.13A and B. In the size-exclusion chromatography (Figure 13C and D), the protein was eluted with 81.4 mL. This process had an overall yield of 1.02 mg of scaffold 5 per culture medium liter.

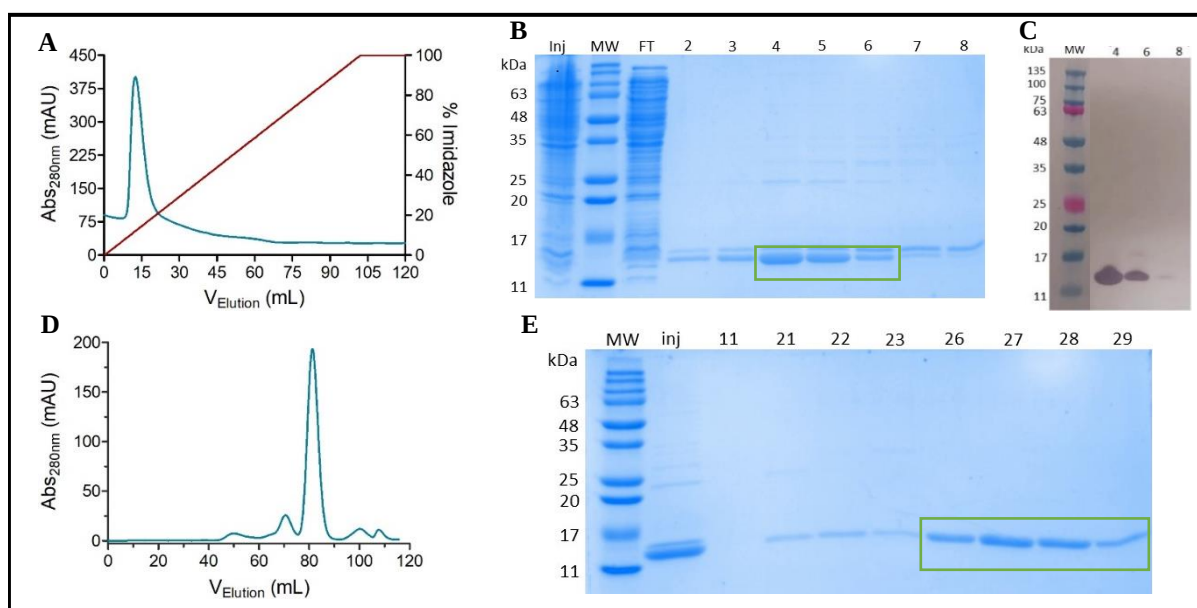


figure 4.13 | Chromatograms and respective SDS-PAGE gels and WB of 5XEF purification steps. (A) Elution profile of the affinity chromatography with HiTrap TALON Crude column. (B) SDS-PAGE gel of the fractions collected during the affinity chromatography step. (C) WB of some of the affinity chromatography fractions. (D) Chromatogram obtained after the size-exclusion chromatography with Superdex 75 prep grade column. (E) SDS-PAGE gel where the fractions collected during the size-exclusion chromatography were examined. Blue line – absorbance at 280 nm. Red line – elution gradient using buffer containing 0.00 mM imidazole. MW - NZYColour Protein Marker II. Inj – sample injected in each purification step. FT- flowthrough collected during the washing step. Numbers – number corresponding to the fractions collect during the chromatographies.

To sum up, the target RICH proteins and respective controls (a total of seven proteins) were successfully expressed and purified. However, each protocol involved a different number of steps and ended up having distinct yields. Additionally, each protein was purified in a specific oligomeric state. These informations are presented in Table 4.1.

Table 4.1 | Overall characteristics of each protein purification procedure, including the oligomeric state.

Protein	Purification Steps (nr.)	Yield (mg.L ⁻¹ culture medium)	Oligomeric state
RICH 1	2	0.5	Monomer
RICH 2	2	0.26	Monomer
RICH 3	3	0.84	Monomer / Dimer
RICH 4	2	0.22	Dimer
RICH 5	3	16.12	Dimer
Scaffold 1	4	0.06	Monomer
Scaffold 5	2	1.02	Monomer

4.1.3. Structural characterization

Secondary structure characterization

One of the main concerns in recombinant protein expression is the folding of the target protein. Since protein structure is closely related to function it is of utmost importance to ensure the right folding before proceeding with further experiments. Therefore, RICHs CD spectra presented in solid lines in Figure 4.14 were collected and compared with literature ones known to be typical of some secondary structure elements. The CD spectra reflect the expected folding

for each protein with RICH 1,2,4 and 5 being dominated by α -helical content suggested through the presence of negative bands near 208 and 222 nm, while RICH 3 spectrum exhibits β -strands features as indicated by the broad negative band center at around 210-220 nm.

Moreover, the CD spectrum of each RICH was analyzed with two different web server-based tools, CAPITO and BeStSel, to estimate the secondary structure elements content. The employment of more than one prediction algorithm aims to overcome possible limitations or biases enabling to compare both evaluations. Although these two computational approaches use different data sets of reference proteins it is possible to judge the given prediction by the normalized root mean square deviation (NRMSD) or “goodness of fit” parameter defined by $[\Sigma(\theta_{exp} - \theta_{cal})^2 / \Sigma(\theta_{exp})^2]^{1/2}$ which ranges from 0 to 1 corresponding to perfect fit or no fit at all, respectively.

In addition to the NRMSD above 0.36, a visual inspection of the fitted data revealed that CAPITO was not suitable for the secondary structure content prediction of RICHs. The experimental data versus the fit by CAPITO is presented in Figure A7 in the annex section. Therefore, the experimental results were fitted with BeStSel as shown in Figure 4.14 in dashed lines.

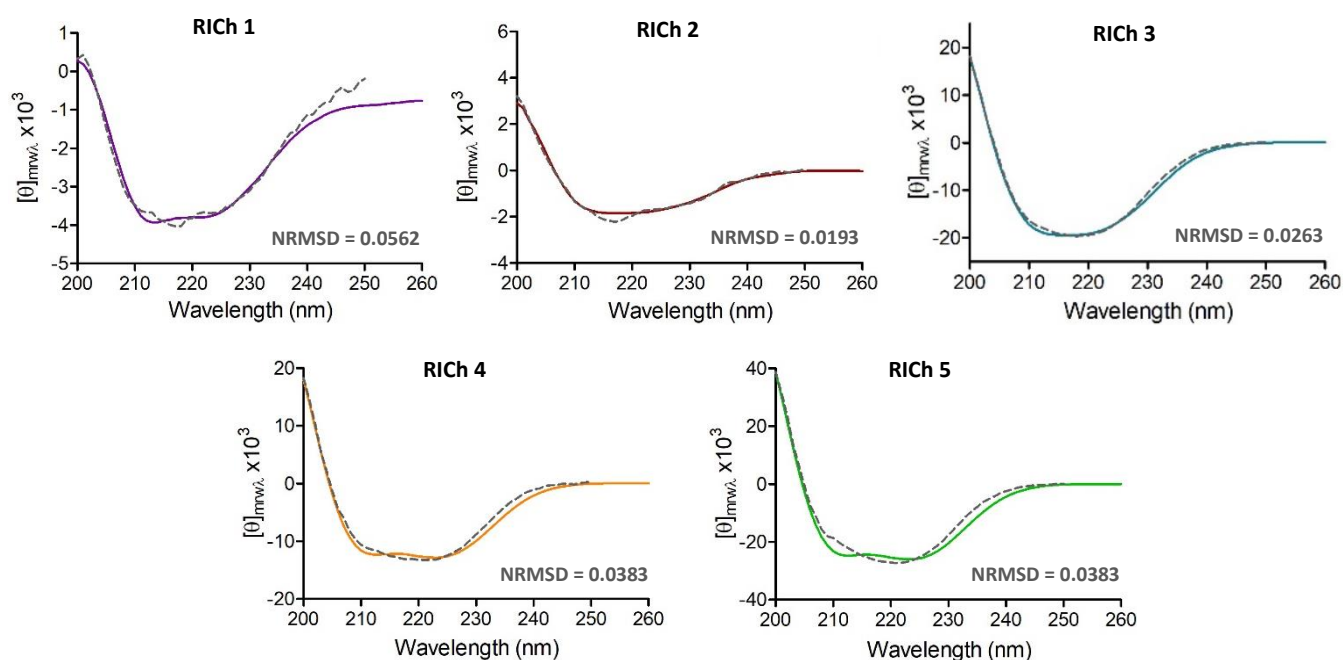


Figure 4.14 | CD spectra of RICHs. (Solid line) CD spectrum acquired for each RICH represented as the mean residue ellipticity in $\text{deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}$ from 200 to 260 nm. (Dashed line) Fit of the experimental data from 200 to 250 nm performed using the BeStSel web server-based tool. The NRMSD of each fit is presented.

BeStSel defines four main secondary structure elements derived from the Dictionary of Secondary Structure of Proteins (DSSP): α -helix, β -strand, turns and others. The “others” category includes 3-10 helix, bends, π -helix, β -bridge, loops and irregular or disordered structures. Despite presenting better NRMSD than CAPITO, BeStSel still does not fit the experimental data perfectly and tends to accentuate the negative band in the 210-220 nm

region. This trait may influence the secondary structure content prediction shown in Figure 4.15 contributing to a higher β -strand percentage.

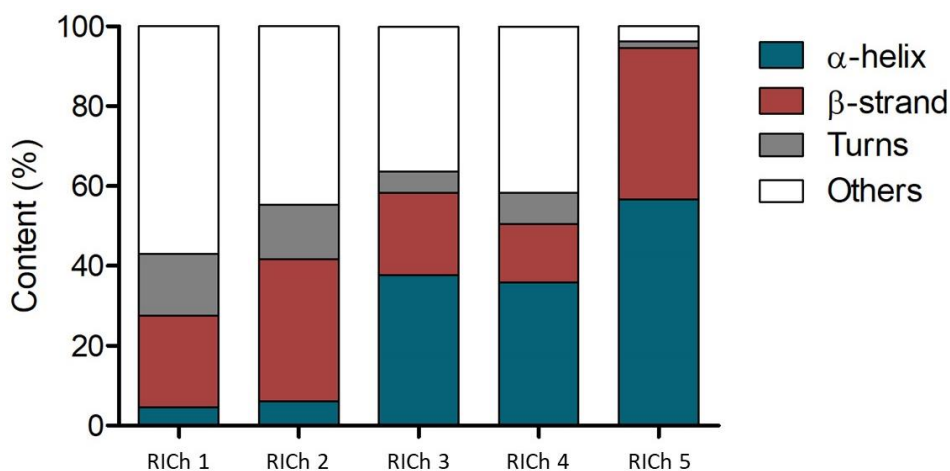


Figure 4.15 | Content of each secondary structure element predicted by BeStSel web server-based tool for each RICH establish from the CD spectra collected.

The CD spectra and the secondary structure prediction shown above indicate that the recombinant RICHs proteins were expressed with the adequate folding to proceed to further assays. However, if the aim was to study detailed structural features, spectra below 200 nm would be required as well as optimization of the acquisition parameters.

Since RICHs binding to the RBD may be affected by the three-dimensional arrangement of the interacting-amino acid residues, it is important not only to determine their secondary structure but also to evaluate their thermal resistance. To address this, spectra were collected at 20 °C, after heating to 90 °C and after recooling the protein sample (Figure 4.16). At 90 °C, RICHs 1 and 2 exhibit a CD spectrum with a negative band center around 200 nm associated with a disorder structure meaning both proteins are partially or completely unfolded. After recooling, RICH 1 only recovers a small portion of its folding while RICH 2 regains nearly its initial secondary structure. RICH 3 is the exception because it does not reveal significant evidence of denaturation at 90 °C indicating that the designed protein maintained the high thermostability of its scaffold protein. RICH 4 and 5 share a similar behavior, losing irreversibly the native secondary structure elements after thermal denaturation.

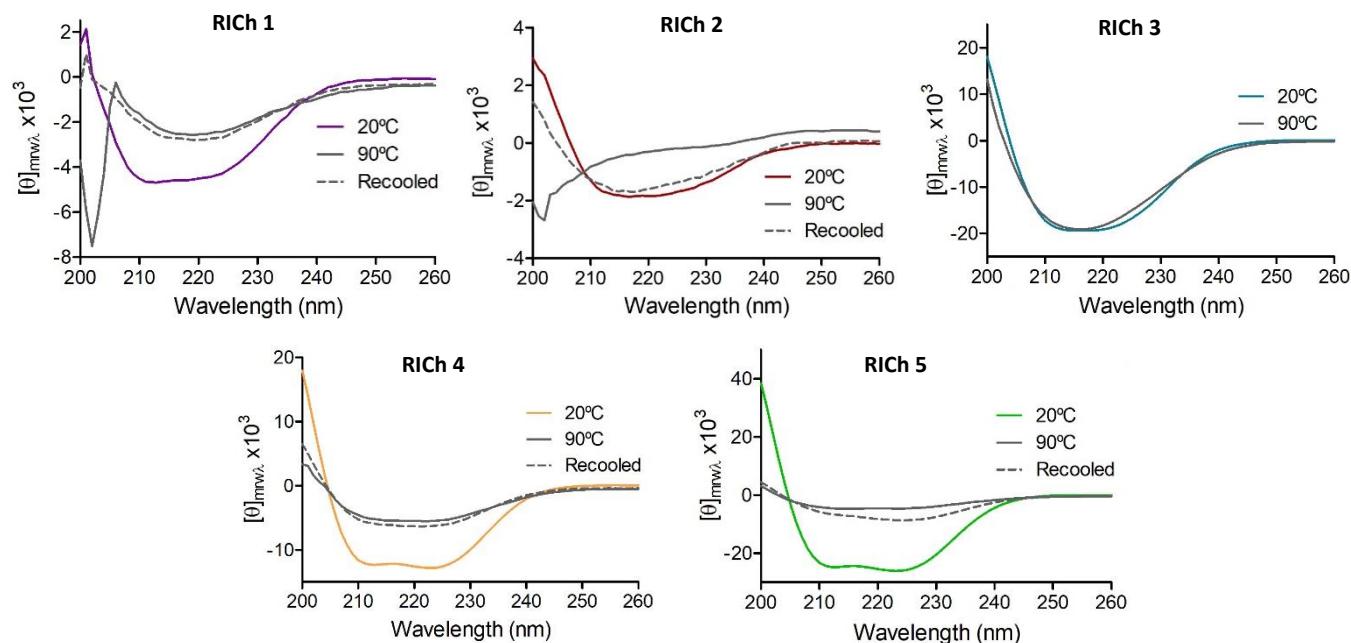


Figure 4.16 | CD spectra collected at 20 °C (solid colored line), 90 °C (solid grey line) and after recooling to 20 °C (dashed line) of each RICh. The mean residue ellipticity in $\text{deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}$ is represented as a function of the wavelength.

4.1.4 Functional characterization

RIChs affinity for the SARS-CoV-2 RBD

To assess the potential of RIChs to prevent the RBD-ACE2 interaction, the relative affinity for the RBD was evaluated by SPR. The SPR experiments allowed us to determine the relative specificity by comparing the RIChs sensorgrams with the controls, as well as characterize the binding kinetics, determine the binding constants of association (k_a), dissociation (k_d), and the equilibrium dissociation constants (K_D).

The sensorgrams shown in Figure 4.17, revealed that RIChs 2 and 4 were not able to bind to the RBD resulting in a sensorgram similar to the negative control, bovine serum albumin (BSA). On the other hand, RIChs 1,3 and 5 were shown to interact with the RBD with different affinities and kinetics. Comparing the slopes of the sensorgram in the association and dissociation steps, RICh 1 and 5 present a slow binding kinetics, while RICh 3 has a fast complex formation as well as a quick dissociation rate.

The binding constants in Table 4.2 were estimated with the Biacore Insight Evaluation Software. RICh 1 and 5 have the best K_D values of 954 ± 203 nM and 456 ± 238 nM, respectively while RICh 3 presents a K_D of 3.8 ± 0.8 μM . Additionally, RICh 3 and 5 reveal a heterogeneous binding kinetics. Although RICh 1 and 3 bound to the RBD with K_D values in the high nM range, hACE2, used as the positive control, presented a lower K_D of 139 nM.

Moreover, to verify if the RIChs-RBD interactions were being accomplished through the designed region, scaffolds 1 and 5 affinity for the RBD was also evaluated. The resulting sensorgrams are presented in Figure A8 in the annex section and show no significant scaffold-RBD binding.

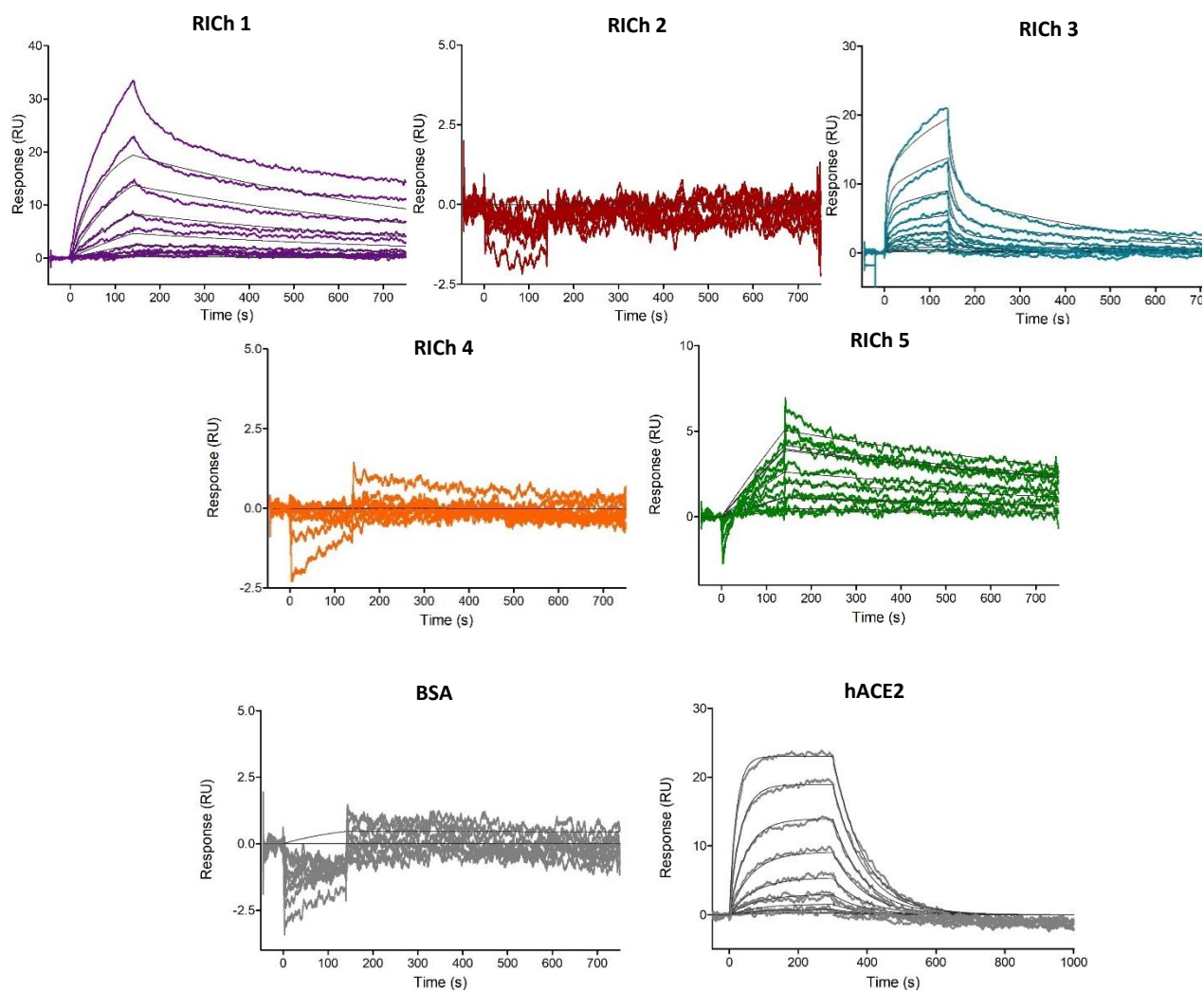


Figure 4.17 | Sensorgrams for the binding of RICHs, BSA (negative control) and hACE2 (positive control) to the SARS-CoV-2 RBD obtained by surface plasmon resonance (SPR).

Table 4.2 | Binding constants determined by the Biacore Insight Evaluation Software for the proteins which bound to the SARS-CoV-2 RBD.

Protein	k_a ($M s^{-1}$)	k_d (s^{-1})	K_D (μM)
RICH 1	1.23E+03	1.17E-03	0.954 ± 0.203
Rich 3	7.88E+02	3.06E-03	$3.88 \pm 0.78^*$
Rich 5	8.58E+02	3.93E-04	$0.458 \pm 0.238^*$
hACE2	7.92E+04	1.07E-02	0.135 ± 0.006

* Heterogenous binding behavior

RICH 1 and RBD structural stability changes induced by complex formation

RICH 1 was selected as the best candidate due to its estimated affinity determined from the SPR experiments. To further characterize the RICH 1-RBD complex, changes in the structural stability of RICH 1 and RBD induced by the complex formation were monitored using dye-free

DSF. In Figure 4.18 the intrinsic tryptophan fluorescence at 330 nm, 350 nm and for the 350/330 ratio is exhibited as a function of the temperature for the complex RICH1-RBD and for each individual protein.

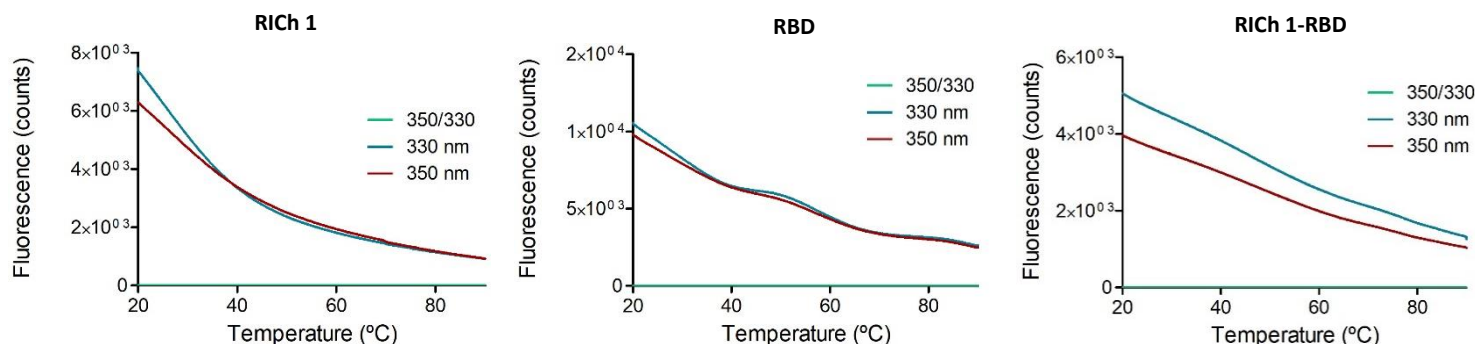


Figure 4.18 | Thermal stability of RICH 1, RBD and the RICH1-RBD complex monitored using dye-free DSF following the intrinsic fluorescence changes at 330 nm (blue), 350 nm (red) and for the 350/330 ratio (green).

The fluorescence intensity at 330 and 350 nm decreased with the increase in temperature for all the three samples tested. Since there is a decrease in both wavelengths, the signal of the 350/330 ratio leads to a flat curve and presents no meaningful information for further analysis. In these cases, a combination of red- and blue-shift upon protein unfolding may be occurring and as a result both effects might cancel each other.

The most common effect of protein unfolding is a red-shift of the spectra resulting in a decrease of the intensity detected at 330 nm in a dye-free DSF assay. The red-shifts are generally related to tryptophan residues which are located in hydrophobic cores in the native folding and become exposed to the solvent environment upon thermal unfolding. Because of the change in tryptophan environment, a shift of the maximum fluorescence intensity towards higher wavelengths occurs.

On the other hand, the blue-shift upon protein denaturation induces a shift of the emission peak towards lower wavelengths and thus a decrease in the fluorescence intensity at 350 nm. Two main phenomena are associated with this shift: either the surface-exposed tryptophans become surrounded by other hydrophobic residues after protein denaturation or a change in the inner tyrosine residues happens.

NanoTemper PR.ThermControl software accurately estimates the temperature at which the proteins unfold (T_m) based on the collected data. Additionally, T_m may also be determined in accordance with the inflection points of the first derivative of the 330 nm fluorescence as a function of the temperature. Since none of the samples presented a classical sigmoidal-shape curve, T_m values could not be determined but rather temperature points at which the environment of tryptophans and/or tyrosines residues suffer an alternation conducting to a change in the fluorescence intensity recorded (Figure 4.19).

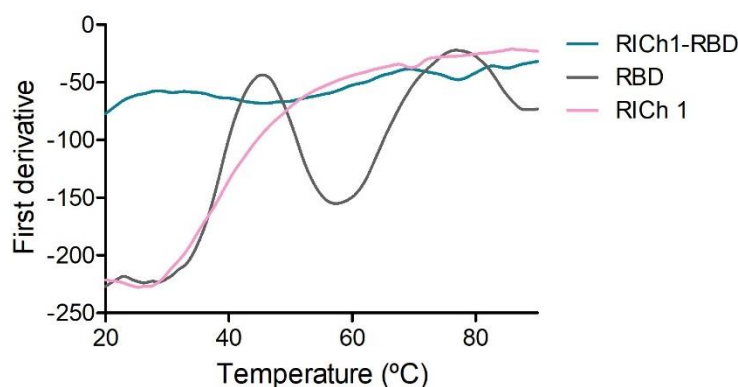


Figure 4.19 | First derivative of the fluorescence data collected at 330 nm during the thermal denaturation assay. (Blue) Complex RICH1-RBD. (Grey) RBD. (Pink) RICH 1.

For RBD alone four inflection points were detected at 32.2, 45.4, 59.8 and 74.6 °C while RICH 1 showed two at 33.3 and 70.1 °C. The analysis of the complex-containing samples pointed out two inflection points at 43.4 and 60.8 °C.

The difference from four to two inflection points when comparing the RBD by itself and the complexed version, suggests less events of changes in the environment of tryptophans and/or tyrosines residues of the RBD when complex to RICH 1. This might be the result of an increase in the RBD structural stability due to the binding of RICH 1. As for RICH 1, the number of inflection points is maintained. However, the temperature at which they occur differs for the RICH 1 alone and complexed, indicating that the binding with the RBD generated a different environment around tryptophans and/or tyrosines residues. Since the RICH 1 and RBD binding is supposed to be mainly achieved through hydrophobic interactions, it is possible that the main alterations in the thermal denaturation profiles are due to the changes in the environment of the hydrophobic residues forming the complex interface.

Neutralization ability

The RICHs which showed evidence of binding to the SARS-CoV-2 RBD on the SPR experiments were selected for neutralization assays. The main goal here was to evaluate the ability of RICH 1, 3 and 5 to prevent viral infection by blocking the interaction between the RBD and ACE2 receptor on the 293T-ACE2 cells.

RICHs were incubated with spike-pseudotyped lentiviral particles at different concentrations ranging from 2.1 to 500 $\mu\text{g}\cdot\text{mL}^{-1}$ and then the mixtures were added to the plated cells. The relative infection calculated is shown in Figure 4.20. The effect of RICH 1, 3 and 5 on the infection process was too subtle to be considered relevant or efficient, possibly due to their lower affinity for the RBD than ACE2.

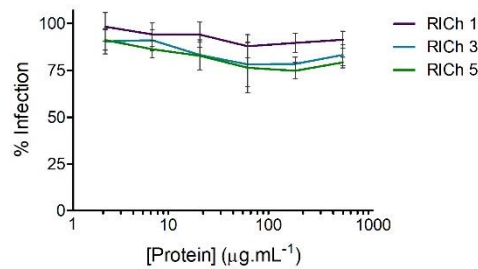


Figure 4.20 | Effect of RICH 1 (purple), RICH 3 (blue) and RICH 5 (green) in the infection of 293T-ACE2 cells by spike-pseudotyped lentiviruses.

Although none of the RICHs could prevent infection to occur, it was possible to validate the design strategy adopted since all the proteins were successfully expressed in *E.coli*, purified with a suitable folding and three out of five shown evidence of binding to the target. For future work some considerations may be applied, for example: (I) taking into consideration the absence of glycans and other pos-translational modifications in proteins expressed in *E.coli* systems that may be contributing to the RBD-ACE2 interaction; (II) a high-throughput method to evaluate the candidates' affinity to the target would be beneficial to rapidly discard the non-binding proteins.

Currently, two main strategies are being applied in the lab: 1) improvement of the binding affinity of the best RICHs candidates; 2) developing a new generation of protein inhibitors of the RBD-ACE2 interaction by applying *de novo* design to construct a protein around the binding motif, discarding the need of protein scaffolds.

4.2 Immunogenic construct

4.2.1. *In silico* design

The immunogenic construct was designed according to the motif-grafting method as described in the methods section. The selected epitope is a fragment of the SARS-CoV-2 S protein, comprising the S2' cleavage site and a portion of the FP, known to be immunogenic as well as conserved among coronaviruses. Therefore, antibodies targeting this region could be able to block either the S2 subunit activation or the interaction of the FP with the host membrane.

The immunogenic construct that resulted from the design process was named TSSFP (Figure 4.21). TSSFP and TSS are identical in sequence and structure except for the α -helix where the immunogenic fragment was grafted. Part of the fragment was predicted to be helical while the other resembles an unordered structure and is, thus, more flexible. The main aims in the design were to stabilize the epitope but maintain it exposed and accessible to the immune system.

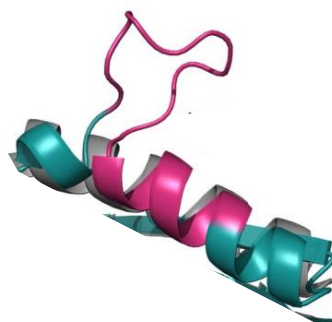


Figure 4.21 | Designed α -helix of the TSSFP or immunogenic construct structure overlayed with its native scaffold. (Blue) TSSFP. (Pink) Immunogenic fragment selected. (Grey) Original TSS scaffold.

4.2.2. Expression and purification

TSSFP is a novel engineered protein so its recombinant expression in *E.coli* was evaluated through expression tests. Aliquots of the cellular extract collected during the expression tests were examined by western blot (WB) to evaluate the TSSFP expression levels and solubility in the different conditions. The WB present in Figure 4.22 revealed that TSSFP is expressed in all conditions but mainly in the insoluble fractions or pellet.

The expression condition to scale-up was selected according to the amount of TSSFP present in the soluble fraction. Therefore, the expression condition using 2xYT medium, 0.1mM IPTG to induce protein expression and incubation at 18 °C for 18h after induction was the selected one.

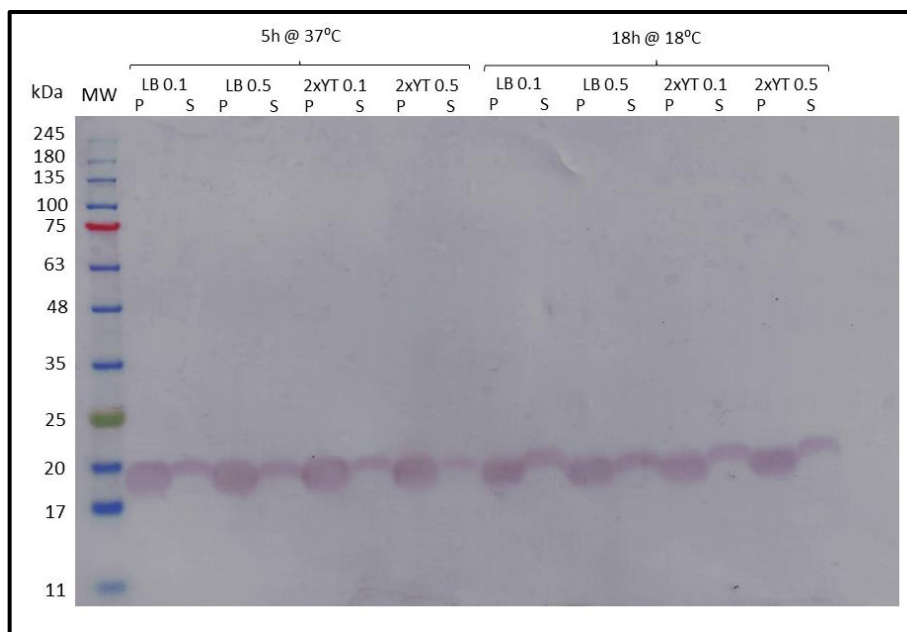


Figure 4.22 | TSSFP expression profile evaluated by WB. MW: NZYColour Protein Marker II. P and S: pellet and supernatant collected after cell lysis. Each lane corresponds to a distinct expression condition, accordingly to the labels: LB and 2xYT – culture media; 0.1 and 0.5 – IPTG concentration (in mM) used to induce protein expression; 5h @ 37 °C and 18h @ 18 °C – incubation conditions after induction.

TSSFP purification occurred in a three-step process which started by an affinity chromatography (HiTrap TALON crude) where the protein was eluted with 89 mM Imidazole, as shown by the chromatogram and SDS-PAGE gel represented in Figure 4.23A and B. The fractions corresponding to the last chromatogram band were concentrated and further purified with a size-exclusion chromatography (Figure 4.23C and D). The TSSFP elution volume (18.54 mL) corresponds to the monomeric form of TSFP. The last purification step consisted of an ionic-exchange chromatography and allowed us to obtain pure fractions of TSFP eluted with 204 mM NaCl, as observed in Figure 4.23E and 4.23F. The overall yield of TSSFP was 0.63 mg per liter of culture medium.

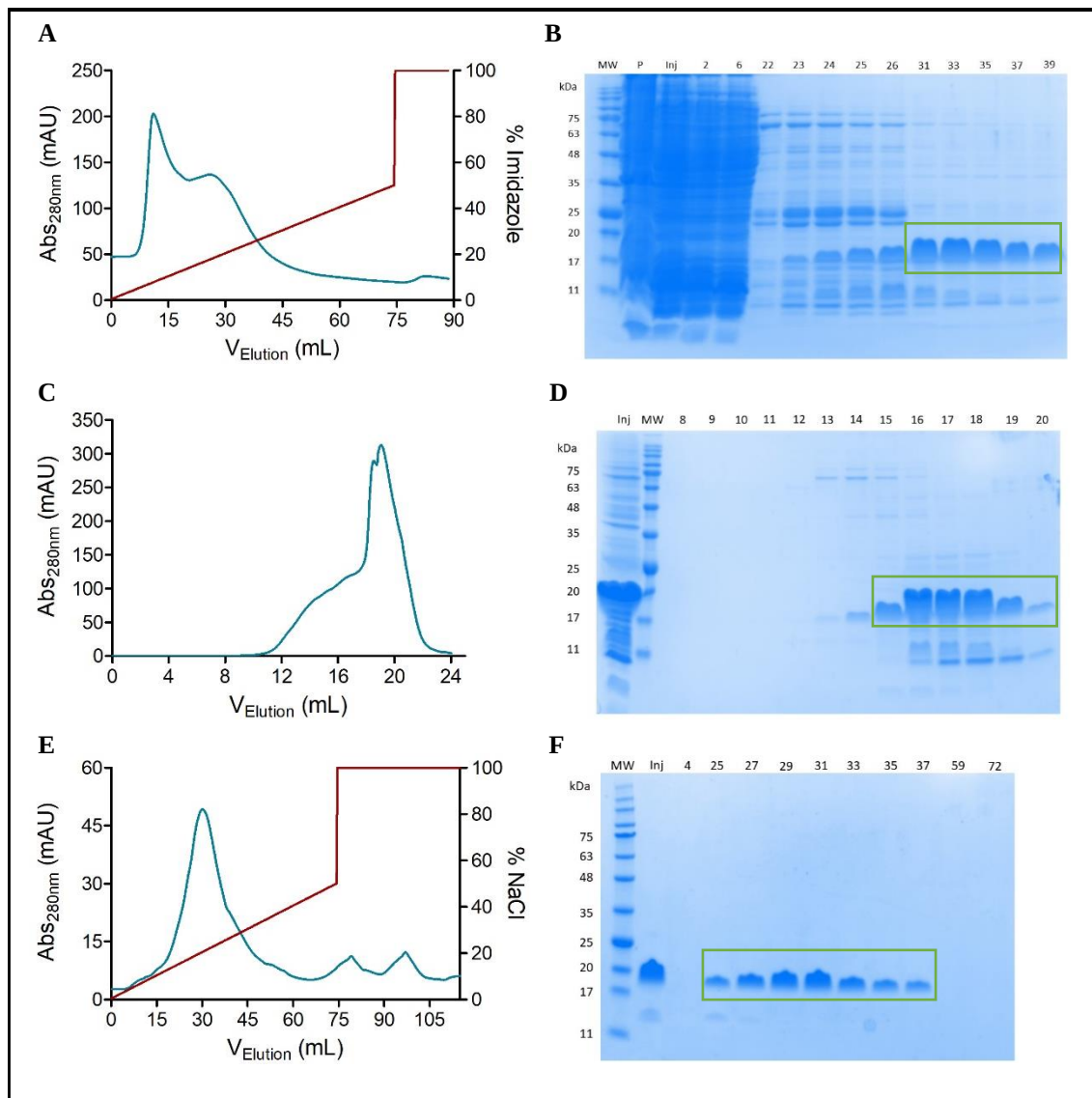


Figure 4.23 | Chromatograms and respective SDS-PAGE gels of TSSFP purification steps. (A) Elution profile of the affinity chromatography with HiTrap TALON crude column. (B) SDS-PAGE gel of the fractions collected during the affinity chromatography step. (C) Chromatogram obtained after the size-exclusion chromatography with Superose 6 column. (D) SDS-PAGE gel where the fractions collected during the size-exclusion chromatography were analyzed. (E) Representative chromatogram of the ion-exchange chromatography with HiTrap QFF column. (F) SDS-PAGE gel of the fractions collected during the ion-change chromatography. Blue line – absorbance at 280 nm. Red line – elution gradient using either the elution buffer containing 500 mM imidazole or 1 M NaCl. MW - NZYColour Protein Marker II. P – pellet collected after cell lysis and centrifugation. Inj – sample injected in each purification step. Numbers – number corresponding to the fractions collect during the chromatographies.

4.2.3. Structural characterization

Secondary structure characterization

The TSSFP secondary structure was analyzed to confirm if the suitable folding was achieved. Thus, the TSSFP CD spectrum (solid line in Figure 4.24) was collected and compared with typical literature spectra of secondary structure elements. The observed negative band near 210-220 nm in the CD spectrum appears to contain a mixture of α -helix and β -strand features.

Additionally, TSSFP CD spectrum was examined with BeStSel and CAPITO to predict the secondary structure elements content. Both predictions were evaluated according to the NRMSD and BeStSel provided the best fitting (dashed line in Figure 4.24) with a NRMSD of 0.0065 while CAPITO presented a NRMSD of 0.65 (Figure A9 in the annex section). The CD spectra and the secondary structure estimated by BeStSel (Figure 4.24) revealed that TSSFP adopted the proper folding to proceed to further assays.

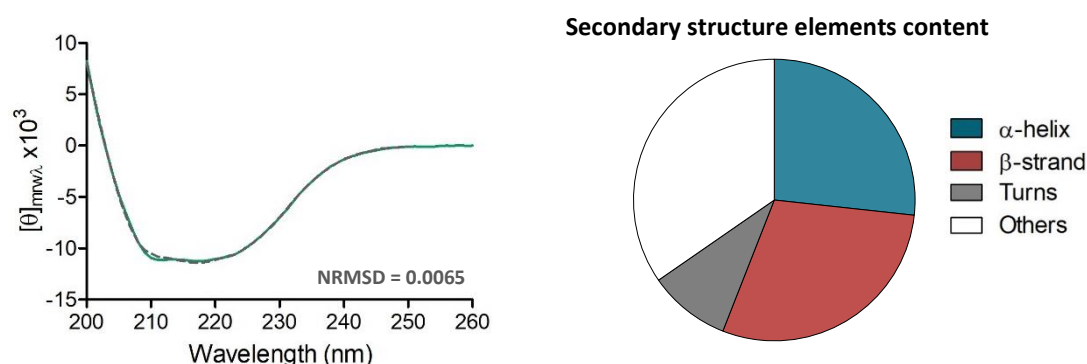


Figure 4.24 | Secondary structure evaluation of TSSFP. (Left side) CD spectra of TSSFP. (Solid line) CD spectrum acquired for TSSFP represented as the mean residue ellipticity in $\text{deg.cm}^2.\text{dmol}^{-1}$ from 200 to 260 nm. (Dashed line) Fit of the experimental data from 200 to 250 nm performed using the BeStSel web server-based tool. The NRMSD of the fit is presented. (Right side) Content of each secondary structure element predicted by BeStSel web server-based tool establish from the CD spectra collected.

Thermal stability

One of the TSS scaffold advantageous characteristic is its high thermal resistance. Thus, to verify if this feature was preserved in the TSSFP construct, a thermal denaturation was performed. The CD signal at 222 nm was monitored as the temperature increased, resulting in the curve presented in Figure 4.25. The 222 nm wavelength was selected because it is one of the wavelengths where the CD signal in Figure 4.23 exhibited the highest value in magnitude and it is related to the presence of ordered secondary structure elements, essentially α -helices.

Although the denaturation curve did not reach an exponential phase or an upper plateau, it was possible to detect the beginning of the unfolding process near 65 °C. To further assess the impact of the high temperature on TSSFP, a CD spectrum was collected at the end of the temperature gradient (90 °C) and plotted against the spectrum acquired at the beginning of the assays (20 °C) (Figure 4.25). The difference between both spectra is less than 4 units, meaning the TSSFP maintained the majority of the initial secondary structure elements.

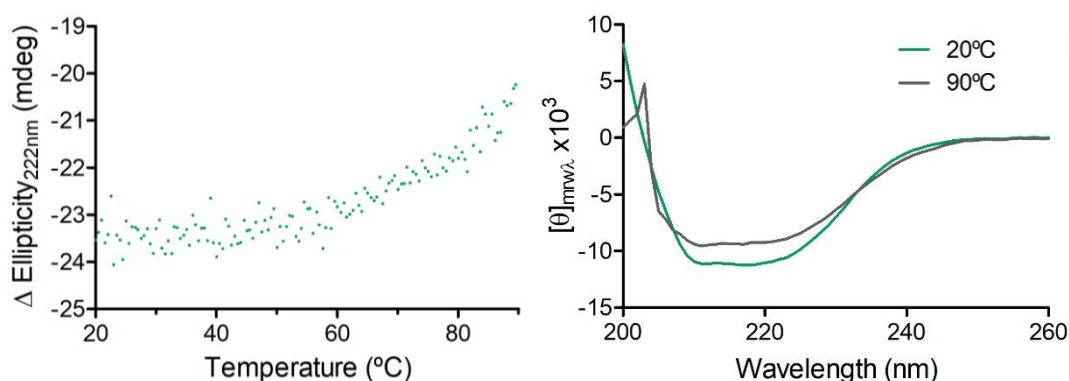


Figure 4.25 | TSSFP thermostability evaluation by CD. (Left side) Variation of the ellipticity at 222 nm as a function of the temperature. (Right side) CD spectra collected at 20°C (green) and 90°C (grey). The mean residue ellipticity in $\text{deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}$ is represented as a function of the wavelength.

Storage stability

TSSFP was designed as a potential biopharmaceutical, hence the importance of studying its thermostability but also its storage stability. To clarify this matter, TSSFP samples were stored at 4 °C and 24 °C (room temperature, RT) for 29 days. CD spectra of the two samples were collected during the storage time to examine the changes in the secondary structure (Figure 4.26).

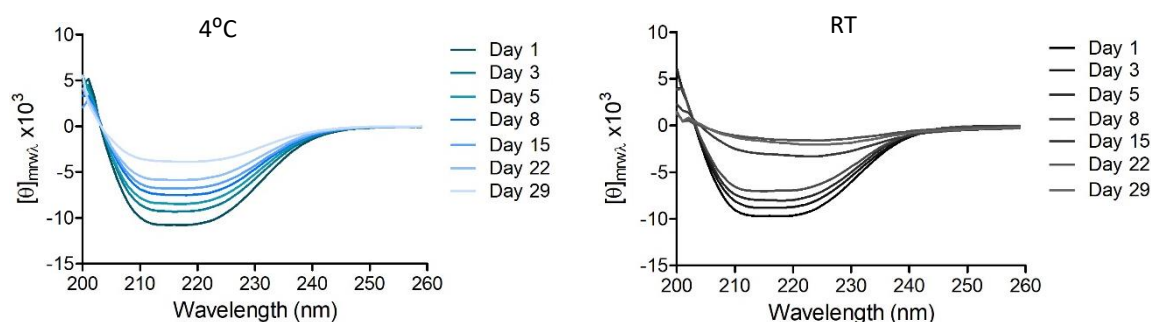


Figure 4.26 | TSSFP storage stability in PBS pH 7.4, 300 mM NaCl, 10 % glycerol evaluated by CD. Spectra were collected at day 1, 3, 5, 8, 15, 22 and 29 of storage. The mean residue ellipticity in $\text{deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}$ is represented as a function of the wavelength. (Right side) CD spectra of the sample storage at 4 °C. (Left side) CD spectra of the sample storage at room temperature.

In both conditions, the protein lost its native secondary structure during the storage time, as suggested by the decrease in the intensity of the band centered around 210-220 nm. However, the loss rate differs between storage temperatures: at 4 °C a gradual loss of the initial structure is exhibited throughout the whole assay time, while at RT there is a more abrupt change in the CD spectra between day 8 and 15, indicating a greater loss of the native structure. By day 29, both samples revealed to have lost the main features of the initial secondary structure. It should be taken into account that the protein samples were stored in the absence of any stabilizing formulation, which are typically employed to increase the shelf life of biopharmaceuticals. Therefore, the observed stability in the absence of such stabilizing substances is *per se* quite encouraging.

The representation of the ellipticity recorded at 222 nm for both samples as a function of the storage time is presented in Figure 4.27. Up to day 8 the changes in the secondary structure for both samples are similar, while from day 15 to 29 the ellipticity value of the two samples diverge. These CD data suggest that, for a week of storage, the robustness of TSSFP secondary structure is not affected by the storage temperatures tested. On the other hand, for storage times above 8 days, the temperature influences the secondary structure being 4°C the best option to delay the loss of native secondary structure.

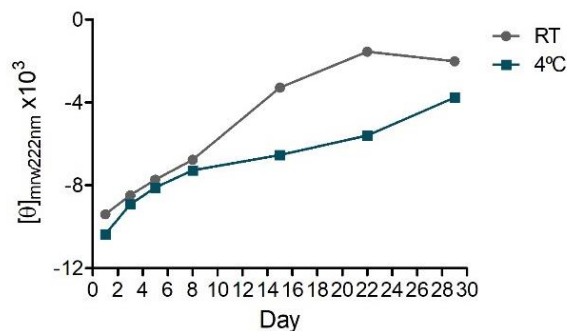


Figure 4.27 | Variation of the mean residue ellipticity at 222 nm (in deg·cm²·dmol⁻¹) as a function of the TSSFP storage time. (Grey dots) Mean residues ellipticity at 222 nm recorded for the sample storage at room temperature. (Blue squares) Mean residue ellipticity acquired at 222 nm for the sample placed at 4 °C. Both experimental data are connected by a solid line.

4.2.4. Functional characterization

From the biopharmaceutical point of view, TSSFP exhibits advantageous structural features, including suitable folding, high thermostability and reasonable storage stability. Therefore, TSSFP was then used in an immunization program (detailed in *Materials and Methods* chapter) to enable its functional characterization.

Sera samples from the two immunized rabbits were collected at distinct days: pre-immune bleed at day 0, small bleed at day 38, medium bleed at day 66, and final bleed at day 87. Additionally, a portion of the final bleed was purified, resulting in the purified antibodies sample. These were examined to assess the presence of antibodies (ABs) generated against the TSSFP and their ability to recognize the immunogenic construct as well as the SARS-CoV-2 S-protein and also their neutralization activity.

Evaluation of the antibodies generated

The existence of ABs in the sera samples able to bind TSSFP and SARS-CoV-2 S-protein was initially assessed by western blot (WB) resulting in the membranes shown in Figure 4.28. Here, WB was used as a qualitative method only to answer the question: do ABs generated bind to both antigen and S-protein?

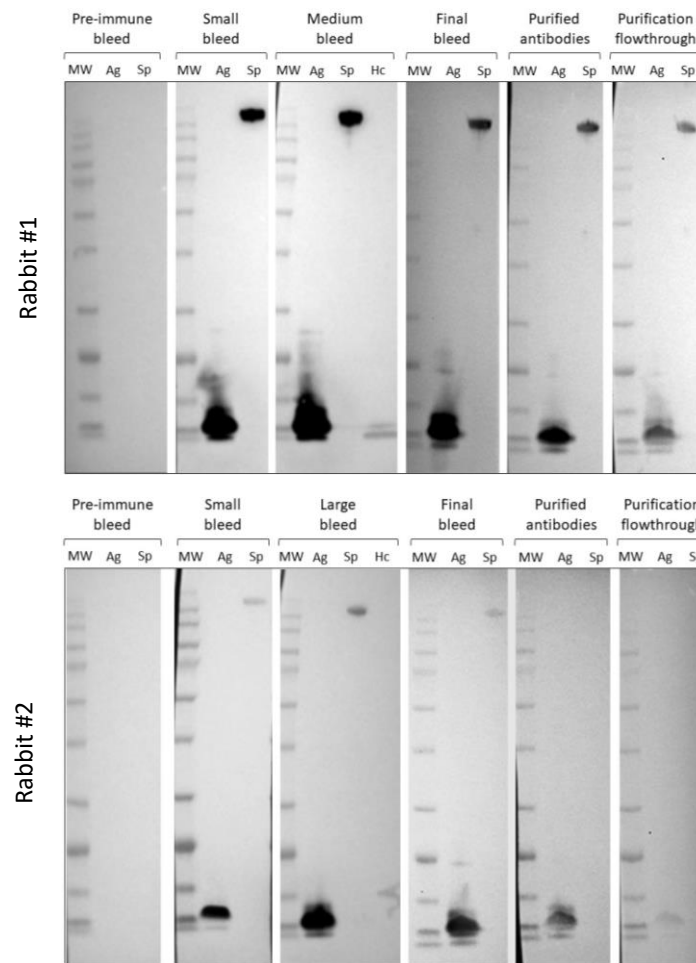


Figure 4.28 | Evaluation of the antibodies generated in the immunization program by western blot.(Top) Sera samples of rabbit 1. (Bottom) Sera samples of rabbit 2. (MW) NZYColour Protein Marker II. (Ag) Antigen, 200 ng of TSSFP was loaded into the well. (Sp) SARS-CoV-2 S-protein, 200 ng loaded into the gel. (Hc) Histidine-Tag control, 400 ng loaded.

Firstly, the intensity of the bands either corresponding to TSSFP or S-protein is higher for the rabbit 1 samples. This could be the result of two distinct reasons: different immune responses of rabbit 1 and 2 due to individual characteristics or technical limitations associated with inconsistency in the WB protein transfer process resulting in less protein in the membrane available for ABs binding. This issue is addressed further on.

Secondly, for both rabbits, ABs were generated in response of TSSFP injections as observed by the appearing of bands in the small, medium and final bleeds. The presence of lower molecular weight bands (~15 kDa) reflects the binding to the TSSFP while the higher molecular weight bands (~180 kDa) reveal the recognition of SARS-CoV-2 S-protein.

Since TSSFP was used in the immunization program with a Histidine-Tag, an additional control was added for medium bleed samples to analyze the presence of ABs targeting this region. A thin band was detected in the rabbit 1 samples meaning that a part of the ABs generated against the antigen are targeting the Tag.

Lastly, the purified ABs samples and the flowthrough collect during the purification process revealed contrasting characteristics. Bands corresponding to TSSFP and S-protein

appear in both rabbit 1 samples. On the other hand, samples from rabbit 2 only seem to bind TSSFP suggesting the loss of ABs able to recognize the S-protein.

The following step consisted in studying the ABs ability to recognize the folded version of the SARS-CoV-2 S-protein, employing enzyme linked immunosorbent assays (ELISA).

The ELISA results in Figure 4.29 revealed that indeed the rabbit 1 generated a greater amount of polyclonal antibodies able to bind and recognize the S-protein when compared with rabbit 2. Moreover, the small and medium bleeds of rabbit 1 exhibited a similar profile suggesting that the antigen boost performed between these two bleeds did not increase the ABs production. However, a difference in the response observed for the latter samples and the final bleed which was collected 30 days after the last antigen boost is detected. In accordance with the previous WB, rabbit 1 purified ABs and flowthrough samples contain S-protein-binding ABs.

The small, medium and final bleeds of rabbit 2 present the same correlation as rabbit 1. Since the final bleed sample contains ABs which became absent after the purification process either in the purified ABs or flowthrough sample, an incident may have occurred during the purification or the transportation.

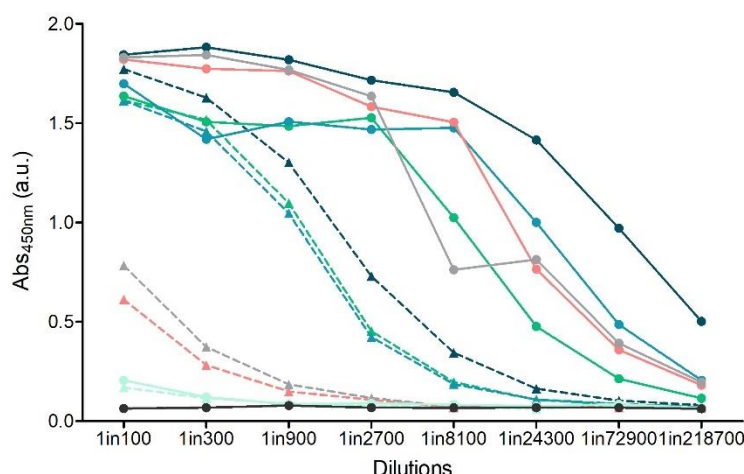


Figure 4.29 | ELISA result showing the ability of the ABs in the samples collected during the immunization program to bind to the SARS-CoV-2 S-protein. The plate was coated with 25 ng of SARS-CoV-2 S-protein. Serial dilution of the pre-immune bleed (light blue), small bleed (green), medium bleed (blue), final bleed (dark blue), purified antibodies (pink), flowthrough (grey) and blank (black) samples. The data collected from the rabbit 1 samples are represented as dots connected by solid lines and the rabbit 2 samples are exhibited as triangles connected by dashed lines. The secondary antibody anti-rabbit IgG-HRP was used.

Nevertheless, the endpoint titers were determined to be: > 218,700 for all rabbit 1 samples; >72,900 for the small, medium and final bleeds of rabbit 2; >2,700 for the purified antibodies and the flowthrough of rabbit 2.

Neutralization ability of the antibodies generated

Once the ABs ability to bind to SARS-CoV-2 S-protein was confirmed, neutralization assays were performed to examine the effect of these in the infection process. 293T-ACE2 cells expressing the human ACE2 receptor on cell surface were employed as host for the spike-pseudotyped lentiviruses expressing the SARS-CoV-2 S-protein.

All the samples studied in the previous experiment were also tested in this assay resulting in the neutralization profiles presented in Figure 4.30. The diverse samples from both rabbits exhibited similar behaviors standing out the unusual shape of the curves corresponding to the bleeds and the decrease in the relative infection induced by the pre-immune bleed. Since none of the pre-immune bleed samples revealed the presence of ABs able to bind to S-protein in the WB or ELISA, the observed behavior in the neutralization assays is unexpected and most likely unspecific. Therefore, an interfering molecule might be present in the rabbit sera.

Contrary to sera samples that are complex and difficult to fully characterize, the purified ABs samples are expected to be simpler and contain only the polyclonal ABs able to bind to TSSFP solubilized in the purification buffer. By comparing the resulting neutralization curves of the purified ABs with both positive and negative controls, a typical sigmoidal shaped curve identical to the positive control is observed indicating a decrease in the infection. NT50 values (Table 4.3) of 79.1 ± 6.2 and 66.6 ± 5.1 were calculated for rabbit 1 and 2 samples, respectively.

According to the WB and ELISA results, the flowthrough samples were anticipated to exhibit a certain effect on the relative infection. Surprisingly, these samples revealed to have the greatest impact on the relative infection resulting in NT50 values of 208.7 ± 5.3 and 278.3 ± 3.0 for rabbits 1 and 2, correspondingly. Some hypotheses may explain these results: (I) the purification process was inefficient in separating the majority of the ABs which bind to TSSFP possibly due to column saturation leading to a strong presence of ABs in the flowthrough; (II) in addition to the latter, the serum components are also present in the flowthrough samples which means the interfering molecule is likely contributing for the decrease in the relative infection.

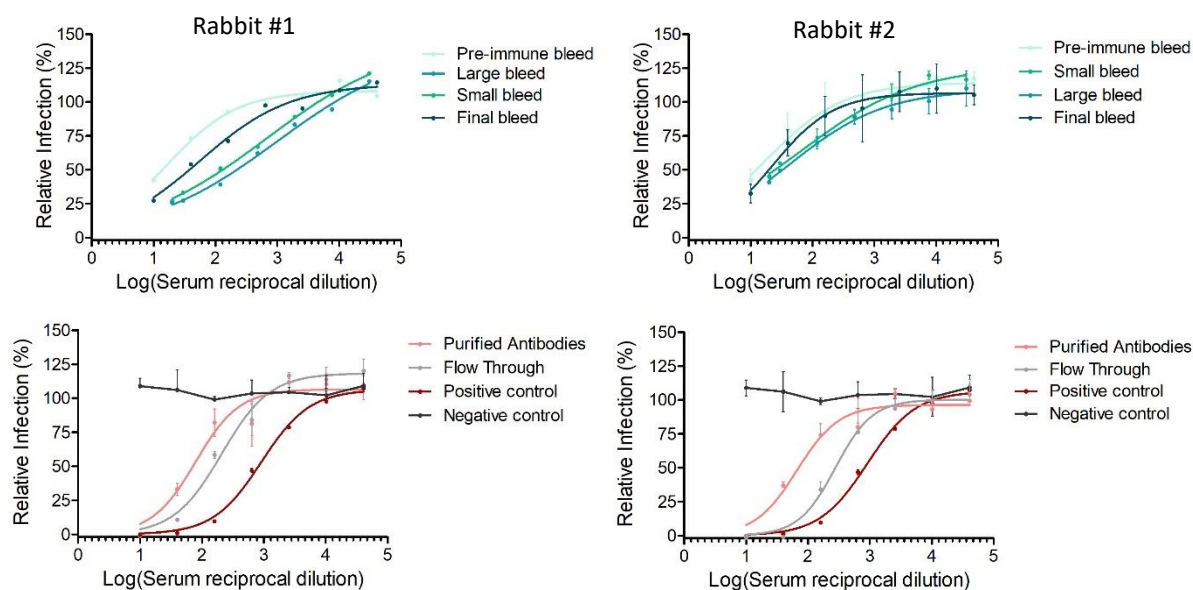


Figure 4.30 | Neutralization curves of the samples collected during the immunization program with TSSFP of rabbit 1 (left) and rabbit 2 (right). (Light blue) Pre-immune bleeds. (Green) Small bleeds. (Blue) Large bleeds. (Dark blue) Final bleeds. (Pink) Purified antibodies. (Grey) Purification flow trough. (Red) Positive control. (Black) Negative control.

Table 4.3 | The 50% neutralization titer determined for all the samples collected during the immunization program using TSSFP.

Rabbit	Sample	NT50
1	Pre-immune bleed	17.8 ± 4.6
	Small bleed	62.9 ± 6.3
	Medium bleed	44.2 ± 3.7
	Final bleed	22.8 ± 2.7
	Purified ABs	79.1 ± 6.2
	Flowthrough	208.7 ± 5.3
2	Pre-immune bleed	16.9 ± 3.4
	Small bleed	486.5 ± 5.9
	Medium bleed	375.0 ± 14.2
	Final bleed	56.2 ± 5.3
	Purified ABs	66.6 ± 5.1
	Flowthrough	278.3 ± 3.0

To further clarify these matters, the focus was directed to the purified ABs samples to determine if the neutralization observed was indeed caused by the ABs binding to the S-protein. The next step consisted in evaluating the effect of the purification buffer (PBS, 0.01% thimerosal, 0.1% BSA) in the neutralization assay. The plots in Figure 4.31 reveal that the buffer neutralizes the infection with a NT50 of 53.6 ± 2.3 which is similar to the one determined for the purified antibodies. Moreover, when diluting the positive control in the purification buffer its neutralizing ability is slightly increased from a NT50 of 230.6 ± 6.1 to 271.1 ± 10.1 . These findings do not corroborate the presence of neutralizing antibodies in the samples.

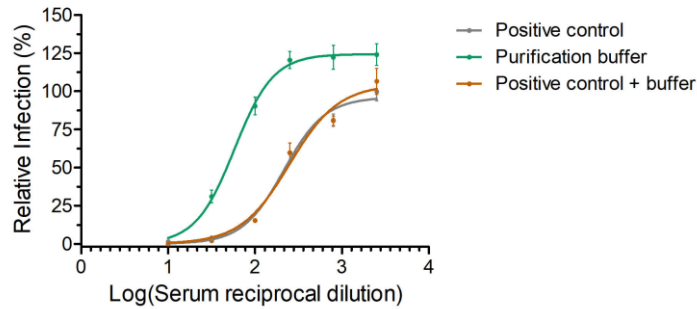


Figure 4.31 | Neutralization curves of the control samples. (Grey) Positive control (right). (Green) Purification buffer PBS, 0.01% thimerosal, 0.1% BSA. (Orange) Positive control diluted in purification buffer.

Effect of the antibodies in fusion mechanism steps

Although the ABs generated during the immunization program against the TSSFP were able to bind to the desired target, the SARS-CoV-2 S-protein, no neutralizing features were confirmed. As mentioned before, these ABs were thought to block the infection process by binding to the S-protein S2' cleavage site and prevent the S2 activation followed by the FPs exposure.

To better understand the interaction between ABs and the S-protein, the S-protein cleavage by the membrane protease of 293T-ACE2 cells in the presence and absence of the purified ABs was examined by WB (Figure 4.32). None of the purified ABs or incubation times seemed to affect the S-protein cleavage, meaning the ABs binding was not enough to prevent S2 activation, which explains the lack of neutralization.

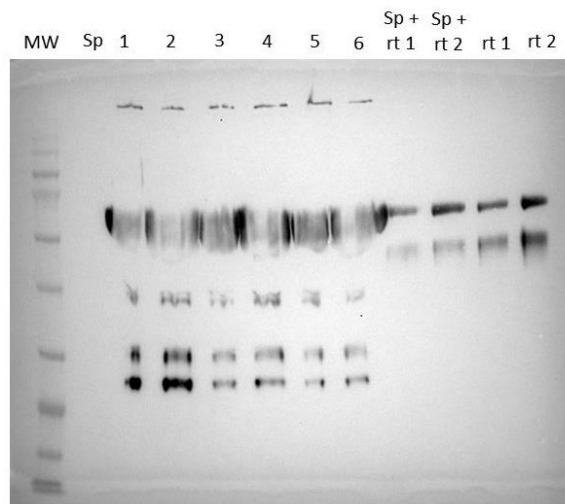


Figure 4.32 | Evaluation of the purified antibodies to prevent S2' cleavage. (MW) NZYColour Protein Marker II. (Sp) SARS-CoV-2 S-protein, 120 ng. (1 and 2) S-protein samples after incubation with 293T-ACE2 cells for 1h and 6h, respectively. (3 and 4) S-protein pre-incubated with the purified antibodies from rabbit 1 and then incubated with 293T-ACE2 cells for 1h and 6h, respectively. (5 and 6) S-protein pre-incubated with the purified antibodies from rabbit 2 and then incubated with 293T-ACE2 cells for 1h and 6h, respectively. (Sp + rt 1) Mixture of S-protein and the purified antibodies from rabbit 1. (Sp + rt 2) Mixture of S-protein and the purified antibodies from rabbit 2. (rt 1) Purified antibodies from rabbit 1. (rt 2) Purified antibodies from rabbit 2.

Even though no neutralizing ABs were obtained, the *in silico* design approach was validated since the antigen constructed induced the production of ABs in the host targeting the desired region and the resulting ABs recognize the S-protein. Additionally, a highly thermostable and resistant scaffold suitable to build future immunogenic constructs was identified.

4.3. SARS-CoV-2 putative fusion peptides

4.3.1. Identification of SARS-CoV-2 putative fusion peptides

Considering the characteristics mentioned in the *Materials and Methods* section shared among FPs, the sequence conservation of different CoVs and the state-of-the-art research^{91 34 92}, four putative FPs were identified as represented in Figure 4.33, comprising the features in Table 4.3.

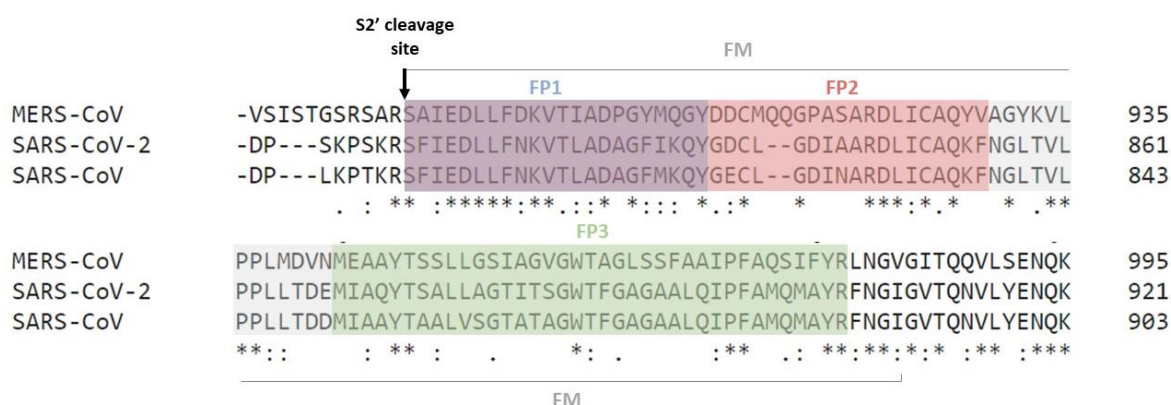


Figure 4.33 | SARS-CoV-2 putative FPs identified. (Lilac) FP1 corresponding to residues 816-837. (Pink) FP2 comprising the residues 816-855). (Green) FP3 composed of the residues 869-905. (Grey) Fusion Module or FM covering the residues 816-905.

Table 4.3 | SARS-CoV-2 putative FPs main characteristics: length in number of residues, residues conservation in percentage, percentage of glycine and alanine content and hydrophobic residues content in percentage.

Fusion peptide	Residues	Amino acid sequence	Length (nr. residues)	Conservation (%)	G & A content (%)	Hydrophobic residues content (%)
FP1	816-837	SFIEDLLFNKVTADAGFIKQY	22	50	14	41
FP2	816-855	SFIEDLLFNKVTADAGFIKQYGDCLGDIAARDLICAQKF	40	45	20	48
FP3	869-905	MIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYR	37	19	32	57
FM	816-905	SFIEDLLFNKVTADAGFIKQYGDCLGDIAARDLICAQKFNGLTVPPLLTDEMIAQYTSALLAGTITSGW	90	33	26	61

Among the putative FPs identified, FP1 is the most explored in the literature followed by FP2 and FP3. Additionally, some authors proposed that the adjacent peptides FP1 and FP2 cooperate during the membrane fusion process.^{36 93} The fusion module (FM) is a novel candidate which presumes the requirement of more than one small FP to induce the fusion between viral and host cell membranes. Thus, the fusion module encompasses all the proposed FPs and is constituted by two moderately hydrophobic regions (FP1 and FP3) separated by a hydrophilic region. Our hypothesis is that both hydrophobic regions insert into the host membrane, whereas the hydrophilic region protrudes to the solvent.

FP1, FP2 and FM common sequence begins right after the S2' cleavage site and includes the conserved LLF motif identified, by mutagenesis assays in previous studies, to be a key structural element for membrane insertion⁹⁴. Moreover, one Ca²⁺ binding site established

between E819/D820 and D830 thought to increase the fusogenicity is also observed in these three peptides.

The amino acid residues shared by FP2 and FM comprise other possible Ca^{2+} binding sites formed between D839, D843 and D848. Two other relevant features are also present: the disulfide bond formed by C840 and C851 contributing to the peptide's structural stability and the GxxxG motif which is found in many high-affinity associations between transmembrane helices.

Lastly, the FP3 and FM common region is the most hydrophobic enabling interactions with phospholipid's tails. Additionally, this region is rich in G and A amino acid residues known to improve the flexibility of peptides.

To experimentally evaluate the FPs effect on lipid membranes, the smaller peptides (FP1, FP2 and FP3) were synthesized and the FM was recombinantly expressed in *E.coli* and purified with chromatographic techniques explained below.

4.3.2. Expression and purification of the FM

FM is a segment of SARS-CoV-2 S-protein whose expression in *E.coli* was still unexplored. Therefore, expression tests were performed. The soluble fraction of each collected cellular extract aliquot was examined by the WB in Figure 4.34 revealing that only two conditions were suitable for the FM expression. The expression scale-up was conducted with *E. coli* BL21(DE3)star strain, LB medium, 0.1mM IPTG, at 18 °C for 18h.

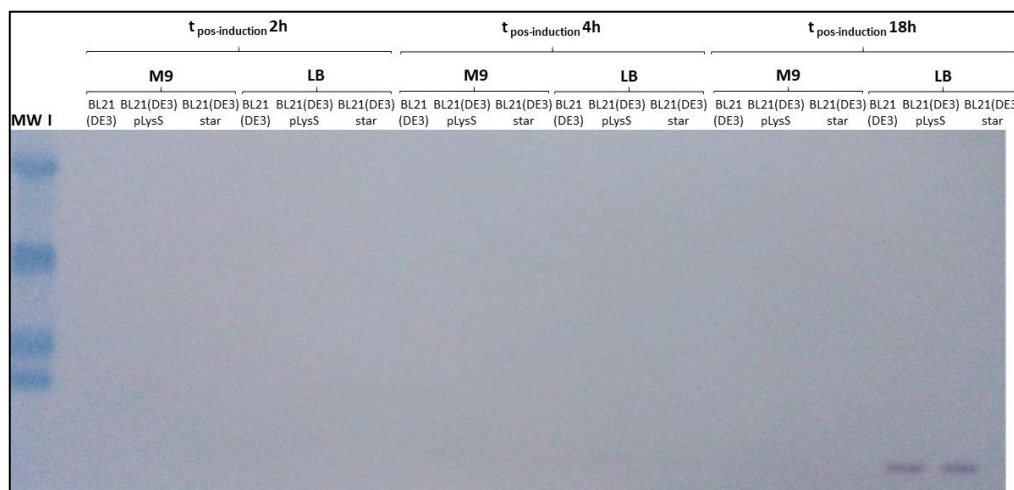


Figure 4.34 | FM expression profile evaluated by WB. MW: NZYColour Protein Marker II. Each lane corresponds to a distinct expression condition, accordingly to the labels: BL21(DE3), BL21(DE3)star, BL21(DE3)pLysS - *E. coli* strains; M9 and LB – culture media; $t_{\text{pos-induction}}$ 2h, 4h and 18h – expression time. The expression was induced with 0.1 mM IPTG and at 18°C .

To improve the purification process, the tag MGSSHHHHHSQDPNSWELQ- was added to the peptide N-terminus. The FM purification protocol began with an affinity chromatography, but the peptide did not bind to the column matrix, being eluted in the flow-through as shown in the SDS-PAGE gel in Figure 4.35 A. Next, the sample was filtered with an ultra-filtration concentrating filter with a Mw cut-off of 30 kDa to remove the contaminants with higher Mw.

Then, a concentrator with a Mw cut-off of 3 kDa was used to concentrate the FM-containing fraction as presented in Figure 4.35B. Finally, a size-exclusion chromatography was performed to further separate the FM which was eluted with 92.8 mL (Figure 4.35C and D). The overall yield of FM was 0.56 mg per liter of culture medium.

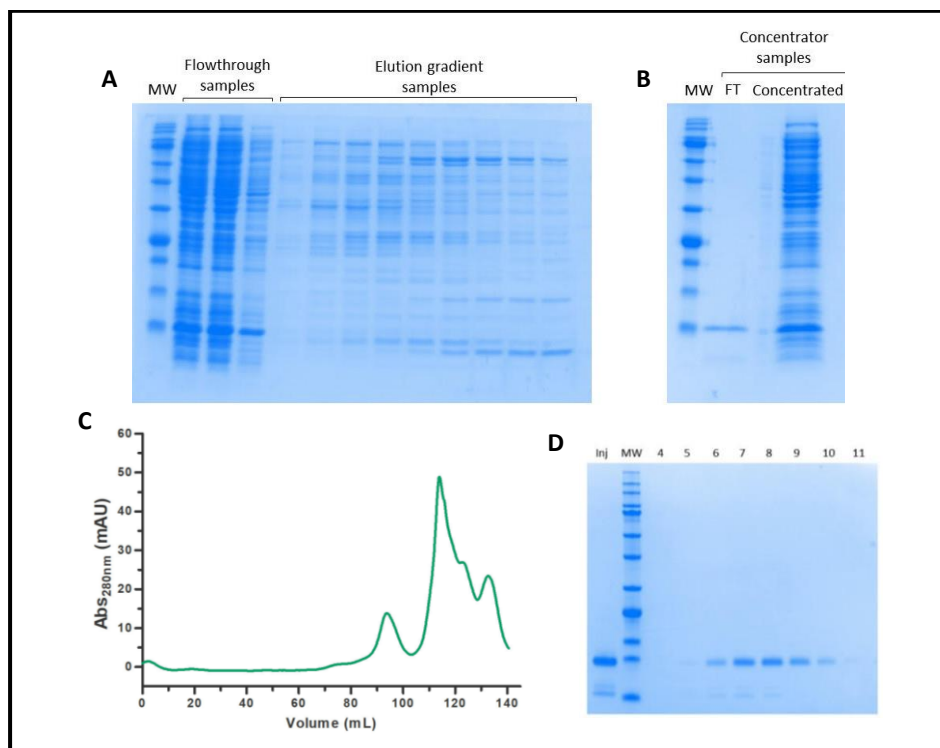


Figure 4.35 | SDS-PAGE gels (12.5%) and chromatogram of FM purification steps. (A) SDS-PAGE gel of the fractions collected during the affinity chromatography step. (B) SDS-PAGE gel of the concentrator samples, flowthrough and concentrated. (C) Chromatogram obtained after the size-exclusion chromatography with Superdex 75 prep grade column. (D) SDS-PAGE gel where the fractions collected during the size-exclusion chromatography were examined. Green line – absorbance at 280 nm. MW - NZYColour Protein Marker II. Inj – sample injected in the purification step. FT- flowthrough collected during the washing step. Numbers – number corresponding to the fractions collect.

4.3.3. FPs aggregation in aqueous solution

In contrast with FM, the smaller FPs were chemically synthesized so their aggregation in solution required examination. DLS and fluorescence spectroscopy were employed to address this matter. First DLS was used to monitor the particles hydrodynamic size with the increase of the peptides concentration, and the results obtained are presented in Figure 4.36. The appearance of populations with higher hydrodynamic size along with the increase in the derived count rates recorded indicate aggregation.

For FP1, the results obtained show evidence of aggregation at concentrations above 50 μ M, while for FP3, aggregation starts at a much lower concentration (10 μ M). FP2 exhibits a special feature, since the increase in concentration seems to shift the hydrodynamic size to slightly higher values, suggesting the occurrence of a controlled aggregation or oligomerization.

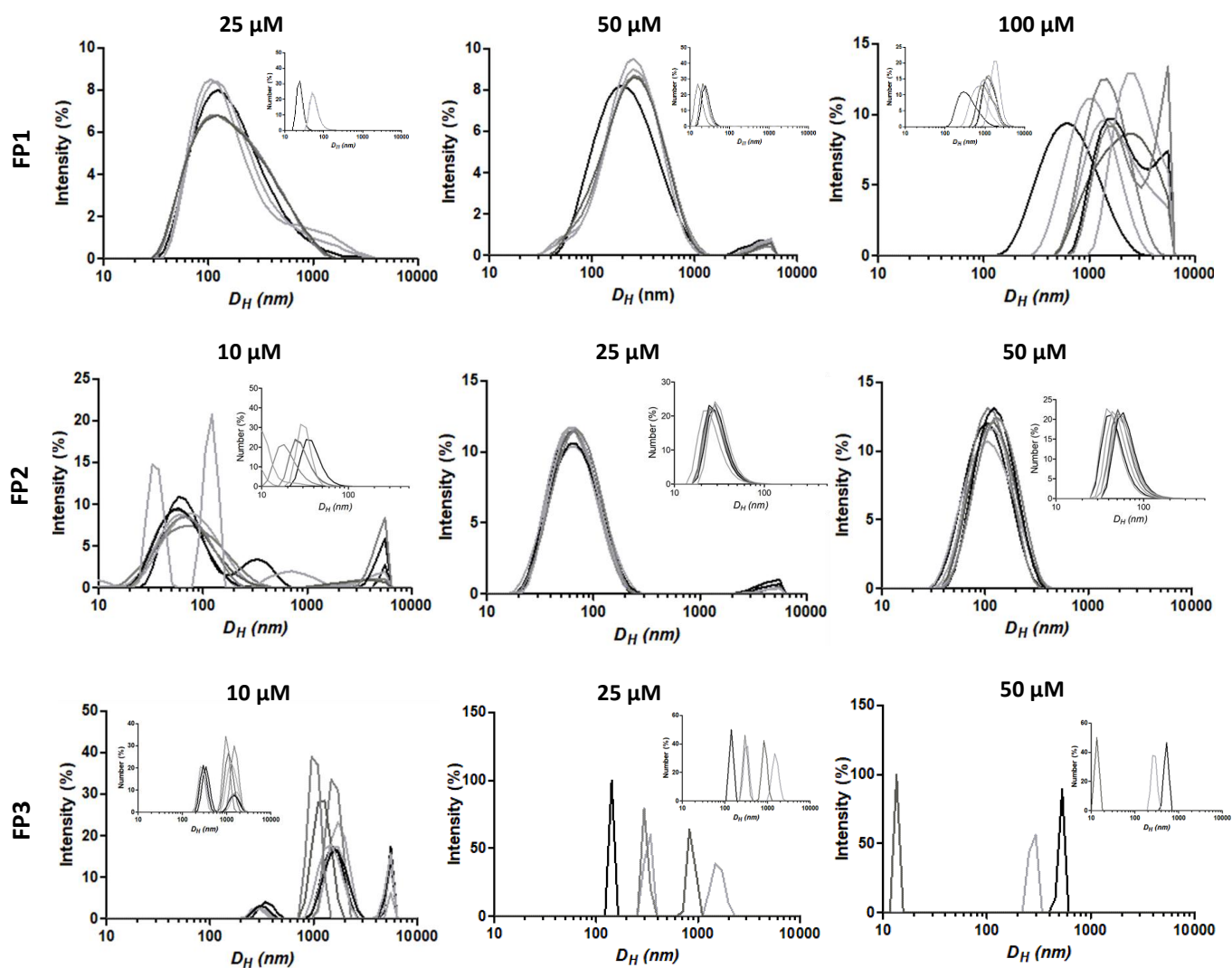


Figure 4.36 | DLS profile of samples containing increasing concentrations of FP1, FP2 or FP3 in 10 mM HEPES 150 mM NaCl pH 7.4.

To further clarify this matter, fluorescence spectroscopy assays using ANS probe were performed. ANS probe in aqueous solution presents a fluorescence emission peak centered at 520 nm with low fluorescence intensity. However, when the environment is hydrophobic, the fluorescence emission intensity increases, and the emission peak shifts to the blue region around 450 nm. The aggregation of a peptide usually tends to create hydrophobic cores where the hydrophobic residues are shielded from the aqueous environment. The results obtained (Figure 4.37A and B) show that for FP3, an increase in the peptide concentration results in an increase of the fluorescence emission intensity of ANS as well as a shift in the fluorescence emission peak wavelength ($\lambda_{\text{emi max}}$) from 545 to 468 nm, indicating the occurrence of uncontrolled aggregation. On the other hand, for FP1 no changes were observed even at concentrations up to 50 μM . In the case of FP2, a gradual increase in the fluorescence emission intensity of ANS is detected at peptide concentrations from 5 to 50 μM , along with a shift in $\lambda_{\text{emi max}}$ from 548 to 501 nm. Interestingly, this shift appears to reach a plateau between 5 and 15 μM which may also be indicative of an oligomerization process.

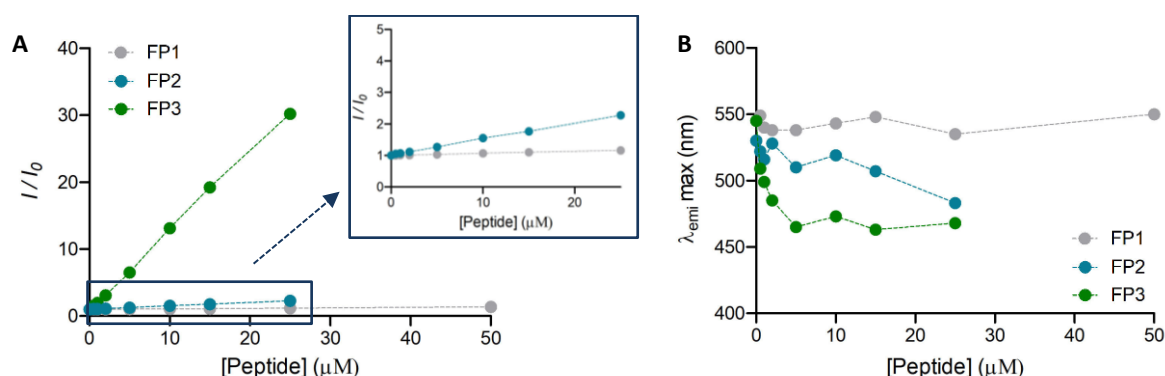


Figure 4.37 | Evaluation of the FP1 (grey), FP2 (blue) and FP3 (green) aggregation using ANS probe. (A) Maximum of the emission intensity as a function of each peptide concentration. (B) Effect of the peptide concentration in the fluorescence emission peak wavelength ($\lambda_{\text{emi max}}$).

4.3.4. SARS-CoV-2 putative FPs effect on membrane model systems

Although some studies have already been conducted with different SARS-CoV-2 FPs proposals, little is known about their mechanism of interaction with host membranes. General effects of viral FPs in membrane model systems including fusion between vesicles, vesicles-content release and vesicles aggregation are established and may be useful to shed light into this matter.

FPs-induced fusion and perturbation on lipid vesicles

FRET-based assays were selected to evaluate the FPs-induced fusion between NBD-RhB doped vesicles and non-doped vesicles. Fluorescence spectroscopy assays were also used to follow the FPs-induced leakage of vesicles containing 5,6-CF. LUVs composed of POPC, POPC:POPG (4:1) and POPC:POPG:Chol (3:1:1) were used to assess if a different lipid composition influences the FPs-membrane interaction. Moreover, diverse ratios of lipid/peptide

(lip/pep) ranging from 2 to 80 were employed to evaluate if the effects on the vesicles were concentration dependent. The lip/pep ratio of 2 or 4 were considered as low, the lip/pep ratio of 20 represents the intermediate ratio and the lip/pep ratios of 57 or 87 were established as high.

When examining Figure 4.38, it is possible to detect FP1-induced fusion and low perturbation on vesicles. FP1 fusogenicity reveals to be identical in POPC and POPC:POPG vesicles being null at the high lip/pep ratio. At the intermediate ratio the FP1 effect is similar on all the lipidic compositions tested. On the other hand, a maximum fusion percentage of 15.2 ± 0.6 % is observed for POPC:POPG:Chol vesicles at the low lip/pep ratio. The FP1-induced fusion on these vesicles was not significantly altered by increasing the lip/pep ratio from intermediate to high, indicating the absence of a concentration dependency. The role of cholesterol in viral membrane fusion is extensively explored in the literature, being established as a modulator of membrane organization and dynamics. Additionally, previous studies have shown that cholesterol promotes the oligomerization of SARS-CoV FP increasing its fusogenicity on lipidic vesicles.^{95 96 97} Although FP1 reveals evidence of preferentially interacting with cholesterol-containing vesicles, further experiments would be necessary to confirm the same profile.

The FP1-induced membrane leakage is subtle reaching the highest leakage percentage of 4.5 ± 0.8 % on POPC vesicles at the intermediate lip/pep ratio. In this case, the effect on POPC:POPG and POPC:POPG:Chol vesicles are similar for both ratios used. Since the highest and lowest values determined only differ in approximately 4 %, no clear assumptions were done.

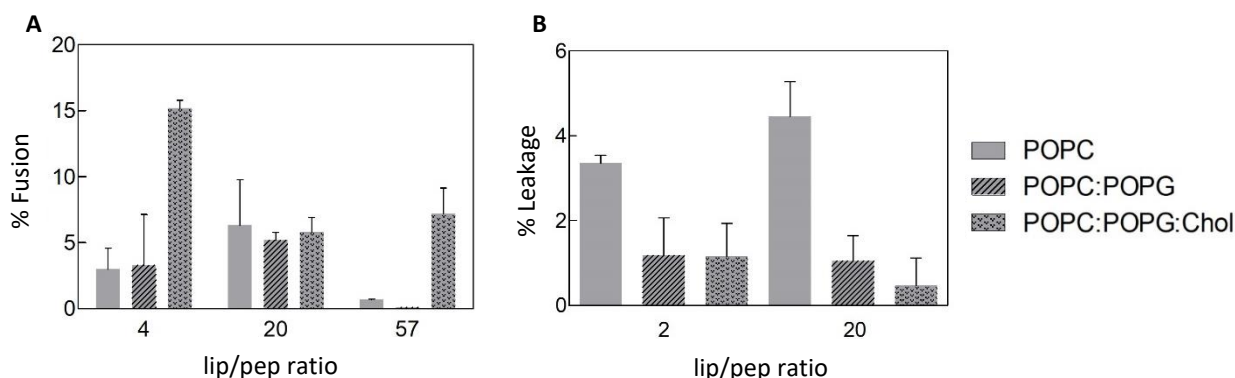


Figure 4.38 | Evaluation of the FP1-induced fusion (A) and perturbation (B) in LUVs composed of POPC (solid), POPC:POPG (4:1) (striped) and POPC:POPG:Chol (3:1:1) (dotted).

The analysis of Figure 4.39 allows identifying the FP2 effect on vesicles by promoting fusion as well as enabling content leakage. The FP2-induced vesicle fusion exhibits high standard deviation between replicates, especially at low and intermediate ratios, which might suggest diverse types of interaction with the lipidic membrane. For instance, peptide-peptide interactions may be occurring and, consequently peptide populations with different affinity for membrane binding arise. At high lip/pep ratio, FP2 promotes 9.4 ± 0.1 % of fusion on POPC vesicles, followed by 6.9 ± 2.4 % on POPC:POPG vesicles and 2.6 ± 1.8 % on POPC:POPG:Chol vesicles. Interestingly, the pattern observed for the membrane perturbation phenomena which led to 5,6-CF leakage differs, reaching higher percentages of leakage at low lip/pep ratio. Additionally, the perturbation is dependent on the lipid composition, being greater in

POPC:POPG and POPC:POPG:Chol at the low ratio presenting leakage values of 18.0 ± 9.8 % and 9.0 ± 3.7 %.

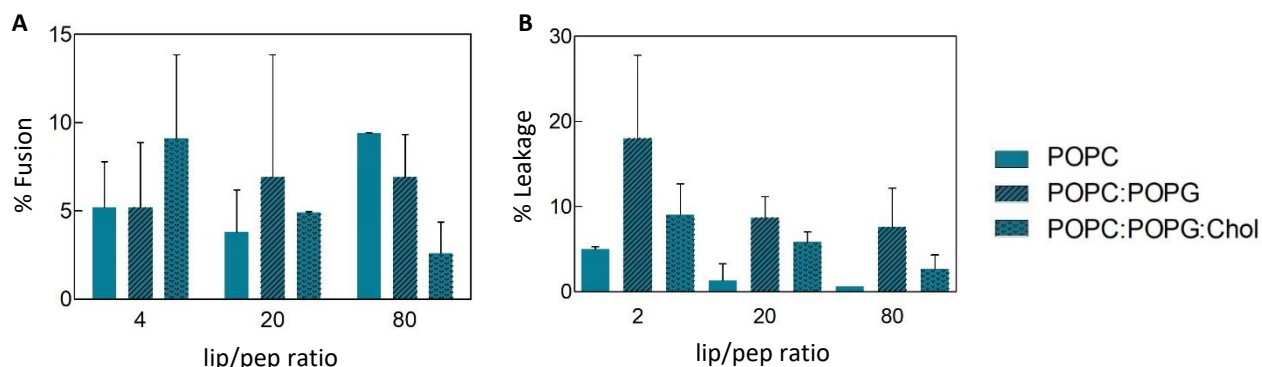


Figure 4.39 | Evaluation of the FP2-induced fusion (A) and perturbation (B) in LUVs composed of POPC (solid), POPC:POPG (4:1) (striped) and POPC:POPG:Chol (3:1:1) (dotted).

The FP3 effects on vesicles is shown in Figure 4.40. Both FP3-induced fusion and membrane perturbation are more efficient at low lip/pep ratio. At the higher ratio almost no activity is detected, suggesting that these phenomena are influenced by local concentration of peptide. When comparing the influence of the lipidic composition on both assays, different profiles are observed. The fusogenic activity appears to be higher on POPC (23.8 ± 3.8 %), than on POPC:POPG (17.8 ± 1.2 %) and lastly on POPC:POPG:Chol (10.2 ± 1.1 %) vesicles. In contrast, the vesicle content release promoted by the FP3 does not exhibit clear evidence of lipid composition dependence achieving 16.7 ± 3.7 % on POPC vesicles, 24.2 ± 9.7 % on POPC:POPG vesicles and 17.9 ± 9.4 %.

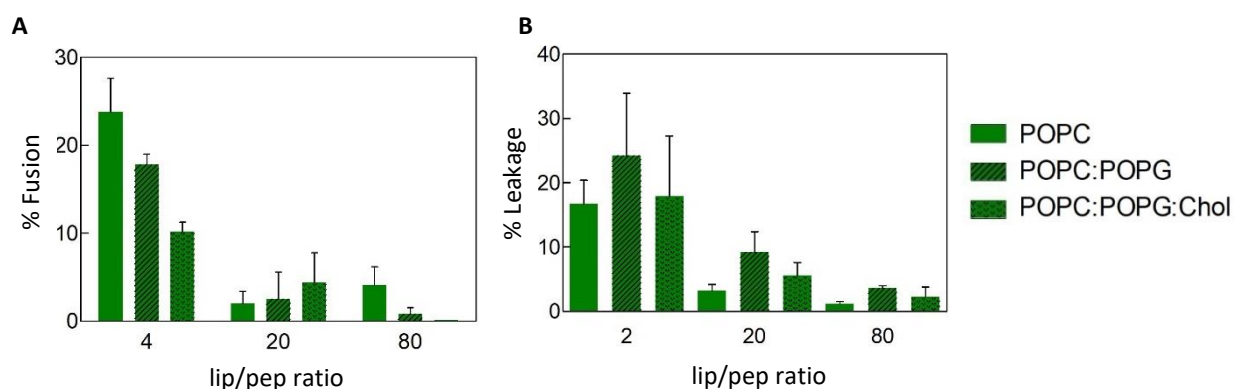


Figure 4.40 | Evaluation of the FP3-induced fusion (A) and perturbation (B) in LUVs composed of POPC (solid), POPC:POPG (4:1) (striped) and POPC:POPG:Chol (3:1:1) (dotted).

The interpretation of Figure 4.41 enables to evaluate the FM effect on the three vesicle compositions tested. Regarding FM-induced fusion, the highest activity is detected on POPC vesicles at the intermediate lip/pep ratio achieving 19.4 %, followed by a fusion percentage of 10.9 % on POPC:POPG:Chol. Similarly to the previous FPs, FM perturbation and fusion effects do not share the same patterns. The leakage promoted by FM is greater at low lip/pep ratio and similar on POPC:POPG (11.8 %) and POPC:POPG:Chol (10.8 %). In this case, the internal content

release of POPC (7.2 %) vesicles was the lowest determined. FM also exhibits evidence of concentration dependency.

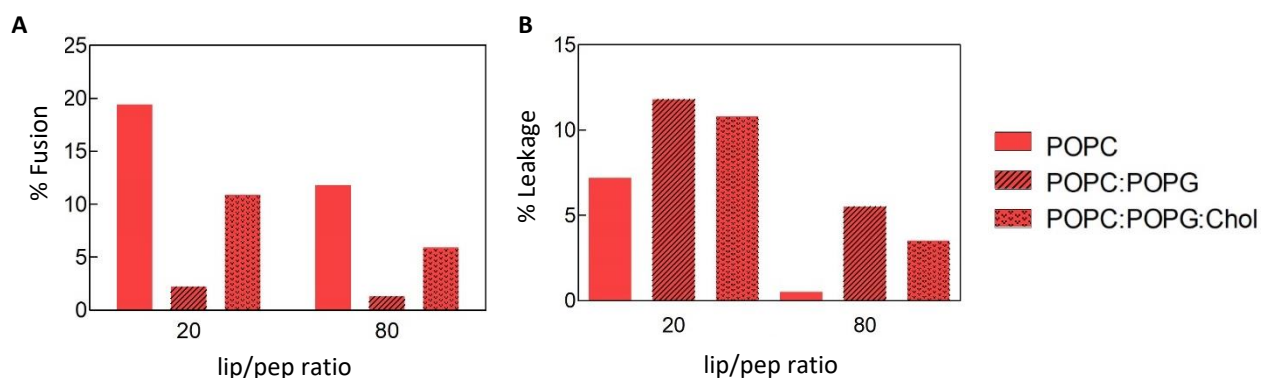


Figure 4.41 | Evaluation of the FM-induced fusion (A) and perturbation (B) in LUVs composed of POPC (solid), POPC:POPG (4:1) (striped) and POPC:POPG:Chol (3:1:1) (dotted).

When comparing the FPs activity on vesicles fusion and perturbation at intermediate lip/pep (Figure 4.42), FM stands out as the most fusogenic peptide on POPC and POPC:POPG:Chol vesicles. In terms of inducing internal content leakage, FM activity is the highest on POPC and POPC:POPG:Chol vesicles and it assumes similar values as FP2 and FP3 for POPC:POPG vesicles.

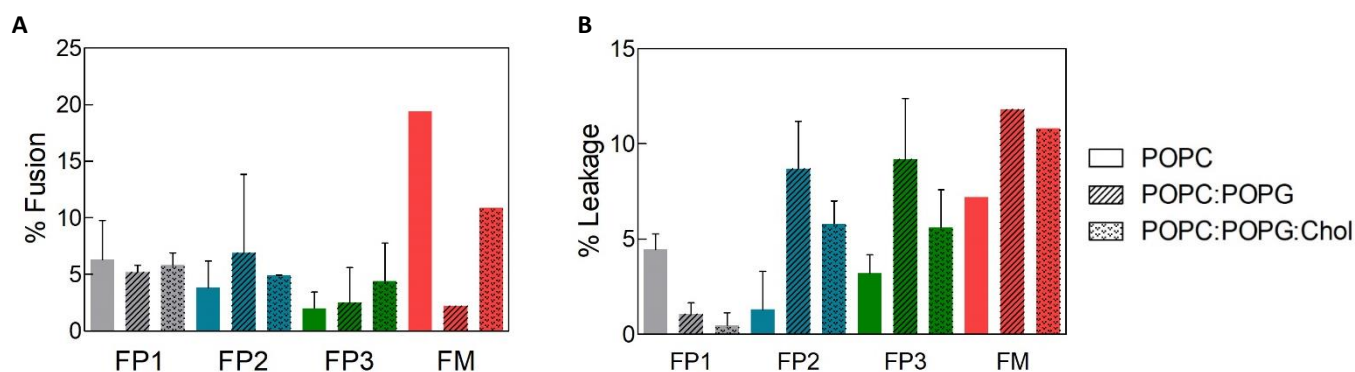


Figure 4.42 | Comparison of the percentage of fusion and leakage induced by each FP at a ratio of lip/pep of 20 on vesicles. (Grey) FP1. (Blue) FP2. (Green) FP3. (Pink) FM. LUVs composed of POPC (solid), POPC:POPG (4:1) (striped) and POPC:POPG:Chol (3:1:1) (dotted).

FPS-induced vesicle aggregation

Usually, fusion between vesicles is also associated with a macromolecular organization of these structures which results in vesicle aggregation. DLS was employed to monitor the FPS-induced vesicle aggregation in order to better characterize the FPS-vesicles interaction. Lip/pep ratios from 2 to 40 were tested on POPC and POPC:POPG vesicles.

Despite the low lip/pep used, the addition of FP1, FP2 and FM to the LUVs samples did not change the hydrodynamic size or the derived count rate recorded as shown in the representative DLS profile in Figure 4.43. Only one population centered around 100 nm was detected either in the absence or presence of these FPS.

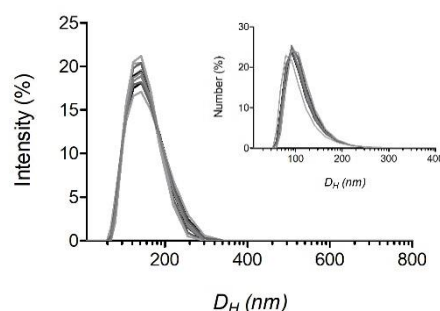


Figure 4.43 | Representative DLS profile of a mixture containing FP1 or FP2 and LUVs.

On the other hand, FP3 induced substantial changes in the LUVs DLS profile for both POPC and POPC:POPG vesicles. As presented in Figure 4.44, the decrease in lip/pep led to the appearance of populations at higher hydrodynamic diameters more evident in POPC:POPG vesicles. Comparing the FP3 in solution and the FP3-LUVs profiles acquired at the same FP3 concentration, it is possible to conclude that the resulting FP3-LUVs pattern is not the sum of the individual profiles. Moreover, the derived count rate is similar in all samples, meaning that the number of detected particles did not drastically decrease. Therefore, the fusion between vesicles cannot be confirmed. However, the formation of larger structures through the FP3-vesicles interaction might be occurring. The highest effect on the anionic POPC:POPG vesicles may be promoted in a charge dependent manner, since FP3 has a net charge of +1.

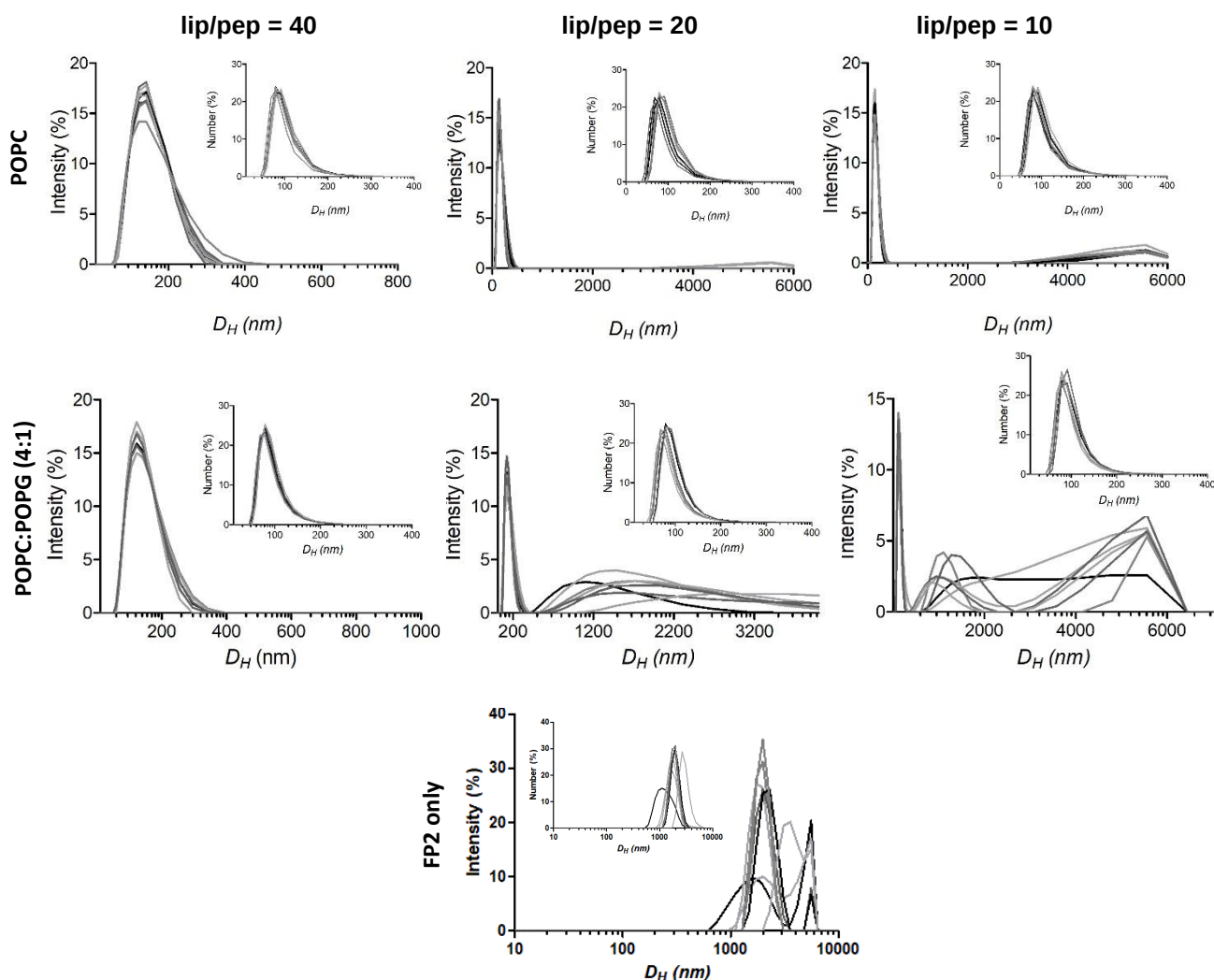


Figure 4.44 | DLS profiles of samples containing either POPC or POPC:POPG (4:1) LUVs with increasing concentration of FP3 and the FP3 profile.

Although the experimental characterization allowed to gain valuable insights about the SARS-CoV-2 putative FPs interaction with membrane model systems, they do not provide a detailed characterization of these systems at the molecular level. Therefore, molecular dynamics simulations were performed to better understand these phenomena.

4.3.5. *In silico* characterization of SARS-CoV-2 putative FPs

In this project, four different systems were studied *in silico* corresponding to the proposed SARS-CoV-2 putative FPs: FP1, FP2, FP3 and FM. Each FP structure was obtained by truncating the SARS-CoV-2 S-protein (PDB ID: 6ZGE) in distinct places as shown in Figure 4.45. The four peptides were placed in a membrane bilayer composed of POPC:POPS (4:1) using CHARMM-GUI Membrane Builder and simulated through 0.3 μ s in the FP1 case and 2.5 μ s for the remaining peptides in triplicates.

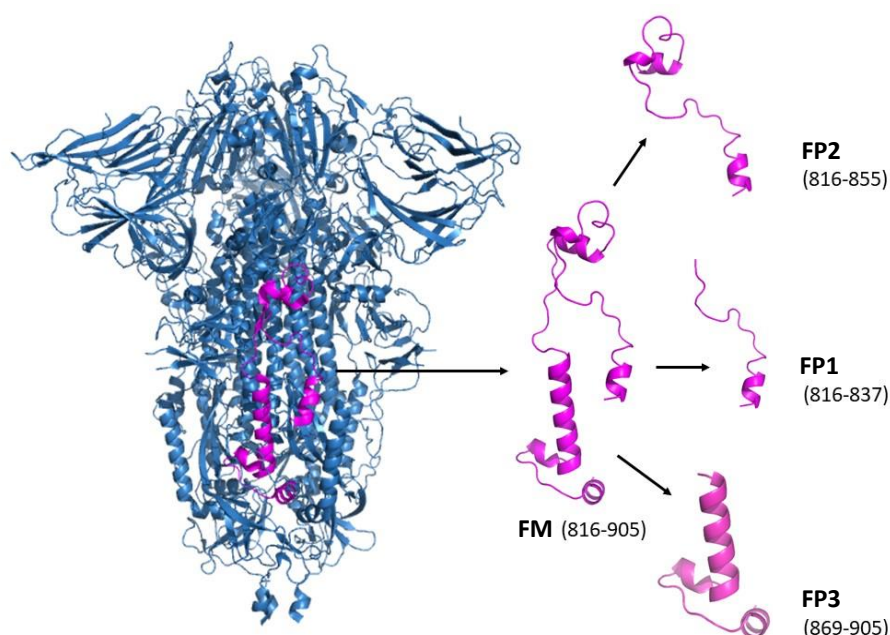


Figure 4.45 | SARS-CoV-2 spike-protein (PDB ID: 6ZGE, blue) with the four putative FPs highlighted (pink). Three smaller FPs (FP1, FP2 and FP3) and a Fusion Module (FM) were identified.

RMSD

The root mean square deviation (RMSD) of the atomic positions measures the similarity between a reference structure and the structures sampled during the simulation. In this case, the reference structure for each FP was their structure in the context of the S-protein structure determined by cryo-electron microscopy (cryo-EM). Given its definition, lower RMSD values are associated with smaller deviations from the initial conformation while higher values correspond to greater conformational changes. The temporal evolution of RMSD calculated for all the four FPs in a membrane bilayer system is presented in Figure 4.46. When analyzing the RMSD of the FPs simulated in a membrane bilayer, three main observations stand out: (I) FP1 has the lowest RMSD values; (II) FP2 and FP3 exhibit similar RMSD profiles with high heterogeneity within replicates and large temporal oscillations, and (III) FM seems to reach a plateau after 500 ns of simulation.

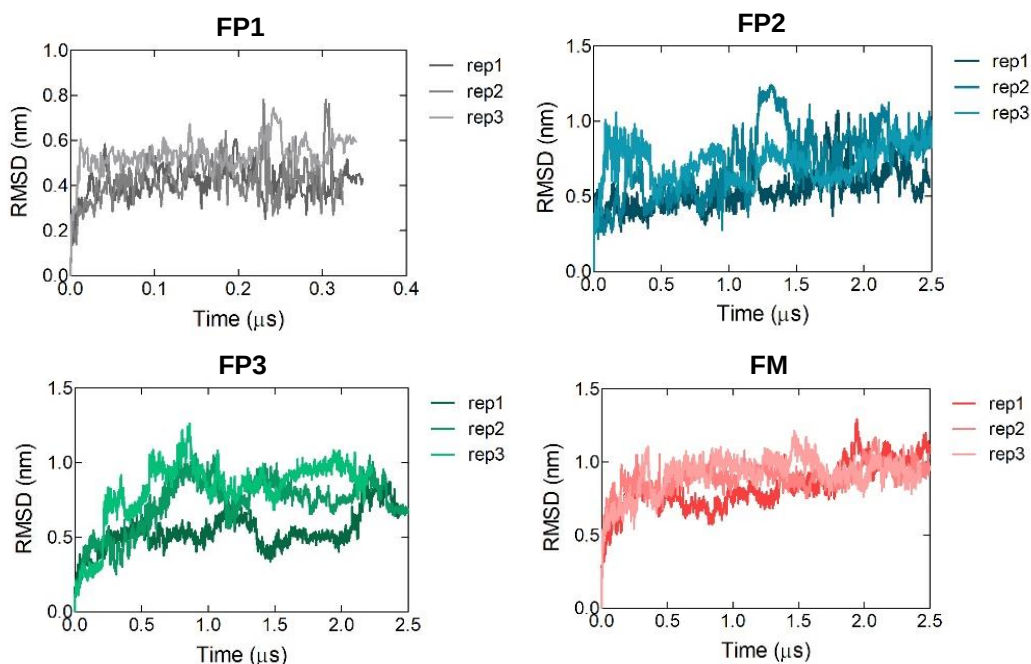


Figure 4.46 | Temporal evolution of the RMSD for all the replicas of each putative FPs. (Grey) FP1. (Blue) FP2. (Green) FP3. (Pink) FM.

The first observation arises from a small deviation between the reference and the simulated structure due to the short length of the FP1 (22 residues). The higher RMSD values observed for FP2, FP3 and FM reflect greater conformational changes during the simulation, usually associated with structural instability. However, keeping in mind that the references structures were acquired for the whole S-protein in solution and the individual peptides are being simulated in a membrane bilayer, it is reasonable to justify the conformational changes by the shift in the environment and its effects on the peptides structure. These data suggest that FP2, FP3 and FM suffer a structural rearrangement when placed in a lipidic environment.

Secondary structure

The analysis of the secondary structure of each SARS-CoV-2 putative FPs was performed using the GROMACS tool *do_dssp*. This tool allowed to compute the secondary structure characteristics for each time frame according to the information in the trajectory file as shown in the figures below.

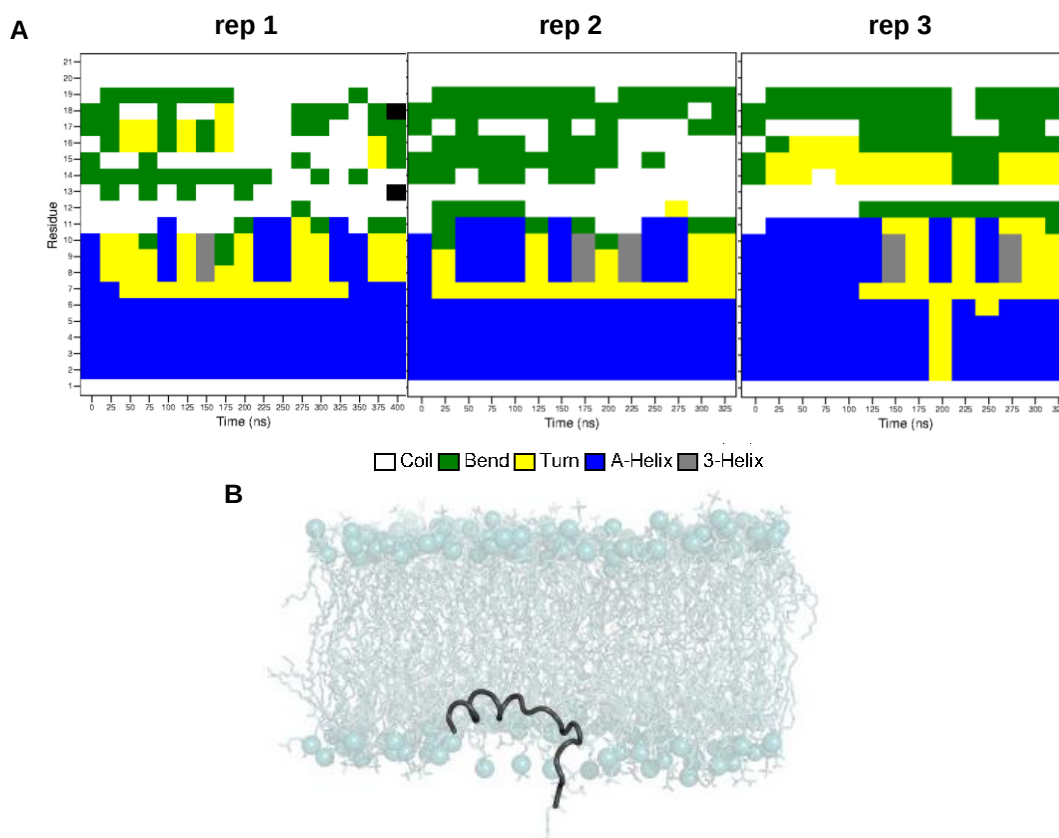


Figure 4.47 | Evaluation of the secondary structure of FP1. (A) Content of each secondary structure element per residue during the simulation time. (B) Snapshot illustrating the FP1 conformation.

When analyzing the secondary structure of the three FP1 replicates (Figure 4.47A), it is possible to observe a similar behavior. In all replicas the FP1 can be roughly divided into two sections according to the secondary structure elements: the 1-10 region, mainly composed of α -helices and turns, and the 11-22 region characterized by bends, turns and coil. To better examine this matter, the conformational features are illustrated in Figure 4.47B. The 1-10 region is positioned beneath the phosphate headgroups of the membrane bilayer while the 11-22 region is partially solvent exposed. Although both regions share a similar content of hydrophobic and charged residues, the 1-10 region position allows to stabilize the hydrophobic residues. In contrast, the region 11-22 localization provides a less stabilizing environment for this type of residues resulting in a more flexible secondary structure.

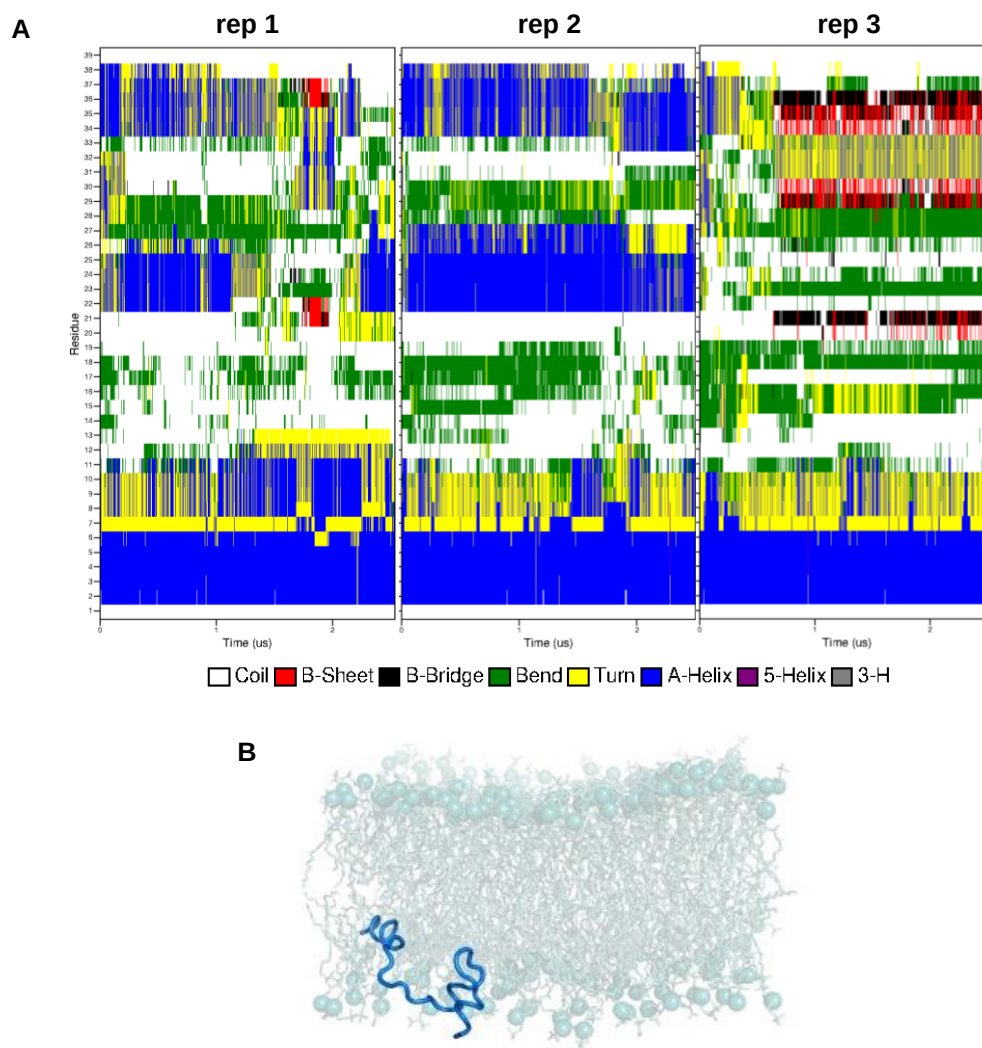


Figure 4.48 | Evaluation of the secondary structure of FP2. (A) Content of each secondary structure element per residue during the simulation time. (B) Snapshot illustrating the FP2 conformation.

FP2 share its first 22 residues with FP1, revealing the same structural behavior described above. The secondary structure of the remaining FP2 residues (23-40) varies between replicates. In replicas 1 and 2, this region adopts α -helix and bends features while in replica 3 bends, turns and β -strands are gained (Figure 4.48A). As observed in the secondary structure illustration of FP2 (Figure 4.48B), the 23-40 region is located at the phosphate head groups level and throughout the simulation time it switches from being totally exposed to the solvent to being partially inserted in the membrane bilayer, which reflects the structural changes observed.

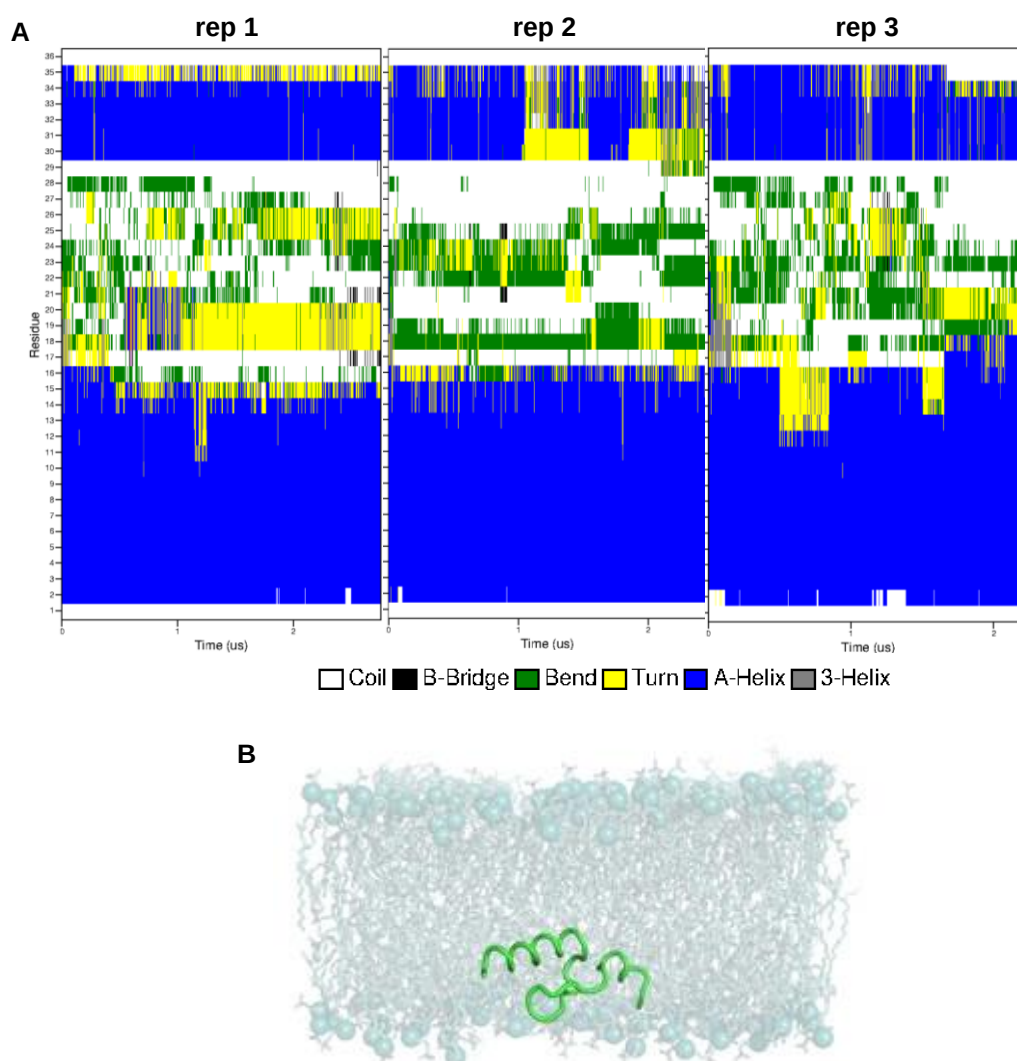


Figure 4.49 | Evaluation of the secondary structure of FP3. (A) Content of each secondary structure element per residue during the simulation time. (B) Snapshot illustrating the FP3 conformation.

The FP3 secondary structure determined is identical for all the three replicas (Figure 4.49A) enabling to divide the peptide into three main regions: 1-16, 17-29 and 30-39. The 1-16 and the 30-39 regions present a secondary structure dominated by α -helices connected by the 17-29 region which is composed of bends, turns and coil. When examining Figure 4.49B and the simulation trajectory, it is possible to observe that the more flexible region (17-29) is either partially in contact with the solvent or shielded from it by the phosphate head-groups. These changes in the FP3 placement in the membrane bilayer may explain the structural features observed.

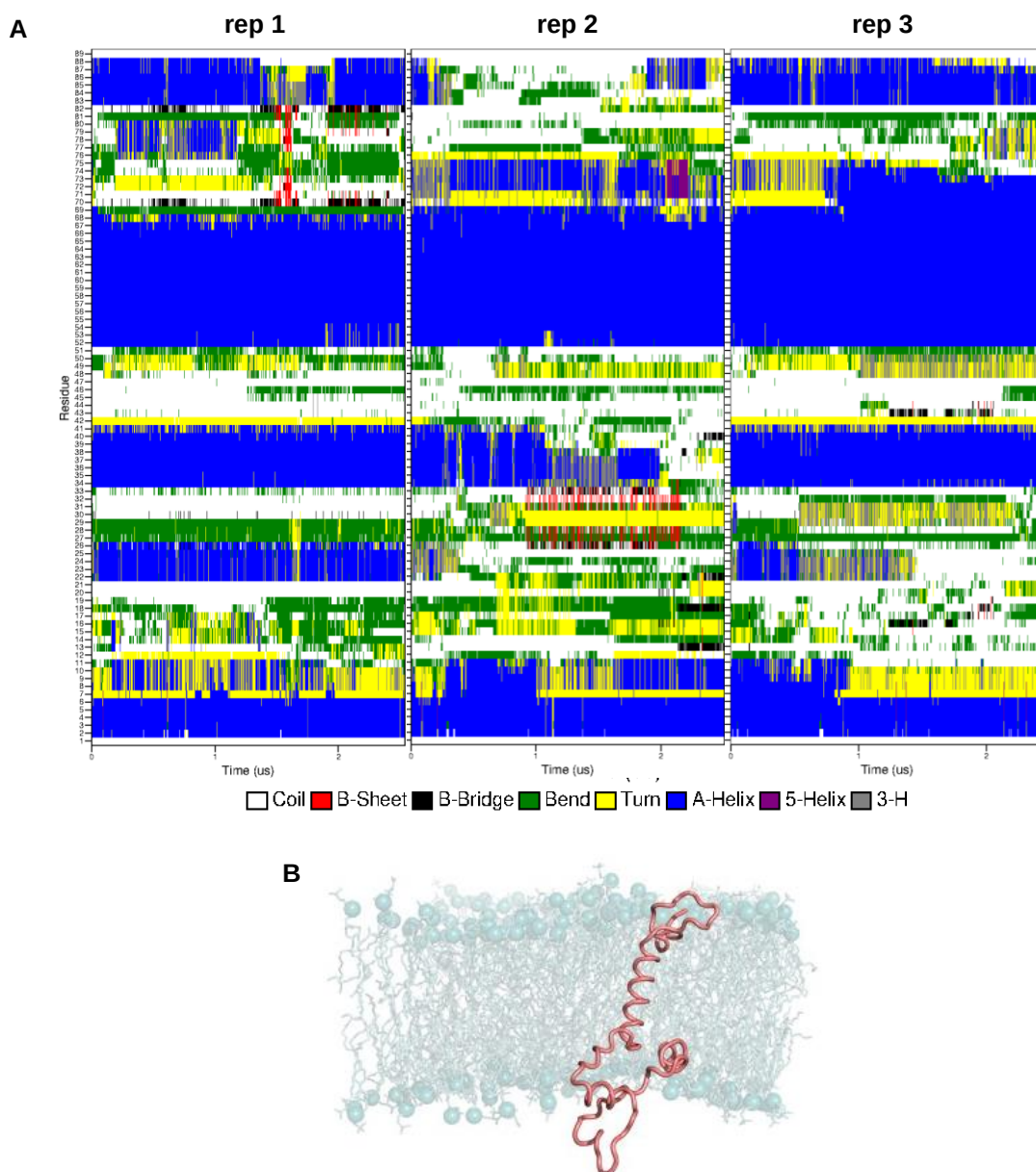


Figure 4.50 | Evaluation of the secondary structure of FM. (A) Content of each secondary structure element per residue during the simulation time. (B) Snapshot illustrating the FM conformation.

FM is composed of the FP2, FP3 and the linking residues in between. The resulting secondary structure (Figure 4.50A) presents identical features to its building units: the 1-40 region resembles the FP2 behavior; the 41-51 region is dominated by flexible secondary structure elements, including turns, bends and coil; the 52-90 region has the same structural profile as FP3. As shown in Figure 4.50B, the FP2 and linking regions are placed between the phosphate headgroups level and the solvent which may contribute to the high conformational flexibility. On the other hand, the FP3 region seems to be stabilized as a nearly transmembrane α -helix with only its flexible residues partially in contact with the outside of the membrane bilayer.

Insertion of the FPs in the membrane bilayer

The FP region of a fusion protein must insert into the host membrane to enable the fusion between viral and host membranes. Therefore, the insertion ability of the SARS-CoV-2 putative FPs was evaluated in order to better characterize the candidates.

The first approach aimed to widely describe the FPs insertion depth in the membrane bilayer and consisted in computing the distribution of the density of key elements along the Z-axis, including the peptides, the phosphate head-groups and water molecules (Figure 4.51). The average of the density distribution was determined for the last 200 ns of FP1 simulations and the last 1 μ s of FP2, FP3 and FM simulations.

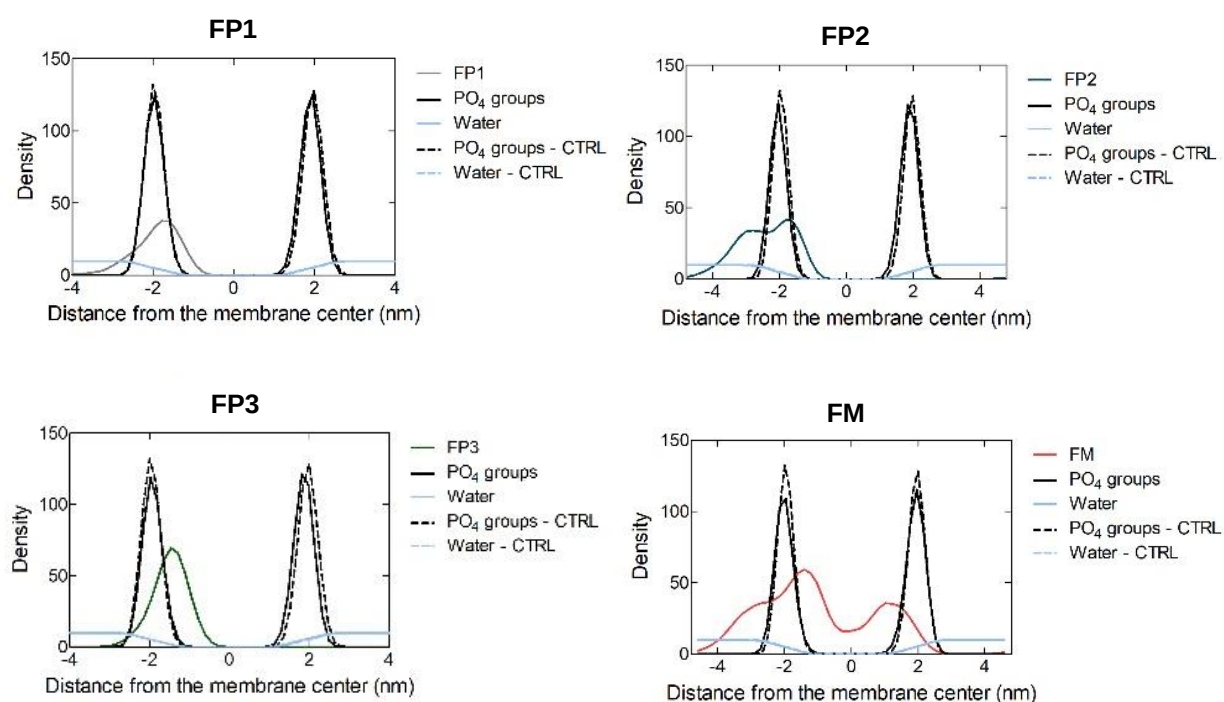


Figure 4.51 | Density distribution along the Z-axis of the putative FPs, the phosphate groups and water molecules.

By analyzing the figure above, it is possible to identify different insertion depths among the four FPs, especially when comparing the smaller peptides with the FM. FP1, FP2 and FP3 remain at the phosphate head group level of the lower leaflet although in slightly distinct manners: FP1 is distributed at the same level as the phosphate groups; FP2 is partially positioned outside the membrane bilayer being exposed to the solvent and; FP3 is almost shielded from the solvent through insertion below the phosphate groups. In contrast, FM is distributed along the membrane bilayer reaching both the top and bottom leaflets. FM is also partially exposed to the solvent possibly due to the region shared with the FP2. Additionally, the density distribution of phosphate groups and water of a control membrane (without any peptide) revealed that no major changes were induced by the FPs presence in the systems.

To evaluate the insertion extent in detail, the distance of each amino acid residue and the average position of the phospholipid head groups of the lower leaflet was determined for all the four FPs. The values presented in Figure 4.52 correspond to the average distance

calculated over the last 200 ns of simulation for FP1 and 1 μ s for the remaining FPs. The standard deviation among the three replicas of each system is also presented.

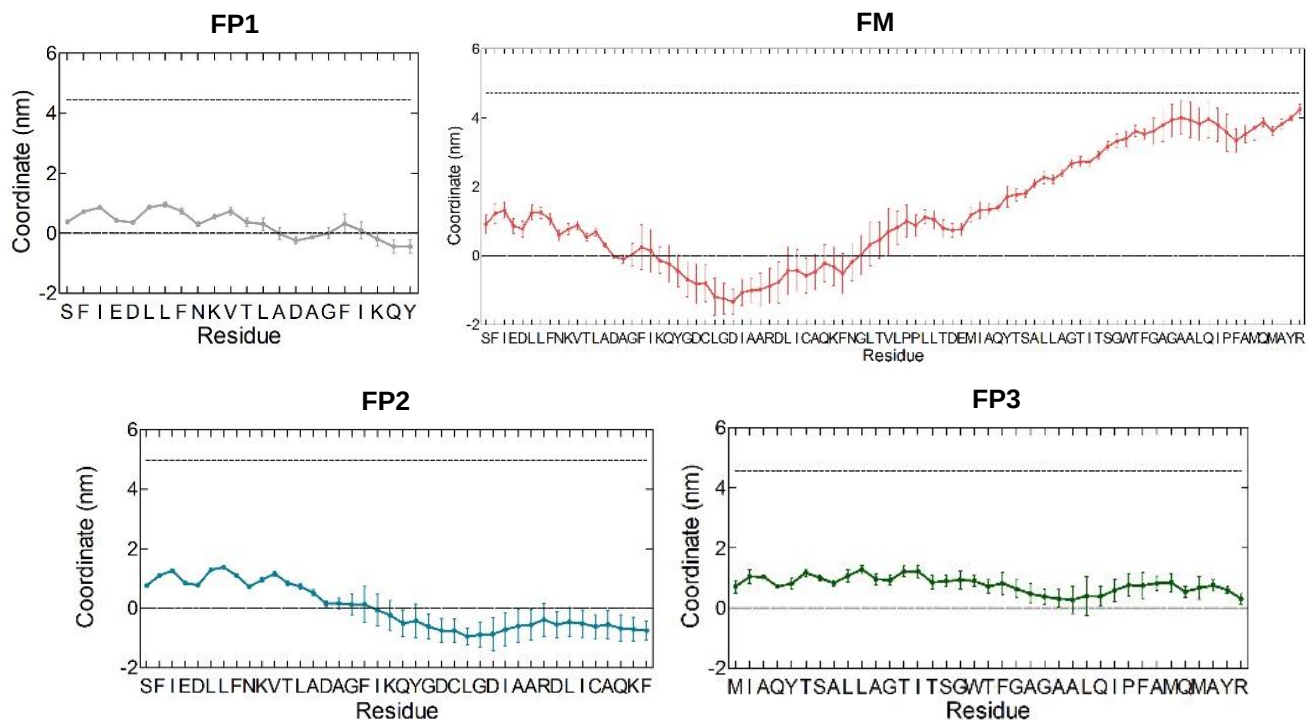


Figure 4.52 | Average insertion of the FPs per residue in the membrane bilayer considering the phosphate groups of the lower leaflet as the reference. (Dashed lines) represent the phosphate groups of both lower and upper leaflets. (Grey) FP1. (Blue) FP2. (Green) FP3. (Pink) FM. The standard deviation between the three replicas is presented in the form of error bars.

The residues shared among FP1, FP2 and FM exhibited similar behaviors and can be divided into three main regions according to the insertion depth: SF-TL, AD-FI and KQ-KF.

In the simulations performed with FP1, FP2 and FM, the SF-TL region was stably inserted in the lower leaflet adopting a nearly parallel position to the phospholipids headgroups. This region insertion in the membrane might be led by the strong contribution of hydrophobic residues (including two Phe) which could balance the presence of two negatively charged residues (Gln and Asp). Moreover, the LLF motif conserved among SARS-CoVs was identified, by mutagenesis assays in previous studies, to be a key structural element for membrane insertion and consequently fusogenicity.

The position of the residues comprised in AD-QY region fluctuates between the phospholipids headgroups levels and the outside of the membrane, as implied by the error bars. This region is composed of small amino acids such as Ala and Gly which confer flexibility, followed by a partially exposed Asp suggested in the literature to be a conserved Ca^{2+} binding site. Phe833 seems to be the insertion point of this region in the membrane bilayer, possibly due to its hydrophobic nature.

KQ-KF region is almost fully exposed to the solvent during the simulations of either FP2 or FM. Two positively charged Lys flank the region becoming potential interaction points with the phosphate groups of the lower leaflet. A disulfide bond established by the two Cys is also a conserved structural feature among SARS-CoVs which might contribute to the stabilization of

this long loop outside the membrane bilayer. In addition, KQ-KF contains three Asp and two Gln, the latter being a polar residue and therefore preferably interacting with the water molecules outside the membrane bilayer.

When analyzing the linker region of the FM which connects the sequences shared with FP2 and FP3 (NG-TE region), it is possible to observe a variability in the residues position between replicas. However, the region is practically fully inserted beneath the phospholipid headgroups of the lower leaflet.

On the other hand, the residues common to FP3 and FM revealed a different insertion profile. In the FP3 simulations, the residues are inside the membrane bilayer parallel to the phosphate head-groups of the lower leaflet while in the FM system the same residues adopt a transmembrane position slightly diagonal to the Z-axis. Moreover, the Gly₈₇₉-Ala-Gly-Ala-Ala₈₈₃ motif presents higher fluctuation in its positions in the FM simulations than in the FP3 ones. This might result from interactions with the phospholipids of the upper leaflet or be induced by the neighbor residues positions.

FPs-membrane bilayer interactions

As shown above, all the FPs are inserted, at least partially, in the membrane bilayer. Therefore, to better understand the insertion process and to identify key amino acid residues, the analysis of the contacts between the peptides and the phospholipids was performed. The POPC and POPS phospholipids were analyzed separately and further divided into “Heads” and “Tails” according to Figure 4.53.

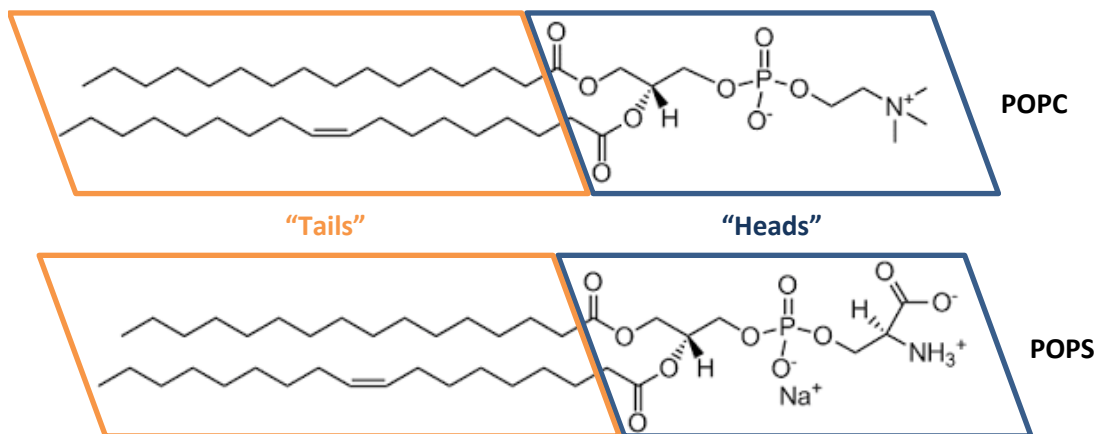


Figure 4.53 | POPC and POPS structural division into “Heads” and “Tails”

The number of contacts between amino acid residues and the lipidic groups in each replica were calculated using *gmx_mindist*. The average number of contacts for each residue as well as the corresponding standard deviation observed among replicas is presented for the four SARS-CoV-2 putative FPs. The specific contacts established as hydrogen bonds between the residues and the “Head” groups of the phospholipids were also determined.

The contacts formed by FP1 residues and “Heads” or “Tails” follows a similar pattern for both POPC and POPS, as shown in Figure 4.54A. The analysis of the same figure enables to identify single residues or motifs with defined preference for either the hydrophobic or hydrophilic part of the phospholipids. For instance, the N-terminal S₈₁₆ residue, the ADAG₈₃₂

segment and the C-terminal motif KQY₈₃₇ presents a higher number of contacts with the “Heads” while the I₈₁₈, the LLF₈₂₃ motif and the V₈₂₆ establish more contacts with the “Tails”. These preferences might be explained by the biochemical properties characteristic of each amino acid, such as hydrophobicity, which influence the positioning in the membrane bilayer during the simulation. The screenshot in Figure 4.54B illustrates the FP1 contacts with the phospholipids, placing the mentioned hydrophobic residues buried in tails rich areas and in contrast the terminals and the ADAG motif locate near head groups.

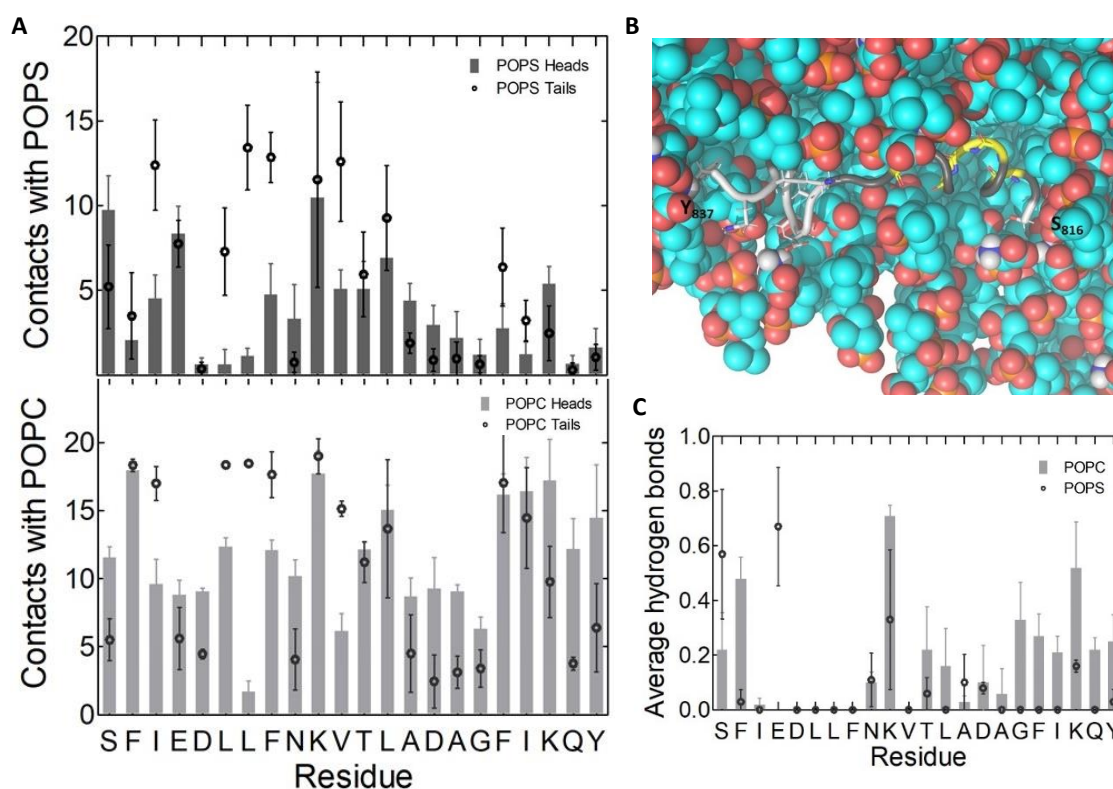


Figure 4.54 | Interactions established between FP1 and the membrane bilayer. (A) contacts with POPC and POPS phospholipids “Heads” and “Tails” groups. (B) Representative snapshot of the FP1 localization within the membrane bilayer. (C) Hydrogen bonds formed between each FP1 residue and the POPC and POPS “Heads” group. Error bars represent the standard deviation between the three replicas.

Additionally, the analysis of Figure 4.54C allows to pinpoint key FP1 residues for the interaction with the phospholipid “Heads”. F₈₁₇, K₈₂₅ and K₈₃₅ form hydrogen bonds mainly with POPC while S₈₁₆ and E₈₁₉ establish almost exclusively bonds with POPS.

When examining Figure 4.55A, the contacts formed by the FP2 residues seem to be evenly distributed among “Heads” and “Tails”, except for some residues, namely, the FI₈₁₈, LLF₈₂₃ and KV₈₂₆ hydrophobic regions which are mainly in contact with the “Tails”. Moreover, the residues D₈₃₀ and R₈₄₇ prefer to contact with the “Heads” maybe due to their charged side chain. Figure 4.55B shows that the residues mostly identified as Tails-interacting are placed in a helix inserted below the phospholipid headgroups. Regarding the hydrogen bonds between FP2 and phospholipids “Heads”, residues E₈₁₉, D₈₃₀, A₈₄₆ and R₈₄₇ play a major role in interacting with POPS while K₈₂₅ and R₈₄₇ are responsible for the bonds with POPC.

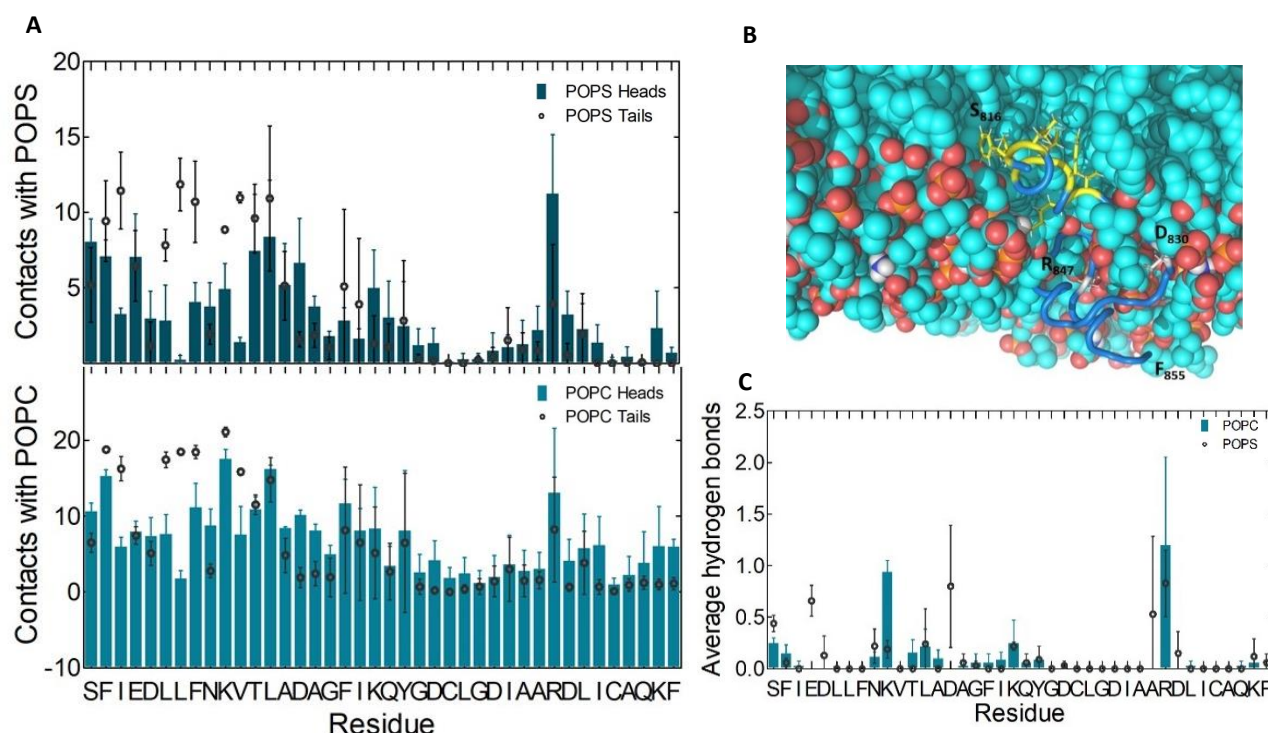


Figure 4.54 | FP2-membrane bilayer interactions. (A) contacts with POPC and POPS phospholipids “Heads” and “Tails” groups. (B) Snapshot representative of the FP2 localization within the membrane bilayer. (C) Hydrogen bonds form between each FP2 residue and the POPC and POPS “Heads” group. Error bars represent the standard deviation between the three replicas.

Since FP3 is more hydrophobic peptide than FP1 and FP2 it is expected to acquire a distinct contacts’ profile, as shown in Figure 4.56A. Although some residues including M₈₆₉, Q₈₇₂, AGAA₈₉₃, Q₈₉₅ and R₉₀₅ formed contacts with the “Heads”, a greater extent of interactions with the “Tails” are established by the segments I₈₇₀-I₈₈₂ and P₈₉₇-A₉₀₃. Gutamines are polar residues known to interact preferentially with water and in this case with the polar “Head” groups. Therefore, these residues might be necessary to assist the peptide transition from water to the lipid environment. AGAA₈₉₃ takes the form of a flexible exposed linker between the two hydrophobic segments I₈₇₀-I₈₈₂ and P₈₉₇-A₉₀₃, allowing them to adjust and insert deeply below the phospholipid head groups, as illustrated in Figure 4.56B. When analyzing the average of hydrogen bonds recorded for FP3 in Figure 4.56C, three residues standout: M₈₆₉, W₈₈₆ and R₉₀₅, among which R₉₀₅ establishes a higher number of hydrogen bonds possibly due to a closer positioning of the arginine side chain and the phosphate group led by Coulomb’s attraction.

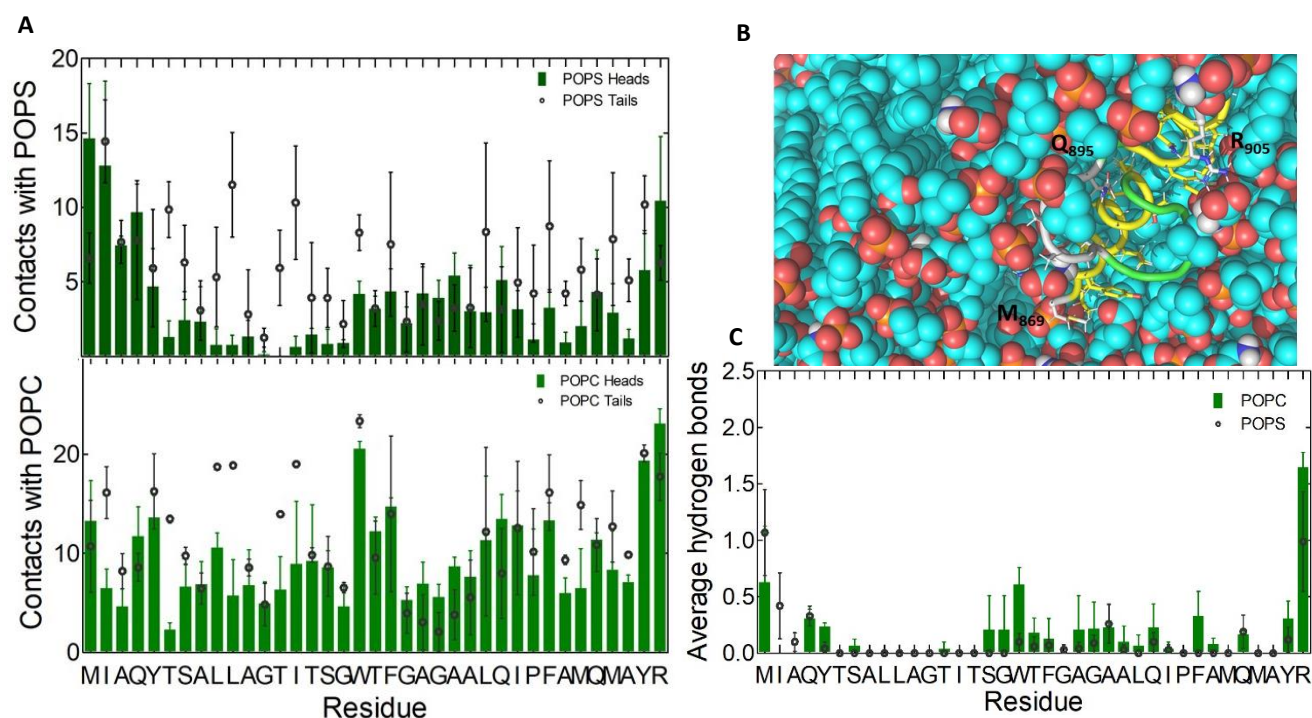


Figure 4.56 | FP3-membrane bilayer interactions. (A) contacts with POPC and POPS phospholipids “Heads” and “Tails” groups. (B) Snapshot representative of the FP3 localization within the membrane bilayer. (C) Hydrogen bonds form between each FP3 residue and the POPC and POPS “Heads” group. Error bars represent the standard deviation between the three replicas.

The FM residues shared with FP2 reveal similar contact patterns where FI, LLF, V and L preferentially interact with the “Tails” while S, N, ADA, K, R and KF are in close contact with the “Heads”. As shown in Figure 4.57A the FM region corresponding to the FP2 sequence is distributed between the phospholipids headgroups level and the outer membrane. Interestingly the GDC-AAR segment is solvent exposed during most of the simulation time. On the other hand, the region in common with FP3 is inserted across the membrane bilayer and thus establishes more contacts with the “Tails”. However, Q₈₉₄ and R₉₀₅ still represent key residues in the contact with the “Heads”. In addition, the LTVLPPLL segment in the linker forms extensive contacts with the “Tails” and the E₈₆₈ residue of the same linker interacts preferably with the “Heads”.

When examining the FM hydrogen bonds in Figure 4.57C, two residues stand out: K₈₂₅ and R₉₀₅. Some other residues in the middle of the peptide also contribute to the formation of hydrogen bonds mainly with POPC “Heads”.

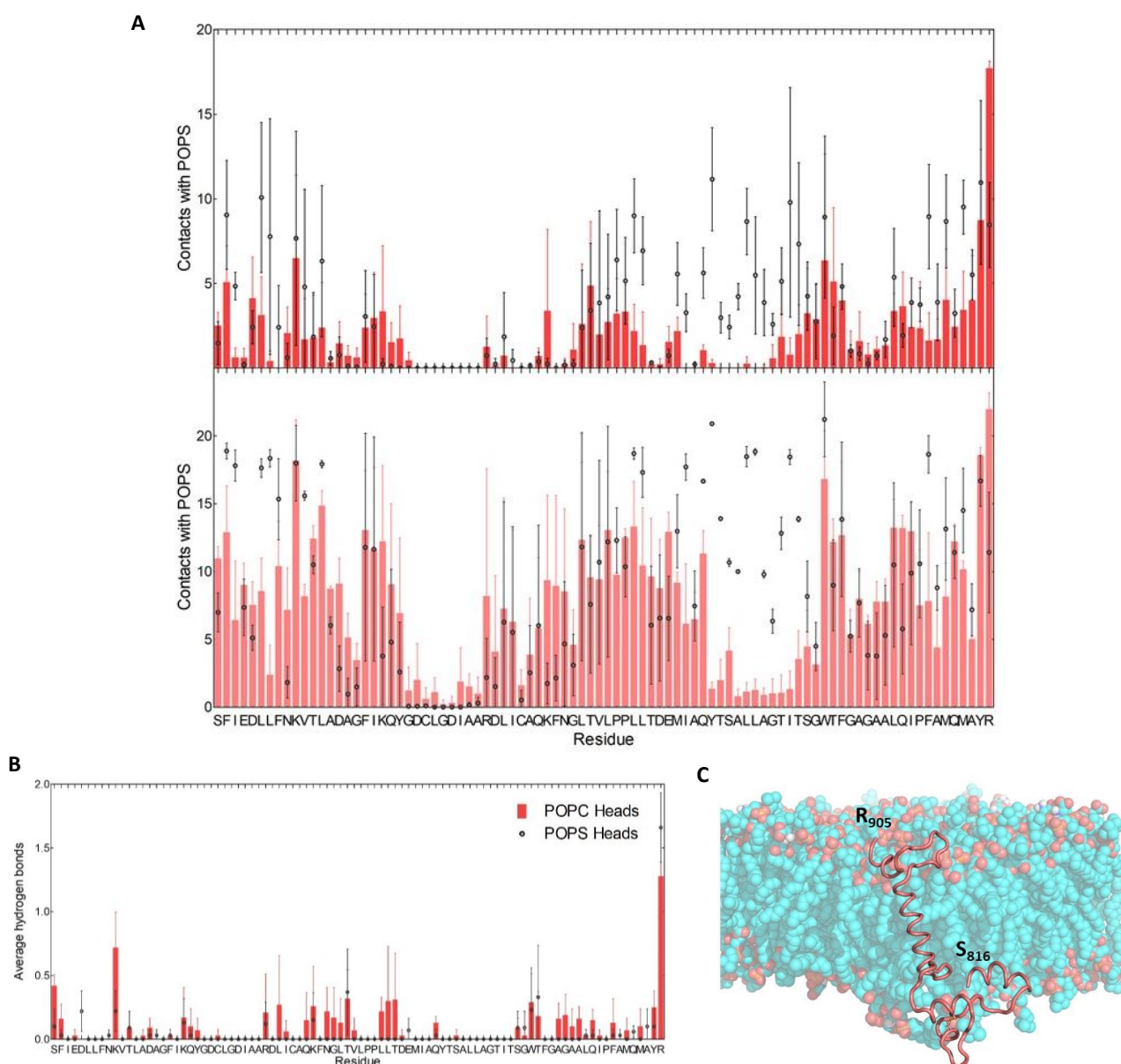


Figure 4.57 | Interactions established between FM and the membrane bilayer. (A) contacts with POPC and POPS phospholipids “Heads” and “Tails” groups. (B) Hydrogen bonds formed between each FM residue and the POPC and POPS “Heads” group. Error bars represent the standard deviation between the three replicas. (C) Representative snapshot of the FM localization within the membrane bilayer.

FPs-induced changes in the membrane proprieties

Once the viral fusion peptides insert into a membrane bilayer, usually a change in the some of the membrane properties occur, including the alteration of the order parameters of the carbons in the lipids tails, a fluctuation on the membrane thickness and even protrusion events may also be induced.

Here, the average order parameter of the POPC:POPS (4:1) membrane bilayer was determined for each carbon atom of both sn1 and sn2 chains in the lipid tails using *gmx order*. Carbon atoms in the sn1 chain are fully saturated while sn2 comprises two unsaturated carbon atoms. The order of a lipid tail increases with the increase in the order parameter calculated.

When comparing the order parameters profile of a pure membrane (Figure 4.58) with the ones obtained for the FPs containing systems, a slight decrease is observed especially for POPC. This phenomenon is more pronounced for FP2 and FM systems and happens at the beginning of both sn1 and sn2 between C5-C9 and C3-C7 respectively. This might be caused by the insertion mechanism of the amino acid residues shared by FP2 and FM whose position oscillates from being just below the phosphate head groups or being partially exposed to the outside of the membrane forcing the lipids to lose their original order profile.

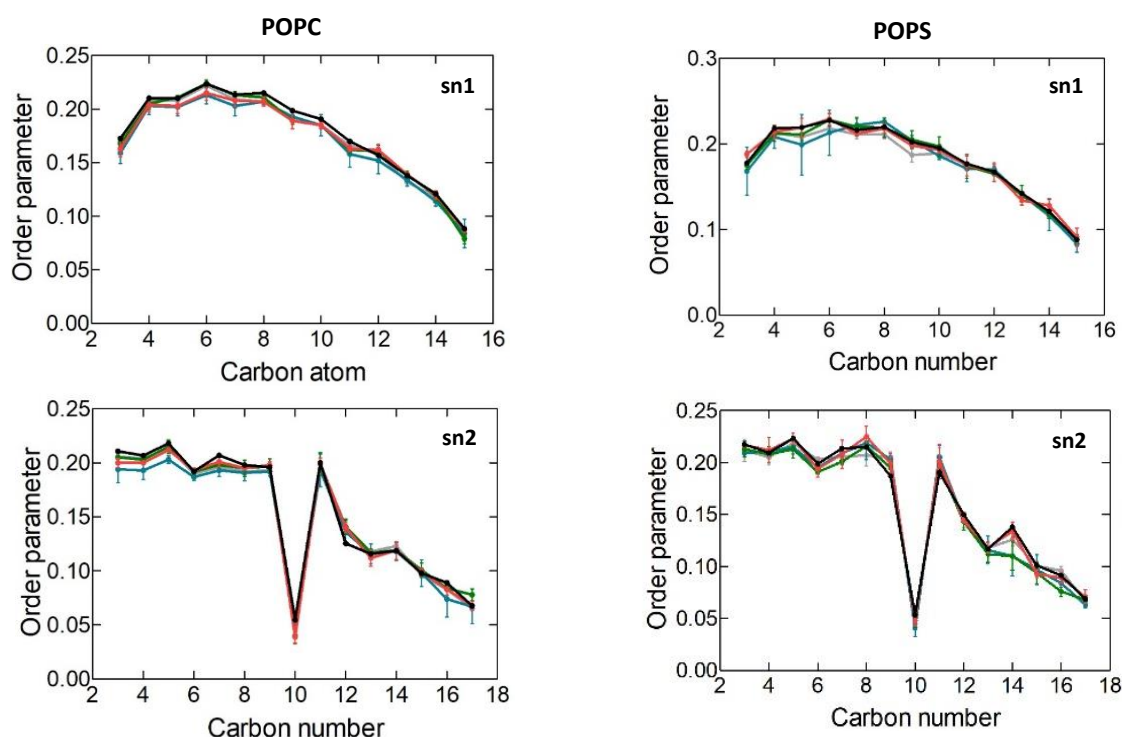


Figure 4.58 | Order parameters determined for the POPC (left) and POPS (right) tails.

Lipid-tail protrusion events may be used as indicatives to measure the potential of a molecule or peptide to trigger a membrane fusion process. These events can be detected when at least one of the carbon atoms in the lipid tail protrudes 0.1 nm or more above the phosphate head group. The comparison of the average number of protrusion events in a pure membrane and in FPs systems (Figure 4.59) allows to clearly identify a disturbing effect of all the peptides. Interestingly, the extent of this effect appears to be slightly greater for FP2, FP3 and FM than in for FP1. A correlation with the results obtained for other properties examined, including the insertion might be established. For instance, the FP1 is the peptide whose insertion is shallower and, therefore, less prone to induce greater lipid-tail protrusion. On the other hand, both FP2 and FP3 insert below the phosphate head groups and revealed to form more contacts with the lipid tails as well as with fixed hydrogen bonds with the heads groups. Although FM comprises FP2 and FP3 residues, its induced protrusion events are of the same order of magnitude as those of the FP2 and FP3. This might be the result of a different insertion and positioning of the FP3 residues in the membrane bilayer. As mentioned before, the FM residues which are shared with the FP3 are inserted in a transmembran manner and this may led to a decrease in the perturbing effect.

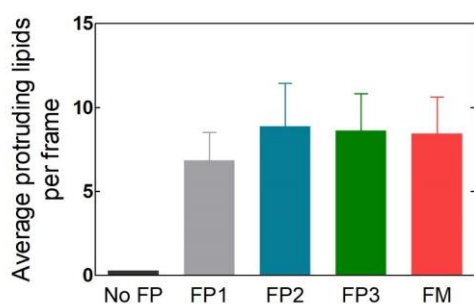


Figure 4.59 | Average protruding lipids per frame induced by each of the FPs.

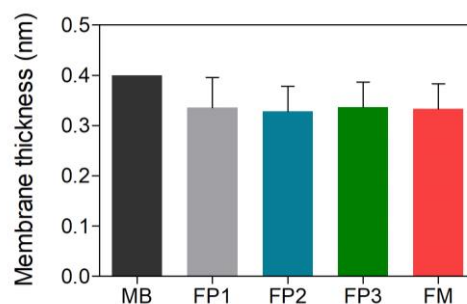


Figure 4.60 | Global membrane thickness determined for each system.

The membrane thickness is thought to be an important property that may change during the membrane fusion process. Therefore, this feature was calculated as the distance between the phosphate groups in the top and bottom membrane layers throughout the simulation. The global membrane thickness of all the FPs-containing systems exhibited in Figure 4.60 did not change significantly. However, FM promotes a visible narrowing of the membrane in its surroundings as illustrated in Figure 4.61. This phenomenon is less pronounced for the systems containing the smaller FPs.

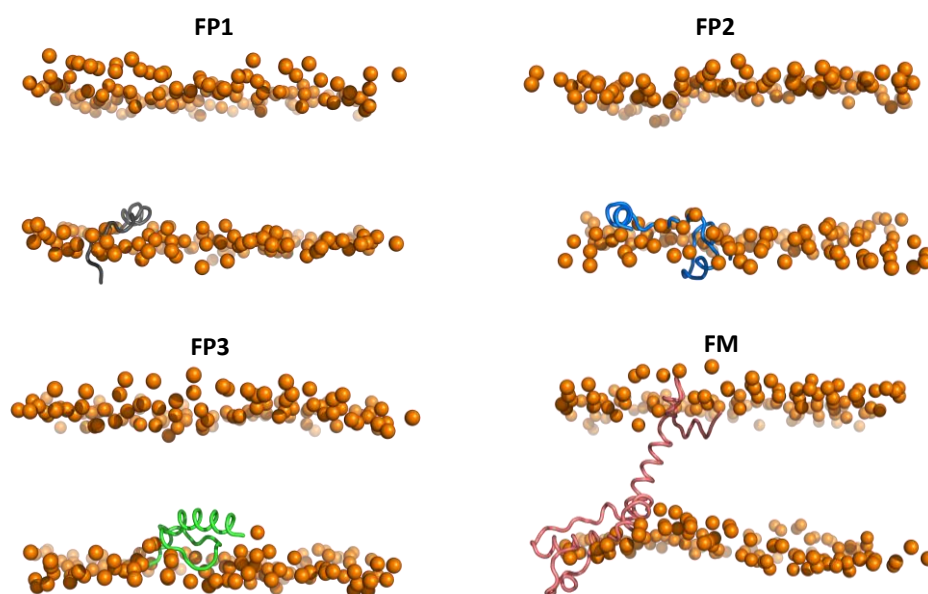


Figure 4.61 | Snapshots of the FPs-induced perturbation in the membrane bilayer. The phosphate headgroups of the top and bottom layers are represented as orange spheres. The putative FPs are presented as cartoon. (Grey) FP1. (Blue) FP2. (Green) FP3. (Pink) FM.

4.3.6. Comparison with previous studies of SARS-CoV-2 putative FPs

As mentioned before, previous studies on SARS-CoV-2 putative FPs gave insights about these peptides' structural features and their effect on membrane model systems. The insertion of FP1 and FP2 in a membrane bilayer determined by MD simulations in this project is in agreement with the NMR experiments in bilayered micelles. Since in both cases the S816-D830 region is more deeply inserted into the membrane bilayer while the D838-F855 region is solvent exposed protruding from the bilayer. The FM amino acid residues shared with FP1 and FP2 exhibited a similar insertion pattern in the MD simulations.

Although the FP3 structure was not yet resolved, its effect on LUVs was already addressed in previous works. According to these, the FP3 induced more leakage on vesicles than the other two small peptides. However, in the experiments performed in this project the FP3 and FP2 exhibited similar leakage effect which was greater than the FP1 effect. Additionally, and in concordance with the literature, the FP2 and FP3 activity was enhanced in the presence of anionic phospholipids.

5. Conclusions and future perspectives

The main objectives of this thesis were to i) shed light into and ii) prevent the SARS-CoV-2 mechanism of infection by a combination of computational and experimental approaches. A more applied approach aimed at preventing the infection, targeting either the RBD-ACE2 interaction or the S2' cleavage and consequent FP exposure. A more fundamental approach aimed at unveiling the molecular determinants of SARS-CoV-2 putative fusion peptides (FPs).

Within the applied research scope, five designed RBD-interacting chimeric proteins (RIChs) were successfully produced in *E. coli* with the predicted fold. Three out of the five RIChs interacted with the SARS-CoV-2 RBD through the designed region, fulfilling one of the objectives of this topic. Moreover, two of the candidates exhibited reasonable affinity constants, although higher than hACE2, strengthening the design protocol quality. However, no significant neutralizing effect was observed, perhaps due to the lower affinity or the possible competition of RIChs and ACE2 for the same RBD region. This may also suggest that the interaction between RBD and ACE2 depends on more than just the interface residues and other considerations must be done for future design of interacting proteins.

Additionally, in the same line of research, an immunogenic construct comprising the S2' cleavage site residues and the highly conserved FP1 region was designed to become a possible broad-spectrum vaccine. After expression in *E. coli*, the construct revealed useful features from the biopharmaceutical point of view, including a high thermostability and the ability to maintain its main secondary structure characteristics for eight days of storage at either 4 or 24 °C even in the absence of any stabilizing compound. Furthermore, this immunogenic construct triggered antibody production in two rabbits confirming its immunogenicity. The resulting antibodies were able to recognize the immunogenic sequence in the spike-protein context not only in its denatured form but also in its folded version with high endpoint titers. Their neutralizing ability was not fully clear, and some hypotheses could be proposed to explain this issue. The antibodies' inability to prevent the S2' cleavage was confirmed. Although no neutralizing antibodies were obtained, important insights were brought to light. For instance, we should cover a larger region within the S2'cleavage site and the FP or target other conserved regions which experience less conformational changes during the infection process.

Regarding the fundamental line of research, four putative FPs were proposed and characterized both *in silico* and experimentally revealing distinct insertion modes and membrane perturbing activities. FP1 was the peptide with lower membrane activity, followed by FP2 and FP3. The FM was the only peptide which inserted in a transmembrane manner and was able to induce fusion between vesicles to the greatest extent. The findings might suggest the requirement of a fusion module to trigger and conduct the membrane fusion process. Additionally, FP2 and FP3 oligomerization should be clarified and this phenomenon must also be evaluated for the FM in order to verify the existence of membrane perturbation dependent on oligomerization.

6. References

1. Rothan, H. A. & Byrareddy, S. N. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *J. Autoimmun.* **109**, 102433 (2020).
2. WHO. *COVID-19 weekly epidemiological update*. World Health Organization (2021).
3. Muralidar, S., Ambi, S. V., Sekaran, S. & Krishnan, U. M. The emergence of COVID-19 as a global pandemic: Understanding the epidemiology, immune response and potential therapeutic targets of SARS-CoV-2. *Biochimie* **179**, 85–100 (2020).
4. Shi, Y. *et al.* COVID-19 infection: the perspectives on immune responses. *Cell Death Differ.* **27**, 1451–1454 (2020).
5. Cui, J., Li, F. & Shi, Z.-L. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* **17**, 181–192 (2019).
6. Gorkhali, R. *et al.* Structure and Function of Major SARS-CoV-2 and SARS-CoV Proteins. *Bioinform. Biol. Insights* **15**, (2021).
7. Adapted from ‘Human Coronavirus Structure’, by BioRender.com (2022). Retrieved from <https://app.biorender.com/biorender-templates>.
8. Arya, R. *et al.* Structural insights into SARS-CoV-2 proteins. *J. Mol. Biol.* **433**, 166725 (2021).
9. Walls, A. C. *et al.* Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **181**, 281–292.e6 (2020).
10. Walls, A. C. *et al.* Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proc. Natl. Acad. Sci.* **114**, 11157 LP – 11162 (2017).
11. Adapted from ‘An In-depth Look into the Structure of the SARS-CoV2 Spike Glycoprotein’, by BioRender.com (2022). Retrieved from <https://app.biorender.com/biorender-templates>.
12. Zhao, X., Chen, H. & Wang, H. Glycans of SARS-CoV-2 Spike Protein in Virus Infection and Antibody Production. *Front. Mol. Biosci.* **8**, 1–10 (2021).
13. Hernandez, E. R. *et al.* The Role of the SARS-CoV-2 S-Protein Glycosylation in the Interaction of SARS-CoV-2/ACE2 and Immunological Responses. *Viral Immunol.* **34**, 165–173 (2021).
14. Walls, A. C. *et al.* Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **181**, 281–292.e6 (2020).
15. Bestle, D. *et al.* TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. *Life Sci. Alliance* **3**, 1–14 (2020).
16. Park, J.-E. *et al.* Proteolytic processing of Middle East respiratory syndrome coronavirus spikes expands virus tropism. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 12262–12267 (2016).
17. Shulla, A. *et al.* A transmembrane serine protease is linked to the severe acute respiratory syndrome coronavirus receptor and activates virus entry. *J. Virol.* **85**, 873–882 (2011).
18. Samavati, L. & Uhal, B. D. ACE2, Much More Than Just a Receptor for SARS-COV-2. *Front. Cell. Infect. Microbiol.* **10**, 1–9 (2020).
19. Lan, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215–220 (2020).
20. Shang, J. *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581**, 221–224 (2020).
21. Wu, K., Peng, G., Wilken, M., Geraghty, R. J. & Li, F. Mechanisms of host receptor adaptation by severe acute respiratory syndrome coronavirus. *J. Biol. Chem.* **287**, 8904–8911 (2012).
22. Wan, Y., Shang, J., Graham, R., Baric, R. S. & Li, F. Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS

- Coronavirus. *J. Virol.* **94**, (2020).
23. Benton, D. J. *et al.* Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature* **588**, 327–330 (2020).
 24. McCallum, M., Walls, A. C., Bowen, J. E., Corti, D. & Veasler, D. Structure-guided covalent stabilization of coronavirus spike glycoprotein trimers in the closed conformation. *Nat. Struct. Mol. Biol.* **27**, 942–949 (2020).
 25. Yuan, Y. *et al.* Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nat. Commun.* **8**, 15092 (2017).
 26. Gui, M. *et al.* Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. *Cell Res.* **27**, 119–129 (2017).
 27. Harrison, S. C. Viral membrane fusion. *Nat. Struct. Mol. Biol.* **15**, 690–698 (2008).
 28. Gregory, S. M. *et al.* Structure and function of the complete internal fusion loop from Ebolavirus glycoprotein 2. *Proc. Natl. Acad. Sci.* **108**, 11211 LP – 11216 (2011).
 29. Basso, L. G. M., Zeraik, A. E., Felizatti, A. P. & Costa-Filho, A. J. Membranotropic and biological activities of the membrane fusion peptides from SARS-CoV spike glycoprotein: The importance of the complete internal fusion peptide domain. *Biochim. Biophys. Acta. Biomembr.* **1863**, 183697 (2021).
 30. Lozada, C., Barlow, T. M. A., Gonzalez, S., Lubin-Germain, N. & Ballet, S. Identification and Characteristics of Fusion Peptides Derived From Enveloped Viruses . *Frontiers in Chemistry* **9**, (2021).
 31. Pattnaik, G. P., Meher, G. & Chakraborty, H. Exploring the Mechanism of Viral Peptide-Induced Membrane Fusion. *Adv. Exp. Med. Biol.* **1112**, 69–78 (2018).
 32. Epand, R. M. Fusion peptides and the mechanism of viral fusion. *Biochim. Biophys. Acta* **1614**, 116–121 (2003).
 33. Koppiseti, R. K., Fulcher, Y. G. & Van Doren, S. R. Fusion Peptide of SARS-CoV-2 Spike Rearranges into a Wedge Inserted in Bilayered Micelles. *J. Am. Chem. Soc.* **143**, 13205–13211 (2021).
 34. Lai, A. L. & Freed, J. H. SARS-CoV-2 Fusion Peptide has a Greater Membrane Perturbing Effect than SARS-CoV with Highly Specific Dependence on Ca²⁺. *J. Mol. Biol.* 166946 (2021). doi:<https://doi.org/10.1016/j.jmb.2021.166946>
 35. Gorgun, D., Lihan, M., Kapoor, K. & Tajkhorshid, E. Binding mode of SARS-CoV-2 fusion peptide to human cellular membrane. *Biophys. J.* (2021). doi:<https://doi.org/10.1016/j.bpj.2021.02.041>
 36. Khelashvili, G., Plante, A., Doktorova, M. & Weinstein, H. Ca²⁺-dependent mechanism of membrane insertion and destabilization by the SARS-CoV-2 fusion peptide. *Biophys. J.* **120**, 1105–1119 (2021).
 37. Hartenian, E. *et al.* The molecular virology of coronaviruses. *J. Biol. Chem.* **295**, 12910–12934 (2020).
 38. Tsang, H. F. *et al.* An update on COVID-19 pandemic: the epidemiology, pathogenesis, prevention and treatment strategies. *Expert Rev. Anti. Infect. Ther.* **19**, 877–888 (2021).
 39. Han, F. *et al.* Current treatment strategies for COVID-19 (Review). *Mol. Med. Rep.* **24**, 1–12 (2021).
 40. Jayk Bernal, A. *et al.* Molnupiravir for Oral Treatment of Covid-19 in Nonhospitalized Patients. *N. Engl. J. Med.* 1–12 (2021). doi:10.1056/nejmoa2116044
 41. Pfizer Inc. Pfizer’s Novel COVID-19 Oral Antiviral Treatment Candidate Reduced Risk of Hospitalization or Death by 89% in Interim Analysis of Phase 2/3 EPIC-HR Study | Pfizer. *Pfizer Website* (2021).
 42. Kyriakidis, N. C., López-Cortés, A., González, E. V., Grimaldos, A. B. & Prado, E. O. SARS-CoV-2 vaccines strategies: a comprehensive review of phase 3 candidates. *npj Vaccines* **6**, (2021).
 43. Guo, Y. *et al.* A SARS-CoV-2 neutralizing antibody with extensive Spike binding coverage

- and modified for optimal therapeutic outcomes. *Nat. Commun.* **12**, 2623 (2021).
44. Shrestha, L. B., Tedla, N. & Bull, R. A. Broadly-Neutralizing Antibodies Against Emerging SARS-CoV-2 Variants. *Front. Immunol.* **12**, 752003 (2021).
 45. Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280.e8 (2020).
 46. Wong, S. K., Li, W., Moore, M. J., Choe, H. & Farzan, M. A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2. *J. Biol. Chem.* **279**, 3197–3201 (2004).
 47. Monteil, V. *et al.* Inhibition of SARS-CoV-2 Infections in Engineered Human Tissues Using Clinical-Grade Soluble Human ACE2. *Cell* **181**, 905-913.e7 (2020).
 48. Riaño-Umbarila, L. *et al.* Generation of a Broadly Cross-Neutralizing Antibody Fragment against Several Mexican Scorpion Venoms. *Toxins* **11**, (2019).
 49. Cao, L. *et al.* De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. *Science* **370**, 426–431 (2020).
 50. Sitthiyotha, T. & Chunsriviro, S. Computational design of SARS-CoV-2 peptide binders with better predicted binding affinities than human ACE2 receptor. *Sci. Rep.* **11**, 15650 (2021).
 51. Basit, A. *et al.* Designing Short Peptides to Block the Interaction of SARS-CoV-2 and Human ACE2 for COVID-19 Therapeutics. *Front. Pharmacol.* **12**, 731828 (2021).
 52. Kinsella, C. M. *et al.* Preparedness needs research: How fundamental science and international collaboration accelerated the response to COVID-19. *PLoS Pathog.* **16**, 1–12 (2020).
 53. Silva, D., Correia, B. E. & Procko, E. Chapter 17 Motif-Driven Design of Protein – Protein Interfaces. **1414**, (2016).
 54. Alford, R. F. *et al.* The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017).
 55. Poh, C. M. *et al.* Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralising antibodies in COVID-19 patients. *Nat. Commun.* **11**, (2020).
 56. Shmulevitz, M., Epand, R. F., Epand, R. M. & Duncan, R. Structural and functional properties of an unusual internal fusion peptide in a nonenveloped virus membrane fusion protein. *J. Virol.* **78**, 2808–2818 (2004).
 57. Apellániz, B., Huarte, N., Largo, E. & Nieva, J. L. The three lives of viral fusion peptides. *Chem. Phys. Lipids* **181**, 40–55 (2014).
 58. Van Gunsteren, W. F. *et al.* Biomolecular modeling: Goals, problems, perspectives. *Angew. Chemie - Int. Ed.* **45**, 4064–4092 (2006).
 59. Valverde, J. R. Molecular Modelling: Principles and Applications. *Briefings in Bioinformatics* **2**, 199–200 (2001).
 60. Kumar, H. & Maiti, P. K. Introduction to Molecular Dynamics Simulation. **23**, 161–197 (2011).
 61. Keith, J. A. *et al.* Combining Machine Learning and Computational Chemistry for Predictive Insights Into Chemical Systems. *Chem. Rev.* **121**, 9816–9872 (2021).
 62. Nauš, J. Becker, O.M., McKerell, A.D., Jr., Roux, B., Watanabe, M.: Computational Biochemistry and Biophysics. *Photosynthetica* **39**, 34 (2001).
 63. Wu, E. L. *et al.* CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *J. Comput. Chem.* **35**, 1997–2004 (2014).
 64. Jo, S., Kim, T. & Im, W. Automated builder and database of protein/membrane complexes for molecular dynamics simulations. *PLoS One* **2**, e880 (2007).
 65. Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: Resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, (2012).
 66. Abraham, M. J. *et al.* Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).

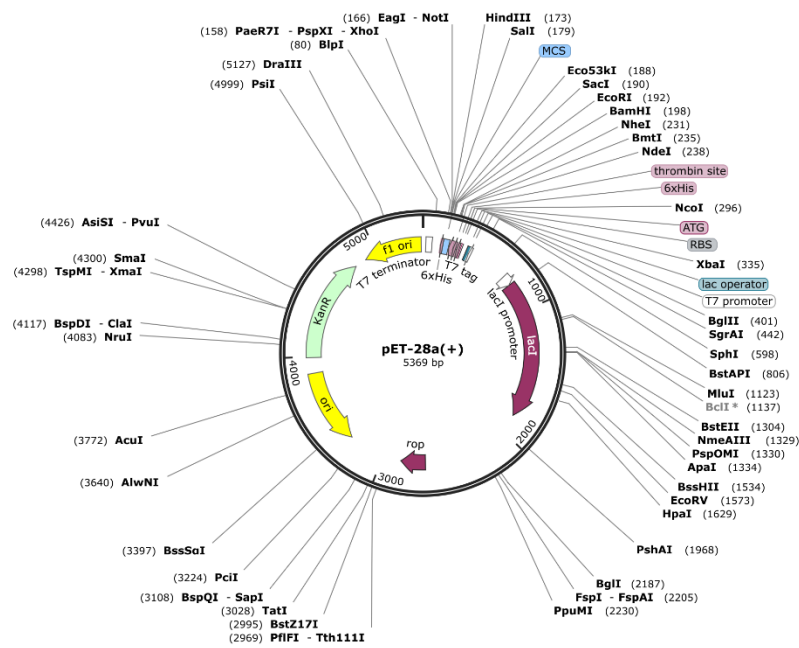
67. Huang, J. *et al.* CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Biophys. J.* **112**, 175a-176a (2017).
68. Boonstra, S., Onck, P. R. & Giessen, E. van der. CHARMM TIP3P Water Model Suppresses Peptide Folding by Solvating the Unfolded State. *J. Phys. Chem. B* **120**, 3692–3698 (2016).
69. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
70. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **81**, 511–519 (1984).
71. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695–1697 (1985).
72. Petersen, H. G. Accuracy and efficiency of the particle mesh Ewald method. *J. Chem. Phys.* **103**, 3668–3679 (1995).
73. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
74. Li, C. *et al.* FastCloning: A highly simplified, purification-free, sequence- and ligation-independent PCR cloning method. *BMC Biotechnol.* **11**, (2011).
75. Kruger, N. The Bradford Method For Protein Quantitation. in 15–21 (2002). doi:10.1385/1-59259-169-8:15
76. Rodger, A. & Marshall, D. Beginners guide to circular dichroism. *Biochem. (Lond)*. **43**, 58–64 (2021).
77. Kelly, S. M., Jess, T. J. & Price, N. C. How to study proteins by circular dichroism. *Biochim. Biophys. Acta - Proteins Proteomics* **1751**, 119–139 (2005).
78. Micsonai, A. *et al.* BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic Acids Res.* **46**, W315–W322 (2018).
79. Wiedemann, C., Bellstedt, P. & Görlach, M. CAPITO - A web server-based analysis and plotting tool for circular dichroism data. *Bioinformatics* **29**, 1750–1757 (2013).
80. Stetefeld, J., McKenna, S. A. & Patel, T. R. Dynamic light scattering: a practical guide and applications in biomedical sciences. *Biophys. Rev.* **8**, 409–427 (2016).
81. Bhattacharjee, S. DLS and zeta potential - What they are and what they are not? *J. Control. Release* **235**, 337–351 (2016).
82. Falke, S. & Betzel, C. *Dynamic Light Scattering (DLS)*. (Springer International Publishing, 2019). doi:10.1007/978-3-030-28247-9_6
83. Provencher, S. W. CONTIN: A general purpose constrained regularization program for inverting noisy linear algebraic and integral equations. *Comput. Phys. Commun.* **27**, 229–242 (1982).
84. Shang, J. *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581**, 221–224 (2020).
85. Chattopadhyay, G. & Varadarajan, R. Facile measurement of protein stability and folding kinetics using a nano differential scanning fluorimeter. *Protein Sci.* **28**, 1127–1134 (2019).
86. Real-Hohn, A., Groznica, M., Löffler, N., Blaas, D. & Kowalski, H. nanoDSF: In vitro Label-Free Method to Monitor Picornavirus Uncoating and Test Compounds Affecting Particle Stability. *Front. Microbiol.* **11**, 1–12 (2020).
87. Alenquer, M. *et al.* Signatures in SARS-CoV-2 spike protein conferring escape to neutralizing antibodies. *PLoS Pathog.* **17**, 1–25 (2021).
88. Lakowicz, J. R. *Principles of fluorescence spectroscopy. Principles of Fluorescence Spectroscopy* (2006). doi:10.1007/978-0-387-46312-4
89. Lis Thomas, T. & Mishra, A. K. ANS fluorescence as a tool to monitor cross-linking polymerization of acrylamide. *Eur. Polym. J.* **38**, 1805–1810 (2002).
90. Lee, C. *et al.* Crystal structure of the flagellar chaperone FliS from *Bacillus cereus* and an

- invariant proline critical for FliS dimerization and flagellin recognition. *Biochem. Biophys. Res. Commun.* **487**, 381–387 (2017).
91. Madu, I. G., Roth, S. L., Belouzard, S. & Whittaker, G. R. Characterization of a Highly Conserved Domain within the Severe Acute Respiratory Syndrome Coronavirus Spike Protein S2 Domain with Characteristics of a Viral Fusion Peptide. *J. Virol.* **83**, 7411–7421 (2009).
 92. Borkotoky, S., Dey, D. & Banerjee, M. Computational Insight into the Mechanism of SARS-CoV-2 Membrane Fusion. *J. Chem. Inf. Model.* **61**, 423–431 (2021).
 93. Lai, A. L., Millet, J. K., Daniel, S., Freed, J. H. & Whittaker, G. R. The SARS-CoV Fusion Peptide Forms an Extended Bipartite Fusion Platform that Perturbs Membrane Order in a Calcium-Dependent Manner. *J. Mol. Biol.* **429**, 3875–3892 (2017).
 94. Madu, I. G., Roth, S. L., Belouzard, S. & Whittaker, G. R. Characterization of a highly conserved domain within the severe acute respiratory syndrome coronavirus spike protein S2 domain with characteristics of a viral fusion peptide. *J. Virol.* **83**, 7411–7421 (2009).
 95. Meher, G., Bhattacharjya, S. & Chakraborty, H. Membrane Cholesterol Modulates Oligomeric Status and Peptide-Membrane Interaction of Severe Acute Respiratory Syndrome Coronavirus Fusion Peptide. *J. Phys. Chem. B* **123**, 10654–10662 (2019).
 96. Liu, K. N. & Boxer, S. G. Target Membrane Cholesterol Modulates Single Influenza Virus Membrane Fusion Efficiency but Not Rate. *Biophys. J.* **118**, 2426–2433 (2020).
 97. Pattnaik, G. P. & Chakraborty, H. Cholesterol alters the inhibitory efficiency of peptide-based membrane fusion inhibitor. *Biochim. Biophys. Acta. Biomembr.* **1861**, 183056 (2019).

7. Annex

7.1. Plasmids

A



B

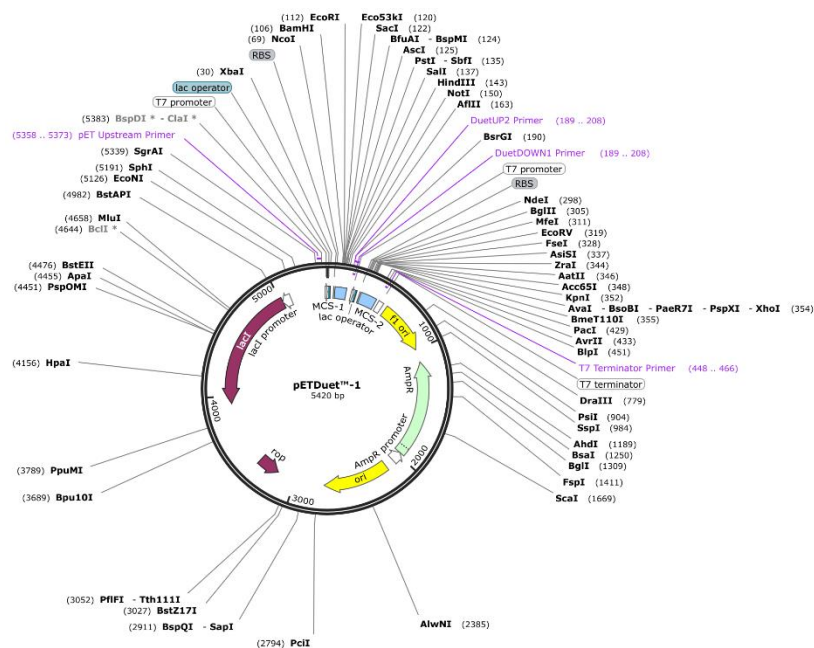


Figure A1 | Maps of the expression plasmid. (A) pET28a(+) used for RIChs, scaffolds 1 and 5, and TSSFP expression. (B) pETDuet for FM expression. Created with SnapGene®

7.2. FM expression vector produced by FastCloning method

FastCloning was introduced by *Li. et al*⁷⁴ in 2011 as a simple, versatile and effective method to perform several molecular biology tasks, such as: molecular cloning or inserting any DNA fragment into a plasmid vector.

As schematized in Figure A2, this is a ligase-independent technique in which the insert and the vector ends have around 16-base overlapping and only three main steps are needed to produce the desire plasmid. First, both vector and insert are amplified by PCR with a high-fidelity DNA polymerase. The resulting PCR product is incubated with *DpnI* to remove the DNA templates used in PCR reactions. Finally, the mixture is transformed into *E.coli* competent cells aiming to obtain the suitable clones.

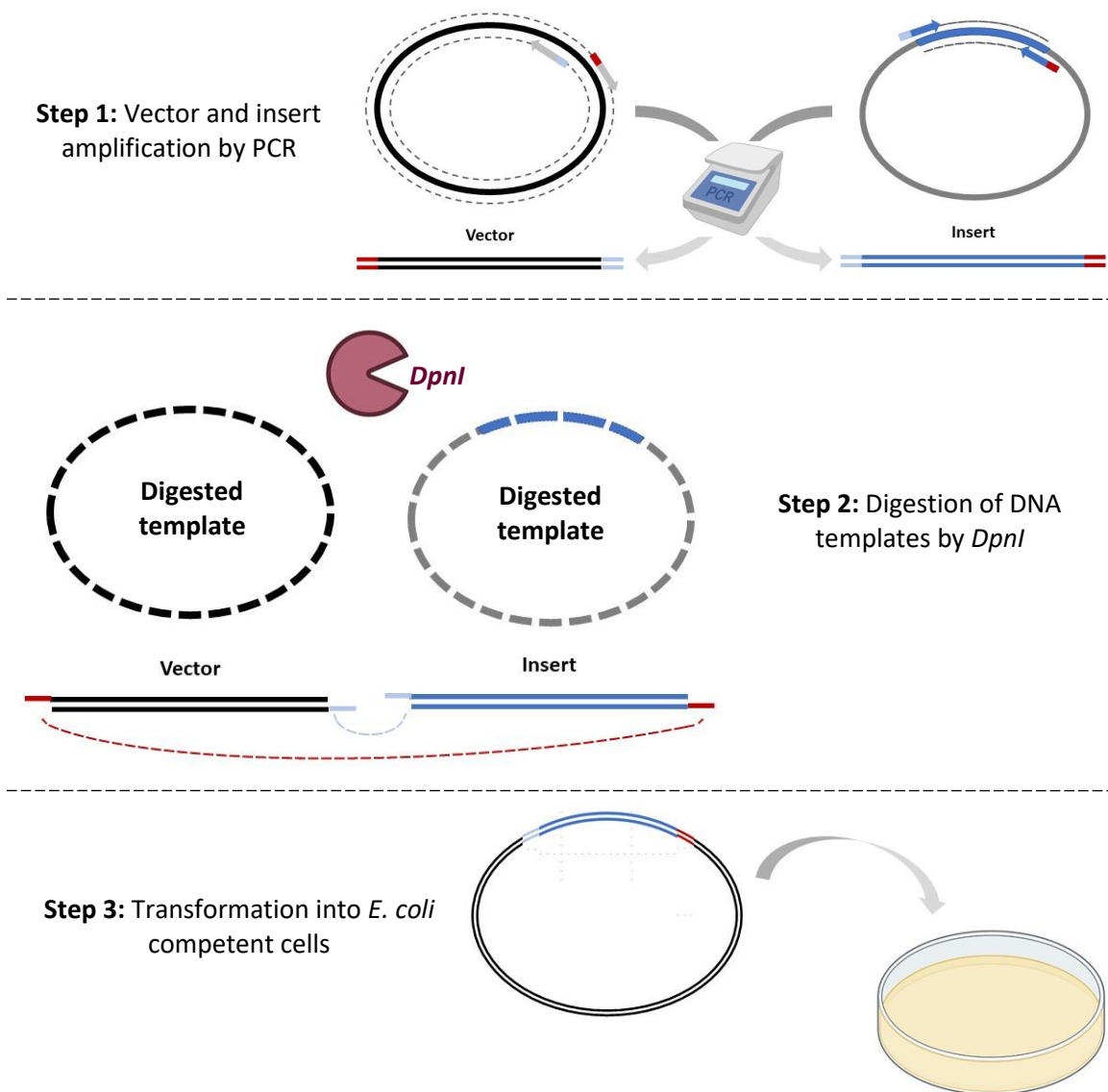


Figure A2 | FastCloning procedure. Step 1: PCR amplification of insert and vector using primers pairs which guarantee the approximately 16-base overlapping at the end of both. Step 2: *DpnI* digestion to eliminate the DNA templates. Step 3: Transformation of the mixture into *E.coli* competent cells.

7.2.1. PCR amplifications

The primers for vector and insert amplification in Table A1 were designed with NEBuilder Assembly Tool (<https://nebuilder.neb.com>) and have an annealing temperature around 64.5 °C and 56 °C, respectively.

Table A1 | Primers used in the PCR amplification with correspondent sequence, length and guanine-cytosine content in percentage. Lower-case letters in the insert primers represent the overlapping with the vector.

Primer	Sequence	Length	%GC
Vector forward	AAGCTTGC GGCCG CATAATG	20	55
Vector reverse	CTGTAATCCCAAGAATTCGGATCC	25	44
Insert forward	cgaattcttg ggaattacagAGCTTTATTGAAGATCTGC	39	39
Insert reverse	cattatgcggccgcaagcttTAGCATACGCCATCTG	38	53

The PCR reaction components for a total volume of 50 μ L were: 0.5 μ L Phusion DNA polymerase (New England Biolabs), 10 μ L Phusion® HF buffer 5x (New England Biolabs), 10 mM dNTPs (Thermo Fisher Scientifics), 50 ng of DNA template, 0.5 μ M of each primer. The PCR cycling parameters for the vector amplification were 98 °C for 30 sec, (98 °C for 5 sec, 64.5 °C for 15 sec, 72 °C for 2.75 min) $\times$ 18 cycles, 72 °C for 10 min, and 16 °C as final temperature. The insert PCR cycling parameters were 98 °C for 30 sec, (98 °C for 5 sec, 56 °C for 15 sec, 72 °C for 30 sec) $\times$ 18 cycles, 72 °C for 10 min, and 16 °C as final temperature. 5 μ L of each PCR product were analyzed in a 1% agarose gel electrophoresis with GreenSafe staining (NZYTech) ran at 100 V for 30 min. The resulting gel present in Figure A3 was visualized under a UV trans-illuminator. The presence of bands with the expected base pairs of 5408 bp and 272 bp for the vector and the insert, respectively, confirmed the PCR success.

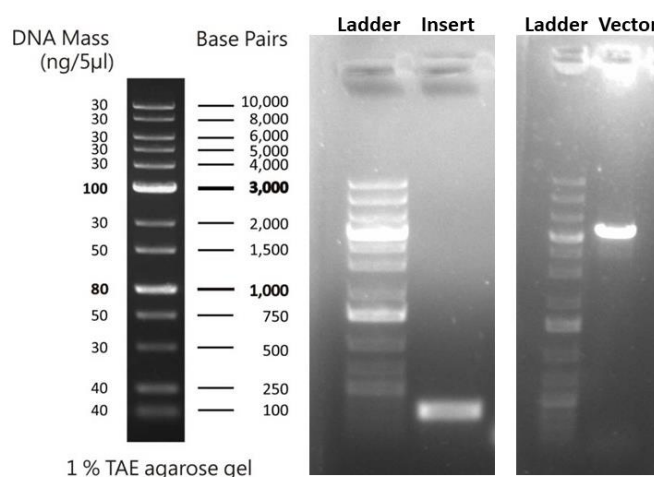


Figure A3 | PCR products analyzed with 1% agarose gel electrophoresis.

After PCR products confirmation, 10 μ L of amplified vector and insert were co-digested with 2 μ L of *DpnI* (Thermo Fisher Scientifics) at 37 °C for 2h. Then, 2 μ L of the resulting mixture were added to *E. coli* DH5 α competent cells and the transformation was performed as explained in section 3.3.1. In the next day, colonies were picked and re-suspended in 20 μ L milli-Q® H₂O for PCR confirmation using: 0.125 μ L DreamTaq DNA Polymerase (Thermo Fisher Scientific), 2.5 μ L DreamTaq™ Green Buffer 10 $\times$ (Thermo Fisher Scientific), 10 mM dNTPs (Thermo Fisher Scientific), 2 μ L cells and 1 μ M of each primer presented in Table A2 for a total volume of 22.5

μL. The PCR cycle parameters were 95°C for 5 min, (95 °C for 30 sec, 54.6 °C for 30 sec, 72 °C for 1 min) ×30 cycles, 72 °C for 10 min, and 16°C as final temperature. DNA sequencing was performed by STAB VIDA. After this confirmation, an overnight culture was prepared and, then used to extract the expression plasmid with the NZYMiniprep Kit (NZYTech).

Table A2 | Primers used in the colony PCR amplification with correspondent sequence, length, guanine-cytosine content in percentage and the annealing temperature.

Primer	Sequence	Length	%GC	T _{annealing} (°C)
Fwd	CAAAGTGACCCTGGCGG	17	65	54.6
Rv	GCTAGTTATTGCTCAGCGG	19	53	

7.3. RICH 1 expression tests

Table A3 lists the conditions tested for the RICH 1 expression. The aliquots collect during the RICH1 expression tests were evaluated in terms of the protein solubility and quantity. The observation of the resulting WB in Figure A4 shows that the majority of the expressed RICH 1 is in the insoluble fractions or pellets.

Table A3 | Conditions of the RICH 1 expression tests where the induction started at an OD_{600nm} of 0.6 and was conducted at 18°C for 18h.

<i>E.coli</i> strain	Culture medium	[IPTG] (mM)
BL21(DE3)	LB	0.1
		0.5
	2xYT	0.1
		0.5
BL21(DE3)pLysS	LB	0.1
		0.5
	2xYT	0.1
		0.5
BL21(DE3)gold	LB	0.1
		0.5
	2xYT	0.1
		0.5
BL21(DE3)star	LB	0.1
		0.5
	2xYT	0.1
		0.5

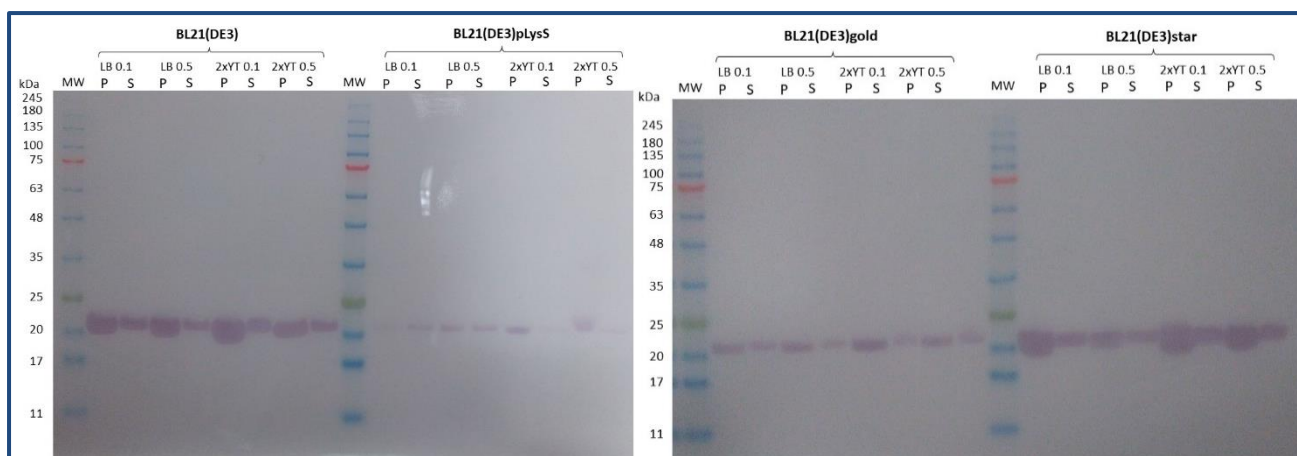


Figure A4 | RICH 1 expression evaluated by WB. MW: NZYColour Protein Marker II. P and S: pellet and supernatant collected after cell lysis. Each lane corresponds to a distinct expression condition, accordingly to the labels: LB and 2xYT – culture mediums; 0.1 and 0.5 – IPTG concentration in mM used to induce protein expression; BL21(DE3), pLysS, gold and star – E.coli strains.

7.4. Calibration curves

The calibration curve of the S75 10/300 GL column was determined using the standards in Table A4 included in the gel filtration kit LMW (GE 28-4038-41). 100 μ L of sample was loaded at a flow rate of 0.5 mL $\cdot$ min $^{-1}$ into the column previously equilibrated in PBS.

The calibration curve equation was calculated by representing the logarithm of the molecular weight ($\log MW$) of each standard as a function of the elution volume (V_{Elution}) as presented in Figure A5.

Table A4 | Protein standards used to determine the calibration curve. Each protein molecular weight (Da) and the respective elution volume (mL) is presented.

Protein	Molecular Weight (Da)	V_{Elution} (mL)
Aprotinin	6500	15.761
Ribonuclease A	13700	13.187
Carbonic Anhydrase	29000	11.300
Ovalbumin	44000	10.030
Conalbumin	75000	9.250
Blue Dextran 2000	2000000	8.129

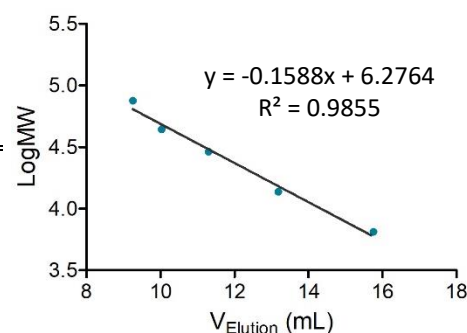


Figure A5 | Calibration curve of the S75 10/300 GL column and the correspondent equation

7.5. Structural characterization

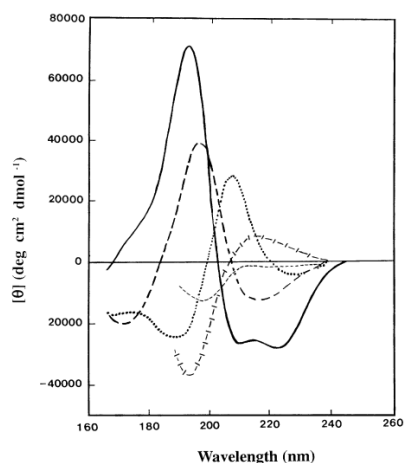


Figure A6 | Typical CD spectra of each secondary structure element. (Solid line) α -helix; (Long dashed line) anti-parallel β -sheet; (dotted line) β -turn; (cross dashed line) extended 31-helix ; (short dashed line) irregular structure. Retrieved from reference

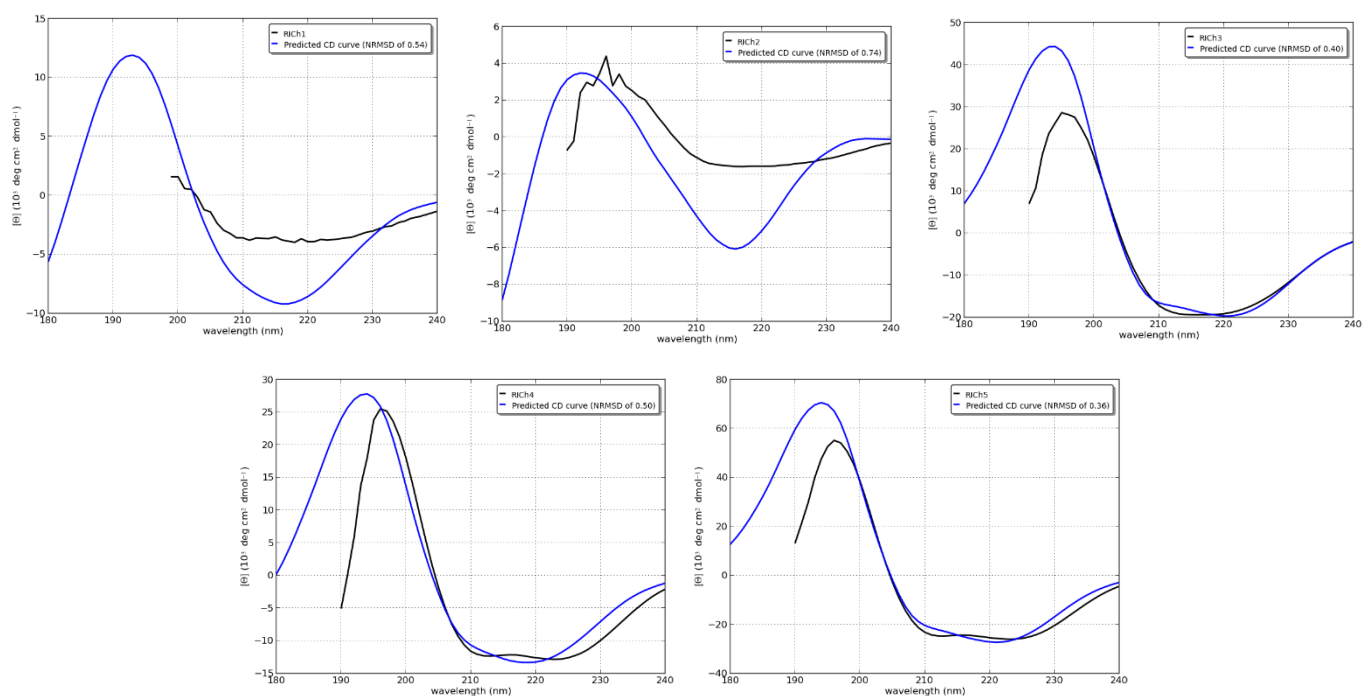


Figure A7 | CD spectra of RICHs and CAPITO fit. (Black) CD spectrum acquired for each RICH represented as the mean residue ellipticity in $\text{deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}$. (Blue) Fit of the experimental data performed using the CAPITO web server-based tool and the respective NRMSD.

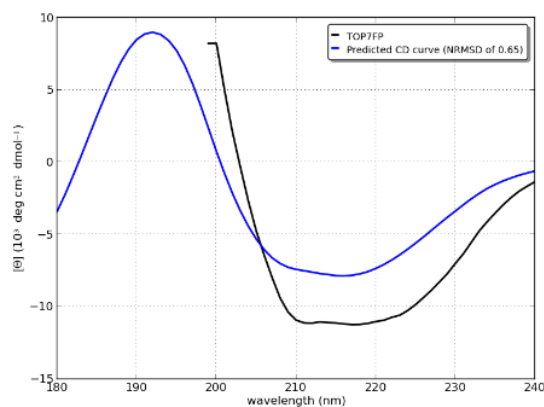


Figure A9 | CD spectra of TSSFP and CAPITO fit. (Black) CD spectrum acquired represented as the mean residue ellipticity in $\text{deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}$. (Blue) Fit of the experimental data performed using the CAPITO web server-based tool and the respective NRMSD.

7.6. Functional characterization

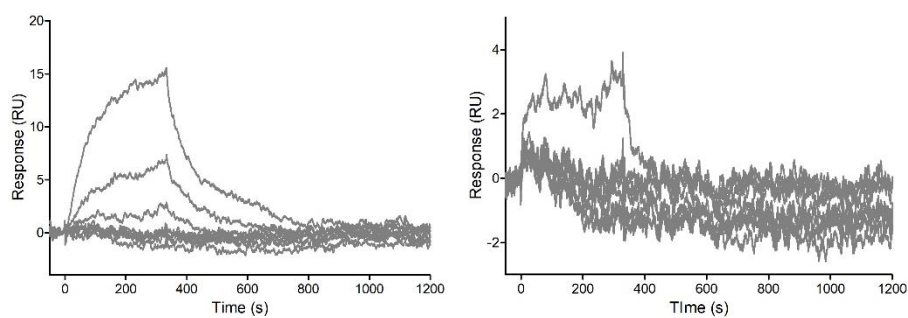


Figure A8 | Sensorgrams for the binding of scaffold 1 (left) and scaffold 5 (right) to the SARS-CoV-2 RBD obtained by surface plasmon resonance (SPR).