
THE UNIVERSITY *of* EDINBURGH

# Stapled by design, New peptide-based therapeutic leads targeting protein-protein interactions

**Marie BLUNTZER**

Dissertation for the degree of doctor of philosophy (PhD)
at the University of Edinburgh, UK

2021

Dissertation date: November, 2021

I, Marie BLUNTZER, declare that this thesis titled, “Stapled by design, New peptide based therapeutic leads targeting protein protein interactions” and the work presented in it are my own. I confirm that:

- This thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree.
- Except where stated otherwise by reference or acknowledgment, the work presented is entirely my own.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Part of this work (Chapter 1) have been published in : Bluntzer, MTJ, O’Connell, J, Baker, TS, Michel, J, Hulme, AN. Designing stapled peptides to inhibit protein-protein interactions: An analysis of successes in a rapidly changing field. Peptide Science. 2021; DOI 10.1002

Year: 2021
Title: Stapled by design, New peptide-based therapeutic leads targeting protein-protein interactions
Author: Marie BLUNTZER

# Lay Abstract

The work in this thesis describes the development and use of computational methods to design and predict the structure of stapled peptides. Peptides, like insulin, can be used as drugs to treat various diseases. However, peptides can be quite easily degraded in the blood by proteases, which can act like scissors cutting the bonds between the amino acids, allowing peptides to be cleared by the kidneys. Peptides also generally need to be injected (e. g. insulin) as they are not able to cross the gastrointestinal wall before being digested. Adding staples to peptides is an emerging technology to enhance the pharmacological properties of peptides, making peptides more stable and thus more efficient as drugs. Some stapled peptides might even be taken orally.

In the stapling technique, unnatural amino acids are included in the peptide sequence. Through various types of chemical reactions, the side chains of two specific unnatural amino acids are linked and act as a brace. Such staples reduce the flexibility of the peptide by constraining it into a certain shape. Introducing staples can sometimes reduce the effectiveness of proteases or/and produce peptides that binds to their target with more affinity, making them more active. It is thought that the fact a peptide adopts a certain shape, and in particular a helix shape, plays a large part in its enhanced properties. However not every stapled peptide forms a helix, thus being able to predict the peptide structure, could potentially help to design such peptides. Given the small set of available data, machine learning methods cannot be employed because they require large amounts of data to be trained efficiently. Thus, *ab-initio* method, that predicts secondary structures without prior knowledge must be employed and Molecular Dynamic (MD) simulations have been proven useful in different reported studies.

MD simulation is a technique which uses Newton's classical equations of motion to simulate the motions of atoms in large molecules, and it can be applied to proteins and peptides. Following the many successes of MD to design ligands for proteins, the goal of this project is to validate the utility of MD for the prediction of stapled peptides structure. However MD relies on the use of forcefields in which every molecule and amino acid is described using pre-established parameters. As stapled peptide technology is recent, there are no available parameter set for the unnatural amino acids used to introduce the staples into peptides. In this work, we implemented a method to facilitate the calculation these parameter sets for unnatural amino acids. We then produced parameter sets for the unnatural amino acids commonly used in staple chemistry which had also to be compatible with the existing sets of parameters

for natural amino acids.

To predict the conformation of peptides, MD methods that sample a large number of conformations must be employed, thus in this work we implemented a Solute Tempering approach. Solute tempering is a form of MD, where several simulations are run in parallel, with some of these simulations being progressively heated. Using this method facilitates the observation of structured and unstructured conformations in peptides. From our MD simulations we have many snapshots of the peptide and we can deduce which conformations occur more often, which in turn can be linked to the conformation of the peptide.

When designing a molecule or peptide which can bind to the surface of a protein, it might block its interaction with other proteins. This is a desirable effect if a biomolecule is malfunctioning, or, if altering the function will have some therapeutic effect.

One main goal of computer-aided drug design is to find compounds which bind into a specific binding site, through design and computational screening. Here we developed tools to design sequences to bind stapled peptides to these proteins, and build computational models.

We designed peptides to target the 'spike protein', present at the surface of the SARS-CoV-2 virions, which is responsible for the on-going COVID-19 pandemic. By designing a peptide able to bind to this protein and by using fluorophores in the peptide sequence, it allows, in theory, an easy detection of the virus in patient samples (e. g. saliva...), making the diagnosis of COVID-19 cheaper and more rapid. These design strategies were then tested through MD. By simulating a peptide bound to a protein active site, insights into the strength of its interaction with the protein can be gained, by measuring relevant quantities, such as distance variations between the receptor and peptide atoms, interactions between the peptide and protein atoms. Values to quantify the conformational changes of the peptide bound and the dissociation of the peptides are also useful. Over a hundred sequence variations of peptides sequences were performed, thus producing a large amount of data. Using PCA (Principal component analysis) reduces the complexity of looking at each variable separately in the data set and also produces more relevant descriptors. PCA implements a linear combination of variables into a fewer number of "collective variables" and help the analysis of large numbers of simulations.

IL-1R is a protein implicated in inflammatory response. We used tools to design stapled peptides using different type of chemistry. We then used free Energy methods, in which the binding energy is quantified *via* MD. The Free Energies can be related to a compound affinity. And used the simulation to select the most stable design *in-silico*. After synthesising the peptides and testing them against IL-1R we compared *in-vitro* and *in-silico* affinity values of stapled peptides. Here we report the development of stapled peptides able to bind to the IL-1R protein with a better affinity in cell assay, than any reported peptides for this protein.

In this work we thus explored a number of MD methods to design and predict the properties of stapled peptides.

# Scientific environment

The author has carried out the research reported in this dissertation at the Research Group of the Professor Alison Hulme (*LaMoRe*), Department of Chemistry, Edinburgh, UK and at UCB, Slough, UK.

We thank Medical Research Scotland (Studentship to MTJB, ref PhD-1042-2016) and UCB for funding.

# Acknowledgements

First, I would like to express my gratitude to my supervisors. I would like to thank Prof. Alison Hulme for her patience, support and insightful suggestions and Dr. Julien Michel, for his guidance and advice. I would like to thank Dr. James O'Connell for his help, input and suggestions on the project. I would like to extend my sincere gratitude to Dr. Terry Baker, who, although no longer with us, continues to inspire future generations of scientists by his wisdom and dedication to science.

I would also like to thank everyone at UCB for their warm welcome during my placements, and I am especially grateful to Dr. Melanie Wong and Martin Domville for their technical support and their patient explanations. I would also like to thank Dr. Ben Cossins, who introduced me to molecular dynamics and encouraged me to start a PhD in the first place. He also gave me substantial advice during this project.

From the University of Edinburgh, I would like to thank Prof. Manfred Auer and Dr. Paul Barlow for their collaboration on the COVID-19 project. Also, many thanks to Dr. Phan Nhan who developed a CONA assay, to test some of our hypotheses.

I would like to thank all present and past members of the Hulme and Michel groups who have helped me on the project, in particular, Fergus, Richard, Toni, Salome, Jordi, Alan and Stefano. In addition, I would like to thank Lisa, Maria, Edith and Sally for their friendship, and for making the chemistry lab and computational office nice places to work.

I would like to thank my flatmates Ritumba, Jarvi, Stefanie, Angelo and Errikos for keeping me sane and making our flat a cosy and social place to come home to. I would like to thank my friends Amaëlle, Anne, Blandine for checking in on me, and for taking my mind away from my PhD at times. Thanks to Lana, Tiem and Scottie for the board games nights. Also, a big thank you to all the members of the MDCC for the lovely Sunday morning cycling rides (and coffees breaks!) over the last year.

Je voudrais remercier ma famille. Et en particulier je voudrais remercier ma mère, qui, certainement, est ma plus grande supportrice et admiratrice. Merci pour ton soutien continu et infaillible durant ces trois dernières decenies. Last but not least I would like to thank my partner Seppe for his incredible support. You have tolerated many of my periods of doubts and stress during my PhD and the writing of the thesis. I am not sure how I would have went through without you. Nogmaals bedankt.

# Abstract

Interest in the development of peptide-based therapeutics has increased in recent years, mainly due to the relatively low toxicity of both the peptides and their metabolites, and the ready availability of building blocks and ease of synthesis, which allows for facile structural variation. Peptide binding sites generally provide larger contact surfaces than small molecules, and sometimes more exposed interaction sites, making the regulation of large protein-protein interfaces (PPIs) possible. However, in contrast to antibodies and other proteins, in most cases natural short peptides (10 to 20-mers) do not retain a stable secondary conformation, which can lead to a decreased affinity. Indeed, most of the peptide drugs that have reached the market are exceptions, with the ability to retain a compact, stable conformation (for example macrocyclic peptides or peptides stabilised with disulphides bridges). Thus, constrained peptides, which are forced to retain a rigid conformation, might be better candidates for protein binding. 'Stapling' peptides can induce an $\alpha$-helical structure by introducing 'stapled' residues acting as a synthetic brace, holding two residues of adjacent helix-turns close together. Interest in the development of stapled peptides has markedly increased over the last decade, due to early successes and the increasing availability of staple precursors. The handful of successful strategies to design these novel peptides that have already emerged are discussed in Chapter 1, together with significant examples that have influenced research in the field.

Among computer-aided drug design techniques, Molecular Dynamics (MD) simulations and associated methods have earned increased attention within the pharmaceutical sector due to improvements made in the last decades. However, there are still limitations to the use of these methods for peptide development and design, due to the lack of protocols for using unnatural amino acids in MD workflows. In this thesis, MD simulations are used to design peptides containing non-proteinogenic amino acids and assess their properties. In Chapter 2, the fundamental principles of computational chemistry and MD used in this project are introduced.

Chapter 3 describes the development of a pipeline to generate and validate parameter sets automatically, which can be applied in MD simulations for modelling unnatural amino acids commonly used in staple chemistry. The parameters were found to be in good agreement with quantum mechanics data and fully compatible with the AMBER14SB forcefield. The unnatural amino acids also reproduced the behaviour of natural amino acids in terms of psi/phi distribution. In Chapter 4, these parameters were integrated into a custom solute

tempering protocol to sample the conformational space of stapled peptides. This method was benchmarked against the more 'traditional' T-REMD and was found to be computationally less intensive while producing similar sampling. A set of well-documented stapled peptides, based on a helical peptide from the tumour suppressor protein p53, which binds the negative regulator MDM2, was used to probe the accuracy of the MD simulations. It was found that the experimentally determined helical content of these stapled peptides, reported in the literature, could be successfully reproduced.

This first validation of the new forcefields for stapled residues, allowed the development of a computational platform to design helical peptides from known PPIs called 'Stapline'. This pipeline extracts key binding interactions of known protein binders and builds models that replicate these interactions using the side chains of stapled helical peptides. The application of the 'Stapline' platform to the in-silico design of two classes of stapled peptides that selectively recognise the receptor binding domain of the "Spike" protein on the SARS-CoV-2 virus is described in Chapter 5. The work was extended to a third class of stapled peptides, using studies of coil systems to design peptides to the tail of the Spike protein. The binding affinity of these fluorescent, stapled peptides to the SARS-CoV-2 virus were assessed against known Spike-protein binding peptides in biophysical assays.

This 'Stapline' procedure was then integrated into a second peptide design workflow to optimise the binding selectivity of IL1-R inhibitors as described in Chapter 6. IL1-R is a protein involved in many diseases including rheumatoid arthritis, as it is one of the key proteins in the inflammatory pathway. Peptide inhibitors of the IL-R1/IL-1 interaction developed to date have nanomolar potency but have short half-life. Thus, there is a need for new peptides with improved drug-like properties. Following MD simulations, a MMPBSA (Molecular Mechanics Poisson-Boltzmann Surface Area) and MMGBSA (Molecular Mechanics Generalised Born Surface Area) protocol was employed. The efficiency of this protocol in assessing the binding of the peptides to their target, as well as predicting their interactions was evaluated. By combining MD descriptors to quantify peptide stability and relative free energy of binding obtained from MMPBSA and MMGBSA, we suggested favourable stapled peptide designs. A selection of highest-ranked stapled peptide inhibitors were characterised using biophysical techniques, including FRET, Biacore and a cell-based assay, leading to the successful identification of novel potent binders.

# Contents

# Acronyms

**AMBER** Model Building with Energy Refinement

**CD** Circular Dichroism

**CMAP** correction map

**COVID-19** COrona VIrus Disease 2019

**CPU** Central Processing Unit

**ESP** ElectroStatic Potential

**FRET** Förster Resonance Energy Transfer

**GAFF** General AMBER Force Field

**GPU** Graphic Processing Unit

**HF** Hartree-Focks

**HR** Heptad Repeat

**HTRF** homogeneous time-resolved fluorescence

**IL-1** Interleukine-1

**IL-1R** Interleukine 1 receptor (type I)

**MD** Molecular Dynamics

**MEP** Map of Electronic Potentials

**MetaD** Metadynamic Simulation

**MS** Mass Spectrometry

**Ni-NTA** Nickel- Nitrilotriacetic acid

**PCA** Principal Component Analysis

**PME** Particle Mesh Ewald

**PPI** Protein-Protein Interface

**QM** Qunatum Mechanics

**REMD** Replicas Exchange Molecular dynamic

**RMSD** Root Mean Square Deviation

**SCF** Self-consistent field

**uAA** unnatural amino acid

**vdW** Van der Waals

# 1

# Stapled Peptide Design Strategies

## 1.1 Introduction

Despite heavy investment in the early 21st century, a number of challenges have stalled the development of marketable peptide therapeutic drugs. As a consequence, between 2003 and 2019 only 13 synthetic peptide drugs reached the European market.[1] But with 140 peptides in clinical development in 2019,[2] interest in this class of therapeutics is clearly growing and constrained peptide technologies have attracted attention from larger pharmaceutical companies as well as academic laboratories and smaller start-ups.[3] Short peptides with natural amino-acids usually have poor drug-like properties since they frequently have high conformational variability, low cell penetration, and undergo rapid proteolysis; whilst longer peptides can be challenging both to produce and deliver in cells. Some of the earliest peptide drugs developed over 50 years ago, were natural hormones such as insulin,[4] erythropoietin,[5] oxytocin,[6] secretin[7] and calcitonin,[8] which all have comparatively high molecular masses (3 – 7 kDa). However, shorter peptides were also marketed during the same period; including vasopressin,[9] a natural 10-mer peptide hormone. Notably the first seven FDA-approved peptidic drugs (Insulin, Adrenocorticotropic hormone, Calcitonin, Oxytocin, Vasopressin, Octreotide and Leuprorelin)[10] all had a stabilised, or constrained, secondary structure, which is linked to improved resistance to proteolysis. Thus stabilisation of secondary structure, be it through the introduction of non-peptidic fragments, backbone modifications, or unnatural amino acids to the sequence has been a prominent feature of peptide drug development since the 1990's. Stapled peptides, whereby cross-linking of two or more side-chains is carried out via chemical synthesis, generally have a more compact structure, enhanced cell penetration, and are more resistant to proteolysis.[11–13] Moreover, their metabolites are relatively safe,[14] and recent studies suggest that stapled peptides offer an advantage over traditional drugs by averting the development of drug resistance.[11, 15] Hence stapled peptides have predictably sparked a growing interest from the scientific community since the early 2000's and have emerged as a potential new class of drugs. A 2019 review by Ali *et al.* found 78 stapled peptides where structural information has been submitted to the protein databank (PDB),[16] of which 55 also had binding affinity reported.[16] In parallel, the chemistry landscape for stapled peptides is continuously diversifying with a dozen "novel" linker chemistries reported in publications in 2019.[17–20] In some instances, only very sub-

tle differences in the staple chemistry, or stereochemistry, trigger a significant change in the binding affinity, the pharmaceutical profile of the peptide, or give added functionality.[21–27] Despite multiple publications and reviews describing reliable methods for the synthesis of stapled peptides[28–33] and the commercialisation of the most common unnatural amino acids, their synthesis remains costly and until a few years ago, non-automated. Consequently, the synthesis and screening of stapled peptides has usually been restricted to a few dozen peptides, limiting access to extensive libraries. Stabilized peptide scaffolds sit at the interface of biologics and small-molecule drugs, and guidance for their development can be found in both fields. In this Chapter, we focus on the successful approaches that have been published for the design of stapled peptides that inhibit specific PPIs, and do not cover the design of stapled peptides that act as antimicrobial peptides (AMPs). Since many stapled peptides are designed to inhibit intracellular PPIs, they necessarily share common features with another class of peptide therapeutic, the cell-penetrating peptides (CPPs).[15, 18, 34–39] We cover some of the notable strategies that have been used to optimise the pharmacokinetic properties of stapled peptides sequences, enhance their activity and reduce their proteolysis. But whilst we have focussed this Chapter on these enhancements, as with other drug classes, in some cases more 'classical approaches' might provide a better alternative to stapling.[40, 41] Rather than highlighting solely standard PPI targets (e.g. p53-MDM2/MDMX[42–46] BCL9/$\beta$-catenin;[47] BCL-2(BAX)/BCL-XL,[48] HIV gp41/10E8 antibody,[49] LEDGF-p75/HIV-IN[50]), we have included a broad range of examples from the very recent literature to show how widely peptide stapling is being adopted.

## 1.2 Defining the Primary Sequence

Cell signalling is achieved principally through a cascade of protein-protein interactions that assemble functionally related proteins into complexes, activating signal transduction pathways. The protein interaction network of an organism, or interactome, generally gives a better indication of its biological complexity than its genome. In 2017, Hutlin *et al.* estimated the number of candidate interactions in the human interactome as 56,000; [51] specificity at the binding interface is thus essential to ensure cell functionality. As for natural peptides, the specificity of stapled peptides is largely ensured by the choice of the amino acids interfacing with the target protein. The synthetic landscape is vast: using the 20 natural amino acids to build peptides of up to five residues gives over three million possible sequences. The synthesis of stapled peptides is more challenging than linear peptides, as they often contain unnatural amino acid (uAA) residues that are not biosynthesised by bacteria, and can be costly to prepare. In this section, we have analysed recent successes in the design of stapled peptides to identify specific techniques used for their primary sequence optimisation.

### 1.2.1 Natural peptide interaction sequences

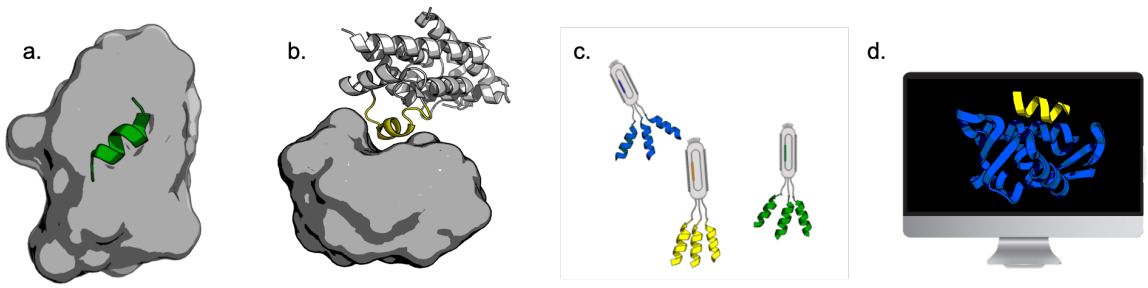



**Figure 1.1:** Approaches to defining the primary sequence of stapled peptides. A Starting from a natural peptide interaction sequence. B Excising a helical fragment from a protein-protein interface. C Developing stapled peptide sequences using screening. D Computational design of stapled peptide primary sequences.

Starting from a lead peptide with poor pharmacokinetics and improving its properties by using stapling techniques is the most obvious route to advance the development of peptide therapeutics. Multiple successful examples have been reported in which stapling has been achieved across a variety of secondary structures such as helical peptides, beta-hairpins, and 'extended' peptides.

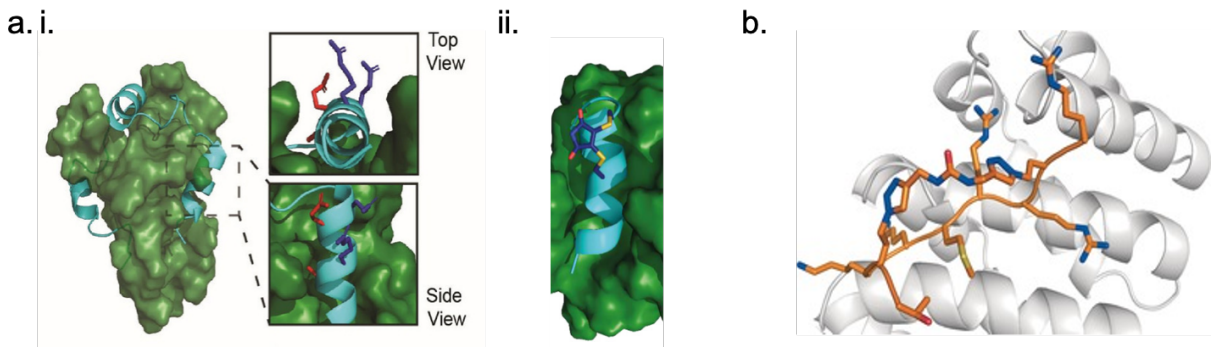



**Figure 1.2:** Peptide conformations constrained by stapling. a. (i) Crystal structure of HIF-1$\alpha$ (blue) binding to p300 (green); (ii) Model of a stapled peptide (blue) derived from HIF-1$\alpha$ binding to p300 (green) reproducing similar contacts. Reproduced from reference [52] under Creative Commons Licence CC-BY. b. Crystal structure of an extended stapled peptide (orange) derived from the HNF1$\beta$ NLS peptide which binds to mImportin $\alpha$1. Reproduced from reference [53] under Creative Commons Licence CC-BY.

In 2017, Wu *et al.* investigated the cell penetrant helical peptide, melittin.[**?** ] Melittin is a naturally occurring peptide from honeybee venom that has been found to inhibit the proliferation of hepatocellular carcinoma (HCC) cancerous cells. Wu *et al.* prepared a library of stapled peptides by truncation and optimisation of the staple position and length, using Verdine's ring-closing metathesis method for all-hydrocarbon staple production.[54] The stapled peptides only differed by their sequence length, the residues replaced by the staple, and the length of the staple linker. Some of the stapled peptides exhibit enhanced binding affinity

and significantly improved resistance to proteolysis over the parent peptide. The enhanced helicity of a stapled peptide in solution often promotes its activity; stapled peptides based on the sequence of helical, estrogen receptor (ER) coactivator peptides, show a marked increase in helicity and enhanced Kd relative to the initial peptide whilst retaining an almost identical binding mode.58 The HIF-1$\alpha$/p300 PPI plays a key role in tumour metabolism. The HIF-1$\alpha$ peptide is a 40-mer containing 3 $\alpha$-helical segments each of which binds p300 at a specific site. By focusing on the largest helical segment, and using a dibromomaleimide stapling strategy, a competitive binder of the HIF-1$\alpha$ peptide with increased helicity in the bound state was reported (Figure 1.2.a).[52] Non-helical stapled peptides targeting nuclear PPIs have also been developed in recent papers. Wiedmann *et al.*[53] used peptides based on an HNF1$\beta$ NLS sequence developed in the 1990's by Lin *et al.*[55] and "double click" CuACC staple chemistry,[56–58] to produce a series of constrained extended stapled peptides (Figure 1.2.b). Initial bis-triazole stapled peptides had reduced binding affinity for their protein target relative to the peptide lead. After further optimisation of the linker length, a stapled peptide with an equivalent low micromolar Kd was generated. Crucially, cell penetration was only observed for the stapled peptide. McGrath *et al.* have designed a constrained peptide that binds to the WDR domain of Transducin-like Enhancer (TLE). The design was based on a peptide derived from a transcription factor binding partner.[59] Stapling this short peptide was achieved by linking the indole moiety of the N-terminal tryptophan to the C-terminal proline via ring-closing metathesis. X-ray analysis showed that the binding mode was almost identical to that of the initial peptide with an approximately 6-fold higher affinity of the stapled peptide over its acyclic counterpart.[55]

### 1.2.2 Excision of a helical fragment from a protein-protein interface

Frequently structural data for a PPI implicated in a disease is available, but no peptide binders have been reported. If the binding interface includes one short contiguous protein segment, this becomes a convenient starting point for the development of a therapeutic peptide. If the secondary structure of the interacting protein segment promotes the positioning of the PPI interacting residues, peptide stapling is most likely to reproduce the native contacts. Indeed, McWhinnie *et al.* have reported three hydrocarbon-stapled peptide series which reproduce the secondary structure of the native protein they were extracted from ($\alpha$Syn, CKS1 and CK1$\alpha$) with high fidelity, whereas the unstapled natural sequences displayed almost no secondary structure content (Figure 1.3.a).[60] Similarly, Wang *et al.* have isolated a minimum helical motif from the coiled-coil region identified in the crystal structure of the respiratory syncytial virus fusion (RSV-F) protein,[61] and then enhanced the helicity of this initial peptide sequence using computational mutagenesis and all-hydrocarbon peptide stapling, achieving a corresponding two-fold improvement in binding affinity.

Successful examples of a peptide excision strategy applied to the generation of all-

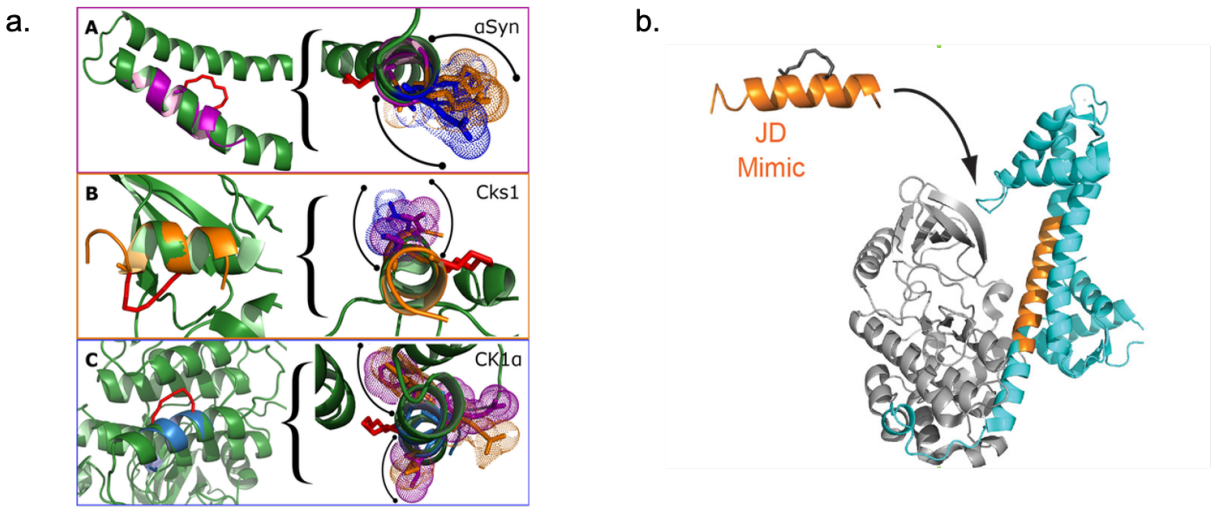



**Figure 1.3:** Excision strategies in the design of stapled peptides. a. NMR Structures of three stapled peptides derived from (A) $\alpha$Syn (purple), (B) Cks1 (orange) and (C) CK1$\alpha$ (blue) aligned with their initial protein segment (green). PDB codes 1XQ8, 2AST and 1CK1. Reproduced with permission from reference [60]. b Design of a stapled peptide (orange) from the excision of the junction domain of Plasmodium falciparum Calcium-Dependent Protein Kinase 1 (orange) in a catalytically inactive state. Calmodulin-like domain and kinase domain shown in blue and gray respectively. PDB code 3KU2. Reproduced with permission from reference [62].

hydrocarbon stapled peptides include: stapled peptides which block the helix-helix interfaces found in the coiled-coil region of the gp41 protein with potent anti-viral activity against HIV-1;[63] stapled peptides which mimic the junction domain of Plasmodium falciparum calcium-dependent protein kinase 1 (PfCDPK1) (Figure 1.3.b) to disrupt J-domain binding and provide allosteric inhibition of PfCDPK1 activity, blocking malarial parasite development;[62] stapled peptides extracted from the dimerization interface of bone morphogenetic protein-2 (BMP-2) which is a possible target for bone repair therapies;[64] stapled peptides which target the nuclear transcription factor NF-Y, a therapeutic target implicated in various diseases such as cancers and neurodegeneration.[65] Clearly, the success of these strategies has relied on the availability of crystallographic data and amino acid building blocks to readily synthesise stapled peptide sequences.

### 1.2.3 Stapled peptide sequences from biological screening

Until very recently, many popular biological approaches used for the development of peptide therapeutic leads (e.g. phage display,[66] ribosome display,[67] mRNA-display,[68] bacterial or yeast surface display[69]) could not be applied to the development of stapled peptides due to the need to incorporate unnatural amino acids, and stapling reaction protocols harmful to bacteria or phages. Instead, such screening methodologies are used to identify initial peptide leads that are then optimised into stapled variants. To illustrate, Achek *et al.* discovered the first peptide inhibitor of TLR4, a key target for the treatment of rheumatoid arthritis, using phage display techniques and lactam stapling which increased the helicity of the initial

peptide and improved its $IC_{50}$ two-fold over its unstapled variant.[70] Developments in stapling methodologies now allow biological screening to be applied to stapled peptides, where the staple is introduced through cysteine alkylation. In this instance, no uAA incorporation is required, and the cross-linking reaction can be conducted under mild, dilute conditions that are not deleterious to phages. Using a phage library based on the axin $\alpha$-helix (Figure 1.4.a), Diderich *et al.* were able to generate a cysteine-stapled peptide PPI inhibitor for $\beta$-catenin which showed nanomolar activity in a fluorescence polarization competition assay.[71] Using the same approach with biphenyl linkers, Anananuchatkul *et al.* produced micromolar binders of Galectin-3.[72] The cysteine alkylation approach has been extended to the incorporation of $\alpha$-methyl substituted cysteine using a strategy based on the mRNA display of peptides (Figure 1.4.b). This uAA has been shown to increase the helical propensity of the cysteine-stapled peptide, giving similar properties to related all-hydrocarbon stapled peptides with $\alpha$-methyl substitution.[73] Since mRNA display can create libraries that are several orders of magnitude larger than on-bead or phage display libraries, this could be a very powerful strategy for the generation of new stapled peptide leads.

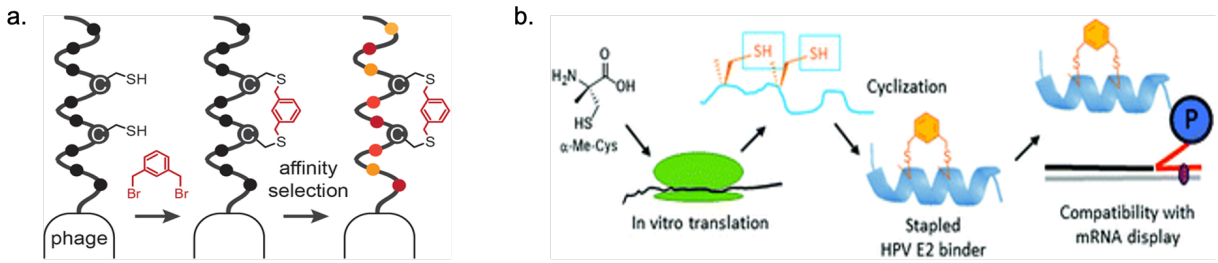



**Figure 1.4:** Adapting biological methods to the rapid screening of stapled peptides. a. cysteine crosslinking stapling procedure allows phage display screening of stapled peptides without the need for uAA incorporation. Reproduced with permission from reference [71]. b. The display of peptides incorporating $\alpha$-methyl substituted cysteine on mRNA allows screening of large libraries of stapled peptides. Reproduced from reference [73] under Creative Commons Licence CC-BY-NC.

### 1.2.4 Computational design of stapled peptide sequences

Analysis of structures of protein-peptide complexes derived from X-ray diffracted crystals, NMR measurements or MD simulations is another effective approach to design of stapled peptide sequences (Figure 1.1.d). The most straightforward analyses rely on interactive 3D visualisation of structures using popular molecular graphics software such as as Pymol.[74] The position and spacing of hydrogen bonds and hydrophobic contacts at a binding interface allow key residues to be inferred. When multiple structures of related peptides are available it may be possible to deduce conformational changes that occur upon binding linear sequences to direct staple optimisation.[75, 76] A complementary computational approach to this structure-based design of stapled peptides involves extracting functionally significant patterns from the sequences of naturally occurring peptides and re-using these leads to design *de*

*novo* peptides. This type of bioinformatics approach requires large databases of peptides with known activity and consequently has not yet played a major role in stapled peptide design, owing to the comparatively small number of such datasets in the literature. Howerer, fast and reliable methods such as AGADIR[77] could potentially be used prior to the introduction of staples to enhance the overall propensity of a peptide sequence to fold into an $\alpha$-helix.[78] The reader is directed to an extended review of AMP databases and data mining by Porto *et al.* for further details.[79] MD simulations enhance structure-based visualisation methods for peptide design by providing information about molecular flexibility, and estimates of binding energetics. MD simulations may be used on their own, but also in conjunction with other computational methods, such as free energy calculations, metadynamics, steered MD, parallel MD, and other methods which overcome limitations inherent to Boltzmann sampling.[80–85] In theory, any MD method for the design of peptides could be adapted to stapled peptides. However, a major limitation currently is the availability of high quality parameter sets to model the energetics of unnatural amino acids. To date, only a few MD methods have been applied to stapled peptides. Fragment-based approaches, such as those applied to the development of small molecule binders and to predicting binding pockets, have been successfully transposed to the design of stapled peptides by Tan *et al.* who used a fragment screen on MDM2 to optimise the placement of both aromatic and staple residues in their peptide lead (Figure 1.5.a).[86] The MELD accelerated sampling technique, which uses a Hamiltonian and temperature replica exchange approach, accurately predict binding poses of stapled peptides where key interactions are known, and also predicts the binding energy of peptides (Figure 1.5.b).[87, 88] Free energy perturbations (FEP) can be used to improve a given sequence; typically a virtual alanine scan is conducted and the free energy is computed to measure the effect of changes,[89] or to enhance the binding properties of the peptide.[90] Despite advances in FEP methods for macrocycles,[91, 92] application of this method to stapled peptides could suffer from complications if the mutations occurred at the staple residues. As an example, Valiente *et al.* describe the use of FEP methods for sequence selection, combined with umbrella sampling calculations to estimate binding free energies of the modelled stapled-peptides the target receptor BCLXl.[93] The binding energy between a peptide and protein can also be estimated with the Molecular Mechanics/Poisson Boltzmann Surface Area (MM/PBSA) or Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) methods.[94] In these methods, molecular dynamics simulations of the receptor-ligand complex in explicit, or implicit, solvent are run to calculate the free energy of solvation of the complex. A few examples of application of MM/GBSA to stapled peptides have been reported, which helped understand the role of hydration and flexibility in stapled-peptide binding.[44, 95–97]

Using docking algorithms to predict specific binding poses for large ligands and peptides is more challenging than for typical drug-like small molecules. Firstly, the PPI binding surface is usually larger and lacking in distinct features, with more hydrophobic contacts than binding pockets for small molecules.[90] Secondly, the conformational landscape of peptides

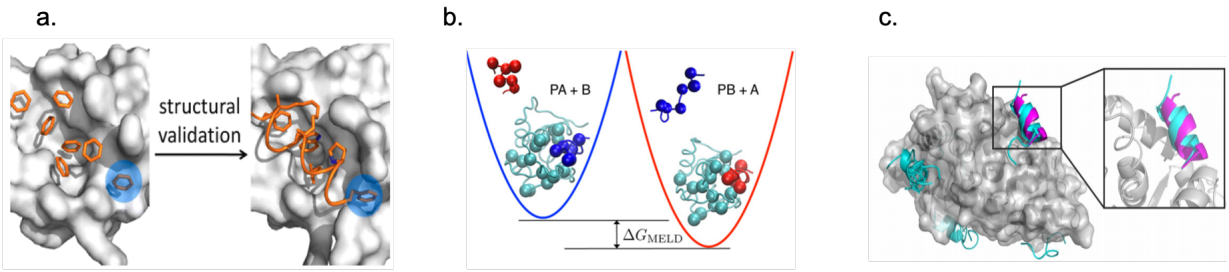



**Figure 1.5:** Adapting computational methods to the study of stapled peptides. A Location of fragments observed in MD simulations (orange) shows the initial hotspots used for the design of a stapled peptide. Reproduced with permission from reference [86]. B MELD sampling swaps two states: peptide A bound to the target with B in an unbound reference state (left); and peptide B bound with A in a reference state (right). The ratio of populations in the two states relates to the free energy difference ΔGMELD. Reproduced with permission from reference [87]. C Peptide docking software, such as CABS-dock, shows that the peptide (blue) docks in the original binding site (purple). Reproduced from reference [98] under Creative Commons Licence CC-BY.

is broader due to the large number of backbone and side-chain torsional degrees of freedoms that typically exceeds the sampling capabilities of most docking algorithms. Several methods have been designed to overcome the difficulties arising from docking peptides, such as allowing side-chain flexibility, or rearrangement, during docking.[83] In practice the more effective docking protocols for peptides are based on protein docking algorithms rather than small molecule docking algorithms. Applications of docking algorythms include publications by Ciemny *et al.* who successfully modelled the flexible regions during p53 peptide binding to the MDM2 protein receptor, using the CABS-dock method which accommodates large-scale structural rearrangements (Figure 1.5C);[98] and Tiwari *et al.* who successfully modelled docking of the p53-activating peptide ATSP-7041 to albumin,[35] using the protein-protein docking program ATTRACT.[44, 99]

### 1.2.5 Analysis of the primary sequences of PDB-deposited stapled peptides

An analysis of crystallographic data for protein-bound stapled peptides allows several conclusions to be drawn regarding their primary sequences. The comparatively high content of unnatural amino acids reflects the relatively short length of stapled-peptide sequences (12-20 residues) that incorporate one, or sometimes two, staples (Figure 1.6). Previous analysis of protein-protein interactions found that hotspots are enriched in tryptophan, arginine, and tyrosines compared to the rest of the protein.[100] Positively charged amino acid residues are over-represented in bound stapled peptides (Figure 1.6.B), consistent with their introduction to enhance solubility,[101–103] and cell-permeability. However, as noted by Bird *et al.*,[104] an important balance must be struck between achieving cellular uptake by stapled peptides, whilst avoiding membrane disruption. Tryptophan, which seems to enhance cellular uptake through favourable membrane interactions,[105–107] and also promotes helicity,[108]

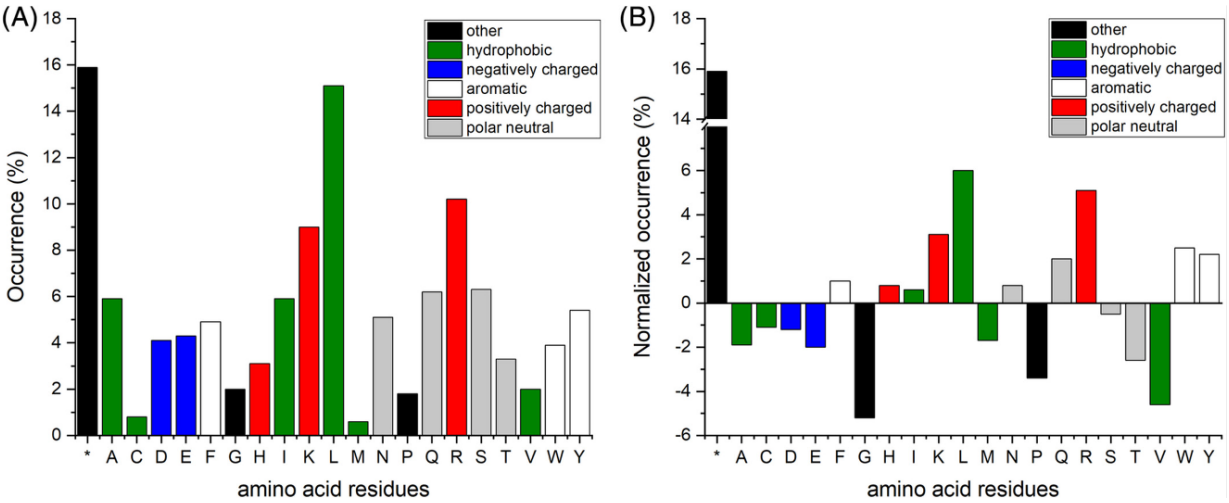



**Figure 1.6:** Analysis of the primary amino acid sequences of stapled peptides targeting PPIs (see ESI). (A) Amino acid distribution in bound stapled peptides from 63 structures in the PDB (Table A.2). (B) Occurrence of amino acid residues in stapled peptides normalized to their natural abundance in proteins.

is over-represented in stapled peptides in the PDB compared to its natural abundance in proteins. Other values are consistent with amino acids identified as either inducing (leucine)[109] or breaking (glycine, proline, aspartate, threonine, glutamate) helicity.[110, 111]

## 1.3 Selection and placement of the staple residues

The selection and placement of the staple in stapled peptides requires many considerations, including how to stabilise a desired secondary structure,[112] whilst retaining sufficient flexibility to allow binding; how to retain, or even enhance, biological activity through protein-staple interactions; and how to influence the activity of the stapled peptide by enhancing cell penetration. Typically these parameters are explored sequentially and as for natural peptides, libraries of stapled peptides can be constituted and scanned against the target protein via different detection methods (PCR, ELISA, NMR HSQC).[113] For example, Araghi *et al.* optimised the sequence of MS1, a native peptide binding to Mcl-1, by first introducing an RCM staple which gave a two-fold improvement in binding affinity.[114, 115] A library of hundreds of stapled peptides was then generated, which focussed on amino acid mutations at key positions using natural and unnatural amino acids; fluorescence-based on-bead screening was used for hit detection.[115]

### 1.3.1 Stabilising secondary structure

The $\alpha$-helix is a common motif in protein-protein recognition. A survey of the Protein Data Bank conducted in 2009, suggested that 62% of protein complexes (out of 9,339) present $\alpha$-helical content at their interfaces, and that less than 2% of the studied cases could be targetable

using a small molecule strategy.[116] $\alpha$-helical peptides, and helical peptides mimetics present some of the most attractive alternative approaches for the design of biological tools and therapeutics. Staple chemistries which favour $\alpha$-helicity are ones which link residues spaced by one or more turns of the helix (3/4 residues for one turn, 7 residues for two turns, 11 residues for 3 turns). Due to the increasing popularity of this strategy, stapling chemistries that give rise to $\alpha$-helices have been extensively reviewed in recent years.[16, 58, 117–122] They can be grouped into one-and two-component strategies depending on whether the staple is formed using an external cross-linking agent or not. One-component strategies using natural amino acids consist mostly of lactam bridge formation, whilst those employing uAAs are dominated by all-hydrocarbon staples generated by RCM reactions between two alkene side-chains and "click" staples generated through CuAAC reactions between azide and alkyne side-chains, or thiol-ene reactions between thiol and alkene side-chains. Two-component strategies tend to rely on cross-linking two natural cysteine residues with a bifunctional linker, or on the use of double-click reactions coupling across two azide-containing uAAs. Over 75% of $\alpha$-helical stapled peptides in the PDB have a clear 'hydrophobic moment' with one helical face enriched in hydrophobic residues and one face enriched in hydrophilic residues (Table A.2).

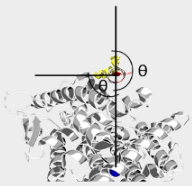
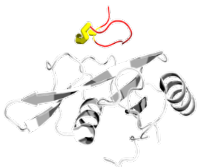
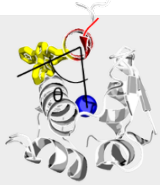
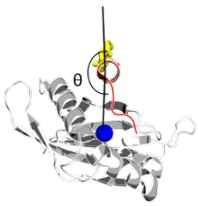
The introduction of a staple generally improves the therapeutic properties of individual peptides as well as their helicity, but increased helicity does not necessarily correlate with either an increase in binding affinity or in cell permeability.[123–126] Sim *et al.* demonstrated that a staple could change the dynamics of peptide-protein interactions not only through constraining the conformation of the peptide but also by altering the hydration properties.[127] In their computational study of the interactions of a stapled peptide with MDM2, they determined that a hydrophobic staple creates a confined space between the protein and the staple that traps waters, thus decreasing the 'dewetting' barrier of the binding event. However, there are only a very few studies in which the effect of different staple chemistries on helicity have been explored (e.g. de Araujo *et al.*,[128] Tian et al.[129]), with most studies to date based on the preferred stapling methodology of the chemist generating the peptide. Thus, it is not yet possible to conclude which is the "best" staple for stabilising the structure of $\alpha$-helical peptides. Whilst most staple chemistries have been optimised for helical peptides, some other motifs have been explored. For example, a series of stapled tankyrase inhibitors have been generated based upon a published 10 amino acid sequence; the most potent stapled peptide was shown to bind through the same key interacting residues, with a $\beta$-hairpin secondary structure.[97] Loops are another structural feature that can easily be achieved by multivalent linkers by thio-alkylation of cysteine residues as described by Brown *et al.*[130] and Timmerman *et al.*[131] No specific residue spacing is required and loops of different sizes can be obtained with this method; multi-functional cysteine cross-linkers permit the formation of multiple loops. Peptides in loop conformations display a larger and more complex contact area with the protein, and usually, the linker interacts directly with the protein surface.[97, 132]

### 1.3.2 Optimising biological activity

In general, the first step in optimising stapled peptides for a macromolecular target is the identification of appropriate sites for the incorporation of the non-natural amino acids used to form the cross-link. This is usually achieved by studying crystallographic, NMR or computational structural data; residues that are not involved in target recognition can be selected as potential sites for staple positioning, typically with i,i+3; i,i+4; or i,i+7 spacing depending on the staple chemistry to be employed. Biological or computational alanine scanning is frequently used to determine how to minimise the effect of introducing a staple on the desired biological activity.[96, 115, 133] Residues are successively mutated to alanine, and the residues conferring the least loss of activity upon mutation are chosen for the placement of staple residues. The length of linker used for the stapling process may be adjusted to match the optimum spacing identified by such scanning methods (e.g. stapling across residues spaced at i,i+7 rather than i,i+4).[41, 89, 134, 135]

Lau *et al.* investigated the p53/MDM2 interaction with peptides constrained using a double click staple chemistry. Variations in the staple position and length were explored with a library of 24 peptides and a competitive fluorescence anisotropy assay to detect peptide binding.[136] Correspondingly, Lalonde *et al.* scanned all possible lactam staple positions (i,i+4 and i,i+7) on a series of ghrelin peptides; testing their affinity for the growth hormone secretagogue receptor type 1a (GHS-R1a) to determine their potential use as imaging probes.[137] Examination of deposited stapled peptide sequences in the PDB (Table A.2) suggests there is no inherently preferred position for the staple in the overall amino acid sequence. Another design consideration is matching the hydrophobicity of the staple to the protein interface. If the protein interface presents an extended, open hydrophobic surface, energetically favourable contacts with the staple may be possible. This situation is particularly favourable for all-hydrocarbon staples formed through RCM. The estrogen receptor (ER) has an extensive hydrophobic binding interface, permitting hydrophobic interactions between the all-hydrocarbon staple of stapled SRC2-BCP1 peptide with the protein (Table 1.1, PDB code 5WGQ).[12] Similarly it was shown for the non-helical stapled bicyclic peptide Grb7-B4 binding to the Grb7-SH2 subdomain that interactions between the RCM staple and protein were required for high-affinity binding (Table 1.1, PDB code 5EEQ).[132] However, staple-protein hydrophobic contacts have been observed in other types of stapled peptides, e.g. to staples formed using a two-component double strain-promoted cyclisation with a cyclodiyne cross-linker which also have a large hydrophobic contact surface (Table 1.1, PDB code 5AFG).[138] In contrast, if the binding surface of the protein is narrower, or cleft-like, the staple may be directed towards the solvent. The binding interface of nuclear receptor coactivator 1 (NCOA1) has a more restricted surface and the hydrophobic staple of a stapled peptide derived from its binding partner YL-2 directly faces the solvent (Table 1.1, PDB code 5Y7W).[139]

**Table 1.1:** Examples of stapled-peptide protein interactions. Protein-peptide interaction PDB Code Staple Staple interaction with protein surface Staple angle

| Protein-peptide interaction | PDB Code | Staple | Staple interaction with protein surface | Staple angle θ[a] |
|---|---|---|---|---|
|  | 5WGQ | RCM i,i+4 and Lactam i,i+4 | Hydrophobic interaction of all-hydrocarbon staple with the surface; lactam staple faces the solvent | 93 and 175° |
|  | 5EEQ | RCM i,i+7 | Close contacts of bis-allyl serine staple with Met495, Asp496, Asp497 backbone, and side-chains of EF loop and Ile 518 of BG loop | ND[b] |
|  | 5AFG | Bis click i,i+7 | Hydrophobic interaction of cyclodiene staple with the surface | 34° |
|  | 5Y7W | RCM i,i+4 | No interaction with the protein: all-hydrocarbon staple faces the solvent | 162° |

### 1.3.3 Analysis of the primary sequences of PDB-deposited stapled peptides

An analysis of crystal structures in the PDB (Table A.2) indicates that staples can adopt a broad range of orientations towards the protein surface or solvent (Figure 1.7.a). The most frequent depositions arise from the use of all hydrocarbon staples at either i,i+4 or i,i+7 spacing (red and blue points, Figure 1.7.a). The surface area of the protein in contact with the stapled peptide varies from 700 Å$^2$ to 3000 Å$^2$, with an average value of 1550 Å$^2$. The mean value for the surface hydrophobicity was found to be close to zero, indicating that stapled peptides can bind both polar and hydrophobic surfaces. The staple angle ($\theta$) is defined as the angle from the centre of mass of the peptide, to two points representing the centre of mass of the protein, and the centre of mass of the staple residues (when all three are translated into the same plane as shown in Figure 1.7.b). This angle is thus an approximation of the angle between the side-chain of the stapled residues and a normal vector to the surface of the protein. In the dataset $\theta$ varies between 40° and 180°, with an average value of 103°. A weak correlation (0.34) is observed between this staple angle and the hydrophobicity of the protein surface to which it is bound. This relatively modest correlation is in part due to the difficulty in comparing bound

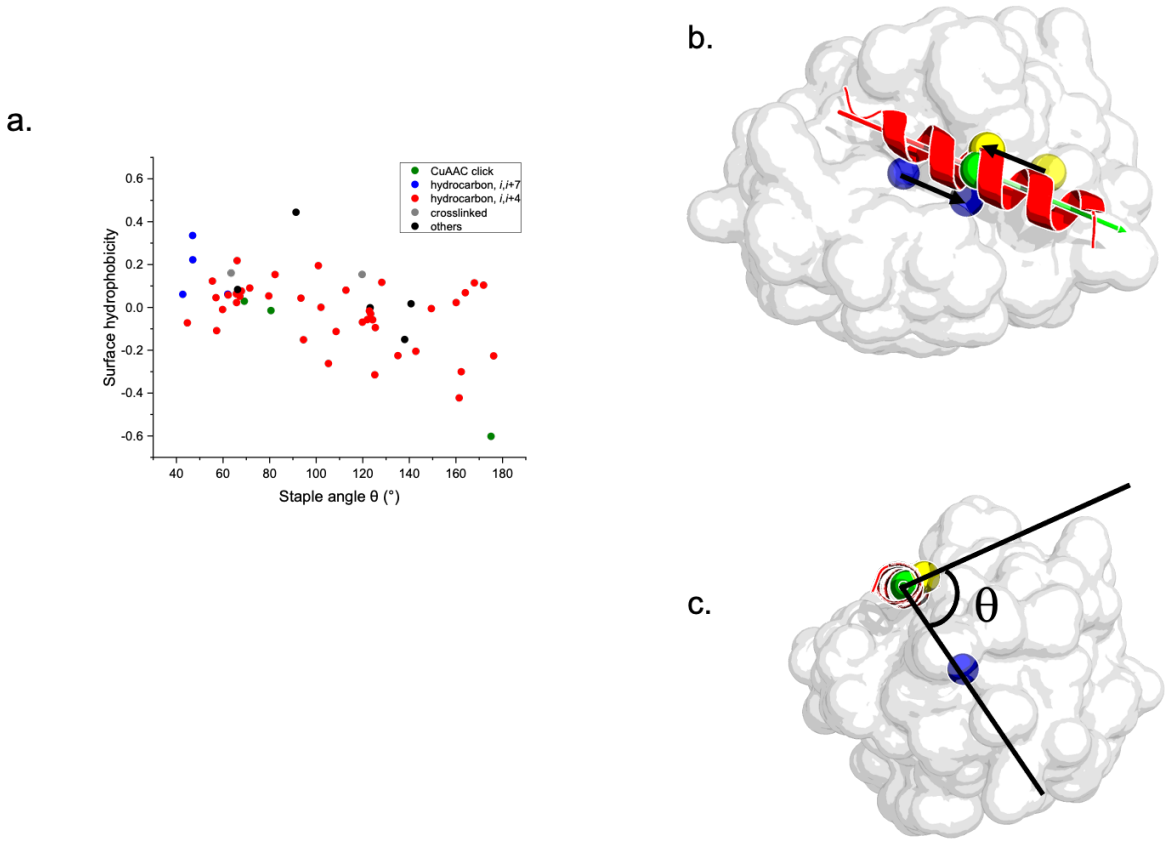



**Figure 1.7:** Analysis of the staple angle $\theta$ for 42 protein-bound, stapled peptide structures found in the PDB (see ESI). A Plot of staple angle $\theta$ against the protein surface hydrophobicity with staple residues coloured by staple type. Surface hydrophobicity was calculated using the solvent-accessible surface area (SASA) of each amino acid present at the binding surface, multiplied by its Eisenberg value (Table A.1)[140] divided by the total area of the binding surface. More positive values denote more hydrophobic interfaces. Staple angle $\theta$ was calculated by translation (black arrows) of the centre of mass of the protein (blue sphere) and the centre of mass of the staple residues (yellow sphere) to the plane defined by the vector of the helix (red/green axis) and the centre of mass of the helix (green sphere). Illustrated for PDB structure 3MK8 in B (i) and (ii).

stapled peptides, where proteins wrap around the stapled peptides differently, with one, two, or sometimes even three, interfaces with the stapled peptide. Thus the angle of the staple residues to the protein surface is not always directly equivalent to the angle made by the centre of mass of the staple residues to the centre of mass of the protein used in this analysis. Nevertheless, there is a trend which indicates that for all staple chemistries, the more hydrophobic the PPI surface is, the more the staple will tend to lean towards it (low $\theta$) and the more polar the surface, the more the staple will be oriented away from it (high $\theta$). Overall this indicates that consideration of the protein surface hydrophobicity may be a useful strategy in general, but detailed investigation of specific binding site features remains important to guide staple placement.

### 1.3.4 Ranking cell penetration by staple chemistry

In their study, Lau *et al.* developed helical peptides with nanomolar affinity to MDM2 using an i,i+7 double-click stapling technique. However, some of the most potent binders of MDM2 were found to be inactive in a p53 reporter cellular assay due to poor uptake.[136] Only a few studies have directly compared the effects of different types of staple chemistry on similar sequences, therefore it is difficult to generalise about the impact of staples on factors such as helicity, or cell uptake.[29] Maximising helicity has been one of the principal aims of many chemical studies, but it has been shown that flexibility in macrocyclic structures is often critical in allowing a peptide to cross the cellular lipid membrane.[141] To compare the effect of staple chemistries, Araujo *et al.* studied the stapling of a penta-alanine peptide and concluded that a lactam bridge leads to the most helical peptides, followed by hydrocarbon and triazole staples.[128] Tian *et al.* also concluded that for a 12-mer peptide sequence in an aqueous solution, the highest helicity was observed for lactam and all-hydrocarbon stapled peptides followed by triazole stapled peptides, while the helicity of a m-xylene bridged peptide was only increased by 10 % over the natural sequence.[129] But notably they discovered that the highest cell permeability was recorded for all-hydrocarbon and perfluorobenzene stapled peptides, which also have the highest hydrophobicity; a result which was consistent across four cell lines tested. Muppidi *et al.* used a series of aryl and vinylaryl groups to cross link i,i+7 cysteine residues by alkylation; they also found a correlation between cell permeability and hydrophobicity.[142] Finally, Nielsen *et al.* found that whilst orally bioavailable cyclic peptides frequently violate the "Rule of 5" for small molecules with a MW >500, their lipophilicity still tends to conform, with values in the range LogP = 0 to 5.147

## 1.4 Adding functionality to stapled peptides

Functionality may be readily added to stapled peptides through the addition of further residues to the primary sequence. For example, Dougherty *et al.* have shown that the addition of cell-penetrating peptide (CPP) sequences, can be used to enhance the cellular uptake of stapled peptides.[143] Addition of a cyclic CPP to the C-terminus of a stapled peptide which binds MDM2, resulted in improved cytosolic delivery and cellular $EC_50$ values equivalent to the small molecule inhibitor, nutlin 3a. Alternatively, the addition of E3 ligase targeting peptide sequences to stapled peptides has been explored by Jiang *et al.*, in the construction of a proteolysis-targeting chimera (PROTAC).[144] TD-PERM, a stapled peptide with proven activity against Estrogen Receptor $\alpha$ (ER$\alpha$), was linked to a pentapeptide which binds the Von Hippel–Lindau (VHL) E3 ligase. The resultant hetero bi-functional TD-PROTAC induced degradation of ER$\alpha$ in a proteasome-dependent manner inhibited the proliferation of ER$\alpha$-positive breast cancer cells and led to tumour regression in an MCF-7 mouse xenograft model. Successful strategies for functionalising the staple component of a stapled peptide require

chemistry which is orthogonal to that of the peptide. It is generally easier to introduce additional functionality to staples formed using a bi-component strategy (cysteine cross-linking, double CuAAC etc.) than their mono-component counterparts (all-hydrocarbon RCM, mono CuAAC etc.). Staple functionalisation may be used to add functional groups to track, or solubilise, the stapled peptide (Figure 1.8.a); to modulate the activity of the stapled peptide (1.8.b); or to polymerise the peptide, giving rise to multivalent species (Figure 1.8.c).

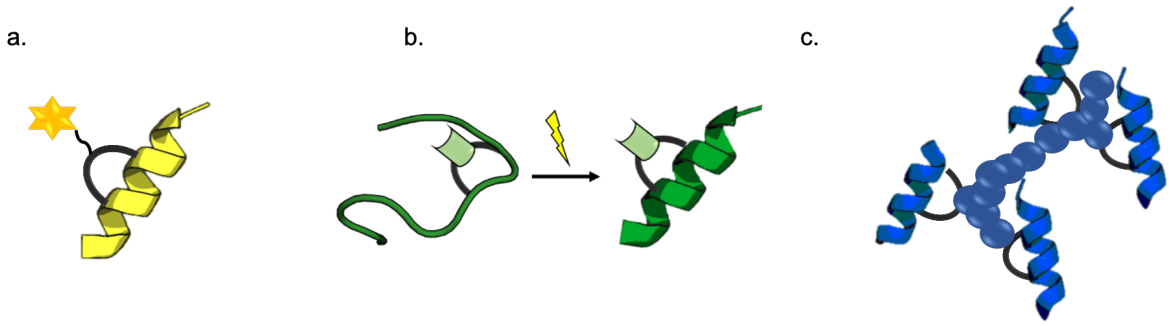



**Figure 1.8:** Adding functionality to the stapled peptide. A Attachment of functional groups. B Modulation of stapled peptide function. C Stapled peptide polymerisation.

### 1.4.1 Attachment of functional groups

Fluorescence is often used to follow the uptake of peptides into cell and tissues and a range of methods have been developed to attach fluorescent probes to peptides.[145] However, it has been shown than the fluorescent moiety itself can induce changes in peptide uptake.[146] Todorovic *et al.* have developed a one-pot stapling strategy which couples a cysteine and an amine-terminated amino acid in the presence of ortho-phthalaldehydes forming an isoindole-bridged peptide which is inherently fluorescent (Figure 1.9.a).[20] Functionalisation is not limited to fluorescence, and Wu *et al.*, describe a library of functionalised dialkynyl linkers compatible with double CuAAC stapling of peptides (Figure 1.9.b); these include linkers with additional non-peptidic polyamines and nuclear targeting peptides.[**?** ] It was shown that in vitro binding affinity for MDM2 was not significantly affected by the linker chemistry; but positively charged residues enhanced both cellular uptake and target engagement in cell reporter assays. Two different approaches to modifying the pharmacokinetics of stapled peptides through staple PEG-ylation have been explored. Xiao *et al.* used a one-component strategy to explore the effects of PEG-ylation on the conformational stability of the $\beta$-sheet WW domain, using amino acids containing both PEG motifs and an alkene function and RCM to cross-link the staple.[147] In contrast, Tian *et al.* used a bis-cysteine alkylation strategy and a linker containing the PEG motif to probe the dual agonist activity of a series of stapled peptides against GLP-1R and GCGR (Figure 1.9.c).[148] A large selection of chemistries have been adapted to peptides and allow some modulation for specific added functionality. Vasco

*et al.* developed an Ugi reaction operating on modified side-chains can both stabilize an $\alpha$-turn and introduce N-functionalization on the lactam bridge (Figure 1.9.d).[149] Assem *et al.* reported the chemical linking of thiol by acetone further functionalised on the ketone moiety with diverse molecular tags by oxime ligation (Figure 1.9.e).[150] In addition to stabilizing helical structures, this functionalisation provide opportunities to improve both peptide-target interactions and other peptide properties.

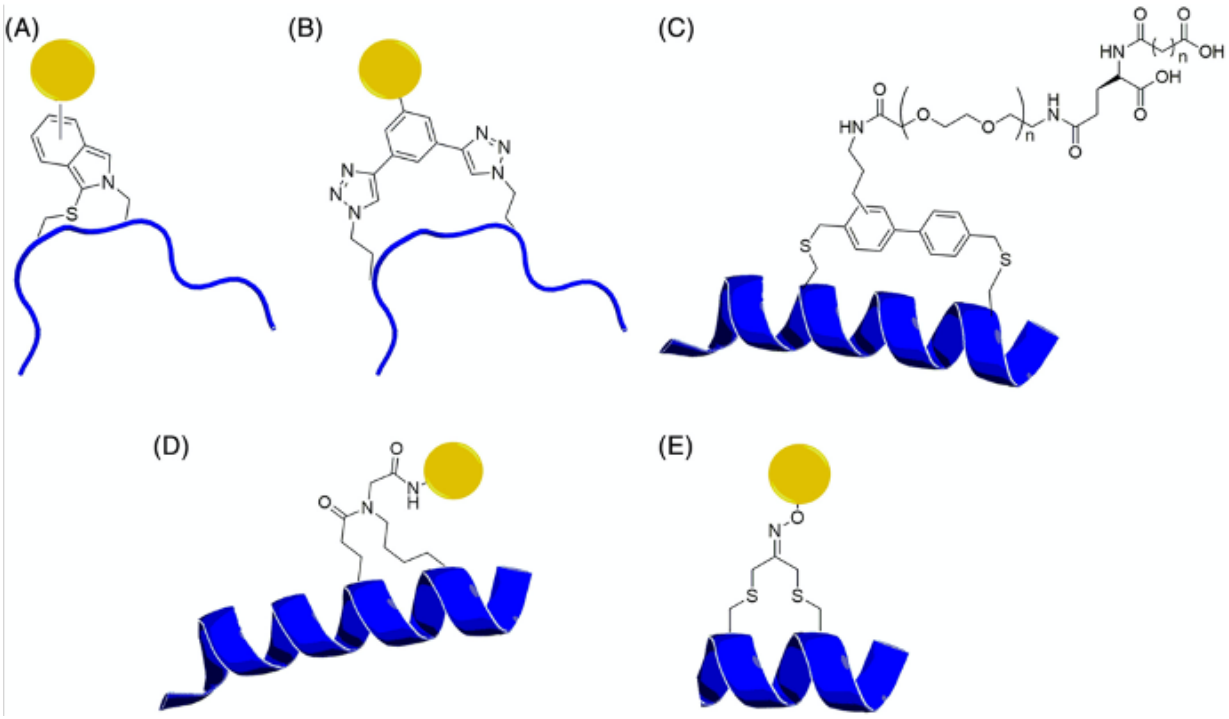



**Figure 1.9:** Attachment of functional groups using: A. an isoindole bridging group; B double CuAAC stapling; C bis-cysteine alkylation; D an Ugi multicomponent reaction; and E acetone crosslinking with oxime functionalisation.

## 1.4.2 Modulation of stapled peptide function

The helicity of stapled peptides can be made photoswitchable by incorporating a photo-isomerisable group into the staple itself. UV or visible light irradiation enables both spatial and temporal control in investigating the effects of helicity on, for example, target binding.[151] Brendenbeck *et al.* reported the first azobenzene-based photoswitchable stapled peptide,[152] in which the central azobenzene linker was coupled to i,i+7 spaced cysteine residues via two 2-iodoacetamide groups. In a related approach, Hoppmann *et al.* developed a photoswitchable azo-benzene based click amino acid which could be incorporated into peptide sequences and stapled to a cysteine residue at the i+4 position using thiol-ene chemistry.[153] In an alternative approach to photoswitching, Madden *et al.* developed a non-reversible staple which can be formed in situ by a nitrile imine-mediated cycloaddition reaction when peptides are irradiated at 302 nm.[154] The benzyl-pyrazole group produced by this stapling process is fluorescent which allows it to be tracked in cells.[154]

### 1.4.3 Stapled peptide oligo-/polymerisation

Oligo-/polymerisation enhances the protease resistance of stapled peptides and provides multivalent interactions; hence several different strategies have been reported. Tran *et al.* have developed a CuAAC-based polymerisation approach in which 1,3,5-triethynylbenzene is used to staple across i,i+7 spaced azido residues on the peptide. This leaves a further alkyne on the staple which can be dimerised using a bis-azido linker and a second CuACC reaction.[155] In an alternative approach, Lee *et al.* have developed a 'stapling polymerisation' procedure which allows the synthesis of multimers with 3-16 embedded helical peptides.[156] Acryloyl groups were added to the side-chains of lysine residues at i,i+7 positions in the peptide sequence and free radical polymerisation was achieved in the presence of acrylamide monomer.

## 1.5 Enhancing therapeutic delivery

The four main routes of drug administration are oral, intravenous and subcutaneous injection, and transdermal delivery (Figure 1.10). Oral and transdermal delivery are often considered as the best routes of administration as they are painless, non-invasive and do not require trained medical personnel.[157] Most peptide therapeutics are subject to rapid proteolysis, which explains their generally poor pharmacokinetics; proteases are found throughout the body, but mostly in the gastrointestinal (GI) tract, liver, kidneys, and blood.[122] Major contributors to metabolism (e.g. trypsin and chymotrypsin) generally hydrolyse peptide bonds between specific amino acids; thus to improve short half-lives mutation and/or rearrangement of the primary amino acid sequence can be performed to avoid protease-cleavage sites. Kim *et al.* introduced single amino acid substitutions in GNU derivatives and produced AMPs resistant to proteases.[158] Tools such as Pepcutter,[159] which uses the initial sequence and a library of the most common proteases and their associated cleavage sites to predict the susceptibility of a peptide towards hydrolysis, can be used to optimise the primary sequence. Constraining a peptide through stapling has been shown to enhance the stability of peptides towards proteolysis in several studies.[160, 161] A second challenge for hydrophilic peptide therapeutics is rapid renal clearance from the blood, which prevents efficient uptake in tissues. Binding to albumin and other proteins reduces renal filtration, and peptide conjugation to PEG, or a lipophilic moiety, can also improve serum half-life. Pessi *et al.* produced a doubly lactam-bridged stapled peptide conjugated to cholesterol which acts as fusion inhibitor for the Ebola virus;[162] they reported an extended serum half-life for the cholesterol-conjugated peptide allowing daily dosing. Finally, the amino acid sequence of a peptide affects its propensity to aggregate, to form either amorphous or amyloid-like fibrils; a number of programmes are now available to predict this, based on contributing factors such as hydrophobicity, charge state, and $\beta$-sheet forming propensity. [163–166]

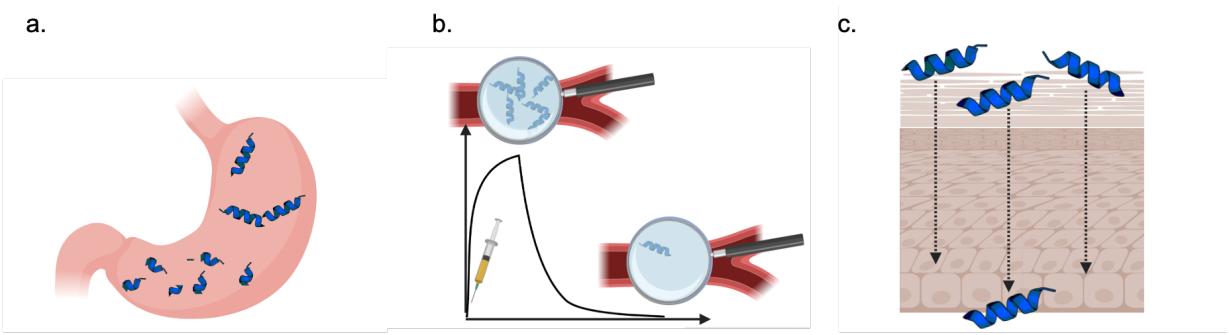



**Figure 1.10:** Barriers to the therapeutic delivery of stapled peptides. A Peptides administered orally are subject to low pH and degradation by proteases; B Peptides administered intravenously appear as a high initial burst, but are subject to degradation by serum proteases. C Peptides administered transdermally must pass multiple layers of the epithelium.

### 1.5.1 Oral delivery

Oral delivery is the most frequently employed drug delivery mode, but it is also the most challenging for peptides (Figure 1.10.a). Under the acidic conditions of the GI tract (as low as pH 1.0 – 3.0) some amino acids form by-products, for example cysteine, methionine and tryptophan are all susceptible to oxidation. Substitution of these sensitive amino acids (e.g. norleucine for methionine, and serine for cysteine) can increase peptide stability whilst retaining activity. N-terminal glutamine is also unstable at low pH and replacing this amino acid, or using an acetate capping group, can prevent its cyclisation.[167, 168] Orally available peptides drugs have on average 8.7 HBA and 4.5 HBD;[169] it is thought that the highly acidic conditions of the GI tract disrupt backbone hydrogen bonding, destabilizing helical and beta-sheet conformations and making peptides more susceptible to proteases. The number of rotatable bonds in a peptide is also inversely correlated with oral bioavailability and most successful orally-available peptides marketed to date are cyclic peptides (e.g. cyclosporine).[170, 171] In the case of stapled peptides, the improved stability arising from side-chain cross-linking can enable absorption, as seen with a 36-residue double-stapled peptide following oral gavage.[172] Finally, enhanced lipophilicity can be used to increase the half-lives of peptides; fatty acid conjugation of GLP-1 analogues has been shown to increase their half-life from 10-12 h to days.[138]

As a second challenge, the gut epithelium is composed of multiple layers of cells and mucus, which the peptide therapeutic must cross. Four modes of transport are available: passive diffusion through the cells (transcellular) and through intercellular junctions (paracellular); by endocytosis followed by endosomal release; or by receptor-mediated mechanisms.[173] Peptides with MW >700 Da have only limited paracellular permeation in intestinal epithelia,[174, 175] and an upper limit to permeation has been suggested as MW 3.5 kDa by Rubas *et al.*[176, 177] However, Ji *et al.* have reported the oral absorption of a 5.3 kDa cyclotide targeting intracellular MDM2.[178] The doses used in this study were high (40 mg/kg), and the peptide was administered in 5% dextrose, which is known to enhance intestinal perme-

ability; but evidence of activity after oral administration against this intracellular target was reported. The MW of current oral peptide therapeutics is typically around 800 Da and does not generally exceed 2000 Da.[169]

### 1.5.2 Intravenous and subcutaneous injection

Intravenous or subcutaneous injections currently represent the main delivery method for peptide therapeutics.[179] Drugs injected intravenously have a high initial bioavailability (Figure 1.10.b) as this is defined as the proportion of a drug that is available in the bloodstream. Subcutaneous injection does lead to a slightly greater variability in bioavailability, since peptides with a higher molecular weight have been linked to a greater absorption by the lymphatic system rather than the blood.[180] Overall, the most notable limitation of these delivery methods is the initial 'burst' encountered when the peptide is injected, which is followed by the rapid degradation of peptide.

### 1.5.3 Transdermal delivery

Transdermal applications include topical (skin), buccal (mouth), corneal and intranasal delivery; as for oral administration, one of the major challenges is transport of the stapled peptide through multiple layers of cells (Figure 1.10.c). Enhanced cell penetration often increases trans epithelial bioavailability.[175, 181] Some transdermal drugs such as cortisone, or anti-inflammatory drugs are limited to a local distribution usually for skin or tissue-related conditions. However there are many cases where application of a drug to the skin have been proven efficient for systemic delivery (nicotine, oestrogen). Despite typically low absorption, dermal administration avoids the first-pass liver metabolism that occurs for oral drugs. The stratum corneum, constituting the top layer of the skin limits the transdermal absorption of polar and large peptides.[182, 183] The lipid structure of the stratum corneum can be disrupted by natural peptides (e.g. magainin) when these are applied in combination with a surfactant chemical enhancer leading to enhanced skin permeation.[183] Several hundred peptides that enhance the skin permeability of drugs are collated in the database TopicalPdb.[184] Among the reported sequences, arginine and lysine combined with lipophilic natural and unnatural amino acids are found extensively. Chemical modifications of peptides, such as the addition of lipophilic moieties, have also been shown to increase stability and skin permeability.[162, 185, 186]

Transmucosal nasal drug delivery has also emerged as an important field in drug delivery technology due to the high vascularity and large surface area, which enable drug uptake.[175] Intranasal delivery can allow the absorption of peptides of up to 2 kDa without absorption enhancers, and peptides of 2-6 kDa with absorption enhancers.[187] Peptides can also be used as modulators to facilitate the uptake of drugs; in general, cell penetrant peptides display excellent absorption.[188] Peptide drugs on the market using intranasal delivery include

desmopressin, a cyclic peptide used to treat diabetes insipidus. However, intranasal delivery is not appropriate for the treatment of infants, for whom alternative delivery methods, such as buccal delivery, must be sought.[189] Although stapled peptides have not yet entered clinical trials using intranasal delivery, several stapled peptides have shown efficacy in mouse models. Gaillard *et al.* reported a double-stapled RSV peptide, which disrupts viral cell entry and significantly decreases pulmonary infection in mice following intranasal delivery.[41]

## 1.6 Conclusion

This chapter has focused on the many considerations required for the design of biologically active stapled peptides, drawing from the extensive recent literature, and from analysis of PDB deposited structures containing protein-bound stapled peptides. The growing recognition of the utility of stapled peptides is fuelling the development of bespoke computational tools, synthesis procedures and biological assays, that are accelerating their adoption as a research tool. With dozens of new publications every month from research groups across the world, and several stapled peptides entering clinical trials, there is ample evidence that stapled peptide technologies will play an increasingly important role in pharmaceutical applications.

# 2

# From Quantum Mechanics to Molecular Dynamics, Applications of Molecular Dynamics to Peptide Folding

## 2.1 Quantum Mechanics. Basis and Approximations in the SCF method.

Quantum Mechanics (QM) methods refers to a set of methods that provides a description of the physical properties of molecules at the scale of atoms and subatomic particles. In QM, energy, momentum, angular momentum, and other quantities of a system are restricted to discrete values, that satisfy the electronic Schrödinger equation (Equation 2.1). This equation allows the calculation of the potential energy surface of a particle.

$$\hat{\mathrm{H}}\,|\Psi\rangle = E|\Psi\rangle \tag{2.1}$$

Equation 2.1 is the time independent Schrödinger equation, where $\Psi$ is the state vector of the quantum system, and $\hat{H}$ the Hamiltonian operator which can be an observable (mass, position or velocity) or an eigenvalue. E is the energy of the system. Using the time-independent Schrödinger equation enables the use of simpler wave-functions. Generally, the real wave function of a system is too complex to be calculated directly, but wave functions can form standing waves, called stationary states, which are a quantum states with all observables independent of time. The self-consistent field method (SCF) is a method which permits the determination of the the wave function and the energy of a quantum many-body system in a stationary state. In this method an initial Hamiltonian is approximated. The Hartree-Fock (HF) method, which is a SCF method relies on several approximations:

- The Born–Oppenheimer approximation is assumed: the molecular orbital approximation which stipulates that electron movement can be studied separately to the nuclei movement.
- The electronic Hamiltonian is expressed as a sum of the one-electron Hamiltonian, and the overall wave-function is a product of one-electron wave-functions.

- The mean-field approximation is used. In a multi-body system an averaged spin is calculated and used instead of calculating the pairwise interactions of each particle spin separately. The Hamiltonian each particle is simplified using the mean-field term. This method allows both Coulomb and Fermi electron-electron correlations to be accounted for using the same level of complexity as for particle in gas phase, or non-interacting particle.

- Each energy eigenfunction is assumed to be described by a single Slater determinant. And the variational solution is assumed to be a linear combination of a finite number of basis functions.

Most theories in classical physics can be derived from QM, often relying on a number of approximations, that allow the study of macro-scaled systems, such as proteins. In Molecular Dynamics, potential energies are analytically calculated by the derivation of values obtained from Quantum Mechanics. Although more accurate basis sets have been developed, Hartree-Fock is still widely used for the parameterisation of molecules. Basis set are set of functions used to represent the electronic wave function in the Hartree–Fock method. Multiple basis sets can be used, such as the STO basis sets, the Pople basis sets, the correlation-consistent basis sets. In the Pople basis set '6-31G*', 6 is the number of primitive Gaussian functions used to represent the wave function of the core electrons, and '31' indicates that the valence orbitals are composed of two basis functions each, the first one composed of a linear combination of 3 primitive Gaussian functions, the other one composed of one primitive Gaussian function. Interestingly Kuyper *et al.* suggested that the HF/6-31G* "over-polarises" bond dipoles in gas phase, and the resulting charge distribution approximates the atom polarisation occurring in an aqueous phase. These potential energies have been found to reproduce the relative free energies of solvation found experimentally.[190] Ever since, the mainstream versions of the AMBER forcefield have been consistently derived from the Hartree Fock method, using the 6-31G* basis set.[191, 192] In this work we used the HF/6-31G* basis set, (Chapter 3), to derive forcefield parameters, for consistency and compatibility within the current forcefields.

## 2.2 Principles of Molecular Dynamics

Molecular Dynamics (MD) simulations represent a class of established computer methods to assess molecular properties. In large and complex bio-molecules, such as proteins, several phenomena can be studied using MD, such as loop motions, the opening of cavities, the stability of inter-molecular interactions, or the interactions of the system with solvent molecules.[193–195] Other common usages of MD include the derivation of kinetic and thermodynamic properties for free energy estimations, (e.g. of a conformational change or upon binding of a ligand)[196] and the generation of conformational ensembles distributed according to statistical distribution functions.[197] In this Section we present the principles of MD

simulations and forcefields.

MD methods rely on the Born Oppenheimer approximation of the Schrödinger's equation which separates vibrational, rotational and electronic energies. Molecular Mechanics calculate the motions and energy of molecules based on nuclear coordinates only, as electrons are not considered explicitly. MD relies on the assumption that the nuclei of atoms are heavier, move slower and that electrons can, in effect, adjust instantly to any nuclei movement. Thus in MD each atom (or group of atoms in coarse models) is modelled as a single particle. A MD system consists of a set of particles that move through space in response to their interaction which each other. In the simulated system, every particle is given a velocity, mass and position. This allows to compute a potential energy surface. In Molecular Mechanics an analytical, simplified representation of the inter-atomic energies is used, called forcefield. The potential energy of a MD system is attributed using the forcefield parameters and a set of spatial coordinates. The basic functional form of potential energy in MD includes bonded terms for interactions of atoms linked by covalent bonds and non-bonded terms (also termed non-covalent) that describe the long-range electrostatic and van der Waals forces. In a forcefield-based MD simulation, the movement of atoms is constrained to replicate the natural behaviour of molecules, and these rules are defined by the forcefields (see Section 2.3). Thus MD rely heavily on the correct parameterisation of the forcefields.

The most common class of forcefields are the non-polarisable forcefields which are also called additive forcefields. Given their relative performance and accuracy, they have been used prominently for the last 40 years. In this thesis, only additive forcefields have been exploited. As a very brief mention, two main other classes of forcefields currently show promising developments. Polarisable forcefields are adapted to the simulation of metal complexes and the exploration of electrostatics (e.g. electric fields). However, they require much more computing time to simulate a similar size system than additive forcefields.[198] More recently, machine learning forcefields have made some considerable improvements, but they are still limited to relatively small systems, and more systematic validation is needed for them to be used routinely for simulating bio-molecules.[199]

Molecular dynamics is a deterministic technique: the evolution of a simulated system in time is, in principle, completely governed by the given initial atomic coordinates and velocities. However, in practice two simulations will most often diverge and produce different outcomes. This is due to two principal phenomenons. The first being that several algorithms used in MD simulations rely on the generation of random numbers (e. g. in the generation of the initial atoms velocities, the Monte Carlo barostats and Anderson thermostats), this could potentially result in different trajectories. However, the generation of these 'random' numbers is generally achieved using random number generators derived from cryptography applications. A key (a seed) turns a highly regular counter into a reproducible stream of random numbers.[200] This makes it possible to quickly set the counter in the random engine based, e.g. on the time-step and atom index, and obtain the same random numbers stream, but in practice this

sequence is also dependent on the processor used for the simulation, thus the stream will only be reproduced if ran on a computer with an identical processor and operating system. The second aspect of the non-reproducibility in MD simulations, results from the sequential round-off of errors during the integration of the system. For longer simulations, successive small differences during the integration step will lead to diverging simulations, and ultimately to chaotic behaviors. Some MD engine can use double precision, (which double the length of float numbers) to counteract this effect in part.

### 2.2.1 Integrators

Particles in MD follow Newton's equations of motion to model to the fullest extent the physical movements of atoms and molecules, giving a view of the dynamic "evolution" of the system. Newton's second law links the acceleration and the forces exerted on a particle as seen in Equation 2.2

$$a_i = -\frac{F_i}{m_i} \tag{2.2}$$

In equation 2.2, $F_i$ being the forces applied into the particle i and $m_i$ being the mass of the particle.

Numerous algorithms exist for integrating the equations of motion. Many of these are finite difference methods in which the integration is partitioned into small steps. Each step is separated in time by a specific period called time-step. An initial configuration of the bio-molecule or system under study defines the positions of each particle. The Verlet algorithm is the most commonly used algorithm to integrate Newton equations. Even if more sophisticated algorithms have been developed since, Verlet algorithms are still widely used and are generally sufficient to produce acceptable motions. The algorithm considers that the positions of the particles in the system can be approximated by a Taylor series:

$$r_i(t+\delta t) = r_i(t) + v_i(t)\delta t + \frac{a_i(t)}{2}\delta t^2 + \mathcal{O}\delta t^3 \tag{2.3}$$

Where $r_i$ $v_i$, and $a_i$ denote the i–th particle's position, its velocity, and its acceleration. t is the time, and $\delta$t is the time-step and $\mathcal{O}\delta t^3$ is the rest of the Taylor serie, which is neglected. Similarly, positions for time t − $\delta$t can then be defined as:

$$r_i(t-\delta t) = r_i(t) - v_i(t)\delta t + \frac{a_i(t)}{2}\delta t^2 + \mathcal{O}\delta t^3 \tag{2.4}$$

The summation of Equations 2.3 and 2.4 gives :

$$r_i(t+\delta t) = r_i(t-\delta t) + 2r_i(t) + a_i(t)\delta t^2 + \mathcal{O}\delta t^3 \tag{2.5}$$

In Equation 2.5, we can observe that the position of the atoms at time t + $\delta$t, can easily

be deduced from their acceleration and their positions at time t - $\delta$t and t. Thus, combining equations 2.2 and 2.5 connects the forces applied to a system and the movement of its particles.

### 2.2.2 Time-step selection

A straightforward approach to decrease the computational cost of a MD simulation is to increase the time-step, therefore reducing the number of integrations needed to simulate motions over the same duration. However the time-step $\delta t$ must be selected carefully. Long time-steps allow 'faster' MD, but also provide numerical instabilities. If the time-step is longer than the timescale of the fastest motion, the energy of the system will not be conserved. The vibrational time period associated with the stretching of the bond between carbon and hydrogen might become smaller than the time-step, which is a major issue when calculating hydrogen positioning. Typically time-steps need to be approximatly ten times smaller than the fastest frequency for expansions, for atom positions to be reasonably accurate. Constraining the vibration of carbon-hydrogen bond lengths is usually employed to counterbalance this effect. Constraints can also be applied on all bonds to permit the use of even longer time-steps. The application of angle constraints may permit a further small increase in the time-step but is generally not recommended. SHAKE, and LINCS algorithms are among the most popular methods for constraining bonds.[201, 202]

Usually, hydrogen mass re-partition (HMR) or virtual interaction sites ($V_{sites}$) are used with larger time-steps. The method of hydrogen mass re-partitioning (HMR) is a potentially useful tool for accelerating MD simulations. By re-partitioning the mass of heavy atoms into the bonded hydrogen atoms, the highest-frequency motions of the macromolecule under study are slowed down, thus allowing the simulation's time-step to increase by up to 2 fs. An alternative method is to use so-called virtual interaction sites in conjunction with bond constraints. Here, the masses, charges, and velocities of the hydrogen atoms are redistributed to the adjacent heavy atoms. Integration of the equations of motion is performed based on the heavy atoms alone. Only after the positions of the heavy atoms have been updated are the hydrogen positions determined from the positions of the bonded heavy atoms.

### 2.2.3 Building the unit cell: Ensembles and Periodic Boundary Conditions

A thermodynamic system has several variables that describe its macroscopic states, such as the number of particles, volume, temperature, pressure, and total energy. Ensembles are artificial constructs fixing some of these variables while leaving others free to fluctuate. Usually, in MD the number of particles is fixed, and the three commonly used ensembles are described below. The constant-energy, constant-volume ensemble (NVE), is known as the micro-canonical ensemble. As its name suggests, the total energy of the system is fixed as well as its volume. It

is the easiest system to implement, but in practice, a constant energy ensemble does not relate to many real systems as energy is normally exchanged extrinsically. The canonical, constant-volume, constant-temperature, ensemble (NVT) allows control over both the temperature and volume. The unit cell vectors are fixed and the pressure can vary. The constant-temperature, constant-pressure ensemble (NPT) allows control over both the temperature and pressure. The unit cell vectors can change, and the pressure is maintained by adjusting the volume of the system. The pressure is controlled by a barostat. Multiple barostat algorithms exist, such as the Berendsen[203] and Anderson algorithms.[204] The NPT ensemble is the closest to experimental procedures.

Boundary conditions are used to define the borders of simulated system. They refer to the properties of the unit cell edges, upon interaction with the simulated system. Usually, for the simulations of bio-molecules in water or solvent, periodic boundary conditions (PBC) are implemented. With PBC, the surface of the unit cell is removed, assuming a continuous ensemble of identical replicas of the unit cells. In this system, an atom crossing the surface of a unit cell will be transposed to the opposite face assuring the modelling of an 'infinite' system.

## 2.3 Molecular dynamics forcefields: The case of the AMBER forcefield

The basic functional form of potential energy in molecular mechanics additive forcefields includes a sum (Equation 2.6) of bonded and non-bonded terms. The bonded terms describe interactions between pairs of atoms linked by covalent bonds, as well as angles and dihedrals. Non-bonded (non-covalent) terms describe electrostatic and Lennard-Jones energies, and a term for long-range interactions is usually added (PME or RF). Each type of potential is decomposed and summed individually.

$$V_{system} = V_{bonded} + V_{non-bonded} + V_{long-range} \tag{2.6}$$

Different additive forcefields use slightly different descriptors, the sections below are based on the AMBER forcefield formalism.

### 2.3.1 Non-bonded interactions

$$V_{non-bonded} = \sum_{pairs} V_{LennardJones} + \sum_{pairs} V_{Electrostatic} \tag{2.7}$$

The sum of two different energetic potentials, namely the steric (Lennard-Jones) potential and the electrostatic (Coulombic) potential describe the interactions between non-bonded atoms (Equation 2.7).

### Lennard-Jones potential

The Lennard-Jones potential is a simplified model that describes the essential interaction between two atoms: Two interacting particles repel each other at very close distances, attract each other at a moderate distance, and do not interact at infinite distance.

$$\sum_{j=1}^{N-1}\sum_{i=j+1}^{N}\left[A_{ij}\left(\frac{r_{ij,m}}{r_{ij}}\right)^{12}-2B_{ij}\left(\frac{r_{ij,m}}{r_{ij}}\right)^{6}\right] \tag{2.8}$$

Equation 2.8 refers to the form of the Lennard-Jones potential, used in the AMBER molecular forcefield. $A_{i,j}$ and $B_{i,j}$ depend on the atom type, and $r_{ij}$ is the distance between the atoms i and j. This potential has a minimum for an optimal inter-atomic distance $r_{m,ij}$ where the energy is at its minimum. After a certain atomic distance, the energy becomes asymptotic to the x-axis (Figure 2.1).

The Lennard-Jones potential is a pair potential, meaning that this potential is calculated between a pair of atoms. $A_{i,j}$ and $B_{i,j}$ can also be expressed as function of $\varepsilon$ (kcal.mol[1]) and $\sigma$ (Å) using the relationship in Equation 2.9, where $\varepsilon$ is the depth of the potential 'well' (usually referred to as 'dispersion energy', and $\sigma$ is the distance at which the particle-particle potential energy is zero (often referred to as 'size of the particle'). The Lennard-Jones potential has its minimum at a distance of $r = r_m = 2^{1/6}\sigma$, where the potential energy has the value $V = -\varepsilon$ .

$$A = 4\varepsilon_{ij}\sigma_{ij}^{12},\;\; B = 4\varepsilon_{ij}\sigma_{ij}^{6} \tag{2.9}$$

In AMBER forcefields atoms are given an atom type by taking into account their element, their hybridisation state and the atoms they are bonded to (e.g. $sp^3$ carbon in an aliphatic chain, $sp^2$ aromatic carbon ...). When building the model, $\varepsilon_{ij}$ and $\sigma_{ij}$, which are the values between two atoms with different atom types, are deduced from the values of $\varepsilon_{ii}$ and $\sigma_{ii}$, which are the values between two atoms with the same atom type. Lorentzt-Bertelot combining rules (Equations 2.10 and 2.11) are used to generate these parameters. AMBER then converts these $\varepsilon$ and $\sigma$ values into values of A and B prior to the simulation (Equation 2.9).

$$\sigma_{ij} = \frac{\sigma_{ii}+\sigma_{jj}}{2} \tag{2.10}$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}} \tag{2.11}$$

The Lennard-Jones potential as shown in Equation 2.8 is calculated for atoms separated by three or more covalent bonds. For pairs of atoms separated by three covalent bonds (so-called 1,4-interactions), the steric energy is frequently divided by two, and a dihedral torsional term is added. The latter will be described in Section 2.3.2

### Coulombic potential

As for the Lennard-Jones energy, the electrostatic potential is calculated between charged atoms separated by at least three covalent bonds. The potential is given by the Coulombic term (Equation 2.12).

$$\sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \tag{2.12}$$

In Equation 2.12, N is the total number of pairs of non-covalent covalent atoms (spaced by at least two covalent bonds) in the system, $q_i$ and $q_j$ are the partial atomic charges of the atoms i and j. $\varepsilon_0$ is the permittivity in vacuum, and $r_{ij}$ is the distance between the two atoms i and j. For additive forcefields, the partial charges are kept constant and are determined by the forcefield parameters set.

### Long range interactions

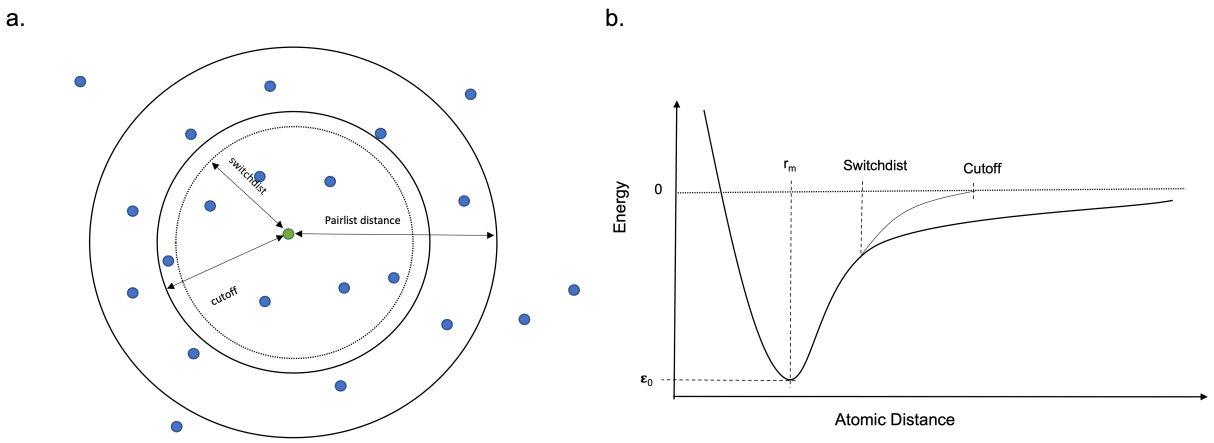



**Figure 2.1:** a. Representation of the pair list, cut-off and switch distances, b. Lennard Jones Potential with application of a switchdist and cutoff: the potential energy is smoothly bring to zero (dashed line) between the switchdist and the cutoff.

A non-bonded term is, in principle, calculated for all pairs of atoms. However, for systems with a large number of atoms, this is too computationally demanding. As non-bonded interactions become negligible with larger inter-atomic distances, Verlet lists and cut-offs are generally introduced to improve performance. Only pairs of atoms within a certain pair list distance are included in the Verlet list and will be considered for the calculation of the non-bonded interactions. A number of pairwise interactions are further excluded from the Verlet list using a cut-off. Above this cut-off distance, the non-bonded interactions are considered to be negligible and thus set to zero. To reduce abrupt energy changes near the cut-off distance, in some instances, a switching distance ('switchdist') is introduced, (Figure 2.1) to smoothly reduce the interaction to zero before the cut-off is reached. As the atoms move during the simulation, the Verlet list is updated after a certain number of MD steps (usually 5-10).

The treatment of the Coulombic term at long distances is less straightforward, as electrostatic interactions decay slowly compared to dispersive interactions. Different methods can be applied, such as the Particle Mesh Ewald (PME) or the Reaction Field method. In the PME method, electrostatics are treated differently depending on whether they occur at longer or shorter distances. While the short-range interactions are calculated with the direct summation, the longer-range distances are calculated using a summation in Fourier space. PME is a method that allows the approximation of this Ewald sum to be made numerically.

$$E_{long}(r) = \sum kU_{long}(k)|p(k)|^2 \tag{2.13}$$

In Equation 2.13, representing the long range interactions, $U$ is the Fourier transform of the potential and $p(k)$ is the Fourier transform of the charge density.

### 2.3.2 Bonded interactions

The bonded interactions are a combination of four terms:

$$V_{bonded} = V_{bond} + V_{angle} + V_{dihedral} + V_{improper} \tag{2.14}$$

They describe bond elongations, angle deformations and periodic and improper dihedral torsion angles.

#### Bond stretching term

The energy of a covalent bond between two atoms is calculated by analogy with a harmonic oscillator (Hooke's law) (Equation 2.15).

$$V_{bonds} = \sum_{i \in \text{bonds}} k_{bi}(l_i - l_i^0)^2 \tag{2.15}$$

In this equation, $l_i$ is the distance between two atoms, $l_i^0$ n is the equilibrium distance and $k_{bi}$ n is the force constant. The forcefield bonds parameters are predetermined for different atom types. Two terms are used in the bond potential: $k_{bi}$ and $l_i^0$ for each atom.

#### Bond-angle bending

The bending potential captures the energy of the angle deformation between three covalent atoms $i$, $j$ and $k$. Similarly to the bonded term, the bond-angle term is also calculated as an anharmonic potential penalising bond angles deviating from the equilibrium value $\theta_i^0$:

$$V_{angle} = \sum_{i \in \text{angles}} k_{ai}(\theta_i - \theta_i^0)^2 \tag{2.16}$$

In Equation 2.16, $\theta$ refers to the angle between three atoms, where $\theta_i^0$ is the reference angle where the energy is at its lowest, and $k_a^i$ is the associated force constant.

### Dihedral term

The third energy term concerns the dihedral angle of two planes defined by three covalent bonds and involving four atoms. The corresponding potential is defined as a truncated Fourier series expansion around the equilibrium dihedral angle (Pitzer's potential):

$$\sum_{i \in \text{torsions}} \sum_{n} \frac{1}{2} V_i^n [1 + \cos(n\omega_i - \gamma_i)] \tag{2.17}$$

Each Fourier series in Equation 2.17 represents a dihedral angle, where n is the order, $V_{i,n}$ the dihedral energy at its highest value, $\gamma$ the phase and $\omega$ the dihedral angle between the four atoms. A similar equation is used to calculate the energetic contribution of the improper dihedral angle potential.

In the recent version of the AMBER forcefield (AMBER19SB),[205] a new term was introduced to correct the torsional energy profile of backbone atoms. While the torsional energies in the side-chains are still calculated using the Pitzer potential, the backbone torsional energies are obtained through the use of corrections maps (CMAP) This approach is very similar to the CMAPs which have been used in the CHARMM forcefield for a longer period of time (CHARM22 and CHARM36).[206] In AMBER19SB $\psi$ and $\phi$ torsion terms are calculated *via* the CMAP term corresponding to their residue. As a side note, the Pitzer potential is still present for these torsions, but the $V_i^n$ are set to zero. CMAP correspond to a 3D energy surface function of the $\psi$ and $\phi$ angle values. The energy of a given $\phi, \psi$ couple is obtained by extrapolation. In Figure 2.2.a, the correction map of Leucine in the AMBER19SB is plotted. CMAPs are amino-acids specific, and the arginine and leucine CMAPs have been extracted from a topology file parameterised using AMBER19SB forcefield and compared in Figure 2.2.b.i and ii.

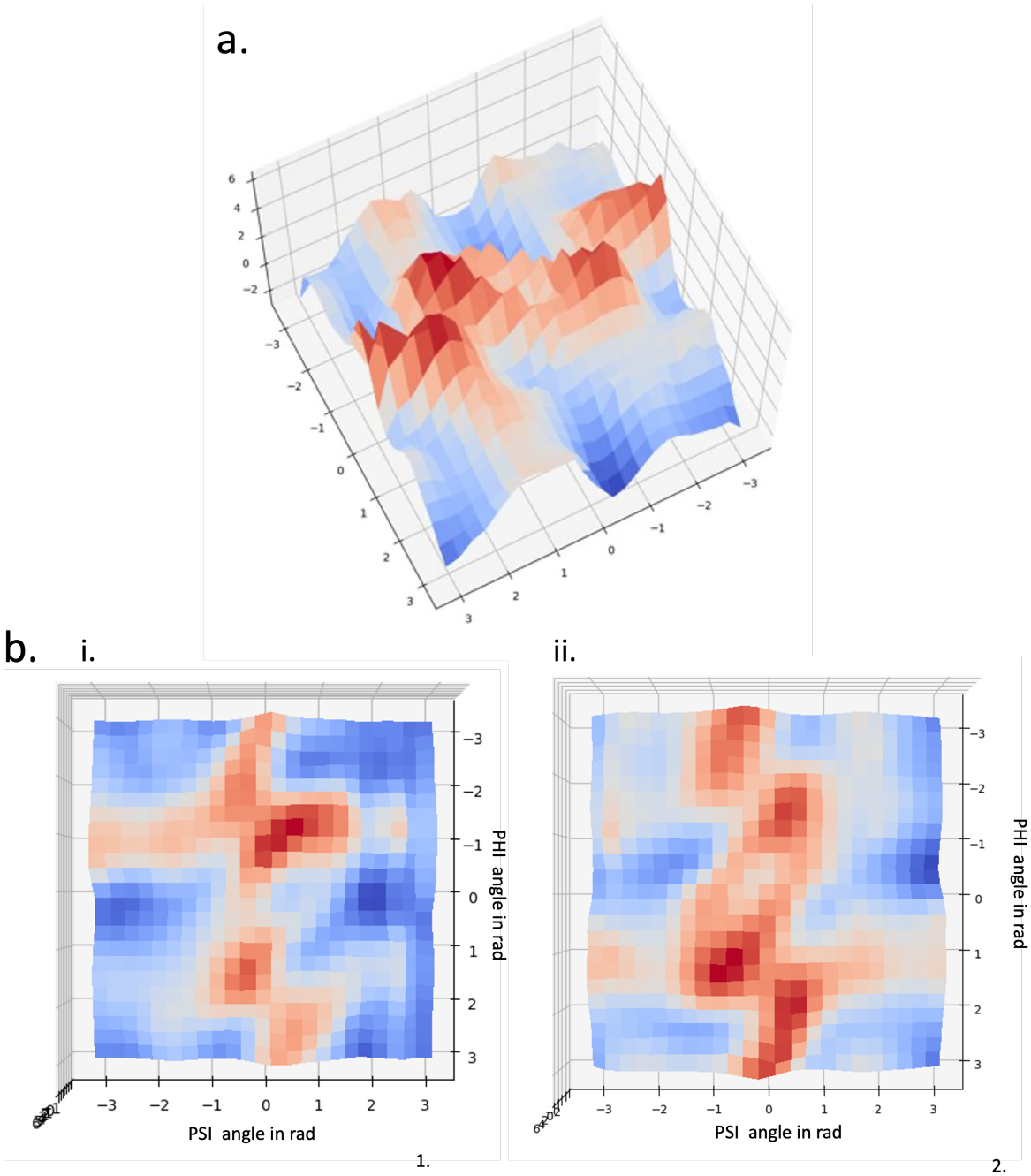



**Figure 2.2:** a. 3D plot of the Leucine CMAP (AMBERSS19SB forcefield) b. Comparison of the CMAPs of i. Arginine and ii. Leucine

## 2.4 Secondary structure prediction of bio-molecules

Secondary structure prediction is an important field of research. Initially, secondary structure prediction algorithms were developed to model missing parts in protein crystal structures. More recently, they have been applied to the prediction of peptide conformations. There have been a growing interest in the use of peptide therapeutics, so an accurate prediction of their secondary structure is useful as it is directly linked with their biological properties. Two main categories of prediction methods are currently in use: knowledge-based methods and physics-based methods.

### 2.4.1 Knowledge-based methods

One approach to predict peptide secondary structure relies on a statistical approach based on the calculation of propensities of each residue to form an $\alpha$-helix or $\beta$-strand. This idea was initiated by the work of Chou and Fasman in the 70s and had a very low accuracy (around 50 %).[207] Amino acid residues appear with different frequencies in bio-molecules. 'Secondary structure potentials' are derived from residue frequencies in the structural elements of known structures. These 'secondary structure potentials' (e.g. $\alpha$-helix, or coil forming potential) are then assigned to the residues of a sequence to deduce its secondary structure.[207–209] This method was improved by Deleage *et. al* by assigning first the secondary structure class (e.g., all-$\alpha$, all-$\beta$), followed by further refinement using specific parameters dependent on the given secondary structure class.[210, 211]

Other methods are based on the observation that polar and apolar residues frequently form a periodic pattern.[212] For example, alternation of polar and apolar residues often indicates that $\beta$-strands are formed. Foremost the hydrophobic residues on one side of a $\beta$-strand interact with the other hydrophobic residues located in the core of the protein, while the residues forming a hydrophilic side are facing the solvent. Likewise, apolar residues in every third or fourth position suggest the presence of an $\alpha$-helix, where the apolar side of the helix is facing the hydrophobic core of the protein. However, these secondary structure predictins preferences are not strong enough on a single residue level. The occurrence of a pattern comprising several residues preferring one type of secondary structure element suggests the formation of such structures.[213] The accuracy of prediction methods on peptide sequences is about 65%. Since these predictions are performed using single residues or small segments, the effect of long-range interactions is often ignored. Neglecting long-range interactions is often considered as a limiting factor of the accuracy of secondary structure prediction methods.[214]

Machine learning-based approaches, which learn "by examples", have shown unprecedented success for the prediction of secondary structures. With the expansion of crystallographic structures available in the last 30 years, notably in the Protein Data Bank (PDB), diverse implementations of machine learning algorithms have been reported, including Neu-

ral Networks,[215] Hidden Markov Models,[216] and Support Vector Machines.[217] Hidden Markov models were one of the first methods to appear and use predicted local structures for fold recognition. The essential principle is to create a model that satisfies the most observed behaviour of an amino acid considering the adjacent residues.[218] The Garnier–Osguthorpe–Robson (GOR) method considers windows of 17 positions and compiles the expected conformation of an amino acid, including its 16 neighbouring residues (8 in each direction). Probability functions are built to yield the final state assignment of the secondary structure for the sequence. The most successful approaches using Hidden Markov Models achieve an accuracy ranging from 63% to 71% when predicting secondary structure.[160] In neural networks, each 'neuron' receives many simultaneous inputs and computes a weighted sum to assign the prediction. The weights are adjusted based on training data sets, which can include features such as packing patterns, or long range effects of residues spaced further along in the sequence.[219, 220] The most accurate neural methods achieve a prediction accuracy of around 72 %. Support-vector machines are supervised machine learning models and are another promising approach to predict the secondary structure of bio-molecules.[221]

Since the 1990s, several databases have been produced to gather information on peptides, such as their helicity and cell penetration. Recent examples of such databases include StraPep, a database for bioactive peptides;[222] SATPdb, a database for therapeutic peptides;[223] Neuropep, a database for neuropeptides;[117] and CPPsite, a database for cell-penetrating peptides.[36] The main limitation of empirical methods is that they rely on previous knowledge available in these databases. Hence, they are not suited to predict the secondary structure of sequences containing unnatural amino acids, for which only limited data is available. Hybrid approaches such as PEPstrMOD couple results from machine learning methods with molecular dynamics.[224] Based firstly on a knowledge-based estimation of the propensity of a peptide to form an $\alpha$-helix, a refinement is made using a forcefield. This approach has seen promising results when applied to D-amino acids, cyclic peptides, and peptides with variable capping groups.

### 2.4.2 Physics-based methods

Physics-based methods attempt to predict the most prevalent secondary structure for any given amino acid sequence by simulating a physical model. MD-based methods are especially popular, as they allow a realistic consideration of experimental variables such as pressure and temperature, as well as solvation. In principle, MD methods can be adapted for *ab-initio* prediction of protein structures by sampling several folding events of a peptide chain from a random coil. In practice, however, a plain MD simulation may struggle to reach timescales of microseconds or longer that are necessary to observe significant conformational changes in peptides.

### Sampling issues in MD and advanced Molecular Dynamics methods

Convergence to equilibrium is an essential requirement for molecular simulation outputs to be accurate and reproducible. Yet testing for convergence is challenging, especially in instances where kinetic trapping and the consequent plateauing of ensemble quantities can falsely suggest premature equilibrium. To study MD with atomistic detail, femtoseconds time-steps are needed during the simulations. However, many problems in biochemistry involve events that only spontaneously occur after a nanosecond or even a millisecond, and when these events do occur, they are often rapid. MD simulations are efficient at conserving energy. Thus, if a conformational change requires non-stable intermediate conformations, the system might get trapped within a local minimum, and the observed states are thus likely to be highly dependent on the initial system. A collective variable (CV) is a low dimension variable that describes a conformational change or motion of interest. Multiple flavours of CVs can be used, and sometimes combined, to describe a motion. They can be as trivial as interatomic distances, angles or torsion, to more sophisticated variables such as a root mean square deviation, contact maps, and path collective variables. Figure 2.3 shows a hypothetical system energy landscape. Using an unbiased MD only a small range of conformations depending on the initial conformation can be explored and sampled, and a large number of conformations will not be represented in the trajectory obtained through MD.

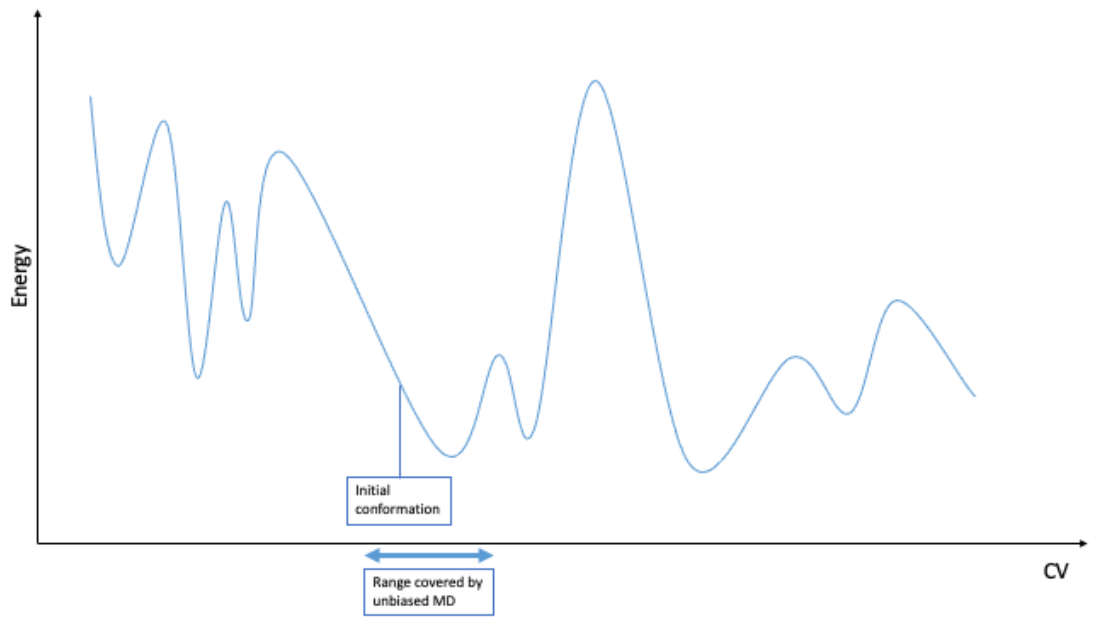



**Figure 2.3:** (Theoretical) conformational landscape of a biosystem: energy profile against a reaction coordinate or collective variable (CV): only a small portion of the energy landscape is explored with MD

Therefore, methods sometimes referred to as biased molecular dynamics which can cross energetic barriers have been implemented. Generally, these methods aim to lower energetic

barriers so they can be crossed in order to explore more of the conformational landscape.[225] A few MD techniques that have proven popular in the prediction of the secondary structure of peptides are presented below.

### Umbrella Sampling

In this MD technique, reaction coordinates define different states along the path characterising a conformational change. In a general umbrella sampling procedure, multiple windows are set with initial structures using different reaction coordinate values. A bias potential is applied to each window according to a harmonic (or adaptive) bias function.[226] The bias will constrain the system to sample a narrow window along the reaction coordinate, ensuring that the potential energy distribution overlaps between adjacent windows. Some post-processing methods such as the weighted histogram analysis method (WHAM) can recover the unbiased, free energy.[227]

### Metadynamics

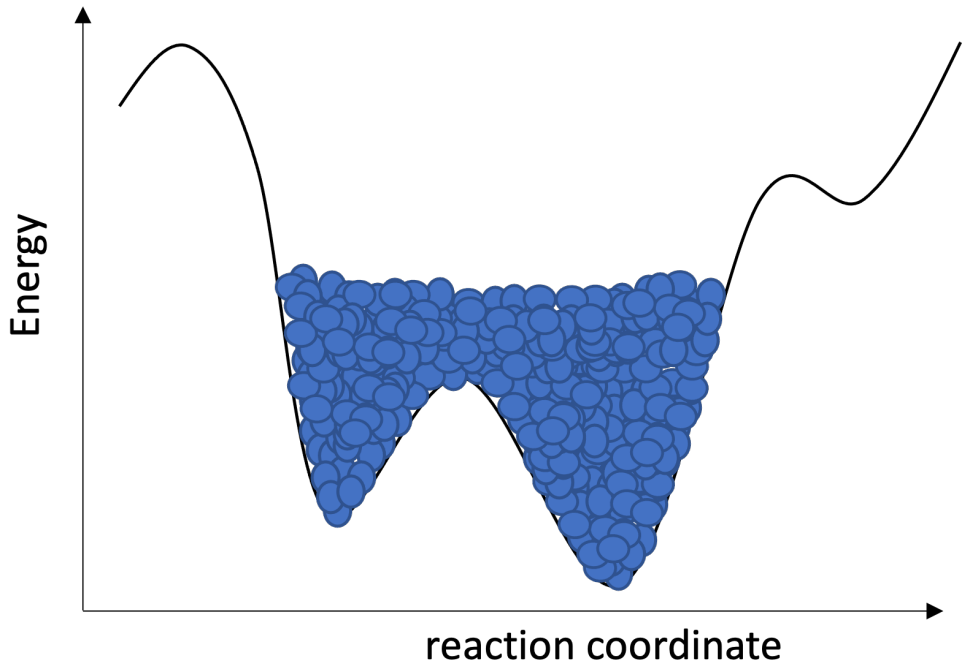



**Figure 2.4:** Energetic profile during a Metadynamics run: energy Gaussians (blue spheres) are added during the simulation 'filling' the energy profile

Meta-dynamics (MetaD) was originally developed by Parrinello and co-workers[228] to enhance the sampling of rare events. In MetaD simulations, an external history-dependent bias potential is imposed on the system during the simulation. Gaussian repulsive potentials centred on sampled values of the collective variables are deposited at regular intervals. As these potentials accumulates, the energy in the most explored region of the conformational

space raises, to encourage the system to visit configurations that have not already been sampled. The free energy of each region can be recovered as the sum of Gaussian added along the collective variable space is recorded (Figure 2.4). In theory, the user can use as many collective variables as desired, but in practice, one, two or three collective variables are usually employed.

**Accelerated molecular dynamics**

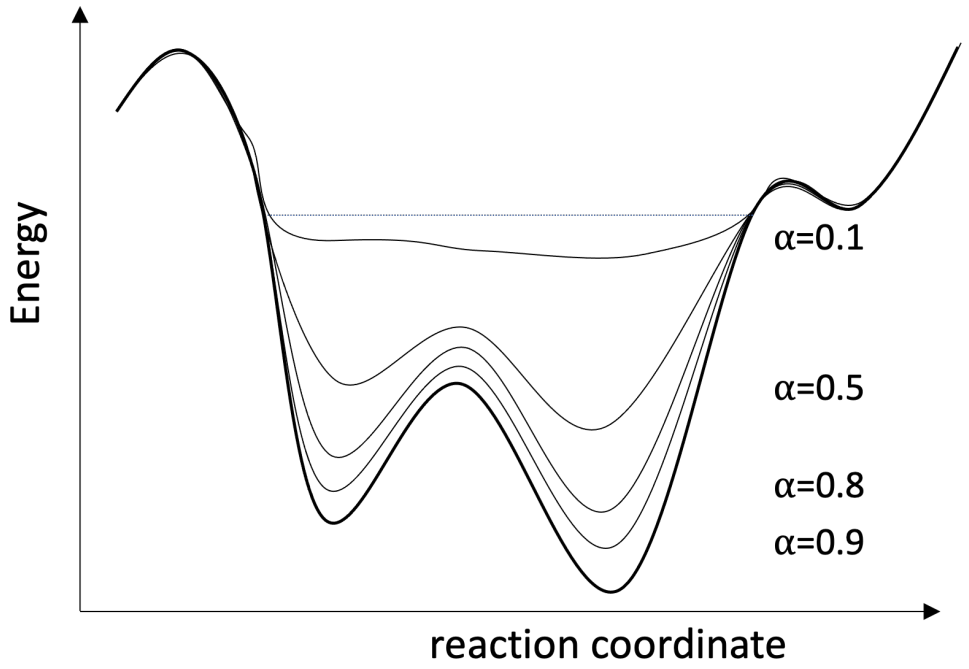



**Figure 2.5:** Energetic profile during an accelerated MD run: Boost potentials are added when energy falls below a threshold.

Accelerated molecular dynamics (aMD) is an enhanced-sampling method that improves conformational space sampling by reducing energy barriers separating different states of a system. One of the main advantages in comparison with other methods is that no collective variable or reaction coordinate is necessary. In aMD, the potential energy landscape is altered by adding a bias potential to the true potential such that the escape rates from potential energy wells are enhanced. The bias is applied for conformations with the lowest potential energy. The potential energy landscape is modified by raising energy wells that are below a certain threshold level while leaving those above this level unaffected (Figure 2.5). In the original form of aMD, when the system potential energy falls below a threshold energy, $E$, a boost potential is added. As a result, barriers separating adjacent energy basins are reduced, allowing the system to sample conformational space that cannot be easily accessed in a classical MD simulation.

**Replica exchange simulations**

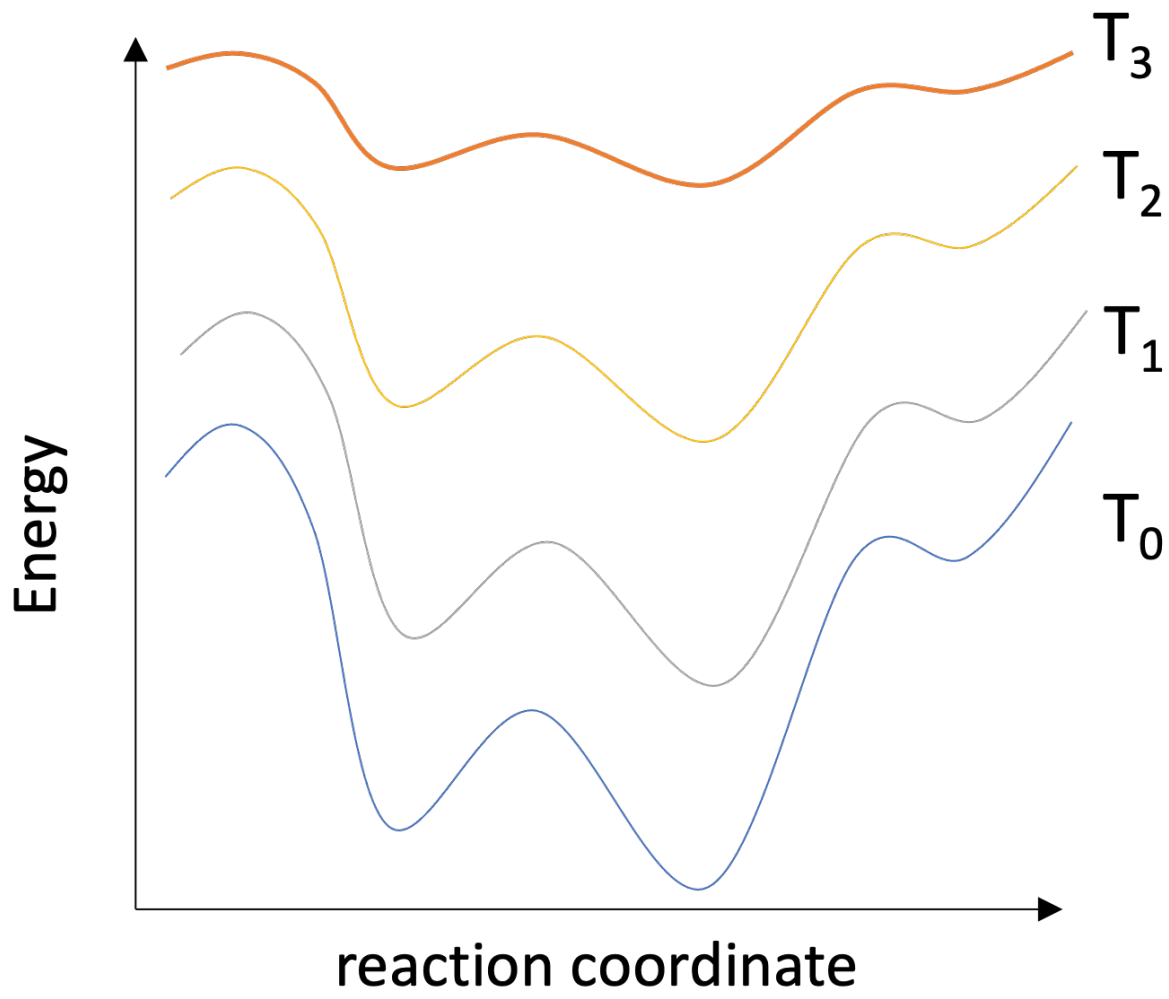



**Figure 2.6:** Energetic profile of parallel simulations during T-REMD: The energy profile is flattened with highest temperatures.

REMD simulations have shown to achieve effective sampling of systems that have a free energy landscape with many local minima. It has especially been shown to be potent in describing the folding of multiple protein systems.[229–231] By using more 'excited' thermodynamic states, the energetic profile will remain the same, except that the energy barriers will 'flatten' and be lower to cross (Figure 2.6), allowing the system to diffuse more easily from one state to another. The idea of parallel tempering is to run MD simulations that belong to different thermodynamic states, in parallel. More details about this method will be given in Chapter 4. REMD can be implemented in multiple and flexible ways. The most-reported implementation and earliest implementation are temperature replica exchange molecular dynamics (T-REMD) simulations which have shown efficacy for predicting the folding of peptides because they do not require prior knowledge of the folded state of the peptide. Even though T-REMD is computationally demanding, it has demonstrated flexibility in the setup as well as

a good sampling for rare events.[219] With this method, various thermodynamic quantities can be retrieved as a function of temperature. Other implementations of REMD exist, such as Hamiltonian Replica exchanges (H-REMD). Examples of H-REMD include Solute tempering, SWISH (Sampling Water Interfaces through Scaled Hamiltonians) and REMD with adaptive dihedral torsion scaling. [232–234] In general, because each replica can be simulated using its own processor, the various REMD methods are well suited for running on parallel computers.

### 2.4.3 Forcefield performance for the folding of peptides

Forcefields are the principal component of MD simulations, and despite more than 60 years of work in this field, they are still under constant improvement. The most frequently employed forcefields are GROMOS,[235] CHARMM,[236] OPLS,[237] and AMBER,[191]. Each of them exists in different versions and variants, as they have been corrected and improved multiple times. The AMBER forcefield for the natural amino acids was developed and published in 1986 (FF86) and was since revised several times.[238–243]. The addition of improved protein backbone parameters led to the creation of AMBER FF99SB, and the subsequent improvement of side-chain torsion potentials resulted in AMBER FF99SB-ILDN. Force fields optimised against NMR data are AMBER FF99SB*, CHARMM22* and CHARMM36. This section focuses on relevant peptide simulation studies carried out with diverse forcefields. For the sake of brevity and owing to their relevance to the work in this thesis, only explicit solvent simulation studies from the last 10 years are covered. The papers will be discussed in a chronological manner.

Lange *et al.* (2010) compared microsecond long simulations of ubiquitin and the gb3 domain of Protein G for 6 different forcefields (AMBER03, AMBER99SB, OPLS/AA, CHARMM22, GROMOS43a1, GROMOS53a6) to NMR data.[244] Large differences in sampling were observed. They reported that simulations longer than 100 ns had an increased probability of sampling non-native conformational states using the AMBER03 and GROMOS forcefields. Hydrogen bonds were also found to be weaker than in experimental results. From the tested forcefields, AMBER99SB was the best at reproducing the experimental data.

Lindorff-Larsen *et. al* (2010) tested 8 forcefields (AMBER99SB-ILDN, AMBER99SB-ILDN, AMBER03, AMBER03*, OPLSAA, CHARMM22, CHARMM27, CHARMM22*) on two peptides $(AAQAA)_3$ and CLN025, a 10-mer peptide.[241] Then in 2012 they compared the agreement of these forcefields to NMR data and circular dichroism, modeling ubiquitin, the gb3 protein and the FiP35 WW domain. They reported FF99SB*-ILDN and CHARMM22* to be better at stabilising the native structures over the other forcefields for both studies. In agreement with these studies, Piana *et al.* also concluded that FF99SB*-ILDN and CHARMM22* stabilise the native structures of the villin protein over the other forcefields.[245]

Cino *et al.* (2012) compared microsecond long MD simulations of Nrf2, a $\beta$-hairpin forming peptide, starting from an extended structure and using ten different forcefields (AMBER99SB-ILDN, AMBER FF99SB*-ILDN, AMBER99SB, AMBER99SB*, AMBER03, AMBER03*, GROMOS96 43a1p, GROMOS96 53a6, CHARMM27, and OPLS-AA/L).[246] OPLS-AA and (to a lesser extent) CHARMM-27 did not fold the peptide into a hairpin. In contrast, AMBER FF99SB*-ILDN, GROMOS96-43a1p, and GROMOS96-53a6 simulations yielded hairpin signatures throughout large parts of the trajectories. Overall, GROMOS (both 53a6 and 43a1p) performed the best, with GROMOS 43a1p giving the lowest RMSD to the NMR structures. They also analysed the trajectories using DSSP and C$\alpha$–C$\alpha$ contact maps. Interestingly, they noticed a different behaviour between capped and uncapped versions of the Nrf2 peptide using the tested AMBER forcefields. Experimentally, capping groups at the N-terminus were shown to have a modest stabilising effect on helical conformations (acetyl capping group having the largest effect), while capping groups at the C-terminus have little effect on the overall helical propensity.[**?** ]

Beauchamp *et al.* (2012) published an extensive study of 524 NMR experimental measurements and compared these to simulations using 9 AMBER forcefields variants (AMBER96, AMBER99, AMBER03, AMBER03*, AMBER03w, AMBER99SB*, AMBER99SB-ILDN, AMBER99SB-ILDN-phi, AMBER99SB-ILDN-NMR), the CHARMM forcefield (CHARMM27) and the OPLSAA forcefield.[247] They used 5 different solvent models (GBSA, TIP3P, SPC/E, TIP4P-EW, TIP4P/05). The dataset comprised simulations of dipeptides, tripeptides, tetra-alanine, and ubiquitin. AMBER96, AMBER99 and OPLS-AA performed the worst, while AMBER99SB-ILDN-phi, FF99SB-ILDN-NMR gave the closest results to the NMR data, followed by AMBER99SB-ILDN. They also reported that the explicit solvent models (TIP3P, SPC/E, TIP4P-EW, andTIP4P/2005) outperformed the implicit solvent model GBSA. However, each water model performed differently for each forcefield. Although the differences between the explicit solvent models are small, TIP4P-EW generally performed the best, especially for FF99SB variants.

Henriques *et al.* (2015) reported that eight forcefields (AMBER9915, AMBER99SB, AMBER99SB-ILDN, AMBER FF99SB-NMR, AMBER99SB* and AMBER03*, CHARMM22, and CHARMM36) were unable to reproduce the secondary structure of the largely disordered peptide, Histatin 5.[248] Moreover, they report that the forcefields used in these simulations exhibit considerable bias towards overly compact conformational ensembles (which were also found to differ between each forcefields) and favour certain secondary structure motifs (which was also found to be forcefield dependent). Finally, they emphasised on discrepancies observed between water models. The forcefield parameters derive from the study of structured proteins, which are mostly globular and compact, and exhibit a preference for helical structures. The authors suggest that these forcefields have been designed to stabilise helical structures, hindering their transferability to studying more unstructured entities by restricting their natural, highly flexible character.

Robustelli *et al.* (2018) developed a new implementation of the AMBER99SB-ILDN called AMBER99SB-disp and compared it to AMBER and CHARMM variants (AMBER99SB*-ILDN, CHARMM22*, CHARMM36m, AMBER03ws, AMBER99SB-UCB, AMBER99SB-ILDN).[249] They tested the forcefields with different water models. An extensive range of ordered and disordered proteins was used as test systems. The forcefields produced results in good agreement with experiments in many cases, but the simulations of the disordered proteins led to less reproducible results. The new AMBER99SB-disp forcefield yielded the best forcefield score. However, six of the systems simulated by the various forcefields for comparison were also used for training of AMBER99SB-disp, which could have provided an unfair advantage. For folded peptides, all but AMBER03ws performed well, with no significant score differences. For disordered proteins, AMBER99SB-disp, AMBER99SB-ILDN/TIP4P and CHARMM36m outperformed the other forcefields.

Gao *et al.* (2017) studied the folding of $\alpha$-helical and $\beta$-hairpin peptides using different AMBER forcefields on $\alpha$-helical and $\beta$-hairpin peptides.[250] Eight AMBER forcefields were compared (AMBER03, AMBER99SB, AMBER99SB-ILDN, AMBER99SB, AMBER12SB, AMBER14SB, and AMBER14-ipq, AMBER99SB-NMR) and used REMD to accelerate the sampling, starting with linear conformation and tweaking the temperature scale. For peptide 219M (an $\alpha$-helical peptide), AMBER03, AMBER99SB-NMR, AMBER12SB, and AMBER14SB forcefields sampled more folded states than the other three forcefields. For 1LE1 (a $\beta$ hairpin peptide), using the AMBER14ipq and AMBER99SB-ILDN forcefields, more beta-sheet propensity was observed, while the other forcefields mainly sampled extended or random coils for this peptide, and using the AMBER12SB forcefield, even $\alpha$-helical fragments were detected in place of $\beta$-sheets. Their conclusion was that the lack of a polarisation effect was the limiting factor in those forcefields for reproducing ordered secondary structures. AMBER99SB-NMR had the best performance among the seven common AMBER forcefields with the lowest RMSD to NMR structures and the highest probability of native hydrogen bonds.

Finally a recent paper from Tian *et al.* (2019) compared the AMBER14SB forcefield with the most recent version AMBER19SB forcefield. They compared the propensity of helicity of residues in MD simulations and in NMR.[205] They reported a better correlation of helicity per residue using AMBER19SB in conjunction with the OPC water model, while AMBER14SB gave a better reproducibility of the helical propensity when used with the OPC water model (Figure 2.7).

In conclusion, the accuracy of the MD forcefields has increased significantly over the 10 last years and older forcefields are now obsolete. The accuracy of a forcefield varies depending on the system under study, and the water model can impact the results significantly. A large amount of development that recently went into AMBER makes the forcefield a good choice. It is noteworthy that none of the above forcefields was found to be particularly good at predicting the conformation of largely disordered systems. In Chapter 3 we present a bespoke

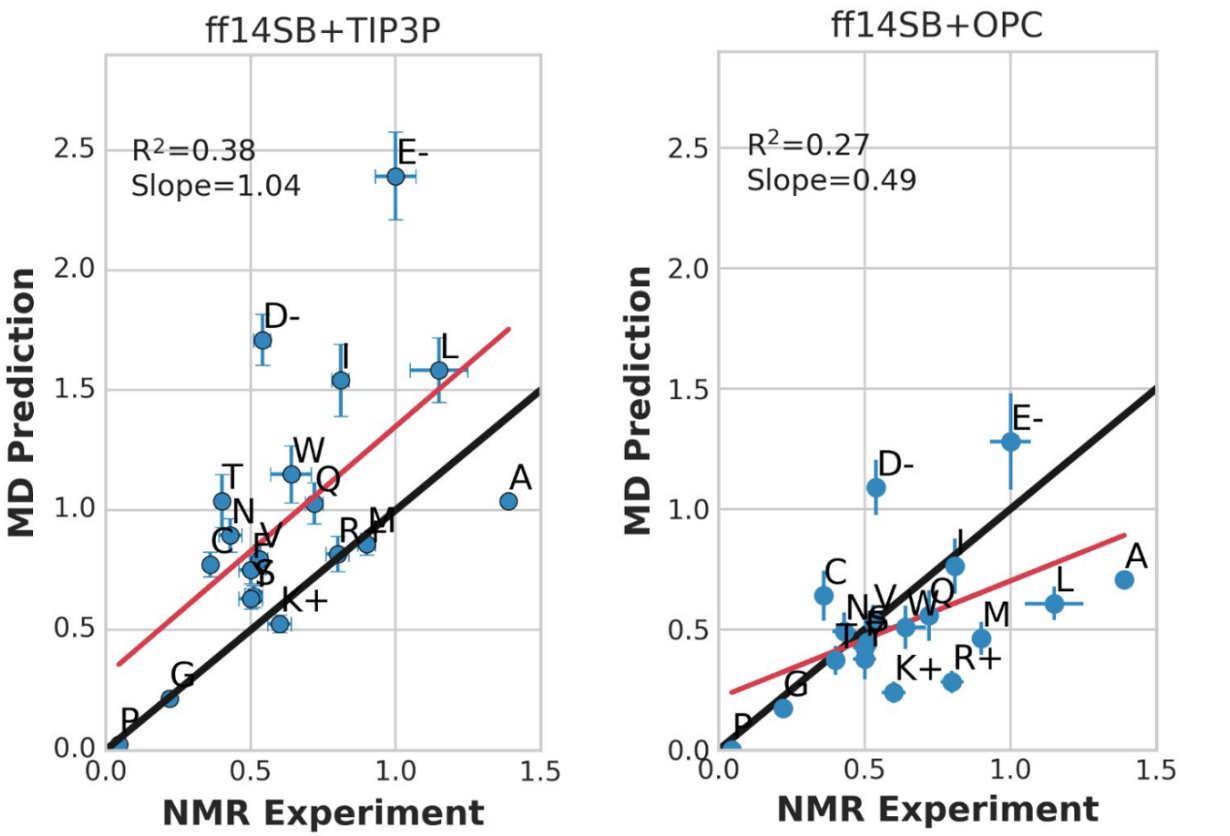



**Figure 2.7:** Helical propensity Agreement between NMR and molecular dynamics, Reproduced with permission from Tian *et al.*[205]

protocol for *de novo* parameter generation for unnatural amino acids which is compatible with AMBER14SB and later tested with the AMBER19SB forcefield.

# 3

# Development of a Forcefield Parameterisation Protocol for non Proteogenic Amino-acids

## 3.1 Determination of forcefield parameters

Current molecular dynamic simulation methods primarily rely on empirical potential energy functions (forcefields) to evaluate the potential energy of bio-molecules. Therefore an accurate description of the energetic properties of the molecules is crucial to enable the computation of physical quantities in good agreement with experimental data and the observation of molecular processes.

The AMBER family of forcefields contains parameters for all-natural amino acids. However, parameters for amino acid modifications found in proteins and peptides are lacking from the general forcefields. This includes covalent binders, oxidation and cyclisation of natural amino acids and unnatural amino acids. Consequently, many groups have developed 'add-on' parameter sets that can be combined with default forcefield parameter sets to simulate non-natural amino acids. To the best of our knowledge, to date, no AMBER compatible stapled peptides library add-on forcefield has been published. A non-exhaustive list of add-on forcefields compatible with the AMBER forcefields includes :

- AMBERDYE, a forcefield add-on for dyes commonly used in labelling experiments.[251]
- SWISSSidechains, a forcefield add-on for some of the most common unnatural amino acids.[252]
- Forcefield-PTM, a forcefield add-on for post-translational amino acid modifications.[253]
- Forcefield-NCAA, a forcefield add-on for unnatural amino acids found in complement inhibitors of the compstatin family.[254]
- Phenylalanine and Tyrosine Derivatives add-on forcefield.[255]

- Azido and alkynyl containing amino-acids forcefield.[256].

In Chapter 2, Section 2.3, we described the different terms and parameters of the AMBER forcefield. Historically several methods have been used to parameterise molecules for MD. The bond (and respectively angle) parameter is expressed with a distance (respectively angle), corresponding to the relaxed distance (respectively angle) between two (respectively three) atoms and a harmonic potential, which translate the energy necessary to stretch or compress the bond (or angle) from its optimum value. As an example, for the AMBER94 forcefield, equilibrium distances and angle for the bonds and angles were obtained by averaging bonds distances and angles measured in protein X-ray crystallography structures. Harmonic force constants were computed from QM vibrational analyses.[257] The Lennard-Jones parameters were derived from Monte-Carlo liquid simulations by adjusting the Lennard-Jones parameters of simple molecules such as methane, ethane, propane and butane to reproduce the densities and enthalpies of vaporisation of these liquids.[257] Parameters for the bonded and Lennard-Jones interactions have been fairly conserved until the most recent versions of the AMBER forcefield.[205, 257] Atomic partial charges were calculated through Quantum Mechanics (QM) by fitting the (gas phase) electrostatic potential, calculated at the HF 6-31G* level. Indubitably, the torsional terms have been through the most refinements over the different versions of the forcefields. Different data-sets and methodologies for the parameterisation of the dihedral terms have been developed over time. In AMBER94, the torsion parameters were obtained by fitting against energetic profiles from QM calculations for a small number of low-energy conformations of glycine and alanine dipeptides.[257] In AMBER96, identical parameters were used initially for the $\phi$ and $\psi$ dihedral terms, and these were then empirically adjusted to reproduce the energy difference between extended and constrained $\alpha$-helical energies for alanine tetrapeptides. AMBER99SB reflects another attempt at refitting backbone dihedral parameters by including eleven representative structures of alanine tetrapeptides along with alanine dipeptides.[192] AMBER99SB-ILDN is a variant of the AMBER99SB forcefield where NMR data has been used to refine the torsional term of the backbone to NMR data.[241] Therefore, in this work, parameterisation of the dihedral terms has been given the most attention.


**Figure 3.1:** General pipeline for *de novo* parameterisation of amino acids

In this work, Lennard-Jones, bond and angle parameters for unnatural amino acids were extracted from the GAFF2 forcefield. This is justified by the observation that very few changes in these terms have been made in the recent iterations of the AMBER forcefields. Given the

relatively small amount of experimental data for stapled peptides, we focused on producing data sets from quantum mechanics to fit the remaining partial charges and torsional (dihedral and improper) terms. Figure 3.1 describes the different steps of our parameterisation protocol. The first stage is to produce a number of conformations and to derive the atomic partial charges; details of the protocol are given in section 3.2.1. The second stage produces Lennard-Jones parameters and bonds and angles parameters, as briefly described in section 3.3; this is trivially achieved by using tools already available. Finally, we developed a method for the parameterisation of torsional angles, which required a carefully generated set of conformations. The generation of those conformers and the dihedral fitting is described in section 3.4.

## 3.2 Derivation of atomic partial charges

For biomolecular additive non-polarisable forcefields, the most commonly used charge determination methods are based on fitting the electrostatic potential (ESP). Other variations include bond charge increments[258] and electro-negativity equalisation methods.[259]

### 3.2.1 The ESP method

In the AMBER forcefield, the partial charges are summed through a Coulombic potential (as described in Chapter 2 Section 2.3.1). Difficulties in determining such charges arise from the fact that they are not observable and are dependent on a given conformation. In non-polarisable forcefields, it is required that these charges are time and conformation independent, which is an approximation.

ESP calculations require the conduction of quantum mechanics (QM) calculations of the electronic environment of the molecule. A preliminary geometry optimisation operated by QM calculations allows the system to be as close as possible to a low energy conformation to limit artefacts due to a high energy conformation. Then the map of electrostatic potential (MEP) is computed: during the calculation of the MEP, a virtual unit charge, generally positive, is placed at a certain distance from the molecule. The electrostatic potential is defined by the interaction (or perturbation) energy between the charge and the molecule. This calculation is performed many times around the molecule, and the electrostatic potentials are then assembled to produce a 3D map of the electrostatic potential, the MEP. The last step is to fit the molecule in this 'electronic shape' and assign the partial charges which reproduce this shape at best.

#### Limitations of ESP

The standard ESP charges derivation method has several shortcomings. First, Williams *et al.* studied the variation of ESP charges calculated for an alanine dipeptide using 12 different

conformations.[260] The partial charges generated by these conformations varied considerably. Second the charges of "buried" atoms are statistically under-represented, as MEP points must lie outside the molecular van der Waals (vdW) surface. Thus, buried atoms (such as quaternary carbon atoms) are not represented by a significant number of MEP points. In contrast, atoms at the molecular surface contribute to a more substantial number of MEP points. Third, equivalent atoms (e.g. the hydrogens of a methyl group, or atoms in symmetric groups) might be given different partial charges values, which is only relevant for the specific conformation used for the calculation.

### 3.2.2 The Restrained Electrostatic Potential (RESP)

Bayly *et al.* developed the RESP (restrained ESP-fit) charge model.[261] The RESP model involves a chi-square of the charges to the electrostatic potential (ESP, $\chi^2_{ESP}$), but with the addition of hyperbolic restraints on charges for equivalent atoms ($\chi^2_{restrain}$).

$$\chi^2_{RESP} = \chi^2_{ESP} + \chi^2_{restrain} \tag{3.1}$$

with

$$\chi^2_{restrain} = a \sum_{n_{atoms}} (\sqrt{(q_i^2 + b^2) - b} \tag{3.2}$$

$b$ determines the tightness of the hyperbola around its minimum and $a$ is the weight factor for the strength of the restraint function. Charge values are fitted to reproduce the MEP, but these restraints serve to re-evaluate the charges of buried atoms without impacting the fit. The statistically defined charges are only negligibly affected by the restraint. In contrast, atoms with fewer MEP points are restrained toward a null partial charge, reducing unintended artefacts. RESP charges determination lead to smaller variations of the charge among related functional groups. RESP charges are used in all contemporary AMBER forcefields. The final RESP model is produced via a two-stage fit, with the first stage recovering equivalent atoms by molecular symmetry (e.g. C1 and C3 in propane) and the second stage needed to fit equivalent charges on hydrogen atoms.

### 3.2.3 The Restrained Electrostatic Potential (RESP) using multiple conformations

Atoms being polarisable, partial charges are dependent on the conformation chosen for the RESP calculation and this has an impact on the final partial charge values. Multi-conformational approaches can be used to tackle this issue. Multiple RESP charge estimations are produced and averaged. This approach has been described by Dupreadeau *et al.*,[262] but was only available via the pyR.E.D web-server, which was down from October 2018 to June 2019. In this work, we implemented our own RESP derivation protocol relying on different

conformations of the molecule.

### 3.2.4 In-house implementation of a multi-conformation RESP

**Implementation**

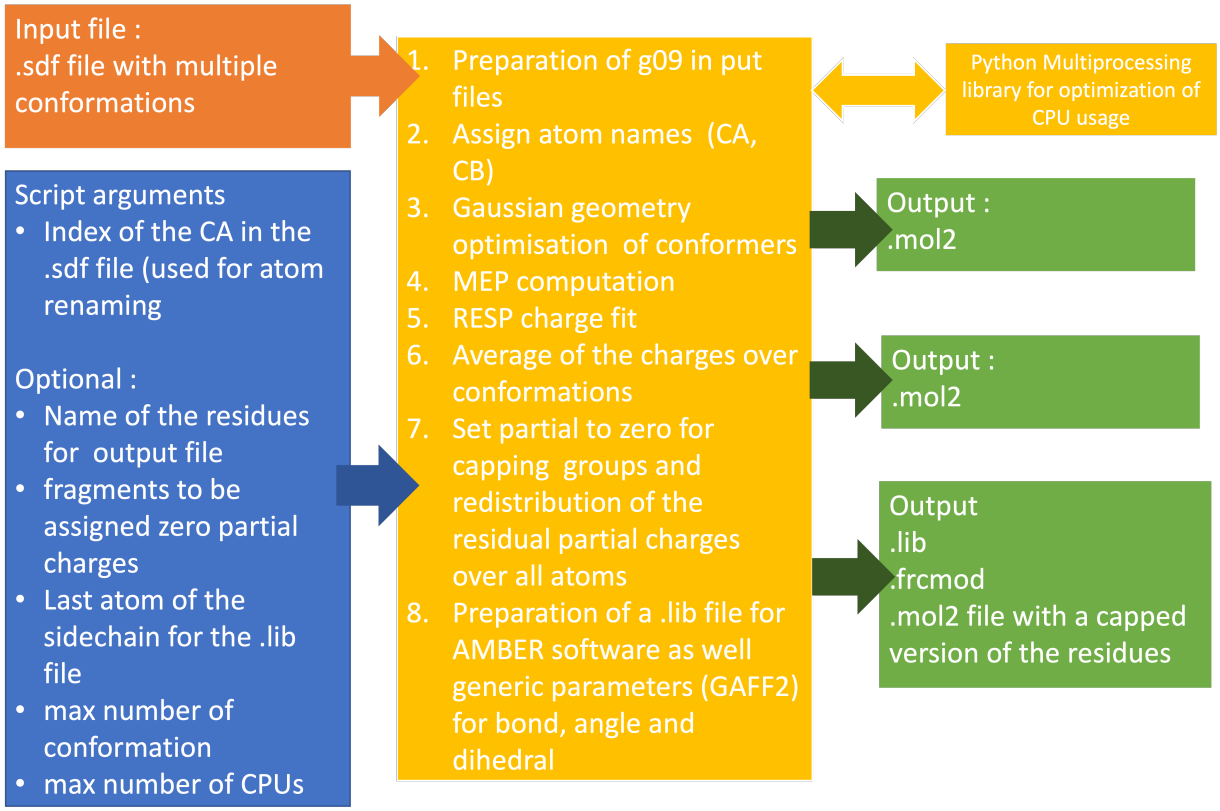



**Figure 3.2:** Pipeline followed by the multi-RESP.py script

We ran parallel RESP derivations of multiple conformations of the molecule. The resulting partial charges were then merged to reduce the error linked with the choice of the conformation given as input. In the multi-RESP implementation, the user provides an input file, containing a certain number of conformers, which can be generated independently or using the script provided. Of note, capping groups are used during QM calculations to avoid 'halved' molecular bonds and to mimic adjacent amino acids. The capping groups acetyl at the N-terminal and methyl-amine at the C-terminal were used. After QM, the capping groups are removed, and any non-integral residual partial charge is redistributed evenly over all remaining atoms. Typically, this residual charge does not exceed 0.05 *e* as seen in Figure 3.3. Stapled residues were parameterised using two residues chemically bound by their side-chain and two backbone fragments (Figure 3.3). For two charge estimation repeats using two sets of 50 conformations, discrepancies of around 0.05 *e* are observed. Thus charges are rounded after two digits. The script returns a forcefield library file which can be used directly to setup MD simulations. The multi-RESP implementation can found on github : michellab/Stapline

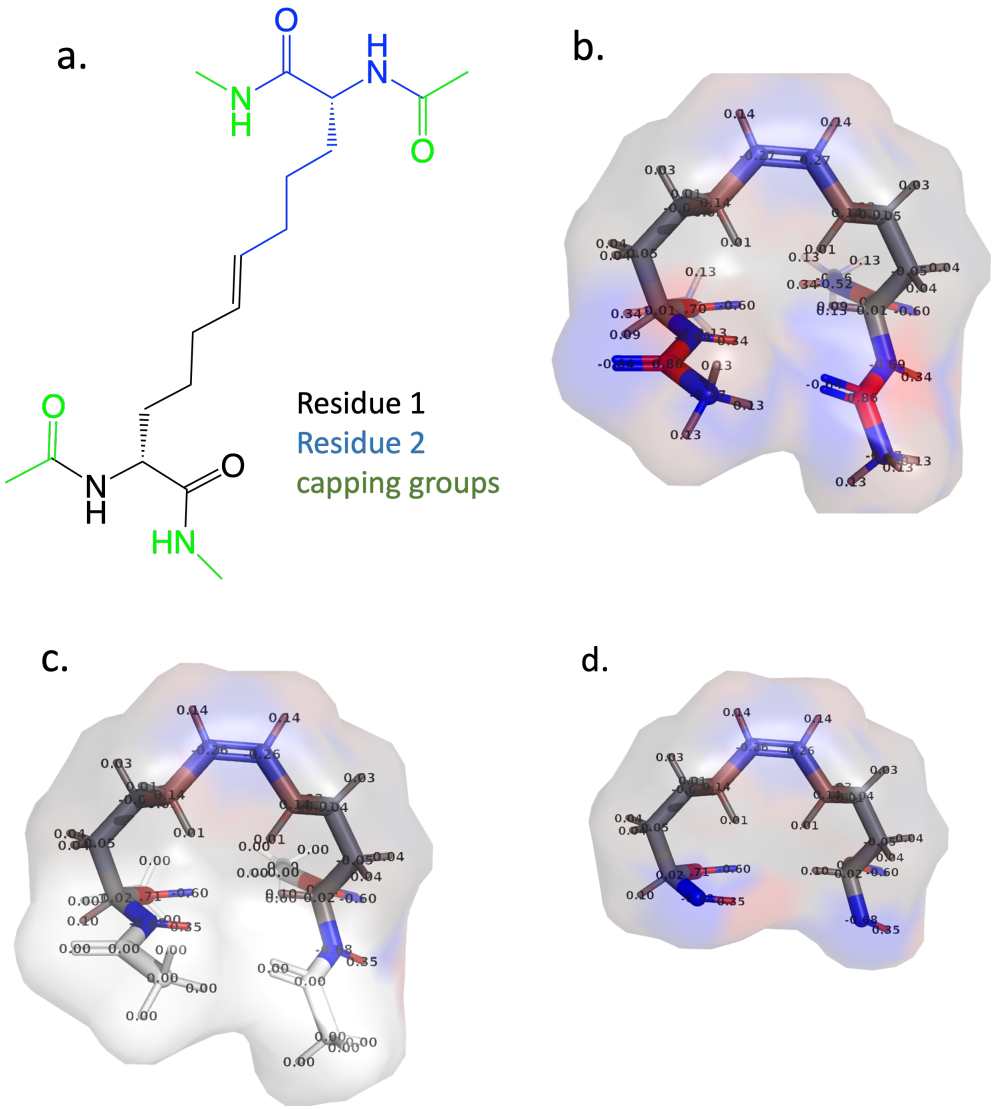



**Figure 3.3:** Partial charges calculated for (2*R*,11*R*,*Z*)-2,11-diacetamido-N1,N12-dimethyldodec-6-enediamide a. Structure and the stapled residues produced as well as the capping groups are highlighted b. Partial charges including capping groups c. Partial charges of the capping group are set to zeros, and residual charges are distributed to all atoms of the residues. d. Final fragment kept for building the library.

All intermediate stages were scripted so the user could run the multi-RESP on one line; the script directly returned a .lib file and could thus be used directly in simulation, with generic torsion terms from the General AMBER forcefield (GAFF2). Atoms are renamed to follow the AMBER forcefield naming formalism.

### Practical Considerations

The pipeline developed here not only produced a RESP fit, but also produced AMBER compatible libraries for the amino acid, (given the user provides the C$\alpha$ atom of the amino acid). These forcefield add-ons are meant to be combined with the AMBER general forcefield. Thus it is important to follow the nomenclature of this forcefield. All amino acids of the AMBER

force-field have integer charges (-1, 0 or +1). Hence any combination of amino acids will always result in an integer charge. Net charges are usually neutralised with counter-ions in an explicit solvent system. This is particularly important as the Ewald method used to compute electric fields will only converge if the net electric charge of the system is zero.

When sidechains atoms are absent of an input file, they will be added automatically by tleap (AmberTools),[263] which is a powerful software to prepare simulations. We took advantage of that feature extensively in the rest of our work to introduce stapled residues in our designs. However, the software does not rely on a library of rotamers and uses a single representative side-chain conformation. We prepared libraries with sidechains as extended as possible to produce the least clashes when added into a peptide.

**Comparison of our in-house multi-resp with pyR.E.D**

The pyR.E.D server is a reference for calculating RESP charges on multiple conformations.[262] We compared the partial charges obtained with our method with partial charges obtained using the pyR.E.D server. In both cases, the same basis set and level of theory were used. Of note, our implementation differs slightly from the R.E.D implementation. In our approach, the charges are computed from separate RESP and subsequently averaged. We compared two RESP runs using the RED server, and a run obtained using our method, to determine an acceptable root main square (RMS) value between two consecutive runs. For each run, we used an independently generated conformer set of 20 conformations of pentenyl-alanine. We also compared the effect of the C$\alpha$ stereochemistry on (*S*)-pentenyl-alanine and (*R*)-pentenyl-alanine partial charges. Thus six independent runs were compared. RMS were calculated pairwise between the runs (Table 3.1). As seen in Table 3.1, RMS values between two runs on the R.E.D server can result in partial charge variations of up to 3.1 *e*. In comparison, the RMS values between our in-house implementation and the R.E.D yielded differences of up to 3.4 *e*. Thus the difference between the partial charges generated with our approach was similar to the difference between two consecutive runs on the R.E.D server. The partial charges computed with our approach yielded similar results to the ones of the R.E.D server.

Moreover, the partial charges between the (*R*) and (*S*) enantiomers were also found to be comparable, as the RMS between the (*R*) and (*S*) enantiomers using our in-house approach yielded a RMS of 1.2 *e* (and 1.9 *e* and 1.4 *e* for the R.E.D server) which is lower than the RMS obtained between two consecutive runs of the *R* enantiomer on the R.E.D server (3.1 *e*)

| | | RED charges deriv | | | in-house | |
|---|---|---|---|---|---|---|
| | | *(S)*-2 | *(R)*-1 | *(R)*-2 | *(S)*-3 | *(R)*-3 |
| RED charges deriv | *(S)*-1 | 1.3 | 1.9 | 1.2 | 1.6 | 2.2 |
| | *(S)*-2 | | 2.1 | 1.4 | 1.2 | 1.3 |
| | *(R)*-1 | | | 3.1 | 2.1 | 1.8 |
| | *(R)*-2 | | | | 1.4 | 1.5 |
| in-house | *(R)*-2 | | | | | 1.2 |

**Table 3.1:** RMS (in *e*, elementary charge) between partial charges of 2-acetamido-*N*-methylhept-6-enamide calculated for different runs, using the *(S)* and *(R)* enantiomers using the R.E.D. server (runs *(S)*-1 *(S)*-2 *(R)*-1 *(R)-2*), and our in-house implementation (runs *(S)*-3 *(R)*-3)

### Determining the number of conformer needed to achieve a sufficient sampling

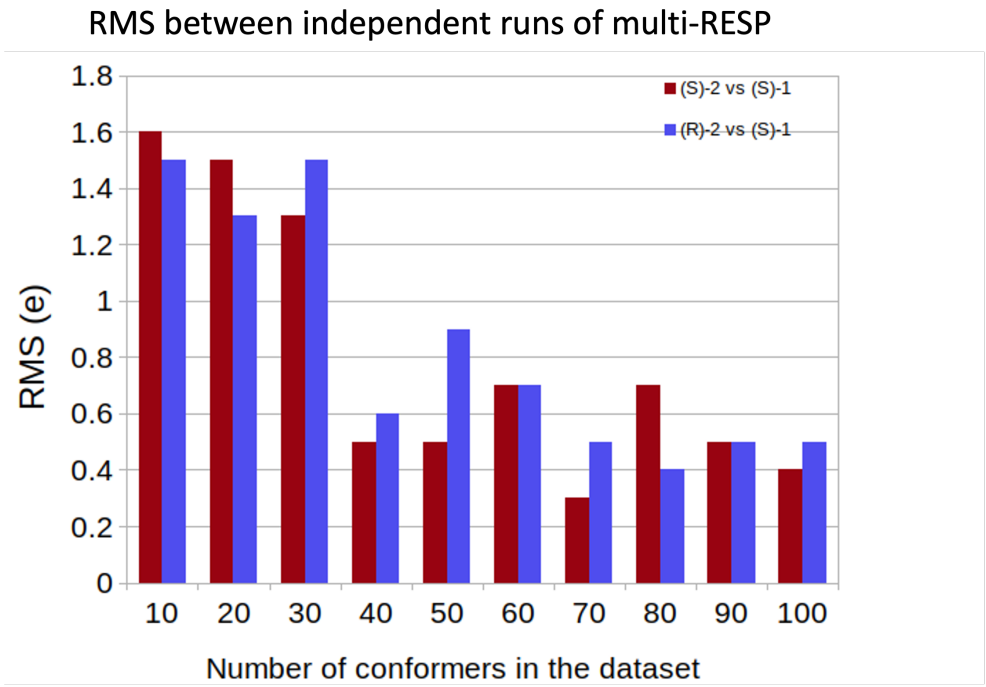



**Figure 3.4:** Variation of the partial charges when increasing the number of conformers in RESP calculations. Partial charge RMS values were calculated using three consecutive runs obtained by our in-house approach, and compared pairwise. *(S)*-2 vs *(S)*-1: RMS between *(S)*-pentenyl-alanine, run 1 and *(S)*-pentenyl-alanine, run 2; *(R)*-2 vs *(S)*-1: RMS between *(R)*-2 : *(R)*-pentenyl-alanine, run 1 , vs *(S)*-pentenyl-alanine, run 2;

We ran the multi-RESP pipeline for a number of different conformers to generate the partial charge of the dimerised *(S)*-pentenyl-alanine and *(R)*-pentenyl-alanine. A RMS value was calculated by comparing the partial charges values for each atom, aiming to understand if lower variations of the partial charges could be achieved between consecutive runs by using more conformers. Rdkit was used to generate the conformer sets. We used two independent runs using *(S)*-pentenyl-alanine, and one run using *(R)*-pentenyl-alanineglycine for this study. The MEP calculations were conducted for the 300 conformations using Gaussian09.[264]

Above 40 conformers, a drop of RMS from 1.3 *e* to around 0.6 *e* is achieved (Figure

3.4). However, above 60 conformers the RMS between the consecutive runs is not reduced any further. Thus using between 40 to 60 conformers is an adequate number to calculate the partial charge of the molecules. This study also reiterates the fact that *(R)* and *(S)* enantiomers have similar partials charges as the RMS between the (*R*) and (*S*) conformers is similar to the RMS between two runs using the (*S*) conformer, for any number of conformations. Unless otherwise stated 50 conformers where used for every other molecule. *(S)* and *(R)* conformers where given the same partial charges, to allow the use of the enantiomers interchangeably, without introducing residual charges.

## 3.3 Determination of Bond and Angle parameters

In the AMBER forcefield, only a small number of bonds are present : 29 bond types are defined for the 20 natural amino acids. As an example every $sp^3 - sp^3$ carbon-cabon uses a relaxed distance of 1.52 Å and a harmonic constant of 310 $kcal.mol^{-1}$.Å$^{-2}$. Similar observations are made for the angle parameters, with only 99 different types of angle within the 21 natural amino acids (e.g. sp3-sp3 carbon-carbon bond has a relaxed angle of 109.50° and a force constant of 50 $kcal.mol^{-1}$). Noteworthy, the general AMBER forcefield only contains chemical groups present in natural amino acids. As an example, non-aromatic alkenes, which are present in RCM staple chemistry or the triazole group present in click chemistry staples, are not represented in the general forcefield. We thus used GAFF2 (generalised AMBER forcefield) to attribute bond and angle parameters as it contains more varied chemical groups such as heteroaromatic and alkyne functions.[242] AmberTools possess a module, parmchk2, to automatically assign bond and angle parameters to the molecule of interest.[263]

## 3.4 Determination of parameters for dihedrals

### 3.4.1 Introduction

Dihedral torsion parameters require careful optimisation as small changes can significantly influence conformational preferences of proteins and oligonucleotides.[192, 265–267]. The use of different versions of the AMBER forcefield result in very different helical propensities in MD simulations (see Chapter 2, Section 2.4.3). At the start of the project, the program Paramfit,[268] developed by Betz *et al.* appeared very promising for the parametrisation of the dihedrals for our stapled residues library. We thus implemented a systematic and semi-automated protocol relying on Paramfit. We noticed long calculation times to obtain dihedral parameters. We also observed that poor parameters were occasionally produced by this approach (some of the results obtained using this approach will be presented in section 3.4.4). This motivated us to develop our own approach, shown in section 3.4.6. Our protocol produces more robust and accurate fits than Paramfit, with the additional benefit of faster runtime

execution. The backbone dihedral parameters were improved by initialising the calculations with the parameters of the GAFF2 forcefield and fitting to potential energies determined by quantum chemistry calculations using a least-square approach. Molecular dynamics simulations were subsequently performed to determine the preferred backbone conformations and compare them with the proteogenic amino-acids.

### 3.4.2 Dihedral fitting

In a forcefield, the change in potential energy upon rotation of a dihedral angle is dependent on non-bonded interactions and an additional bonded-term dependent on the four atoms connected in the dihedral bond.

$$V_{dihedral} = \frac{1}{2} \sum_{i \in \text{torsions}} V_{1-4_{(non-bonded)}} + \sum_{i \in \text{torsions}} \sum_{n} \frac{1}{2} V_i^n [1 + \cos(n\omega_i - \gamma_i)] \tag{3.3}$$

In AMBER forcefields, Equation 3.3 is used to calculate the potential energy of a dihedral torsion in a molecule. Two terms contribute to this energy. The first term is the non-bonded term, $V_{1-4}$, which is defined by the Lennard-Jones and Coulomb energies, and in our protocol no partial charge adjustment was made at this stage. In the AMBER forcefield, the non-bonded energies calculated between atoms (1-4) of the dihedral angle term are halved. The second term, the bonded energy of a dihedral angle, is parameterised using a Fourier series. Each Fourier term is defined by : a force constant, $V_i^n$; a multiplicity, $n$ and a phase shift $\gamma_i$. For bonded dihedral terms, in the AMBER forcefield, the force constant of a given Fourier term generally has a value between 0.00 and 25.00 $kJ.mol^{-1}$. The phase shift values is either 0 or 180°. The multiplicity determines how many terms will be used for each dihedral. Usually, in the AMBER forcefield 1, 2 or 3 Fourier terms are used for each torsion with multiplicity values of 1, 2, 3, 4 or 6. As an example, a torsion with two $sp^3 - sp^3$ central atoms generally has two Fourier terms with multiplicity values of 1 and 3, while a torsion with a central bond comprising $sp^3 - sp^2$ hybridised atoms generally has 1 or 2 Fourier terms with multiplicity values of 1 and 2 (or 4).

### 3.4.3 Least square fitting method principle

A mathematical procedure for finding a function that best fits a given set of points is to minimise the sum of the squares of the offsets of the points from the function (sometimes called target or objective function). For the non-linear fitting, several iterations are needed to find the parameters minimising this sum. In general, several refinement cycles are required to obtain best-fit parameter values. Importantly, initial values can have an impact on the final set of parameters as poor initial values can result in the iterative least-squares refinement to converge to a local minimum. The objective function should return the Chi function, which corresponds to the difference between the data and model. The chi-square ($\chi^2$) is often de-

fined as seen in equation 3.4. Weighting factors can also be applied to minimise the effect of outliers in the data set.

$$\chi^2 = \sum_i^N [\frac{(y_i^{data} - y_i^{model})^2}{w_i^2}] \tag{3.4}$$

### 3.4.4 Paramfit approach

Paramfit was developed for the parameterisation of molecules, the method relies on a least square method to fit the parameters.[268] The function to minimise is the sum of differences between MM single point energies and QM energies for each conformation is presented in Equation 3.5.

$$f(N, E_{QM}, K) = \sum_n^{orders} [(E_{MM}(i) - E_{QM}(i) + K)^2] \tag{3.5}$$

With $E_{QM}$ and $E_{MM}$ being respectively the quantum and molecular dynamic energies. K is a constant to compensate for differences in the relative energy values between the quantum mechanics calculations and the Molecular dynamic calculations. A dihedral fitting problem usually possesses many variable in comparison to a typical use case of the least square method. Indeed, the number of terms per torsion is rather large: with one carbon-carbon bond, considering all bonds are saturated, the dihedral term is composed of 9 'torsions' (4 atoms linked), each consisting of several terms; one per order (usually 1 to 3 orders). Paramfit implements a genetic and simplex algorithm for the convergence of the least square fitting method. The simplex method is one of the most versatile methods used for least-square fitting, and it is still the standard method employed for many optimisation problems. After the objective function is defined, the set of rules to apply to the system is translated into linear functions following certain rules called slack variables. These variables define the boundary of the search area. The method starts from an extreme point of that boundary. Next, a test determines whether that extreme point is optimal. If the test is not passed, an adjacent extreme point is sought along an edge in the direction for which the value of the objective function increases at the fastest rate; a new extreme point is reached. The sequence described is then repeated. Termination occurs when an optimal extreme point is found. The main advantage is a fast convergence rate as the method typically converges on the optimal solution in a number of steps that is only a small multiple of the number of extreme points. However, the simplex algorithm is not very efficient for systems having a lot of local minima, as it will converge to the closest, being thus highly dependent on the initial set of parameters. In order to cover more search space, Paramfit implements a hybrid genetic algorithm to conduct the minimisation, with refinement using a simplex algorithm to accelerate convergence. A genetic algorithm is used to generate multiple solutions on the energy landscape. The genetic algorithm is a procedure inspired by biological evolution. First, an initial number (population) of parameter sets are

generated. Parameters in each set are changed ('mutated') at random in a specific range of the initial parameters. The sets are then ranked according to their performance on the fitness function. 'Mutations' that improve the fit are retained. For recombination, two sets of parameters ('parents') are chosen uniquely from the selection pool and are combined in one of two ways, selected at random. Then recombination with other beneficial parameters occurs before a new round of mutations is introduced in the population. Successive generations are generated until an optimum is reached. Convergence is reached when the best fitness within the population remains unchanged for a threshold number of generations. Paramfit implements a hybrid genetic algorithm, where once the best fitness remains unchanged from one generation to the next using the genetic algorithm, a simplex algorithm is run. In their paper Betz *et al.* [268] concluded that their algorithm could find local minima in poorly sampled landscapes and was also able to recover parameters used in the ff99SB forcefield.

### 3.4.5 Paramfit wrapper

A wrapper script was written to facilitate and automate the use of Paramfit. Paramfit requires specific input files to specify the torsion to be optimised and the parameters of the fit. We wrote a script that recognises the torsions in a residues (PSI, PHI, CHI1, CHI2, CHI3...), given the C$\alpha$ is correctly labelled by the user. The script then produces a conformational scan around each bond: each bond of the backbone is rotated using a given angle step within 360 degrees, and produces the Paramfit input files. Input files for Gaussian09 are auto-generated for each conformer. The QM energy of each conformer is then computed and the conformations are optimised, with the positions of the heavy atoms in the dihedral of interest kept constant. The QM outputs are parsed directly by the script, and Paramfit is used to produce a set of torsional energy parameters for the torsion of interest. We found that Paramfit performed better if relatively good parameters were used in the input topology. As non-aromatic alkene (present in the RCM staples) parameters are found in the GAFF2 forcefield, we used the GAFF2 forcefield parameters in the input.

#### Results

Overall we found that the MM values obtained using the GAFF2 forcefield parameters offer generally a pertinent energy profile when compared to the QM values. It allowed us to assess if Paramfit produced a pragmatic improvement over these parameters. A set of quantum mechanical energies for different conformers was generated for each torsion of the molecule of interest. We used the genetic algorithm and subsequently the simplex algorithm implemented in Paramfit to refine those parameters. However, the results obtained with this method were unsatisfactory. For instance, Figure 3.5 shows a case where Paramfit produces erratic parameters for the CHI3 torsion. The MM profile of the CHI3 torsion is around 5-fold the energy found using QM or MM with the GAFF2 parameters. To overcome this problem, we devel-

oped a 'hybrid' set of parameters (Figure 3.5), where Paramfit parameters were adopted only if they offered an improved fitting against the QM energy profiles, in comparison to the MM values obtained with the GAFF2 parameters. This approach provided a MM energy profile closer to the QM data even if some discrepancies in the energetic profile were observed. We can see in Figure 3.5 the MM values obtained with our hybrid method gives a better fit when compared with GAFF2 parameters.

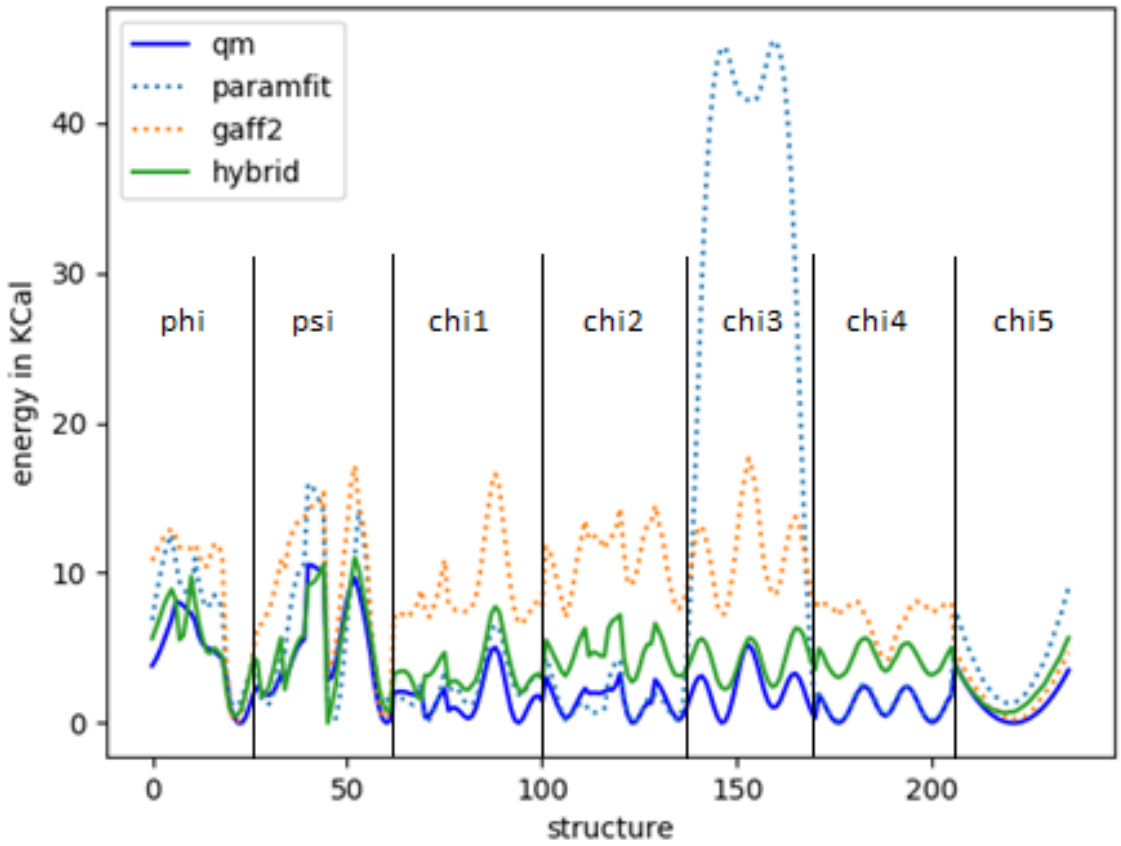



**Figure 3.5:** Fitted parameters by torsion for the pentenyl glycine residue (stapled) the blue curve represents the QM data used for the fit of the new parameter (light blue); Non-fitted GAFF2 parameters are found in orange; Hybrid parameters are a set partially taken from newly optimised parameters and GAFF2 parameters (green)

However, this hybrid approach was relatively slow, with 10 to 54 hours needed for the fitting of each parameters (using a 24-core workstation); usually, after the QM energies were obtained, the fitting would take two days to complete for all of the torsions, but in some instances, convergence was not reached after a few days of refinement, and the values obtained showed large discrepancies with QM data.

A major bottleneck of this approach is that Paramfit calculates the least square to quantum data using single-point energies. That means that all MM energies are computed for each QM point at each iteration of the least square. This slows down the process in two ways. Firstly calculating single point energies for a relatively large number of conformations is a costly

operation. Isolating one single point energy calculation on Paramfit is arduous. Using Sire, another MM engine supporting AMBER topologies, we calculated that one MM single point energy required an average of 0.4 seconds. A part of this execution time is used to load topologies and coordinates from a file, and a lower time should be expected when in this run iteratively. However a non-negligible time should be expected on each refinement of Paramfit to calculate single point energies. Calculating a large number of single points slows down this approach independently of the algorithm chosen for convergence. The second reason is that fitting to single points energies does complicate the convergence as a discrete function is used.

Besides the performance issues, we had several concerns using the results obtained from the Paramfit optimisation procedure. First, the absolute values returned by Paramfit for the dihedral amplitude term could be 10 to 50 higher than GAFF2. Second, some of the calculated values were negatives. Negative values are technically acceptable, but no negative dihedral amplitude terms are found in the AMBER forcefield. Large amplitude values are more problematic because they may lead to poor torsional energies for conformations not present in the training set. In Figure 3.6.a, we compared the QM energy profile and the MM energy profiles using parameters obtained from Paramfit and using GAFF2 parameters for the (S)-penthenyl-alanine molecule. Paramfit produces a tight fit of the rotamer of the CHI1 torsion used to generate the parameters (training data set). However, using another conformation of the molecule and calculating the single-point energies of the same torsion with different rotamers, we observe significant differences in the profile using the parameters produced by the latter fitting, as seen in Figure 3.6.b. For some conformations, energies greater than to 22 $kcal.mol^{-1}$ were obtained for several conformations of this torsion, which is 2.3 the maximum energy obtained in QM.

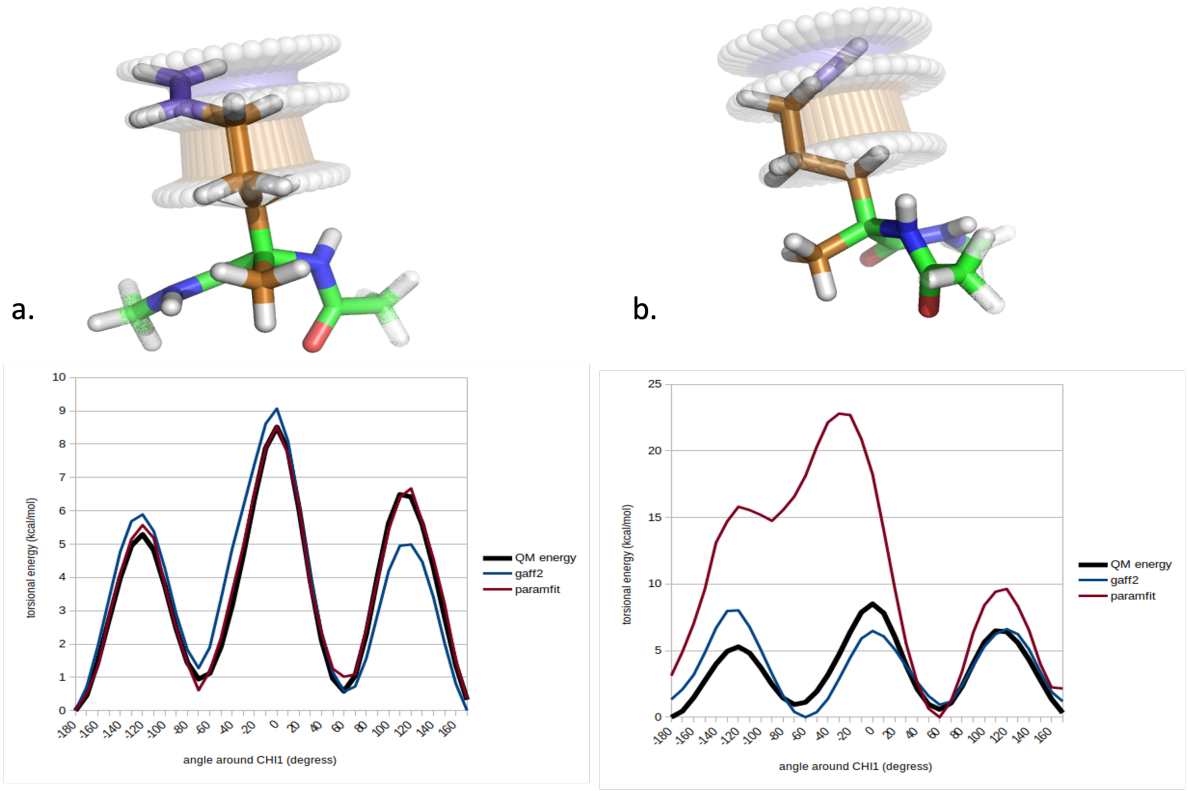



**Figure 3.6:** Comparison of MM energies calculated with Paramfit and GAFF2 parameters and QM energy for *(S)*-pentenylalanine (stapled) a. Using the set of conformations used during the fitting (training dataset) : a good fit is obtained. b. using an independent set of conformations (untrained dataset) : large discrepancies are observed with energy jumps.

### 3.4.6 A custom dihedral parameterisation protocol

Following the setbacks encountered whilst using Paramfit, we decided to build an in-house unnatural amino-acid parameterisation tool. Our approach relies on the least-square fitting method, isolating each dihedral one by one in a sequential manner. The number of required QM data points are similar to Paramfit. In our method, the single point energy for each torsion is calculated only once as our algorithm fit the forcefield parameters directly to an energy profile obtained via quantum mechanics. Single point energies are calculated at the start of the optimisation. The bonded and non-bonded MM energy components are determined and subtracted from the QM energy so that the residual energy is fitted using the dihedral MM equation. The energy resulting from the difference between the total QM energy and the MM non-bonded energy is referred as 'QM torsional energy' for simplification in the next section. In comparison with Paramfit, fittings made directly on the dihedral term, whereas Paramfit performs the fit on discrete values of single point energies.

#### Theoretical details of the parameterisation strategy

The dihedral parameterisation of the amino acids has been performed using a least-squared-fit method. For each dihedral, 36 conformations spaced by 10 degrees were produced, covering

a full rotation around the two central atoms of the considered torsion. Afterwards, the QM energy of each conformation is computed with a rigid or flexible scan, (the difference will be discussed in the next section). The MM energy is computed using the partial charges obtained with the multi-RESP procedure, Lennard-Jones and bonded terms from the GAFF2 forcefield. The amplitude terms of the dihedral angle describing the torsions to be optimised are set to zero. This MM energy is subtracted from the QM energy to give the QM dihedral energy, which will be used to fit the data.

$$E_{MM_{dihedral}}(i) = \sum_{i}^{d} \sum_{n}^{orders} A_{i,n}(1 + \cos(n_i\theta_i + \phi) + K) \tag{3.6}$$

The objective function for the fit is presented in Equation 3.6, where d is the number of the different dihedral types involved in the torsion (up to 9), orders is the maximum order of each dihedral, $\theta$ is the dihedral angle and $\phi$ the phase. Similarly to Paramfit, K is a constant to compensate for differences in the relative energy values between the quantum mechanics calculations and the Molecular dynamic calculations. It is added to the function as only the energy variation in the dihedral term is considered here. This constant is similar to the K constant also found in Paramfit and has the same function. $A_{i,n}$ is the dihedral energy barrier for a given dihedral term and order. In the AMBER forcefield, this barrier is positive for all natural amino acids present in the forcefield. $\phi$ is the phase that is equal to 0 or $\pi$. In theory, we could fit directly to this function. However, this equation produces a large number of parameters. Additionally, the phase terms are not continuous in AMBER forcefields (either 180 or 0 degrees). Instead, looking at equally spaced torsions around one bond, we can thus derive Equation 3.6. Following the same notation, Equation 3.6 thus becomes

$$\div dd\theta E_{MM_{dihedral}}(i) = \sum_{i}^{d} \sum_{n}^{orders} -n_i.A_{i,n} \div dd\theta E_{MM_{dihedral}}(i) = \sin(n_i\theta_{n,i} + \phi_{n,i})) \tag{3.7}$$

Depending on the value of $\phi$ Equation 3.7 can be simplified to

$$\sum_{i}^{d} \sum_{n}^{orders} -n_i.a_{i,n}\sin(n\theta_i)) \tag{3.8}$$

with

$$a_{i,n} = \begin{cases} +A_{i,n}, & \text{if } \phi_i = 0 \\ -A_{i,n}, & \text{if } \phi_i = \pi \end{cases} \tag{3.9}$$

Using the Equation 3.8 as the target for the least squares method requires fewer variables than 3.6 and enforce adoption of discrete phase values. The least squares fitting approach must use continuous parameters in order to converge. In our approach, the parameters are fitted on

the derivative function of the dihedral energy, to allow the use of continuous parameters.

The main limitation of this approach was that continuous dihedral profiles need to be obtained. The derivation of the energy is more prone to be erroneous when high energy conformations are derived. Procedures used to produce relevant energies profiles will be detailed in the following section.

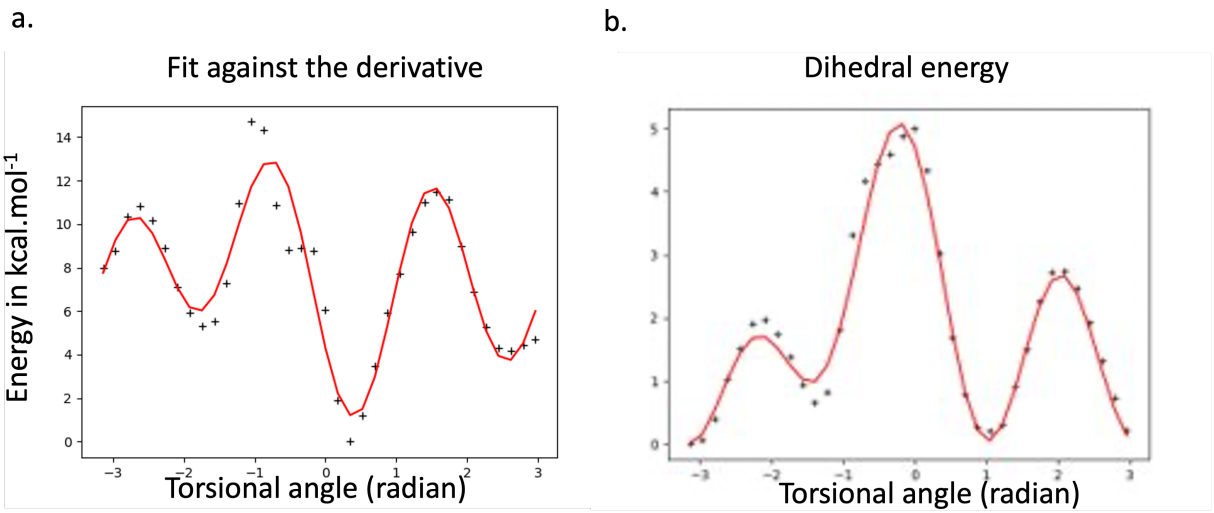



**Figure 3.7:** Dihedral parameter fitting. Fit (red line) obtained using the leastsquare method on the QM energy derivative (black crosses) and b. MM energy (red line) using the parameters obtained with the QM data (black crosses)

In Figure 3.7.a we can see an example of dihedral parameters fitted against the derivative of the energy. The target Function is reported in Equation 3.8. The phases $\phi$ for each term was then retrieved using equation 3.9. For this example, the dihedral term was then calculated using conformations upon a 360° rotation of the central atoms and plotted against the QM data (converted to kcal). Overall the plotting the dihedral derived against the QM energy produced a satisfactory fit to the original QM data, and the fitting calculations were short, at generally less than 10 s per dihedral.

### Rigid potential energy surface scan

A rigid potential energy scan has a few advantages. Rigid Potential energy surface (PES) scans are much faster to run than relaxed PES scans. Once a preliminary geometry optimisation is conducted, no further geometry optimisation is necessary which reduces the computational cost and limits convergence failures. Another advantage of not performing individual optimisation, is that all atom bond distances and angles are strictly identical during the QM scan. Thus the MM bonded terms for bonds and angles will remain constant allowing the energy variations induced by the dihedral bond rotation to be better isolated. However, care needs to be taken when choosing the initial conformation, as in most cases, taking a random conformation and rotating one of its dihedral bonds leads to steric clashes and the generation

of high energy conformations. In Figure 3.8 we can see an example of clashes found in the structure of pentenyl alanine upon rotation of the chi1 dihedral bond. While the energy profile returned by the QM calculation does not have sudden rise, high MM energy is calculated upon rotation around the dihedral, with energies peaking around 300 $kcal.mol^{-1}$, and attributed to non-bonded interactions upon decomposition of this energy using Sire. Obviously this is not a relevant energy profile to fit parameters to, as it features very improbable molecular states. Usually the energies of lower states are more relevant as they will constitute most of the states sampled during a MD simulation.

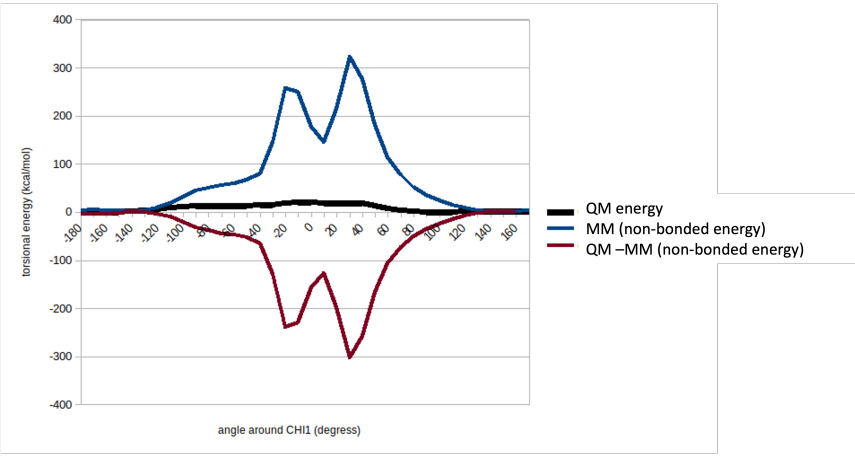



**Figure 3.8:** Clashes observed for conformations around the CHI1 torsion of (*S*)-pentenylalanine when a random conformation is chosen and no optimisation is used during QM

### Generation of low energy rotamers

One of the main difficulties in this method was the generation of low energy conformers whilst rotating a dihedral. A first solution was to check the energy of the rotamers generated before running the QM calculations and to discard series where clashes occurred upon rotation. This additional step generally led to smoother energy profiles. It should be noted that a relatively large number of initial conformations need to be produced at first (3000 conformers) for the molecule of interest. For each of the 36 structures generated during the dihedral rotation (one structure every 10°) we checked if any clashes were detected (using a van der Walls radii distance based approach).

In Figure 3.9 the energy profiles upon rotation of the dihedral bonds of (*S*)-pentenyl alanine are reported. Sampling among the lower energy conformations leads to smoother rigid PES profiles, providing smooth energy surfaces for the fitting of the dihedral.

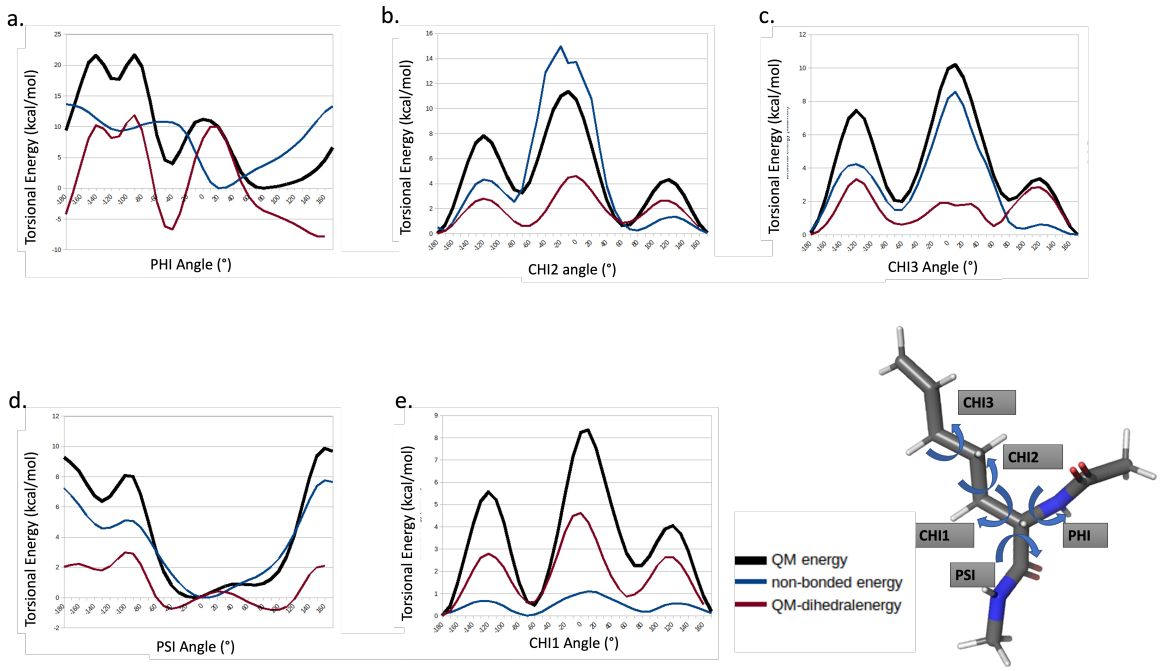



**Figure 3.9:** QM and MM energies using rigid PES scans around the different torsions of pentenyl alanine a. Phi angle, b. Chi2 angle, c. Chi3 angle, d. Psi angle, e. Chi1 angle.

### Flexible potential energy surface scan

When computing flexible potential energy surfaces, the geometry of the structure is optimised for each conformation either independently, or sequentially. During the optimisation all degrees of freedom are allowed, except for the atoms participating in the dihedral torsion of interest, which are frozen. The dihedral torsion is varied in 10° increments. Flexible potential energy scans lead to longer QM calculation times (around 2 to 18 hours per conformation for a 32 atom molecule).

In Figure 3.10.a we can see that freezing a dihedral angle while optimising the geometry of the rest of the molecule does not always produce relevant energy profiles if the scan is performed in independent runs (one per dihedral angle). Optimising the structure for each conformation independently also leads to large discrepancies in the positions of atoms. In Figure 3.10.a, The non-bonded MM energy is non-continuous due to the variety of conformations. Removing this MM energy from the PES produces a jump in the resulting 'QM dihedral energy' (at 200°). This leads to a dihedral energy profile that cannot be fitted using the dihedral term.

In Figure 3.10.b the molecule was optimised sequentially, using the scan method implemented in Gaussian. During the scan large conformational changes during optimisation can lead to abrupt energy drops, leading to artefacts in the energy profile. But overall this profile is in better accordance with the AMBER forcefield. Furthermore these conformational rearrangements can easily be identified as they produce drops in the energy, and the energy profile

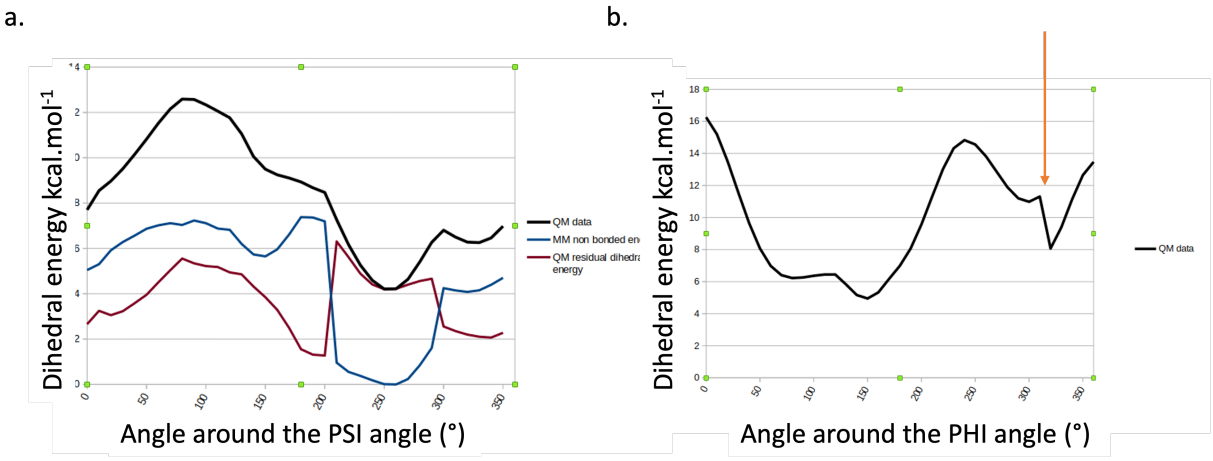



**Figure 3.10:** PES of the Phi angle in pentenyl alanine,. a. The structures have been optimised independently (separate QM runs) b. The structures have been optimised sequentially (one QM run). The orange arrow indicates a major conformational rearrangement of the structure leading to a lower energy conformation.

can be corrected. Generally these drops vary from 2 to 6 $kcal.mol^{-1}$. This last approach has been the approach used for the parameterisation of the residue library presented in Section 3.5.

### 3.4.7 Performing the fit on multiple conformations of the molecule

Once the MM bonded energy was removed from the QM energy, discrepancies were observed among the profiles for the same torsion. Thus, in this implementation, we decided for each torsion to run more than one PES scans and combine those scans during the fitting.

The script fits the dihedral term using Equation 3.6, and the dihedral angle values of the molecule (calculated once during the fit) to several torsional QM energies. This approach overall led to energetic profiles closer to the ones found in general forcefields. In Figure 3.11 the MM torsion profiles using the GAFF2 parameter and the parameters obtained with our approach are compared to the QM profiles for the pentenyl-glycine residue. In Figure 3.7, the QM energy attributed to the dihedral energy ($QM_{total-energy}$ - $MM_{nonbonded-energy}$) is deduced by subtracting the non bonded MM energy to the QM energy. To visualise how the fitting process performed we plotted the MM energy profile, using the optimised dihedral parameters and the GAFF2 parameters against the QM energy profiles (Figure 3.11). By comparing the energy differences between a set of five QM torsion profile to their MM dihedral profiles, for the pentenyl-glycine residue, the rms was 5.6 kcal.mol$^{-1}$ for GAFF2 and 3.6 kcal.mol$^{-1}$ for our in-house approach, which thus offer an improvement when compared to the general force field.

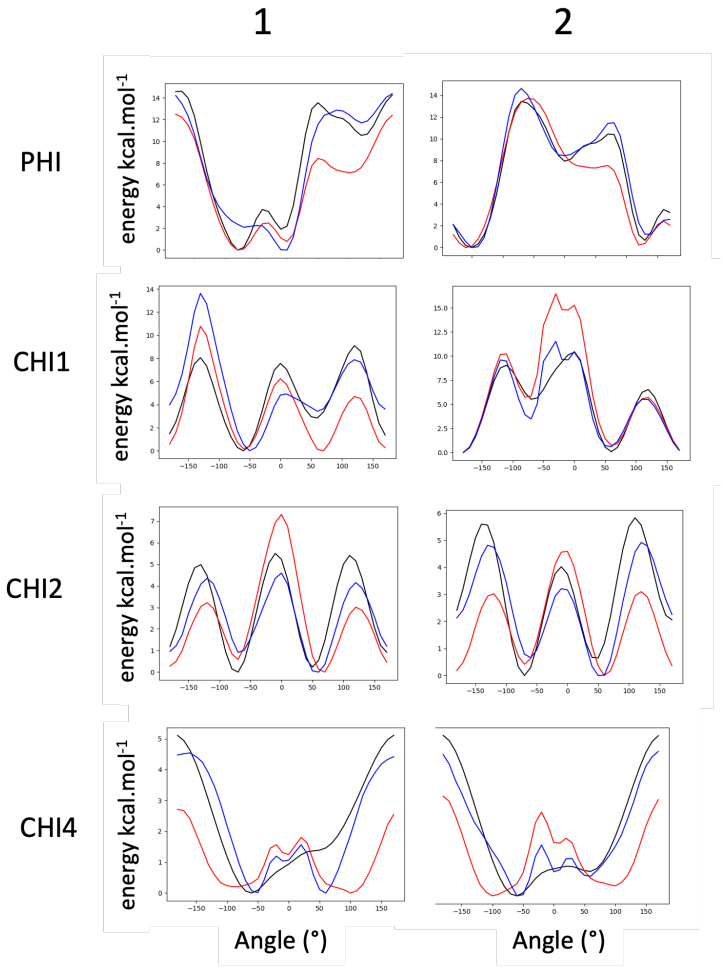



**Figure 3.11:** Comparison of QM energy with the MM energy using in-house parameter are represented in blue and GAFF2 parameters are represented in red

### Additional remarks about the implementation

The main script, 'dihedral-fitting.py', first analyses the molecule and deduces the torsion names, using the atoms names present in the mol2 file. It then generates a series of low energy conformers starting from the input conformations given in input and produces the input files for Gaussian09, using the scan function within Gaussian.

Once all Gaussian calculations are run, energies and stationary coordinates are extracted from the output files. Stationary structures from the Gaussian09 output file are used to calculate MM single point energies. Topology files are prepared where dihedral energies for the torsion are set to zero. Then using the software mdtraj, angles for every four atom combination are retrieved and using Equation 3.7 as the target function, the least square fit is conducted.

The number of dihedral terms used was based on the nature of the dihedral to fit. Dihedrals with two central $sp^3$ carbons were fit with 3 terms of order 1, 2, and 3. Planar dihedral (e.g. amide group, alkene, aromatics) were fit with 3 terms of order 1, 2 and 4. To perform the least square fit method, we used the lmfit python library from the python3 package.

The variation of energy observed in the QM profile not only depends on torsional param-

eters but also on the non-bonded interactions ( bonds, angles and every other dihedral were kept constant). Fitting the dihedral torsion directly to the QM energies would lead to over-estimating the dihedral energies. Therefore, single-point energy calculations were performed using Sire to subtract non-bonded energies from the total energies before fit.[269] Clashes were determined by computing inter-atomic distances and applying atomic element-specific cutoffs. Conformers with one or more interatomic distances under a cutoff were discarded. The general pipeline is presented in Figure 3.12

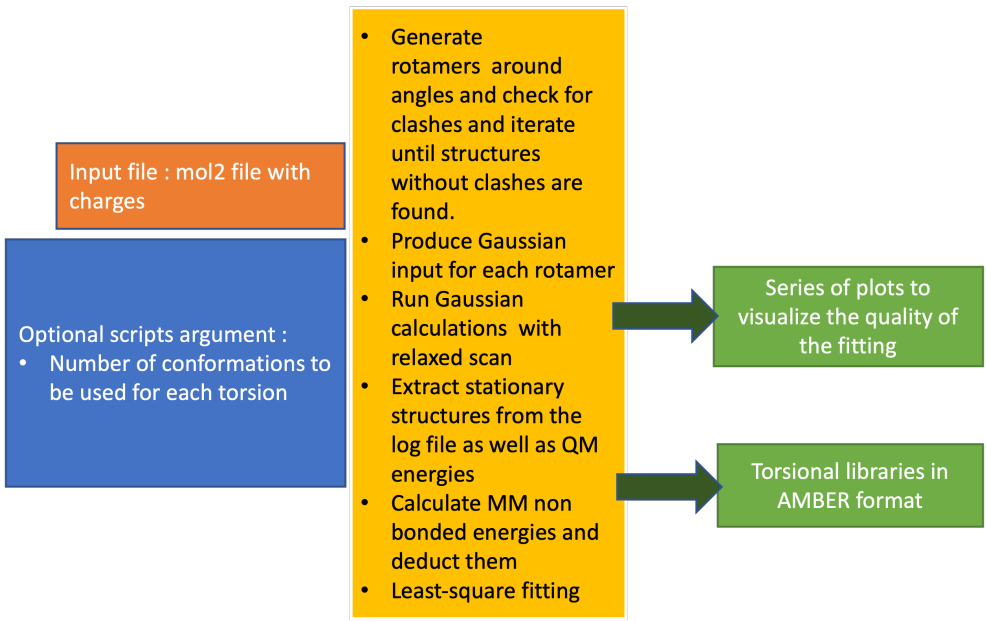



**Figure 3.12:** A general pipeline used for the parameterisation of the dihedral angle.

## 3.5 Building a library of stapled residues

A library of stapled peptides residues was prepared using the forcefield development pipeline presented in this Chapter (Figure 3.14), The parameters for these uAAs are readily available on Github. This library comprises the most commonly used staple chemistries, and includes :

- Ring Closing Metathesis (RCM) staple residues (non stapled) : (*S*)-4-pentenyl)glycine (RCM-S-G), (*R*)-4-pentenyl)glycine (RCM-R-G) (Figures 3.13.a and 3.14.a).
- Ring Closing Metathesis (RCM) staple residues (stapled) (Figures 3.13.a and 3.14.a): (*S*)-4-pentenyl)glycine (RCM-S-G), (*R*)-4-pentenyl)glycine (RCM-R-G), (*S*)-4-pentenyl)alanine (RCM-S-A), (*R*)-4-pentenyl)glycine (RCM-R-A).
- CuAAC CLICK residues : the triazole is formed during the CLICK reaction of propargylglycine and Lys-$N_3$ (Figures 3.13.b and 3.14.b).

- Cysteine Cross-linked residues (CYS). The xylene staple is formed during the alkylation of cysteine residues with *meta*-methylxylene (CYS-M) and *para*-methylxylene (CYS-P) (Figures 3.13.c and 3.14.c).

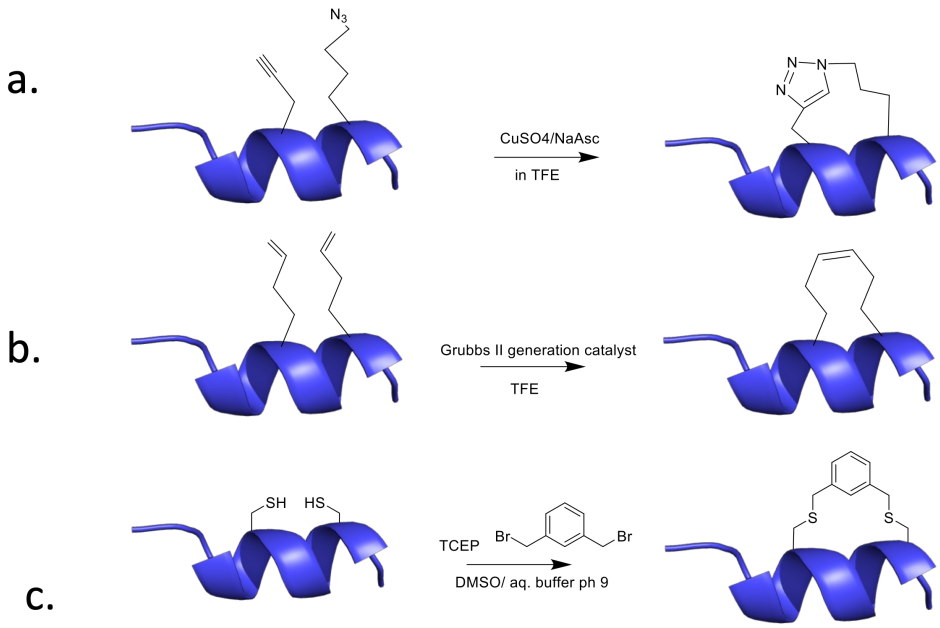



**Figure 3.13:** Most commonly used stapled chemistries : a. RCM staple chemistry, b. click staple chemistry, c. cysteine cross linking chemistry

### 3.5.1 Backbone conformational preferences of the parameterised residues

To validate the reliability of the parameters generated, MetaD (Metadynamics) simulations were carried out to establish the backbone conformational preference of the residues in our library. MetaD is an established MD method that allows the rapid exploration of free energy surfaces using collective variables (Chapter 2, Section 2.4.2). The present simulations used the backbone $\phi$ and $\psi$ torsion angles as collective variables. The well-tempered MetaD variant was used, meaning that the Gaussian potentials added become smaller with time. Post-processing of the deposited potentials allows reconstruction of the free energy surface as a function of the collective variables to produce a Ramachandran plot.

For natural amino acids, Ramachandran plots may be estimated from statistical analyses of protein structure databases. Such analysis is not possible for non-standard amino acids owing to the lack of experimental structural data. However, we can compare the Ramachandran plots of the *de-novo* parameterised amino-acids with the amino acids from the AMBER forcefield. For amino acids made of linear alkyl chains, it is reasonable to expect a profile comparable to common amino acids (Proline and Glycine excepted). Vitalini et al. al.[270]

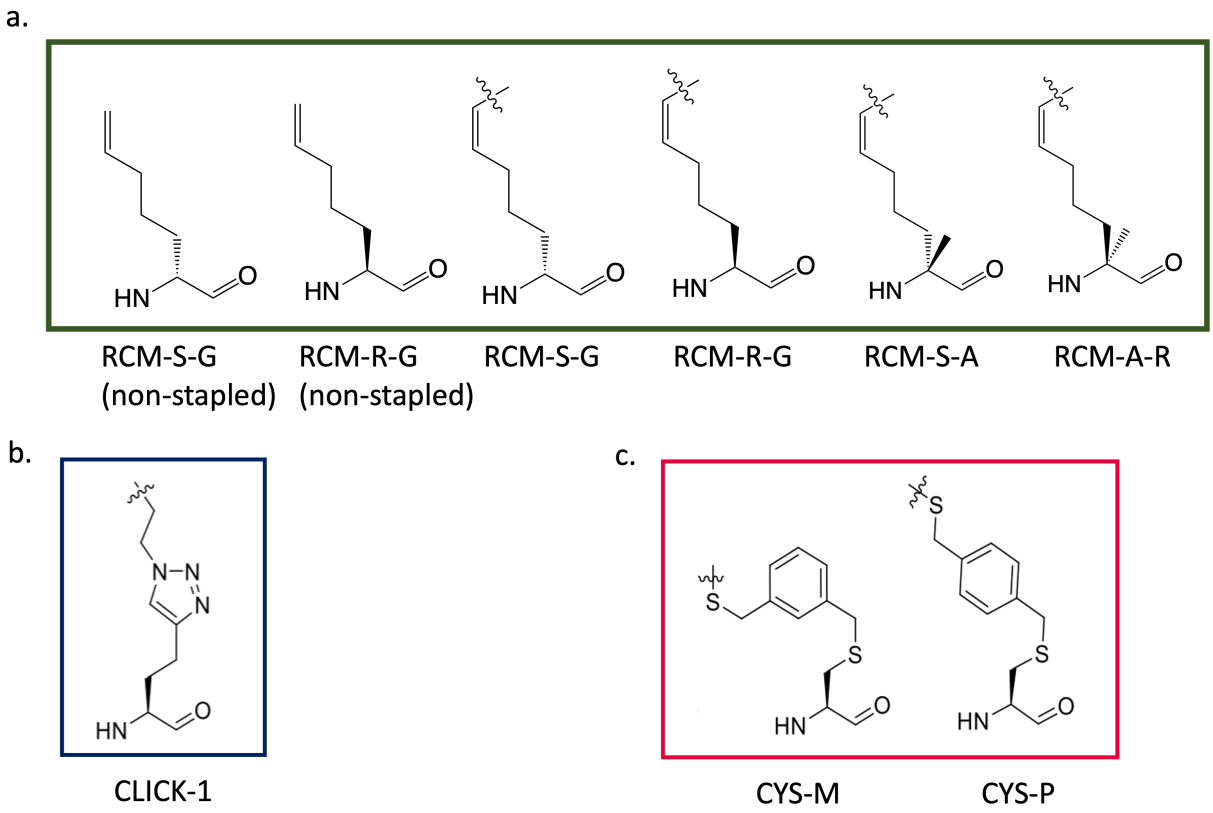



**Figure 3.14:** Stapled residues parameterised in this work by staple chemistry a. RCM chemistry b. click staple chemistry, c. cysteine cross linking chemistry

have tested the AMBER forcefield for different amino acids and report that computed Ramachandran plots were in agreement with crystallographic data. Dominant features of the free energy surface for monosubstituted amino acids include a large basin around phi-psi values of 60°/160° (corresponding to $\beta$-sheet regions), another basin is located near phi-psi values of 60°-0° (corresponding to right-handed helices), and a smaller basin near phi-psi values of 60°-60° (corresponding to left-handed helices)(Figure 3.15). Most natural amino-acids (except glycine and proline) have similar Ramachandran plots (Figure 3.15 a. b. and c.)

Figure 3.16 shows that the Ramachandran plots produced with the new forcefield parameters are very similar to those of natural amino acids described by the AMBER forcefield for the monosubstitued amino acids (Figure 3.15). It is unsurprising that Ramachandran plots resemble the Ramachandran plots of the natural amino acid as the geometry at the C$\alpha$ is similar. The residues with (*R*) stereochemistry are symmetric to the residues with (*S*) stereochemistry (Figure 3.14 a and b / c and d). Howerer, since Pentenyl alanines residues have a disubstituted C$\alpha$ (RCM-S-A, RCM-R-A in Figure 3.14 a) the Ramachandran plots are slightly different (Figure 3.16.e), exhibiting some central symmetry for both stereochemistry. In this Ramachandran plot, only two main low energy region are present. One low energy region correspond to the the $C\alpha$ region. For the monosubsitued amino-acids the RMS between the surface energies were calculated, as seen in Table 3.2. Even if paramfit performed slightly

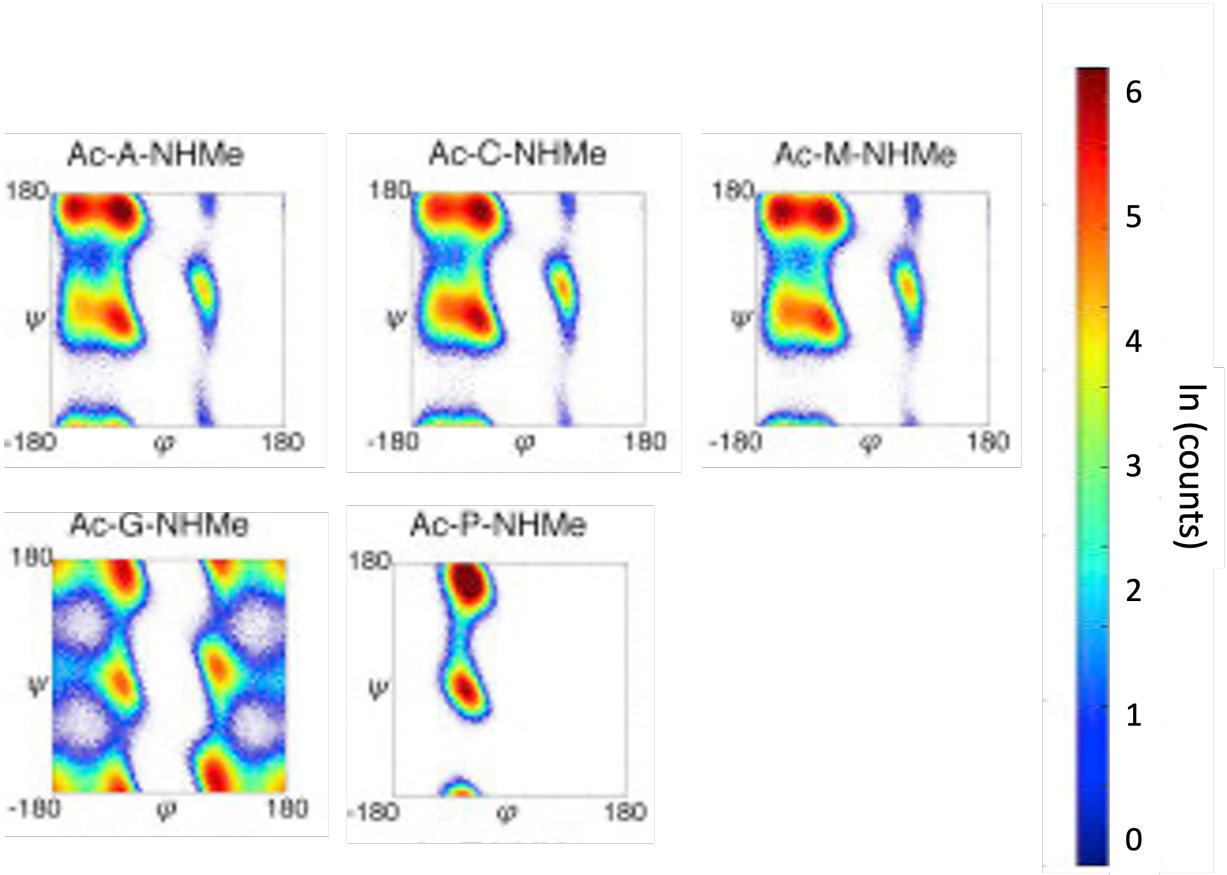



**Figure 3.15:** Ramachandran plots for residues alanine, cysteine, methionine, glycine and proline. Reproduced with permission from [270]

better for one of the residues ((S)-pentenyl alanine (stapled) with a RMS of 27.2 (kcal) against 28.7 (kcal)), the lowest RMS between de *de-novo* parameterised residues is observed for the in-house approach (22.1 kcal vs. 92.3 kcal and 25.4 kcal vs 45 kcal).

**Table 3.2:** RMS (kcal) between parameters generated with paramfit and alanine (AMBER14SB) (column Paramfit) and between parameters generated with in-house parameters and alanine (column In-house)

| | RMS (kcal) (PARAMFIT) | RMS (kcal) (In-house) |
|---|---|---|
| (S)-pentenyl alanine (unstapled) | 45.4 | 25.4 |
| (R)-pentenyl alanine (stapled) | 92.3 | 22.1 |
| (S)-pentenyl alanine (stapled) | 27.2 | 28.7 |

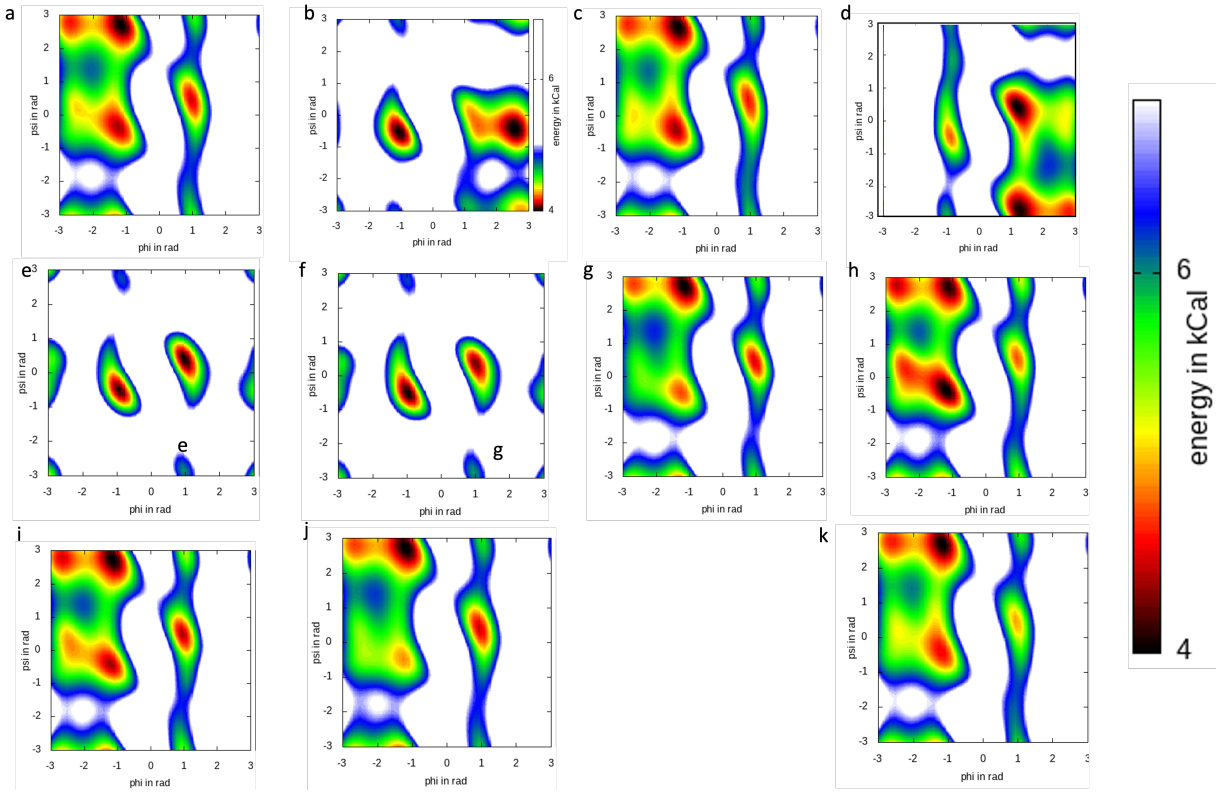



**Figure 3.16:** Ramachandran plots obtained by 200 ns metadynamic for the *de-novo* parameterised stapled residues : a. (*S*) pentenyl glycine(unstapled) b. (*R*) pentenyl glycine(unstapled) c. (*S*)pentenyl glycine (stapled) d. (*S*)pentenyl alanine (stapled) e. (*S*)pentenyl alanine (stapled), f. Click (triazole) (azido-lysine), i. Click (triazole) (proparglycine), g. cysteine *meta*-benzene, h. cysteine *para*-benzene, k. alanine (for comparison)(AMBERFF14SB)

## 3.6 Conclusion

This work has focused on the development of a pipeline for the parameterisation of unnatural amino acids. The resulting pipeline featured a high degree of automation and was shown to produce parameters more rapidly than the alternative Paramfit software, during the fitting of the parameters to the QM data. Our RESP protocol is a portable protocol to calculate RESP on multiple conformations and the partial charges obtained with this method are in accord with the partial charges obtained with the R.E.D server partial charges. Our results indicate that GAFF2 parameters were often able to reproduce QM torsional energy profiles reasonably well, and can constitute a first rapid approximation of the dihedral parameters. However some discrepancies can be observed when molecules parameterised with the GAFF2 parameters are compared with QM. In this work, we built a pipeline able to improve these dihedral parameters, and our bespoke parameters improved the fit with QM data. MD simulations performed with the resulting parameter sets indicate backbone conformational preferences in broad agreement with canonical amino acids for unnatural amino acids monosubstituted at $C_\alpha$. In contrast, disubstituted unnatural amino acids show significantly different conformational preferences, which is to be expected.

## 3.7 Methods

### Conformer generation

RDkit was used to generate a range of conformers for a given input structure.[271] The conformers generated using RDkit produces a diverse range of plausible conformations. Other software to produce the conformers such as balloon and frog2 were tested but the molecular geometry of alkene, alkyne and amide groups were not always respecting the basic Valence shell electron pair repulsion (VSEPR) rules.[272, 273]

### Meta-dynamic simulations

Metadynamics simulations were run using AMBER20 and plumed software. Simulations were prepared using AmberTools. All input solutes were solvated in a box extending at least 10 nm away from the solute's edges. Energy minimisation was carried out with 1000 steepest descent steps. The temperature was then increased to 300 K using a 200 ps NVT simulation. The pressure was next equilibrated using a 200 ns NPT simulation. Production simulations were run for 500 ns at 300K and 1 bar using the Berendsen thermostat and a V-scale barostat. Metadynamics Gaussians were deposited along the phi and psi angles every 500 timesteps with height equal to 1.2 kJ/mol and width 0.35 rad for both CVs. Hills were summed and plotted to give Ramachandran plots.

# 4

# Pipeline for the Secondary Structure Prediction of Helical Stapled Peptides

## 4.1 Introduction

In Chapter 1 we discussed how the staple position, as well as the chemical nature of the staple, has a dramatic impact on the helicity of a peptide. More intriguingly, recent studies highlighted that the stereochemistry of the C$\alpha$ significantly affects the overall helicity, and variations in secondary structure are also induced by the sequence alone.[60] Bioinformatics methodologies to predict the secondary structure of peptides are well established, but only for sequences that contain proteinogenic side-chains. There is currently no data-driven methodology to predict the secondary structure preference of stapled peptides owing to a dearth of experimental data. To address this gap, we implemented an *ab-initio* structure prediction method. Our approach features parallel MD simulations to predict the helicity of a given stapled peptide sequence. In general, because replicas can be spread among separate computer processing units (CPU or GPU), the various REMD methods are well suited for running on parallel computers. Initially, we used temperature Replica Exchange Molecular Dynamics simulations (T-REMD) as some successes have been reported using this method to reproduce the helicity of unstapled peptides.

In the first part of this project, we hypothesised that implementing a TFE/water solvent model, based on the TFE model published by Vymetal *et al.* to the AMBER forcefield would allow experimental measurements to be reproduced more accurately.[274] After discussing the principle of REMD simulations, our solute tempering implementation will be discussed. Then the REMD and REST methods will be compared for sampling of the peptide Cks1, a natural peptide. Finally, the results on helicity prediction for the p53 peptide targeting MDMX using the solute tempering method will be presented.

### 4.1.1 Principles of REMD

Replica exchange molecular dynamics (REMD) is an established simulation methodology that has been used to study peptide and protein folding mechanisms.[229–231] In REMD, a set of MD simulations (replicas) of the same system are carried out in parallel (synchronously or

asynchronously). Whilst in molecular dynamic (MD), simulations can sometimes be 'trapped' in local minima when the energy barrier to other conformations is too high, in REMD, even if a trajectory is temporarily trapped in a local minimum, the system can escape from this minimum via an exchange with a higher temperature simulation. When the temperature increases, the potential energy profiles 'flatten', and the energetic barriers lower. (see Chapter 2, Section 2.4.2)

Usually, REMD simulations include a Markov chain Monte Carlo algorithm that attempts to exchange neighbouring trajectories at a regular time interval. Instead of swapping two simulations, one at a very low thermodynamic state and the other at a high thermodynamic state, exchanges are made between an ensemble of simulations with a lower temperature difference between each other to allow a potential energy overlap between replicas. If the overall potential energy of a replica in a more thermodynamically excited state ('higher replica') is lower than the overall potential energy of a replica in a less exited thermodynamic state ('lower replica'), the exchange is accepted. To permit further exchange between replicas that possess higher and lower potential energy, a Metropolis Factor is implemented to allow exchanges with a probability proportional to the energy difference between two replicas. The average probability of exchange ($\langle P(T_1 \leftrightarrow T_2)\rangle$) between any two neighboring temperatures over the entire simulation is dependent on the energy distribution, $U_1$ and $U_2$ (equation 4.1).

$$\langle P(T_1 \leftrightarrow T_2)\rangle = \int_{+\infty}^{-\infty} P(T_1 \leftrightarrow T_2)\rho_{U_1-U_2}(u)du = \int_{0}^{-\infty} e^{Cu}\rho_{U_1-U_2}(u)du + \int_{+\infty}^{0} \rho_{U_1-U_2}(u)du \tag{4.1}$$

In Equation 4.1, $\rho_{U_1-U_2}$ is the probability of having a certain energy difference between the two replicas. Considering the Gaussian distribution of the energy, with $\sigma_1$, $\sigma_2$ and $\mu_1$, $\mu_2$ being respectively the averages and standard deviations of the energy distribution of the simulated system. For a system in equilibrium, the average energy depends linearly on the temperature as well as the number of atoms in the simulated system. A large difference in temperature implies large differences in the potential energies of the simulations to be exchanged, and thus less probability of 'successful' exchanges between simulations. The rate of successful exchanges refers to the number of total accepted exchanges in comparison to the total number of attempts and is a relatively important variable in REMD. In order to make REMD efficient, a certain rate of successful exchanges (or rate of exchanges) must be achieved.

$$P_{i,j} = min(1, exp[(\beta_i\beta_j)(E_jE_i)]) \tag{4.2}$$

Literature suggests that a exchange rate between 0.3 and 0.6 is a good range for an exchange aptempts every 10 ps.[275] If the exchange rate is below that range, sampling will not be efficient, and the REMD would be equivalent to parallel MDs using different thermostats. If the exchange rate is too high, the efficiency is not optimal either, as systems might not be

able to stabilise their conformation at lower temperatures or explore different conformations at higher temperatures. In addition fewer replicas could be used to to cover the same range of temperature. The temperature scale having a direct impact on the energy of each replica, it also has a direct impact on the probability of exchange. The probability of exchange for a tempered replica simulation can be simplified as seen in equation 4.2.

REMD can be implemented in multiple and flexible ways. Tempered-REMD (T-REMD) is a form of REMD where the individual replicas are simulated at different temperatures. T-REMD simulations have been frequently used to study the preferred conformations of peptides and proteins.[276] The ease with which coordinates between replicas may be exchanged is related to the fluctuations of the potential energy between the replicas. This REMD implementation tends to require a large number of replicas. Indeed, the standard deviation of the energy is typically linearly related to the square root of the number of degrees of freedom in the system. The degree of freedom (noted $N_{df}$) is dependent on the number of atoms and the number of constraints and virtual sites in the system, as seen in Equation 4.3.

$$N_{df} = 9N_{water-molecules} + 3N_{protein-atoms} - N_c - N_v \tag{4.3}$$

In Equation 4.3, $N_{water-molecules}$ represents the number of water molecules, $N_{protein-atoms}$ represents the number of protein atoms. $N_c$ represents the number of virtual sites and $N_v$ represents the number of constrained internal degrees of freedoms. In explicit solvent simulations a large number of water molecules are required to solvate a typical bio-molecule. An increase in temperature increases the potential energy of the system, this leads to large fluctuations in the total potential energy, especially because of the increased number of water molecules. Consequently, replicas need to have similar temperatures for their exchange to be accepted. T-REMD simulations require a large amount of computing resources, typically 40 to 100 simulations must be used, to allow overlap in energy distribution.

### 4.1.2 The solute tempering method

Hamiltonian replica exchange molecular dynamics (H-REMD) simulations are REMD simulations where replicas differ by their Hamiltonian. Here the term Hamiltonian differs slightly of the Quantum Mechanic definition as the term Hamiltonian refers to a scaling of the potential energies of the atoms of a system. H-REMD is said to be a more efficient alternative over T-REMD, allowing the sampling of systems using fewer replicas by only exciting a restrained number of atoms in the system, in contrast to T-REMD where the higher thermostat excites every atom uniformly. H-REMD can be implemented in multiple and flexible ways. A REST2 implementation will be discussed in this chapter.[277] The premise of solute tempering is that only the solute (the protein in a bio-molecular system) needs to be heated across replicas (Figure 4.1). As a result, the overall potential energy of the system does not increase between replicas. Thus, compared to REMD, fewer intermediate replicas are needed between

the highest and lowest replicas to keep the energy potential distributions of the simulations in a range where the REMD exchange rate is efficient.

**REST1 and REST2 methodologies**

The original REST method, which will be referenced as 'REST1' for clarity, was developed by Lui *et. al*,[232] and involves both a thermostat and a Hamiltonian scaling. The whole system is heated uniformly using a thermostat, including the solvent molecules. The energy of the solvent atoms is then reduced (atoms are 'cooled down') by increasing their Hamiltonian (an increase in the Hamiltonian of an atom decreases its energy) (Figure 4.1).

$$V_{total} = V_{protein-protein} + \sqrt{\frac{T}{T_0}} \times V_{protein-solvent} + \frac{T}{T_0} \times V_{solvent-solvent} \tag{4.4}$$

In Equation 4.4, $T_0$ refers to the 'target' temperature, of unbiaised temperature and $T$ to the temperature of the replica. $V_{total}$ refers to the potential energy of the system, $V_{protein-protein}$ refers to the energy between protein atoms, $V_{protein-solvent}$ refers to the energy between solvent and solute atoms and $V_{solvent-solvent}$ between solvent atoms. Thus different factors are applied for protein-protein interactions and solvent-protein interactions, while interaction between solvent atoms are not scaled.

$$V_{total} = \frac{T_0}{T} \times V_{protein-protein} + \sqrt{\frac{T_0}{T}} \times V_{protein-solvent} + V_{solvent-solvent} \tag{4.5}$$

In REST2, developed by Wang *et. al* [278] all simulations use the same thermostat, and only the Hamiltonian of the solute atoms are affected. In Equation 4.5, $T_0$ refers to the 'target' temperature, in our case 290 K, and $T$ to the temperature of the replica. $V_{total}$ refers to the potential energy of the whole system, $V_{protein-solvent}$ to the energy between solvent and solute atoms and between solvent atoms and $V_{solvent-solvent}$ to the potential energy between solute atoms. Equation 4.5 shows that the protein-protein interactions are scaled more significantly than the protein-solvent interactions.

The potential energies described in the previous equations ($V_{protein-protein}$, $V_{protein-solvent}$, $V_{solvent-solvent}$) can further be decomposed as the sum of the Coulomb, Lennard Jones and bonded interactions. REST2 was used in this work, and an implemetntation for AMBER is available on github (michellab/Stapline), compatible with the AMBER14SB (and previous versions) and AMBER19SB forcefields. Only a NAMD implementation was available at the time of study.[279]

### 4.1.3 Implementation of the REST2 method for AMBER

To scale the simulations according to Equation 4.5 we modified the forcefield parameters in the input topologies prior to running MD simulations. A Python script library was written to manipulate the AMBER topology files (relying on parmed).[280]

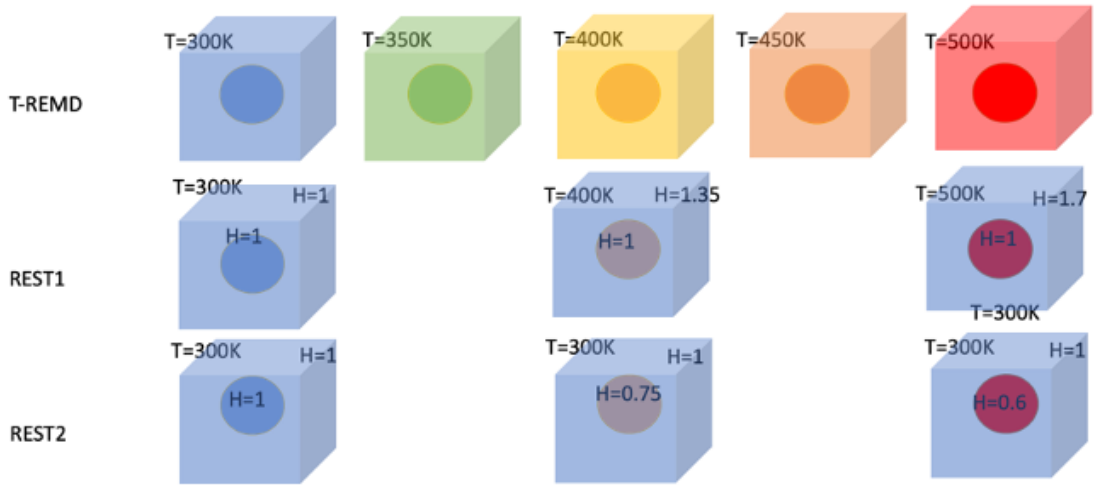



**Figure 4.1:** Overview of the solute tempering simulations: Only the atoms of the solute (protein) are heated up, in comparison to a standard T-REMD where both protein and solvent are heated up: less replicas are needed in the solute tempering method

Equation 4.5 shows that inter-molecular Coulombic interactions are differently scaled depending on whether they occur between pairs of solute atoms or solute-solvent atoms. If f is the scaling factor (with $f = \frac{T}{T_0}$), for a given replica. The Coulomb interaction between a pair of atoms i,j is given by Equation 4.6.

$$E_{coulomb} = \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \tag{4.6}$$

Thus if the protein partial charges are scaled by a factor $\sqrt{f}$:

$$q_{i,scalled} = \sqrt{f} q_i, \tag{4.7}$$

$$E^{protein-protein,\ scaled}_{electrostatic} = \frac{\sqrt{f}q_i\sqrt{f}q_j}{4\pi\varepsilon_0 r_{ij}} = f q_i q_j 4\pi\varepsilon_0 r_{ij} = f E^{protein-protein}_{electrostatic} \tag{4.8}$$

then, the electrostatic potential between protein and solvent becomes

$$E^{protein-solvent,\ scaled}_{electrostatic} = \frac{q_{i,scaled} q_j}{4\pi\varepsilon_0 r_{ij}} = \sqrt{f} * \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} = \sqrt{f} E^{protein-solvent}_{electrostatic} \tag{4.9}$$

Thus scaling the partial charges of the protein atoms by $\sqrt{f}$ allows a simple scale of the electrostatic energy according to Equation 4.5. However, for charged solutes, scaling the partial charges leads to introduction of a net residual charge in the system. Since simulations of electroneutral systems are preferable, the residual charges are re-distributed over all ions

of the opposite charge present in the system.

Similarly, protein-protein Lennard-Jones interactions must be scaled by a factor $f$ while protein-solvent atoms are scaled by a factor $\sqrt{f}$. The AMBER forcefield implementation relies on the Lorentz-Bertelot combination rules (Chapter 2, Section 2.3.1, Equation 2.10). The sigma and epsilon parameters defined to calculate the Lennard-Jones parameters for a given heterogeneous atom pair are directly calculated using a combination of the sigma and epsilon of a homogeneous atom pair. The $\sigma$ value corresponds to the distance the force takes effect, thus the $\sigma$ value was left unaffected. The Lennard-Jones interactions are scaled by changing the value of the $\varepsilon$ parameter of only the protein atoms, by the factor $f$. Indeed for Lennard-Jones interactions between two protein atoms, following the Lorentz-Bertelot-rules:

$$\varepsilon_{ij}^{protein-protein,\ scaled} = \sqrt{f\varepsilon_{ii}.f\varepsilon_{jj}} = f\sqrt{\sigma_{ii}.\sigma_{jj}} = f\varepsilon_{ij}^{protein-protein} \tag{4.10}$$

and for interactions involving one protein atom and one solvent atom:

$$\varepsilon_{ij}^{protein-solvent,\ scaled} = \sqrt{f\sigma_{ii}.\sigma_{jj}} = \sqrt{f}\sqrt{\sigma_{ii}.\sigma_{jj}} = \sqrt{f}\varepsilon_{ij}\varepsilon_{ij}^{protein-solvent} \tag{4.11}$$

$$V_{\mathrm{LJ}}(r) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right] \tag{4.12}$$

With the Lennard-Jones potential written as a function of $\varepsilon_{i,j}$ and $\sigma_{i,j}$, this potential is directly proportional to $\varepsilon_{i,j}$. Thus the Lennard-Jones potential is correctly scaled according to Equation 4.5 by scaling the protein atoms $\varepsilon$ by the factor $f$.

The scaling of bonded interactions is even more straightforward as, by definition, no intermolecular interactions exist. Thus for the protein bonded interactions, for the bonds and angles, the equilibrium distance and angle values are left while the harmonic restraints are scaled. For dihedrals, the scaling factor is applied to the dihedral barriers, leaving phases and multiplicity untouched. Improper dihedrals are treated analogously to proper dihedrals.

Cmaps are potential energy surfaces tabulated at different discrete values. A Cmap value for a given pair of $\phi$ and $\psi$ angles is determined by interpolation of the closest values in the map. The Cmap term is an energy and can be scaled by scaling every tabulated value of the energy surface in the topology file by the factor $f$. The parmed module does not yet possess the functionality to modify Cmaps, thus a parser for the AMBER topology was implemented, to scale the Cmaps fields. In Figure 4.2 we can see that applying scaling on the Cmap values flattens the surface of alanine Cmap uniformly.

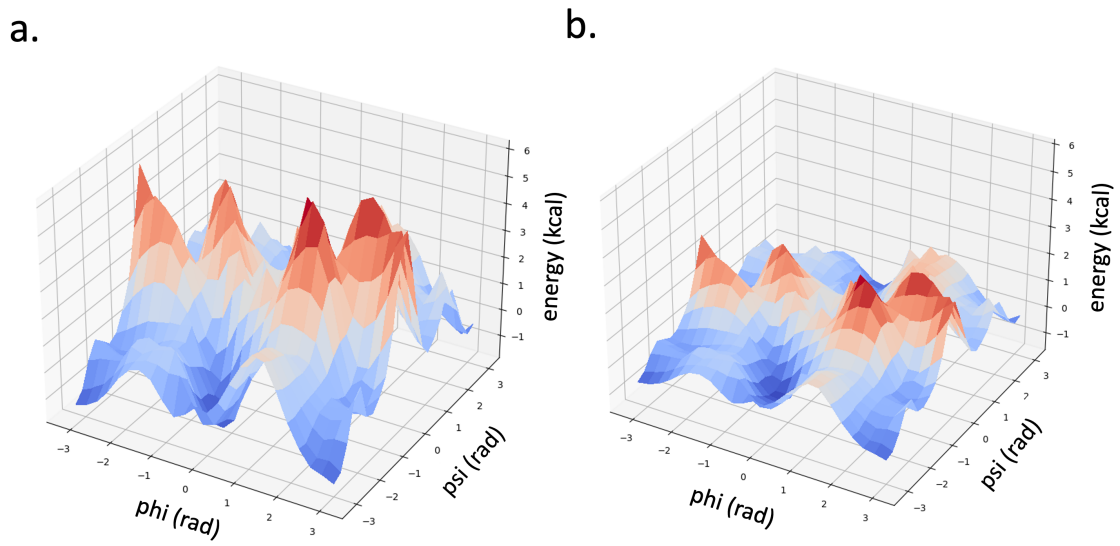
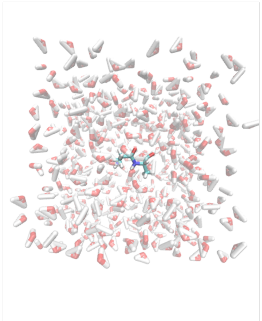



**Figure 4.2:** a. Alanine Cmap, no scaling applied, b. Alanine Cmap, scaling of 0.5 applied

### 4.1.4 Test implementation using alanine dipeptide

Our implementation was tested initially on a solvated alanine dipeptide. We used single point energy values, which are a snapshot of the potential energy for one structure of a parameterised molecule. We used the same coordinates to measure the single points, to better compare the parameter scaling. In Table 4.1 the 'expected' values were calculated as 0.6 times the potential energy for the protein-protein interaction and $\sqrt{0.6}$ for the protein solvent interaction. Every scaled term was in agreement with the expected value for this term, and no discrepancies were observed. Thus our implementation produces the expected behaviour.

a.



b.

| | | Single Point decomposition (no scaling factor) (kcal/mol) | Single Point decomposition (scalling factor of 0.6) (kcal/mol) | Expected Value (kcal/mol) |
|---|---|---|---|---|
| Solute-intrabonded angle | $x10^{-1}$ | 3.62 | 2.17 | 2.17 |
| Solute-intrabonded bond | $x10^{-2}$ | 2.06 | 1.24 | 1.24 |
| Solute-intrabonded dihedral | | 14.40 | 8.66 | 8.66 |
| Solute intrabonded improper dihedral | $x10^{10}$ | 1.21 | 0.73 | 0.73 |
| Total intrabonded solute | | 14.80 | 8.89 | 8.88 |
| LJ solute | | 2.60 | 1.56 | 1.56 |
| Coulombic solute | | -56.10 | -33.60 | -33.60 |
| Total non bonded Solute | | -53.50 | -32.10 | -32.10 |
| LJ solute:solvent | | -6.93 | -5.37 | -5.52 |
| Coulombic solute solvent | | 0.69 | 0.54 | 0.55 |
| Total non bonded Solute solvent | | -6.24 | -4.83 | -4.97 |
| Total bonded solvent | $x10^{-2}$ | 1.56 | 1.56 | 1.56 |
| Total non bonded (solvent) | $x10^{-2}$ | -33.90 | -33.90 | -33.90 |
| Total Energy of the system | $x10^{-2}$ | -32.80 | -32.70 | -- |

**Table 4.1:** a. System Studied: Alanine di-peptide in a TIP3P water box (410 water molecules). b. Single point energy decomposition of the scaled topology.

### 4.1.5 Validation of a trifluoroethanol model for aqueous simulations

#### Motivation

Alcohols have been shown to denature proteins by destabilising their tertiary structure.[281] This destabilisation is accompanied by stabilisation of secondary structure motifs. Fluorinated alcohols in particular 2,2,2-trifluoroethanol (TFE) and hexafluoroisopropanol (HFIF) are known to stabilise $\alpha$-helices.[282–284] TFE typically increases the helical content of a peptide sequence by 25 to 60%.[283] Consequently, TFE and HFIF are among the most frequently used co-solvents for peptides helicity determination in experimental studies (NMR, CD). However, the molecular properties underlying peptide folding remain unclear. Some studies using molecular dynamics suggest that the stabilising effect of TFE is due to preferential aggregation of TFE molecules around the peptides.[285] This coating would displace water, and the resulting low dielectric environment would favour the formation of intra-peptide hydrogen bonds. Because TFE interacts only weakly with non-polar residues, hydrophobic interactions within the peptides are not disrupted. Therefore, TFE promotes stability rather than inducing denaturation. A TFE/water mixture would enhance helix folding by decreasing the initial barrier for helix nucleation and reducing the energetic cost for extension of the helix; however, the origin of this mechanism is not completely understood. Vymetal *et al.* recently developed a trifluoroethanol model compatible with the AMBER forcefield and demonstrated on two model peptides (AK-17 and the antimicrobial peptide HAL-1) that TFE selectively increased the propensity for $\alpha$-helix-like conformation of backbone torsion.[274, 286] A non-complete coating of the peptide by the TFE molecules was observed in their simulations, disrupting the local water molecular network. Disruption of this network could lead to an

increase in the intramolecular peptide interactions and to an enhanced secondary structure formation.

### Model validation

Vymetal *et al.* reported parameters for their TFE model in a GROMACS compatible format.[274] These parameters were converted into AMBER input files for compatibility with the rest of our H-REMD pipeline. We carried out a thorough validation study to ensure that the conversion was correct and that the simulated properties in TFE reproduced experimental values. Density, temperature and energies did not show any unexpected large variations over time during the simulation. The first 10 ns were removed from analyses to allow for equilibration. Properties such as the density of pure TFE and mixtures, the proportion of the *gauche* and *trans* population in pure TFE, excess enthalpy of mixing and excess enthalpy of vaporisation of TFE/water mixtures were calculated.

The density of a solvent box is a good indicator of the quality of the parameters. Both in experiments and in the simulations, the density decreases at increased temperatures, which is a property of liquid phases. The boiling point of TFE is 347.1 K. Using the TIP4P/ew water model, data from the simulation was found to be in accordance with the physical values in the temperature range 290-350K (Figure 4.3).[287, 288]

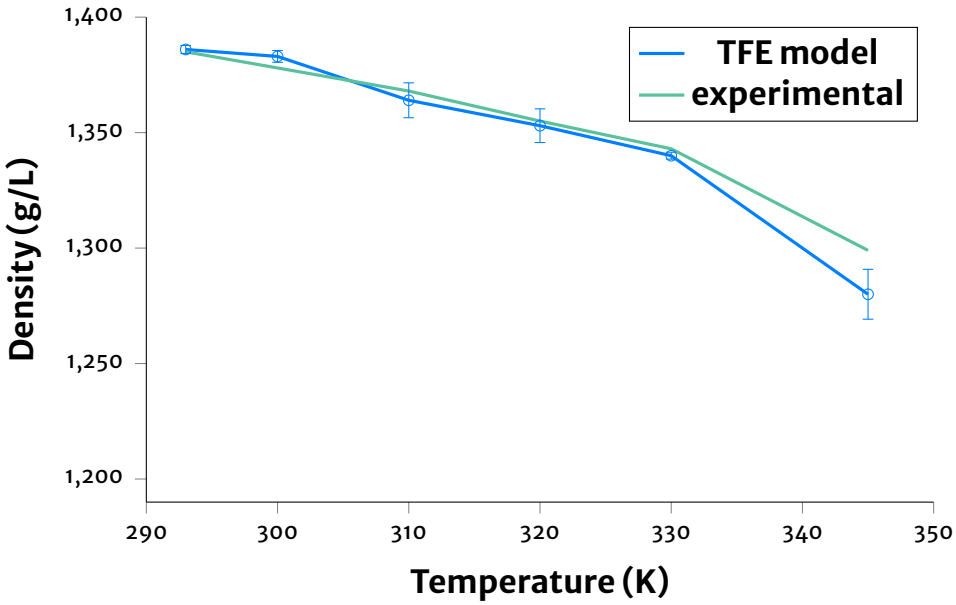



**Figure 4.3:** Density of TFE and TFE model vs temperature

Further validation of the accuracy of the parameters was obtained from the *trans/gauche* equilibrium of the O-H bond in TFE. Figure 4.4 shows the probability distribution of the different TFE conformations. The *trans* conformation is driven by steric repulsions between the hydrogen and the fluorine, while in the *gauche* conformation, weak hydrogen bond interactions occur with the fluorine atoms and stabilise this conformation. The population of the

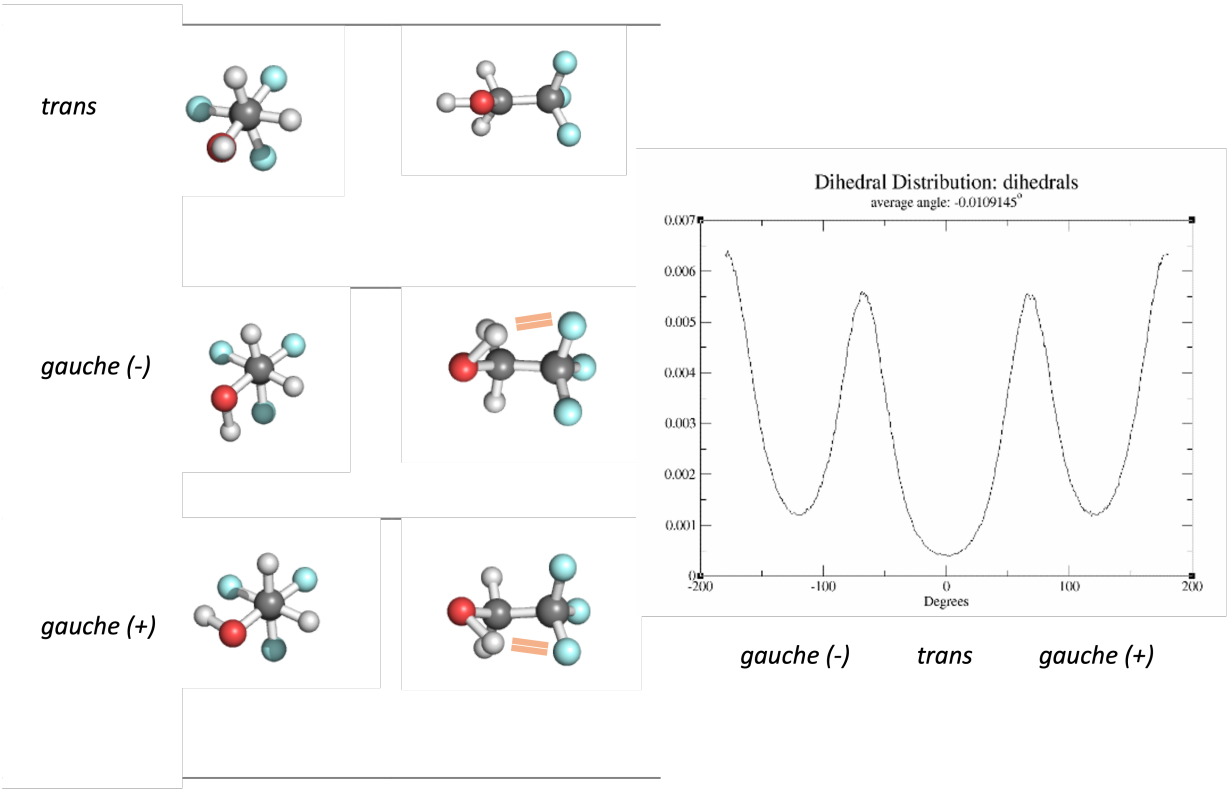



**Figure 4.4:** a. *Gauche* and *trans* conformations of TFE ; b. Distribution of *gauche/trans* conformations for the OH bond obtained during MDsimulations

C-C-O-H angle was estimated to be 39.7 % (± 2 %) for the simulation compared to 39.9 % for the literature and 40 for the experimental value (Figure 4.4). From these results, we can assume that the behaviour of the TFE molecules is in accordance with experimental data (Figure 4.4).

The enthalpy of vaporisation represents the amount of energy that must be transferred to a liquid substance to transform a quantity of that substance into a gas.

$$\Delta H_{vap} = E_{potential}/(N_{molecules}) - E_{TFE(g)} \tag{4.13}$$

Equation 4.13 was used for the calculation of the Enthalpy of vaporisation. $E_{potential}$ represents the total potential energy of the system. The Enthalpy of vaporisation was calculated at 300 K. A simulation of a molecule of TFE in a vacuum was performed to enable calculation of the second term. With a box containing 2500 molecules of TFE $\Delta H_{vap}$ was found to be of -44.2 ± 0.3 $J.K^{-1}.mol^{-1}$ which is in good agreement with both experiment (44.554 $J.K^{-1}.mol^{-1}$ ± 0.8) and previous work by Vymetal *et al.* (44.310 $J.K^{-1}.mol^{-1}$)(Figure 4.6).[274, 288]

The excess volume of mixing ( $\Delta V_{mix}$ and excess enthalpy of mixing $\Delta H_{mix}$ were calculated using Equations 4.14 and 4.15.

$$\Delta H_{mix} = \langle U_{mix} \rangle - x_{TFE} \langle U_{TFE} \rangle + x_{water} \langle U_{water} \rangle \tag{4.14}$$

$$V_{mix} = \frac{x_{TFE}M_{TFE} + x_{wat}M_{wat}}{d_{mix}} - (\frac{x_{TFE}MTFE}{d_{TFE}} + \frac{x_{wat}M_{wat}}{d_{wat}}) \tag{4.15}$$

Where $x_{TFE}$, $x_{water}$, $M_{TFE}$, $M_{water}$, $d_{mix}$, $d_{TFE}$, $d_{water}$, $U_{mix}$, $U_{TFE}$, $U_{water}$ represent respectively the molar fraction of TFE, the molar fraction of water, the molecular weight of TFE, the molecular weight of water, the density of the mixture, the density of TFE, the density of water, the enthalpy of the system and the calculated enthalpy. For TFE fractions between 0 and 0.6, values of excess volume and values of excess energy of mixing are in accordance with literature values (Figure 4.5).[288]

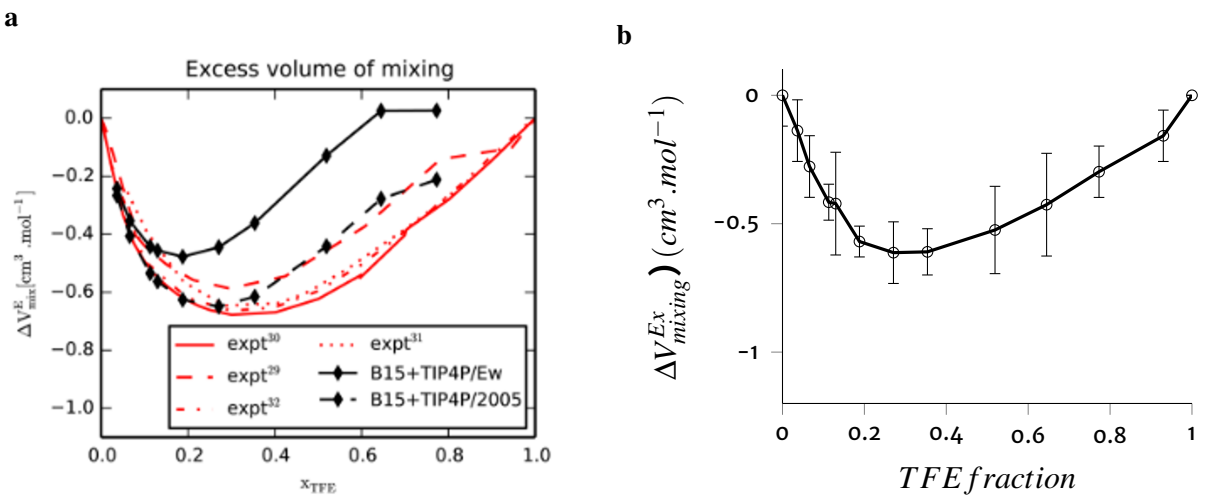



**Figure 4.5:** Excess volume of mixing of the TFE model. a. Reproduced with permission Vymetal *et al.*[274] b. Excess volume of mixing in AMBER simulations using Vymetal *et al.* model, converted to AMBER and the TIP4P/ew water model.

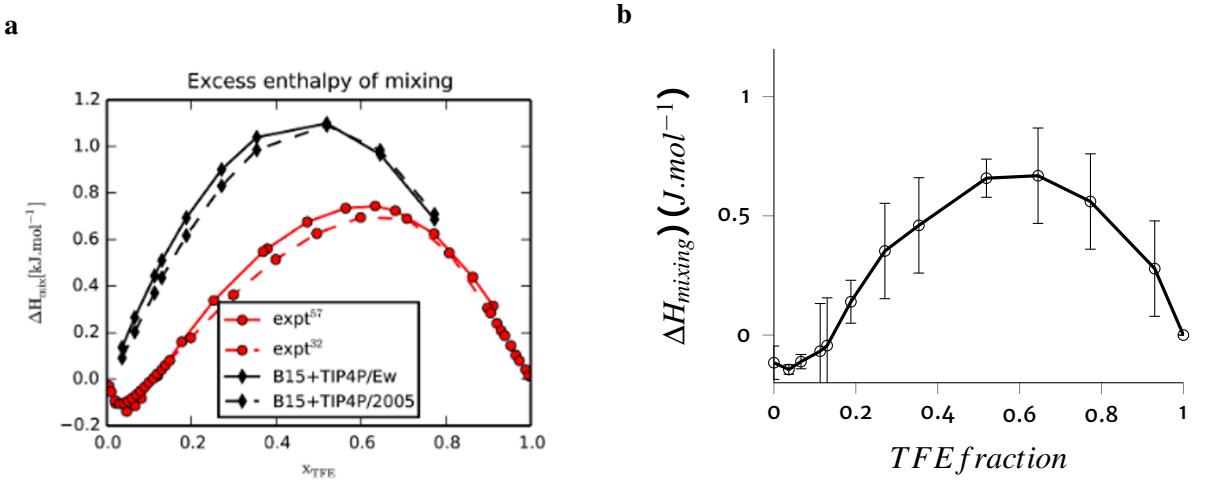



**Figure 4.6:** Excess enthalpy of mixing of the TFE model. a. Reproduced with permission Vymetal *et al.*[274] b. Excess enthalpie of mixing in AMBER simulations using Vymetal *et al.* model, converted to AMBER and the TIP4P/ew water model.

Overall the TFE model adapted from the work of Vymetal *et al.* was judged to be adequate to conduct MD simulations using the TIP4P/ew water model.

## 4.2 Comparison of T-REMD and REST2

The REMD and REST2 methodologies were tested on a short fragment of the Cks1 peptide in a TIP4Pew/TFE (75/25 %) solvent, using the TFE model presented in the previous section. REMD simulation were run using GROMACS and the REST2 simulation were run with AMBER. The objectives were to assess whether the two methodologies gave similar results, in term of sampling and helicity, and whether REST2 is more efficient than T-REMD, in terms of compute time.

### 4.2.1 Energy distributions and walk over replicas

As stated in Section 4.1.1, the exchange rate is an important variable in REMD and depends directly on the energy distribution overlap between replicas. For each methodology, replicas were generated to span temperatures ranging from 300 K to 600 K. For T-REMD simulations the temperature spacing between replicas was determined using the web tool developed by Patricksson *et al.*[289] For the REST simulations, replica's hamiltonians were evenly distributed between 1 and 0.5 (a scaling factor of 0.5 corresponds to doubling the temperature of the thermostat). The distributions of potential energy in the different replicas were plotted after running each replica for a duration of 500 ps, without exchange. A further 1 ns simulation, with exchanges, was run to determine the probability of exchange.

96 replicas were used for the T-REMD simulation to achieve a significant overlap of potential energy distributions (Figure 4.7), and the average exchange rate was 31 %. REST2 simulation, with only 8 replicas achieved an average exchange rate of 29%. Thus the REST2 simulation gave a similar exchange rate to the T-REMD, while minimising the number of replicas.

The next stage was to assess how the replicas explored the ladder of excited states. The diffusion of replicas was monitored across the temperature ladder. In Figure 4.8, the T-REMD protocol the 'first' replica (the one starting with an unbiased state) explores the full ladder, after being limited to the 30 lowest temperatures for the first 40 ns. In the REST2 protocol the replica rapidly diffuses across the ladder multiple times as the ladder only comprises 8 replicas. This also suggests that the REST2 protocol is more efficient, the replicas more rapidly explore the full Hamiltonian ladder.

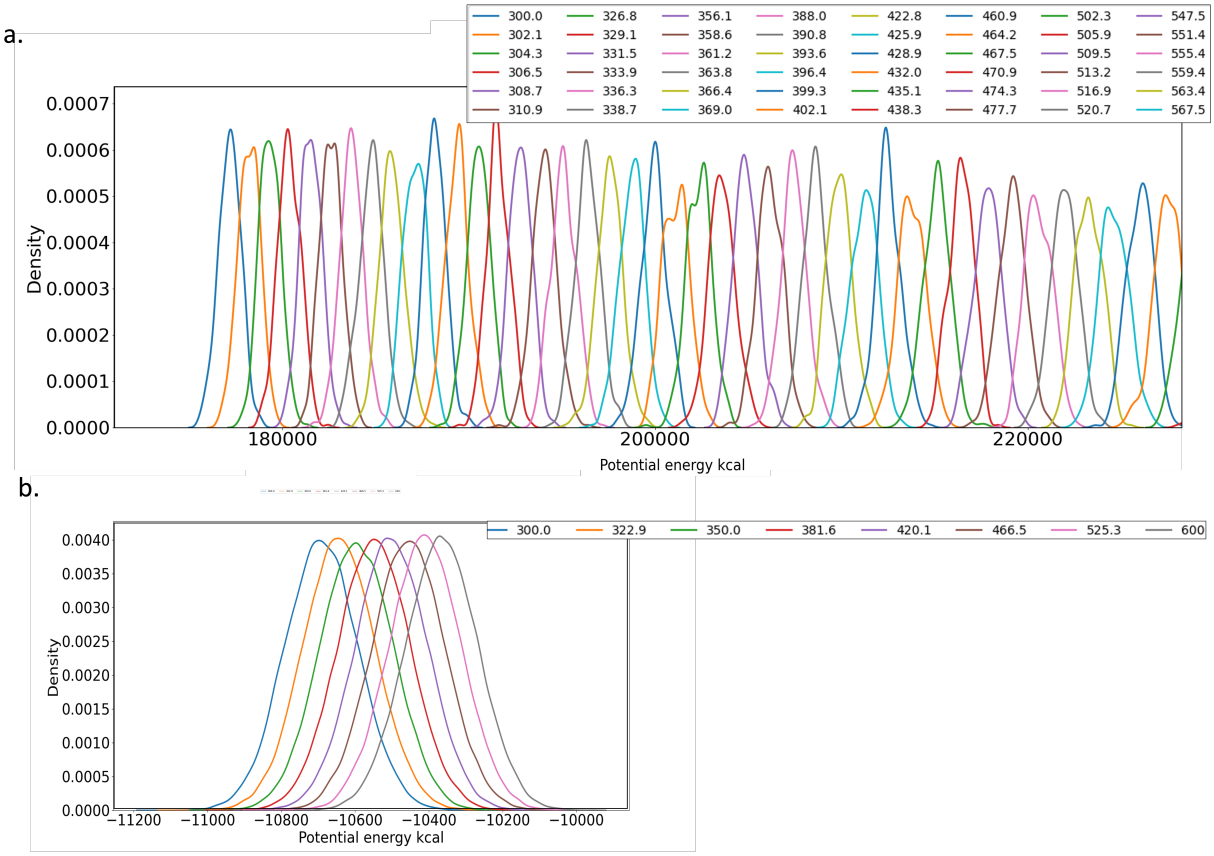



**Figure 4.7:** Energy distribution of the replicas for the CKS1 peptide. a. REMD replicas for temperatures from 300 K to 400 K b. REST replica for Hamiltonian ranging from 1 to 0.5.

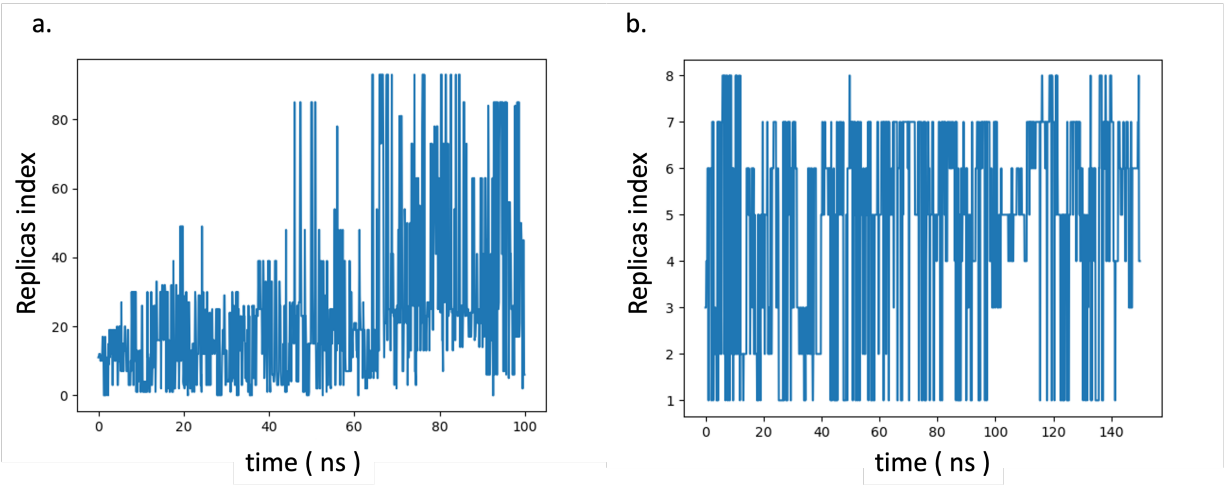



**Figure 4.8:** a. Walk of first replica in the REMD simulation. b. Walk of the first replicas in the REST simulation.

### 4.2.2 Melting curves of peptides

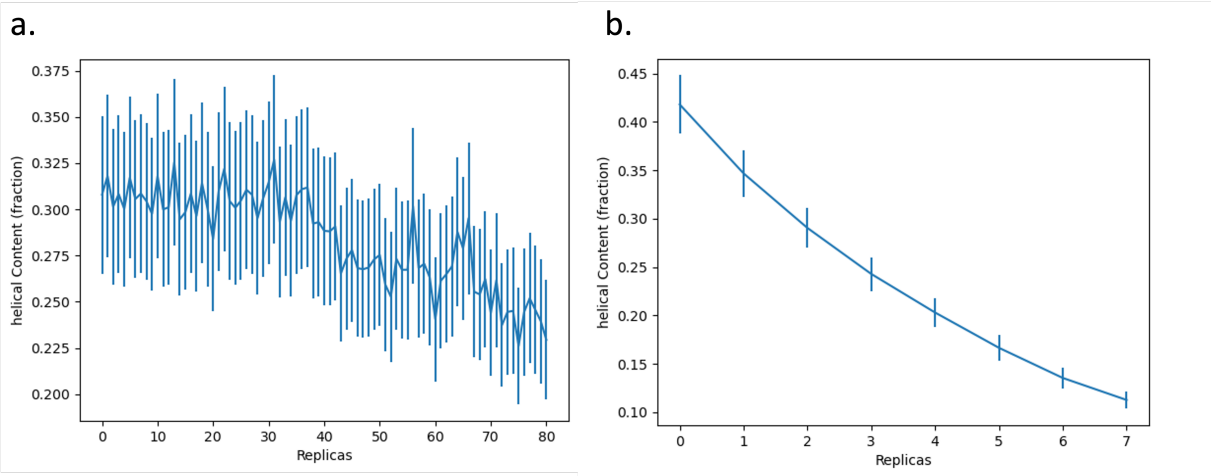



**Figure 4.9:** Cks1 peptide 'melting curve'. a. Helical propensity of the REMD replicas (over 50 ns). b. Helical propensity of the REST replicas (over 300 ns).

The helical propensity of the Cks1 peptide was plotted for each replica. The 'melting curves' obtained does not aim to correlate to experimental data. In the case of the REST simulation, more variations in the helicity of the peptides are observed, as 96 replicas of a 50 ns trajectory are used when plotting the melting curve. In the case of the REST simulation, only 8 replicas are used, using a 300 ns simulations, producing a 'smoother' curve (Figure 4.9). The propensity of helical content decreases with higher temperatures, which is in accordance with the expected experimental property of peptides. We note that the range of simulated temperatures (300-600K) is much larger than the one usually found in experiments. The general shape of the REMD plot is in accordance with the work of Feng *et al.* [290] while the curve of the solute tempering simulation differs, presenting a convex shape (Figure 4.9 b) Both helical contents of replicas at temperature 300 K are in accordance (35 % for the REMD and 41 % for REST). It is a reasonable assumption that neither of these curves reflect the actual experimental melting curve of the peptides, since higher energy replicas are used to increase peptide sampling. The helical propensity of the biased replicas (all replicas beside the lowest replicas) is not related to the actual helicity at the simulated temperatures.

### 4.2.3 Comparison of conformational sampling

A 50 ns simulation was used for REMD and a 300 ns was used for REST2. The Kmeans algorithm was used to cluster the trajectory of each replica using 15 clusters, with a sliding 5 ns window. The backbone RMSD was used as the clustering metric.

We studied how the size of the largest cluster evolves over time, during 5 ns windows, using 15 clusters. In general the largest cluster is populated similarly across the temperature ladder ( around 20 % ) for both protocols (Figure 4.10). There is however one exception, in

the REST2 simulation, the population of the largest cluster increases gradually after 150 ns. No REMD simulations with similar runtimes could be achieved with the computing resources at our disposal, so it cannot be concluded that a similar trend would be observed in REMD after 150 ns. The increase of the size of the dominant cluster could suggest that the simulation converge toward one structure for the lower replica.

In the first 50 ns we can see that the size of the biggest cluster is similar for both protocols. Thus these protocols achieve similar sampling.

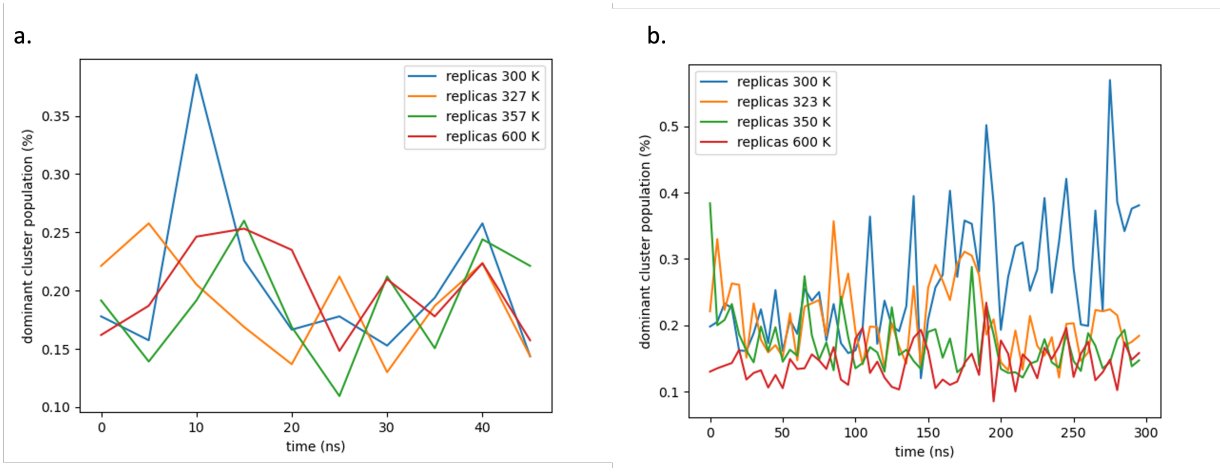



**Figure 4.10:** Percentage of the largest cluster for the a. REMD simualtion, b. REST simulation

### 4.2.4 Benchmark comparison between REST and REMD

Clusters presenting less than 20 GPU running in parallel are fairly common at the time of writing, thus running solute tempered simulations using GPU resources is possible. Simulations presented in this work were run using GTX980 and GTX1080 Nvidia GPUs and the AMBER20 MD software. Simulations speed ranged from 250-300 ns/day. In contrast the T-REMD simulations presented in this work were run on the ARCHER facilities composed of 8-core Xeon E5-2650 v2 (Ivy Bridge) series CPU, using GROMACS-16 and one CPU per parallel simulation and simulations speed ranged from 13-40 ns/day The ten-fold speed difference is mainly due to using GPU code, which was enabled by using less replicas.

## 4.3 Application of solute tempering to predict the helicity of MDMX peptides

Stapled peptides for the protein murine double minute X and 2 (MDMX and MDM2) have been widely studied in the field. The stapled $\alpha$-helical peptides were designed to compete with the p53 tumour suppressor protein in binding MDMX. The disruption of this interaction offers a means to restore the the action of p53 and ultimately induce apoptosis of tumor

cells. Notably, the MDMX inhibitor ALRN-6924 was the first stapled peptide to enter clinical trials.[291] Another stapled peptide, ATSP-7041,[45] binds both MDM2, and MDMX with nanomolar affinities ( Kd=0.9/6.8 nM ), and shows submicromolar cellular activity in cancer cell lines. ATSP-7041 is highly specific and possesses favourable pharmacokinetic and tissue distribution properties.[27, 45, 292]

The reported MDMX stapled peptides are generally designed starting from the sequence of the p53 trans-activation domain that binds the protein MDMX, introducing mutations and selecting sites for introducing a given staple. Here we compare our helicity predictions with CD measurements from Bernal *et al.* to assess the validity of our predictions.[293].

We have performed all-atom explicit solvent molecular dynamics simulations for six peptides and the sequences tested are presented in Table 4.2. We chose to simulate the peptides using a *cis* sterochemistry for the staple double bond. We selected helical (e.g. *Cis*-p53-SAH-8) and non helical peptides (e.g. *cis*-p53-SAH-3) to test if the solute tempering simulations could predict the large differences in their helicity observed experimentally upon modification of staple positioning, and single amino-acid mutations.

| Peptide | Sequence | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| p53 | Ac | L | S | Q | E | T | F | S | D | L | W | K | L | L | P | E | N | $NH_2$ |
| *Cis*-p53-SAH-3 | Ac | L | S | Q | (R)-X | T | F | S | D | L | W | (S)-X | L | L | P | E | N | $NH_2$ |
| *Cis*-p53-SAH-6 | Ac | L | S | Q | Q | T | N | (R)-X | N | L | W | R | L | L | (S)-X | Q | N | $NH_2$ |
| *Cis*-p53-SAH-7 | Ac | Q | S | Q | Q | T | F | (R)-X | N | L | W | K | L | L | (S)-X | Q | N | $NH_2$ |
| *Cis*-p53-SAH-8 | Ac | Q | S | Q | Q | T | N | (R)-X | N | L | W | R | K | K | (S)-X | Q | N | $NH_2$ |
| *Cis*-p53-SAH-8A | Ac | Q | S | Q | Q | T | A | (R)-X | N | L | W | R | K | K | (S)-X | Q | N | $NH_2$ |

**Table 4.2:** Sequences of the p53 derived stapled peptides tested in this work

### 4.3.1 Determination of the helical content of the peptides

We tested several versions of AMBER forcefields to compare their influence on the overall helicity of the peptides. Specifically, we tested the AMBER14SB/TIP3P, the AMBER14SB-so/TIP3P, and the AMBER19SB/OPC forcefields. AMBER14SB-so combines optimised side-chain parameters reported in AMBER14 with backbone parameters from AMBER99SB. This forcefield has been reported to be better suited to modeling peptides.[294] AMBER19SB is the newest AMBER forcefield which includes Cmap correction terms and has been shown to be better at reproducing the helical propensities of natural amino acids.[295] Experimental data were obtained in PBS solution with a pH 7.4.[293] To be closer to these experimental data, and because the TFE model has only be tested with the TIP4P/ew it was not used for modelling the MDMX peptides.

The overall helicity was computed over the whole sequence using the DSSP algorithm. In Table 4.3 the helicity for the lowest temperature replicas is reported. For each simulation, two repeats were performed to assess the reproducibility of the REST2 protocol.

| | exp. helicity | AMBERSB14 | AMBERSB14-so | AMBERSB19 |
|---|---|---|---|---|
| WT | 11 % | 35/42 % | 20/21 % | 25/16 % |
| *Cis*-p53-SAH-3 | 14 % | 10/05 % | 49/35 % | 15/25 % |
| *Cis*-p53-SAH-8 | 85 % | 40/45 % | 40/36 % | 62/53 % |

**Table 4.3:** Helical propensity of p53 derived stapled peptides. Experimental helical propensity retrieves from Bernal *et al.* publication.[293] The values of two independent REST2 runs are reported using AMBER14SB, AMBER14SB-so and AMBER19SB forcefields.

The reported helicity figures are from the last 200 ns of a 300 ns simulation. Capping groups were not included for the calculation of the overall average helicity. The initial input structures used across the temperature ladder consisted of helical and non-helical conformations (see Methods).

The protocol is fairly reproducible, with helicity values reproducible to within ca. 10%. Interestingly there are significant variations in helicity between different forcefields. Of the three forcefields tested, AMBER19SB is the most consistent with experimental values, predicting *cis*-SAH-p53-8 to have the highest helicity at (62/53 %) AMBER14SB-so predicted higher helicity for the stapled peptides but did not predict significant helicity difference between the two stapled peptides. AMBERSB14 was found to over-stabilise the helicity of the WT sequence. We therefore chose to continue this study using the AMBER19SB forcefield. We ran further REST2 simulations on *cis*-p53-SAH-6, *cis*-p53-SAH-7 and *cis*-p53-SAH-8A.

| | exp. helicity [293] | AMBER19 | number of folding events |
|---|---|---|---|
| WT | 11 % | 25/16 % | 70/55 |
| *Cis*-p53-SAH-3 | 14 % | 15/25 % | 19/12 |
| *Cis*-p53-SAH-6 | 14 % | 2 % | 20 |
| *Cis*-p53-SAH-7 | 36 % | 52 % | 31 |
| *Cis*-p53-SAH-8 | 85 % | 62/53 % | 72/35 |
| *Cis*-p53-SAH-8A | 39% | 10 % | 4 |

**Table 4.4:** Helical propensity of p53 derived stapled peptides for 1 or 2 independent Solute tempering runs using the AMBER19SB forcefield, and the number of observed folding events.

The number of folding events is the number of times the helicity of the peptide changes from values lower than 10 % to values higher than 80 %, or drops from an helicity higher than 80 % to less than 10 %. It is calculated on all (eight) trajectories, after processing continuous simulations. Having a high number of folding events ensures that the peptide samples multiple conformations, and that the helical propensity does not depend on the helical propensity of the initial structures. The number of folding events can be very variable (Table 4.4). The low number of folding events for *cis*-SAH-p53-8A suggests that the calculated helical propensity may not be reliable. Nonetheless, the significant decrease of helicity in *is*-p53-SAH-8A in comparison with *cis*-p53-SAH-8 is reproduced by the simulations.

### 4.3.2 Comparison with circular dichroism experiments

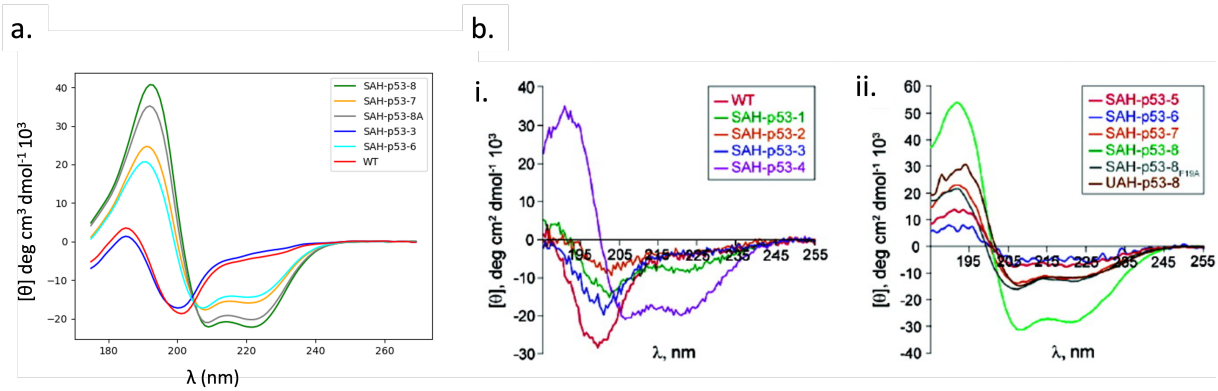



**Figure 4.11:** a. Circular dichroism spectra calculated using the software SESCA,[296] using 50 snapshots along the trajectory produced during REST2 simulations b. Experimental circular dichroism. Reproduced with permission from Bernal *et. al* [293] form the American Chemical Society (2007).

The helicity deduced from a CD spectrum depends on a broad range of parameters. A common approach is to deduce the helicity of a peptide from the intensity of the CD spectrum measured at 210 nm, but measurements at 180 nm are also reported. Temperature during the experiment can change the peptide helicity, as well as peptide concentration, salt, and solvent. Circular dichroism spectra for the conformations observed in the simulations were computed using the software SECSA (Structure-Based Empirical Spectrum Calculation Algorithm )[296] which allows the use of multiple snapshots. The secondary structure classification of SESCA is much more complex (8 types) than that used by DSSP. SESCA also includes the side-chain contributions, as aromatic residues can slightly impact the circular dichroism spectra curves. Thus we aimed to compare the experimental helicity, measured by circular dichroism (CD) measurements of p53 stapled peptides in aqueous solution with the results of the MD simulations. 50 snapshots taken at regular interval in the unbiased replica trajectory (discarding the first 100 ns of the simulation) were used to generate the CD spectra, using SESCA. When compared to the experimental CD spectra, the spectra calculated with SESCA from the simulations are mainly in agreement. Primarily, unfolded and helical peptides can be clearly distinguished. Indeed using CD, the helical structures present characteristic lower values at 210 and 230 nm while unstructured peptides present lower values around 200 nm. The SAH-p53-8 peptide, is also predicted to be the most helical peptide of the series using the REST2 protocol, which is the case experimentally. However, the helicity obtained in the simulations does not finely correlate with the circular dichroism spectra observed experimentally for each peptide and discrepancies appears in the ranking of the peptides by their helicity between the simulated and experimental values. We note that for the other helical peptides, the helical propensity is over estimated using REST2, for this series of peptides.

### 4.3.3 Conclusions and future directions

In summary, after establishing rapidly that T-REMD simulations were computationally too demanding to be run routinely without access to significant high-performance computing facilities, we implemented a REST2 methodology for AMBER simulations. Our implementation provides a generic and flexible interface for applying the REST2 method to a variety of biomolecular simulations and is compatible with the most recent AMBER forcefields. In comparison with T-REMD, the number of replicas used in REST2 simulations can be reduced by 4- to 5-fold without affecting the conformational sampling efficiency. We compared thoroughly conformational ensembles of the Cks1 peptide generated with both methodologies, showing good agreement between the two protocols. Our study should encourage the wider use of the REST2 methodology for peptide folding simulations. The solute tempering method is a reproducible rapidly converging method and is shown in this work to produce a sampling comparable to the one obtained in REMD. We have further used our REST2 implementation to study variations in helical propensities across a series of stapled peptides derived from the p53 trans-activation domain. The helical propensities of the peptides tested initially were reasonably reproducible (within 10 %). Significant variation in helicity was observed when changing the forcefield, with AMBER19SB/OPC performing the best. In general, stapled p53 peptides were predicted to have a higher $\alpha$-helical content than the WT p53 peptide, as confirmed with experimental CD measurements. The relative $\alpha$-helical propensity of most peptides in our simulations was found to be in good qualitative agreement with experimental measurements when compared with Bernal *et al.*[297] However, the DSSP algorithm sometimes misses helical looking structures, and we noticed that some of the helical structures formed by the stapled peptides were somewhat distorted from the ideal $\alpha$-helix geometry. We have also adapted Vymetal *et al.* TFE model for AMBER simulations, however this model is only compatible with the TIP4P/ew model, and, after testing different forcefields, we noticed significant variations in the helicity, induced by the choice of the forcefield. Thus this TFE model could not be used in the latter part of the project.

These positive results on the stapled p53 peptides have to be placed in context; other test systems were evaluated (not presented in this work) where the helicity of the stapled peptides could not be reproduced with REST2. Although very promising, the strong influence of the forcefields and conditions (water model, use of co-solvent), currently limits the use of such methods for the *ab-initio* prediction of the secondary structure of stapled peptides.

## 4.4 Methods

### 4.4.1 Validation of a water/TFE solvent model

A cubic box of 43 $nm^3$ containing 2000 TFE molecules was built. The first minimisation of close contacts was performed (3500 steps of steepest descent), the temperature was linearly increased over 400 ps from 50 K to 298 K in the NVT ensemble using a Berendsen thermostat. Next, the volume of the box was equilibrated under constant pressure over a 300 ps long simulation using a V-scale temperature scale, a 2-fs time-step and a Berendsen barostat. Throughout all pre-equilibration stages, the temperature was controlled by using a Langevin thermostat with a collision frequency of 3 $ps^1$, long-range electrostatics were treated using particle mesh Ewald summation, a cut-off of 4 Å was applied to all non-bonded interactions. The production runs consisted of 50 ns long simulations using a time-step of 2 fs at 300 K. Periodic boundary conditions (PBC) were used along with a Verlet cut-off scheme. The LINCS algorithm was applied to hydrogen bonds.

Seven different water TIP4P/ew/TFE mixtures were prepared by mixing a pre-calculated number of TFE molecules on preequilibrated TIP4P/ew water boxes. Composition of the TFE/water boxes for each simulation is given in Table 4.5.

| Fraction of TFE | N of molecule of water | N of molecule of TFE |
|---|---|---|
| 0.0365 | 137 | 3616 |
| 0.0661 | 230 | 3249 |
| 0.113 | 352 | 2763 |
| 0.13 | 390 | 2608 |
| 0.188 | 501 | 2163 |
| 0.519 | 846 | 784 |
| 0.645 | 915 | 503 |
| 0.773 | 970 | 285 |

**Table 4.5:** Composition of the TFE/water boxes for each simulation

The vacuum simulation was parametrised with a timestep of 1 fs. The hydrogen bonds were constrained using LINC. The simulation was run for 1 ns.

On a trajectory containing 500 molecules, the dihedral angle C-C-O-H distribution of the TFE was computed over time using a python script using the MDTraj module. Angles between -120° and 120° were considered as *gauche* conformation, while angles outside this range were considered as *trans*. The angles were summed to obtain a distribution, and the integral was obtained to give the ratio of *cis/trans* conformation.

### 4.4.2 REMD simulations

### 4.4.3 Solute tempering simulations

A total of 8 replicas were used with linear scaling factors $\in$ [0,5; 1]. An in-house python script was used to produce the topology files for the simulations. The user can directly produce the sequence of the peptide as an input. The peptide will be built by calling tleap in background. Alternatively the user can provide an AMBER topology file. Simulations were conducted using either AMBER14SB and explicit solvent TIP4P/ew or AMBER19SB and the OPC water model. A helical configuration of the peptide was prepared. After 500 steps of steepest descent, a 20 ns pre-equilibration, was run. Then system was 'melted' at 500K using NVT for 100 ns to produce a conformational sampling of the peptide. Random frames of the system were taken from the latter simulation as input for the REST protocol. 100 ns of NVT equilibration were ran before the 300 ns production was carried out. Neighbouring replicas were subject to exchange at every 500 MD steps (10 ps). The simulations were performed under NVT (constant pressure at 1 bar and constant temperature) ensembles and periodic boundary conditions. The temperature was controlled using the V-rescale algorithm.

### 4.4.4 Helicity measurements

Helicity measurements were obtained using the DSSP module of mdtraj using an in-house script. Secondary structure was assigned for all frames of the simulation and the number of helical assignments was averaged over the length of the peptides. The percentage helicity was obtained by dividing the number of helical content by the total number of residues in the sequence.

The software SESCA [296] was used to generate the circular dichroism plots.

# 5

# Development of Stapled Peptides for the Detection of SARS-CoV-2

## 5.1 Introduction

In December 2019, a novel coronavirus was discovered in Wuhan, China. This virus has since spread rapidly across the world and is still an ongoing threat to numerous regions. This novel virus was named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), and the resulting disease was named COVID-19. Given the current state of the pandemic, continued COVID-19 outbreaks are expected. Therefore, it is crucial to have a full toolbox of treatment and diagnostic tools to address the pandemic.[298–300] Cheap, rapid, sensitive and portable SARS-CoV-2 diagnostics are crucial to monitor the ongoing pandemic and prevent transmission. PCR detection tests in which the DNA in a sample is amplified and sequenced are slow and expensive, but rely on a fairly sensitive method. In comparison, antibody tests are relatively cheap but lack sensitivity, especially in the early stages of infection.[301]

Here we report the computer-aided design of SARS-CoV-2 spike protein-peptide binders for use as cell-permeable optical probes in molecular imaging studies. Such molecules could form the basis of inexpensive point-of-care diagnostic kits, be used to study the life cycle of SARS-CoV-2 in real-time, and potentially be developed further as antiviral agents. Stapled peptides can be designed and produced more rapidly than typical drug-like small molecules, and in comparison to traditional peptides, they can achieve higher bio-availability, owing to their increased cell permeability and metabolic stability. They also offer a modular scaffold that can be further functionalised for diagnostic or imaging purposes *via*, for instance, incorporation of fluorescent dyes. Such fluorescent peptides could be used in an aerosol to detect the virus on contaminated surfaces or in clinical samples like nasal or throat swaps.

In this work, we produced full length glycosylated spike protein simulations by combining several structural models and glycosylation studies, available at the time of this project (Section 5.3). We used these glycosylated MD simulations to determine the targetable sites on the protein surface, by studying the sugar coverage at these sites. We then developed three methods to design stapled peptides targeting the spike protein of SARS-CoV-2, the key entry protein to the host cell.

The first approach, which was simultaneously explored by other groups,[302, 303] relies on excising a segment of the spike protein binding partner, the human ACE2 protein. A novel method for introducing targeted mutations and adding stapled residues to the initial sequence was explored. PCA (PCA) was also applied to assess the stability of the peptides in MD simulations (Section 5.4).

The second approach, relies on designing *de novo* peptides using known binders of the spike protein. This approach is completely novel and aimed at building helical stapled peptides from non-helical residues involved in known PPI (PPI) between the spike protein and antibodies (section 5.5).

The final approach was based on the observation that two domains of the spike protein, the heptad repeat regions (HR), form coiled-coil-like structures in the pre-fusion and the post-fusion conformations of the protein. This last approach aimed at designing peptides binding to the HR2 domain in the pre-fusion conformation of the spike protein, to form a hexameric coiled-coil (section 5.6). Tools specially designed for building and characterisation of coiled-coils, such as ISAMBARD were used.[304]

### 5.1.1 SARS-CoV-2 virus

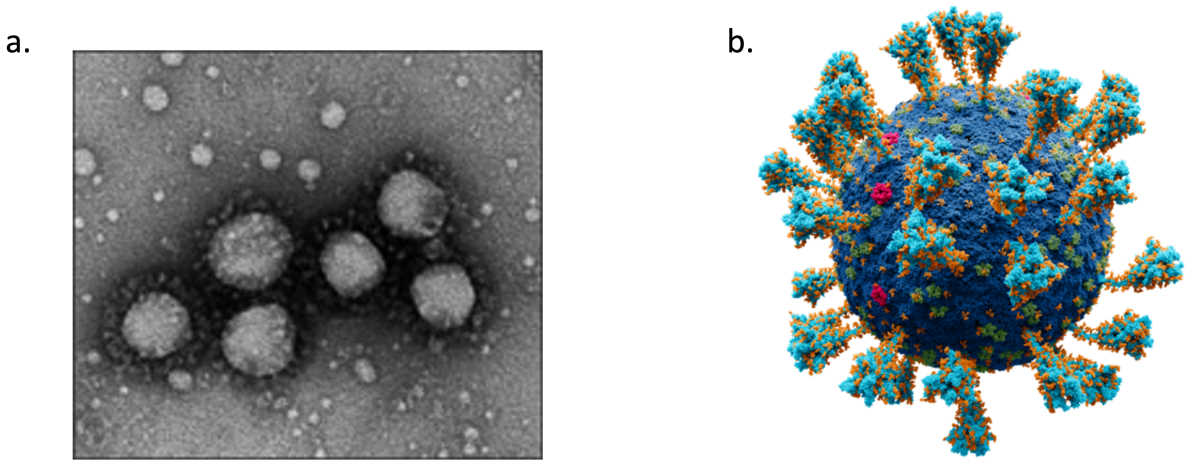



**Figure 5.1:** a. First electron microscope image of the novel SARS-CoV-2 (named 2019-CoV in the initial publication) by Jiang *et. al* Reproduced with permission from The Lancet (License 5182430651367) b. SARS-CoV-2 model. Dark blue: membrane lipids; turquoise: spike glycoproteins (with glycosylation highlighted in orange); red: E proteins, green: M proteins. Figure reproduced in agreement with the license terms CC BY-SA 4.0/Wikimedia Commons. Authors: Alexey Solodovnikov, Valeria Arkhipova

SARS-CoV-2 is the seventh $\beta$-CoV known to infect humans. Jiang *et al.* were the first to image viral particles causing COVID-19, using electron microscopy, (Figure 5.1.a),[305] before Zhu *et al.* identified and characterised SARS-CoV-2.[306] The viral genome was sequenced, showing that this new virus was 75 to 80% identical to SARS-CoV and 50% identical to MERS-CoV. Most of the proteins encoded by SARS-CoV-2 are of a similar length to the

corresponding proteins in SARS-CoV, but SARS-CoV-2 was found to be even more closely related to several bat coronaviruses (RaTG13, RmYN02, ZC45 and ZXC21) as well as one coronavirus found in pangolins.[306] CoVs are characterised by their 'crown-like' appearance due to the presence of large transmembrane proteins, the spike proteins on the virion surface. (Figures 5.1.a and b)

### 5.1.2 Spike protein structure

Spike proteins are conserved to varying degrees across the *Coronaviridae* family, proving their crucial role in the initial recognition of the host cell and later fusion of the virus membrane with the host cell. The spike protein of SARS-CoV-2 shares more than 90% amino acid identity with SARS-CoV.[307] Spike proteins are class I fusion proteins containing 1273 amino acids, which assemble into trimers. Before fusion, each trimer can be divided into three main structural domains, referenced as the head, the stalk and cytoplasmic tail (CT). (Figure 5.2) The spike protein can also be decomposed into an N-terminus S1 domain and a C-terminus S2 domain. These domains are separated by a furin cleavage site (Figure 5.2). The S1 domain is mainly responsible for primary receptor recognition and binding, while the S2 domain is mainly responsible for anchorage and fusion with the host membrane.[308–310]

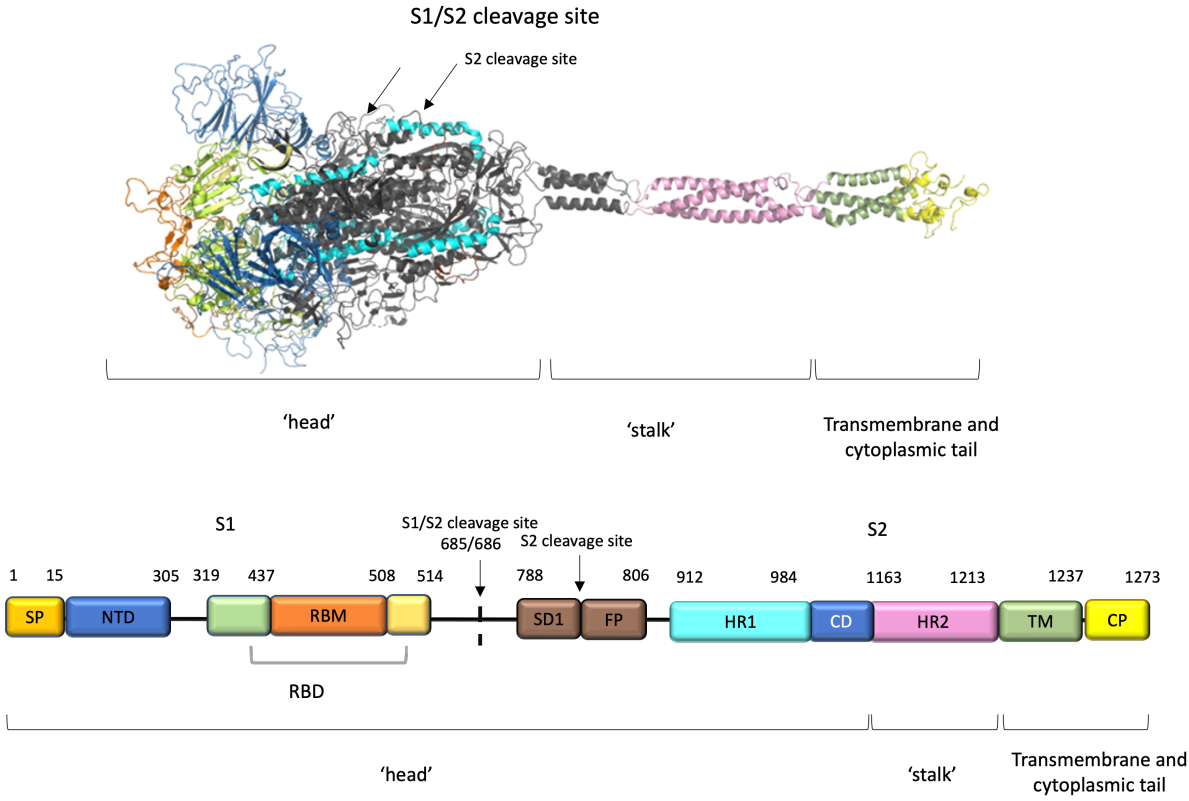



**Figure 5.2:** Representation of the different structural domains, head, stalk, transmembrane and cytoplasmic tail, in relation to the sequence. The different domains have been colored similarly in the structure and sequence.

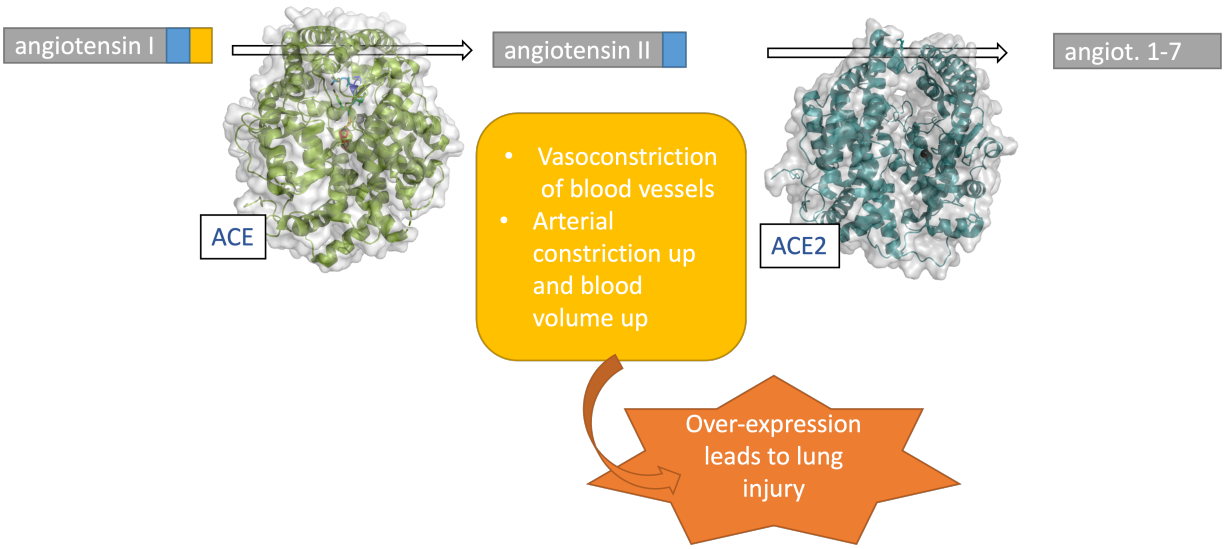



**Figure 5.3:** Role of ACE (PDB 4APH) and ACE2 (PDB 1R42) in the conversion of Angiotensin peptides.

### 5.1.3 ACE2 the main spike protein partner

Similarly to SARS-CoV, ACE2 was the first characterised binder of SARS-CoV-2 spike protein. Crystal structures were rapidly published which gave insight into the PPI between these proteins, further proving the recognition and binding of the spike protein with the ACE2 protein.

ACE2 is expressed by human host cells in various tissues. In the respiratory system, ACE2 is found in type I and type II alveolar epithelial cells in normal lungs and is more abundantly expressed in lungs with fibrotic activity. Furthermore, ACE2 is present in nasal and oral mucosa and the nasopharynx, in enterocyte cells (found in all parts of the small intestine), in the basal cell layer of the epidermis (skin tissues), in endothelial cells and in the membrane of fat cells surrounding various organs.[311] ACE2 is homologous to ACE, both of which are central enzymes in the Renin-Angiotensin Aldosterone System (RAAS). The RAAS is a crucial regulator of systemic blood pressure and has a role in multiple organs. After conversion of the angiotensinogen peptide to angiotensin I by renin, ACE is responsible for the conversion of angiotensin I (a 9-mer peptide) (Figure 5.3) to angiotensin II (an 8-mer peptide). In the lung, angiotensin II influences the pathogenesis of lung injury. Increasing ACE and angiotensin II levels leads to pulmonary hypertension, sarcoidosis, idiopathic pulmonary fibrosis, and acute respiratory distress syndrome. ACE2 acts as a counter-regulatory mechanism by converting angiotensin II to angiotensin(1-7) ((a 7-mer peptide). It has been suggested that the spike protein binding to the ACE2 receptor deregulates the RAAS system, as similar lung lesions have been observed in COVID-19 patients and in patients presenting with ACE2 knock-out.[312]

### 5.1.4 The ACE2 - spike protein PPI

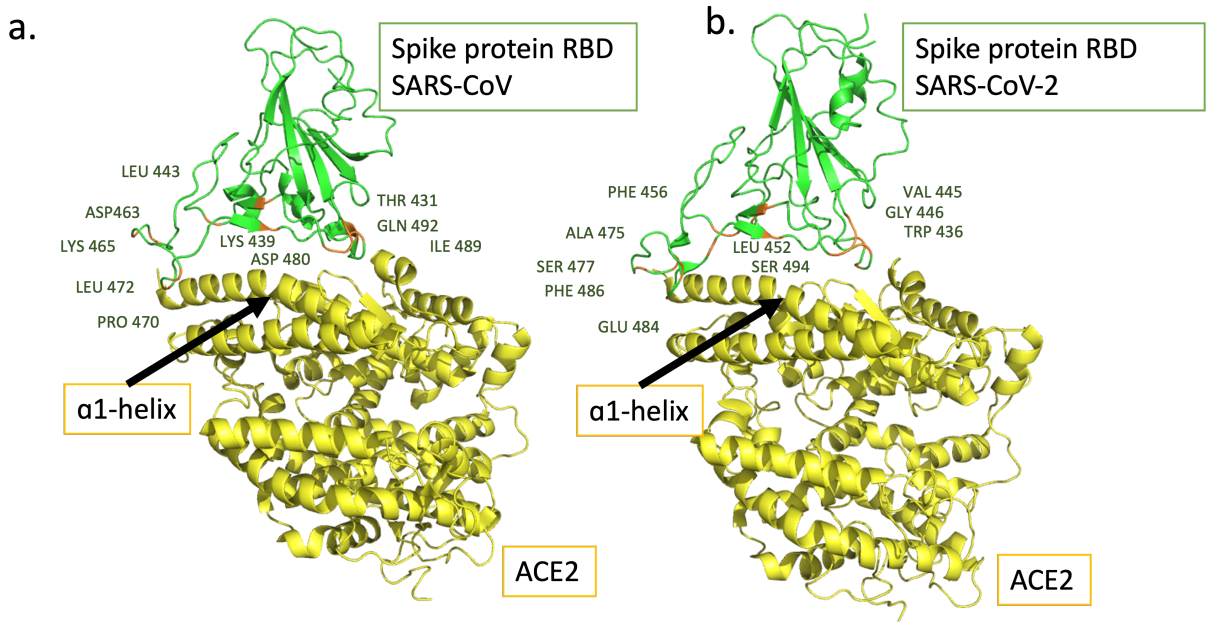



**Figure 5.4:** Amino acid mutations on the SARS-CoV and SARS-CoV-2 PPIs surfaces with ACE2 a. SARS-CoV bound to ACE2, (PDB 6ACK), b. SARS-CoV-2 bound to ACE2 (6M0J); The spike protein RBDs are represented in green with mutations highlighted in orange, ACE2 is represented in yellow

SARS-CoV-2 binds to the ACE2 protein with a protein-protein contact area of only 255.0 Å$^2$.[313] The spike protein mainly interacts with one helical segment of ACE2, the $\alpha$1-helix (residues 19-40) (Figure 5.4). Structural analysis of the hACE2-spike complex (PDB 6M0J, 6M17) reveals that a total of fifteen residues in the hACE2 RBM motif are in contact with the spike protein, of which ten are from the $\alpha$1-helix one from the $\alpha$2-helix (residues 56-83), and the other four from the adjacent $\beta$3-/$\beta$4-strands (residues 347-360) and random coil (residues 323-325). The backbone alignment between SARS-CoV and SARS-CoV-2 RBDs is high, but the binding site of SARS-CoV-2 presents several mutations when compared to SARS-CoV (Figure 5.4), and only a quarter of the PPI interactions are shared with SARS-CoV. The main interfacing positions in the external sub-domain of SARS-CoV are residues Leu472, Arg479, and Tyr487. In SARS-CoV2, these interactions are replicated in SARS-CoV-2 through residues Leu455, Phe486, Gln493, Asn501, and Tyr505.[314] The mutations of these key residues (Leu472, Phe486)(Figure 5.4) might be responsible for the higher affinity of SARS-CoV-2 for ACE2 when compared with SARS-CoV, which could be an aspect of the higher virulence of SARS-CoV-2. SARS-CoV-2 RBD binds to ACE2 with $\mu$molar affinities ($IC_{50}$ value of 3.8 $\mu M$ and $K_D$ value between 30 and 50 nM),[315] which is 5– to 10-fold lower than that of SARS-CoV RBD ($K_D$ value between 180 and 400 nM).

### 5.1.5 Conformational change of the Spike protein upon binding with ACE2

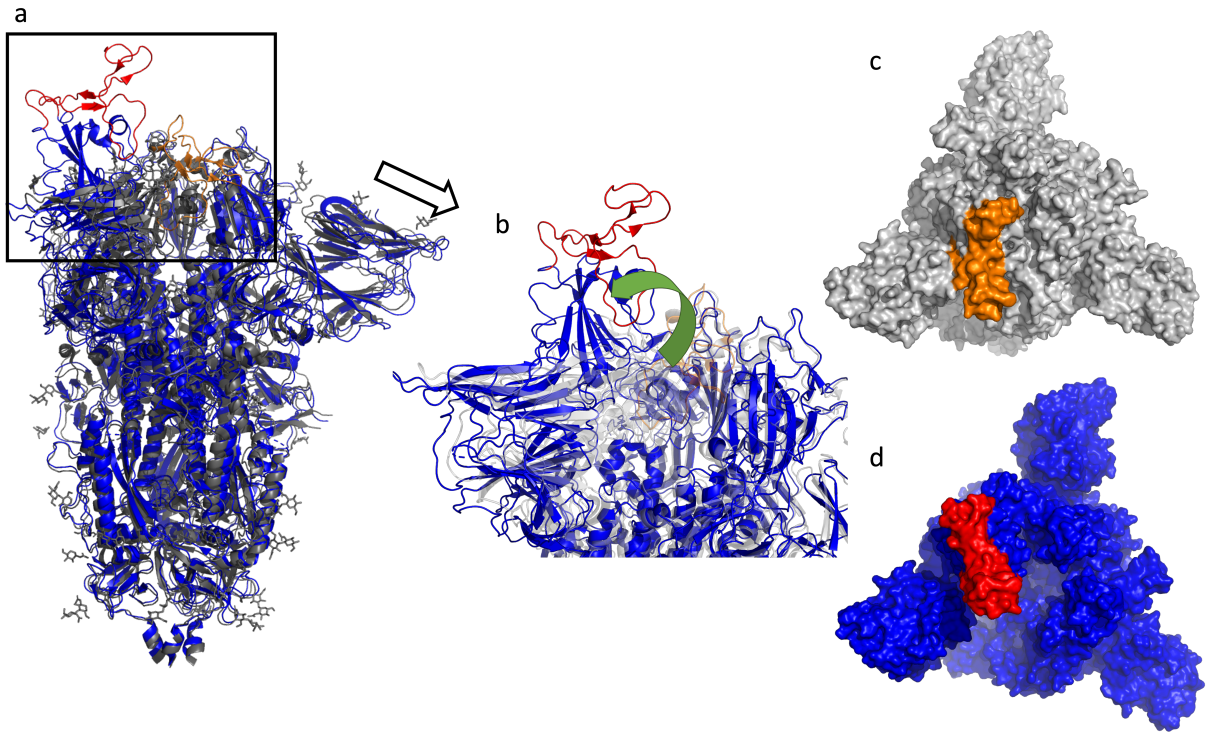



**Figure 5.5:** Overlap of the spike protein in the closed (6ZGG) and opened conformation (6ZGE). The "closed" conformation is represented in grey with the RBM in orange and the opened conformation is represented in blue with the RBM in red a. Head domain of the spike protein b. zoom-in: the RBD movement is highlighted with a green arrow, c. top view of the spike protein in the closed conformation d. top view of the spike in the opened conformation.

The RBD domain of the pre-fusion spike protein of SARS-CoV undergoes a significant conformational change from a "closed" state to an "opened" state (Figure 5.5.a and .b). In the 'closed' conformation the three RBM motifs of the spike are inaccessible. (Figure 5.5.c) The RBD position and orientation is altered through a 'hinge' movement in the 329-336 and 522-529 loops. In the 'opened' conformation, the orientation of the RBM changes and becomes accessible, allowing the binding with hACE2 (Figure 5.5.d).[316, 317] Different rates for this conformational change to occur have been reported in the literature and the values reported vary significantly. Some studies report that 30 % of the spike proteins on the virion surface present at least one RBD in upright position while other studies report values of over 60 %.[317–323]

### 5.1.6 Other spike protein receptors

The exclusive interaction of the SARS-CoV-2 spike glycoprotein with ACE2 might not be enough to explain the 'cytokine storm' experienced by more severe case patients. The bind-

ing of SARS-CoV-2 with additional receptors could compound inflammation and cytokine release, leading to the severe inflammation observed in COVID-19 patients.

Amraie *et al.* suggested that the Myeloid C-lectin like receptor (CLRs) could serve as alternative receptors to hACE2 in non-susceptible cells,[324] but other studies did not report similar results.[325] The interactions of SARS-CoV-2 with CTLs and ensuing endocytosis by resident innate immune cells offers a possible mechanism for asymptomatic COVID-19 clinical manifestation as a result of decreasing the number of free SARS-CoV-2 virions to interact with ACE2.

It is also strongly suspected that SARS-CoV-2 virions interact with integrins through a conserved RGD motif (Arg403-Gly404-Asp405), identified in all SARS-CoV-2 sequence variants of the spike protein. The exact location of the binding site for integrins varies but remains in the proximity of the ACE2 binding site.[326] This RGD motif is also present in many proteins and peptides interacting with integrins.[326] Interestingly, the RGD motif is absent in SARS-CoV could explain the higher virulence of SARS-CoV-2, especially in tissues where ACE2 levels are low. Although the induction of angiogenesis is most often associated with the action of vascular endothelial growth factor (VEGF), at least six integrins, $\alpha V\beta 3$, $\alpha V\beta 5$, $\alpha 5\beta 1$, $\alpha 2\beta 1$, $\alpha V\beta 1$ and $\alpha 1\beta 1$, have been implicated as contributors. If the spike protein of SARS-CoV-2 exhibits affinity for integrins, it may cause dysregulation of angiogenesis, leading to the observed lung injuries in COVID-19 patients.

Other proteins such as Tolls receptors, GRP78, LOX-1 have been reported to interact with SARS-CoV-2 proteins and are linked to potentially facilitating its cell binding and infectivity.[327] However, their exact roles are still under investigation.[308] No Crystal structures of the spike protein in complex with any of these secondary receptors have been yet published.

### 5.1.7 Cell entry mechanism

After binding to the ACE2 protein, the first stage of virus cell entry is mediated by a series of proteolytic cleavages. SARS-CoV-2 possess a cleavage site in the S2 domain site (S2')(Figure 5.2). Transmembrane protease, serine 2 (TMPRSS2), belongs to the serine protease family and has been found to enhance the uptake of the virus by cleaving the protein in this second cleavage site.[309] Intriguingly, SARS-CoV-2 possesses a poly-basic cleavage site (RRAR), absent from SARS-CoV-1. This site (S1/S2) is present in other coronaviruses, with a slightly different motif. It has been proven that this specific motif could lead to a more efficient cleavage by furin which might, in turn, lead to a facilitated fusion with cell membranes.[328] Bestle *et. al* have reported that the cleavage at both sites, S2' and S1/S2, is indispensable for fusion to occur.[329] The role of the cleavage sites is to prime the spike protein, which induce the dissociation of the three S1 sub-units from the spike. Other non-essential proteases have been shown to promote the cell entry of the virus by cleaving part of the spike protein.

One example is Neuropilin-1 (NRP1), which is a membrane-bound co-receptor to a tyrosine kinase. The binding site of Neuropilin-1 has been identified in the vicinity of the S1/S2 cleavage site.[330] Neuropilin-1 has been found to increase the viral entry of SARS-CoV-2 into the host cells, especially in the presence of ACE2 and TMPRSS2.[331, 332] However, the exact molecular mechanism of the NRP1-mediated effect on SARS-CoV-2 infectivity has not yet been fully elucidated.

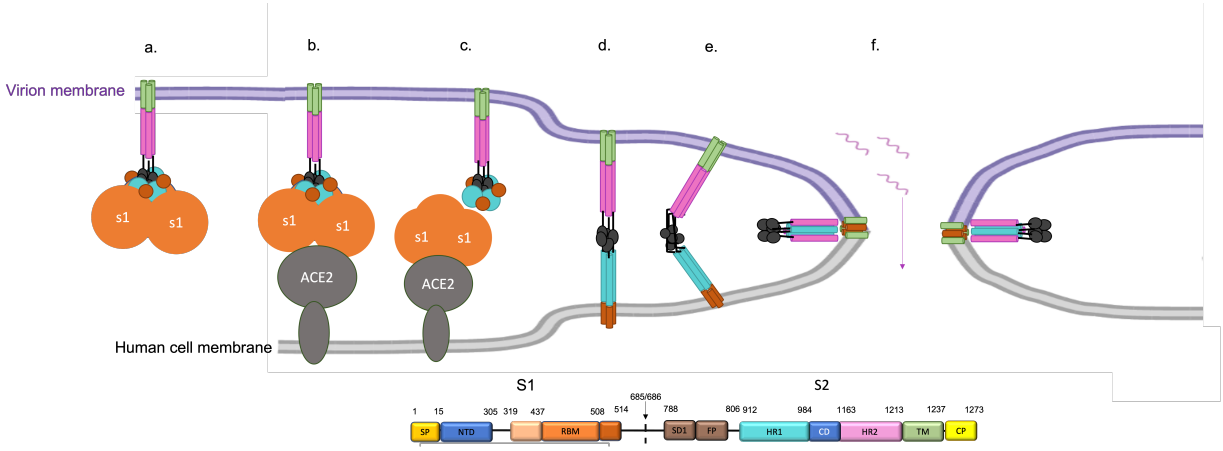



**Figure 5.6:** The different stages of the virus fusion with the human cell membrane a. Approach of the virus with to membrane; b. Binding of the S1 domain to the ACE2 receptor, c. Cleavage of the S1 domain, d. Conformational rearrangement of the HR2 domain and anchorage of the FP domain in the membrane, e. Second conformational rearrangement, virion and human cell membrane are brought closer f. S2 domain in post-fusion conformation, and creation of a pore, entry of virion genetic material.

After recognition of the spike protein and the ACE2 on the host cells (Figure 5.6.a and b the S1 domain detaches from the S2 domain,5.6.c and the remaining S2 trimer undergoes a series of conformational rearrangements to trigger membrane fusion between the host cell membrane and the viral envelope (Figure 5.6.d-f).[333] The fusion peptide (FP), located at the extremity of the S2' cleavage site, is a fairly conserved hydrophobic sequence present in multiple CoVs, which role is to anchor it to the host membrane. In the pre-fusion conformation, three long helices in the spike head protect the fusion peptide by burying it inside the spike protein head, almost at the interface between the three monomers. During fusion, an unstructured linker becomes helical, inducing the formation of a helical segment in the N-terminus. In this pre-hairpin conformation, the fusion peptide is propelled towards the target membrane where it penetrates, anchoring the viral membrane to the target cell membrane. Then the HR1 domain rearranges itself with the trimeric coiled-coil HR2 domain to form a six-helix bundle. In the resulting conformation, the transmembrane domain and the fusion peptide anchored into the target membrane are brought in close proximity until the viral and cell membranes merge together.[306, 334, 335] This route might not be the only route of entry of the virus and Tang *et al.* have described a late 'endosomal route' where the virus enters the cell by endocytosis, after ACE2 binding.[335]

### 5.1.8 Sugar composition of the spike protein

Elucidating the glycosylation of the spike protein and its variations across cell lines is essential for the development of binders and vaccines. Sequence analysis of the SARS-CoV-2 spike protein shows 22 possible N-linked glycosylation sites and four possible O-linked glycosylation sites. But the detection of these sites does not mean that there will be a glycan attached to each site. Usually, the sugar composition is determined by MS analysis, after digestion of the protein of interest into peptide fragments, where the peptides segments produced still carry the glycans. Studies on the glycosylation of the SARS-CoV-2 protein have yielded different results for the occupancy and composition of glycans. It is known that various external parameters impact glycomic analyses, resulting in differences in sugar composition. Determining with certainty the glycosylation of a protein is challenging as glycan synthesis is highly variable and dependent on the type of cell used to express the proteins as well as cell culture conditions. Most of the studies on SARS-CoV-2 were performed with recombinant proteins, and different cell lines can result in major changes in sugar composition. The media and other conditions used to grow the cells expressing the protein can also affect the results, resulting in diverging sugar compositions in the published studies.[336] It is also highly suspected that the sugar composition varies between individual human hosts.

Walls *et al.* were the first to resolve 16 of the glycosylation sites using cryo-EM.[316] Watanabe *et al.* proved that all 22 sites were glycosylated, using a site-specific MS approach.[337] They also analysed the glycosylation pattern of these 22 sites. They found that two sites (Asn234 and Asn709) are principally oligomannose-type. A mixture of more processed oligomannose and complex-type glycans was reported for the Asn61, Asn122, Asn603, Asn717, Asn801, and Asn1074 sites. The remaining glycosylation sites carry even more complex glycans. Zhao *et al.* also reported similar results.[338] Lenza *et al.* characterised the glycan structures of the N-linked glycans in the RBD using NMR spectroscopy, which in contrast to MS, avoids sample digestion.[339] In their work, novel glycan motifs were identified and the importance of N-acetylgalactosylation and hyper-fucosylation at the terminal chains of the RBD N-glycans was uncovered. However, in contrast to these compelling results for N-glycosylation, varying compositions for the O-linked glycosylation sites have been reported. Several groups have reported different O-linked glycosylation patterns for the spike protein. Shajahan *et al.* reported a high level of O-glycosylation of S1 and S2 when expressed independently and detected O-glycosylation at sites Thr323 and Ser325 on the S1 subunit of the spike protein.[340] However, two additional publications reported a low occupancy at most O-glycosylation sites using the spike trimer in the analysis,[337, 338] while Sanda *et al.* reported an O-linked glycosylation site close to the furin cleavage site. In addition, they identified another eight O-glycopeptides.[341] More recent work by Yao *et al.* characterised the spike protein sugar composition using the native virus, which was isolated from a patient.[342] The authors found that the virus exhibited lower levels of oligomannose-type

glycosylation compared with the other studies.

### 5.1.9 Roles of glycosylation

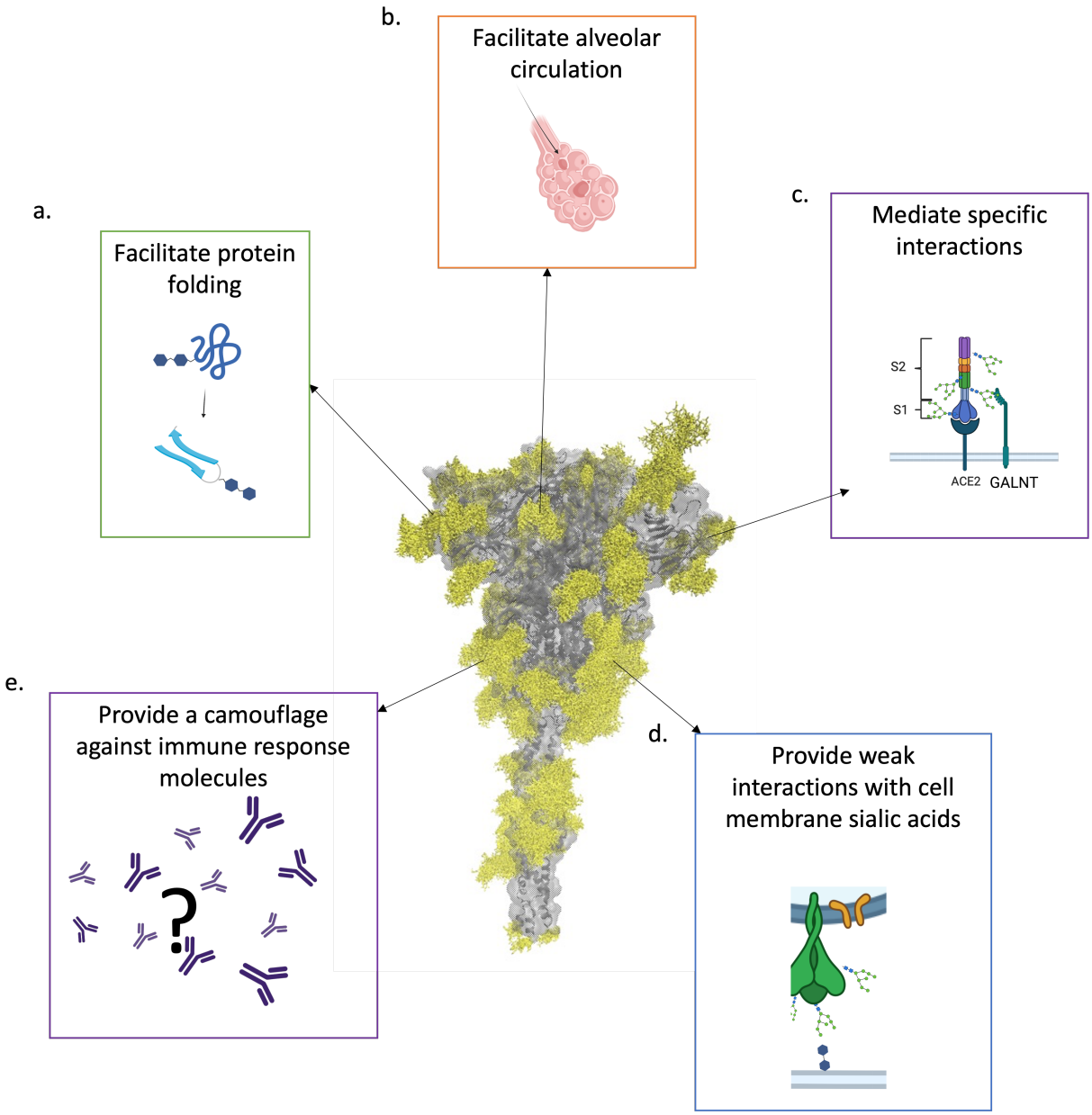



**Figure 5.7:** The different roles of the spike protein glycosylation

A key structural feature of the spike protein is its extensive glycosylation. This coverage is though to have multiple functions for the virus. N-glycans expressed on the viral envelope glycoproteins fulfil diverse biological roles, as seen in Figure 5.7 One primary function of this glycosylation is to ensure the protein correct folding and priming by host proteases.[316] (Figure 5.7.a). Another more contested role is to facilitate the navigation of the virus in the mucus of the respiratory and gastrointestinal tracts (5.7.b). This function is more hypothetical: the mucus is known to offer a protective barrier against MERS-CoV infection, but these results

have not been reproduced with SARS-CoV-2 virions.[343] Cho *et al.* suggested that although glycosylation patterns in the SARS-CoV-2 S1 glycoprotein are more complex than those in SARS-CoV and MERS-CoV, the sugar coverage of the virion proteins is similar and cannot directly the observed difference in contamination rates.[344]

A more widely corroborated function of spike protein glycosylation is the initial viral docking: the initial approach could be mediated by the interaction of specific N-glycan epitopes with sialylated glycans present on the cell membrane.[124, 345–347] The presence of only these sialic acids would not be enough for cell fusion, as protein receptors are necessary for virus entry, but the presence of sialic acids would facilitate the fusion by providing a weak adherence for the virus to the cell surface (Figure 5.7.d). This hypothesis has been proved by experimental assay: on cultivated human cells where sialic acids are absent, the virus entry is 60-fold lower than on cells displaying sialic acids. Variability of expression of sialic acids exists between individuals and could be one of the explanations for the difference in response to infection observed between patients.[348]

As reported in the previous section, the SARS-CoV-2 spike protein contains a unique furin cleavage site regulating the activation of the S protein (T678). A study shows that the O-glycosylation site near the furin cleavage site increases viral infectivity and syncytia formation by decreasing furin cleavage.[341, 349] This glycosylation is mediated by specific members of the GALNT enzyme, a human glycosyltransferase (Figure 5.7.c).

The most widely described role for glycosylation is to provide a highly dense coating of the protein surface (Figure 5.7.e). Complex carbohydrates on viral proteins constitute a perfect camouflage to evade the immune system as antibodies bind protein surface. The sugar forms a barrier, limiting access to the protein surfaces. However, SARS-CoV viruses are not the most densely covered virus. HIV-1 represents an extreme case of sugar shielding where the weight of the sugars accounts for 50% of the mass of the HIV spike protein mass. In contrast, $\beta$-coronaviruses are not shielded as effectively against the immune system, but this shielding still restrains the accessibility of the protein surface and thus the number of possible protein interaction sites with antibodies.[336]

## 5.2 Prospects for using peptides to detect SARS-CoV-2

SARS-CoV-2 affects different tissues, but the most life-threatening conditions are in patients whose lung tissues are infected [350]. Antibodies, small molecules and peptides possess different sizes and physical properties, leading to different possible administration routes and applications. The high selectivity of peptides combined with their safety and tolerability enables their translation into clinical applications. Importantly, degradation pathways are often predictable. The ease of peptide synthesis means that they can be rapidly made using varied sequences. Antiviral peptides have been developed for diverse viruses. An example of a successful antiviral peptide in a clinical application is the anti-HIV peptide T20 (Enfuvirtide),

proving the efficacy of peptides for the treatment of infectious diseases.[351]

### 5.2.1 Efficacy of current diagnosis tools

In their review published in May 2020, Jarom *et al.* analysed and compared the results of 25 different studies on SARS-CoV-2 detection methods.[352] For the RT-PCR tests the authors disclose a sensitivity of 87.8% and a specificity ranging from 92.9 to 100 % among studies. The authors also warned about the lack of consistent references among the studies, which could offset these values. We found no evidence of previous work on the development of optical probes for the detection of COVID-19. Optical probes are interesting as they have multiple uses, such as histological staining, surface detection, or sample detection. Their application might facilitate the setting of such detection methods in larger scales, making them cheaper, and applicable in more settings than other detection methods.

### 5.2.2 Peptides for virus detection on surfaces

Current methods for COVID-19 detection combines an enzyme-linked immunosorbent assay (ELISA) or a reverse-transcription polymerase chain reaction (RT-PCR) with immunofluorescence. RT-PCR is currently the gold standard for SARS-CoV-2 detection, but this method is a multi-step technique that involves purification, nucleic acid amplification, and fluorescence detection. Imaging methods for the detection of virusesinclude computed tomography(CT), single-photon emission computed tomography (SPECT), and positron emission tomography (PET). These methods have low resolution and are costly. Although recently CT has been used for virus diagnosis, it usually can only detect signs of virus infection in tissues. (i.e., lung lesions, pneumonia). Optical biosensors could present an appealing alternative method for virus detection on surfaces.

### 5.2.3 Peptides for diagnosis

Linear and cyclic peptides have been used for the *in vivo* detection of numerous conditions, including cancer,[353, 354] arthritis[355] and viral infections.[356, 357]. One limitation is that a strong affinity is usually required and small ligands with low affinity are not suitable for this approach, even with high specificity.[358] The use of optical biosensors to be used as point-of-care (POC) diagnostic tools, offer the advantage of low costs test manufacturing, and no expensive analysis equipment.

A few studies have proposed peptides as diagnostic agents. For example, antimicrobial peptides have been used to detect Shiga toxin-producing *Escherichia coli* bacteria; and troponin I-specific peptides have been developed to detect heart disease by phage-display, and enzyme-linked immunosorbent assay (ELISA) [357] Colour wide-field fluorescence endoscopy is an emerging technology to detect and appreciate the spread of lung diseases.[359]

Fluorescent peptides could allow the staining of the affected lung tissues after infection by the virus. In April 2020, the University of Edinburgh launched its own research plateform, the QMRI COVID-related Research Hub (QCH) has been established as a rapid response to the current global COVID-19 pandemic,. The QMRI integrates the Edinburgh Royal Infirmary, the University of Edinburgh's College of Medicine and Veterinary Medicine, and the Centre for Inflammation Research (CIR) with NHS Lothian. Discussions with QCH leaders identified fluorescent peptides as a development opportunity and this project was done in relation with the QMRI.

### 5.2.4 Peptides reported to bind the spike protein

| Targeted domain | Name | Sequence derived from | Helicity | IC50 ( micro M ) | Kd / nM | Sequence | Ref | |
|---|---|---|---|---|---|---|---|---|
| HR1 | EK1 | HR2 (HCoV-OC43) | | 15.8 | | SLDQINVTFLDLEYEMKKLEEAIKKLEESYIDLKEL | Xia et al. 2020 | [360] |
| HR1 | EK1-C4 | HR2 (HCoV-OC43) | | 1.3 | | Ac-SLDQINVTFLDLEYEMKKLEEAIKKLEESYIDLKELGSGSG-amino-PEG3-acetyl-Cys(chol-oxycarbonylmethyl | Xia et al. 2020 | [360] |
| HR1 | 2019-nCoV-HR2P | HR2 (1150-1185) (SARS-CoV-2) | | 0.19 | | DISGINASVVNIQKEIDRLNEVAKNLNESLIDLQEL | Xia et al. 2020 | [360] |
| HR1 | SARSHRC-PEG4-chol | HR2 (1150-1185) (SARS-CoV-2) | | 0.01 | | DISGINASWNIQKEIDRLNEVAKNLNESLIDLQEL -PEG4-chol | de Vries et al. 2020 | [361] |
| HR1 | [SARSHRC-PEG4]2-chol | HR2 (1150-1185) (SARS-CoV-2) | | 0.003 | | [DISGINASWNIQKEIDRLNEVAKNLNESLIDLQEL -PEG4]2-chol | de Vries et al. 2020 | [361] |
| HR1 | IPB01 | HR2 (1162-1205) (SARS-CoV-2) | 59 | 1.022 | | ISGINASVVNIQKEIDRLNEVAKNLNESLIDLQEL | Zhu et al. 2020 | [362] |
| HR1 | IPB02 | HR2 (1162-1205) (SARS-CoV-2) | 60 | 0.08 | | ISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELK(Chol) | Zhu et al. 2020 | [362] |
| HR1 | IPB03 | HR2 (1162-1205) (SARS-CoV-2) | 50 | 0.015 | | INASVVNIQKEIDRLNEVAKNLNESLIDLQELK(Chol) | Zhu et al. 2020 | [362] |
| HR1 | IPB04 | HR2 (1162-1205) (SARS-CoV-2) | 60 | 0.033 | | SVVNIQKEIDRLNEVAKNLNESLIDLQELK(Chol) | Zhu et al. 2020 | [362] |
| HR1 | IPB07 | HR2 (1162-1205) (SARS-CoV-2) | 55 | 0.017 | | IDRLNEVAKNLNESLIDLQELKYEQYIK(Chol) | Zhu et al. 2020 | [362] |
| HR1 | SARS-CoV-2 HRC lipopeptide | HR2 (1168 to 1203) SARS-CoV-2 | | 0.01 | | DISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGSGC(Chol) | Outlaw et al. 2020 | [363] |
| RBD | SBP1 | ACE2-(Helix 1)-(24-52) | | | | IEEQAKTFLDKFNHEAEDLFYQS | Zhang et Aal 2020 | [303] |
| RBD | NYBSP-1 | ACE2-(Helix 1)-(24-52) | 94 | 4.1 | | AcTIEEQAKTXLDKXNHEAEDLFYQXSLAXWN-NH2 staples in i/i+4 i/i+4 | Curreli et al. 2020 | [364] |
| RBD | NYBSP-2 | ACE2-(Helix 1)-(24-52) | 61 | 2.9 | | AcTIEEQZKTFLDKXNHEAEDLZYQSSLAXWN-NH2 staples in i/i+4 i/i+4 | Curreli et al. 2020 | [364] |
| RBD | NYBSP-3 | ACE2-(Helix 1)-(24-52) | 50 | 12.9 | | AcTIEEQAKTXLDKXNHEAEDLZYQSSLAXWN-NH2 staples in i/i+4 i/i+4 | Curreli et al. 2020 | [364] |
| RBD | NYBSP-4 | ACE2-(Helix 1)-(24-52) | 80 | 1.97 | | AcTIEEQZKTFLDKXNHEAEDLZYQXSLAXWN-NH2 staples in i/i+4 i/i+4 | Curreli et al. 2020 | [364] |
| RBD | NYBSC-5 | ACE2-(Helix 1)-(24-52) | 19 | >30 | | AcTIEEQAKTFLDKFNHEAEDLFYQSSLAXWN-NH2 | Curreli et al. 2020 | [364] |
| RBD | P7 | ACE2-(Helix 1)-(24-52) | 63 | | 7/6.8 | SALEEQYKTFLDKFLHELEDLLYQLALAL-NH2 | Karoyan et Al. 2020 | [365] |
| RBD | P8 | ACE2-(Helix 1)-(24-52) | 70 | 15 | 0.8/0.09 | SALEEQLKTFLDKFMHELEDLLYQLAL-NH2 | Karoyan et Al. 2020 | [365] |
| RBD | P9 | ACE2-(Helix 1)-(24-52) | 53 | < 1 | 0.3/0.07 | SALEEQYKTFLDKFMHELEDLLYQLSL-NH2 | Karoyan et Al. 2020 | [365] |
| RBD | P10 | ACE2-(Helix 1)-(24-52) | 56 | < 1 | 0.06/0.09 | SALEEQYKTFLDKFMHELEDLLYQLAL-NH2 | Karoyan et Al. 2020 | [365] |
| RBD | hACE2(21-55)A36K-F40E | ACE2(21-55). (Helix 1) | 52 | 2.1 | 3.6 | IEEQAKTFLDKFNHEKEDLEYQSSLASWNYNTNIT | Maas et al. 2021 | [315] |
| RBD | hACE2(21-55)F32K-A36E | ACE2(21-55) | < 38 | 28.4 | | IEEQAKTFLDKFNHEEEDLFYQSSLASWNYNTNIT | Maas et al. 2021 | [315] |
| RBD | hACE2(21-55)F28K-F32E | < 38 | 46.8 | | | IEEQAKTKLDKENHEAEDLFYQSSLASWNYNTNIT | Maas et al. 2021 | [315] |
| RBD | AHB1 | de-novo design (mini protein) | 60 | 0.035 | | DEDLEELERLYRKAEEVAKEAKDASRRGDDERAKEQMERAMRLFDQVFELAQELQ EKQTDGNRQKATHLDKAVKEAADELYQRVRELEEQVMHVLDQVSELAHELLHKLTG EELERAAYFNWWATEMMLELIKSDDEREIREIEEEARRILEHLEELARK | Cao et al. 2020 | [366] |
| RBD | AHB2 | de-novo design (mini protein) | 60 | 0.0155 | | ELEEQVMHVLDQVSELAHELLHKLTGEELERAAYFNWWATEMMLELIKSDDEREIR EIEEEARRILEHLEELARK | Cao et al. 2020 | [366] |
| RBD | LCB1 | de-novo design (mini protein) | 60 | 23.54 $10^{-6}$ | | DKEWILQKIYEIMRLLDELGHAEASMRVSDLIYEFMKKGDERLLEEAERLLEEVER | Cao et al. 2020 | [366] |
| RBD | LCB3 | de-novo design (mini protein) | 60 | 48.1 $10^{-6}$ | | NDDELHMLMTDLVYEALHFAKDEEIKKRVFQLFELADK | Cao et al. 2020 | [366] |
| RBD | 1 | nRNA display | | | 550 | (YNVYELRQSVWVHIVRC)-NH2 | Norman et al. 2021 | [367] |
| RBD | 2 | nRNA display | | | 120 | (YSLIYWYGQLRHVSRGC)-NH2 | Norman et al. 2021 | [367] |
| RBD | 3 | nRNA display | | | 93 | (YSIVYIKGEVRFVGRGC)-NH2 | Norman et al. 2021 | [367] |
| RBD | 4 | nRNA display | | | 15 | (yKAGVVYGYNAWIRC)-NH2 | Norman et al. 2021 | [367] |
| RBD | 5 | nRNA display | | | 76 | (yRIVILSGSRVC)VCC-NH2 | Norman et al. 2021 | [367] |
| RBD | 6 | nRNA display | | | 130 | (yYFDVLLWKAISC)-NH2 | Norman et al. 2021 | [367] |
| RBD | SAP-1 | ACE2-(Helix 1)-(27-42) | | 2.39 | 0.53 | TFLDKFNHEAEDLFYQ | Larue et al. 2020 | [368] |
| RBD | SAP-2 | ACE2-(Helix 1)-(37-45) | | 3.72 | 10.7 | EDLFYQSSL | Larue et al. 2021 | [368] |
| RBD | SAP-3 | ACE2-(Helix 3)-(79-85) | | >7.5 | | LAQMYPL | Larue et al. 2022 | [368] |
| RBD | SAP-4 | ACE2-(Helix 11)-(352-359) | | >7.5 | | GKGDFRIL | Larue et al. 2023 | [368] |
| RBD | SAP-5 | ACE2-(Helix 1)-(24-36) | | >7.5 | | QAKTFLDKFNHEA | Larue et al. 2024 | [368] |
| RBD | SAP-6 | ACE2-(Helix 1)-(37-42) | | 1.9 | 1.36 | EDLFYQ | Larue et al. 2025 | [368] |
| $\alpha 5\beta 1$ integrin | ATN-161 | fibronectin | | | | Ac-PHSCN-NH2 | Beddingfield et al. 2020 | [369] |

**Table 5.1:** Sequences and affinity of peptides reported to bind the SARS-CoV-2 spike protein.

Several groups have been working on developing peptides to mimic the ACE2 binding mode, using the $\alpha 1$ helix sequence. Among the different studies, Curreli *et al.* designed stapled peptides based on ACE2. [364] They used a lactam staple to constrain the peptide and managed to produce peptides with an enhanced helicity content (from 50 to 95 % helical content) compared to the unstapled variant (19% helical content) and a binding affinity in the micromolar range.

It is hypothesised that SARS-CoV-2 spike protein binds to both hACE2 and $\alpha 5\beta 1$. An approach followed by Beddingfield *et al.* was thus to disrupt this interaction using a peptide, and one of their designs, ATN-161, was shown to disrupt SARS-CoV-2 infection in vitro.[369] Treatment using ATN-161 increased cell viability in the presence of SARS-CoV-2 and decreased cytopathic effects associated with a viral infection. However, the binding site could not be identified, and three potential binding sites could be involved. ATN-161 was found to potentially bind the integrin $\alpha 5\beta 1$ ectodomain near the RGD motif, blocking integrin binding to the spike protein. Another plausible site was the interface between the ACE2 and the spike RBD, potentially affecting the binding of RBD with ACE2.

Another class of inhibitory peptides are the spike protein fusion blockers. Several inhibitory peptides binding to the HR domain of different Coronaviridae viruses, have been identified, showing activities in the nanomolar to low micromolar range: HCoV-229E,[370] SARS-CoV,[371–373] and MERS-CoV [374, 375]. The S2 subunits of SARS-CoV and SARS-CoV-2 are highly conserved; they possess 92.6% overall identity for the HR1 region and 100% for the HR2 region. Thus, peptides derived from SARS-CoV HR2 might also inhibit SARS-CoV-2. The fusion blocker peptides usually bind to the HR1 domain to inhibit the fusion process. Such examples include EK1, 2019-nCoV-HR2P and IPB-01, which inhibit SARS-CoV-2 with $IC_{50}$ values of 2.5 $\mu$M, 0.98 $\mu$M and 22 $\mu$M, respectively. Notably EK1, which targets the HR1 domain of several other human coronaviruses, including SARS-CoV and MERS-CoV displays potent prophylactic and therapeutic activity against MERS-CoV and HCoV-OC43 infections *in vivo*, and intranasal application in mice shows promising results.[376] Furthermore, EK1 shows only low immunogenicity and toxicity *in vivo* and was the first broadly active HR-targeting inhibitor against coronaviruses *in vivo*.[360] Derivatives of EK1 have exhibited promising binding activity as reported by the work of Xia *et al.*[360] EK1C4 is 19 to 190-fold more potent than EK1 against several coronaviruses and inhibits infection of TMRPSS2 negative Vero E6 cells by SARS-CoV-2 with an $IC_{50}$ of 37 nM.[360] EK1 and EK1C4 were both tested against HCoV-OC43 *in vivo* and show prophylactic activity as well as therapeutic efficacy when applied intranasally to mice, with better stability, antiviral activity and half-life in the case of EK1C4. Zhu *et al.*[362] modified the peptide IPB-01 by adding a cholesterol group to its C-terminus (IPB-02), which reduces $IC_{50}$ values against SARS-CoV-2 and SARS-CoV pseudoparticles to 0.08 $\mu$M and 0.25 $\mu$M respectively. De Vries *et al.* [361] reported the optimisation of 2019-nCoV-HRP2, varying only by a single amino acid from the initial sequence, and used PEGylation, lipidation and dimerisation, to en-

hance the drug-like properties of the peptide. The authors confirmed the observation that the fusion inhibiting peptide $[SARSHRC\text{-}PEG4]_2$-chol was highly active on TMPRSS2 positive cells with an $IC_{50}$ as low as 3.8 nM. Most notably, intranasal application in ferrets protected them from viral transmission when co-housed with other SARS-CoV-2 infected ferrets. Altogether, these studies show that targeting the HR1 domain by peptides derived from the HR2 region is a promising strategy for generating fusion inhibitors against many class I fusion protein viruses, including SARS-CoV-2. The high HR homology in coronaviruses allowed the production of the broadly active and potent fusion inhibitor EK1C4 for the potential treatment of infection by different coronaviruses including SARS-CoV-2.

Other targets relevant to the fusion peptide have also been highlighted. Membrane interactions, especially those involving the fusion peptide (FP), could be important for viral fusion and thus represent potential antiviral targets. Five peptides analogous to the HR regions were tested for inhibition of SARS-CoV entry. The most potent inhibition of SARS-CoV-1 plaque formation was found for WW-III and WW-IV with $IC_{50}$ values between 2 and 4 $\mu$M. Both peptides are derived from the loop domain between HR1 and HR2, and the authors speculate that they might sterically hinder the extension of the FP or block the fusion process.[377] However, at the time of writing, no peptide interfering with the FP region in the cell membrane have yet been reported as active for SARS-CoV-2.

Numerous other computational approaches to design have been reported for producing better spike protein binders but these are often not backed up with experimental data. As an example, Huang *et al.*[378] designed peptides to block association of the SARS-CoV-2 spike protein with human ACE2, using their in-house EvoEF2 scoring. Roughly, their method relies on specific mutations of the ACE2 sequence to measure changes in the energy of binding. Another example is the work of Han *et al.* [379] who form mini proteins between the $\alpha$ helical scaffolds of ACE2 at the PPI. Their study suggests a much bigger segment than this more structured version of the ACE2-helix1 (residues 30-45) peptide would be necessary to bind to the spike protein.

These approaches targeting the SARS-CoV-2 protein contrast that of Li *et al.*[380], who have also developed fluorescent peptides for the diagnosis of COVID-19 in blood serum. Their method is based on detecting Monoclonal Antibodies (MAbs) generated by the immune response against SARS-CoV-2 by producing peptides derived from SARS-CoV-2 protein fragments, which bind specifically and selectively to human SARS-CoV-2 mAbs. They identified eight peptides derived from the spike protein that showed potential for development of a diagnostic tool. However, this detection approach requires that an effective immune response in the patient has been triggered after infection.

## 5.3 Study of full length spike protein simulations

### 5.3.1 Model building

At the beginning of our study, no MD simulations of the glycosylated spike protein had been made available. The DE Shaw Research group was the first to publish open-access simulations of the spike protein,[381] however, these trajectories did not include the HR2 domain. We also noted a few issues affecting the CoV-2 spike protein in these MD trajectories, probably deriving from the initial structures used. For example, in DESRES-ANTON-11021566, the starting model for the "closed" trajectory, the presence of a few *cis* peptide bonds in each chain of the protein was observed (His245-Arg246, Arg246-Ser247 and Ser640-Asn641, Gln853-Lys854). The His245-Arg246 peptide bond in chain B flipped to a *trans* conformation in the course of the simulation, but the other bonds remained in their *cis* conformations. Any atomistic simulation starting structure should be a low-energy, representative conformation of the protein studied. However, structural data deposited in the PDB data bank are usually imperfect, as some uncertainty in local coordinates, often in flexible regions, is expected. Local conformation errors like these may lead to subtle biases throughout the simulation trajectory, adding more problematic conformational misrepresentations and biasing the simulation in unpredictable ways. The *cis* peptide bond can, for example, prevent the formation of any stable secondary structure by introducing a kink in a loop. Regions modelled as coils are generally more likely to be further away from a low energy structure, primarily if this domain was reported to be structured in other models. Comparison of the structure with its electron density map can also pinpoint areas where the resolution of the structure is low.

In this work, both the opened and closed conformations were modelled. The models were built using two structures (opened and closed) of the full length spike protein provided by the Amaro group [382]. When building the opened conformation model, it was noticed that a loop in the RBD interacting region (residues 673 to 686) was modelled as a coil, and twelve residues of this loop were missing from the original crystal structure used to produce this model (PDB 6VSB). Only four could be inserted in the space left in the cryo-EM structure. Furthermore, the secondary structure of other RBD domains in the closed conformation and the RBD bound to ACE2 (6M0J) presented a $\beta$-sheet secondary structure for this loop. The conformation of this region is thus likely to be erroneous in the provided model. The structure of the spike protein RBD domain of PDB 6M0J, refined by Croll et. al matched the electron density of 6VSB,[383] and grafting this refined 6M0J structure into the structure provided by the Amaro group seemed to produce a more relevant model. No other modification on the protein was made for the opened model, and the closed model was used directly. The sugar composition was deduced by combining the results of two studies on the glycosylation patterns of the spike protein.[382, 384] and can be found in Appendix 5.12. The sugars were built *de novo* using the doglycan tool.[385] Scripts and input files to build the sugars have been

made available on github (michellab/Stapline) The spike protein was simulated with the S1/S2 cleavage site open, as the furin cleavage site (residues 682-685, RRAR) has been shown to be essential for infection of human lung cells [310].

As the simulations aimed to understand the target-able domain of the spike protein, the lipid membrane was not modelled; the TM domain was thus discarded, as solvating this highly hydrophobic domain could result in simulation artefacts. After identifying some unfolding events at the truncated C-terminus, in some preliminary simulations, a 10-residue segment of the transmembrane domain was included in the simulations. The C$\alpha$ atoms of this segment were restrained in an attempt to mimic the restraints resulting from a membrane an the spike protein (Figure 5.8). Restraining forces (5000 $kcal.nm^{-1}$) were applied on the X, Y and Z-axis, to mimic the lipid forces in this region.

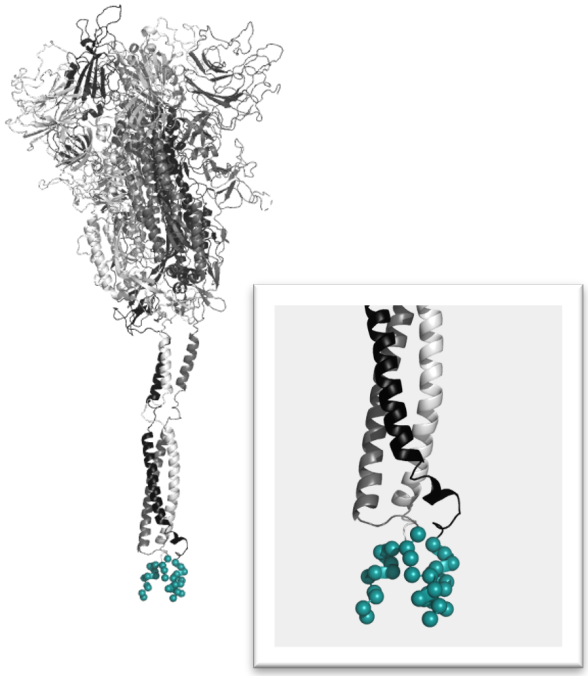


**Figure 5.8:** Spike protein structure used for the simulations: the different chains are highlighted in grey scale and the atoms restrained during the simulation are represented with blue spheres.

### 5.3.2 Backbone flexibility

To visualise flexible regions in the SARS-CoV-2 spike protein, the root-mean-square fluctuations (RMSF) of the protein backbone were calculated. These RMSF studies can give a good idea of the stability of the targeted domains (Figure 5.9). Unsurprisingly, the most flexible regions of the protein were found to be in in loop regions. The NTD domain contains the most flexible segment of the spike protein, in its solvent-exposed loops. The 665-666 cleavage site is also a very flexible area. We noted the presence of a few domains where the RMSF

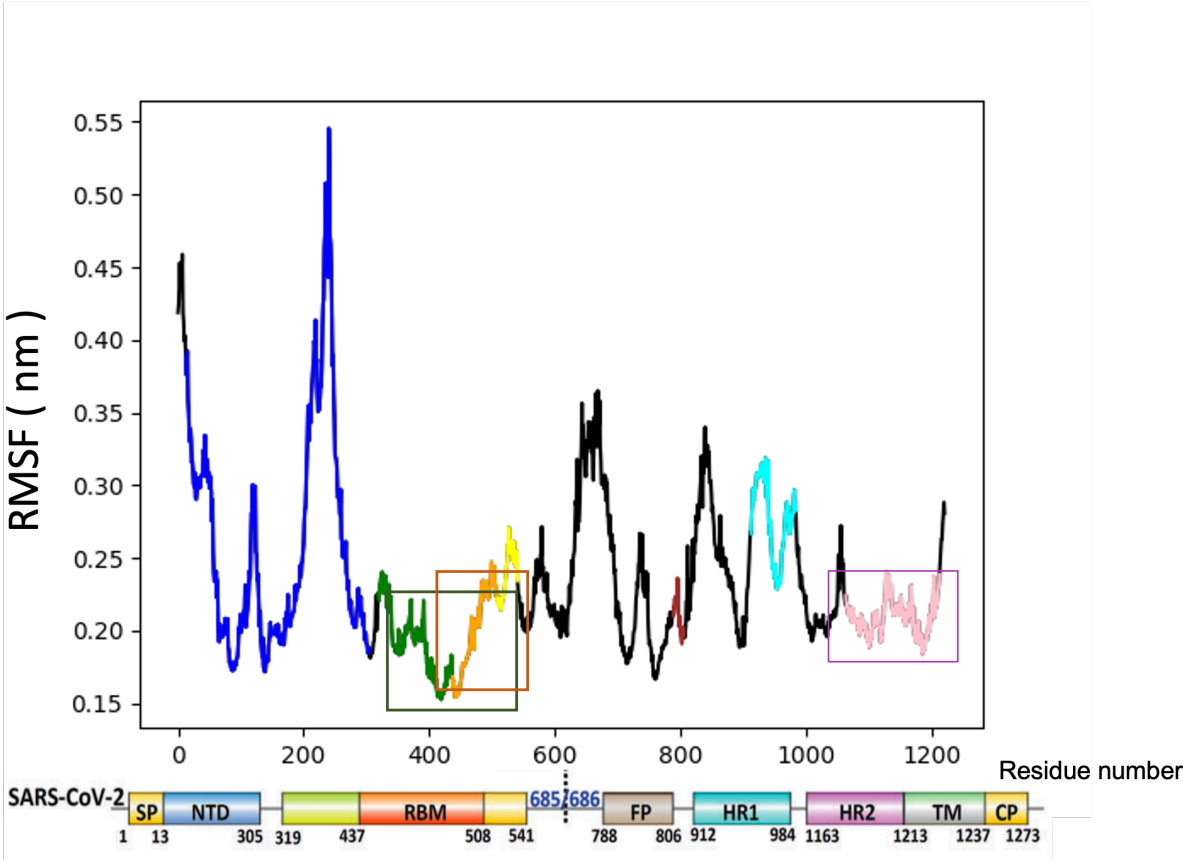



**Figure 5.9:** Study of the RMSF of the spike protein to find area with low fluctuation during a 200 ns simulation. RMSF of the three chains of the spike protein. The curve is coloured by the spike protein domains.

is lower and would favour the binding of ligands (Figure 5.9). Overall the RBD domain is the region with the lowest RMSF, whilst short sections of the NTD region, as well as the HR2 domain, have a lower RMSF. In the simulation, structured domains in both the NTD regions were stable, the peaks of RMSF observed in the NTD region (around residues 1-30 and 180-300, Figure 5.9) correspond to unstructured regions of this domain. The RMSF values in the ACE2 binding site (residues 437-508) are among the lowest values of the spike protein backbone, which validate the fact, that generally the binding sites are in the most rigid part of the protein. Notably the plausible binding site of integrin (residues 400-410) was also found to display low RMSF values.

During the simulation, the 'tail' region (residues 1010-1220) of the spike protein was found to be very flexible, whilst the head region tilted up to 30°. Figure 5.10 shows the distribution of the angle adopted by the head spike with the tail in the simulation. Mostly this angle was around 13° but went up to 20° during the simulation. Sikora *et al.* independently revealed significant head tilting associated with flexing of the HR2 domain in the tail of the spike protein in simulation, which is thus consistent with our observations.[386] These observations are also in accordance with Cryo-EM observations which revealed that the spike protein structure could tilt by up to 60°.[322] However the lower values found in our simulations may reflect a more

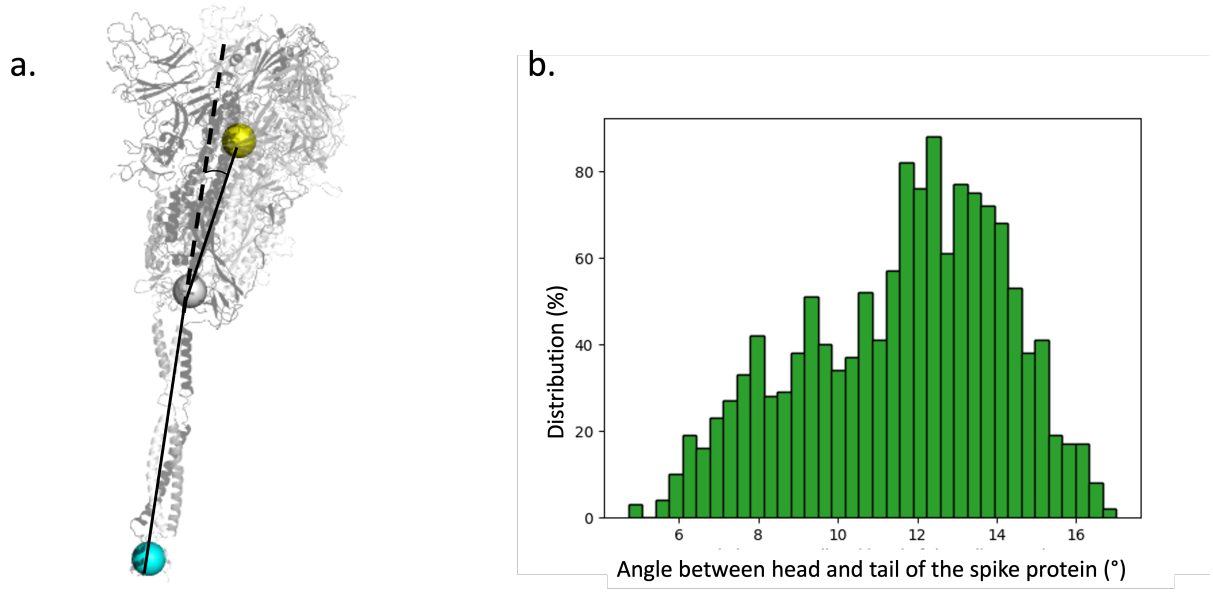



**Figure 5.10:** Distribution of the angle between the head of the spike protein and the tail a. Measured angle b. Obtained distribution in a 200 ns simulation

limited sampling, as the simulations were only conducted for 200 ns.

### 5.3.3 Sugar coverage of the spike protein and peptide design

Study of the coverage of the sugars was conducted to identify areas which might be targeted on the spike protein due to the absence of the protein 'sugar coat'. As anticipated, the sugar coverage of the spike protein varies by domain (Figure 5.11). The S1 domain is overall more densely coated by sugars in comparison to the S2 domain. However in the open conformation, the S1 domain is more exposed, especially around the RBM motif when compared to the closed conformation. This sugar analysis is particularly useful in determining new possible binding sites. In general, the S2 domain appears to be a more target-able surface, offering less sugar coverage. The HR2 domain, in particular, offers 25 % exposed surface, in comparison to the 8-15 % free surface of the S1 domain. Moreover, a small area on the HR2 domain, between residues 1176 and 1196 has been found to be mostly uncovered while the rest of the stalk was found highly glycosylated. Work published at the time of our study is in agreement with these observations.[382, 384]

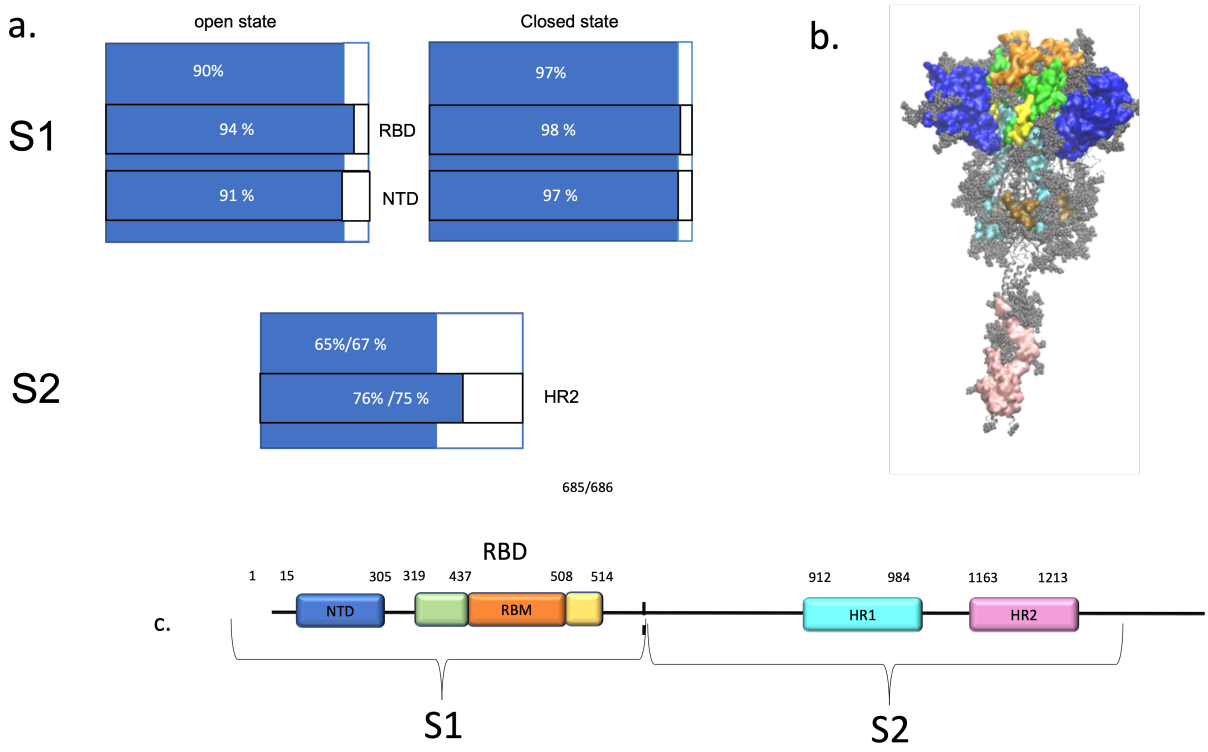



**Figure 5.11:** a. Sugar coverage of the spike protein per domain (using a 1.4 Å radii probe), b. Glycosylated spike protein coloured by domains, the sugars are represented in grey c. spike protein sequence with the domains coured to match those in part b.

## 5.4 Design of peptides mimicking the ACE2 binding

### 5.4.1 Objectives

The analysis of the ACE2 PPI (Section 5.1.4) suggests that the ACE2-spike protein interaction is mediated primarily by interactions between the $\alpha$1-helix of ACE2 (residues 20-45) and the globular RBM domain of the spike protein. This observation provides an initial helical scaffold for the design of peptides. We hypothesised that with some modifications, including mutations to enhance the binding affinity and the addition of staples to maintain helicity, the $\alpha$1-helix would inhibit the ACE2 binding to the spike protein. Usually, shorter peptides do not remain helical; however, shorter stapled peptides taken from longer helical motifs have been shown to conserve their secondary structure.[60] We used different sections of the $\alpha$1-helix of the ACE2 protein, to produce peptide scaffolds with different lengths; smaller peptide scaffolds comprise 15 amino acids of the $\alpha$1-helix, while the longer peptides scaffold comprises the full $\alpha$1-helix (33 a.a.).

Whilst numerous groups are currently working on producing enhanced binders for the spike protein using the $\alpha$1-helix of ACE2; a novel approach to induce mutations in the sequence was explored in this work. When triggered, the immune system produces antibodies, which specifically recognise a pathogen, generally a protein. In the case of SARS-CoV-2, spe-

cific antibodies to the spike protein have been isolated from recovering patients. Nanobodies and antibodies have also been identified by the immunisation of animals. Engineered antibody mimetics and antigen-binding fragments (Fab) have also been reported. Thus at the time of our study, multiple antibodies, antibody mimetics, Fab and nanobody structures were available in the PDB. The umbrella term Antibody-like (Ab-like) will be used to refer to this range of antigen binder molecules in this Chapter. We chose to use the information arising from the binding of Ab-like molecules to the ACE2-RBD to propose mutations to the $\alpha$1-helix peptide scaffold. The peptide side-chains were modified using the side-chains of the Ab-like molecules interacting with the same spike RBD amino acids to build more potent $\alpha$-helical peptides.

### 5.4.2 Preliminary results

In 2020, numerous groups worked on developing peptides to inhibit the binding of the spike protein human cell receptors, using the ACE2 $\alpha$1-helix segment (Table 5.1). We based our approach on one of the first papers published at the time. Zhang *et al* excised the $\alpha$1-helix of ACE2 from residue 21 to residue 36 to test if the resulting peptide would bind the spike protein.[303] No mutations were introduced to the peptide sequence. They initially reported mM binding; however, they subsequently revised their paper, and the ACE2 derived peptide was later shown not to bind the spike protein.[302, 303] Based on the version of this paper, we initially thought that this sequence was binding to the spike protein and could be improved through the addition of staple residues and targeted mutations. With insight, the latter correction to reporting only a weak binding could have been anticipated from our first 200 ns simulation of this segment. Starting from a model of the peptide bound to the spike receptor (using PDB 6M0J) and truncating the ACE2 protein, the simulation rapidly showed instabilities in the binding of the peptide, and dissociation was observed in the peptide termini (mainly the C-terminus) after only 50 ns. The peptide started to rotate on itself after 120 ns, and residues facing the solvent in the initial conformation interacted with the RBD of the spike protein after only 80 ns of the simulation. Most significantly after 160 ns the peptide did not remain bound to the spike protein during our simulations. We hypothesised that this was partially triggered by the hydrophobic nature of the residues facing the solvent, as these amino acids started to interact with the RBD. These hydrophobic residues at the solvent surface of the peptide are buried inside the ACE2 protein in the ACE2/RBD complex (Phe28, Phe32, Phe40, Leu29). These residues were thus mutated to more polar serine residues in a second simulation, which appeared to improve the stability of the peptide in our simulation time scale (200 ns) as no rotation of the peptide was observed. However, it seemed clear that peptides needed some improvement in their sequences, as both head and tail were found to dissociate from the receptor. As seen in Figure 5.12, the extremities of the peptide are subject to unbinding with higher RMSD observed in these regions. Part of the peptide

at the C-terminus had no adherence to the spike receptor (Figure 5.12) Retrospectively, we can conclude that this observed dissociation was in complete accordance with the low affinity observed in experiments.[302, 303] .

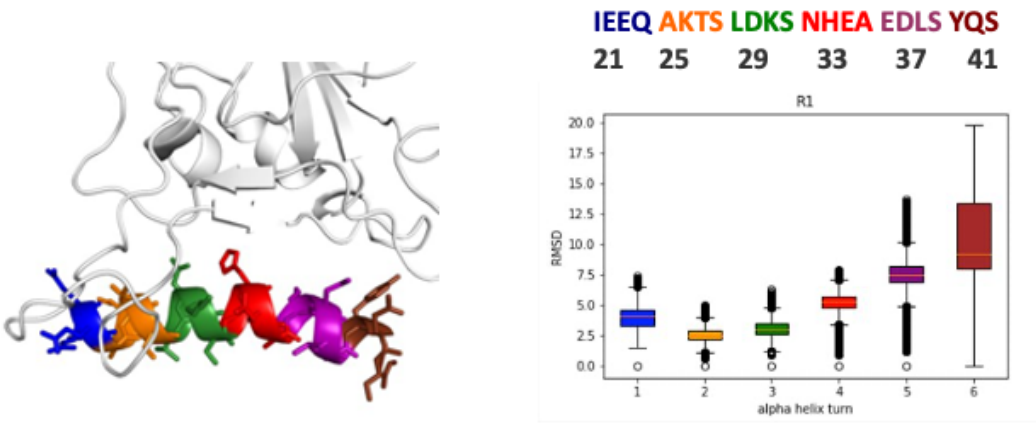



**Figure 5.12:** a. ACE2 (residues 21-45) coloured by helix turns, b. RMSD of each helix turn.

### 5.4.3 Amino acid modifications

The crystal structure of hACE-2 bound to the spike protein (PDB 6M0J) and refined by the COVID-19 structural task force,[383] was used to produce the peptide scaffold. The RBD/Ab-like complex structures, available at the time of our work were collected, to introduce mutations in the ACE2 sequence.

27 structures of Ab-like molecules in complex with the spike protein were available at the time of this work, and all of these Ab-like molecules were bound to the RBD domain. Among these 27 Ab-like molecules, 11 were found to bind the RBM site and could thus be used to introduce mutations into the $\alpha$1-helix peptide sequence (Table 5.2). Based on the results of the preliminary simulations, the mutations to more polar residues of the hydrophobic residues at the solvent-exposed interface were maintained to improve the binding.

First, 6 different peptide scaffolds, with different lengths, deriving from the $\alpha$-1 helix (residues 21 to 45) were modelled in simulations. Following the results of the preliminary simulations, the hydrophobic residues facing the solvent were mutated to serine. After selection of the most stable sequences from unstapled peptides (the peptide selection process will be detailed in Section 5.4.4), staple residues were added. Our approach was to introduce these staple residues in place of residues not implicated in the binding, mostly solvent

| Name | PDB entry | |
|---|---|---|
| CR3022 | 6YOR | [387] |
| H11-D4 | 6YZ5 | [388] |
| H11-H4 | 6ZHD | [388] |
| REGN10933 | 6XDG | [389] |
| BD23-Fab | 7BYR | [390] |
| B38 | 7BZ5 | [391] |
| MR17-K99Y | 7CAN | [392] |
| S309 | 6WPT | [393] |
| CR3022 | 7JN5 | [391] |
| CB6 | 7C01 | [394] |
| VHH-72 | 6WAQ | [395] |

**Table 5.2:** Ab-like molecules binding to the RDB domain at the ACE2 PPI (published before June 2020)

exposed residues, to interfere the least with the binding of the peptide. Between the RBD region of the spike protein and ACE-2, there are fifteen interactions, including 13 hydrogen bonds and two salt–bridge interactions. Out of the 24 amino acid residues, forming the $\alpha$-1 helix derived peptide (residues 21 to 45), Gln24, Asp30, Lys31, His34, Glu35, Glu37, Asp38, Tyr41 and Gln42 interacts with the spike protein RBD. Thus the residues at these positions were not considered when adding the staples. The spike protein RBD surface possesses 71 % of hydrophobic residues. In previous work (Chapter 1) we have established a relation between the hydrophobicity of the receptor surface and the angle which the inserted staple on the peptide makes with it. The more hydrophobic the receptor interface surface is, the more the staple leans towards it. Here we suggest that the angle of the staple should not be directed toward the solvent, but with a 30 to 50° angle with the protein surface. Ten combinations were found that permit the addition of i, i+ 4 staples according to these observations: Glu22-Lys26, Ala25-Leu29, Leu28-Phe32, Leu29-Asn33, Phe32-Ala36, Ala36-Phe40, Arg39-Ser43, Phe40-Ser44, Ser44-Lys47 (the different positions are reported in Figure 5.13 for clarity). After selecting the most stable stapled peptides in simulations (Section 5.4.4), TAMRA fluorophores were added on the sequence. For each peptide, two additional simulations were conducted. One adding the TAMRA fluorophore on the C-terminus and one adding the TAMRA fluorophore at the N-terminus.

After aligning all Ab-like/RBD structures, using the backbone C$\alpha$ atoms of the RBD domains, the interacting residues of the Ab-like molecules were defined as the residues presenting apolar contacts (defined as intermolecular carbon-carbon distance within 4 Å) or presenting hydrogen bonds (Baker-Hubbard definition). Mutations were performed based on the distances between the Ab-like residues C$\alpha$-C$\beta$ and the $\alpha$1-helix residues C$\alpha$-C$\beta$ atoms. If these distance were within a cut-off (3 Å), the mutations were performed. Once added into the sequence, mutated amino acids underwent a side-chain position adjustment, bringing mutated side-chains as close as possible to the side-chain position in the Ab-like complex structure.

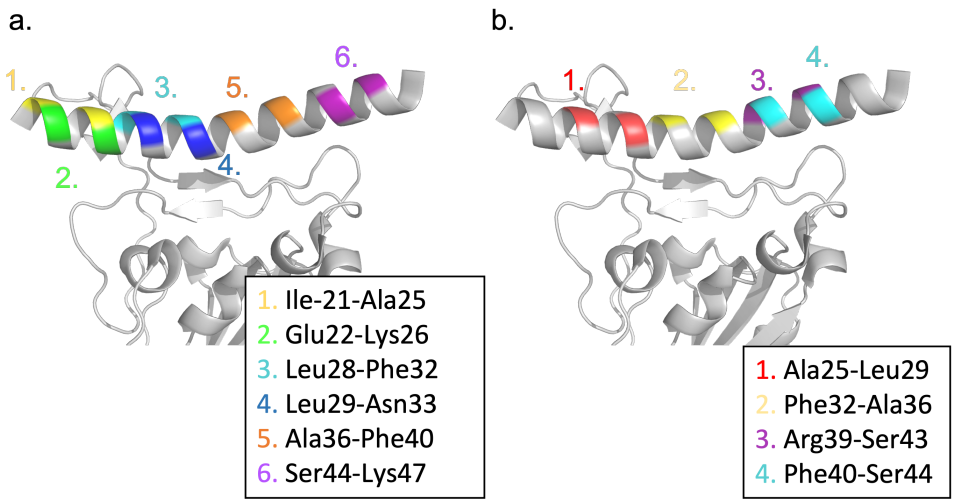



**Figure 5.13:** a. and b. ACE2 bound to the Spike RBD (PDB 6MOJ) the different staples positions have been highlighted in colours

This was achieved though successive rotations of the side-chain dihedral bonds. In total 50 mutations were found, and by combining single and double mutations, simulations of 120 complexed peptides were ran featuring 80 variations around the $\alpha$1-helix of ACE2.

### 5.4.4 Analysis of the trajectories and peptide selection

In a large number of simulations, the $\alpha$-1 helix-derived peptides did not demonstrate tight binding to the spike protein and dissociated from the protein, partially or entirely. These peptides were easily classified as non-binders as they could not maintain their interactions during the 200 ns simulations. Generally, the simulations where the peptide dissociates from the receptor are easily recognisable by the high RMSD values of the peptide backbone (above 10 Å). In contrast, peptides with partial dissociation are harder to differentiate using analytic tools, especially if the peptide re-binds to the spike protein later in the simulation. In other simulations, the peptides rearranged their backbone conformation entirely from the initial binding mode, and started to interact with a different surface of the RBD. Rarely, the new binding conformation of the peptide could offer some added interactions, such as a larger contact area; it also generally led to a loss of helicity. Commonly, we observed that these alternate binding modes were not replicated were starting from the same initial conformation. These alternate binding poses were also generally not maintained during the simulations where these conformations when taken as initial starting points. Only a few peptides showed a tighter interaction with the spike protein epitope and were considered suitable binders. Adding fluorophores to the sequence also increased the instability of the peptides in some instances, with the fluorophore interacting with the protein. In later designs only the fluorophores at the N-terminus were modelled.

A group of simulations were manually classified. Peptides were classified into three groups

based on their visual stability in simulations (Table 5.3).

| Category | Characteristics |
|---|---|
| Category I | 'Stable' binders in MD.<br>Peptides in this Category remained bound to the RBM and most of initial inter-molecular interactions were conserved. |
| Category II | 'Moderately stable' binders in MD.<br>Peptides in this Category remained mostly bound to the RBM. Dissociation of the tail or head of the peptide, rearrangement of inter-molecular interactions or partial unfolding of the peptide could be observed. |
| Category III | 'Unstable binders' in MD.<br>Complete dissociation of the peptide, drastic conformational rearrangements (e. g. unfolding, head to tail bending of the helical scaffold) or complete unfolding could be observed. |

**Table 5.3:** Categories used when visually classifying the stability of the peptide interactions.

To extract information from a large number of MD trajectories, quantitative descriptors that describe the quality of binding must be identified. A series of descriptors were computed for each simulation: the helicity of the peptide, the root mean square deviation (RMSD) of the peptide backbone, the RMSD of key side-chains, the root mean square fluctuation (RMSF) of the peptide backbone, the radius of gyration (Rg) of the peptide, the number of hydrogen bonds, the number of non-polar interactions, the number of conserved hydrogen bonds and non-polar interactions after 100 and 150 ns. Some descriptors (RMSD, RMSF) are helpful to discard the weaker binders in simulations. RMSD Values higher than 10 Å generally meant that the peptides dissociated; RMSF values higher than 5 Å for the temini residues were indicative of their dissociation; but individually, RMS calculations were not sufficient to select the best binders as some peptides with low RMS values were not found to be stable in simulations. In Figure 5.14 we can see the distribution obtained for the different descriptors using the visually-assigned categories described in Table 5.3. Generally, descriptor distributions tend to be narrower for the peptides in Category I, while for peptides in Category III, the distribution tends to be broader. However, the overlap between descriptors distribution (Figure 5.14) prevented the use of a single descriptor to select stable binders or eliminate unstable designs the distributions in any one descriptor.

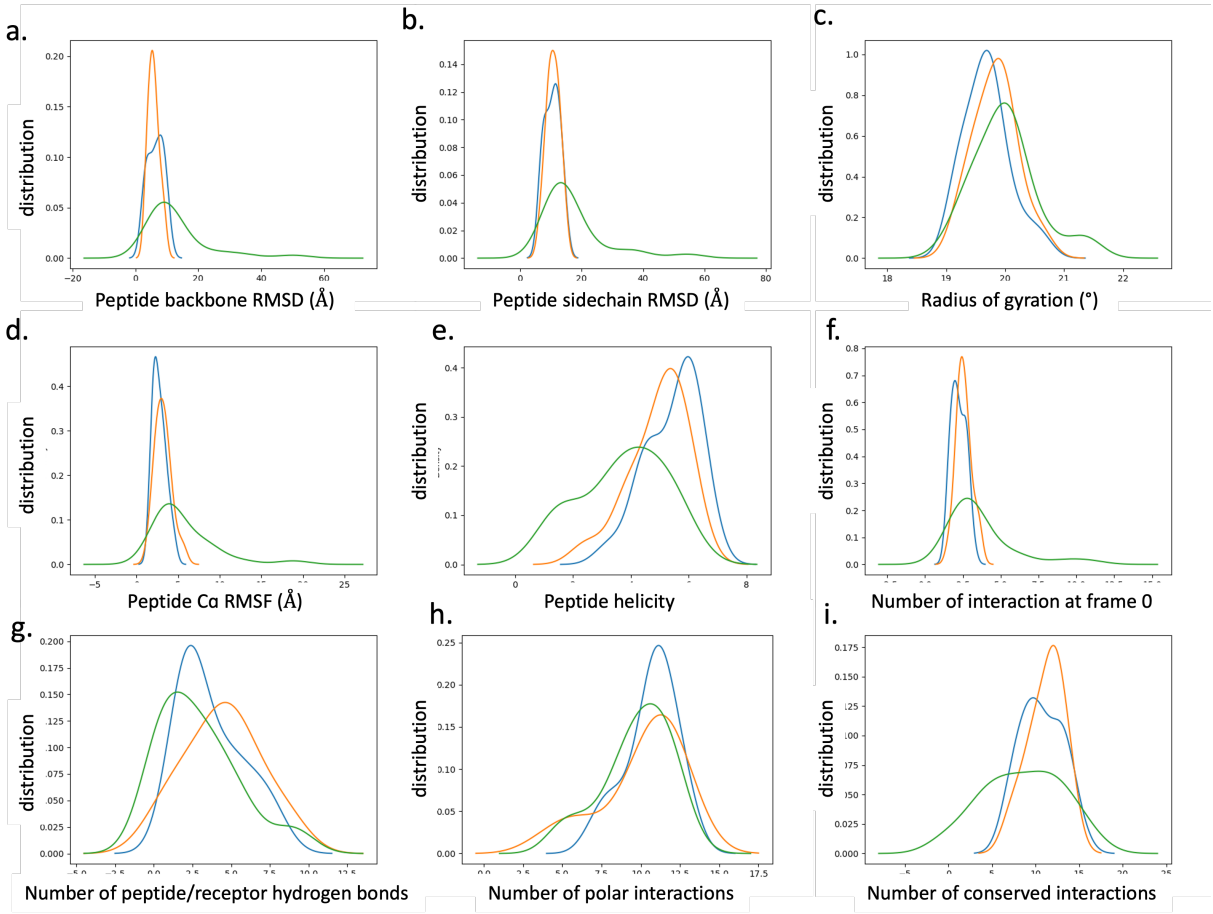



**Figure 5.14:** Distribution of different MD descriptors, for the 3 groups obtained by visual ranking of the simulations. Orange distributions refer to peptides Category I ; green distributions refer to peptides in Category II peptides; green distributions refer to peptides in Category III

### 5.4.5 PCA analysis

Given the large number of simulations to analyse, methods highlighting significant collective phenomena were needed to reduce the dimensionality into fewer components. Here, a PCA analysis was used to reduce the dimension of the different phenomena observed in simulations. First we produced components combining the different descriptors (RMSD, RMSF...) using a small set of simulations ranked visually, using the categories defined in Table 5.3. Then we used these components to classify the rest of the peptide simulations.

PCA components are linear combinations of the simulation descriptors, where each variable gets a loading. In theory, an infinite number of PCAs would give a perfect separation of the original variables. The variability of PCA1 and PCA2 shows that one single component is not sufficient for separating the simulations according their Category (Figures 5.15) and 5.16). Overall, PCA1 and PCA2 reasonably describe the peptides with the first and second PCA components 57 % and 12 % of the variance, respectively. Thus cumulatively, 69 % of the variance is described by the two components. A 2D plot of the PCA components obtained for the simulations (Figure 5.16 shows that the PCA could not produce two components that,

when used together, were able to eliminate 100 % of peptides in Category III. However, the analysis did relatively well at getting the peptides in Category I in a relatively narrow area, and allowed the identification of all binders of the set. When designing a new inhibitor, it is more crucial not to assign a binder as a non-binder than the reverse because it could potentially eliminate potential leads. When adding a third PCA component and plotting the simulations on a 3D plot, simulations were no better separated into the three categories for this data set. Adding a third vector adds 7 % to the total variance description but did not seem to separate the simulations in the best binding groups (The 3D plot is reported in appendix). Thus only two components were used.

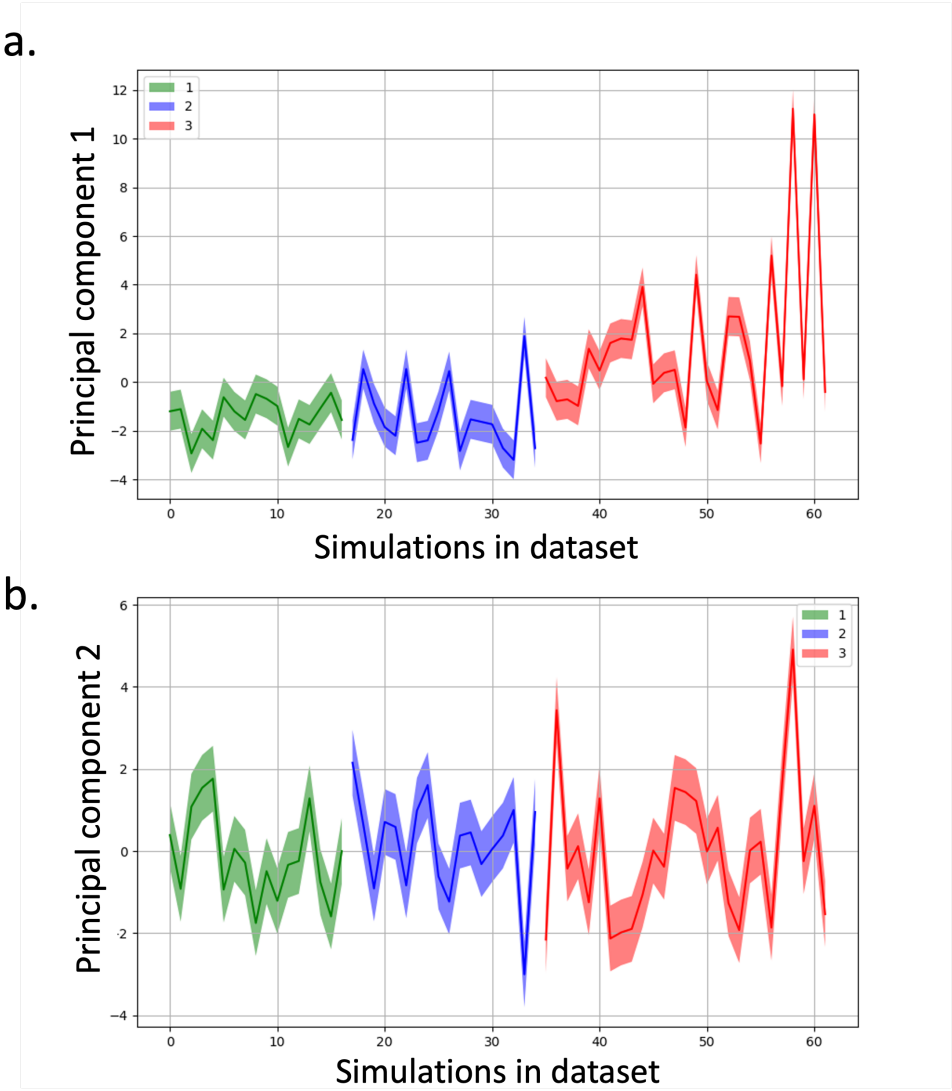



**Figure 5.15:** PCA variances: green peptides in Category I; blue peptides in Category II; red peptides in Category III

The descriptor factor score in a PCA analysis is the impact or importance of a single descriptor on the final PCA component. For the Principal Component 1 (PC-1) the eleven descriptors were found to have similar values of factor score, with absolute values between 0.19 and 0.3. This result reflects the fact that many descriptors need to be used together

to describe the binding of peptides. One exception was the helicity of the peptide with a factor score of 0.08 %. Furthermore the factor score was only 0.05 % for the second PCA component, implying that the peptide helicity was the least impacting descriptor. For PCA2, the descriptor factors varied much largely with values comprised between 0.05 % and 0.41 %. The eigenvector with the largest factor score was observed for the number of conserved interactions after 80 ns, which counted for 41 % of the PCA2 factor score.

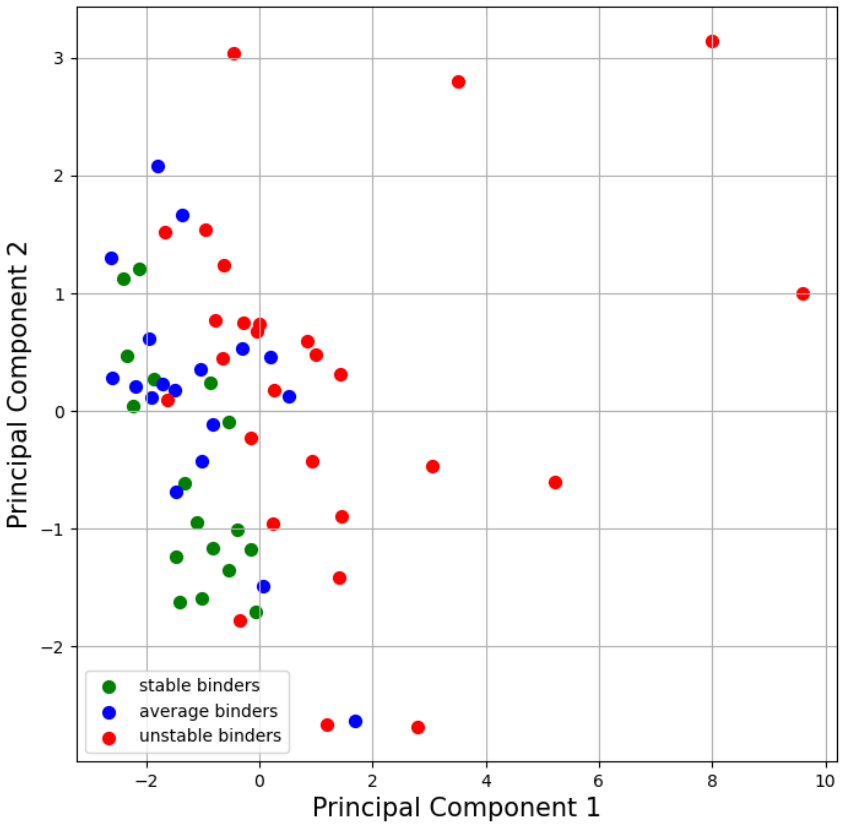



**Figure 5.16:** Separation of the peptides, through PCA into; green good binders; blue mid binders; red weak binders

Generally, for peptides in Category I, PCA component 1 was shown to be <0 and the PCA component 2 <2 for the best binders (Figure 5.16). Peptides in Category II were also observed to have a negative PCA component 1, and the second component was found to be in the 0-2 interval. Peptides in Category III were found to be present in a large area but not overly present for PCA component 1 <0 and PCA component 2 greater than 2. The fact that PCA could not give a complete separation of the peptides is in part due to the fact that peptides in categories II and III were not represented by a unique behaviour, as a conformational change, a dissociation from the receptor, a rotation or translation of the peptide binding, interaction with other parts of the RBD could, or any combination of these phenomena could led to attribute the peptides

to these categories. On the other hand, peptides with stable binding in the simulations had more stable and reproducible metrics and were thus more easily identifiable; therefore the area on the PCA graph is much narrower.

A second set of 25 independently ranked peptides was used to determine if using the previously obtained PCA component could categorise correctly the remaining simulations (Table 5.4). Among these 15 peptides, the three peptides in Category I and the seven peptides in Category II were correctly identified using the PCA analysis. Thus none of the most stable binders were incorrectly identified. This result is very positive as our main concern is not to miss stable binders. That the PCA could recognise peptide in Category II was also a really positive result, given the fact that more than one behaviour can characterise these peptides. The prediction accuracy was lower for peptides in Category III as three peptides were not identified as such. Two of these peptides were classified in Category II, and one in Category I.

| | Category I | Category II | Category III |
|---|---|---|---|
| # of simulations ranked | 4 | 7 | 14 |
| # of simulation correctly predicted | 4 | 7 | 12 |

**Table 5.4:** Result of the PCA components applied to a independently ranked set of peptides, the second line refers to the number of simulation correctly attributed to their Category by the PCA

The PCA method was also applied to published peptides (Table 5.1). A peptide developed by Curreli *et al.* [364] NYBSP-1, with a reported $IC_{50}$ of 4 $\mu M$) and 4 peptides developed by Karoyan *et al.* [365] were simulated: P7 ($IC_{50}$ = 7 $\mu M$), P8 ($IC_{50}$ = 800 $nM$) P9 ($IC_{50}$ = 300 $nM$) and P10 ($IC_{50}$ = 60 $nM$). Thses peptides were shown to be fairly stable in the simulations. Running the PCA analysis, using the same descriptors as precedent placed the peptides in categories I and II, and corresponded to the visual inspections of these peptides. Thus the PCA was found to be satisfactory at classifying the simulations.

### 5.4.6 Peptide selection

Three different design rounds were undertaken. Different staple positions, as well as dye positions, were also explored (Table 5.5 - peptides ACE-P1 to ACE-P4) Following this first selection, 30 mutations at seven positions were found through the scanning of the peptides segment with the known antibody binders. Using the ACE-P1, ACE-P2, ACE-P3, ACE-P4 as scaffold, a total of 4 positions for mutations (introduced using the Ab-like interfaces) were found beneficial for increasing the stability of the peptides in simulations. Mutations with a stabilising effect were then combined. The PCA method was applied to the 150+ peptide simulations. The PCA method selected 30 simulations, where the peptides had more stable interactions with the ACE2 receptor. These simulations were re-run. Around 45 % of the selected peptide simulation duplicates showed a difference of binding stability with a lower ranking in the newer replicas, while 57% showed a behaviour difference with a lower ranking

in at least one replica. Overall we selected 12 mutations which increased the stability of the peptides in simulations. A second screen was run by combining the mutations found to increase the peptide stability. The peptides were again chosen through PCA analysis and a final visualisation of the simulations to ensure the PCA results.

| WT | | | I | E | E | Q | A | K | T | F | L | D | K | F | N | H | E | A | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ACE-P1 | Ac- | Lys(5-TAMRA) | S5 | E | E | Q | S5 | K | T | S | S | D | K | S | N | S | E | S | $NH_2$ |
| ACE-P2 | Ac- | Lys(5-TAMRA) | | | | | | K | T | S | S5 | D | K | S | S5 | S | E | S | $NH_2$ |
| ACE-P3 | Ac- | | S5 | E | E | Q | S5 | K | T | S | S | D | K | S | N | S | E | Lys(5-TAMRA) | $NH_2$ |
| ACE-P4 | Ac- | Lys(5-TAMRA) | S | E | E | Q | A | K | T | S | S | D | K | S5 | N | S | E | S5 | Lys(5-TAMRA) $NH_2$ |
| ACE-P5 | Ac- | Lys(5-TAMRA) | S5 | E | E | Q | S5 | K | T | S | S | D | K | S | M | S | E | S | $NH_2$ |
| ACE-P6 | Ac- | Lys(5-TAMRA) | S5 | E | E | Q | S5 | K | T | S | S | D | K | S | M | S | E | S | $NH_2$ |
| ACE-P7 | Ac- | Lys(5-TAMRA) | S5 | E | E | Q | S5 | K | T | S | S | D | K | Y | N | S | E | S | $NH_2$ |
| ACE-P8 | Ac- | Lys(5-TAMRA) | S5 | E | E | Q | S5 | K | T | F | S | D | K | S | M | S | E | S | $NH_2$ |
| ACE-P9 | Ac- | Lys(5-TAMRA) | S5 | E | E | Q | S5 | K | T | S | S | D | S | S | N | S | E | S | $NH_2$ |
| ACE-P10 | Ac- | Lys(5-TAMRA) | S5 | E | E | Q | S5 | K | T | S | S | D | K | S | M | S | E | S | $NH_2$ |
| ACE-P11 | Ac- | Lys(5-TAMRA) | S | E | E | Q | A | K | T | A | S | D | M | S5 | N | Y | E | S5 | $NH_2$ |

**Table 5.5:** Sequences of the most stable peptides in simulations. Mutations from the $\alpha$1-helix sequence to increase solubility (to serine) have been highlighted in yellow. Mutations introduced from the Ab-like interface screening have been highlighted in green. S5 residues correspond to the pentenyl alanine residues and are coloured in red.

Following the first phase of biological assays, another series of longer peptides was developed. In this peptide series, a longer segment of the ACE2 peptide was chosen to increase the surface interacting with the peptide. Staple positions were conserved from the shorter peptide designs. Single mutations were introduced using the Ab-like side-chain, mutations that produced more stable binders were combined. Overall fewer mutations were found to stabilise the peptide compared to shorter peptides. Three mutations in the C-terminus were found to behighly beneficial for the binding of the peptide (Table 5.6).

| WT | | I | E | E | Q | A | K | T | F | L | D | K | F | N | H | E | A | D | L | F | E | Y | Q | S | S | L | A | S | W | N | Y | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ACE-PL1 | **Ac-Lys(5-TAMRA)** | I | E | E | E | A | K | T | S5 | S | D | K | S5 | N | E | E | A | E | M | R | E | Y | Q | S | S | L | A | K | D | N | D | $NH_2$ |
| ACE-PL2 | **Ac-Lys(5-TAMRA)** | S5 | E | E | Q | S5 | K | T | L | S | D | K | S | N | H | E | A | E | M | R | E | Y | Q | S | S | L | A | K | D | N | D | $NH_2$ |
| ACE-PL3 | **Ac-Lys(5-TAMRA)** | I | E | E | Q | A | K | T | L | S | D | K | S | S5 | E | E | A | S5 | M | R | E | Y | Q | S | S | L | A | K | D | N | D | $NH_2$ |
| ACE-PL4 | **Ac-Lys(5-TAMRA)** | | E | E | Q | S5 | K | T | L | S5 | D | K | S | N | E | E | A | E | D | R | E | Y | Q | S | S | L | A | K | D | N | D | $NH_2$ |

**Table 5.6:** Sequences of the most stable peptides in simulations during round 3. Mutation from the $\alpha$-helix-1 sequence to increase solubility (to serine) have been highlighted in yellow, Mutations introduced by the Ab-like interface screening PPI binding have been highlighted in green. S5 residues correspond to the pentenyl alanine and are coloured in red.

## 5.5 Targeting the binding sites of antibody-like molecules

### 5.5.1 Aim

Many publications emerged in 2020-21 using the ACE2 helix-1 domain as a initial scaffold for peptide design. Here, in contrast, we propose a novel method to design $\alpha$-helical sta-

pled peptides from Ab-like paratopes, which are not helical. Peptides are designed such that they structurally mimic key side-chains interactions of Ab-like molecules bound to the spike protein. 'Stapline', a pipeline for the design of stapled peptides, was built to take advantage of the knowledge of the PPIs of these Ab-like molecules PPI to create *de novo* binder Ab-like molecules usually present 1, 2 or 3 loops implicated in the binding motif. The Stapline protocol allows for the design of peptides presenting similar amino-acid interactions with the RBD surface, reproducing the side-chain contacts. The method builds peptides into helical scaffolds which permits the addition of staple residues.

The amino acid composition of the binding surface of peptides, proteins and Ab-like molecules are remarkably different. Antibody paratopes are rich in aromatic residues such as tyrosine and tryptophan.[396] It also has been reported that only 12 interacting residue pairs contribute to over 40% of the interaction energy in antibody-antigen interactions. In contrast, PPIs use a larger set of amino acids at their binding interfaces.[396, 397] Peptides usually have a slightly higher number of aromatic residues than proteins at their binding interfaces (Chapter 1). Thus utilising the Ab-like residues instead of a protein segment might result in more peptides with amino acid combinations closer to the peptides currently being tested as drug.

### 5.5.2 Pipeline for building *de novo* peptides from PPI interactions

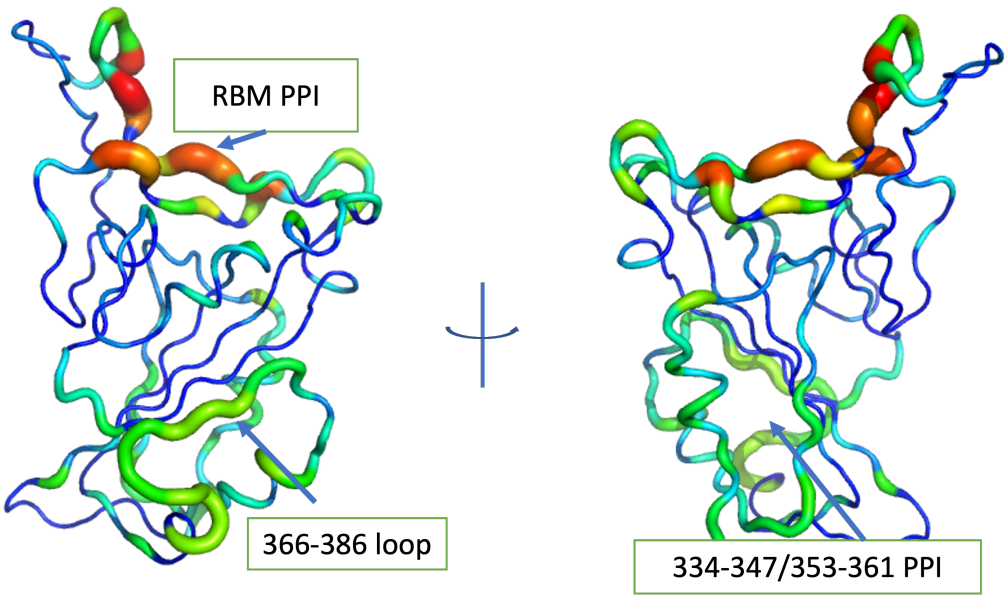



**Figure 5.17:** Binding area of the antibodies with the RBD of the spike protein: the three main PPI are highlighted with a thicker backbone.

All available crystal structures of antibodies in complex with the spike protein were retrieved from the PDB databank, this study took place in June 2020 and since, more antibodies

structures have been published. At the time of the study, 22 structures of Ab-like molecules in complex with the spike protein had been published in the PDB, of which 15 were non-redundant, as some Ab-like molecule structures were present in multiple PDB entries (Table 5.7). Three main interacting sites were found on the RBD to accommodate the binding of Ab-like molecules. However, he Ab-like molecules were found to predominantly bind the RBM motif, although their binding modes varied. Two other PPI were found on the RBD, one in the 366-386 loop and one in the 334-347/353-361 PPI (Figure 5.17). The RBD residues interacting with the most Ab-like side-chains have been highlighted in Figure 5.17 according to the number of interactions between the RBD and the different Ab-like molecules. The thicker red regions indicate more interactions while the thinner blue regions indicate no interaction. Most of the antibodies target the RBM motif or contiguous areas, while the two other binding sites had less interactions with Ab-like molecule side-chains.

| Antibody name | PDB entry | reference |
|---|---|---|
| CR3022 | 6W41, 6Z2M_antibody2 | [398] |
| CR3022 Fab | 6YLA | [387] |
| S309 | 6WPS | [393] |
| REGN10933 | 6XDG_antibody1 | [389] |
| REGN10987 | 6XDG_antibody2 | [389] |
| H11-D4 | 6YZ5, 6Z43 | [388] |
| H11-H4 | 6Z2M_nanobody1 | [387] |
| P2B-2F6 | 7BWJ | [399] |
| BD23-Fab | 7BYR | [390] |
| B38 | 7BZ5 | [391] |
| MR17-K99Y | 7CAN | [392] |
| S309 | 6WPT | [393]] |
| CB6 | 7C01 | [394] |
| SR4-1 | 7C8V,7C8W | [400] |
| VHH-72 | 6WAQ | [395] |

**Table 5.7:** Ab-like structures available at the the time of the study and used for the design of the *de novo* peptides binders

The structures of the complexes involving Ab-like molecules were aligned using the backbone atoms of the RBD. Interacting side-chains were defined as side-chains forming hydrogen bonds or non polar contacts between the Ab-like molecules and the spike protein RBD. Non polar contacts were defined as distances within 4 Å between carbon atoms . (Figure 5.18.a) Over 150 side-chains interacting with the RBD were extracted from the structures listed in Table 5.7. To find the amino acids, which would form the scaffold of the *de novo* peptide, distances between the C$\alpha$ of all extracted amino acids were computed. If a distance between 4.8 and 7.2 Å was found in a pair of side-chains, distances between the second amino acid and every other side-chain were computed. If the distance was again between 4.8 and 7.2 Å with a third amino acid, the angle between the three C$\alpha$ atoms was computed and if a value between 128 and 150° was found the three residues were inserted into an $\alpha$-helical motif

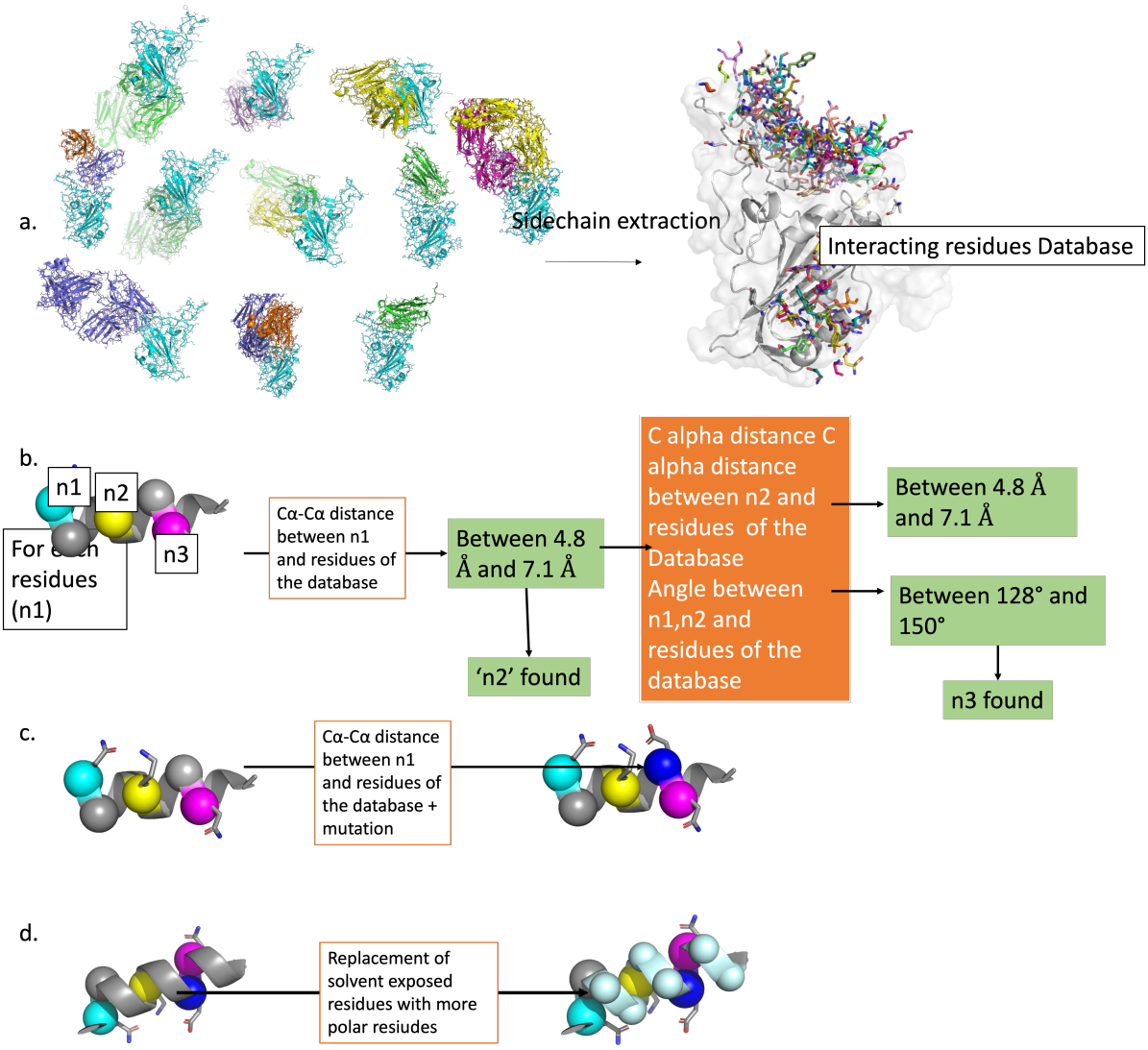



**Figure 5.18:** The steps used to produce *de novo* peptides from the known Ab-like molecule interactions: a. Side-chain extraction, b. Searching for triplets of reasonably aligned residues triplet, c. Additional Mutations; d. Changing solvent exposed amino-acids to more polar residues.

(Figure 5.18.b). The two last steps were repeated to also produce scafolds comprising four Ab-like extracted amino acid side-chains. The three (or four) amino acids were then built into an $\alpha$ helical peptide segment. Distances of 4.8 and 7.2 Å and angles between 128 and 150 °correspond to distance and angles generally found between residues of successive turn in $\alpha$-helical motifs (Figure 5.19). Twenty $\alpha$-helical peptides scaffolds were modelled using this approach. And among these peptides, 12 were $\alpha$-helical peptides where the initial side-chains were extracted from the same antibody paratope. Over 150 sequence variations around these scaffolds were tested by introducing mutations: for all alanine residues in the motif, the distances between the $C_\alpha$ and the Ab-like amino acids $C_\alpha$ were computed. For distances less than 2 Å, mutations were introduced (Figure 5.18.c). Solvent exposed residues were then mutated to more polar residues (Figure 5.18.d). These *de novo* peptide designs were then 'docked' into the structure of the RBD by alignment of the peptide side-chains with the Ab-like side-chain,

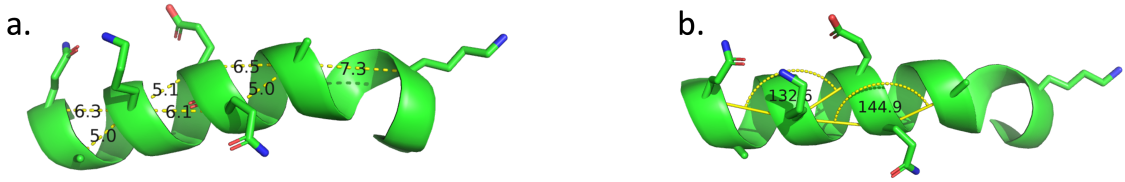



**Figure 5.19:** a. Sample of distances between C$\alpha$ of a typical $\alpha$-helical structure, b. Sample of angles between C$\alpha$ of a typical alpha-helical structure.

then a side chain packer, scwrl4, was used on the side-chains of the peptides to provide lower energy rotamers and minimise clashes in the structure.[401]

### Analysis

Similarly to the approach followed for the ACE2-mimics, the binding stability of the peptides was evaluated through MD. During the MD simulations, most peptides were found to be unstable, and dissociations from the receptor were observed. Due to the variety of binding sites and structures, using a PCA approach was not applicable here. However, some simulation descriptors, especially the RMSD of C$\alpha$, and the RMSD to the original side-chains taken for the design of these peptides, were particularly good indicators of very weak binders, where the peptides did not maintain their binding poses. Among the 200+ peptides sequences tested, only a minimal number of 20 designs were stable in simulations. By re-running these simulations, this list was reduced to less than 10, as the additional simulations did not result in stable binding. Of note, side-chains from different Ab-like molecules could be used to produce the peptides; however, in the three final scaffolds selected, the side-chains originated from a single Ab-like paratope. Eight peptides generated from 3 Ab-like interfaces were found to be the most stable peptides and are reported in Table 5.8.

| Name | initial Ab-like | Sequence | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **AB-P1** | **BD23-FAb** | D | W | A | S5 | K | N | D | S5 | S | T | A | | Lys(5-TAMRA) | $NH_2$ |
| **AB-P2** | **EY6A-Ab** | L | S5 | Y | A | D | S5 | I | K | A | Y | H | G | Lys(5-TAMRA) | $NH_2$ |
| **AB-P3** | **EY6A-Ab** | S5 | Y | A | D | S5 | I | S | K | Y | H | A | G | Lys(5-TAMRA) | $NH_2$ |
| **AB-P4** | **EY6A-Ab** | L | Y | A | S5 | D | I | S | S5 | Y | H | A | | Lys(5-TAMRA) | $NH_2$ |
| **AB-P5** | **EY6A-Ab** | L | S5 | L | A | K | S5 | S | A | D | Y | | | Lys(5-TAMRA) | $NH_2$ |
| **AB-P6** | **H11-D4-nanobody** | S5 | D | S | K | S5 | Y | W | D | S5 | Y | | | Lys(5-TAMRA) | $NH_2$ |
| **AB-P7** | **H11-D4-nanobody** | S5 | D | S | K | S5 | Y | T | S | W | M | | | Lys(5-TAMRA) | $NH_2$ |
| **AB-P8** | **H11-D4-nanobody** | S5 | D | S | K | S6 | I | D | W | Y | | | | Lys(5-TAMRA) | $NH_2$ |
| **AB-P9** | **H11-D4-nanobody** | K | A | S5 | E | S | K | S5 | E | Y | D | T | R | Lys(5-TAMRA) | $NH_2$ |

**Table 5.8:** Sequences of the most stable peptides in simulations using the Ab-like PPI approach. Residues extracted from the Ab-like molecules are coloured in green. S5 residues correspond to the pentenyl alanine and are coloured in red.

In the final design only three different Ab-like molecules were found to produce the most

stable designs. In Figures 5.20, 5.21 and 5.22, the designs, generated by this approach and found to be stable in simulation are represented. The initial Ab-like molecules are also represented and overlayed with the peptides.

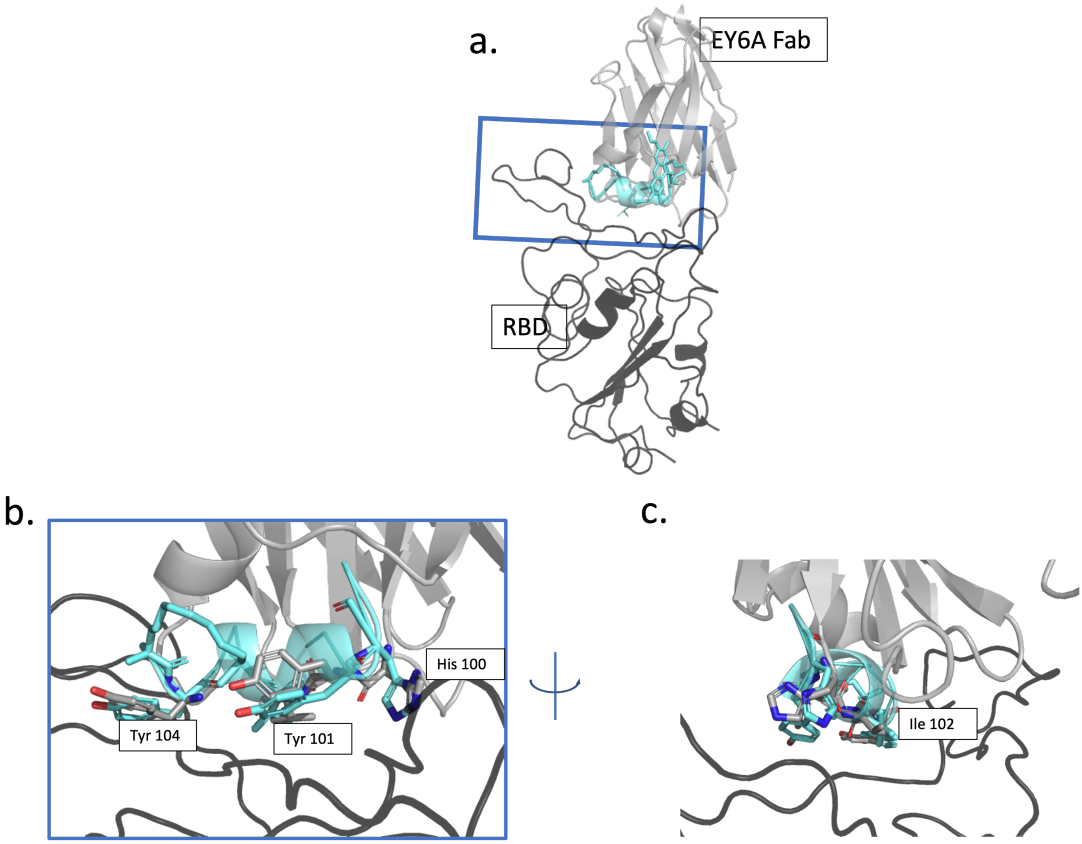



**Figure 5.20:** *De novo* designed peptide mimicking the EY6A Fab antibody bound to the RBD domain (black): the antibody structure (light grey) is overlayed with the peptide (cyan). a. Full view of the complex. b. Enlargement of the blue square in a.; the Ab-like side-chain are indicated c. axial view of the peptide, the Ab-like side-chain are indicated

When the antibody EY6A was overlayed with the *de novo* designed stapled peptide as shown in Figure 5.20. The stapled helical peptide reproduced four of the antibody interactions: Tyr104, Tyr101, His100 and Ile102. A good alignment with the original antibody side chain is obtained. During the simulation, the four side-chains remain fairly stable.

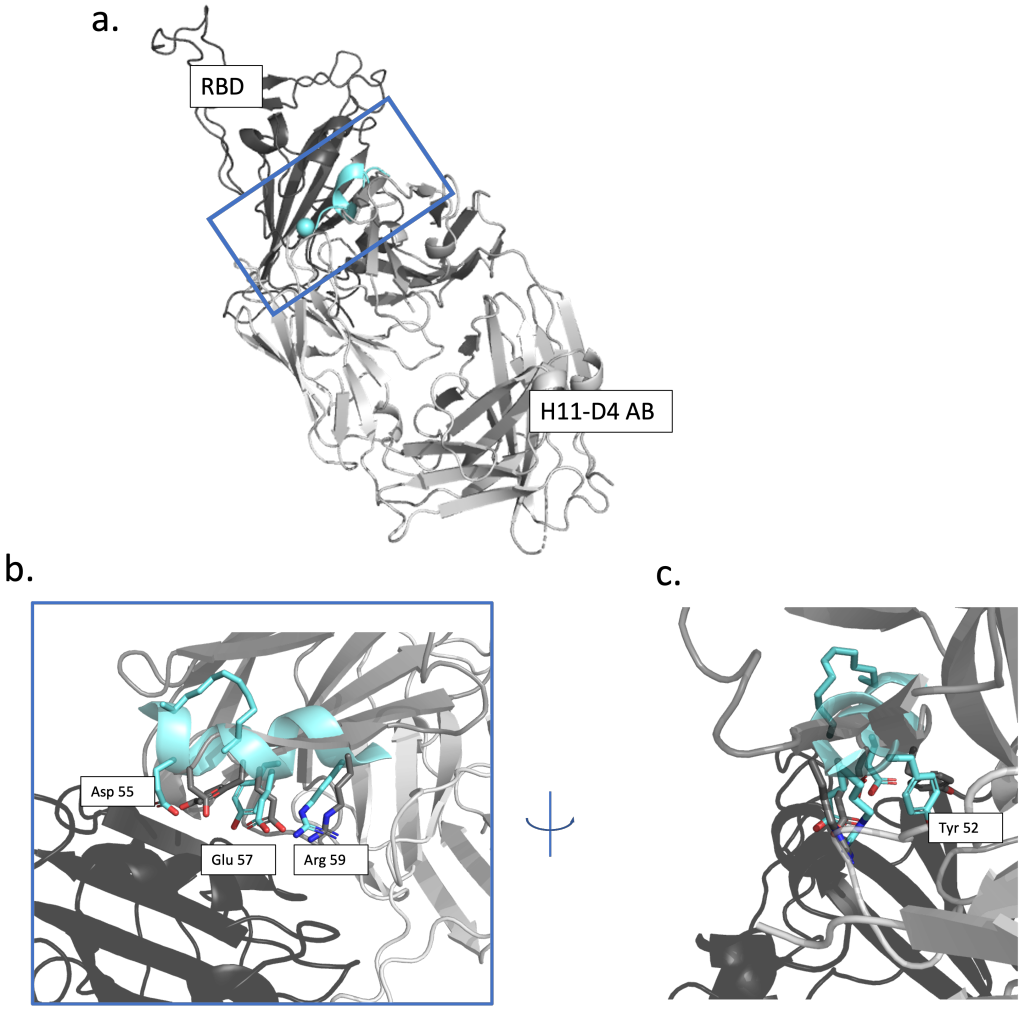



**Figure 5.21:** De novo designed peptide mimicking the H11 Fab antibody bound to the RBD (black): the antibody structure (light grey) is overlayed with the peptide (cyan) a. Full view of the complex. b. Enlargement of the blue square in a. the Ab-like side-chain are indicated c. Axial view of the peptide; the Ab-like side-chains are indicated.

When the Fab BD23 was overlayed with the *de novo* designed stapled peptide ( Figure 5.22), the stapled helical peptide reproduced four of the FAb interactions: the FAb Asp55, Glu57, Arg59 and Tyr52. A good side-chain aligement is obtained went aligning the structure of the AB-like complex with the peptide peptide complex.

Finally the antibody EY6A was overlayed with the *de novo* designed stapled peptide (in Figure 5.21), again four key antibody side-chain interactions were reproduced using the $\alpha$-helical scaffold of the *de novo* peptide: Ser104, Asn54, Thr30 and Trp105. In the Figure the stapled helical peptide reproduce three of the antibody interactions.

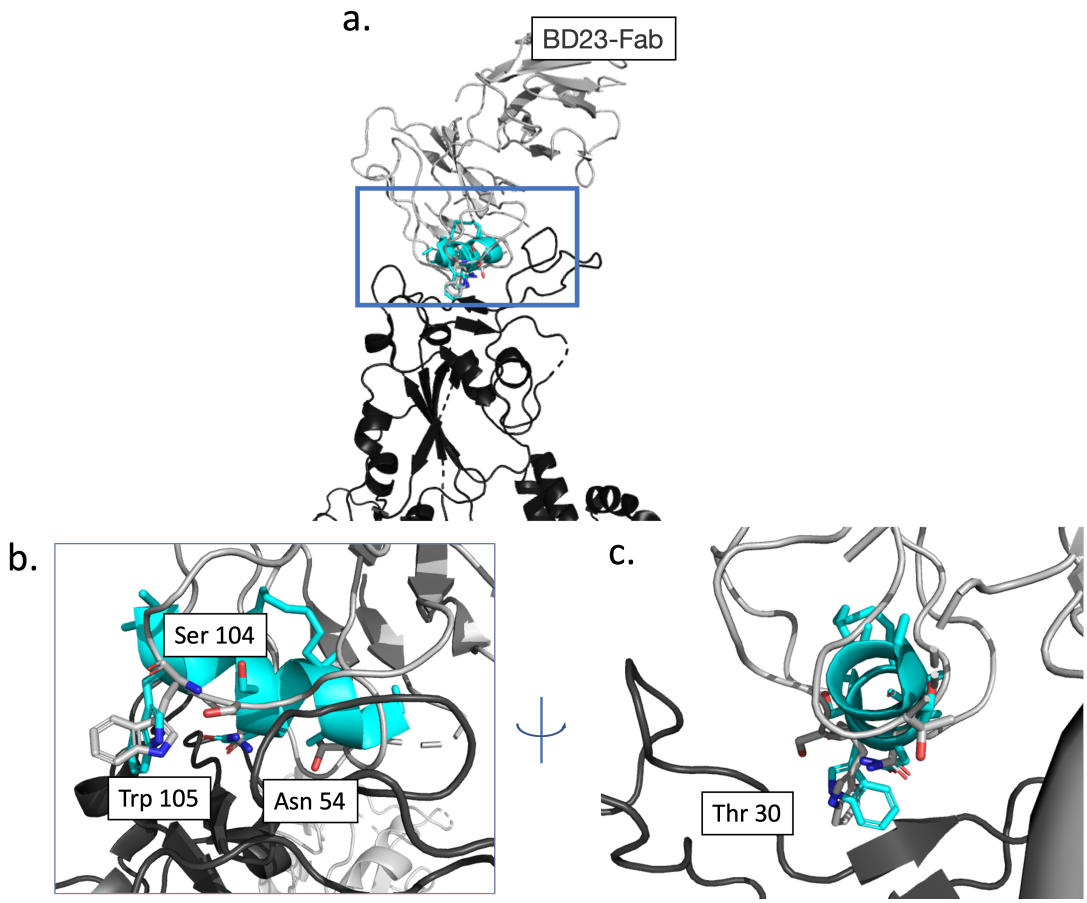



**Figure 5.22:** De novo designed peptide mimicking the BD23 Fab bound to the RBD (black): the Fab structure (light grey) is overlayed with the peptide (cyan) a. Full view of the complex. b. zoom-in (blue square in a.) the Ab-like side-chain are indicated c. axial view of the peptide the Ab-like side-chain are indicated

## 5.6 Targeting the HR2 domain used a coiled-coil approach

### Aims and context

In 1993 and 1994, two groups published groundbreaking peptide designs targeting the HIV HR domains. These peptides blocked the virus entry into cells.[402, 403] In 2004, the first peptide treatment for HIV, Enfurvirtine, a peptide targeting the HR region of HIV, was authorised.[404] Thus developing peptides to block the fusion of viruses spike proteins is an attractive approach. In section 5.2.4 we reported peptides currently known to bind the spike protein specifically, and fusion inhibitors are the second most common peptides that have been developed to target the spike protein, with four different studies reporting peptides binding to the HR1 domain. In section 5.3.3 the full-length simulation revealed weak spots in the glycosylated shield around the tail of the spike protein, and a surface on the HR2 domain that is not entirely covered by sugars was reported. We found that the HR2 domain displays 15 $Å^2$ of solvent accessible area using a probe radius of 1.5 Å. That result was later validated

by the study of spike simulations conducted by the Amaro group.[382] This surface could potentially accommodate the binding of a peptide composed of 2.1 heptad repeats, which is equivalent to 14-16 amino acids and four turns of $\alpha$-helices. Here we report the design of peptides aimed at targeting the HR2 domain by forming a hexameric coil with this segment of the HR2 domain.

### Design considerations

The spike protein undergoes large conformational changes during the fusion process. In the pre-fusion structure, the HR2 domains form a coiled-coil trimer and after fusion a hexameric coiled-coil system is formed. This 6-stranded coiled-coil comprises the HR1 domains, which form an inner trimer surrounded by a outer trimer formed by the HR2 domains, as seen in Figure 5.23. Coiled coils interact through specific helix-helix interactions that follow particular recognition patterns. $\alpha$-helical peptides can be designed *de novo* for an optimal and specific complexation with natural $\alpha$-helices. Heal *et. al* developed a graph theory, iSOCKET to classify coiled-coil systems.[405] Using their method and nomenclature the prefusion coil is a G7, meaning that each segment possess a Knob-into-hole (KIH) which each other, while the post fusion coiled-coil structure is a G94, which means that KIH interactions are present between the each coil of the inner trimer and between coils of the outer trimer and only one coil each in the inner trimer (Figure 5.23).

### Design of peptides targeting the HR2 domain

In this work it was hypothesised that the peptide would bind to the HR2 domain forming a hexameric construct similar to one found in the post-fusion structure. The G94 construct can be decomposed into an inner trimer composed of the HR2 domain and an outer trimer composed of three peptides (Figure 5.24). However the sequence of the internal trimer is not a perfect repeat (Figure 5.24). Heptad repeats comprise a repetitive heptapeptide abcdefg with a and d being hydrophobic residues characteristic of the formation of coiled-coil. Coronavirus spike proteins contain two heptad repeats in their S2 domain, a feature typical of a class I viral fusion proteins.

### Sequence scanning

Our approach was to extract the peptide sequences from the HR1 and HR2 domain. As the domains HR1 and HR2 domains fold into helical motifs it is likely that the extracted peptides have a prevalence to forming this secondary structure. Scanning through the sequences, every 12, 13 and 14-mer was extracted from the HR1-HR2 sequence. A rough filtering to match the residues on HR2 was done according to a set of rules:

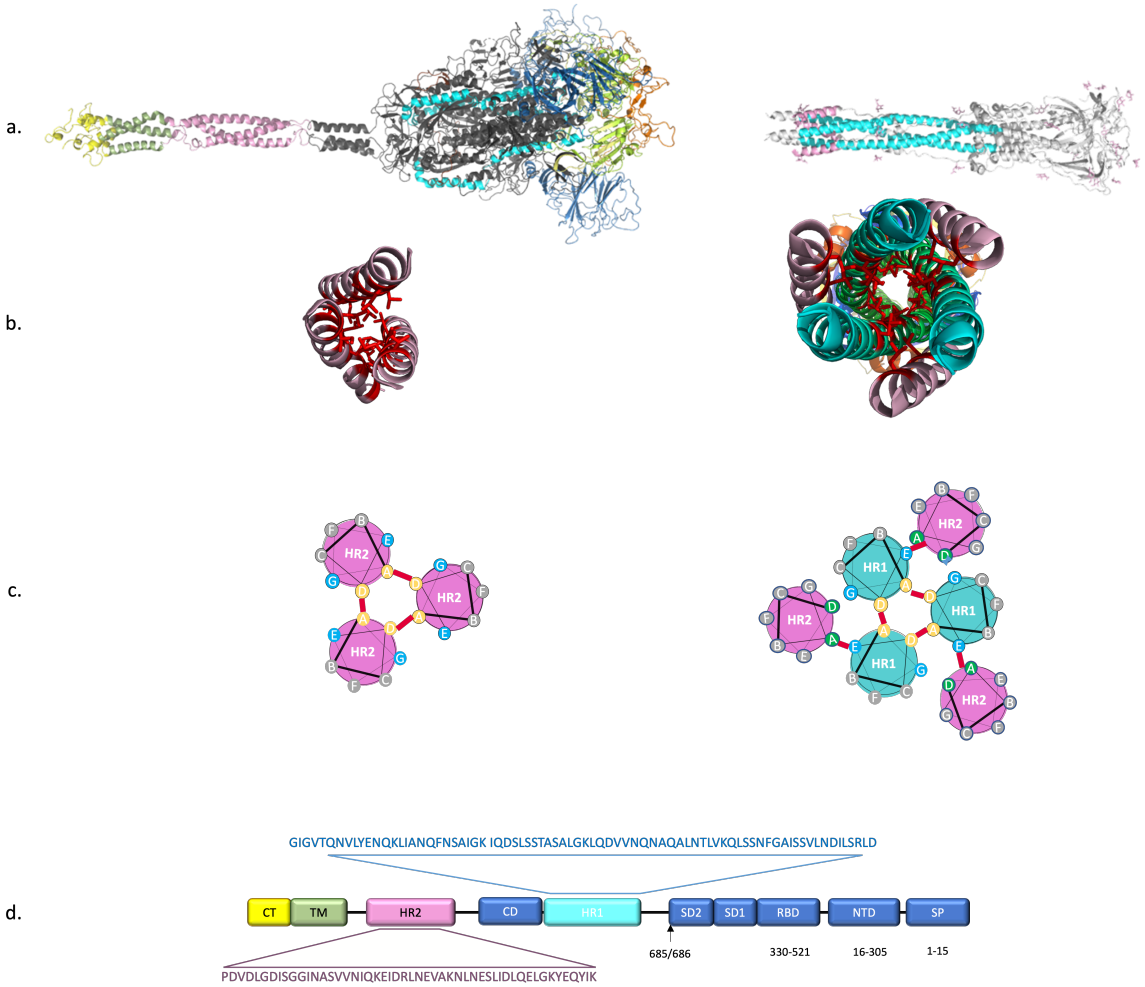



**Figure 5.23:** a. Spike protein structure in the pre-fusion conformation (left), and in the post-fusion (right) b. Coil systems: trimer formed by the HR2 domains in pre-fusion (left), hexameric coiled-coil formed by the HR2 and HR1 domains in post-fusion (right) c. System formed by the coil using iSOCKET. KIH are represented with red lines. d. sequences of HR1 and HR2.

- For the position a and d, sequences presenting amino acids with smaller side chains at these positions were preferred.
- Each a and d position was scanned and sequences which would form polar-polar or hydrophobic-hydrophobic interactions with the corresponding amino acids on the HR2 domain e and g positions were favoured ( Figure 5.25).
- In addition to a/d combinations, residues larger than Ala at the e and g sites of the peptide influence the state adopted,[406] thus small amino acids side chain at these positions were favoured.
- For solvent exposed residues (position c, d, e of the heptad repeat) hydrophobic residues are unfavorable, thus sequences comprising aromatic residues, as well as leucine and isoleucine were discarded.

A first model was built based on a scaffold generated through the alignment of the pre- and post-fusion structures. First the HR2 domain in the pre-fusion structure (the full length spike protein model described in Section 5.3) was aligned with the HR1 central domain in the post-fusion structure (6XRA). After alignment of the backbone of HR2 to the backbone of HR1 the RMSD was found to be 2.5 Å, on the segment that is not masked by the sugar. The HR2 domains of the post-fusion structure served as a scaffold for the peptides, and only the backbone of the post fusion HR2 domain was kept. Every amino acid of the peptide was assigned its position following coiled-coil nomenclature (abcdefg). The sequence was modified using the previously obtained sequences. This aproach was not found to produce stable designs in MD simulations

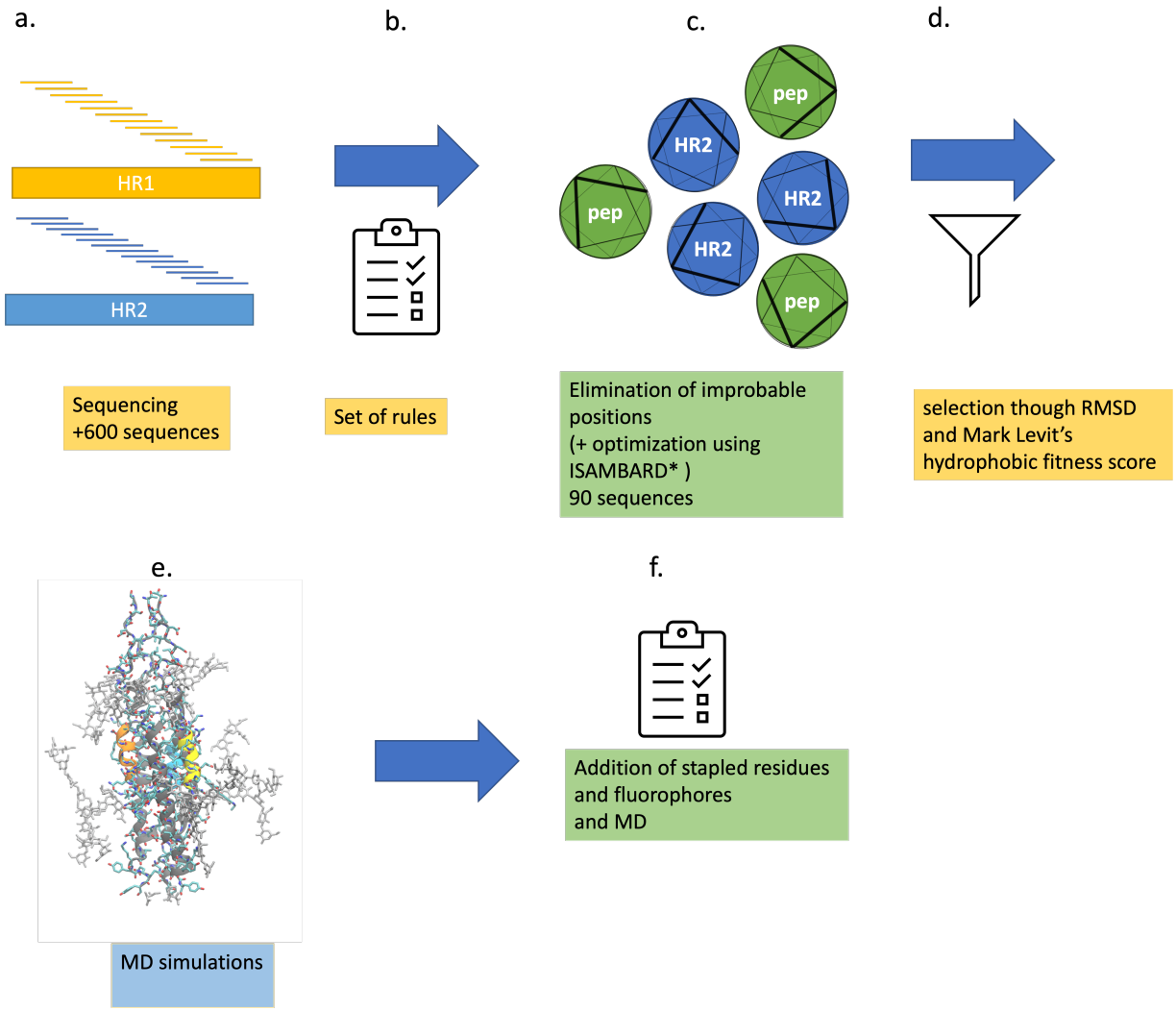



**Figure 5.24:** Extraction of the HR1 and HR2 domain sequences used during peptide scanning.

The second approach relied on the use of ISAMBARD. ISAMBARD has been created to aid parametric protein design by providing a general strategy for modelling protein coiled-coils.[304] Similarly to the approach above, truncated 15-mer peptides from the HR1 and HR2 sequences were generated. A custom class to form this pyramidal G94 coiled-coiled

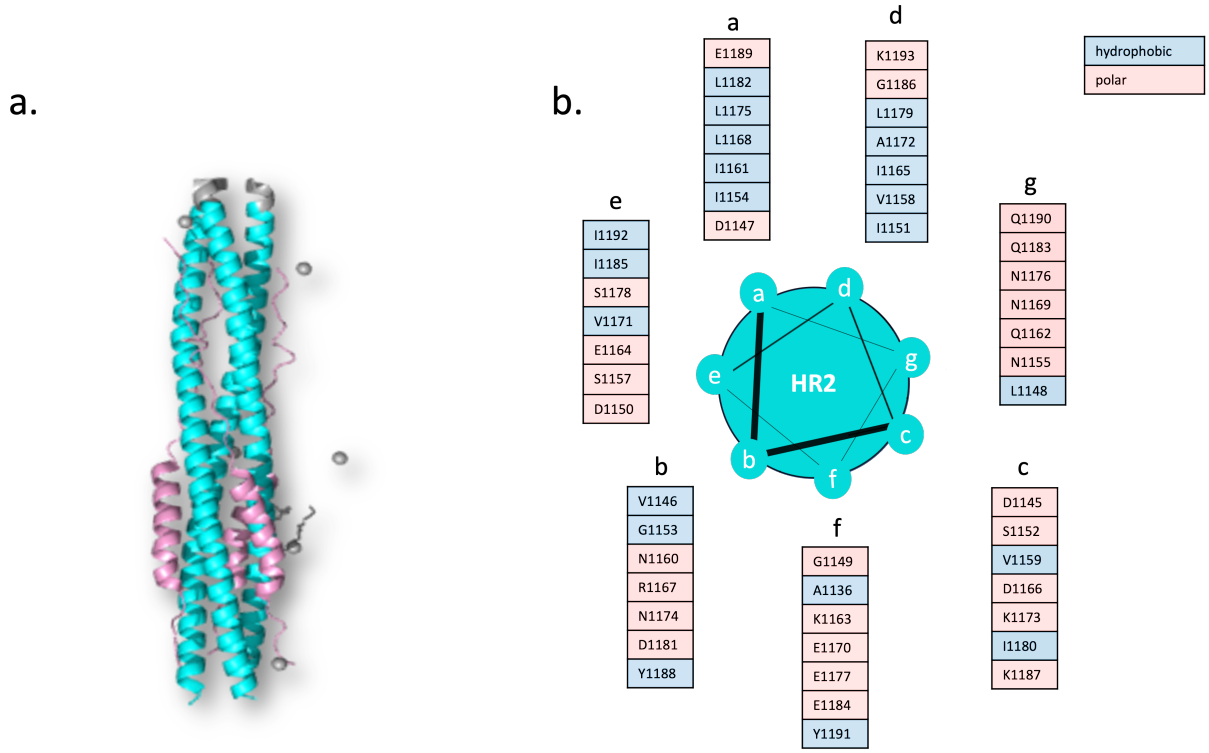



**Figure 5.25:** Evaluation of the sequence of the HR2 domain for coil interaction compatibility. a. Model used b. sequence of the HR2 domain

was created where the parameters of the inner trimer were independent of the ones in the outer trimer. ISAMBARD, possess a in-built optimisation procedure, we used the genetic evolution optimiser in ISAMBARD to fit the parameters for a given sequence. The scoring metric used was the Mark Levitt's hydrophobic fitness as well as the RMSD of the resulting inner trimer coil to the HR2 of the model.

From the 330+ hexameric coiled-coils generated by ISAMBARD, 50 models were selected using structures with the lowest RMSD to the HR2 model, a cut off of less than 10 Åon the $\alpha$-helical coil segaments was taken. The basis for this metric was that the peptides binding to the HR2 domain should not induce a large conformational change. We selected 20 models with the lowest mark Levitt's hydrophobic fitness score. The glycosylations site were added into the the model generated by ISAMBARD. During MD simulations, the models produced with ISAMBARD gave more stable simulations, probably resulting from lesser sidechain clashes and better optimisation of the inner trimer radii.

### 5.6.1 Simulations analysis

To minimise the size of the simulations, only the HR2 domain with glycans at position Asn1173 and Asn1194 were modeled. Early simulations showed an unfolding of the HR2 domains. Thus harmonic constraints were added on both ends of the HR2 domain to mimic

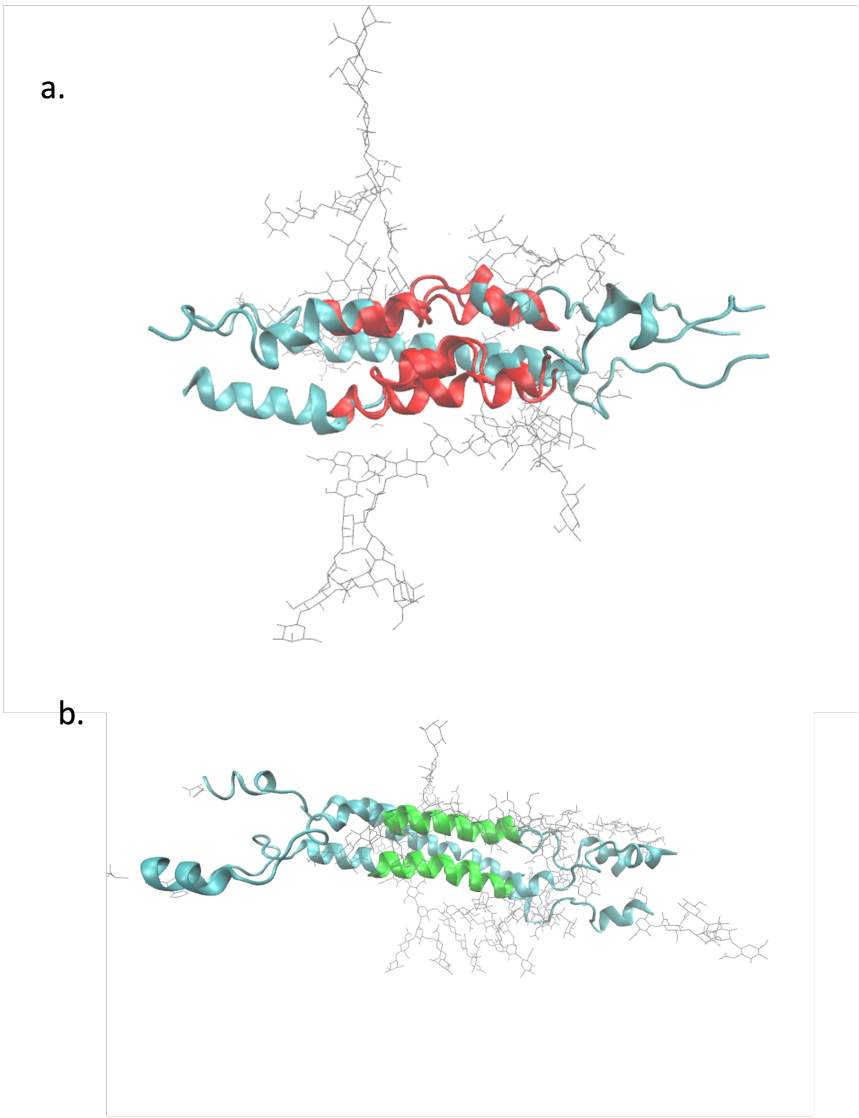



**Figure 5.26:** HR2 domain shown in cyan. The peptides are not represented. Sugars are represented in grey. The binding areas with low RMSF are represented in green; with high RMSF are represented in red. Snapshots at 0, 100 and 200 ns are overlayed. a. The HR2 domain is not affected by the binding of the peptide after 200 ns, b. The HR2 domain is affected by the binding of the peptide after 200 ns.

on one hand the membrane lipids, and at the other end the head of the spike protein. The C$\alpha$ of residues 1140 to 1150 and 1450 to 1460 were constrained. Most of the simulations resulted in the unfolding of the HR2 domain (Figure 5.26.a). In the full length simulation, the HR2 domain is flexible. In simulations of the the HR2 domain without peptides, the integrity of this segment was preserved (Figure 5.26.b). Among the 60 HR1/HR2-derived peptides tested, only two peptides were found not to disrupt the internal trimer and displayed relatively tight binding poses, in comparison with the other peptides.

More descriptors concerning the integrity of the HR2 domains were used to understand its stability, such as the RMSD, and helicity of the inner trimer. For RMSD values higher than 25 Å the structure of the HR2 domain was changing drastically (Figure 5.26 a.) Thus simulation with high RMSD value in the HR2 domain were discarded. The design produced with

ISAMBARD were found to produce peptides which induced fewer conformational rearrangements in the HR2 domain. The RSMD of the peptide backbone, was in average lower than the RMSD of the peptides designed with the first approach (9 Å and 15 AA). This is probably due to the fact, that during the coiled-coil modeling the HR2 domain radii was optimised as well as the side chains, producing structures closer to a low energy structure. For the remaining simulations, the stability of the three peptides were taken into account, as sometimes only one of the three peptides was shown to be unstable, with RMSD higher than 10 AA. The simulation descriptors were calculated for each peptide individually, and the descriptors values were compared for the three peptides of the same simulation. Simulations featuring at least one unstable peptide were discarded. The coiled-coil produced with the first approach produced more diparity between the three peptides of a simulation as the model used does not have an axial symmetry such as the one found in the model generated with ISAMBARD which produces 'perfect' geometrical coiled-coils. The two best peptides are presented in Table 5.9.

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LT113154 | I | D | K | | Q | V | S5 | E | A | I | S5 | K | S | Lys(5-TAMRA) | $NH_2$ |
| LT113155 | N | A | Q | Lys(5-TAMRA) | | L | S5 | A | L | V | S5 | A | L | S | $NH_2$ |

**Table 5.9:** Peptide sequences of the most stable simulations, both sequences result from the coiled-coil approach

## 5.7 Detection of peptides

The Auer group (University of Edinburgh) developed a robust high-throughput confocal nano scanning (CONA) approach; previous applications using this assay have been reported.[407, 408] This assay was adapted to allow the detection of peptides binding to the full-length SARS-CoV-2 spike protein. The development of this assay for screening was particularly advantageous as it is not reliant on BSL 3 laboratories required for handling live virus samples. This technique has previously been shown to detect weak binding ($K_D$ values up to 500 $\mu M$, allowing effective screening of lead compounds. [408] All experiments presented in this section have been developed and carried by Dr Nhan Pham (Universtiy of Edinburgh). This specific method has not been published but is similar to other work published by the Auer group.[408, 409]

### Controls

UPS-confocal fluorescence nanoscanning (UPS-CONA), employs a protein of interest immobilised on a micro-bead and a fluorescently labeled substract which, upon binding to the protein, is quantitatively detected on the bead periphery by confocal imaging (Figure 5.27). Upon conjugation to the protein substrate, a fluorescent "ring" will become detectable in the confocal image plane across the microbeads. (Fig 5.27.b) Images of beads in a 384-well plate,

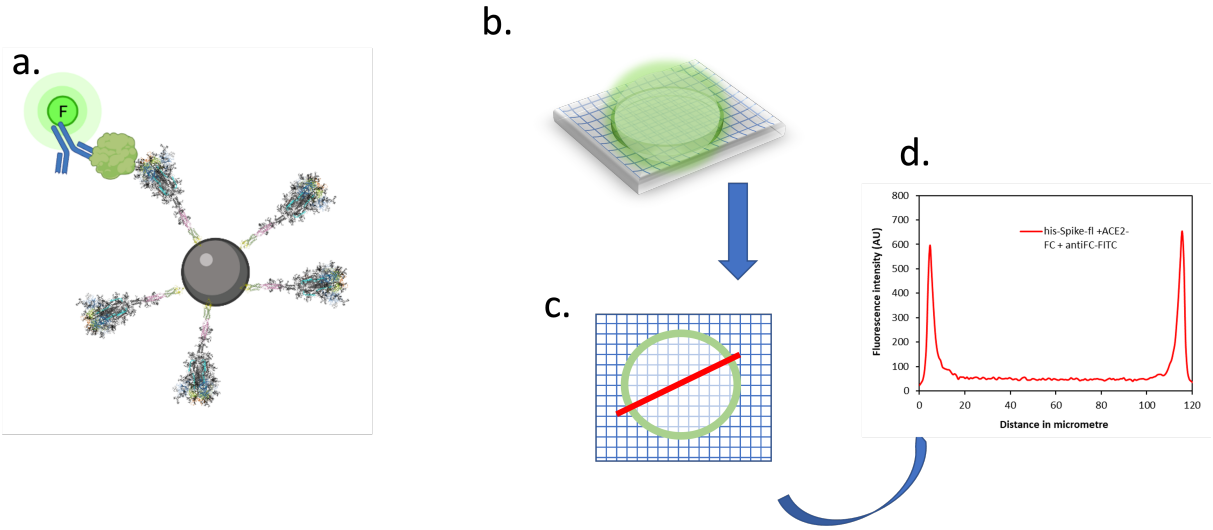



**Figure 5.27:** a. spike protein immobilised on a micro-bead and ACE2 protein and a fluorescent mAb (green). b. fluorescent "ring" c. confocal image images d. Intensity profiles across beads.

acquired using the confocal scanning microscope Opera™(Perkin Elmer) in brightfield and in the fluorescence emission channel. The intensity profiles of over 100 beads are generally analysed and the binding is indicated by a series of fluorescent rings. The intensity of the ring is indicative of the strength of binding (Figure 5.27.d). A control was conducted using ACE2 and a fluorescent anti-ACE2 antibody to detect substrate binders to the spike protein. Full-length, His-tagged spike protein (CPPC) was bound to Ni-NTA agarose beads in PBS (pH 7.4). Validation of the CONA assay was achieved through binding ACE-2 to the receptor-binding domain (RBD) of the spike protein (reported $K_D$ = 120 nM) The detection for the control was indirect as the ACE2 was not labeled and a fluorescent labeled anti-ACE2 was used. These assay were carried out using an indirect approach, by detection of the fluorescent tag of an anti-ACE2 (Figure 5.27) .

Beads alone and beads displaying the spike protein at the surface where shown to not have any fluorescence at 488 nm. (Figure 5.28.a and Figure 5.28.d) A measurement using the anti-ACE2 mAb with and without the spike protein was run to rule out possible non specific binding to the bead of the anti-ACE2 mAb to the beads or to the spike protein. No fluorescent ring was detected as the antibody was washed off the plate (Figure 5.28.b and Figure 5.28.e). The assay was run in the presence of the anti-ACE2 mAb and ACE2 protein to rule out unspecific binding of ACE2 to the beads (Figures 5.28.c). Finally a positive control in presence of the spike protein, ACE2 and the anti-ACE2 mAb-Fl was done. Rings with a high intensity were observed, characteristic of nM binding (Figure5.28.f).

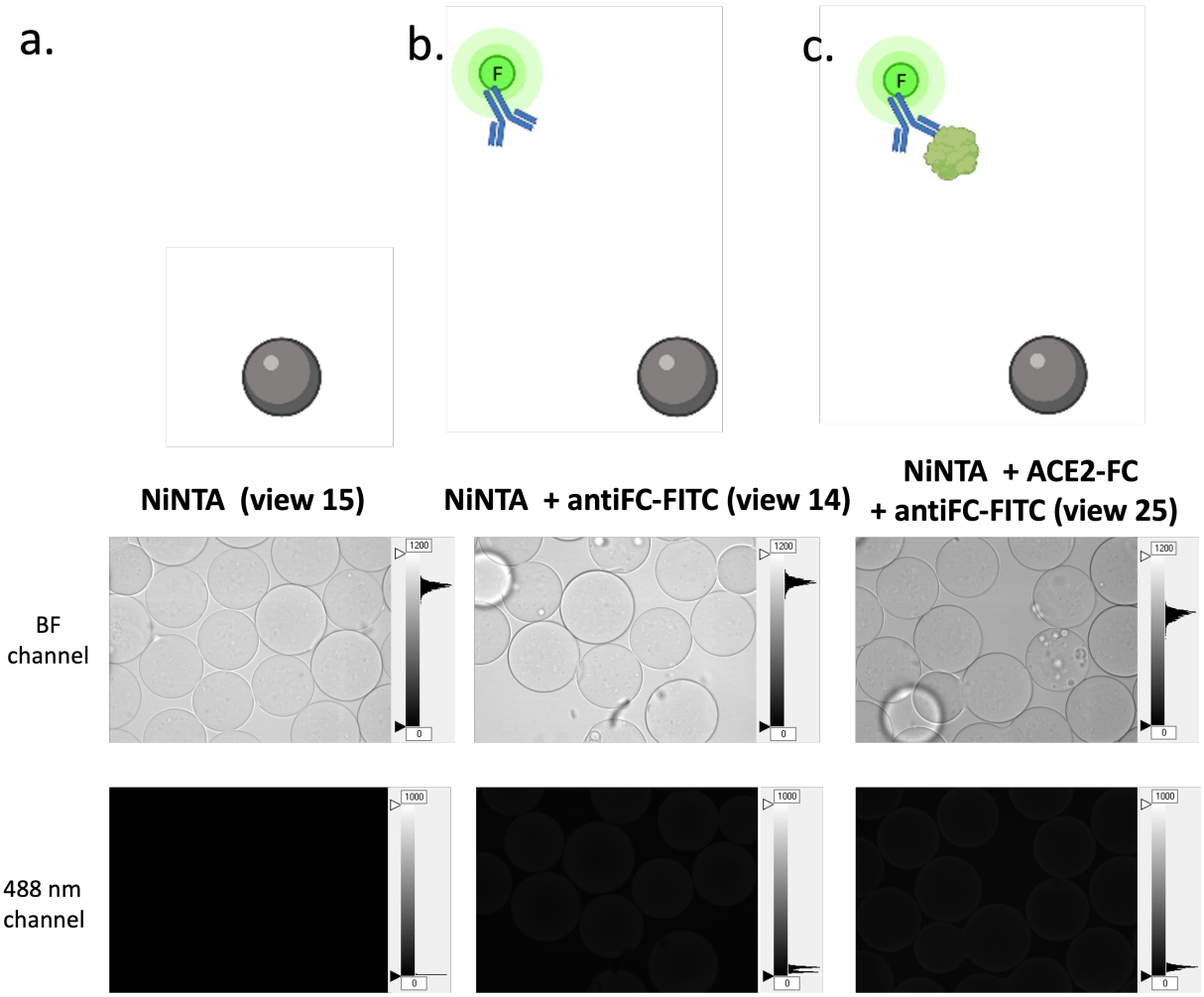



**Figure 5.28:** CONA controls: confocal image images of the brightfield (BF) of the beads (top images) and fluorescence emission channel of the dye at 488 nm (bottom images). a. beads; b. beads and anti-ACE2-Fl; c. beads ACE2 and anti-ACE2-Fl The assays were performed by Dr Nhan Pham, University of Edinburgh.

| | Bead | Protein 1 | Protein 2 | Control for | Observation |
|---|---|---|---|---|---|
| a | x | x | x | Ni-NTA bead fluorescence | no fluorescence |
| b | x | x | anti-ACE2 | unspecific binding of anti-ACE2 to beads | no fluorescence |
| c | x | ACE2 | anti-ACE2 | unspecific binding of ACE2 to beads | no fluorescence |
| d | His-spike | x | x | bead fluorescence | no fluorescence |
| d | His-spike | x | anti-ACE2 | unspecific binding anti-ACE2 to spike protein | no fluorescence |
| f | His-spike | ACE2 | anti-ACE2 | Positive control | bright rings |

**Table 5.10:** CONA assay controls. a,b,c,d ,e,f refer to Figures 5.28 and 5.29 sub-figures

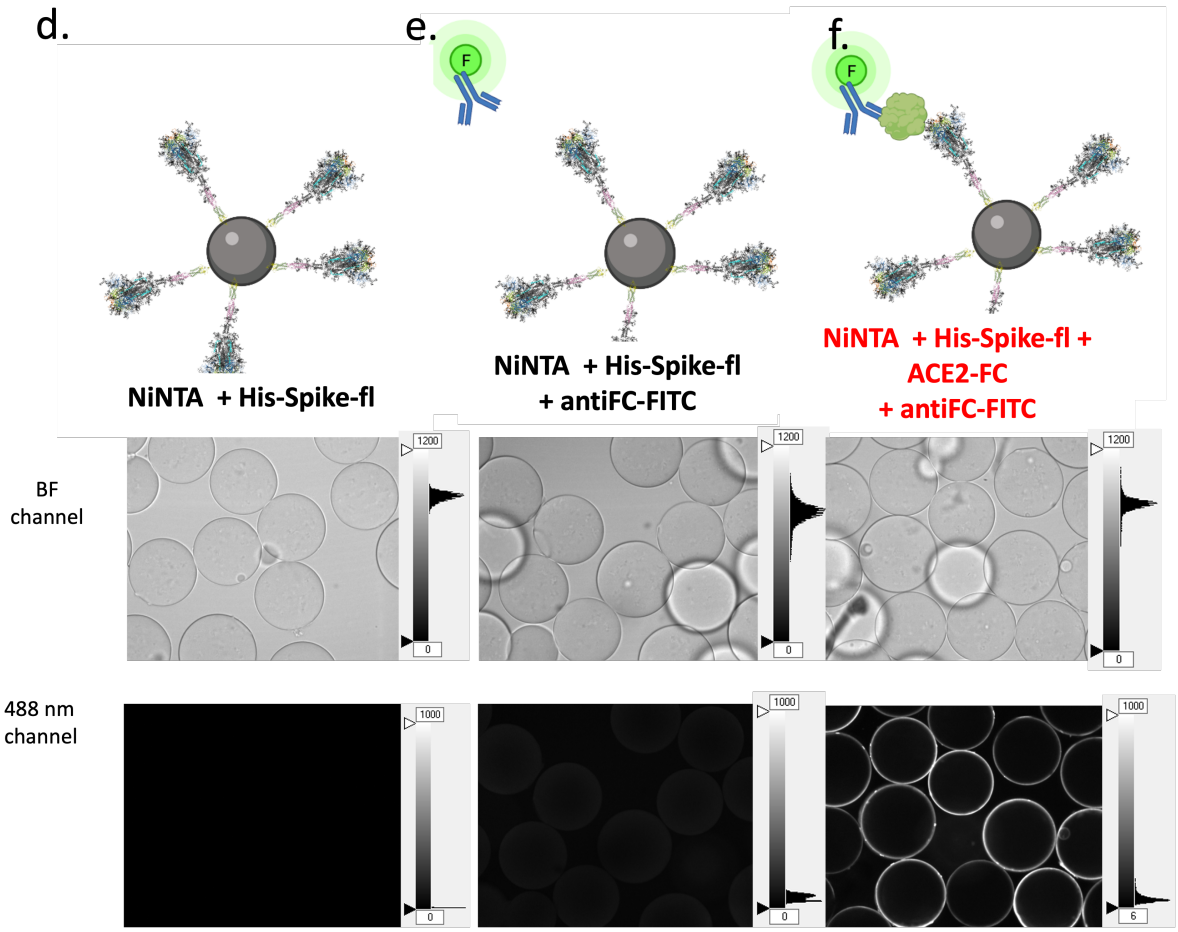



**Figure 5.29:** CONA controls: confocal image images of the brightfield (BF) of the beads (top images) and fluorescence emission channel of the dye at 488 nm (bottom images). d. beads, His-Spike; e. beads, His-SPike and anti-ACE2-Fl; f. beads, His-SPike, ACE2 and anti-ACE2-Fl. The assays were performed by Dr Nhan Pham, University of Edinburgh.

**Measuring the fluorescence of the peptides**

Control experiments made use of an indirect detection method, with fluorescently labelled anti-ACE2. To measure the binding of the peptides a direct binding assay was employed using TAMRA fluorophores linked to the peptides.

| Peptide | Design category | Sequence | CONA assay result |
|---|---|---|---|
| ACE2-P1 | ACE2-mimetic | Ac-Lys(5-TAMRA)-S5-EEQ-S5-KTSSDKSNSES-NH2 | Kd >500 $\mu$ M |
| ACE2-P2 | ACE2-mimetic | Ac-Lys(5-TAMRA)-KTS-S5-DKS-S5-SES-NH2 | Kd >500 $\mu$M |
| ACE2-P3 | ACE2-mimetic | Ac-S5-EEQ-S5-KTSSDKSNSE-Lys(5-TAMRA)-NH2 | Kd >500 $\mu$M |
| ACE2-P4 | ACE2-mimetic | Ac-SEEQAKTSSDK-S5-NSE-S5-Lys(5-TAMRA)-NH2 | Kd >500 $\mu$M |
| AB-P1 | Ab-like-mimetic | Ac-DWA-S5-KND-S5-STA-Lys(5-TAMRA)-NH2 | Kd >500 $\mu$M |
| AB-P2 | Ab-like-mimetic | Ac-L-S5-YAD-S5-IKAYHG-Lys(5-TAMRA)-NH2 | Uptake by bead |
| AB-P3 | Ab-like-mimetic | Ac-S5-YAD-S5-ISKYHAG-Lys(5-TAMRA)-NH2 | Uptake by bead |
| AB-P4 | Ab-like-mimetic | Ac-LYA-S5-DIS-S5-YHA-Lys(5-TAMRA)-NH2 | Uptake by bead |
| AB-P5 | Ab-like-mimetic | Ac-L-S5-LAK-S5-SADY-Lys(5-TAMRA)-NH2 | Kd >500 $\mu$M |
| AB-P6 | Ab-like-mimetic | Ac-S5-DSK-S5-YWDSY-Lys(5-TAMRA)-NH2 | Aggregation |
| AB-P7 | Ab-like-mimetic | Ac-S5-DSK-S5-YTSWM-Lys(5-TAMRA)-NH2 | Aggregation |
| AB-P8 | Ab-like-mimetic | Ac-S5-DSK-S5-IDWY-Lys(5-TAMRA)-NH2 | Aggregation |
| AB-P9 | Ab-like-mimetic | Ac-KA-S5-ESK-S5-EYDTR-Lys(5-TAMRA)-NH2 | Kd >500 $\mu$M |
| HR-P1 | HR2-peptides | Ac-IDKQV-S5-EAI-S5-KS-Lys(5-TAMRA)-NH2 | Kd >500 $\mu$M |
| HR-P2 | HR2-peptides | Ac-NAQ-Lys(5-TAMRA)-L-S5-ALV-S5-ALS-NH2 | Discussed in text |

**Table 5.11:** Summary of fluorescent, stapled peptides by design category, primary sequence and CONA assay result.

### ACE2 peptides

Four peptides (ACE-P1, ACE-P2, ACE-P3 and ACE-P4) (as presented in Table 5.5) designed after the ACE2 binding motif were tested using CONA assay. For the four tested peptides, after 15 min of incubation, no sharp ring was seen around the beads. Instead, the peptides diffused in the media, and the entire channel brightened, as seen in Figure 5.30. Thus no peptide derived from the ACE2 sequences was found to bind to the spike protein.

This low affinity could be the result of one or more factors. Foremost, it is mainly assumed and experimentally observed that truncating a segment of protein results in a loss of secondary structure and thus would produce a less efficient binding. However, work published recently by the Jamieson group [302] reported that the stapled peptides derived from the truncated ACE2 (with no mutations) were conserving a high helical propensity. It is probable that the lower affinity of this segment is not due to a loss of helicity. Furthermore, in the spike protein crystal structures binding to ACE2 (pdb 6M0J), we can observe that the conformation in the interacting segment on the ACE2 side is not a perfect helix. The ACE2-21-44 forms an interrupted helix around residues 30-35 where the helix is bent. This is also observed in the free form of ACE2 (PDB 1R42) and is not directly linked with the binding event, but to the ACE2 internal interactions. Circular dichroism of the peptides is not applicable to observe if the peptide is reproducing this specific conformation, but it is most likely that this conformation cannot not be replicated by a peptide, as it is most likely induced by the protein internal forces and folding arrangement. This bent provides a shape complementary with the spike protein binding domain, and we hypothesise, that if this shape complementary is lost, the binding is altered. The residue His34, at the centre of the binding area, provide three hydrogen bonds to the spike amino acids: Asn479 Tyr440 and Asn 479, and without a bend in the structure of the ACE2 protein, these native contacts would not be reproduced, using a shorter and straight helical peptide.

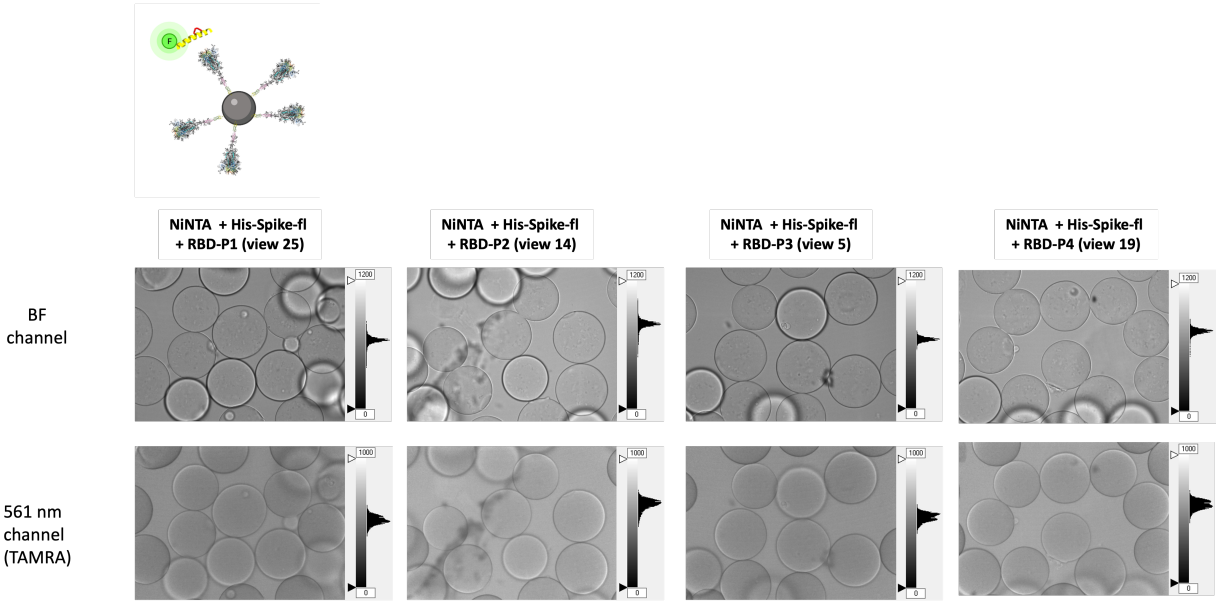



**Figure 5.30:** CONA assay of the ACE2 peptides mimics. The assays were performed by Dr Nhan Pham, University of Edinburgh.

Another hypothesis to this low affinity is the relatively small contact area offered by the $\alpha$-helix to the spike protein. If the helical segment carries most interactions, a network of interactions outside this helix has been described between the ACE2 loops 324-326 and 352-355 and the spike protein RBD loop 499-506. These interactions would not be reproduced by the extracted helical segment.[410] Furthermore, the ACE2 protein could interact with the spike protein at other contact points using sugars. A recognition by sugars could be the first stage of the binding. Further stabilisation could take place during the ACE2 binding by the sugars. Both the spike protein and the ACE2 receptor are heavily glycosylated, and several glycosylation sites are near the binding interface. Whereas the focus has been mainly on amino acid interactions in the ACE2–spike binding interface, the role of glycosylation has often been underestimated. The ACE2 receptor displays seven N-glycosylation sites (N53, N90, N103, N322, N432, N546, and N690) and several O-glycosylation sites. The N90 position might interfere with virus binding and infectivity, which was proved by a recent genetic and biochemical study [411, 412], showing that mutations of N90, which remove the glycosylation, increase the susceptibility to SARS-CoV-2 infection. With an opposite effect, four glycosylation sites (N53, N90, N103, and N322) can interact with the spike protein RBD and stabilise its interaction with ACE2. Two conclusive MD studies corroborate this hypothesis [413, 414]

A final hypothesis is that ACE2 only binds to the spike protein when the RBD of the spike is in the upright position. Different studies have reported different kinetic rates and stability between the conformations, and Amaro *et al.* in their computational study [382] even reported that different sugars bound to N234 result in different rates of of the RBD domain between the up or down position. Binding to a bulkier partner such as a protein in comparison to a

peptide could potentially change this conformational rate, favouring an open state.

### Antibody derived peptides

The CONA assay was used to test the peptides designed from the Ab-like PPI (Figure 5.31). For the AB-P5 peptide, the CONA-assay did not show any fluorescence at 488nm. Thus these peptides could directly be identified as a non-binder.

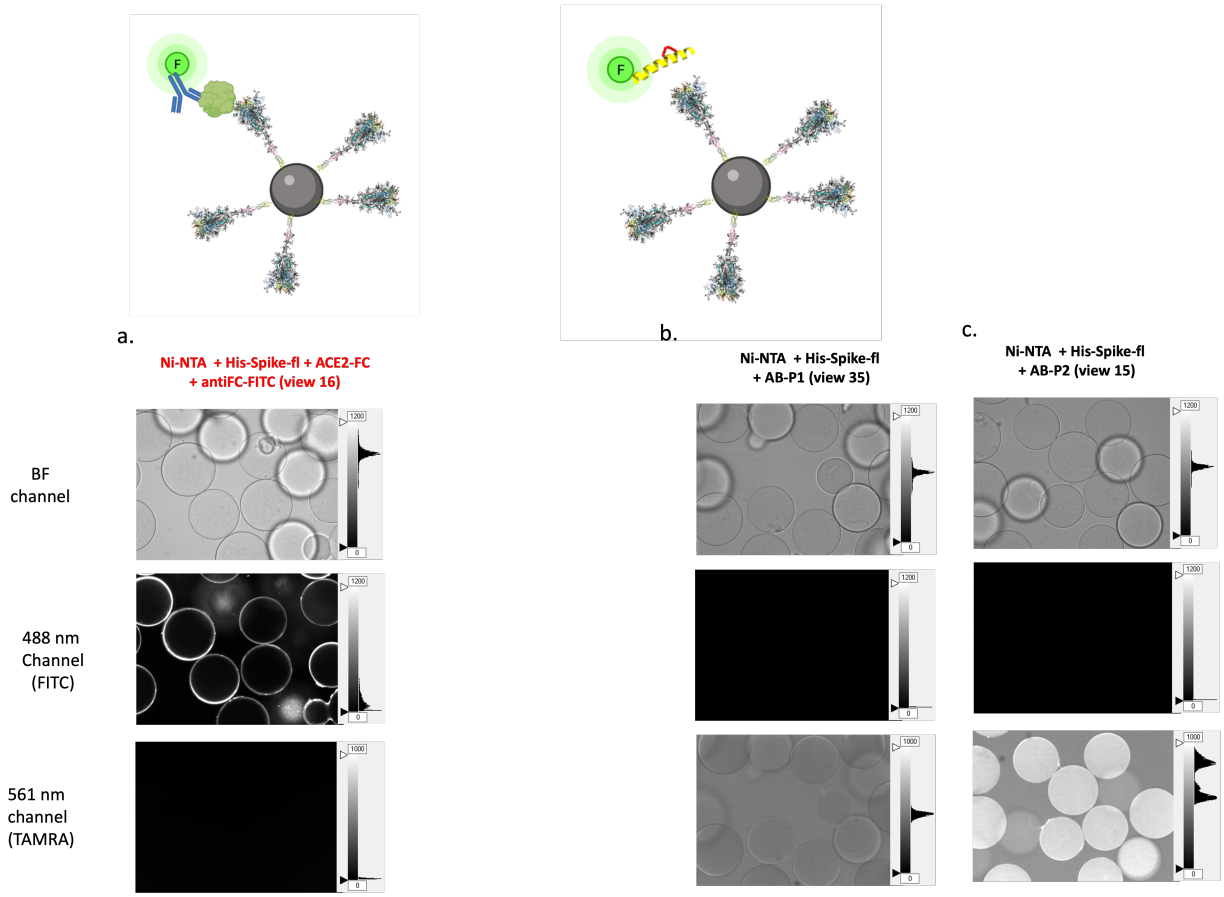



**Figure 5.31:** CONA assay a. positive control: ACE2/anti-ACE2-fl ; b. AB-P1 peptide, c. AB-P2 peptide. The assays were performed by Dr Nhan Pham, University of Edinburgh.

The peptides AB-P2, AB-P3 and AB-P4 CONA assays were found to have an unexpected outcome in which the whole beads became brighter rather than just the periphery (Figures 5.31.b and.c and 5.32 .d). Our hypothesis is that these peptides have an affinity with the Ni-NTA-His-Spike beads and enter inside the beads. The entry of the peptides inside the bead impede the detection of eventual interactions with the his-Spike protein with these peptides. Peptides AB-P6, AB-P7 and AB-P8 aggregate onto the surface of Ni-NTA-his-Spike beads in the assay, which appears as fluorescent rings. However, the rings were discontinuous, and the bead intensity profile did not reproduce the pattern of a binder, which is abnormal if the peptides were indeed binding to the spike protein. (Figure 5.34) A negative control was thus conducted using the Ni-NTA beads without the spike protein. Some fluorescence was observed in the latter experiment, which confirmed the non-specific interaction of the peptide

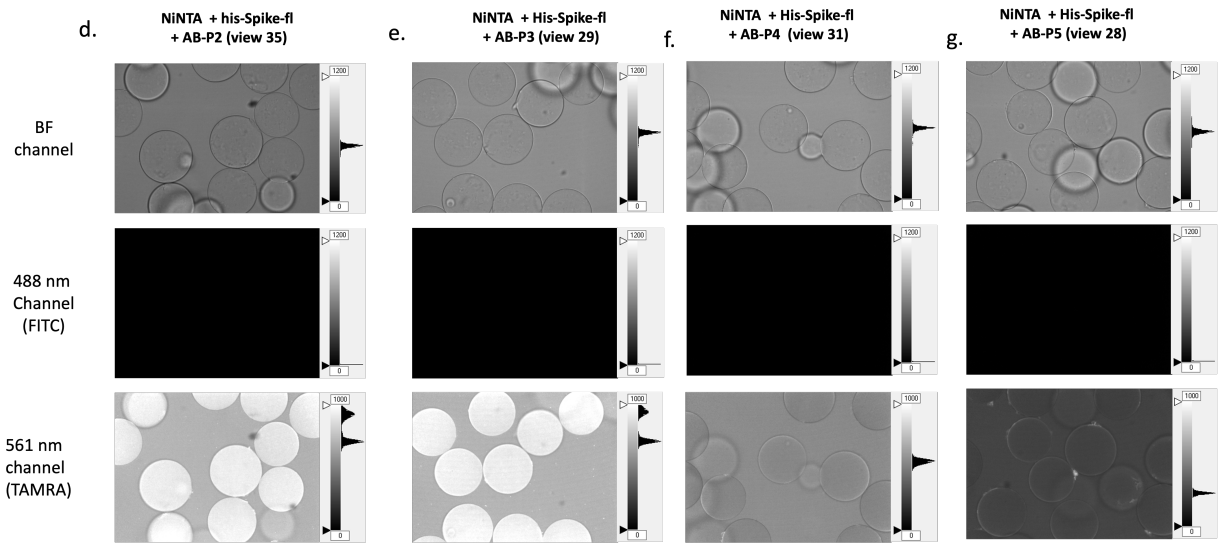



**Figure 5.32:** CONA assay of d. AB-P3 peptide; e. AB-P4 peptide, f. AB-P5 peptide g. AB-P6 peptide. The assays were performed by Dr Nhan Pham, University

with the beads. Thus the observed fluorescence results from an aggregation of the peptide at the surface of the beads, and it is unlikely that the peptides have an affinity with the spike protein.

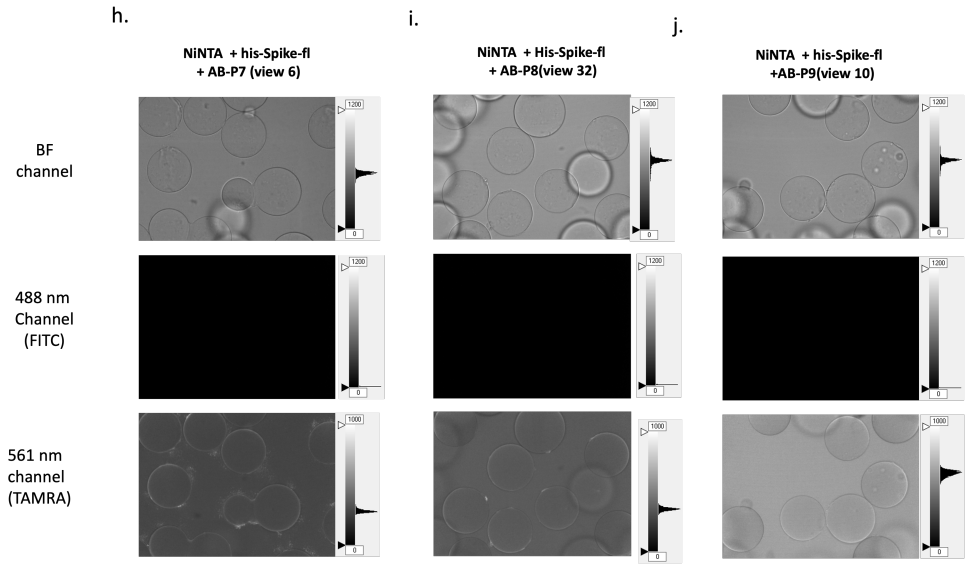



**Figure 5.33:** CONA assay of h. AB-P7 peptide ; i. AB-P8 peptide, j. AB-P9 The assays were performed by Dr Nhan Pham, University

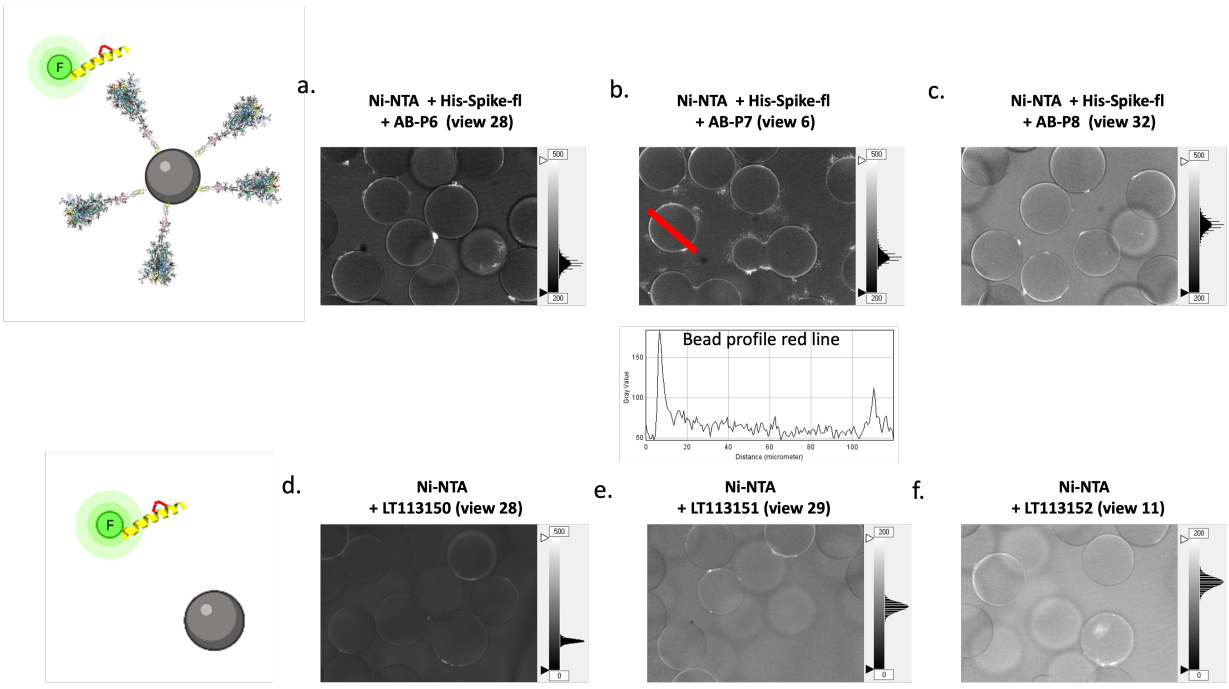



**Figure 5.34:** a. CONA assays for a. and d. AB-P6, b. and e. AB-P8, c. and f. AB-P7. a. b. c. assays in presence of Spike protein ; d. e. f. Negative control without spike protein ; The assays were performed by Dr Nhan Pham, University of Edinburgh.

### Peptides designed to bind the HR2 region

The HR2-P1 and HR2-P2 peptides were tested using the CONA assay. The HR2-P1 peptide was not found to have any affinity for the spike protein as no fluorescence was observed. In contrast, a bright ring was observed for the HR2-P2 peptide. The ring was found to be continuous, and the intensity profile of was found to be closer to the expected intensity of a binder. However, during the CONA assay, some peptide aggregation was observed in the background (Figure 5.35). The peptides were designed to form a hexameric coiled-coil, so it is a viable hypothesis that the peptides were assembling into coiled-coil in the assay without binding to the spike protein. A negative control was performed, using a Ni-NTA bead carrying no spike protein. Some intensity was observed in the negative control, however the intensity was lower than the intensity observed in the assay using the spike protein bound to the beads. While a weak binding of the peptide HR2-P2 could not totally be ruled out, it was acknowledged that most of the fluorescence observed was caused by aggregation of the peptide (and possible formation of coil systems) and non-specific affinity of the peptide for the agarose Ni-NTA beads.

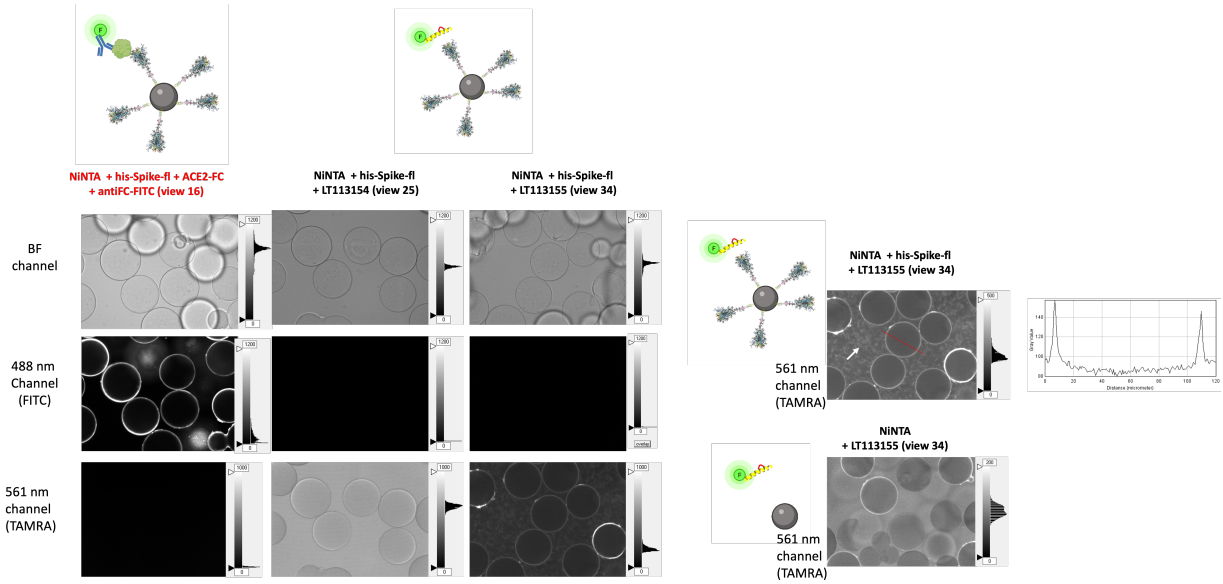



**Figure 5.35:** CONA assay using the HR2-P1 and HR2-P2 peptides. The assays were performed by Dr Nhan Pham, University of Edinburgh.

## 5.8 Conclusion

Fluorescent stapled-peptides have the potential to offer a complementary tool for early detection of COVID-19, with multiple applications to be developed. We conducted a study of an all-atom molecular dynamics simulation of the glycosylated spike protein. As sugars act as a protective shield against antigens, the analysis of these preliminary simulations permitted us to focus on “weak spots” in this shield to design our peptides.

We developed innovative approaches for the design of stapled peptides that were orthogonal to other approaches as part of the worldwide efforts of the scientific community led worldwide in the fight against COVID-19. Force-fields for the peptide staple restraints were used successfully to design a series of hydrocarbon-constrained fluorescent stapled peptides. Three series of peptides were modelled (ACE-2 based, antibody-based and stalk-based) and fluorophore and staple positions optimised computationally. Our aim was the development of a library of fluorescent, stapled peptides that would bind selectively to the SARS-CoV-2 virus spike protein to enable detection of the virus in a range of biomedical and clinical settings. To achieve this, we maximised our use of pre-publication data on the spike protein, exchanged data with colleagues across the world, and used protein samples from consortia such as the COVID-19 Protein Production Consortium (CPPC).

Our initial approach to stapled peptide design was based on reports of medium-length linear peptides derived from ACE2, which demonstrated micromolar affinity binding to the SARS-CoV-2 RBD. Stapled helices based on this sequence were computed using our in-house approach, which allowed point mutations in the sequence and variations in the staple and fluorophore position to be screened *in-silico*. In a second approach, we built a structural database of antibodies known to bind to the spike protein; we aligned the spike-binding region for each crystal structure and then extracted the antibody-binding side-chains. These side-chains were incorporated into helical peptides, scanned to find those that reproduced the key interactions. In the third approach to stapled peptide design, we exploited an exposed surface on the HR2 stalk domain of the spike protein, which our initial simulations had revealed. We designed peptides that would bind this trimeric coil to form a hexamer, similar to that observed in post-fusion structures of the spike protein. For each approach, stapled peptides were scored using MD simulations of binding to the glycosylated spike protein model, and their relative rankings were used to determine which peptides were synthesised.

A total of 14 stapled peptides were synthesised commercially (Lifetein) based upon our computational designs, and these were screened against the full-length SARS-CoV-2 spike protein using the on-bead CONA assay developed in the Auer group laboratories. Antibody-based and ACE2-based peptides showed either no discernible binding ($K_D$ > 500 mM), an unusually high affinity for the agarose resin, or aggregation in solution with precipitation onto the surface of the beads to give a false positive readout. Results recently published by the Jamieson group (Glasgow University) also found a series of ACE-2 based peptides did not bind to the spike protein.[302]

One stapled peptide that appeared to aggregate in solution nonetheless gave a positive CONA assay result and could be adapted for further studies. Since the CONA assay proved robust, sensitive and versatile, this would allow parameters such as peptide dilution, buffer pH, additives etc., to be optimised to circumvent the aggregation issue. This would allow to determine whether we have indeed generated a spike-protein specific, fluorescent, stapled peptide probe that might enable detection of the SARS-CoV-2 virus in a range of biomedical

and clinical settings.

## 5.9 Methods

### 5.9.1 Full length spike protein simulations

Simulations were ran on different hardware. Computing time were allocated by amazon to run on AWS, compute time was also allocated by DiRAC and CCP/HECBioSim for this project.

#### Protein models

Two different models of the spike protein were used.[382, 415] The first model was a 'closed' model, containing the three RBD in the down positions, the model of the protein atoms was used directly from the model. The second model was an 'open' model, with one of the sub-unit in up position, the model was subsequently modified: The spike protein Cryo-EM structures did not resolve the 452-492 region in the open (6VYB) or the closed state (6VXX). These regions were not included in the PDBs in any of the chains. The conformation of the loop 452-492 is in a coil in this model. However, in many crystal structures of the RBD region (6M0J, 6VW1, 7B3O...), this loop is a $\beta$-sheet. In the Figure 5.36 the Amaro model is in orange and the RBD in complex with ACE2 6M17 is shown in purple. The Figure depicts the ACE binding interface seen from top.

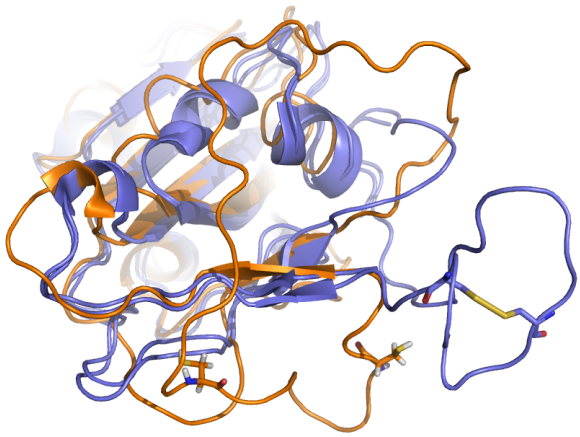


**Figure 5.36:** Alignment of the crystal structure 6M0J with thePDB ID 6M17

Furthermore, in the electron density map of the cryo-EM structure used to produce that initial model, with little to none cryo-Em electron density in this loop, as seen in figure 5.37. Thus the crystal structure 6M0J was 'grafted' into the open model. First a 6M0J refined structure[383] was aligned with the RBD sub-unit in the model. Residues Cys336 and Cys525, were chosen to link both structure as these residues carbon alpha had a RMSD below 2 Å.

**Figure 5.37:** CryoEM Electron density of the structure (orange), overlayed with the electron density

Both cysteine residues forms dissulfides bonds; residues Cys336 forms a dissulfide bond with Cys361 while Cys525 forms a dis-sulfide bond with Cys391.

### Glycolsylation of the spike protein

The spike protein model was provided using the CHARMM forcefield and offered the sugars already built and parameterised. However, using the amber forcefield in our simulation, the GLYCAM forcefield is the glycosylation forcefield the most compatible with the protein forcefield. Sugar monomers are following a different nomenclature in these forcefields; the sugar were rebuilt from scratch. A consensus in the sugar composition was found between Grant *et al.* [336] and Watanabe *et al.* [337] MS sugar characterisation and is been reported in Table 5.12

The sugar composition used in our simulations does not differ much from the one found in Amaro's group model. The sugars were built using the doglycan tool, and the protein was parameterised using AMBER12, GLYCAMFF, and TIP3P was used as a water model. Simulations were run using the GPU build of GROMACS2020

The post-translational glycosylation modification was incorporated into the model.

### Simulations

All simulations were performed with GROMACS16. The AMBER16 forcefield and the GLYCAM forcefield were used to parametrise the full-length spike glycoprotein models. A first non-solvated steepest descent method was used to minimise the system. The system was then solvated with TIP3P water molecules (241436 water molecules) and neutralised with 0.3 M NaCl salt.(1402 Na+, 1366 Cl-). GROMACS16 was used to conduct the simulations.[416] Stepwise minimisation and equilibration were conducted using barostat with a time constant

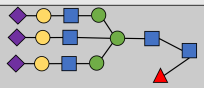
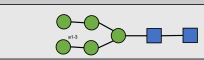
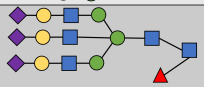
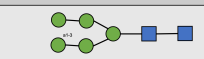
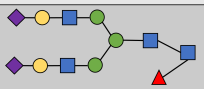
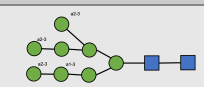
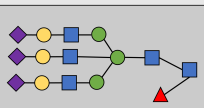
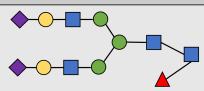
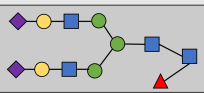
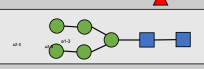
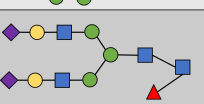
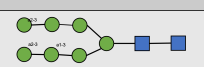
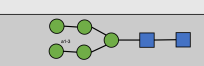
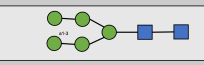
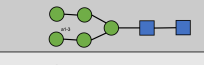
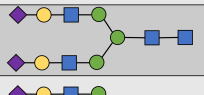
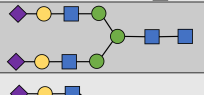
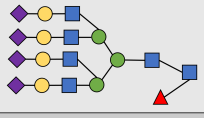
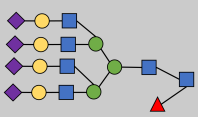
| position | Sugar | Sugar structure for the model |
|---|---|---|
| N17 | FA3 |  |
| N61 | M5 |  |
| N74 | FA3 |  |
| N122 | M5 |  |
| N149 | FA2 |  |
| N234 | M8 |  |
| N282 | FA3 |  |
| N331 | FA2 |  |
| N343 | FA2 |  |
| N603 | M5 |  |
| N616 | FA2 |  |
| N657 | M7 |  |
| N709 | M5 |  |
| N717 | M5 |  |
| N801 | M5 |  |
| N1074 | M5 |  |
| N1098 | A2 |  |
| N1134 | FA2 |  |
| N1158 | A2 |  |
| N1173 | FA4 |  |
| N1194 | FA4 |  |

**Table 5.12:** Sugar composition used in the spike protein model

of 5 ps. A real-space cut-off of 1.2 nm was utilised to calculate the electrostatic interactions, while the Van der Waals interactions were truncated at 1.2 nm and the force switch smoothing function applied between 1.0 to 1.2 nm. During the equilibration, the temperature of the system was gradually increased to 290 K over 500 ps, constrained all covalent bonds with hydrogen atoms, and an integration time step of 1 fs was used. The integration step was then raised to 2 fs for a further 200 ps equilibration. For both equilibration stages, the LINCS algorithm was used to maintain the temperature. For the production stage, the smooth particle mesh Ewald (PME) method and V-rescale thermostat with a time constant of 1 ps A 2-fs integration time step was employed during this run. After equilibration, 200 ns production while the Parrinello-Rahman barostat with a time constant of 5 ps was used to maintain the pressure. Additional 2000 $J/nm^2$ harmonic restraint were applied on the alpha carbon of the 10 residues of the TM.

### 5.9.2 Trajectory analysis

Protein and sugar structural properties were analysed using GROMACS, VMD and MDtraj tools. First the trajectories were aligned and water molecules were removed. using the GROMACS trjconv. The trajectories were visualised using VMD. Solvent-accessible surface area (SASA) was calculated using VMD with a probe radius of 1.4 Å, in 5 ns snapshot intervals. Multiple regions were chosen for SASA calculation: RBM (residues 440–508), NTD (residues 13–310) and HR2 (residues 1164 to 1213) As well as the full S1 domain (residues 1 to 685) and S2 (residues 685 to 1220)

The angle between the head and the tail of the spike protein was determined as the angle between the center of mass of the 'head' (residues 1 to 1050), the center of mass of the residues at the base of the head residues 1090 to 1120) and the center of mass of residues at the extremity of the modelled TM region (1010 to 1020), and was computed using MDtraj.

### 5.9.3 MD simulation setup stapled peptides

The use of unnatural amino acids and DYE prevented the use of multiple tools aimed at peptide/protein modelling, which could not cope with unnatural amino acids. For example scwrl4 does not support uAAs and removes unknown residues from the initial structure (other programs fail). In-house scripts to parse PBD and relying on scwrl4 and tleap were used.

These scripts automated our workflow and can be found on github.

#### ACE2-derived peptides

The glycosylation site 234 was modelled with (see Table 5.12). The PDB structure 6MOJ was used for the simulations.

### Ab-like mimics peptides

The glycosylation site 234 was modelled (see Table 5.12). Different RBD models were used (6YZ7, 7BYR, 6Z2M, 6ZCZ, 7C8W, 7C8W, 7C01). The models were chosen upon the structure of which the side-chains were extracted from. All models were truncated from residues 335 and 525 so they had the same number of residues.

### HR2 targeting peptides

Models were built using building as follows : Radii of both the inner and outer trimers were optimised: the radius distance for the inner trimer was initialised to 6.2 Å and left to vary at $\pm$ 3 Å during optimisation and the outer trimer was initialised at 13 Å and left to vary at $\pm$ 6 Å. The zshift of the outer trimer was left to vary at 4.1 Å. The pitch was optimised between $230 \pm 100$.

The glycosylation sites 1175 and 1194 were modelled (see Table 5.12). During the simulation of the HR2 domain restrains of 20000 $kcal.nm^{-1}$ were imposed on the 10 first C$\alpha$ atoms of the C-terminus (residues 1130-1140) and on the 10 last C$\alpha$ atoms of the N-terminus (residues 1450-1460)

### Simulations

*In-silico* screening of fluorescent stapled peptides required us to combine four different forcefields: the general Amber forcefield (AMBER14SB);[417] the GLYCAM06 sugar forcefield;[418] the AMBER-DYES forcefield;[251] and our in-house stapled-peptide residues forcefield (Chapter 3). The AMBER forcefields files were subsequently modified, and some residue names were modified to avoid redundant residues names.

The models were solvated using a TIP3P water model with a 10 Å around the solute. Counterions were added to the solvent to keep the system neutral. GROMACS20 was used to carry out the simulations. The geometry of the system was minimised in two steps before MD simulation was ran. First, the water molecules were refined through 2500 steps of steepest descent followed by 2500 steps of the conjugate gradient, keeping the protein fixed with a constraint of 2.0 $kcal.mol^1.Å^{-2}$. The complexes were then relaxed by 5000 cycles of conjugate gradient minimization. During the simulation, the particle mesh Ewald method was employed to calculate the long-range electrostatic interactions, while the SHAKE method was applied to constrain all covalent bonds involving hydrogen atoms to allow the time step of 2 fs. A 10 Å cutoff value was used for the non-bonded interactions. The whole system was then heated from 0 K to 300 K running 50 ps molecular dynamics simulation with position restraints at constant volume. An isothermal isobaric ensemble (NPT) simulation was performed for 500 ps to adjust the solvent density followed by 500 ps of constant pressure equilibration at 300 K without constraints to relax the system. The production dynamics at constant pressure

achieved lengths of 200 ns of which snapshots saved at 10 ps intervals were used for further analysis.

### 5.9.4 Trajectory analysis

The RMSD of the backbone, the RMSD of the side-chains the RMSF of the peptide C$\alpha$ were calculated using MDAnalysis.[419] The helicity was calculated using the DSSP module of MDtraj[420] by counting the helical assignment of the residues of the peptide and averaging by the number of residues (removing one residues at the C-terminus and N-terminus). The number of number of hydrophobic interactions was defined as the carbon-carbon distance below 4 Å and limited to one between each receptor-peptide pair. The number of hydrogen bonds was calculated with MDtraj. The number of conserved interactions was calculated as interactions (either polar either hydrophobic) conserved for 80 % of 20 snapshot spaced by 50 ps.

PCA was achieved using Scikit-learn software,[421] using every descriptors calculated coordinates as input dimensions.

# 6

# Development of Stapled Peptides Binding IL-1R

## 6.1 Cytokines and Chronic inflammatory response

Inflammatory responses serve to protect the host from infections after an injury occurs. Inflammation is characterised in the acute phase by an increased blood flow and vascular permeability along with the accumulation of fluid, leukocytes, and expression of inflammatory mediators such as cytokines.[422] Chronic or excess inflammation, lasts longer than the acute phase (months or sometimes years), and can lead to tissue dysfunction, in turn contributing to a multitude of chronic diseases such as cancer, dementia, fibromyalgia, diabetes, heart disease, stroke, liver problems, autoimmune disease such as thyroid disorder, multiple sclerosis, and rheumatoid arthritis.[423]

The acute and chronic phases of the inflammatory response are characterised by the development of specific cellular immune messengers. Many soluble mediators regulate the activation of cells in the injured tissue (fibroblasts, endothelial cells, tissue macrophages, and mast cells) and recruit new specialised cells linked with the inflammatory response (neutrophils, monocytes, lymphocytes and eosinophils). The systemic symptoms of the inflammatory process (hypotension, fever, leukocytosis, cachexia) are the result of these mediators. Most cytokines are multifunctional and are involved in extensive networks involving both synergistic and antagonistic interactions and exhibiting both negative and positive regulatory effects on various target cells. Their effects can be elicited locally or systemically in an autocrine or paracrine manner.[423]

IL-1, TNF-$\alpha$, GM-CSF, IL-6, IL-11, IL-17 and other cytokines are among the proteins involved in the inflammatory response in both the acute and the chronic phase of inflammation.[423] Figure 6.1 summarises the principal cytokines implicated in the inflammatory response, as well as the intracellular pathways. Some of these proteins are of interest for the treatment of numerous diseases related to an over-regulation of the inflammatory system. In particular TNF-$\alpha$, and IL-1, have elicited interest as they are central and extremely potent inflammatory mediators.[424, 425] Indeed, they are central cytokines mediating acute inflammation, and also two of the primary mediators of septic shock.[426, 427] In this work,

peptides were designed to inhibit the IL-1 pathway by binding to the IL-1 receptor.

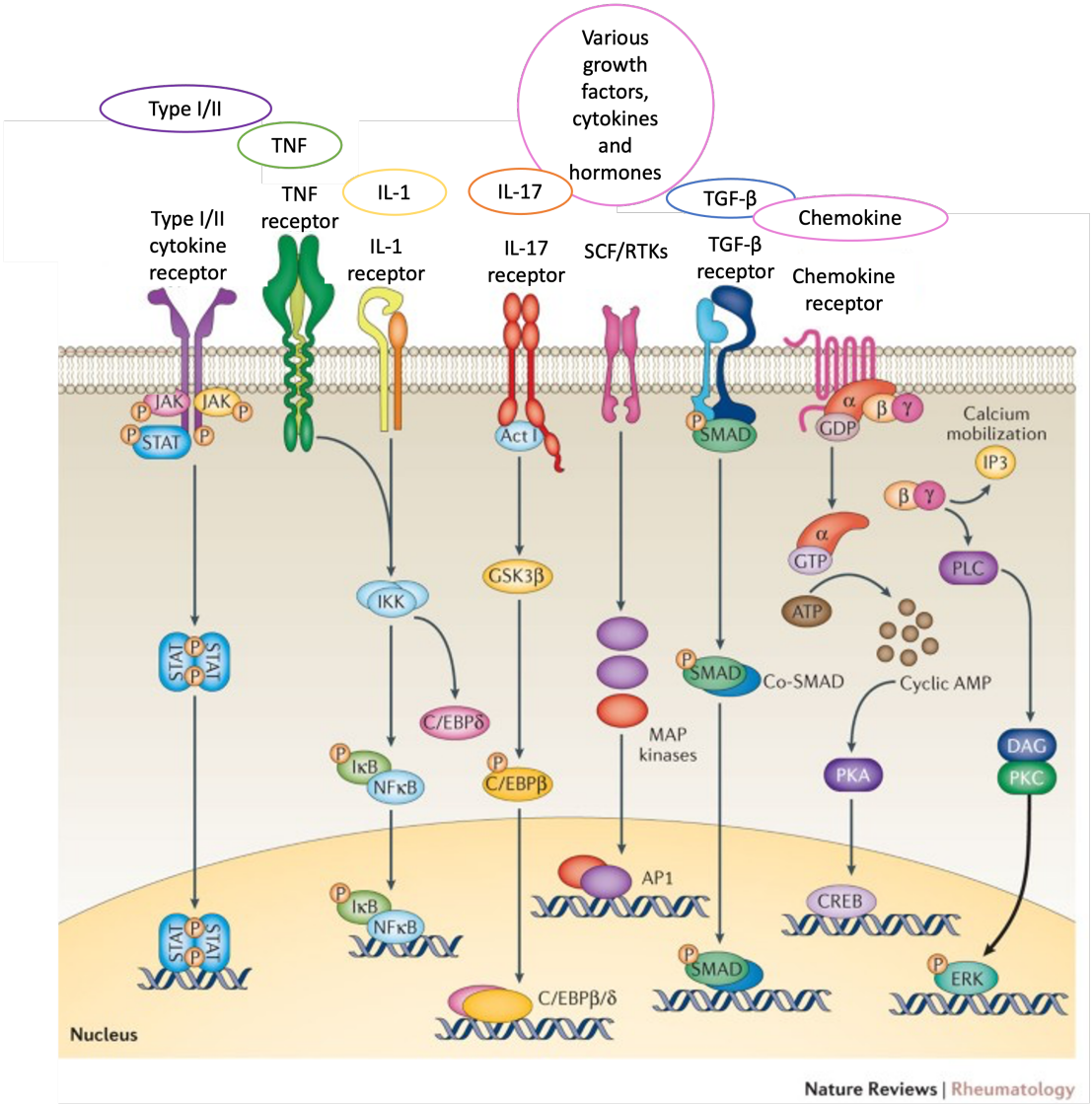



**Figure 6.1:** Principal cytokines implicated in the inflammatory and cellular responses. The major cytokine families are represented: the type I/II cytokines, the Tumour Necrosis Factor family, the IL-1 family, the IL-17 cytokines, various growth factors/ cytokines/ hormones binding to the stem cell factor/receptor tyrosine kinase (STF/RTK) receptors, the transforming growth factor (TGF)-$\beta$ family cytokines, and chemokines. When a cytosine binds to its cognate receptor, the receptor becomes activated, recruiting different pathways, leading the transcription of new proteins in the nucleus. Figure reproduced and modified with permission from Springer Nature, license number 5173020625236. Original authors: Schwartz *et al*[428].

### 6.1.1 IL-1R1 implications in the inflammatory response

IL-1R is a transmembrane protein which is part of the IL-1 pathway. It is a cytokine receptor which binds interleukin 1 $\alpha$ and interleukin 1 $\beta$ (IL-1$\alpha$ and IL-1$\beta$). This binding, in turn activates the Toll pathway.[422, 429] At least ten forms of the IL-1R1 receptor have been discovered and are numbered by their chronological characterisation. In this section, only the

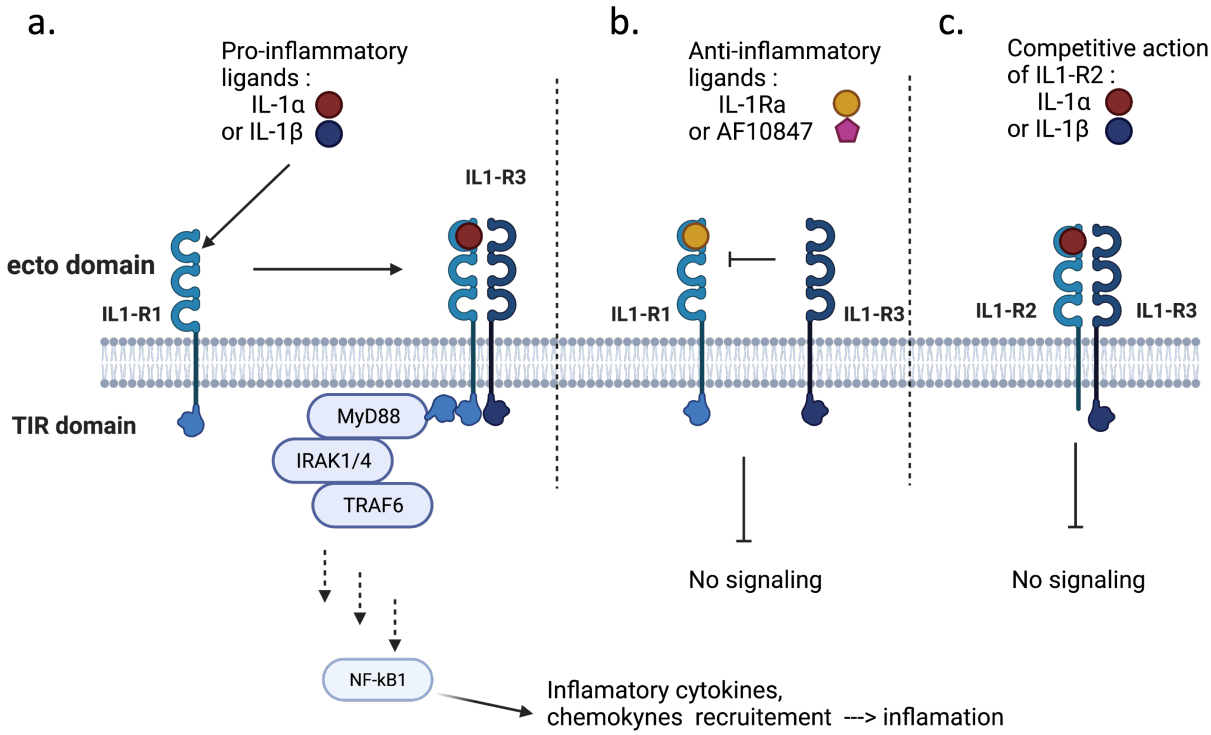



**Figure 6.2:** IL-1R1 activation and inhibition pathways. a. Action of agonist ligands of IL-1R: When binding with IL-1R, dimerisation of the receptor occurs, leading to the binding of the MyD88 TIR domain with the TIR domain of IL-1R1 and to the downstream activation of the NF-κB pathway. b. When an antagonist binds to IL-1R1 dimerisation with IL-1R3 does not occur and no signal is transduced. c. When IL-1$\alpha$ or $\beta$ binds to IL-1R2, no signal is transduced as IL-1R2 does not contain a TIR domain.

role of type 1, 2 and 3 receptors will be discussed as the other types are not implicated in IL-1$\alpha$ and IL-1$\beta$ signal transduction.[430] The type I receptor (IL-1R1, or IL-1R) is a pro-inflammatory receptor and binds to interleukin-1 $\alpha$ and $\beta$ (IL-1$\alpha$ and IL-1$\beta$). The functional domain of the cytosolic component of IL-1R1 is termed the Toll interleukin-1 receptor (TIR) domain and is involved in mediating interactions with cytosolic proteins. TIR domains are also present in cytosolic adaptor molecules acting downstream of these receptors and recognition of TIR domains leads to signal transduction. The activation of IL-1R1, upon binding with IL-1$\alpha$ (or $\beta$), leads to the stabilisation of a heterodimeric form of the receptor with IL-1R3. IL-1R3 is an accessory protein and is also called interleukin-1 receptor accessory protein (IL-1-RAP). IL-1R3 contains a TIR domain but does not bind to IL-1$\alpha$ (or $\beta$). Only when this dimer is formed will the signal be transduced (Figure 6.2.a).

The type 2 receptor (IL-1R2) acts by tempering the inflammatory response by competing for IL-1 $\beta$ binding (Figure 6.2.c). The type 2 receptor closely resembles the type 1 receptor in the d1/d2/d3 domains but does not contain a TIR domain. Consequently, the signal is abrogated upon binding with IL-1-$\alpha$. The IL-1 receptor antagonist (IL-1Ra) is a known antagonist

which selectively binds to IL-1R1, and inhibits the formation of the dimer with IL-1R3 (Figure 6.2.b). IL-1R1 binds IL-1$\alpha$, IL-1$\beta$ and IL-1Ra. IL-1R1 has a higher affinity for IL-1$\beta$ ($K_{dissociation}$ = 2.08 nM) than for IL-1$\alpha$ ($K_{dissociation}$= 28.5 nM) or IL-1Ra ($K_{dissociation}$ = 41.6 nM) [431]

While naturally occurring receptor antagonists (such as IL-1Ra) work well at abrogating IL-1 signaling, an early goal of the field was to discover lower molecular weight antagonists to IL-1R1 that could be delivered orally.

### 6.1.2 IL-1R1 structural characteristics and interactions

Structurally, IL-1R1 is composed of an ectodomain which extends in the extracellular space, a trans-membrane domain and an intracellular domain. The ectodomain of IL-1R1 is composed of 3 subunits, which are referred to as d1, d2 and d3. Several crystal structures of the IL-1R1 ectodomain bound to different partners have been published in the PDB data bank. These structures describe the ectodomain only.

- in crystal structure 1ITB, the type-1 interleukin receptor is complexed with interleukin $\beta$.
- in crystal structure 4DEP, the type-1 interleukin receptor forms the signaling complex by a dimerisation with IL-1-R3, after binding to IL-1$\beta$.
- in crystal structure 1IRA, the type-1 interleukin receptor is complexed with IL-1Ra.
- in crystal structure 1G0Y, the type-1 interleukin receptor is complexed with the peptide inhibitor AF10847.
- in crystal structure 4GAF, EBI-005, a chimera of human IL-1$\beta$ and IL-1Ra forms a complex with the type-1 interleukin receptor.

In the IL-1/IL-1R1 complex, IL-1R1 adopts a conformation where the d1 and d2 subunits are tightly packed against each other and, together, form the main binding interface (PDB 1ITB and Figure 6.3.a). A flexible 6 amino acid linker connects d3 to d2, giving d3 some mobility to fold towards d1 and d2, thus forming a 'closed' conformation of the receptor. The crystal structure also shows d3 in contact with IL-1$\beta$ forming a second interface (Figure 6.3.a). IL-1$\beta$ and IL-1R1 share a contact area of 1755 Å over 47 residues. In the larger interface with the d1/d2 domain, IL-1$\beta$ has a contact surface area of approximately 900 Å$^2$ over 24 amino acids with the d1 domain and the d2 domain. In the d3 domain of the receptor, 16 amino acids contribute to the binding. These PPIs are highlighted in Figure 6.4.a.

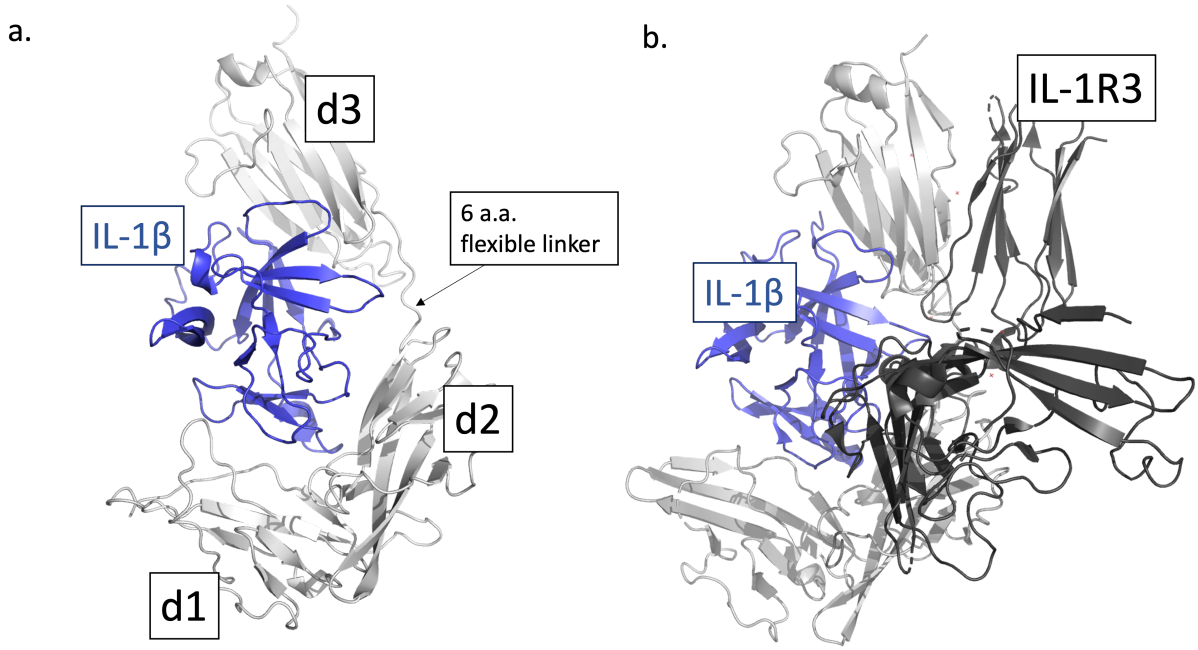



**Figure 6.3:** Binding modes of the ectodomain of IL-1R1 with a. IL-1$\beta$ (Crystal structure 1ITB) and b. IL-1$\beta$ and IL-1R3 (Crystal structure 4DEP)

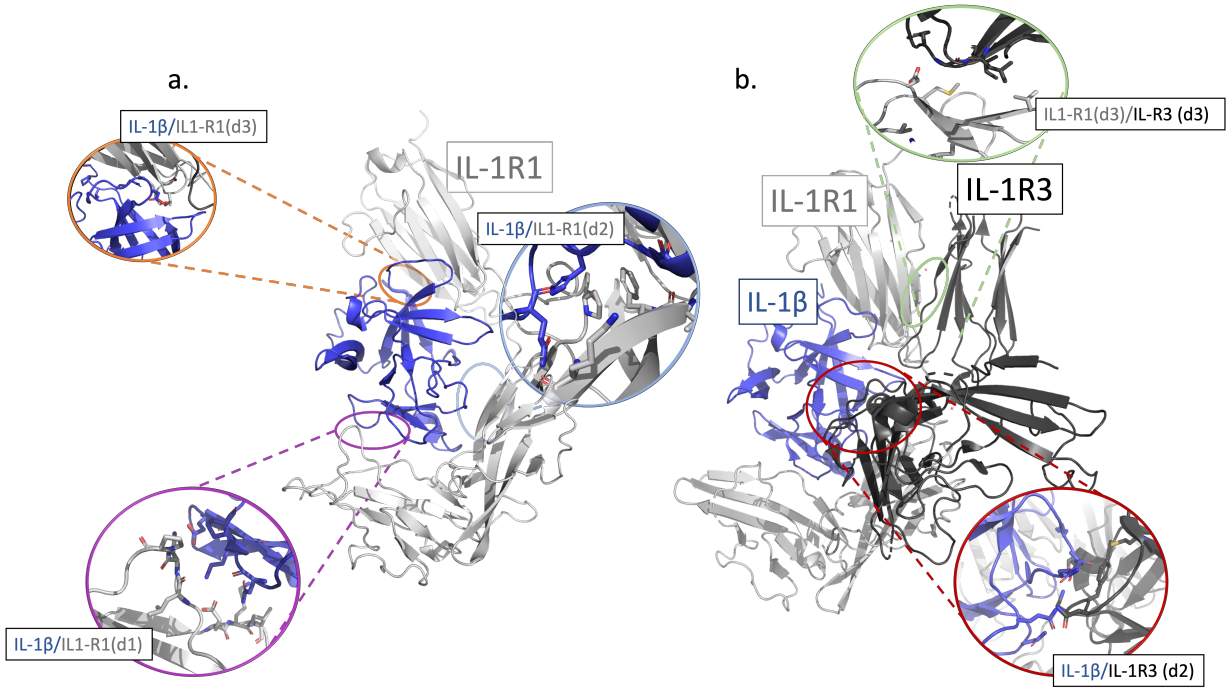



**Figure 6.4:** PPIs between IL-1R1, IL-1$\beta$ and IL-1R3 (Crystal structure 4DEP and 1ITB) a. The 3 PPI between IL-1$\beta$ and IL-1R(1) b. The PPI between IL-1R3 and IL-1R1 and the PPI between IL-1R3 and IL-1$\beta$

A crystal structure of the dimeric receptor complex (4DEP) shows two additional interactions: one between the d3 domains of both receptors, and a second one between the IL-1R3 d2 subunit and IL-1$\beta$ (Figures 6.3.b and 6.4.b). The formation of the binary receptor complex is a key step in the initiation of a functioning signaling complex.[432, 433] The binary

complex is believed to provide a scaffold for the recruitment of cytosolic adaptor proteins through TIR-domain interactions inside the cell, with dimerisation of the TIR domain of the receptor to the TIR domain of the adaptor molecules.[434]

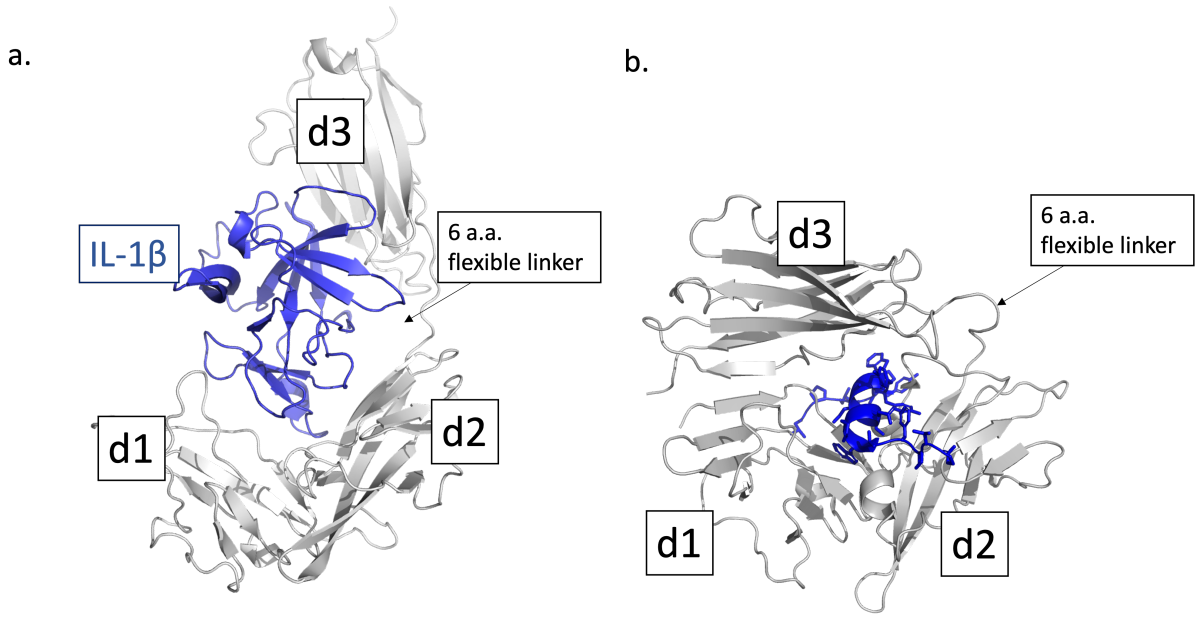



**Figure 6.5:** a. Conformational changes in the d3 domains of IL-1R upon binding with a. IL-1$\beta$ vs b. the AF10847 21-mer peptide. (Crystal structures 1ITB (a.) and 1G0Y (b.)). In the peptide bound structure, the d3 subunit 'clamps' the peptide

Besides IL-1$\beta$, IL-1$\alpha$ and IL-1Ra, the IL-1R1 receptor has been crystallised with a peptide inhibitor, AF10847. The crystal structure, shown in Figure 6.3.b., reveals a large conformational change in the d3 domain of IL-1R. The d3 domain swings by 160° away from the position it would occupy when binding IL-1$\beta$, and instead wraps around the helical motif of the peptide entirely. Importantly, this change of conformation makes the dimerisation of IL-1R1 with IL-1R3 impossible, hence stopping the signal transduction. The surface area shared by the peptide and IL-1R1 is 1364.0 Å, as calculated with PISA.[313] Comparing the structures of IL-1R1 bound to Il-1$\beta$ and bound to AF10847 (Figure 6.3) clearly demonstrates the flexibility of the d3 domain with respect to the d1/d2 domains.

To summarise, the basic mechanism of signal initiation is a stepwise process. (see Figure 6.2) Firstly an agonist cytokine (IL-1 $\alpha$ or $\beta$) binds its cognate receptor. Secondly, this cytokine-receptor complex recruits a secondary receptor forming a dimeric receptor. Finally, the Toll/IL-1 Receptor (TIR) domains of the two receptors are brought into close proximity, initiating an NF-$\kappa$B signal transduction cascade within the cell. Inhibitors, such as IL-1Ra and AF100847, bind to IL-1R1 and block the dimerisation of the receptor and further signal transduction.

## 6.2 Peptide design Strategy

In this work, the aim was to design stapled peptides, based on the peptide AF10847, for binding to IL-1R. This might offer a better bio-availability than the lead peptide, together with a higher - or at least similar - binding affinity.

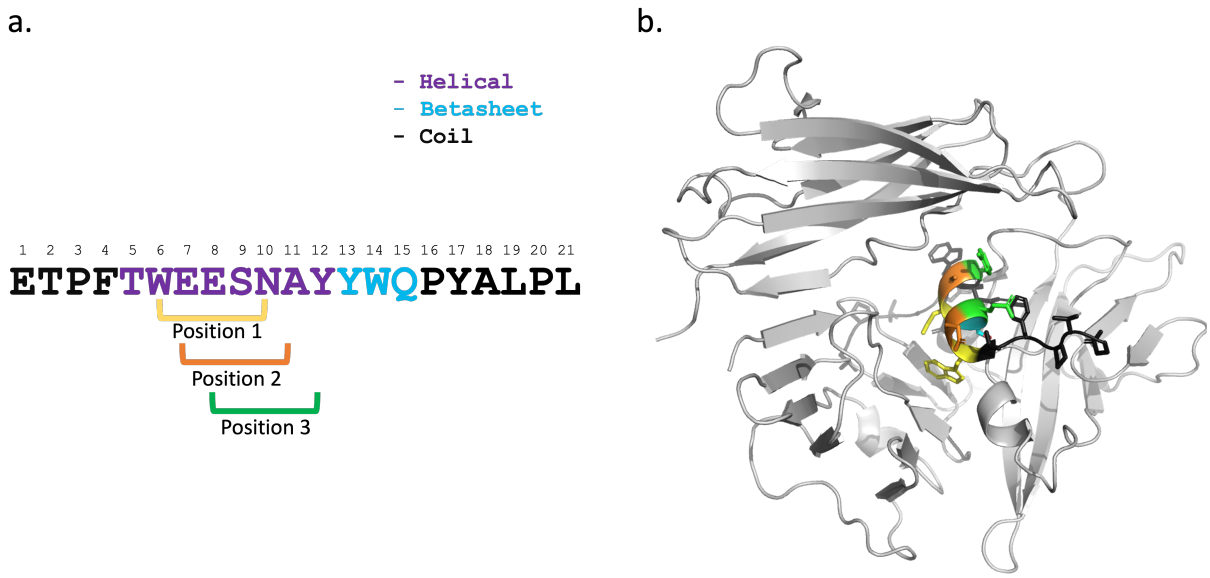



**Figure 6.6:** Sequence of the 21-mer peptide AF10847 (Crystal structure 1G0Y) and staple positions considered a. staple positions on the 21-mer sequence, b. staple positions in context of the crystal structure (yellow position 1, orange position 2 and green position 3).

In the crystal structure 1G0Y (IL-1R1 complexed with AF10847), residues 6 to 13 at the centre of the peptide form the only helical segment (Figure 6.6). At the N-terminus, six amino acids form a coiled conformation with no strong interaction to the receptor. Only a weak interaction with Phe4 is observable. The backbone between residues 13 and 15 of the peptide forms an anti-parallel $\beta$ sheet with the receptor, forming a hydrogen bond between the two backbones. Generally backbone interactions provide more stability as they bind more rigid parts of the protein. The residue Tyr13 of the peptide also forms an hydrophobic interaction, fitting in a small pocket of the receptor. These observations indicate that this segment and structure is important for the binding. At the C-terminus, six amino-acids form a coil, interacting with the protein.

One of the strategies for the placement of the staple was to interfere the least with the binding conformation of the 21-mer peptide AF10847. The helical segment being only 7 amino acids long drastically restricts the number of possibilities for the placement of a staple. Considering that staple positions enhancing helical content are found with a i, i+4 spacing or i, i+7 spacing, choices were limited. Staples geometries requiring a i, i+7 spacing were eliminated, as the helical segment of the peptide is too short to accommodate this staple geometry.

Furthermore, inserting a staple between Tyr5 and Ser9 was found to introduce large clashes with IL-1R, and no viable model could be produced. Only three positions were therefore considered, and these are highlighted in Figure 6.6. The first stapling position is between residues Trp6 and Asn10, the second between residues Glu7 and Ala11 and the third between residues Glu8 and Tyr12. All three positions have strengths and weaknesses. For conservation of the initial charge, the staple in position 1 yields the same net charge as the native peptide, -3, as no charged residue is removed. The isoelectric point is also unchanged at 0.68. A staple in position 2 or 3 removes a negative residue (Glutamine in both cases), reducing the net charge of the peptide to -2. The isoelectric point would be shifted to 0.76 for both positions.

Having in mind conservation of the initial binding position, introducing a staple at position 2 or 3 does not remove any strong interactions, unlike position 1, where the mutation of Trp5 might remove some interactions with d2. In terms of peptide conformation, a staple in position 3 might induce helicity in residues adjacent to Tyr12, which might disturb the $\beta$-sheet conformation and interaction of residues Tyr13 and Trp14 observed in the crystal structure of the 21-mer, which might be detrimental overall. Position 2 is the only position which does not remove any aromatic amino-acids, which generally induce helicity in the sequence. On the other hand, a staple should counteract the loss of helicity and some of the added staples are lipophilic, also balancing the loss of lipophilic residues. Thus from these preliminary observations and considerations none of these three positions presents a clear advantage or disadvantage over the others.

Three commonly used staple chemistries have been used for the staples in our studies: Grubbs ring closing metathesis using *S* and *R* pentenyl-alanine; cysteine cross-linked staples which refers to the formation of thioether bridges via a covalent linkage between Cys and $\alpha$-bromo amide side chains, using *para* and *meta* bromo methyl xylene; and 'click' peptides using alkyne-bearing side chain (proparglycine) and an azidolysine for macrocyclization by Cu(I)-catalyzed azide–alkyne cyclo-addition (CUACC). These staple chemistries were chosen because of the availability of the starting materials as well as the existence of robust procedures to synthesise them. The different sequences tested in the molecular dynamic models are listed in Table 6.1.

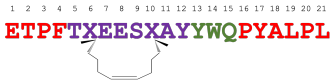
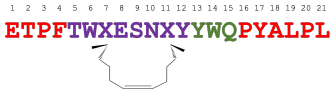
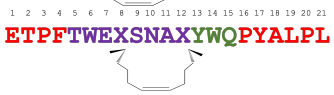
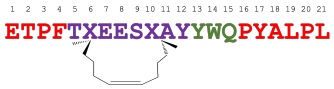
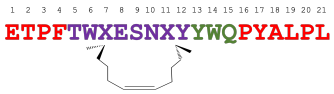
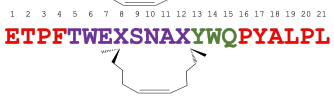
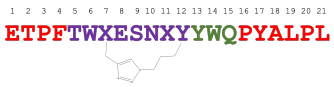
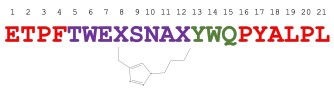
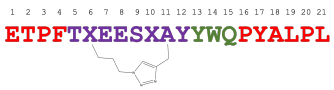
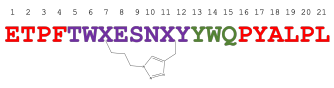
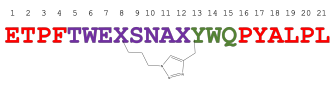
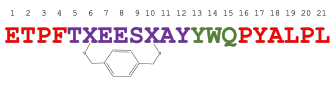
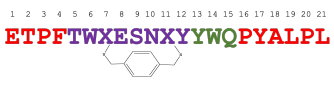
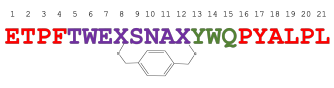
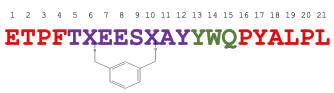
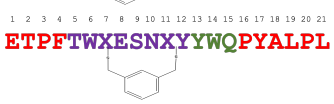
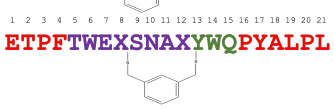
| Name | Sequence | Net Charge/ | Isoelectric Point | Ratio hydrophobic residues/ total residues |
|---|---|---|---|---|
| RCM-SS-1 |  | -3 | 0.68 | 42% |
| RCM-SS-2 |  | -2 | 0.76 | 47% |
| RCM-SS-3 |  | -2 | 0.76 | 42% |
| RCM-RS-1 |  | -3 | 0.68 | 42% |
| RCM-RS-2 |  | -2 | 0.76 | 47% |
| RCM-RS-3 |  | -2 | 0.76 | 42% |
| CLICK-1-1 |  | -3 | 0.68 | 42% |
| CLICK-1-2 |  | -2 | 0.76 | 47% |
| CLICK-1-3 |  | -2 | 0.76 | 42% |
| CLICK-2-1 |  | -3 | 0.68 | 42% |
| CLICK-2-2 |  | -2 | 0.76 | 47% |
| CLICK-2-3 |  | -2 | 0.76 | 42% |
| CYSM-1 |  | -3 | 0.68 | 38% |
| CYSM-2 |  | -2 | 0.76 | 42% |
| CYSM-3 |  | -2 | 0.76 | 38% |
| CYSP-1 |  | -3 | 0.68 | 38% |
| CYSP-2 |  | -2 | 0.76 | 42% |
| CYSP-3 |  | -2 | 0.76 | 38% |

**Table 6.1:** Peptide sequences modelled in this study, and their respective charges, isoelectric points and percentage of hydrophobic residues. Isoelectric points were calculated using the pepCalc server.[435]

## 6.3 Model validation through molecular dynamics

Peptides were simulated in complex with the IL-1R1 using two models. One model, which will be referred to as the 'full length model', comprises the d1/d2 and d3 subunits of IL-1R1 (Figure 6.7.a). A second model, which will be referred to as the 'truncated model', only contains the d1 and d2 subunits (Figure 6.7.b). Two considerations motivated the use of this truncated model where the d3 domain is removed. The first consideration was that for most of the simulation, the binding interactions were found to take place with the d1 and d2 domains in the crystal structure. The second consideration was that the d3 domain is known to display some conformational heterogeneity, thus it is not clear if d3 is always in a closed conformation when binding to the peptide. Furthermore the truncated model might actually be closer to the state of the receptor while binding to the peptide: in our hypothesis the binding of the peptide is a two-step process where the peptide first binds to the d1/d2 domain, followed by closure of the d3 domain.

Molecular dynamic simulations were used to confirm some of the preliminary considerations. Starting with the observation of simulations with the full length protein model in complex with the non-stapled peptide, only small conformational changes were observed in 3 consecutive runs of the molecular dynamic simulations. Some moderate backbone movements were observed in the receptor in the loop 22-28 and the loop 123-132, which form a 'gateway' in the cavity formed by the 3 three subunits. Some flexibility was also observed in the backbone atoms of the receptor in the 116 to 201 segment, which corresponds to the linker between d2 and d3, but no conformational change in the d3 domain was observed. The center of mass of d3 was not found to vary significantly within the 500 ns long simulations, and no conformational change was observed in this region during the 500 ns simulation time.

In term of interactions, the peptide Tyr13 maintained its interaction during simulations as expected. The N-terminus of the peptide was the region with the most flexibility and did not display strong interactions with the receptor. Only weak interactions between the peptide and the d3 domain were observed. When comparing the full model (d1/d2/d3) with the truncated model (d1/d2), the unstapled peptide adopted a very stable conformation in both models: only after a certain time (over 250 ns more) the N-terminal segment started showing more flexibility, while the helical segment of the peptide conserved its secondary structure and remained in a very stable binding position in the simulations.

Moving on to the simulations of stapled peptides in complex with the full length model, most of the designs did not lead to any drastic unfolding of IL-1R, meaning that generally the staple did not add any particular instabilities in the complexes. Also, binding conformations were conserved for almost all peptides, thus validating the hypothesis that the stapled peptides can interact with the receptor similarly to the initial unstapled peptide. Mutation of the residues to stapled residues were found to produce visually relatively stable simulations, especially in the central helical region of the peptide. In Figure 6.8.a and b the RCM peptide

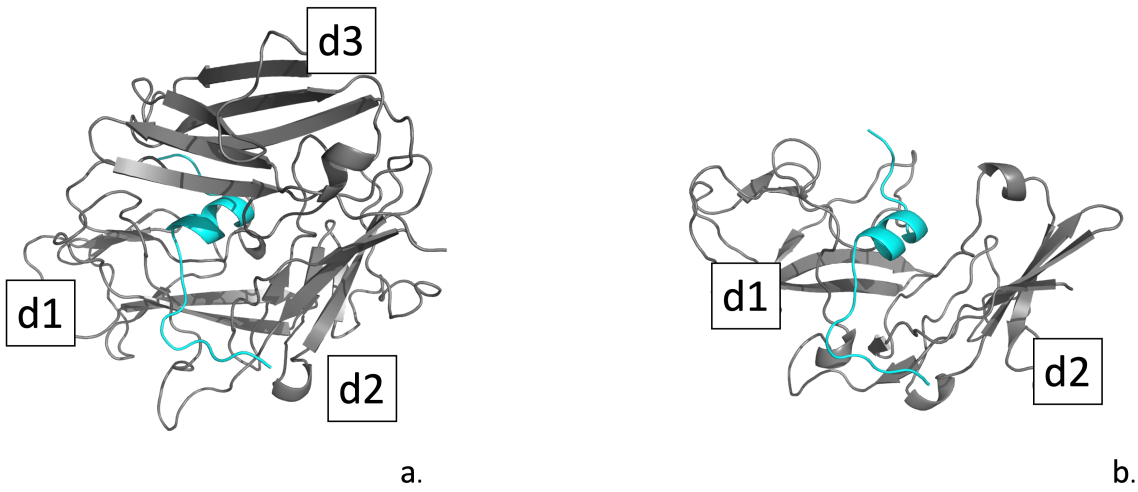



**Figure 6.7:** a. 'full length model ' (d1/d2/d3) ;b. 'truncated model ' (d1/d2) The peptide adopts the same binding pose in both models

is given as an example of a simulation for a staple with low movement observed after 250 ns.

In most cases the staples were not found to directly interact with the receptor. Prior to the simulations, Tyr6 was identified to contribute strongly to the binding, thus concerns were raised when adding the staple in position 1 (replacing the residue Tyr6). However, we observed that modification of this amino acid to a staple residue did not result in a weaker binding for all staples. For instance, simulations of the peptides with the ring closing metathesis staple in this position did not show any interaction of the staple with the receptor and conserved their binding mode. They were found to be relatively stable during the simulations. However, the peptide CLICK-2-1 containing a triazole ring, and the peptide CYS-P-1 with a xylene staple adopted different conformations which led the stapled residues to interact with the protein and the peptide to partially unfold (Figure 6.8.e and f). Using a meta xylene instead of para xylene, or inversion of the staple residues in the CLICK-1-1 peptide (azidolysine-6 proparglycine-10) did not introduce the same change of conformation. Stapled peptides carrying a *meta*-xylene staple and other stapled peptides with click chemistry did not display a loss of helicity in other designs (Figure 6.8.c and d). In positions 2 and 3 no specific interactions were found between the staple and the protein. Thus, interactions between the staple and the protein were mostly found to be detrimental to the binding of the peptides, as they changed the binding pose of the peptide. Undesirable variations in the peptide orientation were also observed. For example, the peptide CLICK-3-1 seemed to adopt a conformation slightly rotated from the initial binding pose (Figure 6.8.g).

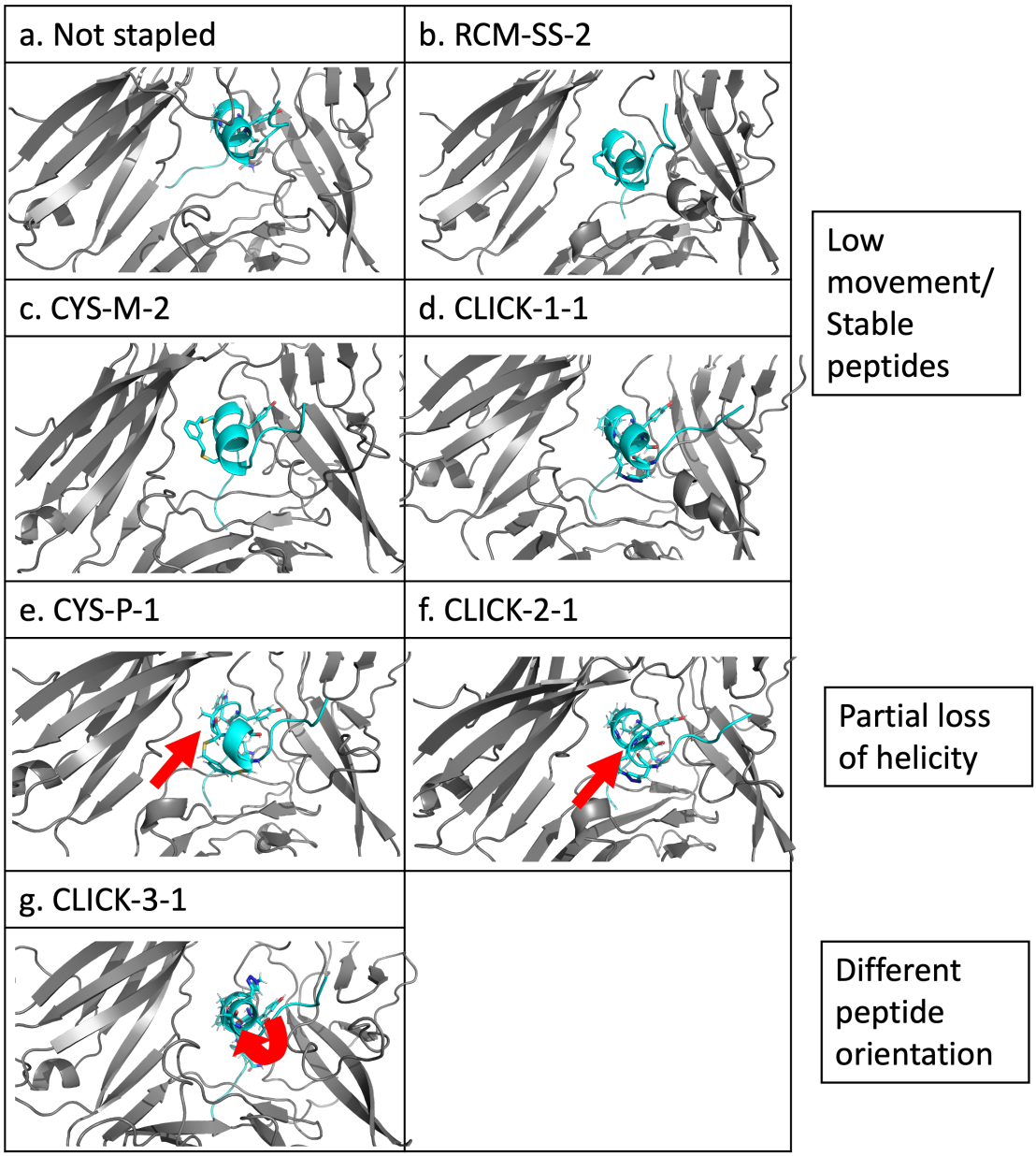



**Figure 6.8:** All snapshots have been taken after 250 ns of a 500 ns simulation. The distortion of peptides CYS-M-1 and CLICK-2-1 is marked with a red arrow. The change of orientation of peptide CLICK-3-1 is highlighted with a curved arrow.

### 6.3.1 Trajectory analysis using trajectory quantitative values

The primary visual inspection of the simulations, gave us some insight into the stability of the peptides. More in-depth and quantitative analyses were performed in an attempt to quantify the binding strength of the peptides in the simulations. The distance variations between the center of mass of the peptide sequences and the center of mass of IL-1R1 suggested that all peptides were reasonable binders, as no distance variation superior to 10 Å was observed for any of the peptides, for both models.

Even if a binding energy cannot be deduced from the observation of a MD simulation, MD simulations provide a wealth of information on the stability of the interactions. Analysis of the dynamics of the peptide, as well as the conservation of its interactions provide a good indication of the strength of its binding. The stability of the peptide interactions can be quantified through trajectory analysis with descriptors such as Root Mean Squared Deviation (RMSD), Root Mean Squared Fluctuation (RMSF), radius of gyration (Rg) and conservation of helicity as measured with the Dictionary of Secondary Structure of Proteins (DSSP) assignment method. A series of simulation descriptors were thus computed and gave access to the binding dynamics of the different peptides.

Small backbone RMSD values relative to the starting structures correspond to small conformational changes of the peptide. Both the RMSD of the peptide backbone to the crystal structure and the RMSD of the side chains involved in key interactions were computed. Most significantly, large differences in RMSD among 3 consecutive runs (greater than 3 times the standard deviation) were only observed for a few simulations (3 sets in the d1/d2/d3 model and only 1 in the d1/d2 model), meaning that simulations generally gave reproducible outcomes. Discrepancies observed in consecutive runs can be a sign of weaker peptide binding. Indeed, a larger sampling allows to identify unfavourable conformations, as simulations will diverge from the initial structure. This could also imply a poorer initial model for these simulations, as starting from a higher energy conformation can lead to diverging simulations. Adding a staple generally introduces bonds with high energy in the building phase, which need to be relaxed before MD. Sometimes, the readjustment of out of equilibrium parameters can force greater conformational changes, which are irreversible in the duration of the simulation.

The RMSD was used to estimate the conservation of the binding pose. As an example the unstapled peptide trajectory showed a gradual rise of its RMSD to 5 Å with a plateau effect. The peptide RCM-RS-1 has a significant difference in final RMSD values when compared to the unstapled peptide (respectively 17 and 5 Å, see Figure 6.9), meaning that RCM-RS-1 settles in a somewhat different binding pose compared to the unstapled 21-mer in the crystal structure. For each peptide, the RMSD variations for the backbone between three simulation runs were found to be consistent with each other, with variations between replicate simulations less than 3 times the standard variation in the first MD run.

The RMSF showed that the head (residues 1-4) and tail (residues 17-21) of the peptides

were more flexible than the helical core (residues 5-12), which was to be expected from the visualisation of the simulated system trajectories. RMSF values reflected the fact that the peptide termini were binding less tightly than the helical core of the peptide (Figure 6.9).

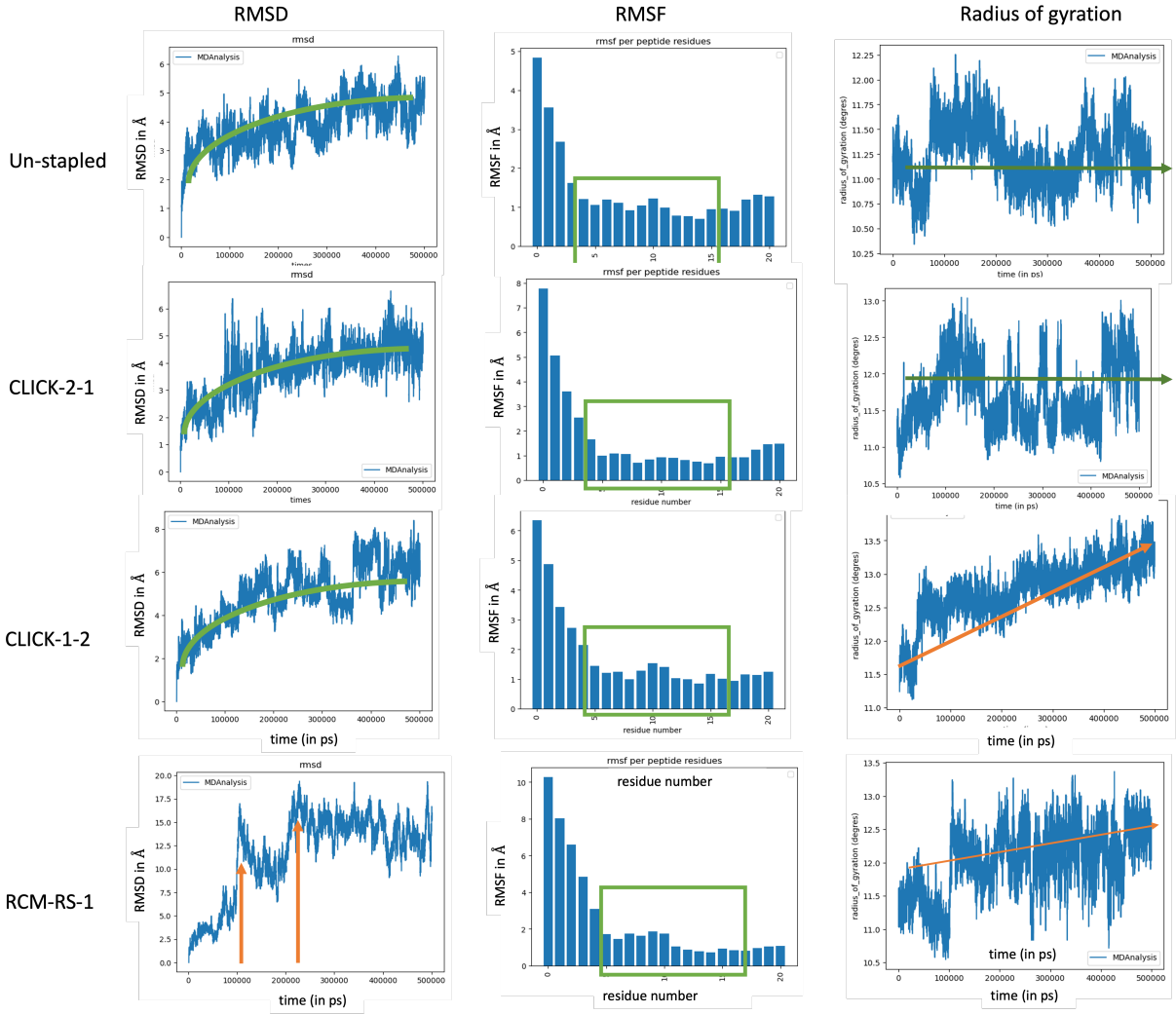



**Figure 6.9:** RMSD values, RMSF values and radius of gyration during 500 ns simulation for the unstapled peptide, CLICK-1-2 peptide and 2 unstable peptides, CLICK-2-1 and RCM-RS-1.

The helicity of peptides was calculated by averaging the helicity of all residues over the trajectory. When comparing the helicity of the stapled peptides with the 21-mer, similar helical content ranges were obtained. However, the helical propensity of a peptide obtained in a bound state, in a MD simulation cannot be linked directly to their helical content in solution. Overall, the helical content in the bound conformation was unaffected by the presence of a staple as the helicity content was already high in the unstapled peptide and remained mostly high in the simulations. However, some discrepancies were observed when comparing the helical propensities obtained in the d1/d2/d3 model with the ones obtained in the d1/d2 model. The helicity of the unstapled peptide was found to be 64 % for the d1/d2/d3 model and the average helicity for the 18 stapled peptides was found to be 61 %. CYS-M-2 had the lowest helicity of 55 %. For the truncated model simulation the helicity was found to be higher than

for the full model, with 73 % helicity for the unstapled peptide. The peptide RCM-RS-3 was found to have the highest helicity with an overall helicity of 83%. Lower values of helicity, especially for the full model, might be caused by the initial clashes with the receptor when the bulky staple residues are added. This moves the peptide from its initial conformation, which can perturb the helical composition. Also, staples can sometimes induce a slightly imperfect $\alpha$-helix, which is not always captured by the DSSP method.

The radius of gyration (Rg) is defined as the radial distance of the center of mass to a fixed point. It is a indicator of the spread of the centers of mass of a molecule. The values of the Rg are not significant for our case use, only the deviations have to be considered. In the MD trajectories for the 18 stapled peptides, a deviation of +/- 1.5 nm in the Rg was linked to a change in the peptide orientation. Importantly, this was not always captured by the RMSD descriptor. To take an example, in CLICK-2-1 simulations where the helical core was observed to deviate from its initial position, the backbone RMSD of the peptide was found to be relatively similar to the RMSD in the unstapled peptide simulation, (Figure 6.9). However, in this case the radius of gyration showed some discrepancies compared to the unstapled sequence, which reflects the rotation of the peptide. Thus, this observation is useful when used in conjunction with the RMSD to pick up more subtle changes.

Hydrogen bonds, as well as a distance based of non-polar interactions, were also computed to understand the link between these interactions and the stability of the peptide. However, the number of interactions (hydrogen bonds and non-polar interactions) were not found to vary significantly among the simulations or to give a hint of correlation with the stability of the peptides. Most simulations scored 18 to 20 stable polar interactions, and generally these values were repeatedly achieved for replica simulations. Therefore, these descriptors were not found to be useful for ranking the peptides as there was almost no variation across the peptides. The different numbers of interactions were found to reflect the staple position (e.g. mutation of Trp4).

Using different MD descriptors and different IL-1R1 models yielded different peptide scoring. Peptides with the lowest RMSD for the d1/d2 model were found to be the RCM-SS-2, CYS-M-3 and the peptides with triazoles, namely CLICK-1-2, CLICK-1-3, CLICK-2-1, CLICK-2-2, and CLICK-2-3. For the d1/d2/d3 model, RCM-RS-3, CLICK-1-1, CLICK-1-2, CLICK-1-3 were found to be the peptides with the lowest RMSD. It was found that using multiple descriptors instead of a single one gave more insight into the viability of the binding poses. Different performances for the two models were also observed, with the d1/d2/d3 model displaying more problems such as diverging replicates and loss of helicity of the stapled peptides. This indicates that the truncated model may be better suited for this system.

In Table 6.2, the best ranked peptides for each MD descriptor are shown. Generally, peptides with the staple in the third position were found to be more unstable, so these peptides were expected to be lower affinity binders. Overall, peptides with the staple in position 2 were found to be more stable and thus were expected to be better binders. Peptides with the staple

| | RMSD of the backbone | | RMSD of the sidechains | | RMSF of the peptide Cα at the interface | | RMSF of the peptide key residues | |
|---|---|---|---|---|---|---|---|---|
| a. d1/d2/d3 model | **CLICK-1-3** | 2.48 ± 0.61 | **RCM-SS-2** | 2.21 ± 0.57 | **CYSP-2** | 1.26 ± 0.65 | **CYSP-2** | 0.88 ± 0.14 |
| | **RCM-SS-3** | 2.94 ± 0.61 | **RS-3** | 2.60 ± 0.55 | **CYSM-3** | 1.38 ± 0.67 | **RCM-RS-3** | 0.97 ± 0.27 |
| | **CYS-P-2** | 3.02 ± 0.71 | **CLICK-1-3** | 2.69 ± 0.58 | **CLICK-1-3** | 1.40 ± 0.62 | **CYS-M-3** | 1.01 ± 0.10 |
| | **CYS-M-2** | 3.08 ± 1.00 | **CYS-P-1** | 2.95 ± 0.48 | **RCM-RS-3** | 1.51 ± 1.10 | **Not stapled** | 1.02 ± 0.19 |
| | **CLICK-1-1** | 3.11 ± 0.90 | **CLICK-1-2** | 3.07 ± 1.09 | **RCM-SS-3** | 1.54 ± 0.79 | **RCM-SS-2** | 1.02 ± 0.17 |
| b. d1/d2 model | **Not stapled** | 2.22 ± 0.39 | **CYS-M-3** | 1.79 ± 0.52 | **Not stapled** | 1.30 ± 0.59 | **Not stapled** | 1.04 ± 0.12 |
| | **RCM-SS-2** | 2.48 ± 0.51 | **RCM-RS-3** | 1.82 ± 0.51 | **RCM-SS-2** | 1.56 ± 0.86 | **CYS-M-3** | 1.12 ± 0.16 |
| | **CLICK-2-2** | 2.79 ± 0.67 | **CLICK-1-3** | 2.09 ± 0.54 | **CYS-P-3** | 1.56 ± 1.00 | **RCM-SS-2** | 1.18 ± 0.14 |
| | **CYS-M-2** | 2.85 ± 0.85 | **RCM-SS-3** | 2.11 ± 0.52 | **CYS-M-3** | 1.58 ± 0.97 | **CYS-P-3** | 1.19 ± 0.27 |
| | **CLICK-1-2** | 2.87 ± 0.62 | **Not stapled** | 2.38 ± 0.38 | **CLICK-1-2** | 1.70 ± 0.95 | **RCM-SS-3** | 1.19 ± 0.18 |

| | RMSF side-chain receptor Cα at the interface | | variation of the radius of gyration | | Helicity of the peptide | |
|---|---|---|---|---|---|---|
| a. d1/d2/d3 model | **CLICK-1-2** | 1.56 ± 0.79 | **RCM-SS-1** | 0.02 ± 0.32 | **CYS-P-3** | 0.74 ± 0.06 |
| | **CLICK-2-1** | 1.74 ± 0.83 | **RCM-RS-1** | 0.07 ± 0.42 | **CYS-M-1** | 0.74 ± 0.01 |
| | **RCM-RS-3** | 1.84 ± 1.11 | **CYS-M-1** | 0.09 ± 0.48 | **CLICK-2-2** | 0.73 ± 0.11 |
| | **RCM-RS-2** | 1.87 ± 1.09 | **CLICK-2-2** | 0.13 ± 0.46 | **CLICK-1-2** | 0.73 ± 0.13 |
| | **CLICK-2-3** | 1.89 ± 1.21 | **RCM-RS-2** | 0.18 ± 0.49 | **CLICK-2-1** | 0.72 ± 0.05 |
| b. d1/d2 model | **CYS-P-1** | 1.22 ± 0.76 | **CYS-M-2** | 0.04 ± 0.32 | **RCM-RS-3** | 0.82 ± 0.17 |
| | **CYS-P-2** | 1.22 ± 0.68 | **Not stapled** | 0.05 ± 0.22 | **CLICK-2-3** | 0.75 ± 0.05 |
| | **CYS-M-2** | 1.24 ± 0.66 | **RCM-RS-2** | 0.10 ± 0.22 | **CLICK-1-3** | 0.74 ± 0.17 |
| | **CLICK-2-2** | 1.24 ± 0.69 | **CYS-M-3** | 0.11 ± 0.50 | **CLICK-2-1** | 0.74 ± 0.09 |
| | **CLICK-2-3** | 1.24 ± 0.62 | **CYS-P-3** | 0.11 ± 0.35 | **RCM-SS-1** | 0.73 ± 0.10 |

**Table 6.2:** Peptide ranking according to their MD descriptors. Only the five best peptides for each descriptor have been reported.

in the first position yielded more variability in their binding stability depending on the staple chemistry. It is important to note that most stapled peptides did not outperform the unstapled peptide during the MD simulation.

No particular type of chemistry was found to give a general advantage, with cysteine cross linked peptides, RCM peptides and CLICK peptides reaching the top 5 of peptide ranking with the lowest RMSD, RMSF and change in Rg in both the truncated and full length model (see Table 6.2) CLICK peptides were found to produce peptides with a slightly higher helical content for both models. We selected peptides giving consistent ranking in the truncated and full model within the descriptors as well as certain peptides displaying unique characteristics such as RCM-RS-3 with the highest helicity in our simulations.

The list of synthesised peptides is as follows: RCM-SS-1, RCM-SS-2, RCM-SS-3, RCM-RS-3, CLICK-1-1, CLICK-1-2, CLICK-1-3, CLICK-2-1, CLICK-2-2, CYS-M-2, CYS-M-1, CYS-P-1 and CYS-P-2. Their synthesis was outsourced to PeptideSynthetics (Fareham, UK) during the pandemic and tested in a series of Bio-assays.

The binding energies were also computed using MM/PBSA and MM/GBSA and the results will be described in depth in section 6.5, where the values will be compared with experimental data. However the ranking was also found to be dependent on the model (d1/d2 or d1/d2/d3) and thus this energetic analysis was not directly used for the selection of the peptides. However

their free binding energy was found to be in the same range or lower that of the unstapled peptide, so a similar $IC_{50}$ was expected.

## 6.4 Bio-assays

### 6.4.1 FRET assays

#### FRET assays principle

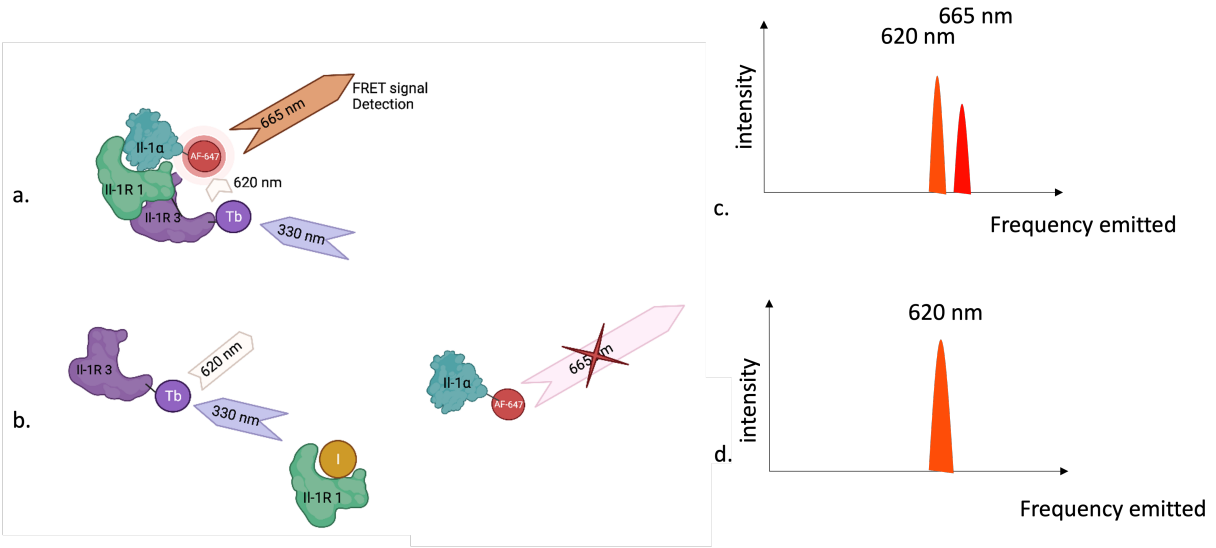



**Figure 6.10:** a. Transmission of the FRET signal: The terbium label is in proximity to the AF-647 fluorescent label and the signal is transmitted, then re-emited and detected. b. Inhibition of IL-1R: the Terbium and AF-647 are too far apart for the energy to be transmitted c. Wavelengths emitted when the Il-1$\alpha$/IL-1R1/IL-1R3 complex is formed d. Wavelengths emitted when the complex is not formed.

In a FRET assay, a donor chromophore is excited at a specific wavelength. While coming back to its low energy state, this chromophore transfers energy to an acceptor chromophore through non-radiative dipole–dipole coupling. FRET is extremely sensitive to small changes in distance as the efficiency of this energy transfer is inversely proportional to the sixth power of the distance between donor and acceptor. Thus, measurements of FRET can be used to determine if two fluorophores are within a certain distance of each other, and FRET is routinely applied to measure the binding of ligands to proteins. A simplified mechanism of the FRET assay is shown in Figure 6.10. Competitive FRET assays were carried out for 13 of the selected stapled peptides, as well as a positive control[436], to determine their $IC_{50}$. Both IL-1R3 and IL-1R1 were present in the assay. The IL-1R3 was labelled with a terbium chelate and IL-1$\alpha$ was labelled with the Alexa Fluor643 (AF-643) fluorophore.

The assay measures the formation of the IL-1R1/IL-1R3 complex of the receptor upon IL-1$\alpha$ binding. A homogeneous time-resolved fluorescence (HTRF) was used. HTRF relies

on the slow emission of the lanthanides which are long-lived luminophores. It allows the measurement to be time resolved, thus removing the interference from the faster emitting buffer or medium components. Terbium chelates are a widely used as donors. In our assay AlexaFluor-665 (AF-665) was used as acceptor. The intensities at the two emitting wavelengths, i.e. 620 nm for the terbium emission and 665 nm for the AF-647 emission, are meaured. The ratio of the intensities is calculated which normalises the difference of emitted wavelengths between each well and removes interference.

### FRET assays results

The dose response curve of the 12 peptides are reported in Figure 6.11.

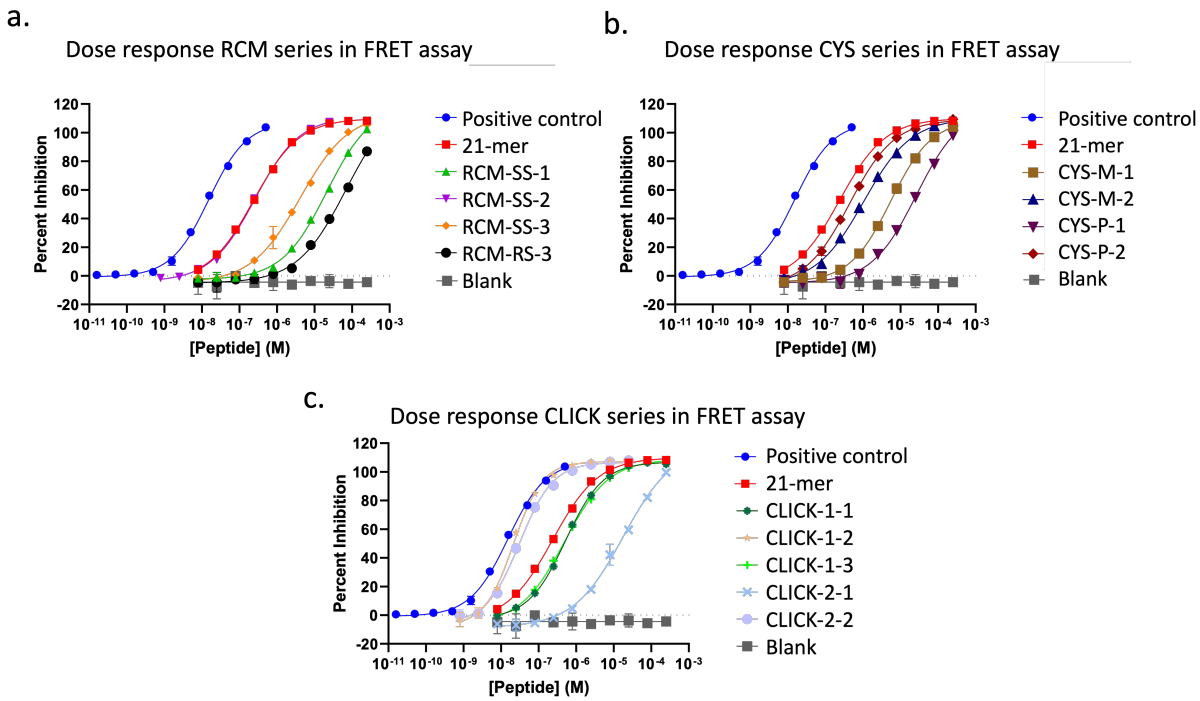



**Figure 6.11:** Affinity vs log concentration obtained by FRET assay, a. for the Ring metathesis stapled peptides, b. for the cysteine crossed-linked stapled peptides and c. for the CLICK stapled peptides

The positive control peptide was used as a positive control for the binding of the *de novo* designed peptides. All peptides were found to be active with a 100 % inhibition below 1 mM, indicating that all peptides were complete competitive inhibitors, except for the RCM-RS-3 which had only 89 % inhibition at the tested concentration of 1 mM. This lower inhibition was attributed to the range of concentrations used in the assay rather than this peptide being a partial inhibitor. The steepness of the slope is quantified by the Hill slope. A dose-response curve with a standard response has a Hill slope of 1.0. The slopes of the curves for the inhibition of IL-1R1 were all found to be between 0.5 and 1.5 which is a satisfactory range and indicates a ”one peptide to one binding site” competitive binding.

The FRET assay was run 3 times independently and the $IC_{50}$ values given in Table 6.3

| n=2 25*/250uM top | FRET | | | SPR | | |
|---|---|---|---|---|---|---|
| | $IC_{50}$ (nM) | Upper confidence limit | Lower confidence limit | $k_{on}$ ($M^{-1}s^{-1}$) x $10^4$ | $k_{off}$ ($s^{-1}$) x $10^2$ | $K_D$ (nM) |
| **Positive Control** | 21.49 | 11.29 | 40.92 | N/A | N/A | N/A |
| **21mer WT** | 187.36 | 131.42 | 267.11 | 21.1 | 0.488 | 26 |
| **RCM-SS-1** | 23014.42 | 6561.7 | 80721 | 3.58 | 3.47 | 970 |
| **RCM-SS-2*** | 238.6 | 54.96 | 1035.9 | 26.2 | 1.12 | 6 |
| **RCM-SS-3** | 5144.38 | 2026.3 | 13061 | 15.6 | 4.39 | 281 |
| **RCM-RS-3** | 66885.71 | 44202 | 101210 | N/A | N/A | N/A |
| **CYS-M-1** | 6156.49 | 5000.5 | 7579.7 | 5.09 | 6.41 | 1260 |
| **CYS-M-2** | 934.69 | 628.65 | 1389.7 | 6.05 | 0.306 | 51 |
| **CYS-P-1** | 27968.34 | 17670 | 44268 | 5.60 | 2.96 | 529 |
| **CYS-P-2** | 387.26 | 276.08 | 543.21 | 10.6 | 0.611 | 58 |
| **CLICK-1-1** | 431.52 | 324.21 | 574.34 | 6.59 | 0.512 | 78 |
| **CLICK-1-2*** | 20.61 | 19.16 | 22.16 | 4.55 | 0.020 | 3 |
| **CLICK-1-3** | 459.55 | 296.5 | 712.26 | 7.03 | 17.2 | 3780 |
| **CLICK-2-1** | 16545 | 14618 | 18726 | 1.08 | 0.220 | 203 |
| **CLICK-2-2*** | 31.31 | 31.00 | 31.62 | 193 | 7.81 | 3 |

**Table 6.3:** Average of the $IC_{50}$ values obtained in 3 FRET assays (one assay was acquired independently by Martin Domville, UCB) with upper and lower confidence limits. $IC_{50}$ values have been colored for readability with the highest values in red and lowest values in green. $k_{on}$, $k_{o}$ff, KD obtained by a Biacore-SPR assay ran by Dr Bruce Carrington at UCB are also reported.

are the averages of the 3 assays. Overall, peptides stapled using CLICK chemistry displayed higher affinity than the other staple chemistry types. Only two peptides were found to have an affinity superior to the 21-mer. CLICK-2-2 had a $IC_{50}$ of 31 nM which is a 6 fold affinity increase and CLICK-1-2 had a $IC_{50}$ of 20 nM which is a 9.1 fold increase when compared to the unstapled sequence. The $IC_{50}$ of the CLICK-1-2 is comparable to the $IC_{50}$ of the positive control peptide. For CLICK-1-2 and CLICK-2-2 the addition of a staple had a positive contribution to the affinity of the peptide, when compared with the unstapled peptide. The RCM-SS-2 had a similar $IC_{50}$ when compared with the unstapled peptide (238 nM). For the other peptides however, the affinity of the peptide decreased, and the staples seemed to contribute slightly negatively (CLICK-1-1 and CLICK-1-3, CYS-P-2, CYS-P-3) or more negatively (CLICK-2-1, CYS-P-2, RCM-SS-1, RCM-SS-3 and RCM-RS-3) to the binding. However all the peptides remained competitive binders. The weakest binder is RCM-RS-3 with a 357-fold loss of affinity compared to the unstapled 21-mer peptide. Adding a staple in position 2 generally resulted in peptides which were better binders than peptides with a staple in position 1 and 3. $K_D$ values obtained from a SPR assay performed by Bruce Carrington at UCB were in good agreement with the $IC_{50}$ values obtained in the FRET assay.

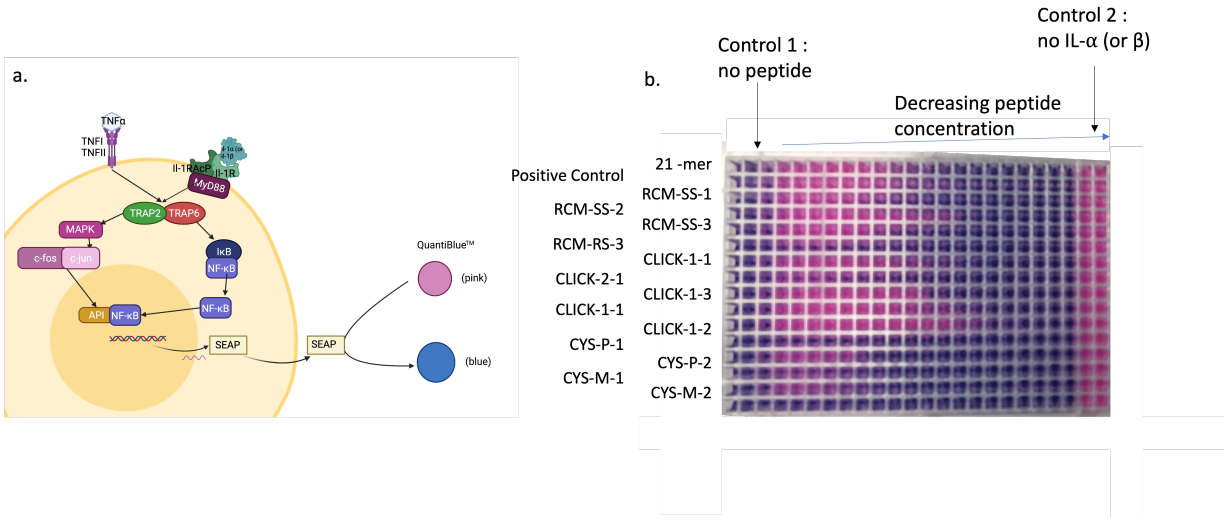



**Figure 6.12:** a. Cell assay method: After binding of IL-1α (or β) to IL-1R, the NF-κB pathway is activated, releasing SEAP from the cell b. plate obtained with the cell assay: pink, low SEAP concentration, blue: high SEAP concentration.

### 6.4.2 Cell assays

HEK-Blue™ TNF-$\alpha$/IL-1$\beta$ cells were used in this assay. They are commercially available and routinely used for the detection of TNF-receptor and IL-1 receptor inhibitors. These cells are engineered to detect the binding of TNF-$\alpha$ and IL-1$\alpha$ (and $\beta$) by monitoring the activation of the NF-κB pathway. (Figure 6.12) They were originally generated by stable transfection of HEK-293 cells with a secreted embryonic alkaline phosphatase (SEAP) reporter gene and sold by InvivoGen (San Diego, US). When the IL-1$\alpha$ binds to its receptor, the NF-κB pathway is activated and SEAP is released (Figure 6.13). Secretion of SEAP by HEK-Blue™ TNF-$\alpha$/IL-1$\beta$ cells is specifically induced by the binding of TNF-$\alpha$ and IL-1$\beta$ (and $\alpha$) to their respective receptors. SEAP levels can be readily determined with QUANTI-Blue™, a reagent that turns purple/blue in the presence of SEAP.

Firstly, none of the peptides was found to be toxic to the cells, by observing the cells under the microscope after 2 h, 24 h, and 48 h of incubation. This was further confirmed by a viability test using a Presto-Blue™ test, which uses resazurin. On entering live cells, the cellular reducing environment reduces resazurin to resorufin, a compound that is red and highly fluorescent. Viable cells continuously convert resazurin to resorufin, increasing the overall fluorescence and color of the media surrounding the cells.

The competitive inhibition of IL-1R1 activation by IL-1$\alpha$ and IL-1$\beta$ was tested at different incubation times (2 h, 24 h, and 48 h) for the peptides. Similarly to the FRET assay, all Hill slopes were found to be between 0.5 and 1.5, which indicates a one to one inhibitory mechanism for the peptides.

The $IC_{50}$ values of the peptides tested against IL-1$\alpha$ were within 1.7 fold of the $IC_{50}$ values of the peptides tested against IL-1$\beta$. Thus, the peptides were found to have a similar inhibitory

effect on IL-1$\alpha$ and IL-1$\beta$. The $IC_{50}$ of the peptides pre-incubated for 2 h were within 3 fold of the $IC_{50}$ pre-incubated for 24 h (presented in appendix) against cells stimulated with 8 pM IL-1$\alpha$. Thus, generally, the incubation time was not found to have a large impact on the $IC_{50}$ of the peptides. One exception was CLICK-2-2, which showed an 4.5 fold increase in potency following 24 h pre-incubation with the cells. Likewise, the $IC_{50}$ values of the peptides pre-incubated for 2h were within 3 fold of the $IC_{50}$ values obtained after pre-incubation for 24h with cells stimulated with 6 pM IL-1$\beta$. All peptides tested remained potent after 24 hours incubation with cells, before the addition of IL-1$_\alpha$ or IL-1$_\beta$.

In Figure 6.13, peptide dose-response curves are shown for inhibition of IL-1R1 activation by IL-1$\beta$. All curves show the effect of pre-incubating the peptides with the cells for 2 hours, before the addition of IL-1$\beta$. Dose-response curves with longer peptide pre-incubation times have been reported in appendix.

| | IL-1α 2 h incubation | | | IL-1β 2 h incubation | | | IL-1α 24 h incubation | | | IL-1β 24 h incubation | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $IC_{50}$ (nM) | Upper confidence limit | Lower confidence limit | $IC_{50}$ (nM) | Upper confidence limit | Lower confidence limit | $IC_{50}$ (nM) | Upper confidence limit | Lower confidence limit | $IC_{50}$ (nM) | Upper confidence limit | Lower confidence limit |
| Positive Control | 123.28 | 64 | 237.46 | 95.11 | 43 | 210.36 | 202.8 | 155.6 | 492.82 | 244.83 | 192.63 | 595.73 |
| 21mer WT | 514.68 | 258.1 | 1026.31 | 585.55 | 267.34 | 1282.52 | 1406.76 | 805.42 | 1727.19 | 1875.47 | 1560.43 | 2258.24 |
| RCM-SS-1 | 1782.51 | 961.43 | 3304.83 | 2251.46 | 1500.21 | 3378.9 | 2104.55 | 1007.38 | 3241.63 | 2919.21 | 1686.32 | 3366.12 |
| RCM-SS-2 | 194.85 | 56.05 | 677.34 | 229.61 | 91.28 | 577.59 | 395.98 | 175.7 | 700.67 | 459.15 | 273.34 | 702.37 |
| RCM-SS-3 | 1062.91 | 472.56 | 2390.75 | 1313.77 | 669.89 | 2576.51 | 2011.04 | 2011.04 | 3773.71 | 2464.47 | 2464.47 | 3682.99 |
| RCM-RS-3 | >30000.00 | 30000 | N/A | 20152.52 | 10979.8 | 36988.3 | >30000.00 | 30000 | N/A | >30000.00 | 30000 | N/A |
| CYS-M-1 | 6069.53 | 5572.77 | 6610.57 | 8158.76 | 2593.07 | 25670.4 | 20835.44 | 13062.8 | 47521.4 | 22258.5 | 15689.5 | 43720.4 |
| CYS-M-2 | 1958.33 | 1202.13 | 3190.21 | 2515.72 | 1556.59 | 4065.83 | 3776.82 | 3776.82 | 5422.35 | 4402.72 | 4402.72 | 5772.6 |
| CYS-P-1 | >30000.00 | 30000 | N/A | >30000.00 | 30000 | N/A | >30000.00 | 30000 | N/A | >30000.00 | 30000 | N/A |
| CYS-P-2 | 1842.78 | 992.52 | 3421.42 | 1843.72 | 1042.03 | 3262.19 | 2745.96 | 2039.64 | 5840.73 | 2978.38 | 1682.87 | 4399.58 |
| CLICK-1-1 | 256.64 | 148.15 | 444.57 | 214.72 | 80.04 | 576.01 | 410.21 | 251.46 | 516.47 | 586.37 | 450.26 | 912.47 |
| CLICK-1-2 | 10.3 | 7.32 | 14.49 | 13.56 | 6.3 | 29.17 | 26.79 | 7.41 | 39.13 | 34.64 | 12.64 | 42.5 |
| CLICK-1-3 | 131.93 | 52.73 | 330.13 | 162.36 | 61.84 | 426.27 | 259.89 | 8.83 | 703.08 | 344.36 | 71.93 | 752.07 |
| CLICK-2-1 | 5496.7 | 3090.74 | 9775.55 | 21782.87 | 6356.6 | 74645.7 | >30000.00 | 30000 | N/A | >30000.00 | 30000 | N/A |
| CLICK-2-2 | 32.28 | 10.81 | 96.39 | 58.33 | 14.76 | 230.52 | 135.73 | 135.73 | 381.49 | 166.55 | 166.55 | 411.61 |

**Figure 6.13:** Dose response curves of the cell assay show inhibition of IL-1βbinding to IL-1R, for the RCM, CYS and CLICK series.

The cysteine cross-linked peptides appeared to have comparable inhibitory activity in the cell and FRET assays. However an increase of affinity for the peptides with CLICK staple chemistry in the cell assay, when compared to the FRET assay was observed. Indeed all CLICK peptides where found to have an enhanced $IC_{50}$ when compared to the native 21-mer sequence in the cell assay. CLICK-2-2 and CLICK-1-2 were found to be respectively 16 and 52 fold more potent than the native sequence and respectively 4 and 12 fold more potent when compared to the positive control peptide. This constitute respectively a 3.4 and 11 fold increase between the cell and FRET assays. Also, the CLICK-1-1 and CLICK-1-3 peptides were found to be respectively 2 and 4 fold more potent than the native peptide, while they were less potent then the 21-mer in the FRET assay. The affinity of the RCM peptides was also increased in the cell assay when compared to the FRET assay, although not as drastically

| | IL-1α 2 h incubation | | | IL-1β 2 h incubation | | | IL-1α 24 h incubation | | | IL-1β 24 h incubation | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $IC_{50}$ (nM) | Upper confidence limit | Lower confidence limit | $IC_{50}$ (nM) | Upper confidence limit | Lower confidence limit | $IC_{50}$ (nM) | Upper confidence limit | Lower confidence limit | $IC_{50}$ (nM) | Upper confidence limit | Lower confidence limit |
| 15mer WT | 123.28 | 64 | 237.46 | 95.11 | 43 | 210.36 | 202.8 | 155.6 | 492.82 | 244.83 | 192.63 | 595.73 |
| 21mer WT | 514.68 | 258.1 | 1026.31 | 585.55 | 267.34 | 1282.52 | 1406.76 | 805.42 | 1727.19 | 1875.47 | 1560.43 | 2258.24 |
| RCM-SS-1 | 1782.51 | 961.43 | 3304.83 | 2251.46 | 1500.21 | 3378.9 | 2104.55 | 1007.38 | 3241.63 | 2919.21 | 1686.32 | 3366.12 |
| RCM-SS-2 | 194.85 | 56.05 | 677.34 | 229.61 | 91.28 | 577.59 | 395.98 | 175.7 | 700.67 | 459.15 | 273.34 | 702.37 |
| RCM-SS-3 | 1062.91 | 472.56 | 2390.75 | 1313.77 | 669.89 | 2576.51 | 2011.04 | 2011.04 | 3773.71 | 2464.47 | 2464.47 | 3682.99 |
| RCM-RS-3 | >30000.00 | 30000 | N/A | 20152.52 | 10979.8 | 36988.3 | >30000.00 | 30000 | N/A | >30000.00 | 30000 | N/A |
| CYS-M-1 | 6069.53 | 5572.77 | 6610.57 | 8158.76 | 2593.07 | 25670.4 | 20835.44 | 13062.8 | 47521.4 | 22258.5 | 15689.5 | 43720.4 |
| CYS-M-2 | 1958.33 | 1202.13 | 3190.21 | 2515.72 | 1556.59 | 4065.83 | 3776.82 | 3776.82 | 5422.35 | 4402.72 | 4402.72 | 5772.6 |
| CYS-P-1 | >30000.00 | 30000 | N/A | >30000.00 | 30000 | N/A | >30000.00 | 30000 | N/A | >30000.00 | 30000 | N/A |
| CYS-P-2 | 1842.78 | 992.52 | 3421.42 | 1843.72 | 1042.03 | 3262.19 | 2745.96 | 2039.64 | 5840.73 | 2978.38 | 1682.87 | 4399.58 |
| CLICK-1-1 | 256.64 | 148.15 | 444.57 | 214.72 | 80.04 | 576.01 | 410.21 | 251.46 | 516.47 | 586.37 | 450.26 | 912.47 |
| CLICK-1-2 | 10.3 | 7.32 | 14.49 | 13.56 | 6.3 | 29.17 | 26.79 | 7.41 | 39.13 | 34.64 | 12.64 | 42.5 |
| CLICK-1-3 | 131.93 | 52.73 | 330.13 | 162.36 | 61.84 | 426.27 | 259.89 | 8.83 | 703.08 | 344.36 | 71.93 | 752.07 |
| CLICK-2-1 | 5496.7 | 3090.74 | 9775.55 | 21782.87 | 6356.6 | 74645.7 | >30000.00 | 30000 | N/A | >30000.00 | 30000 | N/A |
| CLICK-2-2 | 32.28 | 10.81 | 96.39 | 58.33 | 14.76 | 230.52 | 135.73 | 135.73 | 381.49 | 166.55 | 166.55 | 411.61 |

**Table 6.4:** $IC_{50}$ values for cell assays n=5 for 2 h assays; n=3 for 24 h assays. A green to red gradient scale was added for readability. Green highlights low $IC_{50}$ values while red highlights $IC_{50}$ values. Three assays at 2 hours incubation and all assays at 24 hours incubation were run by Dr. Melanie Wong.

as for the CLICK peptides. For example, the affinity of the RCM-SS-2 was found to be 3.5 fold the affinity of the 21-mer WT peptide in the cell assay, while in the FRET assay their affinities were almost identical.

Many hypothesis can enter into play to explain these observations such as an increase of helicity of the peptides, which could increase their membrane permeability. The cysteine cross-linked peptides might also be more cell penetrant due to the large lipophilic group. Therefore, a diffusion of the peptide inside the cell might have occurred, diminishing the availability of the peptide for the IL-1R1 trans-membrane receptor. In contrast, RCM-SS-1, RCM-SS-3 and CLICK-2-2 appeared to be more potent in the cell assay than in the FRET assay, with CLICK-2-2 being 10 fold more potent in the cell assay than in the FRET assay.

## 6.5 Comparison of Free energy-based methods with experimental assays

Cell-based assays can be considered as complex, as the potency of the peptides is measured after a multi-step pathway. In contrast, the FRET assay has less variables in the measurement of the potency of the peptides. Therefore, we only compared the binding energy obtained in our computational approach with the inhibition percentages obtained with the FRET assays. Still, the competitive FRET assay required the formation of an equilibrium between (unlabelled) peptide and labelled IL-1$\beta$ protein binding to the labelled IL-1R, while in simulations only the binding of the peptide to IL-1R1 was taken into account. The systems used in MD simulations are thus much more simplified than the FRET assay, as neither competition with

| name | Experimental $IC_{50}$ (µM) | RMSD of the backbone (Å) | RMSD of the sidechains (Å) | RMSF of the backbone C$\alpha$ (Å) | RMSF of the key side chains (Å) | RMSF sidechain receptor(Å) | varitation of the radius of gyration (°) | Helicity (%) |
|---|---|---|---|---|---|---|---|---|
| **CLICK-1-2** | 0.0206 | 2.87 ± 0.62 | 2.83 ± 0.74 | 1.70 ± 0.95 | 1.30 ± 0.27 | 1.33 ± 0.78 | 0.17 ± 0.34 | 66 ± 0.10 |
| **CLICK-2-2** | 0.0313 | 3.43 ± 1.17 | 2.66 ± 0.55 | 1.89 ± 1.52 | 1.30 ± 0.31 | 1.24 ± 0.69 | 0.27 ± 0.32 | 72 ± 0.13 |
| **Not stapled** | 0.187 | 2.22 ± 0.39 | 2.38 ± 0.38 | 1.30 ± 0.59 | 1.04 ± 0.12 | 1.39 ± 0.77 | 0.05 ± 0.22 | 72 ± 0.01 |
| **RCM-SS-2** | 0.239 | 2.48 ± 0.51 | 2.95 ± 0.84 | 1.56 ± 0.86 | 1.18 ± 0.14 | 1.34 ± 0.70 | 0.28 ± 0.25 | 72 ± 0.05 |
| **CYS-P-2** | 0.387 | 2.95 ± 0.80 | 2.43 ± 0.52 | 1.59 ± 1.11 | 1.14 ± 0.17 | 1.35 ± 0.74 | 0.17 ± 0.32 | 76 ± 0.09 |
| **CLICK-1-1** | 0.432 | 5.22 ± 1.32 | 3.65 ± 0.74 | 2.37 ± 1.80 | 1.78 ± 0.69 | 1.31 ± 0.71 | 0.66 ± 0.47 | 68 ± 0.13 |
| **CLICK-1-3** | 0.460 | 3.26 ± 1.16 | 2.09 ± 0.54 | 1.96 ± 1.18 | 1.68 ± 0.60 | 1.31 ± 0.72 | 0.04 ± 0.32 | 74 ± 0.17 |
| **CYS-M-2** | 0.935 | 2.85 ± 0.85 | 2.86 ± 0.64 | 1.79 ± 1.02 | 1.39 ± 0.42 | 1.29 ± 0.66 | 0.18 ± 0.30 | 71 ± 0.01 |
| **RCM-SS-3** | 5.14 | 3.13 ± 0.81 | 2.11 ± 0.52 | 1.75 ± 1.12 | 1.19 ± 0.18 | 1.29 ± 0.61 | 0.13 ± 0.29 | 72 ± 0.05 |
| **CYS-M-1** | 6.16 | 4.97 ± 3.21 | 3.15 ± 0.72 | 3.37 ± 2.89 | 2.66 ± 1.27 | 1.52 ± 0.81 | 0.66 ± 0.56 | 59 ± 0.05 |
| **CLICK-2-1** | 16.5 | 3.87 ± 1.23 | 3.86 ± 0.83 | 2.16 ± 1.42 | 1.50 ± 0.37 | 1.43 ± 0.74 | 0.39 ± 0.41 | 74 ± 0.09 |
| **RCM-SS-1** | 23.0 | 6.06 ± 1.57 | 3.68 ± 0.73 | 2.78 ± 2.17 | 2.08 ± 0.80 | 1.41 ± 0.76 | 0.68 ± 0.41 | 73 ± 0.10 |
| **CYS-P-1** | 28.0 | 5.59 ± 1.51 | 3.66 ± 0.39 | 2.05 ± 1.83 | 1.37 ± 0.61 | 1.22 ± 0.76 | 0.65 ± 0.57 | 72 ± 0.13 |
| **RCM-RS-3** | 66.9 | 3.42 ± 1.14 | 1.82 ± 0.51 | 1.82 ± 1.45 | 1.22 ± 0.23 | 1.27 ± 0.70 | 0.11 ± 0.35 | 82 ± 0.17 |
| **RCM-RS-1** | N/A | 5.56 ± 1.00 | 3.03 ± 0.78 | 2.33 ± 1.55 | 1.72 ± 0.46 | 1.31 ± 0.73 | 0.11 ± 0.50 | 69 ± 0.01 |
| **RCM-RS-2** | N/A | 4.95 ± 1.18 | 3.20 ± 0.70 | 2.17 ± 1.40 | 1.80 ± 0.43 | 1.37 ± 0.79 | 0.61 ± 0.43 | 62 ± 0.05 |
| **CLICK-2-3** | N/A | 3.94 ± 0.96 | 2.67 ± 0.46 | 2.13 ± 1.77 | 1.49 ± 0.32 | 1.24 ± 0.62 | 0.43 ± 0.41 | 75 ± 0.05 |
| **CYS-M-3** | N/A | 2.79 ± 0.67 | 1.79 ± 0.52 | 1.58 ± 0.97 | 1.12 ± 0.16 | 1.24 ± 0.66 | 0.10 ± 0.22 | 73 ± 0.09 |
| **CYS-P-3** | N/A | 3.30 ± 0.78 | 2.73 ± 0.47 | 1.56 ± 1.00 | 1.19 ± 0.27 | 1.22 ± 0.68 | 0.34 ± 0.26 | 71 ± 0.13 |

**Table 6.5:** Simulation descriptors obtained for the d1/d2 (truncated) model with peptides ranked by affinity in the FRET assay.

IL-1$\alpha$, IL-1$\beta$, nor dimerisation with the receptor type 3 was modelled.

Peptides with the highest and lowest helical content during the simulations were not found to be the best binders (Tables 6.6 and 6.5). CYS-P-2 and CLICK-1-1 have the highest helical content (both 74 %) in the truncated model, and RCM-RS-3 has the highest helical content (82 %) in the full length model. Indeed, these peptides where found to have the lowest affinity in the FRET assay. While an increase in helicity might still have a positive effect on the binding in this study, peptides with high helicity are often not the best binders, as observed in other studies.[437, 438] This might indicate that the peptide needs to maintain a part of its flexibility to facilitate the binding.

The RMSD of the the C$\alpha$ carbon of the peptides gave a more straightforward indication of the stability of the peptides. Peptides with high C$\alpha$ carbon RMSDs (above 4 Å) were found to be the weaker binders, although lower RMSD values were not always linked with tight binders, as the RCM-RS-3 peptide had RSMD values below 3 in both models and yielded the lowest $IC_{50}$ in the FRET assay. Similar observations were made for RMSF values, where in general the lower RSMF values were associated with the best binders like CLICK-1-2, CLICK-2-2, the unstapled peptide, RCM-SS-2 and CYS-P-2, which have RMSF values below 1.2 Å. However a low RMSF value does not always reflect a good binder. For example, RCM-RS-3 also has a RMSF value of 1.3 Å and binds IL-1R1 with a lower affinity ($IC_{50}$ value of 66.9 $\mu$M) compared to the other peptides. Thus, although all good binders display low RMSF values, a low RMSF value does not guarantee a binder with a high affinity. In summary the $IC_{50}$ of the peptide does not strictly correlate with any RMS descriptor.

| name | Experimental $IC_{50}$ (μM) | RMSD of the backbone (Å) | RMSD of the sidechains (Å) | RMSF of the backbone Cα (Å) | RMSF of the key side chains (Å) | RMSF sidechain receptor(Å) | varitation of the radius of gyration (°) | Helicity (%) |
|---|---|---|---|---|---|---|---|---|
| **CLICK-1-2** | 0.0206 | 3.57 ± 1.16 | 3.07 ± 1.09 | 1.95 ± 1.66 | 1.24 ± 0.45 | 1.56 ± 0.79 | 0.23 ± 0.53 | 59 ± 0.13 |
| **CLICK-2-2** | 0.0313 | 3.20 ± 1.31 | 3.08 ± 0.64 | 2.25 ± 1.77 | 1.27 ± 0.23 | 2.12 ± 1.35 | 0.78 ± 0.42 | 58 ± 0.11 |
| **Not stapled** | 0.187 | 4.81 ± 1.20 | 3.60 ± 0.59 | 1.71 ± 1.52 | 1.02 ± 0.19 | 2.27 ± 1.87 | 0.18 ± 0.40 | 64 ± 0.07 |
| **RCM-SS-2** | 0.239 | 4.29 ± 1.52 | 2.21 ± 0.57 | 2.02 ± 1.77 | 1.02 ± 0.17 | 2.33 ± 1.82 | 0.18 ± 0.49 | 65 ± 0.03 |
| **CYS-P-2** | 0.387 | 3.02 ± 0.71 | 3.81 ± 0.58 | 1.26 ± 0.65 | 0.88 ± 0.14 | 1.99 ± 1.46 | 0.30 ± 0.27 | 74 ± 0.01 |
| **CLICK-1-1** | 0.432 | 3.11 ± 0.90 | 3.59 ± 0.61 | 1.69 ± 0.91 | 1.16 ± 0.25 | 1.96 ± 1.26 | 0.30 ± 0.32 | 74 ± 0.05 |
| **CLICK-1-3** | 0.460 | 2.48 ± 0.61 | 2.69 ± 0.58 | 1.40 ± 0.62 | 1.16 ± 0.16 | 1.99 ± 1.28 | 0.23 ± 0.23 | 72 ± 0.09 |
| **CYS-M-2** | 0.935 | 3.08 ± 1.00 | 3.52 ± 0.55 | 1.63 ± 1.04 | 1.07 ± 0.22 | 2.13 ± 1.45 | 0.27 ± 0.29 | 69 ± 0.13 |
| **RCM-SS-3** | 5.14 | 2.94 ± 0.61 | 3.14 ± 0.68 | 1.54 ± 0.79 | 1.14 ± 0.16 | 1.93 ± 1.25 | 0.23 ± 0.26 | 66 ± 0.13 |
| **CYS-M-1** | 6.16 | 6.21 ± 1.38 | 5.12 ± 0.95 | 2.44 ± 1.79 | 1.66 ± 0.69 | 2.26 ± 1.56 | 1.27 ± 0.68 | 57 ± 0.01 |
| **CLICK-2-1** | 16.5 | 6.83 ± 1.28 | 4.13 ± 0.67 | 2.30 ± 1.76 | 1.50 ± 0.48 | 1.74 ± 0.83 | 0.58 ± 0.40 | 60 ± 0.05 |
| **RCM-SS-1** | 23.0 | 4.84 ± 1.59 | 4.36 ± 1.03 | 2.32 ± 1.60 | 1.72 ± 0.52 | 2.25 ± 1.46 | 0.13 ± 0.46 | 66 ± 0.09 |
| **CYS-P-1** | 28.0 | 3.44 ± 0.87 | 2.95 ± 0.48 | 1.86 ± 1.15 | 1.12 ± 0.18 | 2.08 ± 1.35 | 0.39 ± 0.36 | 73 ± 0.13 |
| **RCM-RS-3** | 66.9 | 3.20 ± 0.95 | 2.60 ± 0.55 | 1.51 ± 1.10 | 0.97 ± 0.27 | 1.84 ± 1.11 | 0.02 ± 0.32 | 68 ± 0.17 |
| **RCM-RS-1** | N/A | 6.38 ± 1.99 | 6.58 ± 1.55 | 2.34 ± 1.61 | 1.76 ± 0.45 | 3.96 ± 3.89 | 0.09 ± 0.48 | 69 ± 0.01 |
| **RCM-RS-2** | N/A | 4.17 ± 0.65 | 3.36 ± 0.80 | 1.88 ± 1.07 | 1.45 ± 0.47 | 1.87 ± 1.09 | 0.79 ± 0.30 | 61 ± 0.05 |
| **CLICK-2-3** | N/A | 4.38 ± 1.35 | 3.24 ± 0.78 | 1.86 ± 1.35 | 1.11 ± 0.30 | 1.89 ± 1.21 | 0.19 ± 0.34 | 65 ± 0.05 |
| **CYS-M-3** | N/A | 5.27 ± 0.95 | 4.62 ± 0.76 | 1.38 ± 0.67 | 1.01 ± 0.10 | 2.44 ± 2.03 | 0.23 ± 0.21 | 73 ± 0.05 |
| **CYS-P-3** | N/A | 6.24 ± 1.23 | 4.09 ± 0.62 | 2.47 ± 1.74 | 1.75 ± 0.48 | 1.96 ± 1.15 | 0.07 ± 0.42 | 56 ± 0.06 |

**Table 6.6:** Simulation descriptors of the d1/d2/d3 (full) model with peptides ranked by affinity in the FRET assay.

From the analysis of tables 6.6 and 6.5, it would seem that overall both the full and truncated models were able to separate the weak and tight binders, however in the truncated model the results for the CLICK-1-2 peptides suggest a weaker binding, while the RCM-RS-3 and CYS-P-2 suggest a stronger binding than what was obtained in the experiments.

It is also important to note that the descriptors values are all in a range suggesting the peptides are binders. Indeed, the RMSD values are all relatively low in comparison with those obtained for the peptides designed to bind the SARS-CoV-2 spike protein, which were described in Chapter 5.

### 6.5.1 Energetic analysis and comparison with experimental values

Several techniques exist to predict the binding free energies based on MD simulations, including free energy perturbation (FEP),[439] thermodynamic integration (TI),[440] molecular mechanics Poisson–Boltzmann surface area (MM/PBSA) and molecular mechanics generalized Born surface area (MM/GBSA) techniques.[94] The overall objective of the MM/PBSA and MM/GBSA methods is to calculate the free energy difference between two states which most often represent the bound and unbound state of two solvated molecules or alternatively to compare the free energy of two different solvated conformations of the same molecule. The basic theory of the MM-GBSA approach is that the free energy of binding can be obtained through calculating only the end points of the thermodynamic cycle.

MMPB(GB)SA combines force field based molecular mechanics (MM) with a continuum

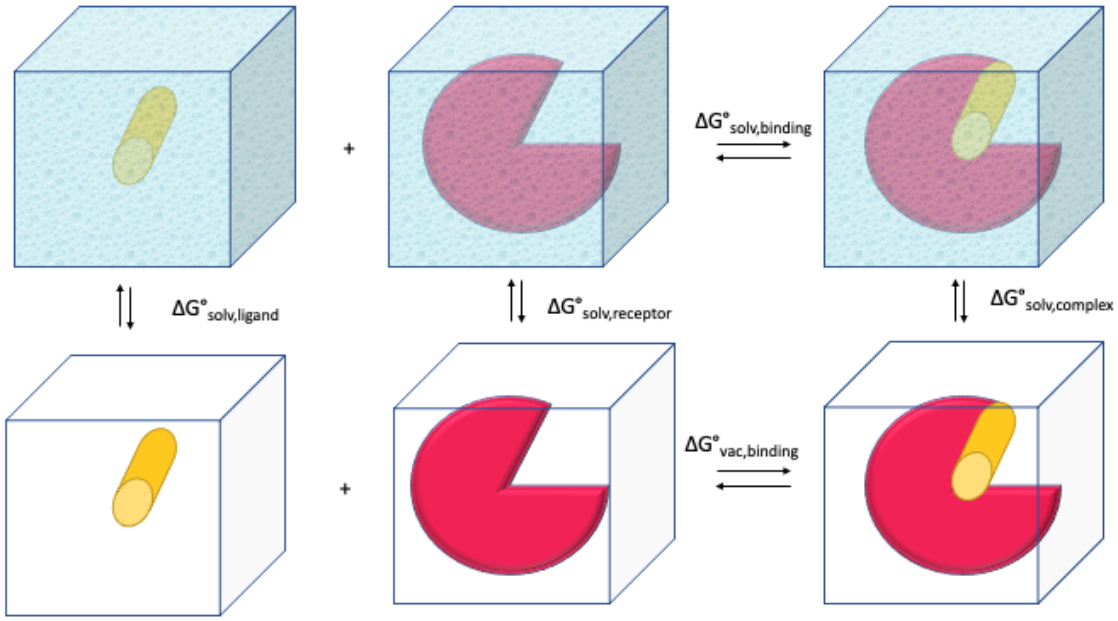



**Figure 6.14:** Thermodynamic cycle MM/PBSA

model, based on the Poisson–Boltzmann (PB) or generalized Born (GB) equation to handle electrostatic interactions, and a non-polar solvation energy correlated with solvent accessible surface area (SA). In principle, the method can be applied to any aqueous binding process: it is not dependent on assumptions of small (perturbative) differences between two similar systems. The configurational entropy due to ligand–protein binding can be calculated by a normal mode (NMODE) method. The MMPB(GB)SA methodology has been applied in various small molecule–protein systems, including drugs binding to their receptors.[61, 441–443]

However, in a solvated system the majority of the energy contributions come from solvent-solvent interactions rather than peptide-protein interactions. The binding energy would be lower in magnitude than the fluctuation in energy within the solvent.

Evidently from the simplified diagram presented in Figure 6.14, the binding free energy $\Delta G_{bind,solv}$ can be calculated with the equation 6.1:

$$\Delta G_{solv,binding} = \Delta G_{vac.,binding} + \Delta G_{solv,complex} - (\Delta G_{solv,ligand} + \Delta G_{solv,receptor}) \tag{6.1}$$

And each term can be estimated as follows:

$$\Delta G = \Delta G_{MM} + \Delta G_{sol} T\Delta S \tag{6.2}$$

where $\Delta G_{MM}$ is the molecular mechanics free energy, $\Delta G_{sol}$ is the solvation free energy, and TΔS represents the entropy term. The molecular mechanics energy was calculated by the

electrostatic and van der Waals interactions, while the solvation free energy was composed of the polar and the nonpolar contributions: $\Delta G_{MM} = \Delta G_{ele} + \Delta G_{vdw}$ MM/PBSA methods have become widely adopted in estimating protein-ligand binding affinities, due to their efficiency and high correlation with experiments. Both MMPSA and MPBSA calculations were run using the established standard parameters, except for dielectric constants, as we explored the impact of using different values of dielectric constants. Several research groups have reported that a solute dielectric constant higher than that used in molecular dynamics (i.e., 4 instead of 1) often leads to a better agreement with experimental data.[444–446] MM/PBSA and MM/GBSA were run on both the truncated and full length model of IL-1R1, and the free energy of binding obtained by these methods was compared with the experimental $IC_{50}$ values, as the ΔG binding should be equal to $-RTlnIC_{50}$.

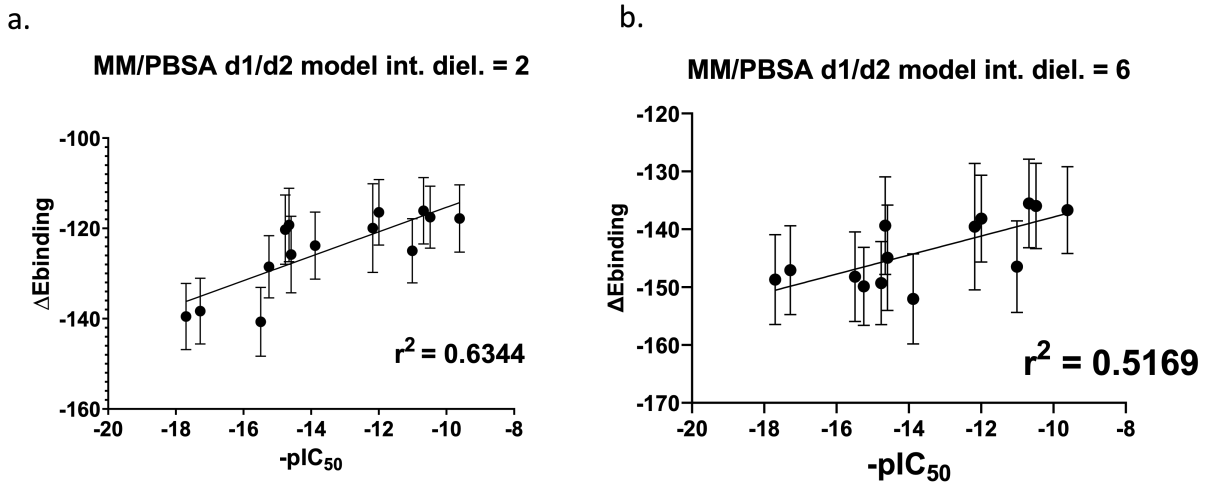



**Figure 6.15:** $\Delta E_{binding}$ using MM/PBSA vs experimental FRET $-pIC_{50}$ a. Using the d1/d2 model and internal dielectric of 2 b. Using the d1/d2 model and internal dielectric of 6.

| | MMPBSA | | | | MMGBSA | | | |
|---|---|---|---|---|---|---|---|---|
| | d1/d2/d3 model | | | | | | | |
| | diel = 1 | diel = 2 | diel = 4 | diel = 6 | diel = 1 | diel = 2 | diel = 4 | diel = 6 |
| Kendal-tau score | -0.32 | 0.00 | -0.32 | -0.43 | -0.32 | -0.32 | 0.00 | 0.54 |
| P value | 0.37 | 0.50 | 0.37 | 0.33 | 0.37 | 0.37 | 0.50 | 0.71 |
| Pearson correlations | 0.24 | 0.46 | 0.18 | 0.19 | 0.00 | 0.01 | 0.01 | 0.01 |
| P value | 0.08 | 0.01 | 0.13 | 0.12 | 0.89 | 0.69 | 0.78 | 0.78 |
| | | | | | | | | |
| | d1/d2 model | | | | | | | |
| Kendal-tau score | -0.43 | 1.30 | 0.54 | 0.00 | 1.30 | -0.32 | 0.00 | 0.54 |
| P value | 0.33 | 0.90 | 0.71 | 0.50 | 0.90 | 0.37 | 0.50 | 0.71 |
| Pearson correlation | 0.32 | 0.63 | 0.44 | 0.52 | 0.16 | 0.16 | 0.36 | 0.14 |
| P value | 0.03 | 0.0007 | 0.01 | 0.0038 | 0.16 | 0.16 | 0.02 | 0.19 |

**Table 6.7:** Pearson and Kendall-tau coefficient calculated between the experimental and free energy values for the MMPBSA and MMGBSA for different dielectric constants. The green values refer to value with a weak correlation according to the Pearson test

The effects of solute dielectric constant in MM/PBSA calculations and standalone MM/PBSA calculations have been analyzed before. It was shown that a solute dielectric

constant higher than that used in molecular dynamics (i.e., 4 instead of 1) often leads to better agreement with experiment. The highest correlations were observed for both the dielectric constant of 2 and 4.[447] Table 6.7 summarises the performance of the MM/PBSA and MM/GBSA at predicting the binding of the peptides for both the full model and the truncated model using different dielectric constants. The Kendall-tau and Pearson coefficients between the *in-silico* free energy of binding and FRET assay -p$IC_{50}$ were used to assess the quality of the MM/PB(GB)SA models. The detailed data of linear regression, Pearson product-moment correlation coefficient, and Kendall-tau pair value correlations are summarised in Table 6.7. $\Delta_{BindingEnergy}$ was plotted against the FRET $pIC_{50}$ values for three calculated series (Figure 6.15). The slopes of the linear regressions were found least often close to zeros in the truncated model, meaning that peptides were attributed similar energies of binding.

The MM/PBSA calculations based on the truncated model gave the best predictions in general. An internal dielectric constant of 2 was found to give the best correlation. These correlations were found to be weak according to the Pearson coefficient with $r^2$ of 0.63 and 0.43 respectively. The second best model was the truncated model with a constant dielectric of 6, which also yielded a weak correlation with the FRET $IC_{50}$. For other models, no correlation was observed.

These observations were in accordance with the work of Wang *et al.* who reported that a relatively higher interior dielectric constant was preferred for MM/PBSA to achieve good predictions for short peptides.[447] In contrast, the MM/GBSA calculations did not show any correlation and we did not observe any improvement using different dielectric constant.

The Kendall-tau coefficients, which measure the relationship between two rankings, disprove any correlation for any of the tested models and dielectric constants. Thus the MMPBSA and MMGBSA can be adequately used to give a prediction of the $\Delta_{binding}$ but are less suited to to produce a suitable ranking of the peptides. The introduction of a higher solute dielectric constant is a reasonable to account for the screening of electrostatic interactions due to polarization of electronics in the MM/PBSA method.

The model was found to have an impact on the final results, with the truncated model having the best correlations in general. This might be related to the fact that the binding is mediated by the d1 and d2 domains. This work thus illustrate the importance of the model in the final results. The higher electronic polarisation contributed to give a better fit to the experiment.

## 6.6 Conclusion

The goal of this study was to develop stapled peptides capable of inhibiting IL-1R1 signalling. Stapled peptides were designed based on the crystal structure of a known peptidic inhibitor, AF100847, in complex with IL-1R. From the crystal structure it became evident that only 3 positions on this 21-mer peptide were available for introducing i, i+4 staples. On these

positions, RCM, Cysteine cross-linking and CLICK chemistry was used to design 18 stapled peptides. It was then investigated how adding a staple to the 21-mer peptide would impact its binding affinity, with a series of molecular dynamics simulations. Two models, d1/d2 and d1/d2/d3, were used to model IL-1R1 in the simulations. For most peptide designs, the interactions during binding were found to be reasonably conserved in the simulations. Upon visual inspection of the simulations, the staples did not participate or interfere with the binding, with the exception of two peptides where the staple introduced conformational changes on the peptides. Besides a visual inspection of the simulations, the peptide binding poses were studied using a series of quantitative trajectory descriptors, including RMSD, RMSF, helicity and Rg. These descriptors proved useful to recognise weaker binders, but could not be used to efficiently categorise moderate and good binders *in-silico*. Thirteen stapled peptides were synthesized and tested with biophysical and cell assays.

Using a FRET assay measuring the interaction between IL-1$\alpha$ and IL-1R, all the peptides tested were found to be competitive inhibitors of IL-1R. However only three peptides showed similar or higher affinity for the receptor, than the 21-mer peptide on which they were based. Especially the CLICK chemistry peptides CLICK-1-2 and CLICK-2-2 performed well in the assay, with a 6- and 9-fold improvement of the $IC_{50}$ values compared to the unstapled 21-mer peptide. Interestingly, CLICK-1-2 and CLICK-2-2 have very similar $IC_{50}$ values as the positive control peptide. In a cell assay conducted with HEK cells, the tested peptides displayed no visible toxicity after 48 hours and were active on the cell receptors. Again, CLICK-1-2 and CLICK-2-2 performed the best in the assay, even showing better receptor inhibition than the positive control peptide. As the structure of the positive control peptide is not disclosed, this difference of affinity cannot be rationalised. While the $IC_{50}$ values of the unstapled peptides increased in the cell assay compared to the FRET assay (6-fold for the positive control peptide and 3-fold for the 21-mer), the $IC_{50}$ values of the CLICK peptides were virtually identical in both assays. It should be noted that the $IC_{50}$ value of the CLICK-2-2 peptide were higher with prolonged incubation times. For all other peptides, the incubation time did not produce large $IC_{50}$ variations. In conclusion, the designed peptides were found to be competitors of both IL-1$\alpha$ and IL-1$\beta$. From the FRET and cell assays, CLICK-2-1 and CLICK-2-2 have been identified as promising leads for the development of anti-inflammatory drugs. Studying these peptides in complex with IL-1R1 by X-ray crystallography would be useful to confirm that they bind in the same conformation as the 21-mer peptide, as expected from the MD simulations. The introduced staple is not expected to have any interaction with the protein based on the performed simulations, and this could also be verified. Since proteolytic instability is a critical limitation for peptide-based products, peptide half-life determination would also provide useful information to assess the pharmacological advantage of using stapled peptides.

Finally, we touched the issue of correlating the results obtained by molecular dynamic simulation to the biophysical and cell assays. Simulation data display some correlation with the experimental $-pIC_{50}$. However the two models (full length and truncated) used in these

simulations have an impact on the results. We found that some descriptors, such as RMSD of the peptide backbone, were found to be useful at eliminating some weak binders, as peptides with high values were generally found to get the lowest $-pIC_{50}$. However no descriptor that efficiently ranked the peptides was identified. Using MM/PBSA, the comparison with experiments shows that correlations between experimental $-pIC_{50}$ and absolute binding affinities are different between the two models. Using different dielectric constants in the MMPSBA and MMGBSA calculations also yield different results. It is notable that it is hard to estimate the energy of binding using the MM/PBSA method, as many molecular mechanisms outside the simplified molecular dynamic model influence the binding. However we found that using the truncated model, MMPBSA and a dielectric constant of 2, the estimation of the binding energy was correlated to the experiment using the Pearson test. This work highlights the fact that models and parameters are not directly transferable to other peptide/receptor systems and that any given system would require an adjustment phase, where identifying the best model to describe the binding requires experimental support. This would allow to identify the best dielectric constant for MM/(GB)PBSA calculations. However, once the parameters used in these calculations are refined experimentally, further *in-silico* work such as the estimation of the binding energy gain upon mutating the peptide sequence could be made more reliable.

## 6.7 Methods

### 6.7.1 Model building

The peptide/protein complex structures for the IL-1R1 stapled peptides complexes were obtained by modification of the crystal structure with PDB identifier 1G0Y. Stapled residues were mutated using an in-house script based on AmberTools, using our in-house staple residue libraries described in Chapter 3. This step generally generates high energy conformations, as only one side chain rotamer is present (even for the natural amino acids). Thus, atoms of side chains in stable residues were relaxed using a geometry optimisation. Each peptide was prepared using the full length model and the truncated model.

### 6.7.2 MD simulation and trajectory analysis

The PMEMD module of AMBERSB20 with the ff14SB forcefield was used to carry out the simulations. The system was solvated with water in an octahedron box using the transferable inter-molecular potential three-point (TIP3P) water model, extending 10 Å in every direction around the solute. Counter ions were added to the solvent to keep the system neutral. The geometry of the system was minimised in two steps before the MD simulation was run. First, the water molecule positions were refined through 2500 steps of steepest descent, followed by 2500 steps of the conjugate gradient, keeping the protein fixed with a constraint of 2.0 $kcal.mol^{1}.Å^{-2}$. The IL-1R/peptide complexes were then relaxed by 5000 cycles of steepest descent and 5000 cycles of conjugate gradient minimisation. During the simulation, the particle mesh Ewald method was employed to calculate the long-range electrostatic interactions, while the SHAKE method was applied to constrain all covalent bonds involving hydrogen atoms to allow the time step of 2 fs. A 10 Å cutoff value was used for the non-bonded interactions. The whole system was then heated from 0 K to 300 K running 50 ps molecular dynamics simulation with position restraints at constant volume. An isothermal isobaric ensemble (NPT) simulation was performed for 500 ps to adjust the solvent density followed by 500 ps of constant pressure equilibration at 300 K without constraints to relax the system. The production dynamics at constant pressure achieved lengths of 500 ns, of which snapshots saved at 10 ps intervals were used for further analysis.

### 6.7.3 Trajectory analysis

Trajectories were processed using cpptraj from AmberTools. The C$\alpha$ of the d1 and d2 subunits of IL-1R1 (residues 1 to 199) were used for a global alignment using cpptraj, and periodic boundary conditions (PBC) were corrected. After globally aligning the structures by their C$\alpha$ carbons, changes in the peptide and protein conformations were assessed via K-means clustering of all snapshots to 10 clusters using the pyEMMA Python module. We used the peptide

backbone atom RMSD as cluster criterion for peptide binding poses. Representative structures were selected for subsequent visual inspection.

Trajectories were aligned using IL-1R1 C$\alpha$. Then a Python script was used on the pre-aligned trajectories to calculate the different descriptors: MDAnalysis was used to calculate the root mean square distances (RMSDs) of the peptide backbone and of side chain atoms in residues Trp6, Glu8, Arn10, Tyr12, Tyr13 and Trp14. RMSF values for the peptide C$\alpha$ and the side chains of the key residues were computed separately and averaged on the full MD trajectory. For the RMSD and RMSF calculations no alignment was used during the calculation, the calculations relied on the pre-alignment of the d1/d2 subunits. The values obtained are thus RSMD values of the peptide atoms relative to IL-1R.

The helical content of the peptides was calculated using the MDtraj DSSP assignment module. Additionally, hydrogen bonding between protein and peptide was characterized using Mdtraj default criteria for angles and distances. A minimum occupancy cut-off of 5.0 % for the analysis of recovered polar contacts was applied. Hydrophobic contact residues were extracted from the trajectory by applying a maximum distance between heavy atoms of 4.0 Å. Native contacts were extracted from the equilibrated structure, and were counted if present in 60 % of the snapshots. Hydrogen bonds were defined by a distance cutoff of 3.2 Å and an angle cutoff of 120°.

### 6.7.4 Energetic analysis

Solvation free energies were calculated by either solving the linearised Poisson Boltzman or Generalized Born equation for each of the three states (this gives the electrostatic contribution to the solvation free energy). $\Delta G_{vacuum}$ was obtained by calculating the average interaction energy between receptor and ligand. As the peptides had similar sequences and binding poses, the normal mode analysis was neglected. The average interaction energies of receptor and ligand were obtained by performing calculations on an ensemble of snapshots collected from the molecular dynamics (MD) simulation. 500 ns simulations were performed and snapshots were extracted every 10 ns and the binding free energies were averaged over the ensemble of conformers produced. MM/PBSA.py was used to carry out the calculations. The ion concentration was set to 0 in agreement with the simulation set-up. The binding free energies were then calculated for the 24 complexes in the dataset using a dielectric constant of 1 (default), 2, 4 or 6. The atomic and solvent probe radii were set at 1.4 Å.

#### FRET assay

Peptides were dissolved in DMSO at 10 mM. From this initial concentration, a 10 point half log dilution series of the peptides in DMSO was dispensed in a PerkinElmer™ White 384-well Opti plate (Cat#607290) using a Tecan D300E. A solution containing 10 nM IL-1$\alpha$-AF647, 10 nM IL-1R1 and 10 nM IL-1R3-Tb was prepared and 30 $\mu$L of this solution was

added to each well of the Opti plate, resulting in a final peptide dilution series for the assay. The top peptide concentration was 250 $\mu$M for all peptides but CLICK-2-1 and CLICK-2-2, where the top peptide concentration was 25 $\mu$M. The plate was sealed and left to incubate on a horizontal plate shaker for 23-25 hours while protected from light. Plates were spun down at 800 g for 2 minutes before being read on a Pherastar FSX Plate Reader (BMG Labtech™) using the Homogeneous Time Resolved Fluorescence (HTRF) module, with an optical height of 10.9 mm. The emitting wavelength of the laser was set at 330 nm. A FRET ratio was generated by normalising the emissions recorded at 620 nm and 665 nm. The percentage inhibition of the peptide at each concentration point was determined by using equation %inhibition = $1 - (P_{obs} - P_{min})/(P_{max} - P_{min})$, with the values of $P_{max}$, $P_{min}$, and $P_{obs}$ in the equation refer to the polarisation of the wells containing the peptides and IL-1R, the polarisation of the free peptide, and the observed polarisation for the wells containing the inhibitors at a range of concentrations under the assay conditions. The $IC_{50}$ of an inhibitor was determined from the plot of %inhibition against inhibitor concentration analysed using GraphPad Prism 6.0 software using a standard 4-parameter logistics fit plotting FRET ratio against concentration on a log scale.

**Cell assay**

HEK-Blue™ TNF-$\alpha$/IL-1$\beta$ cells were purchased from InGenVivo (SanDiego). Cells were frozen at -80 °C. After thawing, cells were maintained in DMEM cell medium containing 4.5 g/L glucose and supplemented with 10% fetal bovine serum and 1% L-glutamine. 35 ml of cells in suspension were incubated at 37 °C in a culture flask (T175), in a 5% $CO_2$-containing, humidified incubator. The medium was removed and rinsed twice with PBS. Cells were treated with trypLE™, a cell dissociation agent, and re-suspended in fresh medium. Cells were then counted and diluted to produce cell suspensions at 5k cells/well (50k cells/ml) (24 h assay) and 10k cells/well (100k cells/ml) (2 h assay). From the 16 peptide solutions diluted at 10 mM (except the positive control peptide, diluted at 3 mM) in DMSO, peptides were then progressively diluted to 200 x final concentrations and dispensed into two PerkinElmer™White 96-well plates to give a 10 point half log dilution with a top concentration of 6 mM. The controls for this assay consisted of a positive control peptide the AF10847, a 21-mer peptide known to inhibit IL-1R1 (dilutions starting from 3 $\mu$M) for this peptide). Peptides were then re-diluted into a second PP 384 well plate using 2 $\mu$l of peptide concentration and 78 $\mu$l cell media.

For the 2 h incubation assay, 10 µL of the peptide concentrations plates, 20 µl 10 K/well HEK cells and 20 µL of 6 pM IL-1$\beta$ or 8 pM of IL-1$\alpha$ were dispensed into a clear 384 wellT/C plate, and incubated for 2 h at 37 °C 5%$CO_2$. For the 24h incubation assay, 10 µL of the peptide concentration plates, 20 µl 5K/well HEK cells and 20 µL of 6 pM IL-1$\beta$ or 8 pM of IL-1$\alpha$ were dispensed into a clear 384 well T/C plate, and incubated for 24 h at 37 °C.

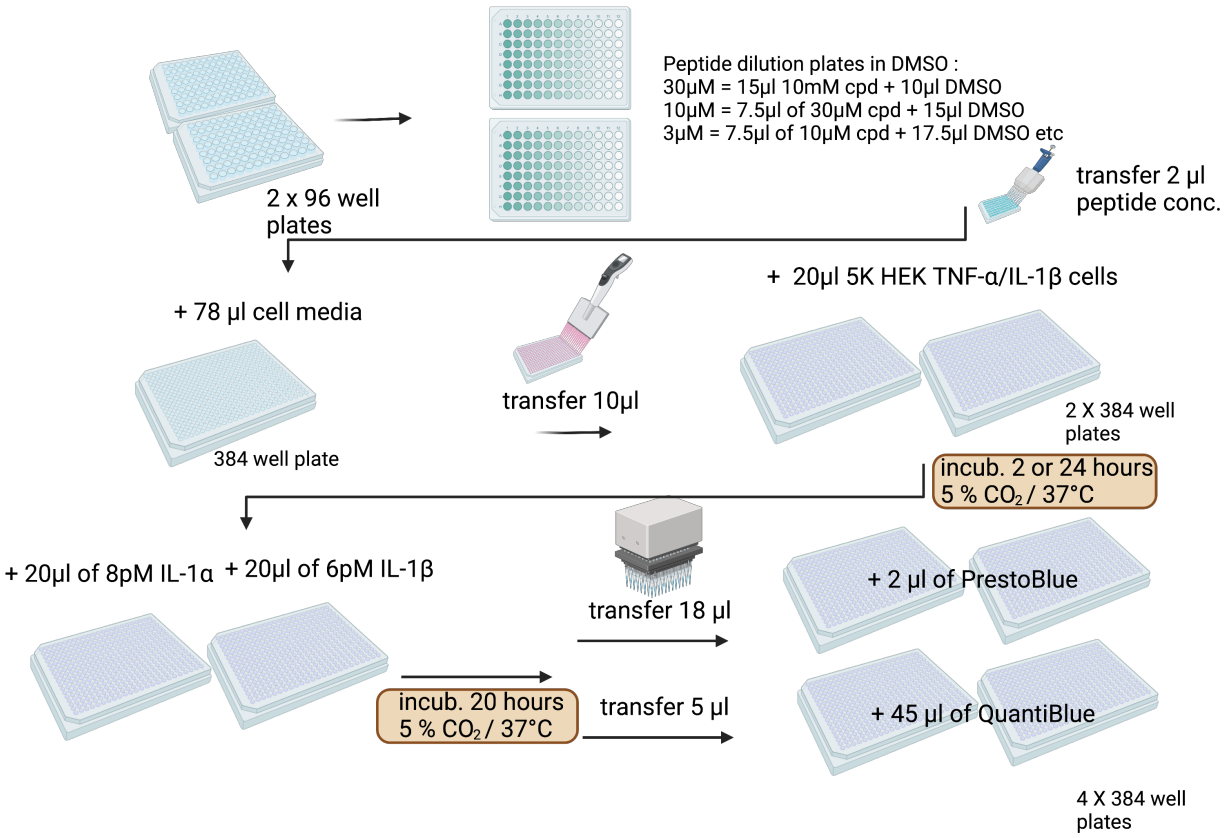



**Figure 6.16:** Methods: Procedure used for the cell assay, created using BioRender™

The plates were further incubated for 20 h at 5%$CO_2$ and 37 °C.

### Viability assay (PrestoBlueCell™ Viability Reagent)

The plates were copied (18 $\mu$l of each well was transferred into a PP 384 well plate) to conduct the viability assay. After the PrestoBlue was warmed up at room temperature, 1/10th volume was added onto the wells. Reading was done at wavelength 580 nm.

### Plate reading (QuantiBlue™)

The QuantiBlue reagent and QuantiBlue buffer were mixed at 1:1 and diluted in sterile water at 2 %. The solution was homogenised using a vortex. A plate was prepared with 180 $\mu$l of QuantiBlue™solution per well into a flat-bottom 96-well plate and 20 $\mu$L of the incubated plates was added. Plates were incubated for 3 hours before reading. The optical density (OD) was measured at 625 nm using a the Pherastar FSX(BMG Labtech™) microplate reader.

The $IC_{50}$ of an inhibitor was determined from the plot of %inhibition against inhibitor concentration on a log scale, which was analyzed using GraphPad Prism 6.0 software using a standard 4-parameter logistic fit. The curves of RCM-RS-3, CLICK-2-1, CYS-P-1 and CYS-M-1 were constrained to 0 and 100% to counter the fact that the peptides did not reach 100 % inhibition at the highest concentration.

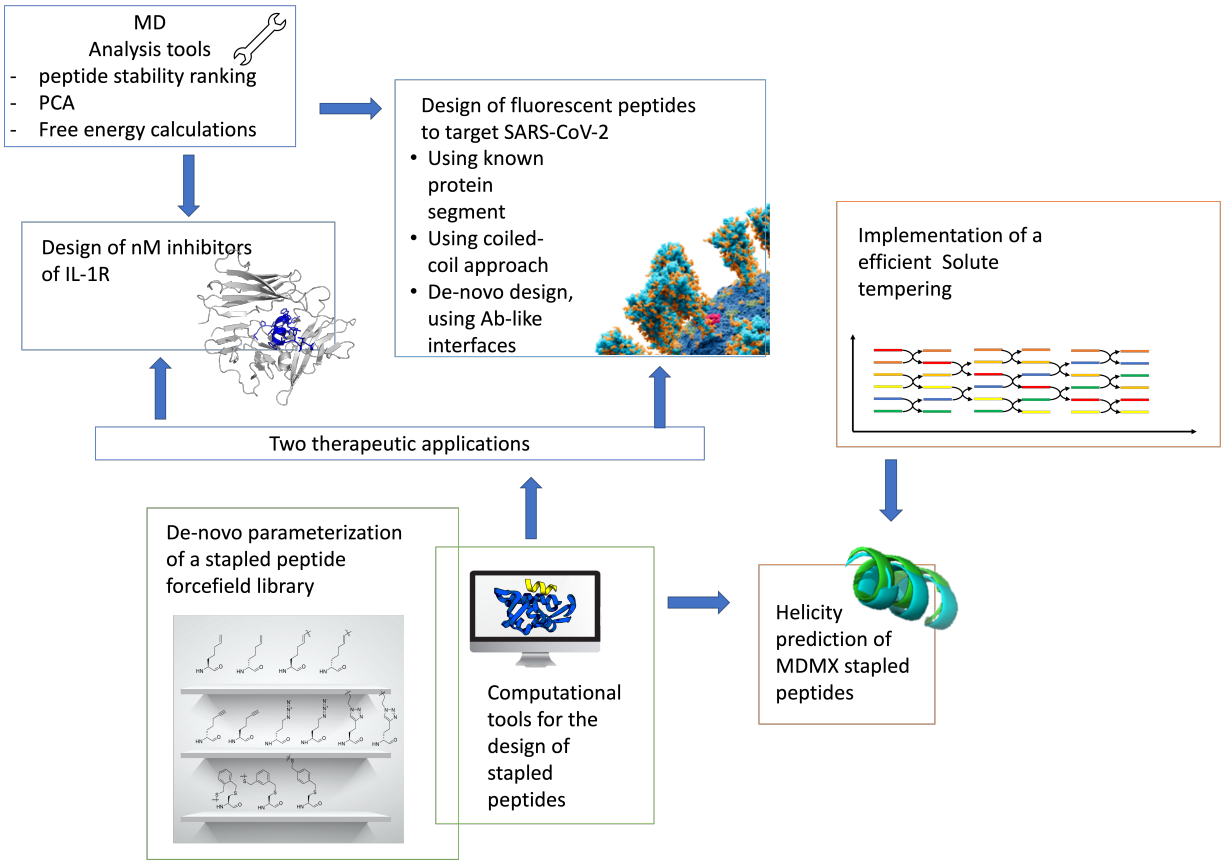
# 7

# Conclusions

With the steady improvement of algorithms and software, and in parallel, the prevailing increase of high-performance computing resources, the use of computational methods such as Molecular Dynamics is a promising approach to better comprehend protein-protein interfaces, binding mechanisms, as well as design and predict physical properties of compounds and peptides. This work has explored a wide range of computational methods applied to the design and properties predictions of stapled peptides.

We developed a new approach to produce molecular dynamics forcefield parameters and applied this method to produce add-on libraries for stapled peptide residues. Before this work, no add-on libraries for staple residues had been reported elsewhere. The new set of parameters reproduced quantum mechanical energy variation among the newly parameterised residues and have been used to model various stapled peptides in Chapters 3-6. The pipeline can readily be used to parameterise other stapled residues, which could help the modelling of novel stapled chemistries. Further work could consider the use of Cmap correction terms in the pipeline to improve agreement with QM data. Such terms have already been introduced in the most recent version of the AMBER19SB forcefield.

As seen in Chapter 1, a popular approach for designing a stapled peptide is to excise an interacting protein motif segment from a known protein-protein interaction to mimic the original binding mode. For the design to be successful the peptide should readily adopt a conformation that resembles the binding mode observed in the protein complex. This is more readily achieved when the binding motif is enriched in specific secondary structures such as $\alpha$-helices or $\beta$-strands. Peptides with a high $\alpha$-helical content have generally been linked to a better affinity as well as higher cell permeability. Producing peptides that are already 'pre-folded' in their binding conformation has also been suggested to reduce binding entropy, thus increasing affinity and availability. However, factors inducing this helicity are still not well understood, as adding a staple to a peptide does not systematically increase its helicity. Data-learning conformation prediction tools for the helicity of stapled peptides have not been developed as the number of published stapled peptides is still too scarce. In this work, we thus explored molecular dynamics simulation methods to allow the rapid exploration of the conformational space of peptides. We implemented a Solute tempered method and used the established T-REMD method as a reference. The REST2 method implemented was found to perform well at exploring the conformational space of peptides, as compared to the REMD




method. It was also shown to be able to reproduce the helicity of the well-described MDMX peptides with the added advantage of being computationally less demanding. This method used fewer replicas, thus needed less processor units and could be used on GPU instead of CPU, which conjointly were the main reasons for this computational gain. The fact that the REST2 method was able to correctly predict the secondary structure of peptides also demonstrated the high quality of the forcefield parameters developed for the unnatural amino acids. It is notable that the Solute tempering method developed in this work is readily applicable to study other slow occurring conformational changes in proteins.

Targeting PPIs has become an important research area in drug design because PPIs significantly expand the target-able space. However, protein-protein interfaces are eminently different from classical ligand-binding sites, which can make the development of small molecules difficult. The most important aim of PPI research is the development and design of novel PPI inhibitors. In this field, MD has become a prevalent tool as it complements experimental screening techniques. In combination with other computational techniques, MD simulations are of great value for the characterisation of protein-protein interfaces, and for detecting and characterising inter-molecular interactions between a ligand and its binding site. In this work we aimed to build computational tools to aid the search and validation of peptide binders. Here we propose tools and insights to design stapled peptides from known peptide binders, from sequences excised from a protein, from known PPIs and using a coiled-coil approach. In

MD multiple peptide-protein conformations can be generated and their stability can ranked. Binding sites on protein surfaces can be characterised in terms of flexibility, and the key interactions between the initial PPI can be identified.

In this work, we attempted to assess the binding of peptides through quantitative measurements in MD simulations. Some descriptors such as RMSD offered valuable insights on the stability of the models. However, separately these did not yield efficient ranking of the peptides in simulations. These descriptors needed to be analysed in more global and comprehensive ways and tailored analysis methods, using different metrics are required to capture the observations made from MD simulations. Molecular dynamics simulations combined with information theory-based analysis, such as PCA, have provided key insights into the translation of MD simulation derived quantities (RMSD, RMSF, Rg...) to peptides binding stability.

In Chapter 5 we developed tools to design fluorescent peptides binding to the spike protein, which could be used as rapid detection agents. As we screened over 200 sequences, the main issue was the ranking of this large number of simulations. The PCA method successfully separated the peptides according to observations with all stable peptides being guessed correctly in the test set. This study provided key insight into descriptors useful for peptide scoring and the potential of PCA for analysis of a large number of simulations. Howerer CONA assays suggested that the tested peptides had $IC_{50}$ higher than 10 µM. Furthermore, some of the tested peptides displayed an unexplained and unprecedented affinity for the agarose beads used in the CONA assay.

We designed peptides with enhanced activity against IL-R1. One of these peptides, CLICK-1-2 demonstrated an $IC_{50}$ of 20 nM, which is a 9-fold increase in comparison to the native peptide. Furthermore, these stapled peptides were found to improve the inhibition of IL-$\alpha$/IL1-R1 triggered cell response. Five peptides showed a $IC_{50}$ lower than the wild type peptide in the cell assay and CLICK-1-2 demonstrated an $IC_{50}$ of 10 nM, which is 50-fold lower then the wild type peptide and 10 fold lower than the positive control.

Interestingly, in the simulations of IL-1R1 and the SARS-CoV-2 peptides similar descriptors of the peptide binding were calculated. We noticed that these descriptors displayed less variability in the IL-1R study than in the spike protein study. For example a descriptor which was calculated for both systems is the RMSD of the peptide backbone, the standard deviation for the IL-R1 simulations for this descriptor was only 3 Å vs 7 Å for the spike protein ACE2 derived stapled peptides. This might reflect the fact that all peptides tested against IL-1R inhibit to some extent the formation of the IL-1R/Il-1$\alpha$ at concentrations $< 5\mu$M.

The most accurate ranking of the affinity of the IL1-R1 peptide was obtained using energetic analysis. We noticed that discrepancies were observed using different models of the peptide bound to IL-1R and different dielectric constants in MMPBSA and MMGBSA, which reflects the need to tailor models, methods and parameters for every system studied. However this setup for IL-1R could readily be used to guide further design improvements of the stapled peptides through mutations of the native sequence.

In conclusion, this work was designed to provide tools to facilitate the development of forcefield parameters for unnatural amino acid and thus extend current molecular dynamic forcefield libraries, predict secondary structure when *ab-initio* methods cannot be applied, and efficiently sample thermodynamically stable conformations for any given peptide. We also developped tools for peptide design based on protein-protein interfaces with direct application for the treatment of rheumatoid arthritis and COVID-19 labelling. We used different methods to score peptide binding, relying on MD simulation analysis, PCA methods as well as MM/PB(GB)SA methods.

Future work in this area should aim to extend the forcefield libraries, as well as increasing the number of models studied in helicity predictions. This will lead to a better understanding of the specific folding properties of stapled peptides. Further use of the design tools developed in this project, all made available on github (michellab/stapline), could help the design of new stapled peptide drug candidates for other PPIs.

# A

# Chapter 1

## A.1 Analytical Methods

Analysis for Figure 1.6: The crystal structures of a total of 65 stapled peptides were extracted from the PDB (as identified in Table A.1). The stapled peptide chains were extracted from the crystal structures and converted to single-letter code sequences using MDTraj.[420] The frequency of occurrence of each amino acid was then calculated with an in-house Python script and was expressed as a percentage of the total number of amino acids in this stapled peptide database. Analysis for Figure 1.7: The crystal structures of a total of 52 helical stapled peptides in complex with protein were extracted from the PDB (as identified in Table A.1) and were analysed in Pymol using an in-house Python script.[74] The stapled-peptide was removed from the crystal structure, and the amino acid residues at the protein interface were defined as those with an increase in solvent-accessible surface area (SASA) in comparison with the original crystal structure. The protein surface hydrophobicity was defined as the sum of the SASA of each amino acid present at the binding surface multiplied by its Eisenberg hydrophobicity value (Table A.1 ),144 divided by the total area of the binding surface (Table A.1). The staple angle was defined as follows: the centre of mass (COM) of the protein, the COM of the "stapled" residue sidechains, and the COM of the stapled peptide helix were computed. The COM of the protein and the COM of the stapled residue sidechains were then projected onto the plane defined by the vector parallel to the peptide helix axis and the COM of the stapled peptide helix (Figure 1.7B (i)). This step ensured the three COM were placed in a plane orthogonal to the binding surface. The angle between the "projected" COM of the protein, the COM of the helix and the "projected" centre of mass of the staple was defined as the staple angle $\theta$. (Figure 1.7B (ii)).

**Table A.1:** Eisenberg values

| Code | Ile | Phe | Val | Leu | Trp | Met | Ala | Gly | Cys | Tyr |
|---|---|---|---|---|---|---|---|---|---|---|
| Value | 0.73 | 0.61 | 0.54 | 0.53 | 0.37 | 0.26 | 0.25 | 0.16 | 0.04 | 0.02 |
| Code | Pro | Thr | Ser | His | Glu | Asn | Gln | Asp | Lys | Arg |
| Value | -0.07 | -0.18 | -0.26 | -0.40 | -0.62 | -0.64 | -0.69 | -0.72 | -1.10 | -1.80 |

**Table A.2:** Stapled peptides extracted from the PDB

| Protein | PDB code | Peptide name | Chemistry | Chain in PDB | Staple Residue PDB Code | Angle (°) | Surface area ( Å$^2$) | Reference |
|---|---|---|---|---|---|---|---|---|
| MDM2 | 6H22 | 28 | double-click | C,D | FL5 | 80 | 1665 | [23] |
| | 6AAW | dPMI-$\delta$(5-12) | RCM/i,i+7 position | B | 0EH, MK8 | 62 | 717 | [448] |
| | 6KZU | dPMI-$\delta$(1-5,9-12) | RCM/i,i+4 position (2 staples) | B | 2JN,DAL,MK8 | 170, 160 | 1058 | [449] |
| | 3V3B | SAH-p53-8 | RCM/i,i+7 position | C,D | MK8 | 47 | 1077 | [42] |
| | 4UMN | M06 | RCM/i,i+7 position | C,D | 0EH, MK8 | 43 | 2489 | [420] |
| | 5AFG | E1 | CuAAC cycloaddition “Click-reaction” | B | PO7 | 69 | 1190 | [138] |
| | 4UE1 | YS-1 Stapled peptide | RCM/i,i+4 position | G,H,J,F | 2JN | 62 | 1730 | [86] |
| | 4UD7 | YS-2 Stapled peptide | RCM/i,i+4 position | G,H,J,F | 2JN | 66 | 1731 | [86] |
| | 5XXK | M011 Stapled peptide | RCM/i,i+4 position | C,D | OEH, 6CW | 45 | 1054 | [450] |
| | 5VK0 | PMI(8,12) | thioester | X,D,L,F,J,H,T,V | 9E7,CYS | 120 | 1224 | [380] |
| | 5VK1 | PMI(4,8) | thioester | X,D,L,F,J,H,T,V | 9E7,CYS | 63 | 1252 | [380] |
| MCL-1/BCL-2 | 3MK8 | MCL-1 SAHBD | RCM/i,i+4 position | B | MK8 | 82 | 1199 | [451] |
| | 5C3F | BID-MM | RCM/i,i+4 position | B | NLE | 102 | 1484 | [452] |
| | 5C3G | BIM-MM | RCM/i,i+4 position | B | NLE | 143 | 1329 | [452] |
| | 5W89 | SAH-MS1-18 | RCM/i,i+4 position | B | MK8 | 57 | 1544 | [116] |
| | 5W8F | SAH-MS1-14 | RCM/i,i+4 position | B | MK8 | 71 | 962 | [116] |
| | 5WHH | D-NA-NOXA SAHB Stapled peptide | RCM/i,i+7 position | B | MK8,0EH | 125 | 1384 | [453] |
| Estrogen receptor | 2YJD | SP1 | RCM/i,i+4 position | C,D | MK8 | 66 | 824 | [437] |
| | 2YJA | SP6 | RCM/i,i+4 position | A | MK8 | 66 | 1093 | [437] |
| | 2LDA | SP2 | RCM/i,i+4 position | A | MK8 | not in complex | not in complex | [437] |
| | 2LDC | SP1 | RCM/i,i+4 position | A | MK8 | not in complex | not in complex | [437] |
| | 2LDD | SP6 | RCM/i,i+4 position | A | MK8 | not in complex | not in complex | [437] |
| | 5DXB | SRC2-SP1 | RCM/i,i+4 position | C,D | MK8/66D | 80 | 1140 | [454] |
| | 5HYR | SRC2-SP2 | RCM/i,i+4 position | G,F | 66D,MK8 | 55 | 1003 | [454] |
| | 5DX3 | SRC2-SP3 | RCM/i,i+4 position | C,D | 5GM,MK8 | 57 | 1294 | [454] |
| | 5DXE | SRC2-SP4 | RCM/i,i+4 position | C,D | MK8 | 67 | 1236 | [454] |
| | 5DXG | SRC2-SP5 | RCM/i,i+4 position | C,D | MH8 | 68 | 1434 | [454] |
| | 5WGD | SRC2-LP1 | RCM/i,i+4 position | C,D | LYS,ASP | 112 | 1691 | [12] |
| | 5WGQ | SRC2-BCP1 | Cross-stitch (olefin & lactam) | C,D | LYS,ASP, MK8 | 93,175 | 1100 | [12] |
| | 6PIT | SRC2-41A | lactam | D | LOU,LYS | not helical | 730 | [455] |
| Aurora-A | 5LXM | Stapled TPX2 peptide 10 | RCM/i,i+4 position | D | MK8 | 135 | 3602 | [456] |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Tankyrase-2 | 5BXO | Cp4n2m3 | RCM/i,i+4 position | C,D | ALA,4XP | not helical | 1280 | [97] |
| | 5BXU | Cp4n4m5 | Double-click Cycloaddition reaction/ i,i+4 position | B | ALA,4XP | not helical | 1279 | [97] |
| GRb7 | 5EEL | G7-B4 peptide | RCM+ Thioether/bicyclic peptide | K,G,J,H,L,I | SER,73C | not helical | 59083 | [132] |
| | 5EEQ | G7-B1 peptide | RCM+ Thioether/bicyclic peptide | K,G,J,H,L,I | SER,73C,48V | not helical | 24441 | [132] |
| $\beta$-catenin | 4DJS | aStAx-35 | RCM/i,i+4 position | B | MK8 | 176 | 1582 | [457] |
| hDcn-1 | 3TDZ | Cul1WHB-Dcn1P-stapled Acetyl(5:9 Staple) | RCM/i,i+4 position | E,F | MK8 | 128 | 1363 | [458] |
| ks-vFLIP | 5LDE | spIKKɣ-Stapled peptide | RCM/i,i+4 position | D | 6ZS | 101 | 1202 | [459] |
| human IgG1 Fc | 5U66 | LH1 | Bi-component lactam | B | 85J,85G | 141 | 1543 | [460] |
| CaV $\beta$ subunit | 5V2P | AID-CAP Stapled peptide | m-xylyl linker macrocyclization/ i,i+5 position | B | 8VY,CYS | 123 | 2994 | [461] |
| | 5V2Q | AID-CEN Stapled peptide | m-xylyl linker macrocyclization/i,i+4 position | B | 8VY,CYS | 138 | 1662 | [461] |
| NCOA1 | 5Y7W | YL-2 | RCM/i,i+4 position | C,D | MK8 | 162 | 1417 | [142] |
| Ctf4 (S. cerevisiae) | 5NXQ | Sld5 CIP A2 | Double-click Cycloaddition reaction (CuAAC)/i,i+6 position | D,E | 9G2,9FZ | 175 | 1478 | [462] |
| HIV-1 | 4NGH | SAH- MPER(671-683KKK)(q)pSer | RCM/i,i+4 position | C | DIV,MK8 | 122 | 1573 | [463] |
| | 4NHC | SAH-MPER(671-683KKK)(q) | RCM/i,i+4 position | C | DIV,MK8 | 123 | 1575 | [463] |
| | 4U6G | SAH-MPER(662-683KKK)(B,q) | RCM/i,i+4 position | E,F | DIV,MK8 | 94 | 1131 | [463] |
| | 2L6E | NYAD-13 | RCM/i,i+4 position | B | MK8 | 172 | 1099 | [464] |
| Zebrafish MDM2/X | 4N5T | ATSP-7041 | RCM/i,i+4 position | B | 0EH, MK8 | 47 | 990 | [43] |
| Myo A tail interacting protein (MTIP) | 4MZJ | pGly[801–805] | RCM/i,i+4 position | B | NLE | 60 | 3339 | [465] |
| (P. falciparum) | 4MZK | pGly[807-811] | RCM/i,i+4 position | B | NLE | 160 | 2856 | [465] |
| | 4MZL | HSB myoA | lactam | C,D | BUA,ASN | 66 | 2925 | [465] |
| eIF4E | 4BEA | sTIP-04 Stapled peptide | RCM/i,i+4 position | B | MK8 | 120 | 1279 | [123] |
| | 5ZJY | sTIP-05 peptide | RCM/i,i+4 position | B | 2JN | 108 | 1825 | [466] |
| | 5ZJZ | sTIP-07 | RCM/i,i+4 position | B | MK8 | 123 | 1763 | [466] |
| | 5ZK9 | sTIP-08 | RCM/i,i+4 position | B | MK8 | 124 | 2478 | [466] |
| | 5ZML | sTIP-09 | RCM/i,i+4 position | B | MK8 | 125 | 1851 | [466] |
| | 5ZK5 | sTIP-10 | RCM/i,i+4 position | B | MK8 | 150 | 1350 | [466] |
| | 5ZK7 | sTIP-14 | RCM/i,i+4 position | C,D | MK8 | 105 | 2542 | [466] |
| NF-Y Transcription Factor | 6QMQ | 2-DN | RCM/i,i+4 position | C | MKD,ABA | 149 | 1441 | [466] |
| | 6QMS | 2-C | RCM/i,i+7 position | C | MKD,J7H | 164 | 1391 | [466] |
| N.A. (Coiled coil) | 4HU6 | GCN4-p1 | Oxime Cross-Links | A,B,C,D | UU5,J7H | not in complex | not in complex | [32] |
| CK2$\alpha$/CK2$\beta$ | 6Q4Q | CKβ-RAD | click | C,D | HEK | not helical | 669 | [467] |
| GRb7 | 5D0J | G7-B1 | RCM/i,i+4 position | L,M | SER | not helical | 30370 | [132] |
| Cullin3 | 2MYL | Cul349-68EN | RCM/i,i+4 position | A | MK8 | not in complex | not in complex | [468] |
| | 2MYM | Cul349-68LA | RCM/i,i+4 position | A | MK8 | not in complex | not in complex | [468] |

# B

# Chapter 3

## B.1 RESP validation

| atom index | atom name | partial charges R.E.D server S enantiomer | R.E.D server R enantiomer | In house S enantiomer |
|---|---|---|---|---|
| 1 | HE | 0.1267 | 0.1248 | 0.0679 |
| 2 | C1 | 0.8513 | 0.7545 | 0.6267 |
| 3 | O1 | -0.6366 | -0.5802 | -0.5579 |
| 4 | C2 | -0.5144 | -0.3726 | -0.3704 |
| 5 | H2 | 0.1389 | 0.1053 | 0.1067 |
| 6 | H3 | 0.1389 | 0.1053 | 0.1067 |
| 7 | H4 | 0.1389 | 0.1053 | 0.1067 |
| 8 | CA | -0.0607 | 0.1139 | -0.0231 |
| 9 | O | -0.5895 | -0.5561 | -0.5855 |
| 10 | CB | -0.0234 | 0.0157 | 0.0427 |
| 11 | C | 0.7141 | 0.604 | 0.0442 |
| 12 | N | -0.6254 | -0.6576 | -0.5796 |
| 13 | CC | -0.0232 | 0.0073 | 0.0051 |
| 14 | CD | -0.0160 | 0.0593 | 0.732 |
| 15 | CE | -0.1056 | -0.1328 | -0.0836 |
| 16 | CF | -0.4941 | -0.4614 | -0.4769 |
| 17 | HA | 0.1014 | 0.0513 | 0.1038 |
| 18 | HB1 | 0.0313 | 0.003 | 0.0158 |
| 19 | HB2 | 0.0313 | 0.003 | 0.0158 |
| 20 | HC1 | 0.0271 | 0.0084 | 0.0164 |
| 21 | HD1 | 0.0479 | 0.0223 | 0.0213 |
| 22 | HC2 | 0.0271 | 0.0084 | 0.0164 |
| 23 | HD2 | 0.0479 | 0.0223 | 0.0213 |
| 24 | H | 0.3242 | 0.3264 | 0.3026 |
| 25 | N2 | -0.5273 | -0.5164 | -0.5137 |
| 26 | C10 | -0.2388 | -0.0708 | -0.0715 |
| 27 | H13 | 0.3459 | 0.3005 | 0.2987 |
| 28 | H14 | 0.1243 | 0.0805 | 0.0796 |
| 29 | H15 | 0.1243 | 0.0805 | 0.0796 |
| 30 | H16 | 0.1243 | 0.0805 | 0.0796 |
| 31 | HF1 | 0.1944 | 0.1827 | 0.1863 |
| 32 | HF2 | 0.1944 | 0.1827 | 0.1863 |

**Table B.1:** Comparison of partial charges for 2-acetamido-N-methylhept-6-enamide]

# C

# Chapter 5

## C.1 FRET assays

| n=1 all 250uM topF | Positive Control | 21mer WT | RCM-SS-1 | RCM-SS-2 | RCM-SS-3 | RCM-RS-3 | CYS-M-1 | CYS-M-2 | CYS-P-1 | CYS-P-2 | CLICK-1-1 | CLICK-1-2 | CLICK-1-3 | CLICK-2-1 | CLICK-2-2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bottom | -3.884 | -10.73 | -3.117 | -15.23 | -3.611 | -1.456 | -1.323 | -7.911 | -2.669 | -9.484 | -2.361 ~ -10.00 | | -6.585 | -6.907 ~ -10.00 | |
| Top | 106.6 | 109.2 | 112.4 | 108.3 | 109.8 | 106.5 | 107.8 | 110 | 113.8 | 108.4 | 105.1 | 105.6 | 107.5 | 110 | 106.4 |
| IC50 | 3.25E-08 | 1.68E-07 | 1.15E-05 | 9.61E-08 | 3.40E-06 | 5.18E-05 | 5.98E-06 | 7.12E-07 | 2.20E-05 | 3.07E-07 | 3.90E-07 | 1.98E-08 | 3.45E-07 | 1.56E-05 | 3.11E-08 |
| HillSlope | 0.8323 | 0.6622 | 0.7091 | 0.6867 | 0.7258 | 0.834 | 0.7592 | 0.6555 | 0.7221 | 0.707 | 0.7892 | 1.37 | 0.7073 | 0.6688 | 1.039 |
| logIC50 | -7.488 | -6.775 | -4.94 | -7.017 | -5.468 | -4.285 | -5.223 | -6.147 | -4.658 | -6.512 | -6.41 | -7.703 | -6.463 | -4.808 | -7.507 |
| Span | 110.5 | 120 | 115.5 | 123.6 | 113.4 | 107.9 | 109.2 | 117.9 | 116.5 | 117.9 | 107.4 ~ 115.6 | | 114.1 | 116.9 ~ 116.4 | |

| n=2 25*/250uM top | Positive Control | 21mer WT | RCM-SS-1 | RCM-SS-2* | RCM-SS-3 | RCM-RS-3 | CYS-M-1 | CYS-M-2 | CYS-P-1 | CYS-P-2 | CLICK-1-1 | CLICK-1-2* | CLICK-1-3 | CLICK-2-1 | CLICK-2-2* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bottom | -0.7895 | -4.493 | -2.65 | -4.106 | -4.791 | -5.03 | -4.393 | -6.259 | -5.027 | -8.004 | -2.749 | -7.161 | -5.958 | -8.525 | -5.662 |
| Top | 107.4 | 110 | 119.7 | 111.2 | 113.5 | 124.8 | 109 | 110.1 | 117.8 | 109.2 | 106.9 | 107 | 108.8 | 117.9 | 106.7 |
| IC50 | 1.53E-08 | 2.389E-07 | 2.105E-05 | 2.585E-07 | 4.147E-06 | 6.814E-05 | 5.535E-06 | 1.051E-06 | 2.651E-05 | 4.45E-07 | 5.246E-07 | 2.16E-08 | 4.944E-07 | 1.796E-05 | 3.147E-08 |
| HillSlope | 0.8635 | 0.7124 | 0.7269 | 0.7286 | 0.6707 | 0.6699 | 0.7964 | 0.6819 | 0.7263 | 0.7254 | 0.8686 | 1.125 | 0.7342 | 0.6518 | 0.9717 |
| logIC50 | -7.815 | -6.622 | -4.677 | -6.587 | -5.382 | -4.167 | -5.257 | -5.978 | -4.577 | -6.352 | -6.28 | -7.666 | -6.306 | -4.746 | -7.502 |
| Span | 108.2 | 114.5 | 122.3 | 115.4 | 118.3 | 129.8 | 113.4 | 116.4 | 122.9 | 117.2 | 109.7 | 114.2 | 114.7 | 126.4 | 112.4 |

| n=3 25*/250uM top | Positive Control | 21mer WT | RCM-SS-1 | RCM-SS-2* | RCM-SS-3 | RCM-RS-3 | CYS-M-1 | CYS-M-2 | CYS-P-1 | CYS-P-2 | CLICK-1-1 | CLICK-1-2* | CLICK-1-3 | CLICK-2-1 | CLICK-2-2* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bottom | 1.57 | -8.925 | 1.917 | -3.255 | 0.7053 | 1.122 | 1.62 | -6.825 | 1.573 | -7.566 | -6.072 | -5.153 | -2.042 | -4.955 | -2.532 |
| Top | 111.4 | 112.8 | 134.3 | 119.3 | 120 | 130.3 | 114 | 114.6 | 124.6 | 111.7 | 111.3 | 108.9 | 111.7 | 121.4 | 109.6 |
| IC50 | 1.997E-08 | 1.641E-07 | 5.041E-05 | 5.457E-07 | 9.648E-06 | 0.0000848 | 7.043E-06 | 1.089E-06 | 3.755E-05 | 4.246E-07 | 3.933E-07 | 2.045E-08 | 5.7E-07 | 1.621E-05 | 3.136E-08 |
| HillSlope | 0.7789 | 0.597 | 0.7099 | 0.6206 | 0.6539 | 0.7209 | 0.6996 | 0.5914 | 0.6786 | 0.6574 | 0.6663 | 1.011 | 0.6948 | 0.6007 | 0.8821 |
| logIC50 | -7.7 | -6.785 | -4.297 | -6.263 | -5.016 | -4.072 | -5.152 | -5.963 | -4.425 | -6.372 | -6.405 | -7.689 | -6.244 | -4.79 | -7.504 |
| Span | 109.8 | 121.7 | 132.4 | 122.6 | 119.3 | 129.2 | 112.4 | 121.5 | 123 | 119.3 | 117.4 | 114 | 113.8 | 126.3 | 112.1 |

**Table C.1:** $IC_{50}$ values for the 3 independents FRET assay

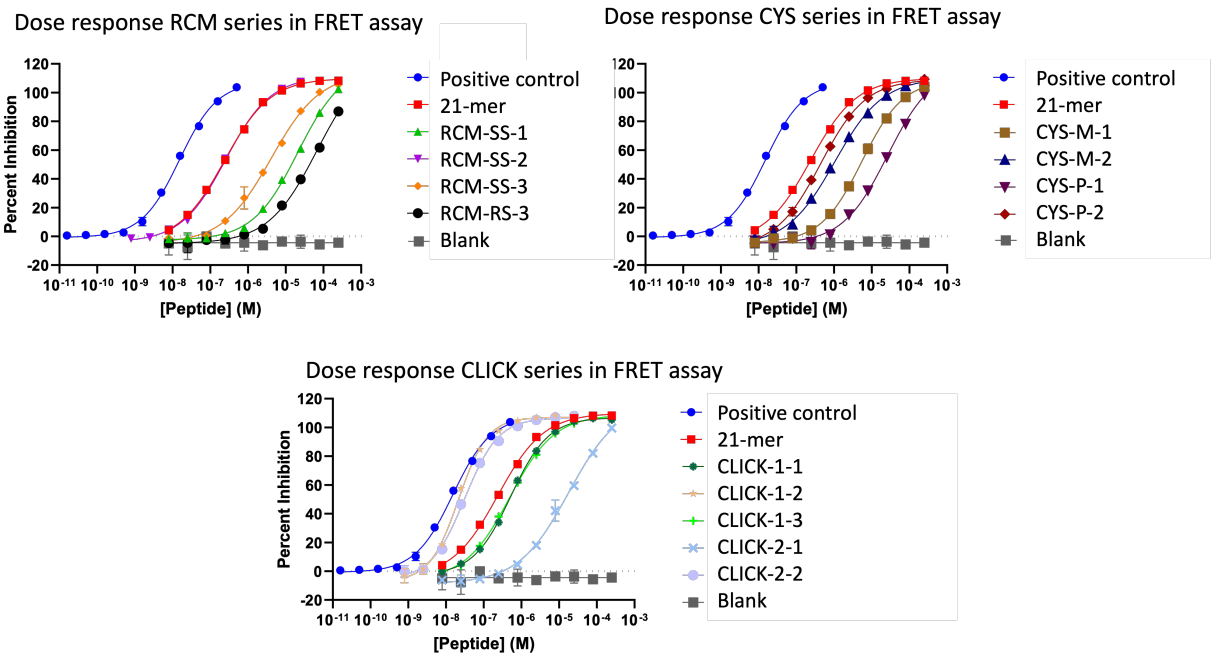



**Figure C.1:** FRET assay Dose response curve for the n=2

# C.2 Cell assays

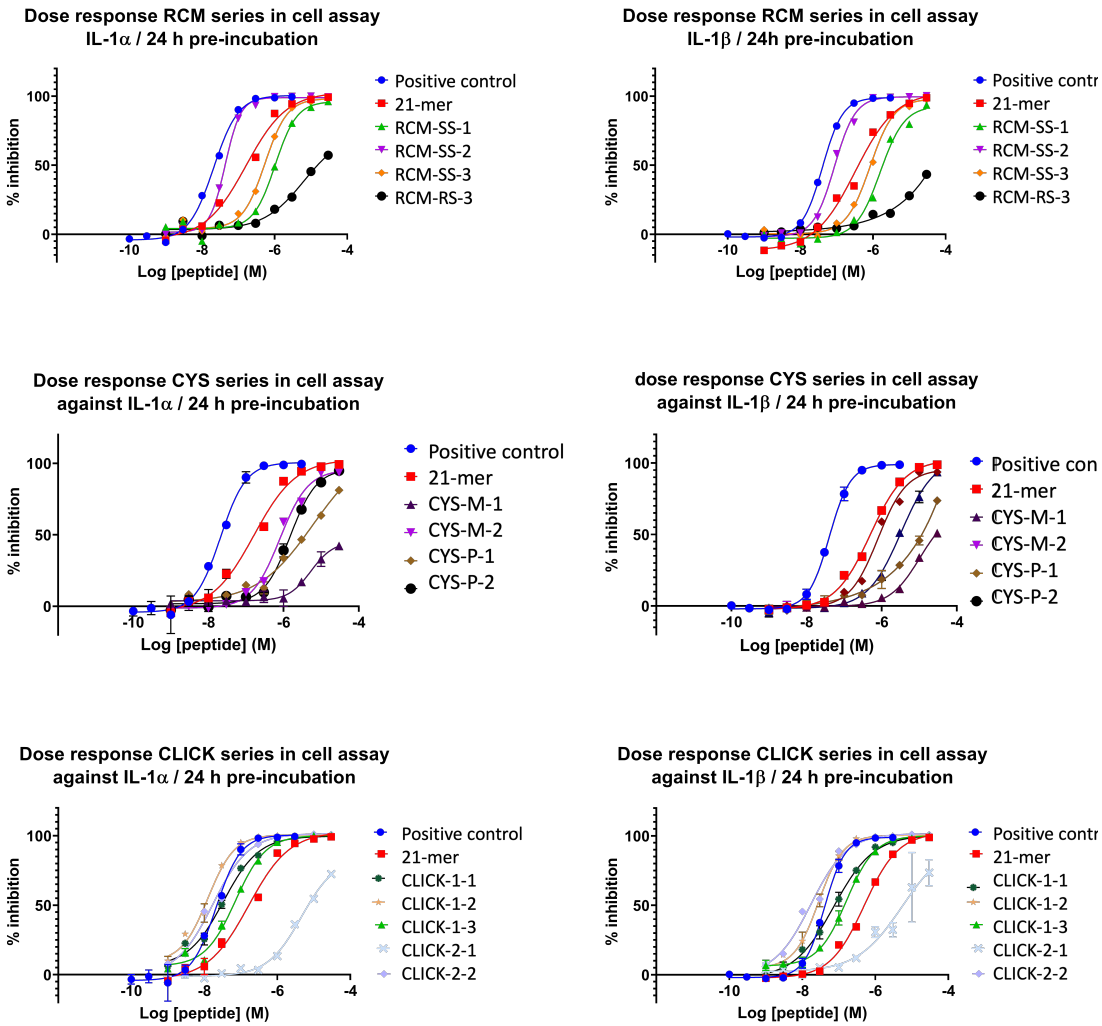



**Figure C.2:** Cell assay dose response curve for Il-1α 24h incubation

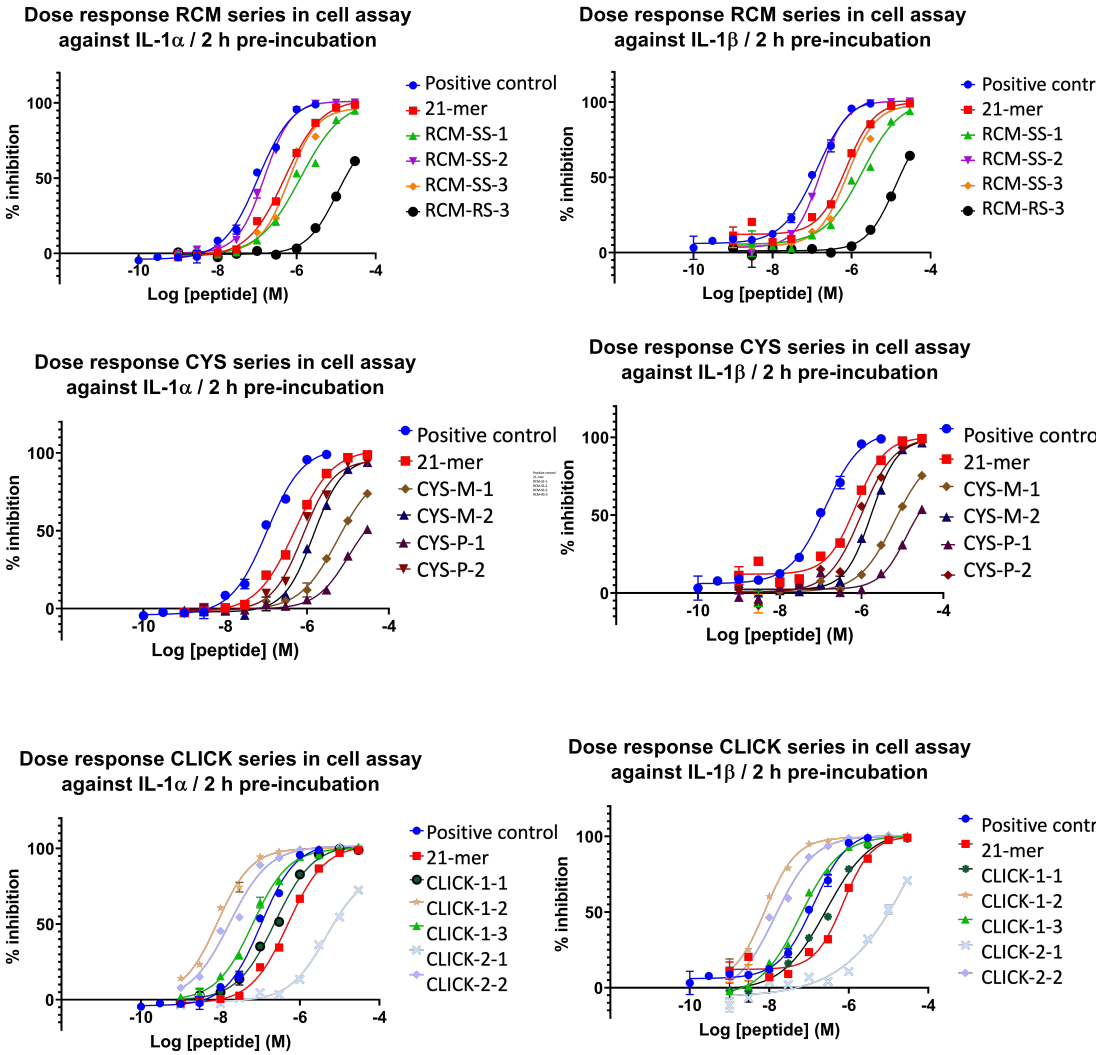



**Figure C.3:** Cell assay Dose response curve Il-1α 2h incubation

# D

# Chapter 6

## D.1 PCA analysis

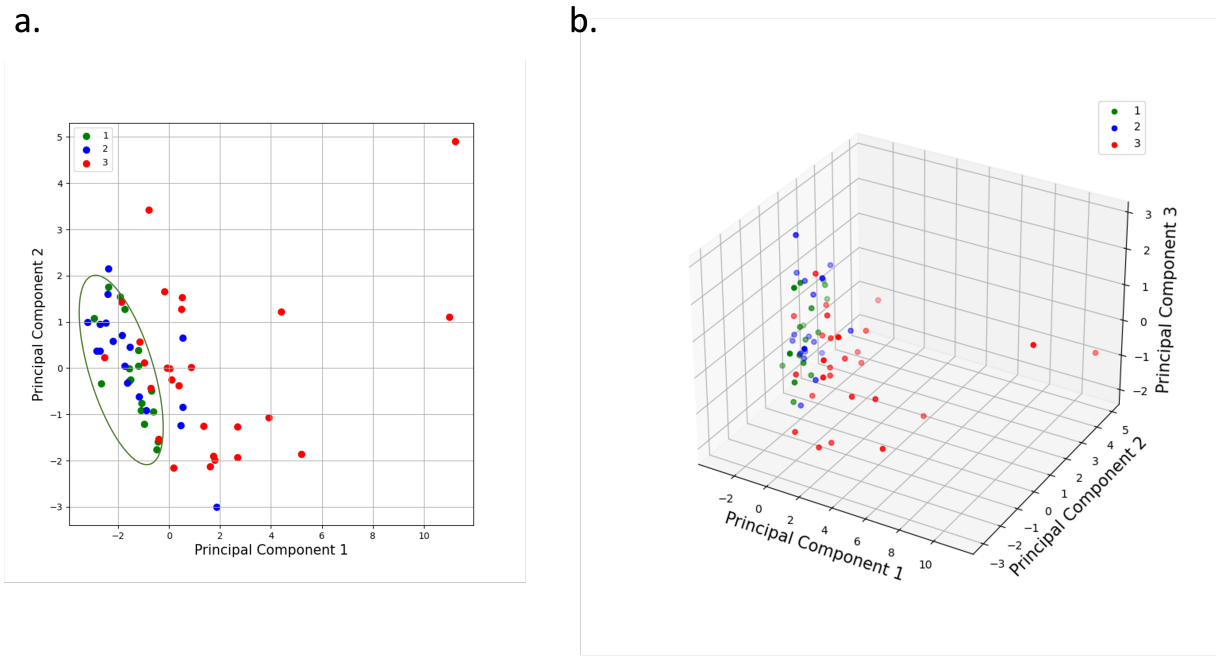



**Figure D.1:** PCA analyses for the ACE2 based peptides, using a. 2 components and b 3 components

# Bibliography

[1] Rene Thurmer. *Peptide Therapeutics: Strategy and Tactics for Chemistry, Manufacturing, and Controls*. Royal Society of Chemistry, 2019. 1

[2] Nagarajan Sangeetha, Palanisamy Selvamani, and Subbiah Latha. Emerging trends in therapeutic peptide pharmaceuticals: Prospects and perspectives. *Journal of Drug Delivery and Therapeutics*, 9(2-s):606–610, 2019. 1

[3] Chris Morrison. Constrained peptides' time to shine? *Nature Reviews Drug Discovery*, 17:531–533, 2018. 1

[4] V. A. Koivisto. Insulin Therapy in Type II Diabetes. *Diabetes Care*, 16(Suppl. 3):29–39, 12 1993. 1

[5] Lawrence T. Goodnough, Terri G. Monk, and Gerald L. Andriole. Erythropoietin therapy, 1997. 1

[6] Joseph Seitchik and Maria Castillo. Oxytocin augmentation of dysfunctional labor. I. Clinical data. *American Journal of Obstetrics and Gynecology*, 144(8):899–905, 12 1982. 1

[7] Eugene Straus, Robert J. Greenstein, and Rosalyn S. Yalow. Plasma-Secretin in Management of Cimetidine Therapy for Zollinger-Ellison Syndrome. *The Lancet*, 312(8080):73–75, 7 1978. 1

[8] Helen E. Gruber, Joel L. Ivey, David J. Baylink, Meredith Matthews, Wil B. Nelp, Karen Sisom, and Charles H. Chesnut. Long-term calcitonin therapy in postmenopausal osteoporosis. *Metabolism*, 33(4):295–303, 1984. 1

[9] Jayanta K Mitra, Jayeeta Roy, and Saikat Sengupta. Vasopressin Its current role in anesthetic practice. *Indian journal of critical care medicine : peer-reviewed, official publication of Indian Society of Critical Care Medicine*, 15(2):71–7, 4 2011. 1

[10] Jolene L. Lau and Michael K. Dunn. Therapeutic peptides: Historical perspectives, current development trends, and future directions. *Bioorganic and Medicinal Chemistry*, 26(10):2700–2707, 6 2018. 1

[11] Yulei Li, Minghao Wu, Qi Chang, and Xia Zhao. Stapling strategy enables improvement of antitumor activity and proteolytic stability of host-defense peptide hymenochirin-1B. *RSC Advances*, 8(39):22268–22275, 2018. 1

[12] Thomas E. Speltz, Christopher G. Mayne, Sean W. Fanning, Zamia Siddiqui, Emad Tajkhorshid, Geoffrey L. Greene, and Terry W. Moore. A "cross-stitched" peptide with improved helicity and proteolytic stability. *Organic and Biomolecular Chemistry*, 16(20):3702–3706, 2018. 11, 194

[13] Zuojun Guo, Udayan Mohanty, Justin Noehre, Tomi K. Sawyer, Woody Sherman, and Goran Krilov. Probing the a-helical structural stability of stapled p53 peptides: Molecular dynamics simulations and analysis: Research article. *Chemical Biology and Drug Design*, 75(4):348–359, 2010. 1

[14] Kathrin Bellmann-Sickert, Tracy A. Stone, Bradley E. Poulsen, and Charles M. Deber. Efflux by small multidrug resistance proteins is inhibited by membrane-interactive helix-stapled peptides. *Journal of Biological Chemistry*, 290(3):1752–1759, 1 2015. 1

[15] Siau Jia Wei, Sharon Chee, Larisa Yurlova, David Lane, Chandra Verma, Christopher Brown, and Farid Ghadessy. Avoiding drug resistance through extended drug target interfaces: A case for stapled peptides. *Oncotarget*, 7(22):32232–32246, 2016. 1, 2

[16] Ameena M. Ali, Jack Atmaj, Niels Van Oosterwijk, Matthew R. Groves, and Alexander Dömling. Stapled Peptides Inhibitors: A New Window for Target Drug Discovery. *Computational and Structural Biotechnology Journal*, 17:263–281, 1 2019. 1, 10

[17] Leslie Reguera and Daniel G. Rivera. Multicomponent Reaction Toolbox for Peptide Macrocyclization and Stapling. *Chemical Reviews*, 119(17):9836–9860, 9 2019. 1

[18] Manuel G. Ricardo, Javiel F. Marrrero, Oscar Valdés, Daniel G. Rivera, and Ludger A. Wessjohann. A Peptide Backbone Stapling Strategy Enabled by the Multicomponent Incorporation of Amide N-Substituents. *European Journal*, 25(3):769–774, 1 2019. 2

[19] Aldrin V. Vasco, Yanira Méndez, Andrea Porzel, Jochen Balbach, Ludger A. Wessjohann, and Daniel G. Rivera. A Multicomponent Stapling Approach to Exocyclic Functionalized Helical Peptides: Adding Lipids, Sugars, PEGs, Labels, and Handles to the Lactam Bridge. *Bioconjugate Chemistry*, 30(1):253–259, 1 2019.

[20] Mihajlo Todorovic, Katerina D. Schwab, Jutta Zeisler, Chengcheng Zhang, Francois Bénard, and David M. Perrin. Fluorescent Isoindole Crosslink (FlICk) Chemistry: A Rapid, User-friendly Stapling Reaction. *Angewandte Chemie*, 131(40):14258–14262, 10 2019. 1, 15

[21] Kengo Hanaya, Jun Ohata, Mary K. Miller, Alicia E. Mangubat-Medina, Michael J. Swierczynski, David C. Yang, Reece M. Rosenthal, Brian V. Popp, and Zachary T. Ball. Rapid nickel( ii )-promoted cysteine S-arylation with arylboronic acids. *Chemical Communications*, 55(19):2841–2844, 2019. 2

[22] Mélanie M. Lorion, Nikolaos Kaplaneris, Jongwoo Son, Rositha Kuniyil, and Lutz Ackermann. Late-Stage Peptide Diversification through Cobalt-Catalyzed C–H Activation: Sequential Multicatalysis for Stapled Peptides. *Angewandte Chemie International Edition*, 58(6):1684–1688, 2 2019.

[23] Krishna Sharma, Alexander V. Strizhak, Elaine Fowler, Xuelu Wang, Wenshu Xu, Claus Hatt Jensen, Yuteng Wu, Hannah F. Sore, Yu Heng Lau, Marko Hyvönen, Laura S. Itzhaki, and David R. Spring. Water-soluble, stable and azide-reactive strained dialkynes for biocompatible double strain-promoted click chemistry. *Organic & Biomolecular Chemistry*, 17(34):8014–8018, 2019. 194

[24] Naomi S. Robertson, Stephen J. Walsh, Elaine Fowler, Masao Yoshida, Sam M. Rowe, Yuteng Wu, Hannah F. Sore, Jeremy S. Parker, and David R. Spring. Macrocyclisation and functionalisation of unprotected peptides via divinyltriazine cysteine stapling. *Chemical Communications*, 55(64):9499–9502, 2019.

[25] Elsa Pflimlin, Sam Lear, Candy Lee, Shan Yu, Huafei Zou, Andrew To, Sean Joseph, Van Nguyen-Tran, Matthew S. Tremblay, and Weijun Shen. Design of a Long-Acting and Selective MEG-Fatty Acid Stapled Prolactin-Releasing Peptide Analog. *ACS Medicinal Chemistry Letters*, 10(8):1166–1172, 8 2019.

[26] Soon-Sun Hong, Jung Ho Choi, Sung Yoon Lee, Yeon-Hwa Park, Kyung-Yeon Park, Joo Young Lee, Juyoung Kim, Veeraswamy Gajulapati, Ja-Il Goo, Sarbjit Singh, Kyeong Lee, Young-Kook Kim, So Hee Im, Sung-Hoon Ahn, Stefan Rose-John, Tae-Hwe Heo, and Yongseok Choi. A Novel Small-Molecule Inhibitor Targeting the IL-6 Receptor $\beta$ Subunit, Glycoprotein 130. *Journal of immunology (Baltimore, Md. : 1950)*, 195(1):237–245, 2015.

[27] Anthony W. Partridge, Hung Yi Kristal Kaan, Yu-Chi Juang, Ahmad Sadruddin, Shuhui Lim, Christopher J. Brown, Simon Ng, Dawn Thean, Fernando Ferrer, Charles Johannes, Tsz Ying Yuen, Srinivasaraghavan Kannan, Pietro Aronica, Yaw Sing Tan, Mohan R. Pradhan, Chandra S. Verma, Jerome Hochman, Shiying Chen, Hui Wan, Sookhee Ha, Brad Sherborne, David P. Lane, and Tomi K. Sawyer. Incorporation of Putative Helix-Breaking Amino Acids in the Design of Novel Stapled Peptides: Exploring Biophysical and Cellular Permeability Properties. *Molecules*, 24(12):2292, 6 2019. 2, 86

[28] Yu Heng Lau, Peterson de Andrade, Yuteng Wu, and David R. Spring. Peptide stapling techniques based on different macrocyclisation chemistries. *Chemical Society Reviews*, 44(1):91–102, 1 2015. 2

[29] Yuteng Wu, Amandeep Kaur, Elaine Fowler, Mareike M. Wiedmann, Reginald Young, Warren R.J.D. Galloway, Lasse Olsen, Hannah F. Sore, Anasuya Chattopadhyay, Terence T.L. Kwan, Wenshu Xu, Stephen J. Walsh, Peterson de Andrade, Matej Janecek, Senthil Arumugam, Laura S. Itzhaki, Yu Heng Lau, David R. Spring, Warren R J D Galloway, Lasse Olsen, Hannah F. Sore, Anasuya Chattopadhyay, Terence T-L Kwan, Wenshu Xu, Stephen J. Walsh, Peterson de Andrade, Matej Janecek, Senthil Arumugam, Laura S. Itzhaki, Yu Heng Lau, and David R. Spring. Toolbox of Diverse Linkers for Navigating the Cellular Efficacy Landscape of Stapled Peptides. *ACS Chemical Biology*, 14(3):526–533, 2019. 14

[30] Young Woo Kim, Tom N. Grossmann, and Gregory L. Verdine. Synthesis of all-hydrocarbon stapled alpha-helical peptides by ring-closing olefin metathesis. *Nature Protocols*, 6(6):761–771, 5 2011.

[31] Michael M. Madden, Claudia I. Rivera Vera, Wenjiao Song, and Qing Lin. Facile synthesis of stapled, structurally reinforced peptide helices via a photoinduced intramolecular 1,3-dipolar cycloaddition reaction. *Chemical Communications*, 37:5588–5590, 9 2009.

[32] Conor M. Haney and W. Seth Horne. Oxime side-chain cross-links in an $\alpha$-helical coiled-coil protein: Structure, thermodynamics, and folding-templated synthesis of bicyclic species. *Chemistry - A European Journal*, 19(34):11342–11351, 8 2013. 195

[33] Florence M. Brunel and Philip E. Dawson. Synthesis of constrained helical peptides by thioether ligation: application to analogs of gp41. *Chemical Communications*, (20):2552–2636, 5 2005. 2

[34] Jean Philippe Richard, Kamran Melikov, Eric Vives, Corinne Ramos, Birgit Verbeure, Mike J Gait, Leonid V Chernomordik, and Bernard Lebleu. Cell-penetrating peptides: A reevaluation of the mechanism of cellular uptake. *Journal of Biological Chemistry*, 278(1):585–590, 1 2003. 2

[35] Erez Koren and Vladimir P. Torchilin. Cell-penetrating peptides: Breaking through to the other side. *Trends in Molecular Medicine*, 18(7):385–393, 2012. 8

[36] Ankur Gautam, Harinder Singh, Atul Tyagi, Kumardeep Chaudhary, Rahul Kumar, Pallavi Kapoor, and G. P.S. Raghava. CPPsite: A curated database of cell penetrating peptides. *Database*, 2012:1–7, 2012. 33

[37] Victoria Sebbage. Cell-penetrating peptides and their therapeutic applications. *Bioscience Horizons*, 2(1):64–72, 2009.

[38] Maria Lindgren, Mattias Hällbrink, Alain Prochiantz, and Ülo Langel. Cell-penetrating peptides, 3 2000.

[39] Kristin L. Horton, Kelly M. Stewart, Sonali B. Fonseca, Qian Guo, Shana O. Kelley, and Grand review research. Peptide Drugs Market Size | Global Share 2019 to 2025, Business Stats, Growth perspective and Forecast 2025 - MarketWatch, 4 2008. 2

[40] Toru Okamoto, Kerry Zobel, Anna Fedorova, Clifford Quan, Hong Yang, Wayne J. Fairbrother, David C.S. Huang, Brian J. Smith, Kurt Deshayes, and Peter E. Czabotar. Stabilizing the pro-apoptotic BimBH3 helix (BimSAHB) does not necessarily enhance affinity or biological activity. *ACS Chemical Biology*, 8(2):297–302, 2 2013. 2

[41] Vanessa Gaillard, Marie Galloux, Dominique Garcin, Jean-François Eléouët, Ronan Le Goffic, Thibaut Larcher, Marie-Anne Rameix-Welti, Abdelhak Boukadiri, Julien Héritier, Jean-Manuel Segura, Elodie Baechler, Miriam Arrell, Geneviève Mottet-Osman, and Origène Nyanguile. A Short Double-Stapled Peptide Inhibits Respiratory Syncytial Virus Entry and Spreading. *Antimicrobial agents and chemotherapy*, 61(4):02241–16, 4 2017. 2, 11, 20

[42] Sohee Baek, Peter S. Kutchukian, Gregory L. Verdine, Robert Huber, Tad A. Holak, Ki Won Lee, and Grzegorz M. Popowicz. Structure of the stapled p53 peptide bound to Mdm2. *Journal of the American Chemical Society*, 134(1):103–106, 1 2012. 2, 194

[43] Anna Czarna, Grzegorz M. Popowicz, Aleksandra Pecak, Siglinde Wolf, Grzegorz Dubin, and Tad A. Holak. High affinity interaction of the p53 peptide-analogue with human Mdm2 and Mdmx. *Cell Cycle*, 8(8):1176–1184, 4 2009. 195

[44] Garima Tiwari and Chandra S. Verma. Toward Understanding the Molecular Recognition of Albumin by p53-Activating Stapled Peptide ATSP-7041. *The Journal of Physical Chemistry B*, 121(4):657–670, 2 2017. 7, 8

[45] Yong S. Chang, Bradford Graves, Vincent Guerlavais, Christian Tovar, Kathryn Packman, Kwong Him To, Karen A. Olson, Kamala Kesavan, Pranoti Gangurde, Aditi Mukherjee, Theresa Baker, Krzysztof Darlak, Carl Elkin, Zoran Filipovic, Farooq Z. Qureshi, Hongliang Cai, Pamela Berry, Eric Feyfant, Xiangguo E. Shi, James Horstick, D. Allen Annis, Anthony M. Manning, Nader Fotouhi, Huw Nash, Lyubomir T. Vassilev, and Tomi K. Sawyer. Stapled $\alpha$-helical peptide drug development: A potent dual inhibitor of MDM2 and MDMX for p53-dependent cancer therapy. *Proceedings of the National Academy of Sciences of the United States of America*, 110(36):3445–3454, 2013. 86

[46] Yeongju Lee, Heeseok Yoon, Sung-Min Hwang, Min-Kyung Shin, Ji Hoon Lee, Misook Oh, Sin-Hyeog Im, Jaeyoung Song, and Hyun-Suk Lim. Targeted Inhibition of the NCOA1/STAT6 Protein–Protein Interaction. *Journal of the American Chemical Society*, 139(45):16056–16059, 11 2017. 2

[47] James Sampietro, Caroline L. Dahlberg, Uhn Soo Cho, Thomas R. Hinds, David Kimelman, and Wenqing Xu. Crystal Structure of a $\beta$-Catenin/BCL9/Tcf4 Complex. *Molecular Cell*, 24(2):293–300, 10 2006. 2

[48] Loren D. Walensky, Kenneth Pitter, Joel Morash, Kyoung Joon Oh, Scott Barbuto, Jill Fisher, Eric Smith, Gregory L. Verdine, and Stanley J. Korsmeyer. A Stapled BID BH3 Helix Directly Binds and Activates BAX. *Molecular Cell*, 24(2):199–210, 10 2006. 2

[49] Gregory H. Bird, Adriana Irimia, Gilad Ofek, Peter D. Kwong, Ian A. Wilson, and Loren D. Walensky. Stapled HIV-1 peptides recapitulate antigenic structures and engage broadly neutralizing antibodies. *Nature Structural and Molecular Biology*, 21(12):1058–1067, 12 2014. 2

[50] Hongtao Zhang, Francesca Curreli, Xihui Zhang, Shibani Bhattacharya, Abdul A. Waheed, Alan Cooper, David Cowburn, Eric O. Freed, and Asim K. Debnath. Antiviral activity of $\alpha$-helical stapled peptides designed from the HIV-1 capsid dimerization domain. *Retrovirology*, 8(1):28, 5 2011. 2

[51] Edward L. Huttlin, Raphael J. Bruckner, Joao A. Paulo, Joe R. Cannon, Lily Ting, Kurt Baltier, Greg Colby, Fana Gebreab, Melanie P. Gygi, Hannah Parzen, John Szpyt, Stanley Tam, Gabriela Zarraga, Laura Pontano-Vaites, Sharan Swarup, Anne E. White, Devin K. Schweppe, Ramin Rad, Brian K. Erickson, Robert A. Obar, K. G. Guruharsha, Kejie Li, Spyros Artavanis-Tsakonas, Steven P. Gygi, and J. Wade Harper. Architecture of the human interactome defines protein communities and disease networks. *Nature*, 545(7655):505–509, 5 2017. 2

[52] Kristina Hetherington, Zsofia Hegedus, Thomas A. Edwards, Richard B. Sessions, Adam Nelson, and Andrew J. Wilson. Stapled Peptides as HIF-1$\alpha$/p300 Inhibitors: Helicity Enhancement in the Bound State Increases Inhibitory Potency. *Chemistry European Journal*, 26(34):7638–7646, 6 2020. 3, 4

[53] Mareike M. Wiedmann, Yaw Sing Tan, Yuteng Wu, Shintaro Aibara, Wenshu Xu, Hannah F. Sore, Chandra S. Verma, Laura Itzhaki, Murray Stewart, James D. Brenton, and David R. Spring. Development of Cell-Permeable, Non-Helical Constrained Peptides to Target a Key Protein–Protein Interaction in Ovarian Cancer. *Angewandte Chemie - International Edition*, 56(2):524–529, 1 2017. 3, 4

[54] Christian E. Schafmeister, Julia Po, and Gregory L. Verdine. An all-hydrocarbon cross-linking system for enhancing the helicity and metabolic stability of peptides. *Journal of the American Chemical Society*, 122(24):5891–5892, 2000. 3

[55] Yao-Zhong Lin, SongYi Yao, Ruth Ann Veach, Troy R. Torgerson, and Jacek Hawiger. Inhibition of Nuclear Translocation of Transcription Factor NF-$\kappa$B by a Synthetic Peptide Containing a Cell Membrane-permeable Motif and Nuclear Localization Sequence. *Journal of Biological Chemistry*, 270(24):14255–14258, 6 1995. 4

[56] Sampat Ingale and Philip E Dawson. On resin side-chain cyclization of complex peptides using CuAAC. *Organic Letters*, 13(11):2822–2825, 2011. 4

[57] Lau Heng Yu, Peterson de Andrade, Soo-tng Quah, Maxim Rossmann, Tze Jing Sum, Pamela J. E. Rowling, Thomas L. Joseph, Chandra Verma, Marko Hyv, Laura S. Itzhaki, Ashok R. Venkitaraman, Christopher J. Brown, P Lane, David R. Spring, Peterson de Andrade, Soo-tng Quah, Maxim Rossmann, Luca Laraia, Niklas Sköld, Tze Jing Sum, Pamela J. E. Rowling, Thomas L. Joseph, Chandra Verma, Marko Hyvönen, Laura S. Itzhaki, Ashok R. Venkitaraman, Christopher J. Brown, David P. Lane, and David R. Spring. Functionalised staple linkages for modulating the cellular activity of stapled peptides. *Chem. Sci.*, 5(5):1804–1809, 4 2014.

[58] Jessica Iegre, Josephine S. Gaynord, Naomi S. Robertson, Hannah F. Sore, Marko Hyvönen, and David R. Spring. Two-Component Stapling of Biologically Active and Conformationally Constrained Peptides: Past, Present, and Future. *Advanced Therapeutics*, 1(7):1800052, 11 2018. 4, 10

[59] Sally McGrath, Marcello Tortorici, Ludovic Drouin, Savade Solanki, Lewis Vidler, Isaac Westwood, Peter Gimeson, Rob Van Montfort, and Swen Hoelder. Structure-Enabled Discovery of a Stapled Peptide Inhibitor to Target the Oncogenic Transcriptional Repressor TLE1. *Chemistry - A European Journal*, 23(40):9577–9584, 7 2017. 4

[60] Fergus S. McWhinnie, Kristel Sepp, Charlotte Wilson, Tilo Kunath, Ted R. Hupp, Terry S. Baker, Douglas R. Houston, and Alison N. Hulme. Mono-Substituted Hydrocarbon Diastereomer Combinations Reveal Stapled Peptides with High Structural Fidelity. *Chemistry - A European Journal*, 24(9):2094–2097, 2 2018. 4, 5, 71, 113

[61] Jinan Wang, Andrey Alekseenko, Dima Kozakov, and Yinglong Miao. Improved Modeling of Peptide-Protein Binding Through Global Docking and Accelerated Molecular Dynamics Simulations. *Frontiers in Molecular Biosciences*, 6, 10 2019. 4, 179

[62] Briana R. Flaherty, Tienhuei G. Ho, Sven H. Schmidt, Friedrich W. Herberg, David S. Peterson, and Eileen J. Kennedy. Targeted Inhibition of Plasmodium falciparum

Calcium-Dependent Protein Kinase 1 with a Constrained J Domain-Derived Disruptor Peptide. *ACS Infectious Diseases*, 5(4):506–514, 4 2019. 5

[63] Guangpeng Meng, Jing Pu, Yue Li, Aixin Han, Yangli Tian, Wei Xu, Tianhong Zhang, Xue Li, Lu Lu, Chao Wang, Shibo Jiang, and Keliang Liu. Design and Biological Evaluation of m-Xylene Thioether-Stapled Short Helical Peptides Targeting the HIV-1 gp41 Hexameric Coiled-Coil Fusion Complex. *Journal of Medicinal Chemistry*, 62(19):8773–8783, 10 2019. 5

[64] Wenqi Song, Kunzheng Wang, Wei Wang, Pei Yang, and Xiaoqian Dang. Grafting, Stripping and Stapling of Helical Peptides from the Dimerization Interface of ONFH-Related Bone Morphogenetic Protein-2. *The Protein Journal*, 38(1):12–22, 2 2019. 5

[65] Sadasivam Jeganathan, Mathias Wendt, Sebastian Kiehstaller, Diego Brancaccio, Arne Kuepper, Nicole Pospiech, Alfonso Carotenuto, Ettore Novellino, Sven Hennig, and Tom N. Grossmann. Constrained Peptides with Fine-Tuned Flexibility Inhibit NF-Y Transcription Factor Assembly. *Angewandte Chemie - International Edition*, 58(48):17351–17358, 10 2019. 5

[66] Erkki Koivunen, Wadih Arap, Daniel Rajotte, Johanna Lahdenranta, and Renata Pasqualini. Identification of receptor ligands with phage display peptide libraries. *Journal of Nuclear Medicine*, 40(5):883–888, 1999. 5

[67] Andreas Plückthun. Ribosome display: A perspective, 2012. 5

[68] Kristopher Josephson, Alonso Ricardo, and Jack W. Szostak. MRNA display: From basic principles to macrocycle drug discovery. *Drug Discovery Today*, 19(4):388–399, 4 2014. 5

[69] Eric T. Boder and K. Dane Wittrup. Yeast surface display for screening combinatorial polypeptide libraries. *Nature Biotechnology*, 15(6):553–557, 6 1997. 5

[70] Asma Achek, Masaud Shah, Ji Young Seo, Hyuk-Kwon Kwon, Xiangai Gui, Hyeon-Jun Shin, Eun-Young Cho, Byeong Sung Lee, Dong-Jin Kim, Sang Ho Lee, Tae Hyeon Yoo, Moon Suk Kim, and Sangdun Choi. Linear and Rationally Designed Stapled Peptides Abrogate TLR4 Pathway and Relieve Inflammatory Symptoms in Rheumatoid Arthritis Rat Model. *Journal of Medicinal Chemistry*, 62(14):6495–6511, 7 2019. 6

[71] Philippe Diderich, Davide Bertoldo, Pierre Dessen, Maola M. Khan, Irene Pizzitola, Werner Held, Joerg Huelsken, and Christian Heinis. Phage Selection of Chemically Stabilized $\alpha$-Helical Peptide Ligands. *ACS Chemical Biology*, 11(5):1422–1427, 5 2016. 6

[72] Teerapat Anananuchatkul, Iou Ven Chang, Takayuki Miki, Hiroshi Tsutsumi, and Hisakazu Mihara. Construction of a Stapled $\alpha$-Helix Peptide Library Displayed on Phage for the Screening of Galectin-3-Binding Peptide Ligands. *ACS Omega*, 107(32):14093–8, 8 2020. 6

[73] Emil S. Iqbal, Stacie L. Richardson, Nicolas A. Abrigo, Kara K. Dods, H. Estheban Osorio Franco, Heather S. Gerrish, Hari Kiran Kotapati, Iain M. Morgan, Douglas S. Masterson, and Matthew C.T. Hartman. A new strategy for the: In vitro selection of stapled peptide inhibitors by mRNA display. *Chemical Communications*, 55(61):8959–8962, 7 2019. 6

[74] W. L. DeLano. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography*, 40:82–92, 2002. 6, 193

[75] So Youn Shim, Young Woo Kim, and Gregory L. Verdine. A new i, i + 3 peptide stapling system for $\alpha$-helix stabilization. *Chemical Biology and Drug Design*, 82(6):635–642, 2013. 6

[76] Tao Wu, Ping He, Wei Wu, Yingli Chen, and Fenglin Lv. Targeting oncogenic transcriptional corepressor Nac1 POZ domain with conformationally constrained peptides by cyclization and stapling. *Bioorganic Chemistry*, 80:1–10, 10 2018. 6

[77] Victor Muñoz, Francisco J. Blanco, and Luis Serrano. The hydrophobic-staple motif and a role for loop-residues in $\alpha$-helix stability and protein folding. *Nature Structural & Molecular Biology*, 2(5):380–385, 5 1995. 7

[78] Zhixia Chen, Xiuli Yu, Aiying Zhang, Fangfang Wang, and Yankun Xing. De Novo Hydrocarbon-Stapling Design of Single-Turn $\alpha$-Helical Antimicrobial Peptides. *International Journal of Peptide Research and Therapeutics*, pages 1–9, 11 2019. 7

[79] William F. Porto, Állan S. Pires, and Octavio L. Franco. CS-AMPPred: An Updated SVM Model for Antimicrobial Activity Prediction in Cysteine-Stabilized Peptides. *PLoS ONE*, 7(12), 12 2012. 7

[80] Aravindhan Ganesan, Michelle L. Coote, and Khaled Barakat. Molecular dynamics-driven drug discovery: leaping forward with confidence. *Drug Discovery Today*, 22(2):249–269, 2 2017. 7

[81] Jacob D. Durrant and J. Andrew McCammon. Molecular dynamics simulations and drug discovery. *BMC Biology*, 9(1):71, 10 2011.

[82] Jérémie Mortier, Christin Rakers, Marcel Bermudez, Manuela S. Murgueitio, Sereina Riniker, and Gerhard Wolber. The impact of molecular dynamics on drug design: Applications for the characterization of ligand-macromolecule complexes. *Drug Discovery Today*, 20(6):686–702, 6 2015.

[83] Mai Suan Li and Binh Khanh Mai. Steered Molecular Dynamics-A Promising Tool for Drug Design. *Current Bioinformatics*, 7(4):342–351, 11 2013. 8

[84] Indrani Bera and Pavan V. Payghan. Use of Molecular Dynamics Simulations in Structure-Based Drug Discovery. *Current Pharmaceutical Design*, 25(31):3339–3349, 9 2019.

[85] Marco De Vivo, Matteo Masetti, Giovanni Bottegoni, and Andrea Cavalli. Role of Molecular Dynamics and Related Methods in Drug Discovery. *Journal of Medicinal Chemistry*, 59(9):4035–4061, 5 2016. 7

[86] Yaw Sing Tan, Judith Reeks, Christopher J. Brown, Dawn Thean, Fernando Jose Ferrer Gago, Tsz Ying Yuen, Eunice Tze Leng Goh, Xue Er Cheryl Lee, Claire E. Jennings, Thomas L. Joseph, Rajamani Lakshminarayanan, David P. Lane, Martin E. M. Noble, and Chandra S. Verma. Benzene Probes in Molecular Dynamics Simulations Reveal Novel Binding Sites for Ligand Design. *The Journal of Physical Chemistry Letters*, 7(17):3452–3457, 9 2016. 7, 8, 194

[87] Joseph A. Morrone, Alberto Perez, Justin MacCallum, and Ken A. Dill. Computed Binding of Peptides to Proteins with MELD-Accelerated Molecular Dynamics. *Journal of Chemical Theory and Computation*, 13(2):870–876, 2 2017. 7, 8

[88] Joseph A. Morrone, Alberto Perez, Qiaolin Deng, Sookhee N. Ha, M. Katharine Holloway, Tomi K. Sawyer, Bradley S. Sherborne, Frank K. Brown, and Ken A. Dill. Molecular Simulations Identify Binding Poses and Approximate Affinities of Stapled $\alpha$-Helical Peptides to MDM2 and MDMX. *Journal of Chemical Theory and Computation*, 13(2):863–869, 2 2017. 7

[89] David J. Diller, Jon Swanson, Alexander S. Bayden, Chris J. Brown, Dawn Thean, David P. Lane, Anthony W. Partridge, Tomi K. Sawyer, and Joseph Audie. Rigorous computational and experimental investigations on MDM2/MDMX-targeted linear and macrocyclic peptides. *Molecules*, 24(24):4586, 12 2019. 7, 11

[90] Kerstin Wallraven, Fredrik L. Holmelin, Adrian Glas, Sven Hennig, Andrey I. Frolov, and Tom N. Grossmann. Adapting free energy perturbation simulations for large macrocyclic ligands: How to dissect contributions from direct binding and free ligand flexibility. *Chemical Science*, 11(8):2269–2276, 2 2020. 7

[91] Haoyu S. Yu, Yuqing Deng, Yujie Wu, Dan Sindhikara, Amy R. Rask, Takayuki Kimura, Robert Abel, and Lingle Wang. Accurate and Reliable Prediction of the Binding Affinities of Macrocycles to Their Protein Targets. *Journal of Chemical Theory and Computation*, 13(12):6290–6300, 12 2017. 7

[92] Vincent Wagner, Linda Jantz, Hans Briem, Kai Sommer, Matthias Rarey, and Clara D. Christ. Computational Macrocyclization: From de novo Macrocycle Generation to Binding Affinity Estimation. *ChemMedChem*, 12(22):1866–1872, 11 2017. 7

[93] Pedro A Valiente, David Becerra, and Philip M Kim. A Method to Calculate the Relative Binding Free Energy Differences of $\alpha$-Helical Stapled Peptides. *Cite This: J. Org. Chem*, 85:1644–1651, 2020. 7

[94] Samuel Genheden and Ulf Ryde. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities, 5 2015. 7, 178

[95] Tatsuya Yamada, Tomohiko Hayashi, Simon Hikiri, Naohiro Kobayashi, Hiroshi Yanagawa, Mitsunori Ikeguchi, Masato Katahira, Takashi Nagata, and Masahiro Kinoshita. How Does the Recently Discovered Peptide MIP Exhibit Much Higher Binding Affinity than an Anticancer Protein p53 for an Oncoprotein MDM2? *Journal of Chemical Information and Modeling*, 59(8):3533–3544, 8 2019. 7

[96] Thomas L. Joseph, David P. Lane, and Chandra S. Verma. Stapled BH3 Peptides against MCL-1: Mechanism and Design Using Atomistic Simulations. *PLoS ONE*, 7(8):1–11, 8 2012. 11

[97] Wenshu Xu, Yu Heng Lau, Gerhard Fischer, Yaw Sing Tan, Anasuya Chattopadhyay, Marc De La Roche, Marko Hyvönen, Chandra Verma, David R. Spring, and Laura S. Itzhaki. Macrocyclized Extended Peptides: Inhibiting the Substrate-Recognition Domain of Tankyrase. *Journal of the American Chemical Society*, 139(6):2245–2256, 2 2017. 7, 10, 195

[98] Maciej Pawel Ciemny, Aleksander Debinski, Marta Paczkowska, Andrzej Kolinski, Mateusz Kurcinski, and Sebastian Kmiecik. Protein-peptide molecular docking with large-scale conformational changes: The p53-MDM2 interaction. *Scientific Reports*, 6, 2016. 8

[99] Sjoerd J. de Vries, Christina E.M. Schindler, Isaure Chauvot de Beauchêne, and Martin Zacharias. A Web Interface for Easy Flexible Protein-Protein Docking with ATTRACT. *Biophysical Journal*, 108(3):462–465, 2 2015. 8

[100] Nir London, Barak Raveh, and Ora Schueler-Furman. Druggable protein-protein interactions - from hot spots to hot segments. *Current Opinion in Chemical Biology*, 17(6):952–959, 8 2013. 8

[101] X. Y. Guo, Z. Q. Peng, Y. W. Zhang, B. Liu, and Y. Q. Cui. The solubility and conformational characteristics of porcine myosin as affected by the presence of l-lysine and l-histidine. *Food Chemistry*, 170:212–217, 3 2015. 8

[102] Tong Shi, Li Yuan, Jianlou Mu, and Ruichang Gao. The effect of Arginine, Lysine and Histidine in the myosin secondary structure by circular dichroism and Raman spectroscopy. *CYTA - Journal of Food*, 17(1):656–660, 1 2019.

[103] Eisuke Takai, Shunsuke Yoshizawa, Daisuke Ejima, Tsutomu Arakawa, and Kentaro Shiraki. Synergistic solubilization of porcine myosin in physiological salt solution by arginine. *International Journal of Biological Macromolecules*, 62:647–651, 11 2013. 8

[104] Gregory H Bird, Emanuele Mazzola, Kwadwo Opoku-Nsiah, Margaret A Lammert, Marina Godes, Donna S Neuberg, and Loren D Walensky. Biophysical determinants for cellular uptake of hydrocarbon-stapled peptide helices. *Nature Chemical Biology*, 12(10):845–852, 10 2016. 8

[105] Marie-Lise Jobin, Marine Blanchet, Sarah Henry, Stéphane Chaignepain, Claude Manigand, Sabine Castano, Sophie Lecomte, Fabienne Burlina, Sandrine Sagan, and Isabel D. Alves. The role of tryptophans on the cellular uptake and membrane interaction of arginine-rich cell penetrating peptides. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1848(2):593–602, 2 2015. 8

[106] Hanna A. Rydberg, Maria Matson, Helene L. Åmand, Elin K. Esbjörner, and Bengt Nordén. Effects of tryptophan content and backbone spacing on the uptake efficiency of cell-penetrating peptides. *Biochemistry*, 51(27):5531–5539, 7 2012.

[107] Hanna A. Rydberg, Nils Carlsson, and Bengt Nordén. Membrane interaction and secondary structure of de novo designed arginine-and tryptophan peptides with dual function. *Biochemical and Biophysical Research Communications*, 427(2):261–265, 10 2012. 8

[108] Avijit Chakrabartty, Tanja Kortemme, S. Padmanabhan, and Robert L. Baldwin. Aromatic side-chain contribution to far-ultraviolet circular dichroism of helical peptides and its effect on measurement of helix propensities. *Biochemistry*, 32(21):5560–5565, 6 1993. 8

[109] C. Nick Pace and J. Martin Scholtz. A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins. *Biophysical Journal*, 75(1):422–427, 7 1998. 9

[110] Nian E Zhou, Oscar D Monera, Cyril M Kay, and Robert S Hodges. $\alpha$-Helical propensities of amino acids in the hydrophobic face of an amphipathic $\alpha$-helix. *Protein and Peptide Lett.*, 1:114–119, 1994. 9

[111] Juan C. Serrano, James Sipthorp, Wenshu Xu, Laura S. Itzhaki, and Steven V. Ley. A New Methodology for Incorporating Chiral Linkers into Stapled Peptides. *ChemBioChem*, 18(12):1066–1071, 6 2017. 9

[112] Timothy A. Hill, Nicholas E. Shepherd, Frederik Diness, and David P. Fairlie. Constraining Cyclic Peptides To Mimic Protein Structure Motifs. *Angewandte Chemie International Edition*, 53(48):13020–13041, 11 2014. 9

[113] Hongtao Zhang, Francesca Curreli, Abdul A. Waheed, Peter Y. Mercredi, Mansi Mehta, Pallavi Bhargava, Daniel Scacalossi, Xiaohe Tong, Shawn Lee, Alan Cooper, Michael F. Summers, Eric O. Freed, and Asim K. Debnath. Dual-acting stapled peptides target both HIV-1 entry and assembly. *Retrovirology*, 10(1):1–20, 11 2013. 9

[114] Raheleh Rezaei Araghi, Jeremy A. Ryan, Anthony Letai, and Amy E. Keating. Rapid Optimization of Mcl-1 Inhibitors using Stapled Peptide Libraries Including Non-Natural Side Chains. *ACS Chemical Biology*, 11(5):1238–1244, 5 2016. 9

[115] Raheleh Rezaei Araghi, Gregory H Bird, Jeremy A Ryan, Justin M Jenson, Marina Godes, Jonathan R Pritz, Robert A Grant, Anthony Letai, Loren D Walensky, and Amy E Keating. Iterative optimization yields Mcl-1–targeting stapled peptides with selective cytotoxicity to Mcl-1–dependent cancer cells. *Proceedings of the National Academy of Sciences of the United States of America*, 115(5):E886–E895, 2018. 9, 11

[116] Andrea L. Jochim and Paramjit S. Arora. Systematic Analysis of Helical Protein Interfaces Reveals Targets for Synthetic Inhibitors. *ACS Chemical Biology*, 5(10):919–923, 10 2010. 10, 194

[117] Yan Wang, Mingxia Wang, Sanwen Yin, Richard Jang, Jian Wang, Zhidong Xue, and Tao Xu. NeuroPep: A comprehensive resource of neuropeptides. *Database*, 2015(0), 4 2015. 10, 33

[118] Andrew Jamieson and Naomi Robertson. Regulation of protein-protein interactions using stapled peptides. *Reports in Organic Chemistry*, 5:65, 8 2015.

[119] Marco J. Klein, Samuel Schmidt, Parvesh Wadhwani, Jochen Bürck, Johannes Reichert, Sergii Afonin, Marina Berditsch, Tim Schober, Roland Brock, Manfred Kansy, and Anne S. Ulrich. Lactam-Stapled Cell-Penetrating Peptides: Cell Uptake and Membrane Binding Properties. *Journal of Medicinal Chemistry*, 60(19):8071–8082, 10 2017.

[120] Mark Klein. Stabilized helical peptides: overview of the technologies and its impact on drug discovery. *Expert Opinion on Drug Discovery*, 12(11):1117–1125, 11 2017.

[121] Philipp M. Cromm, Jochen Spiegel, and Tom N. Grossmann. Hydrocarbon Stapled Peptides as Modulators of Biological Function. *ACS Chemical Biology*, 10(6):1362–1375, 6 2015.

[122] M. Werle and A. Bernkop-Schnürch. Strategies to improve plasma half life time of peptide and protein drugs, 6 2006. 10, 17

[123] Dilraj Lama, Soo T. Quah, Chandra S. Verma, Rajamani Lakshminarayanan, Roger W. Beuerman, David P. Lane, and Christopher J. Brown. Rational optimization of conformational effects induced by hydrocarbon staples in peptides and their binding interfaces. *Scientific Reports*, 3(1):1–10, 12 2013. 10, 195

[124] Erin E. Gallagher, James M. Song, Arya Menon, Lauren D. Mishra, Alyah F. Chmiel, and Amanda L. Garner. Consideration of Binding Kinetics in the Design of Stapled Peptide Mimics of the Disordered Proteins Eukaryotic Translation Initiation Factor 4E-Binding Protein 1 and Eukaryotic Translation Initiation Factor 4G. *Journal of Medicinal Chemistry*, 62(10):4967–4978, 5 2019. 103

[125] Tara Rao, Gloria Ruiz-Gómez, Timothy A. Hill, Huy N. Hoang, David P. Fairlie, and Jody M. Mason. Truncated and Helix-Constrained Peptides with High Affinity and Specificity for the cFos Coiled-Coil of AP-1. *PLoS ONE*, 8(3):e59415, 3 2013.

[126] Kuan Hu, Hao Geng, Qingzhou Zhang, Qisong Liu, Mingsheng Xie, Chengjie Sun, Wenjun Li, Huacan Lin, Fan Jiang, Tao Wang, Yun-Dong Wu, and Zigang Li. An In-tether Chiral Center Modulates the Helicity, Cell Permeability, and Target Binding Affinity of a Peptide. *Angewandte Chemie International Edition*, 55(28):8013–8017, 7 2016. 10

[127] Adelene Y. L. Sim and Chandra Verma. How does a hydrocarbon staple affect peptide hydrophobicity? *Journal of Computational Chemistry*, 36(10):773–784, 4 2015. 10

[128] Aline D. de Araujo, Huy N. Hoang, W. Mei Kok, Frederik Diness, Praveer Gupta, Timothy A. Hill, Russell W. Driver, David A. Price, Spiros Liras, and David P. Fairlie. Comparative $\alpha$-Helicity of Cyclic Pentapeptides in Water. *Angewandte Chemie International Edition*, 53(27):6965–6969, 7 2014. 10, 14

[129] Yuan Tian, Yanhong Jiang, Jingxu Li, Dongyuan Wang, Hui Zhao, and Zigang Li. Effect of Stapling Architecture on Physiochemical Properties and Cell Permeability of Stapled $\alpha$-Helical Peptides: A Comparative Study. *ChemBioChem*, 18(21):2087–2093, 11 2017. 10, 14

[130] Stephen P. Brown and Amos B. Smith. Peptide/Protein Stapling and Unstapling: Introduction of s-Tetrazine, Photochemical Release, and Regeneration of the Peptide/Protein. *Journal of the American Chemical Society*, 137(12):4034–4037, 4 2015. 10

[131] Peter Timmerman, Joris Beld, Wouter C. Puijk, and Rob H. Meloen. Rapid and Quantitative Cyclization of Multiple Peptide Loops onto Synthetic Scaffolds for Structural Mimicry of Protein Surfaces. *ChemBioChem*, 6(5):821–824, 5 2005. 10

[132] Menachem J. Gunzburg, Ketav Kulkarni, Gabrielle M. Watson, Nigus D. Ambaye, Mark P. Del Borgo, Rebecca Brandt, Stephanie C. Pero, Patrick Perlmutter, Matthew C.J. J. Wilce, and Jacqueline A. Wilce. Unexpected involvement of staple leads to redesign of selective bicyclic peptide inhibitor of Grb7. *Scientific Reports*, 6(1):1–13, 6 2016. 10, 11, 195

[133] Shashi Gupta, Masao Hirota, Sheela M. Waugh, Ikuo Murakami, Tomoki Suzuki, Masahiro Muraguchi, Masafumi Shibamori, Yuichi Ishikawa, Thale C. Jarvis, Jeffrey D. Carter, Chi Zhang, Bharat Gawande, Michael Vrkljan, Nebojsa Janjic, and Daniel J. Schneider. Chemically modified DNA aptamers bind interleukin-6 with high affinity and inhibit signaling by blocking its interaction with interleukin-6 receptor. *Journal of Biological Chemistry*, 289(12):8706–8719, 2014. 11

[134] Andrew A. Bogan and Kurt S. Thorn. Anatomy of hot spots in protein interfaces. *Journal of Molecular Biology*, 280(1):1–9, 7 1998. 11

[135] J A Vila, D R Ripoll, and H A Scheraga. Physical reasons for the unusual alpha-helix stabilization afforded by charged or neutral polar residues in alanine-rich peptides. *Proceedings of the National Academy of Sciences of the United States of America*, 97(24):13075–9, 11 2000. 11

[136] Yu Heng Lau, Peterson de Andrade, Niklas Sköld, Grahame J. McKenzie, Ashok R. Venkitaraman, Chandra Verma, David P. Lane, and David R. Spring. Investigating peptide sequence variations for 'double-click' stapled p53 peptides. *Org. Biomol. Chem.*, 12(24):4074–4077, 5 2014. 11, 14

[137] Tyler Lalonde, Trevor G. Shepherd, Savita Dhanvantari, and Leonard G. Luyt. Stapled ghrelin peptides as fluorescent imaging probes. *Peptide Science*, 111(1):e24055, 1 2019. 11

[138] Yu Heng Lau, Yuteng Wu, Maxim Rossmann, Ban Xiong Tan, Peterson De Andrade, Yaw Sing Tan, Chandra Verma, Grahame J. McKenzie, Ashok R. Venkitaraman, Marko Hyvönen, and David R. Spring. Double Strain-Promoted Macrocyclization for the Rapid Selection of Cell-Active Stapled Peptides. *Angewandte Chemie - International Edition*, 54(51):15410–15413, 12 2015. 11, 18, 194

[139] Xiong An Lee, Chandra Verma, and Adelene Y.L Sim. Designing dual inhibitors of Mdm2/MdmX: Unexpected coupling of water with gatekeeper Y100/99. *Proteins: Structure, Function, and Bioinformatics*, 85(8):1493–1506, 8 2017. 11

[140] David Eisenberg, Robert M Weiss, Thomas C Terwilliger, and William Wilcox. Hydrophobic Moments and Protein Structure. Technical report, Institute of Molecular Biology and Department of Chemistry,University of California, Los Angeles, California 90024, U.S.A., 8 1982. 13

[141] Jonathan E. Bock, Jason Gavenonis, and Joshua A. Kritzer. Getting in shape: Controlling peptide bioactivity and bioavailability using conformational constraints, 3 2013. 14

[142] Avinash Muppidi, Kenichiro Doi, Carlo P. Ramil, Hong Gang Wang, and Qing Lin. Synthesis of cell-permeable stapled BH3 peptide-based Mcl-1 inhibitors containing simple aryl and vinylaryl cross-linkers. *Tetrahedron*, 70(42):7740–7745, 10 2014. 14, 195

[143] Patrick G. Dougherty, Jin Wen, Xiaoyan Pan, Amritendu Koley, Jian Guo Ren, Ashweta Sahni, Ruchira Basu, Heba Salim, George Appiah Kubi, Ziqing Qian, and Dehua Pei. Enhancing the Cell Permeability of Stapled Peptides with a Cyclic Cell-Penetrating Peptide. *Journal of Medicinal Chemistry*, 62(22):10098–10107, 11 2019. 14

[144] Yanhong Jiang, Qiwen Deng, Hui Zhao, Mingsheng Xie, Longjian Chen, Feng Yin, Xuan Qin, Weihao Zheng, Yongjuan Zhao, and Zigang Li. Development of Stabilized Peptide-Based PROTACs against Estrogen Receptor $\alpha$. *ACS Chemical Biology*, 13(3):628–635, 3 2018. 14

[145] Ching Hsuan Tung. Fluorescent peptide probes for in vivo diagnostic imaging. *Biopolymers - Peptide Science Section*, 76(5):391–403, 1 2004. 15

[146] Hazel H. Szeto, Peter W. Schiller, Kesheng Zhao, and Guoxiong Luo. Fluorescent dyes alter intracellular targeting and function of cell-penetrating tetrapeptides. *The FASEB Journal*, 19(1):118–120, 1 2005. 15

[147] Qiang Xiao, Natalie A. Bécar, Nathaniel P. Brown, Mason S. Smith, Kimberlee L. Stern, Steven R.E. Draper, Katherine P. Thompson, and Joshua L. Price. Stapling of two PEGylated side chains increases the conformational stability of the WW domain via an entropic effect. *Organic and Biomolecular Chemistry*, 16(46):8933–8939, 2018. 15

[148] Yulin Tian, Huafei Zou, Peng An, Zhihong Zhou, Weijun Shen, and Qing Lin. Design of stapled oxyntomodulin analogs containing functionalized biphenyl cross-linkers. *Tetrahedron*, 75(2):286–295, 1 2019. 15

[149] Aldrin V. Vasco, Carlos S. Pérez, Fidel E. Morales, Hilda E. Garay, Dimitar Vasilev, José A. Gavín, Ludger A. Wessjohann, and Daniel G. Rivera. Macrocyclization of

Peptide Side Chains by the Ugi Reaction: Achieving Peptide Folding and Exocyclic N-Functionalization in One Shot. *Journal of Organic Chemistry*, 80(13):6697–6707, 7 2015. 16

[150] Naila Assem, David J. Ferreira, Dennis W. Wolan, and Philip E. Dawson. Acetone-Linked Peptides: A Convergent Approach for Peptide Macrocyclization and Labeling. *Angewandte Chemie - International Edition*, 54(30):8665–8668, 7 2015. 16

[151] Laura Nevola, Andrés Martín-Quirós, Kay Eckelt, Núria Camarero, Sébastien Tosi, Artur Llobet, Ernest Giralt, and Pau Gorostiza. Light-Regulated Stapled Peptides to Inhibit Protein-Protein Interactions Involved in Clathrin-Mediated Endocytosis. *Angewandte Chemie International Edition*, 52(30):7704–7708, 7 2013. 16

[152] Jens Bredenbeck, Jan Helbing, Janet R Kumita, G Andrew Woolley, and Peter Hamm. $\alpha$-helix formation in a photoswitchable peptide tracked from picoseconds to microseconds by time-resolved IR spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 102(7):2379–2384, 2 2005. 16

[153] Christian Hoppmann, Ronald Kühne, and Michael Beyermann. Intramolecular bridges formed by photoswitchable click amino acids. *Beilstein Journal of Organic Chemistry*, 8:884–889, 2012. 16

[154] Michael M. Madden, Avinash Muppidi, Zhenyu Li, Xiaolong Li, Jiandong Chen, and Qing Lin. Synthesis of cell-permeable stapled peptide dual inhibitors of the p53-Mdm2/Mdmx interactions via photoinduced cycloaddition. *Bioorganic and Medicinal Chemistry Letters*, 21(5):1472–1475, 3 2011. 16

[155] Phuong Thu Tran, Christian Ørnbøl Larsen, Tobias Røndbjerg, Martina De Foresta, Micha B. A. Kunze, Ales Marek, Jacob Hartvig Løper, Lotte-Emilie Boyhus, Astrid Knuhtsen, Kresten Lindorff-Larsen, and Daniel Sejer Pedersen. Diversity-Oriented Peptide Stapling: A Third Generation Copper-Catalysed Azide-Alkyne Cycloaddition Stapling and Functionalisation Strategy. *Chemistry European Journal*, 23(14):3490–3495, 3 2017. 17

[156] Young-Joo Lee, Sanghun Han, and Yong-beom Lim. Simultaneous Stabilization and Multimerization of a Peptide $\alpha$-Helix by Stapling Polymerization. *Macromolecular Rapid Communications*, 37(13):1021–1026, 7 2016. 17

[157] L. Zhu, L. Lu, S. Wang, J. Wu, J. Shi, T. Yan, C. Xie, Q. Li, M. Hu, and Z. Liu. Oral absorption basics: Pathways and physicochemical and biological factors affecting absorption. In *Developing Solid Oral Dosage Forms: Pharmaceutical Theory and Practice: Second Edition*, pages 297–329. Elsevier Inc., 1 2017. 17

[158] H. Kim, J. H. Jang, S. C. Kim, and J. H. Cho. De novo generation of short antimicrobial peptides with enhanced stability and cell specificity. *Journal of Antimicrobial Chemotherapy*, 69(1):121–132, 1 2014. 17

[159] Elisabeth Gasteiger, Christine Hoogland, Alexandre Gattiker, S'everine Duvaud, Marc R. Wilkins, Ron D. Appel, and Amos Bairoch. Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*, pages 571–607. Humana Press, 2005. 17

[160] Philipp M. Cromm, Jochen Spiegel, Philipp Küchler, Laura Dietrich, Julia Kriegesmann, Mathias Wendt, Roger S. Goody, Herbert Waldmann, and Tom N. Grossmann. Protease-Resistant and Cell-Permeable Double-Stapled Peptides Targeting the Rab8a GTPase. *ACS Chemical Biology*, 11(8):2375–2382, 8 2016. 17, 33

[161] Steven A. Kawamoto, Adriana Coleska, Xu Ran, Han Yi, Chao-Yie Yie Yang, and Shaomeng Wang. Design of triazole-stapled BCL9 $\alpha$-helical peptides to target the $\beta$-catenin/B-cell CLL/lymphoma 9 (BCL9) protein-protein interaction. *Journal of Medicinal Chemistry*, 55(3):1137–1146, 2 2012. 17

[162] Antonello Pessi, Sandra L. Bixler, Veronica Soloveva, Sheli Radoshitzky, Cary Retterer, Tara Kenny, Rouzbeh Zamani, Glenn Gomba, Dima Gharabeih, Jay Wells, Kelly S. Wetzel, Travis K. Warren, Ginger Donnelly, Sean A. Van Tongeren, Jesse Steffens, Allen J. Duplantier, Christopher D. Kane, Pascale Vicat, Valerie Couturier, Kent E. Kester, John Shiver, Kara Carter, and Sina Bavari. Cholesterol-conjugated stapled peptides inhibit Ebola and Marburg viruses in vitro and in vivo. *Antiviral Research*, 171, 11 2019. 17, 19

[163] Greet De Baets, Joost Schymkowitz, and Frederic Rousseau. Predicting aggregation-Prone sequences in proteins. *Essays in Biochemistry*, 56(1):41–52, 2014. 17

[164] Kyle Trainor, Aron Broom, and Elizabeth M Meiering. Exploring the relationships between protein sequence, structure and solubility. *Current opinion in structural biology*, 42:136–146, 2 2017.

[165] Fabrizio Chiti, Massimo Stefani, Niccolò Taddei, Giampietro Ramponi, and Christopher M. Dobson. Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature*, 424(6950):805–808, 2003.

[166] Benedetta Bolognesi and Gian Gaetano Tartaglia. Physicochemical principles of protein aggregation. In *Progress in Molecular Biology and Translational Science*, volume 117, pages 53–72. Elsevier B.V., 2013. 17

[167] Janice Reimer, Dmitry Shamshurin, Michael Harder, Andriy Yamchuk, Vic Spicer, and Oleg V. Krokhin. Effect of cyclization of N-terminal glutamine and

carbamidomethyl-cysteine (residues) on the chromatographic behavior of peptides in reversed-phase chromatography. *Journal of Chromatography A*, 1218(31):5101–5107, 8 2011. 18

[168] Dirk Chelius, Kay Jing, Alexis Lueras, Douglas S. Rehder, Thomas M. Dillon, Alona Vizel, Rahul S. Rajan, Tiansheng Li, Michael J. Treuheit, and Pavel V. Bondarenko. Formation of pyroglutamic acid from N-terminal glutamic acid in immunoglobulin gamma antibodies. *Analytical Chemistry*, 78(7):2370–2376, 4 2006. 18

[169] Bradley Croy Doak, Björn Over, Fabrizio Giordanetto, and Jan Kihlberg. Oral druggable space beyond the rule of 5: Insights from drugs and clinical candidates, 9 2014. 18, 19

[170] Pascal R. Beauchesne, Nancy S.C. Chung, and Kishor M. Wasan. Cyclosporine A: A review of current oral and intravenous delivery systems. *Drug Development and Industrial Pharmacy*, 33(3):211–220, 1 2007. 18

[171] G.F Cooney, V Jeevanandam, S Choudhury, G Feutren, E.A Mueller, and H.J Eisen. Comparative bioavailability of neoral and sandimmune in cardiac transplant recipients over 1 year. *Transplantation Proceedings*, 30(5):1892–1894, 8 1998. 18

[172] Gregory H. Bird, Navid Madani, Alisa F. Perry, Amy M. Princiotto, Jeffrey G. Supko, Xiaoying He, Evripidis Gavathiotis, Joseph G. Sodroski, and Loren D. Walensky. Hydrocarbon double-stapling remedies the proteolytic instability of a lengthy peptide therapeutic. *Proceedings of the National Academy of Sciences of the United States of America*, 107(32):14093–14098, 8 2010. 18

[173] P. Lundquist and P. Artursson. Oral absorption of peptides and nanoparticles across the human intestine: Opportunities, limitations and studies in human tissues. *Advanced Drug Delivery Reviews*, 106:256–276, 11 2016. 18

[174] Lei Diao and Bernd Meibohm. Pharmacokinetics and pharmacokinetic-pharmacodynamic correlations of therapeutic peptides. *Clinical Pharmacokinetics*, 52(10):855–868, 2013. 18

[175] Yildiz Ozsoy, Sevgi Gungor, and Erdal Cevher. Nasal delivery of high molecular weight drugs, 9 2009. 18, 19

[176] Werner Rubas, Mary E.M. Cromwell, Randall J. Mrsny, Gladys Ingle, and Kathleen A. Elias. An integrated method to determine epithelial transport and bioactivity of oral drug candidates in vitro. *Pharmaceutical Research*, 13(1):23–26, 1996. 18

[177] James L. Madara and Kiertisin Dharmsathaphorn. Occluding junction structure-function relationships in a cultured epithelial monolayer. *Journal of Cell Biology*, 101(6):2124–2733, 12 1985. 18

[178] Yanbin Ji, Subhabrata Majumder, Melissa Millard, Radhika Borra, Tao Bi, Ahmed Y. Elnagar, Nouri Neamati, Alexander Shekhtman, and Julio A. Camarero. In vivo activation of the p53 tumor suppressor pathway by an engineered cyclotide. *Journal of the American Chemical Society*, 135(31):11623–11633, 8 2013. 18

[179] Benjamin J Bruno, Geoffrey D Miller, and Carol S Lim. Basics and recent advances in peptide and protein drug delivery. *Therapeutic delivery*, 4(11):1443–1467, 2014. 19

[180] Jiunn H. Lin and Weirong Wang. Role of Lymphatic System in Subcutaneous Absorption of Therapeutic Proteins. In *Pharmaceutical Sciences Encyclopedia*, pages 1–10. John Wiley & Sons, Inc., Hoboken, NJ, USA, 12 2015. 19

[181] F. Veuillez, Y. N. Kalia, Y. Jacques, J. Deshusses, and P. Buri. Factors and strategies for improving buccal absorption of peptides. *European Journal of Pharmaceutics and Biopharmaceutics*, 51(2):93–109, 2001. 19

[182] Heather A.E. Benson and Sarika Namjoshi. Proteins and peptides: Strategies for delivery to and across the skin. *Journal of Pharmaceutical Sciences*, 97(9):3591–3610, 2008. 19

[183] Yeu Chun Kim, Peter J. Ludovice, and Mark R. Prausnitz. Transdermal delivery enhanced by magainin pore-forming peptide. *Journal of Controlled Release*, 122(3):375–383, 2007. 19

[184] Deepika Mathur, Ayesha Mehta, Priyanka Firmal, Gursimran Bedi, Charu Sood, Ankur Gautam, and Gajendra P. S. Raghava. TopicalPdb: A database of topically delivered peptides. *PLOS ONE*, 13(2):e0190134, 2 2018. 19

[185] Sarika Namjoshi, Istvan Toth, Joanne T. Blanchfield, Nicholas Trotter, Ricardo L. Mancera, and Heather A.E. Benson. Enhanced transdermal peptide delivery and stability by lipid conjugation: Epidermal permeation, stereoselectivity and mechanistic insights. *Pharmaceutical Research*, 31(12):3304–3312, 5 2014. 19

[186] Rima Caccetta, Joanne T. Blanchfield, Jessica Harrison, Istvan Toth, and Heather A.E. Benson. Epidermal penetration of a therapeutic peptide by lipid conjugation; stereoselective peptide availability of a topical diastereomeric lipopeptide. *International Journal of Peptide Research and Therapeutics*, 12(3):327–333, 9 2006. 19

[187] Lisa Tang, Adam M. Persky, Gunther Hochhaus, and Bernd Meibohm. Pharmacokinetic aspects of biotechnology products. *Journal of Pharmaceutical Sciences*, 93(9):2184–2204, 2004. 19

[188] Noriyasu Kamei and Mariko Takeda-Morishita. Brain delivery of insulin boosted by intranasal coadministration with cell-penetrating peptides. *Journal of Controlled Release*, 197:105–110, 1 2015. 19

[189] Allison R. Smego, Philippe Backeljauw, and Gutmark Little Iris. Buccally administered intranasal desmopressin acetate for the treatment of neurogenic diabetes insipidus in infancy. *Journal of Clinical Endocrinology and Metabolism*, 101(5):2084–2088, 5 2016. 20

[190] Kuyper. Free energy calculations on the relative solvation free energies of benzene, anisole, and 1,2,3-trimethoxybenzene: theoretical and experimental analysis of aromatic methoxy solvation | Enhanced Reader. *The Journal of Physical Chemistry*, 95(17), 1991. 22

[191] D. Quentin McDonald and W. Clark Still. AMBER torsional parameters for the peptide backbone. *Tetrahedron Letters*, 33(50):7743–7746, 12 1993. 22, 38

[192] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters, 11 2006. 22, 44, 51

[193] Antonija Kuzmanic, Gregory R. Bowman, Jordi Juarez-Jimenez, Julien Michel, and Francesco L. Gervasio. Investigating Cryptic Binding Sites by Molecular Dynamics Simulations. *ACS Applied Materials and Interfaces*, 2020. 22

[194] Jordi Juárez-Jiménez, Arun A. Gupta, Gogulan Karunanithy, Antonia S.J.S. Mey, Charis Georgiou, Harris Ioannidis, Alessio De Simone, Paul N. Barlow, Alison N. Hulme, Malcolm D. Walkinshaw, Andrew J. Baldwin, and Julien Michel. Dynamic design: Manipulation of millisecond timescale motions on the energy landscape of cyclophilin A. *Chemical Science*, 11(10):2670–2680, 3 2020.

[195] Pattama Wapeesittipan, Antonia S.J.S. Mey, Malcolm D. Walkinshaw, and Julien Michel. Allosteric effects in cyclophilin mutants may be explained by changes in nano-microsecond time scale motions. *Communications Chemistry*, 2(1):1–9, 12 2019. 22

[196] Jérémy Fidelak, Jarek Juraszek, Davide Branduardi, Marc Bianciotto, and Francesco Luigi Gervasio. Free-energy-based methods for binding profile determination in a congeneric series of CDK2 inhibitors. *Journal of Physical Chemistry B*, 114(29):9516–9524, 7 2010. 22

[197] Cameron Abrams and Giovanni Bussi. Enhanced Sampling in Molecular Dynamics Using Metadynamics, Replica-Exchange, and Temperature-Acceleration. *Entropy*, 16(1):163–199, 12 2013. 22

[198] VS Sandeep Inakollu, Daan P. Geerke, Christopher N. Rowley, and Haibo Yu. Polarisable force fields: what do they add in biomolecular simulations?, 4 2020. 23

[199] Paul S. Nerenberg and Teresa Head-Gordon. New developments in force fields for biomolecular simulations, 4 2018. 23

[200] David R.C. Hill. Parallel Random Numbers, Simulation, and Reproducible Research. *Computing in Science and Engineering*, 17(4):66–71, 7 2015. 23

[201] Vincent Kräutler, Wilfred F. Van Gunsteren, and Philippe H. Hünenberger. A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *Journal of Computational Chemistry*, 22(5):501–508, 4 2001. 25

[202] Andreas Kukol. Lipid Models for United-Atom Molecular Dynamics Simulations of Proteins. *Journal of Chemical Theory and Computation*, 5(3):615–626, 3 2009. 25

[203] H. J.C. Berendsen, J. P.M. Postma, W. F. Van Gunsteren, A. Dinola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8):3684–3690, 10 1984. 26

[204] Hans C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics*, 72(4):2384–2393, 2 1980. 26

[205] Chuan Tian, Koushik Kasavajhala, Kellon Belfon, Lauren Raguette, He Huang, Angela Migues, John Bickel, Yuzhang Wang, Jorge Pincay, Qin Wu, and Carlos Simmerling. ff19SB: Amino-acid specific protein backbone parameters trained against quantum mechanics energy surfaces in solution, 6 2019. 30, 40, 41, 44

[206] Matthias Buck, Sabine Bouguet-Bonnet, Richard W. Pastor, and Alexander D. MacKerell. Importance of the CMAP correction to the CHARMM22 protein force field: Dynamics of hen lysozyme. *Biophysical Journal*, 90(4), 2006. 30

[207] Peter Y Chou and Gerald D Pasman. Conformational Parameters for Amino Acids in Helical, /3-Sheet, and Random Coil Regions Calculated from Proteins! *Biochemistry*, 13(2):211–222, 1974. 32

[208] B. Robson and R. H. Pain. Analysis of the code relating sequence to conformation in proteins: Possible implications for the mechanism of formation of helical regions. *Journal of Molecular Biology*, 58(1):237–257, 5 1971.

[209] Richard C. Garratt, William R. Taylor, and Janet M. Thornton. The influence of tertiary structure on secondary structure prediction. Accessibility versus predictability for $\beta$-structure. *FEBS Letters*, 188(1):59–62, 8 1985. 32

[210] G. Deléage and B. Roux. An algorithm for protein secondary structure prediction based on class prediction. *"Protein Engineering, Design and Selection"*, 1(4):289–294, 8 1987. 32

[211] S. Selvaraj and M. Michael Gromiha. An analysis of the amino acid clustering pattern in ($\alpha/\beta$)8 barrel proteins. *Journal of Protein Chemistry*, 17(5):407–415, 1998. 32

[212] M. MICHAEL GROMIHA and P.K. PONNUSWAMY. Prediction of protein secondary structures from their hydrophobic characteristics. *International Journal of Peptide and Protein Research*, 45(3):225–240, 1 2009. 32

[213] Jong Park, Sabine Dietmann, Andreas Heger, and Liisa Holm. Estimating the significance of sequence order in protein secondary structure and prediction. *Bioinformatics*, 16(11):978–987, 2000. 32

[214] W. Kabsch and C. Sander. On the use of sequence homologies to predict protein structure: Identical pentapeptides can have completely different conformations. *Proceedings of the National Academy of Sciences of the United States of America*, 81(4 I):1075–1078, 2 1984. 32

[215] Wafaa Wardah, M. G.M. Khan, Alok Sharma, and Mahmood A. Rashid. Protein secondary structure prediction using neural networks and deep learning: A review, 8 2019. 33

[216] Christopher Bystroff, Vesteinn Thorsson, and David Baker. HMMSTR: A hidden Markov model for local sequence-structure correlations in proteins. *Journal of Molecular Biology*, 301(1):173–190, 8 2000. 33

[217] X.-D Sun and R.-B Huang. Prediction of protein structural classes using support vector machines. *Amino Acids*, 30:469–475, 2006. 33

[218] Biljana Mojsoska and Håvard Jenssen. Peptides and peptidomimetics for antimicrobial drug design, 7 2015. 33

[219] Edmund Lin and M. Scott Shell. Convergence and heterogeneity in peptide folding with replica exchange molecular dynamics. *Journal of Chemical Theory and Computation*, 5(8):2062–2073, 8 2009. 33, 38

[220] H Nymeyer and A. E. Garcia. Simulation of the folding equilibrium of -helical peptides: A comparison of the generalized Born approximation with explicit solvent. *Proceedings of the National Academy of Sciences*, 100(24):13934–13939, 2003. 33

[221] Chao Chen, Yuanxin Tian, Xiaoyong Zou, Peixiang Cai, and Jinyuan Mo. Prediction of protein secondary structure content using support vector machine. *Talanta*, 71(5):2069–2073, 3 2007. 33

[222] Jian Wang, Tailang Yin, Xuwen Xiao, Dan He, Zhidong Xue, Xinnong Jiang, and Yan Wang. StraPep: a structure database of bioactive peptides. *Database*, 2018, 1 2018. 33

[223] Sandeep Singh, Kumardeep Chaudhary, Sandeep Kumar Dhanda, Sherry Bhalla, Salman Sadullah Usmani, Ankur Gautam, Abhishek Tuknait, Piyush Agrawal, Deepika Mathur, and Gajendra P.S. Raghava. SATPdb: A database of structurally annotated therapeutic peptides. *Nucleic Acids Research*, 44(D1):D1119–D1126, 1 2015. 33

[224] MM Seibert, Alexandra Patriksson, Berk Hess, and David van der Spoel. Reproducible polypeptide folding and structure prediction using molecular dynamics simulations. *Journal of molecular biology*, 354(1):173–183, 2005. 33

[225] Jovan Damjanovic, Jiayuan Miao, He Huang, and Yu Shan Lin. Elucidating Solution Structures of Cyclic Peptides Using Molecular Dynamics Simulations, 2021. 35

[226] Johannes Kästner. Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(6):932–942, 11 2011. 35

[227] Shankar Kumar, John M. Rosenberg, Djamal Bouzida, Robert H. Swendsen, and Peter A. Kollman. Multidimensional free-energy calculations using the weighted histogram analysis method. *Journal of Computational Chemistry*, 16(11):1339–1350, 11 1995. 35

[228] Alessandro Laio, Antonio Rodriguez-Fortea, Francesco Luigi Gervasio, Matteo Ceccarelli, and Michele Parrinello. Assessing the accuracy of metadynamics. *Journal of Physical Chemistry B*, 109(14):6714–6721, 4 2005. 35

[229] Xavier Periole and Alan E. Mark. Convergence and sampling efficiency in replica exchange simulations of peptide folding in explicit solvent. *Journal of Chemical Physics*, 126(1):014903, 1 2007. 37, 71

[230] S Gnanakaran, Hugh Nymeyer, John Portman, Kevin Y. Sanbonmatsu, and Angel E. García. Peptide folding simulations, 2003.

[231] Nicolae Viorel Buchete and Gerhard Hummer. Peptide folding kinetics from replica exchange molecular dynamics. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 77(3), 2008. 37, 71

[232] Pu Liu, Byungchan Kim, Richard A. Friesner, and B. J. Berne. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proceedings of the National Academy of Sciences of the United States of America*, 102(39):13749–13754, 9 2005. 38, 74

[233] Vladimiras Oleinikovas, Giorgio Saladino, Benjamin P. Cossins, and Francesco L. Gervasio. Understanding Cryptic Pocket Formation in Protein Targets by Enhanced Sampling Simulations. *Journal of the American Chemical Society*, 138(43):14257–14263, 11 2016.

[234] Katja Ostermeir and Martin Zacharias. Hamiltonian replica-exchange simulations with adaptive biasing of peptide backbone and side chain dihedral angles. *Journal of Computational Chemistry*, 35(2):150–158, 1 2014. 38

[235] Wilfred F. van Gunsteren, Xavier Daura, and Alan E. Mark. GROMOS Force Field. *Encyclopedia of Computational Chemistry*, 2:1211–1216, 4 2002. 38

[236] K Vanommeslaeghe, E Hatcher, C Acharya, S Kundu, S Zhong, J Shim, E Darian, O Guvench, P Lopes, I Vorobyov, A D Mackerell, and Jr. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of computational chemistry*, 31(4):671–90, 3 2010. 38

[237] Andreas Bernkop-Schnürch. The use of inhibitory agents to overcome the enzymatic barrier to perorally administered therapeutic peptides and proteins. *Journal of Controlled Release*, 52(1-2):1–16, 1998. 38

[238] Scott J. Weiner, Peter A. Kollman, Dzung T. Nguyen, and David A. Case. An all atom force field for simulations of proteins and nucleic acids. *Journal of Computational Chemistry*, 7(2):230–252, 1986. 38

[239] Dail E. Chapman, Jonathan K. Steck, and Paul S. Nerenberg. Optimizing protein-protein van der waals interactions for the AMBER ff9x/ff12 force field. *Journal of Chemical Theory and Computation*, 10(1):273–281, 1 2014.

[240] Urmi Doshi and Donald Hamelberg. Reoptimization of the AMBER force field parameters for peptide bond (Omega) torsions using accelerated molecular dynamics. *Journal of Physical Chemistry B*, 113(52):16590–16595, 12 2009.

[241] Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L. Klepeis, Ron O. Dror, and David E. Shaw. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function, and Bioinformatics*, 78(8):1950–1958, 6 2010. 38, 44

[242] Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case. Development and testing of a general Amber force field. *Journal of Computational Chemistry*, 25(9):1157–1174, 7 2004. 51

[243] Yuan-Ping Pang. FF12MC: A revised AMBER forcefield and new protein simulation protocol. *Proteins: Structure, Function, and Bioinformatics*, 84(10):1490–1516, 10 2016. 38

[244] Oliver F. Lange, David Van Der Spoel, and Bert L. De Groot. Scrutinizing molecular mechanics force fields on the submicrosecond timescale with NMR Data. *Biophysical Journal*, 99(2):647–655, 7 2010. 38

[245] Stefano Piana, Kresten Lindorff-Larsen, and David E. Shaw. How robust are protein folding simulations with respect to force field parameterization? *Biophysical Journal*, 100(9):L47–L49, 5 2011. 38

[246] Elio A. Cino, Wing Yiu Choy, and Mikko Karttunen. Comparison of secondary structure formation using 10 different force fields in microsecond molecular dynamics simulations. *Journal of Chemical Theory and Computation*, 8(8):2725–2740, 8 2012. 39

[247] Kyle A. Beauchamp, Yu Shan Lin, Rhiju Das, and Vijay S. Pande. Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements. *Journal of Chemical Theory and Computation*, 8(4):1409–1414, 4 2012. 39

[248] João Henriques, Carolina Cragnell, and Marie Skepö. Molecular Dynamics Simulations of Intrinsically Disordered Proteins: Force Field Evaluation and Comparison with Experiment. *Journal of Chemical Theory and Computation*, 11(7):3420–3431, 7 2015. 39

[249] Paul Robustelli, Stefano Piana, and David E. Shaw. Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences of the United States of America*, 115(21):E4758–E4766, 5 2018. 40

[250] Ya Gao, Chaomin Zhang, Xianwei Wang, and Tong Zhu. A test of AMBER force fields in predicting the secondary structure of A-helical and B-hairpin peptides. *Chemical Physics Letters*, 679:112–118, 7 2017. 40

[251] Timo Graen, Martin Hoefling, and Helmut Grubmüller. AMBER-DYES: Characterization of charge fluctuations and force field parameterization of fluorescent dyes for molecular dynamics simulations. *Journal of Chemical Theory and Computation*, 10(12):5505–5512, 12 2014. 43, 153

[252] David Gfeller, Olivier Michielin, and Vincent Zoete. SwissSidechain: a molecular and structural database of non-natural sidechains. *Nucleic Acids Research*, 41(D1):D327–D332, 10 2012. 43

[253] George A. Khoury, Jeff P. Thompson, James Smadbeck, Chris A. Kieslich, and Christodoulos A. Floudas. Forcefield-PTM: Ab initio charge and AMBER forcefield parameters for frequently occurring post-translational modifications. *Journal of Chemical Theory and Computation*, 9(12):5653–5674, 12 2013. 43

[254] George A. Khoury, James Smadbeck, Phanourios Tamamis, Andrew C. Vandris, Chris A. Kieslich, and Christodoulos A. Floudas. Forcefield-NCAA: Ab initio charge parameters to aid in the discovery and design of therapeutic proteins and peptides with

unnatural amino acids and their application to complement inhibitors of the compstatin family. *ACS Synthetic Biology*, 3(12):855–869, 12 2014. 43

[255] Xiaowen Wang and Wenjin Li. Development and Testing of Force Field Parameters for Phenylalanine and Tyrosine Derivatives. *Frontiers in Molecular Biosciences*, 7:608931, 12 2020. 43

[256] Addison K. Smith, Joshua W. Wilkerson, and Thomas A. Knotts. Parameterization of Unnatural Amino Acids with Azido and Alkynyl R-Groups for Use in Molecular Simulations. *Journal of Physical Chemistry A*, 124(30):6246–6253, 7 2020. 44

[257] Wendy D. Cornell, Piotr Cieplak, Christopher I. Bayly, Ian R. Gould, Kenneth M. Merz, David M. Ferguson, David C. Spellmeyer, Thomas Fox, James W. Caldwell, and Peter A. Kollman. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society*, 117(19):5179–5197, 5 1995. 44

[258] Bruce L. Bush, Christopher I. Bayly, and Thomas A. Halgren. Consensus bond-charge increments fitted to electrostatic potential or field of many compounds: Application to MMFF94 training set. *Journal of Computational Chemistry*, 20(14):1495–1516, 11 1999. 45

[259] Michael K. Gilson, Hillary S.R. Gilson, and Michael J. Potter. Fast Assignment of Accurate Partial Atomic Charges: An Electronegativity Equalization Method that Accounts for Alternate Resonance Forms. *Journal of Chemical Information and Computer Sciences*, 43(6):1982–1997, 11 2003. 45

[260] Donald E. Williams. Net Atomic Charge and Multipole Models for the ab Initio Molecular Electric Potential. In *Reviews in Computational Chemistry*, pages 219–271. John Wiley & Sons, Ltd, 1 2007. 46

[261] Christopher I. Bayly, Piotr Cieplak, Wendy D. Cornell, and Peter A. Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *Journal of Physical Chemistry*, 97(40):10269–10280, 1993. 46

[262] François-Yves Dupradeau, Adrien Pigache, Thomas Zaffran, Corentin Savineau, Rodolphe Lelong, Nicolas Grivel, Dimitri Lelong, Wilfried Rosanski, and Piotr Cieplak. The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building. *Physical Chemistry Chemical Physics*, 12(28):7821, 7 2010. 46, 49

[263] Case D. A., Ben-Shalom. I.Y., Brozell. S.R, Cerutti. D. S., Cheatham. T.E, Cruzeiro V.W.D, Darden. T. A, R. E. Duke, D. Ghoreishi, M. K. Gilson, H. Gohlke, A. W.

Goetz, D. Greene, R. Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T. S. Lee, and LeGra S. AMBER, 2018. 49, 51

[264] Frisch M. J., Trucks, Schlegel, Scuseria, Robb, and J. R. Cheeseman. gaussian09, 2016. 50

[265] Andrea Cesari, Sandro Bottaro, Kresten Lindorff-Larsen, Pavel Banáš, Jiří Šponer, and Giovanni Bussi. Fitting Corrections to an RNA Force Field Using Experimental Data. *Journal of Chemical Theory and Computation*, 15(6):3425–3431, 6 2019. 51

[266] Matthias Diem and Chris Oostenbrink. Hamiltonian Reweighing to Refine Protein Backbone Dihedral Angle Parameters in the GROMOS Force Field. *Journal of Chemical Information and Modeling*, 60(1):279–288, 1 2020.

[267] Laurie A. Christianson, Melissa J. Lucero, Daniel H. Appella, Daniel A. Klein, and Samuel H. Gellman. Improved treatment of cyclic ?-amino acids and successful prediction of ?-peptide secondary structure using a modified force field: AMBER*C. *Journal of Computational Chemistry*, 21(9):763–773, 7 2000. 51

[268] Robin M. Betz and Ross C. Walker. Paramfit: Automated optimization of force field parameters for molecular dynamics simulations. *Journal of Computational Chemistry*, 36(2):79–87, 1 2015. 51, 53, 54

[269] C J Woods and J M Michel. Sire: An advanced, multiscale, molecular simulation framework. *siremol.org*, 2016. 64

[270] F. Vitalini, F. Noé, and B.G. G. Keller. Molecular dynamics simulations data of the twenty encoded amino acids in different force fields. *Data in Brief*, 7:582–590, 6 2016. 65, 67

[271] Landrum G. RDkit, 2010. 69

[272] J. Santeri Puranen, Mikko J. Vainio, and Mark S. Johnson. Accurate conformation-dependent molecular electrostatic potentials for high-throughput in silico drug discovery. *Journal of Computational Chemistry*, 31(8):1722–1732, 6 2010. 69

[273] Maria A. Miteva, Frederic Guyon, and Pierre Tufféry. Frog2: Efficient 3D conformation ensemble generator for small compounds. *Nucleic Acids Research*, 38(SUPPL. 2):W622, 5 2010. 69

[274] Jiři Vymětal and Jiři Vondrášek. Parametrization of 2,2,2-trifluoroethanol based on the generalized AMBER force field provides realistic agreement between experimental and calculated properties of pure liquid as well as water-mixed solutions. *Journal of Physical Chemistry B*, 118(35):10390–10404, 2014. 71, 78, 79, 80, 81

[275] Daniel Sindhikara, Yilin Meng, and Adrian E. Roitberg. Exchange frequency in replica exchange molecular dynamics. *Journal of Chemical Physics*, 128(2):024103, 1 2008. 72

[276] Hironori Kokubo and Yuko Okamoto. Prediction of transmembrane helix configurations by replica-exchange simulations. *Chemical Physics Letters*, 383(3-4):397–402, 1 2004. 73

[277] C. Landon, H. Meudal, N. Boulanger, P. Bulet, and F. Vovelle. Solution structures of stomoxyn and spinigerin, two insect antimicrobial peptides with an alpha-helical conformation. *Biopolymers*, 81:92–103, 2006. 73

[278] Lingle Wang, Richard A. Friesner, and B. J. Berne. Replica exchange with solute scaling: A more efficient version of replica exchange with solute tempering (REST2). *Journal of Physical Chemistry B*, 115(30):9431–9438, 8 2011. 74

[279] Sunhwan Jo and Wei Jiang. A generic implementation of replica exchange with solute tempering (REST2) algorithm in NAMD for complex biophysical simulations. *Computer Physics Communications*, 197:304–311, 12 2015. 74

[280] ParmEd — ParmEd documentation. 74

[281] Paul D. Thomas and Ken A. Dill. Local and nonlocal interactions in globular proteins and mechanisms of alcohol denaturation. *Protein Science*, 2(12):2050–2065, 12 1993. 78

[282] Nami Hirota, Kazuko Mizuno, and Yuji Goto. Group additive contributions to the alcohol-induced $\alpha$-helix formation of melittin: Implication for the mechanism of the alcohol effects on proteins. *Journal of Molecular Biology*, 275(2):365–378, 1 1998. 78

[283] Garry W. Buchko, Avijita Jain, Matthew L. Reback, and Wendy J. Shaw. Structural characterization of the model amphipathic peptide Ac-LKKLLKLLKKLLKL-NH2 in aqueous solution and with 2,2,2-trifluoroethanol and 1,1,1,3,3,3-hexafluoroisopropanol. *Canadian Journal of Chemistry*, 91(6):406–413, 6 2013. 78

[284] Michael Jackson and Henry H. Mantsch. Halogenated alcohols as solvents for proteins: FTIR spectroscopic studies. *Biochimica et Biophysica Acta (BBA)/Protein Structure and Molecular*, 1118(2):139–143, 1 1992. 78

[285] Arthur Cammers-Goodwin, Thomas J. Allen, Sherri L. Oslick, Kim F. McClure, Janette H. Lee, and D. S. Kemp. Mechanism of Stabilization of Helical Conformations of Polypeptides by Water Containing Trifluoroethanol. *Journal of the American Chemical Society*, 118(13):3082–3090, 1 1996. 78

[286] Jiří Vymětal, Lucie Bednárová, and Jiří Vondrášek. Effect of TFE on the Helical Content of AK17 and HAL-1 Peptides: Theoretical Insights into the Mechanism of Helix Stabilization. *The Journal of Physical Chemistry B*, 120(6):1048–1059, 2 2016. 78

[287] H D Baehr. Vapor Pressure and Liquid and Gas Densities of 2, 2, 2-Trifluoroethanol. *International Journal of Thermophysics*, 10(3):577–589, 1989. 79

[288] Rajappa Chitra and Paul E. Smith. Properties of 2,2,2-trifluoroethanol and water mixtures. *The Journal of Chemical Physics*, 114(1):426, 2001. 79, 80, 81

[289] Alexandra Patriksson and David van der Spoel. A temperature predictor for parallel tempering simulations. *Physical Chemistry Chemical Physics*, 10(15):2073, 4 2008. 82

[290] Jianwen A. Feng, Jeff Kao, and Garland R. Marshall. A second look at mini-protein stability: Analysis of FSD-1 using circular dichroism, differential scanning calorimetry, and simulations. *Biophysical Journal*, 97(10):2803–2810, 11 2009. 84

[291] Funda Meric-Bernstam, Mansoor N. Saleh, Jeffrey R. Infante, Sanjay Goel, Gerald Steven Falchook, Geoffrey Shapiro, Ki Y Chung, Robert Martin Conry, David S. Hong, Judy Sing-Zan Wang, Ulrich Steidl, Loren D. Walensky, Vincent Guerlavais, Marie Payton, D. Allen Annis, Manuel Aivado, and Manish R. Patel. Phase I trial of a novel stapled peptide ALRN-6924 disrupting MDMX- and MDM2-mediated inhibition of WT p53 in patients with solid tumors and lymphomas. . *Journal of Clinical Oncology*, 35(15_suppl):2505–2505, 5 2017. 86

[292] Vilmos Kertesz, Marissa Vavrek, Carol Freddo, and Gary J. Van Berkel. Spatial profiling of stapled $\alpha$–helical peptide ATSP-7041 in mouse whole-body thin tissue sections using droplet-based liquid microjunction surface sampling-HPLC-ESI–MS/MS. *International Journal of Mass Spectrometry*, 437:17–22, 3 2019. 86

[293] Federico Bernal, Andrew F Tyler, Stanley J Korsmeyer, Loren D Walensky, and Gregory L Verdine. Reactivation of the p53 Tumor Suppressor Pathway by a Stapled p53 Peptide. *J. AM. CHEM. SOC*, 129:2456–2457, 2007. 86, 87, 88

[294] Case Ross C Walker, David A, Roitberg Kenneth Merz Pengfei Li, and Adrian M. Amber 2020 Reference Manual. Technical report. 86

[295] Chuan Tian, Koushik Kasavajhala, Kellon A.A. Belfon, Lauren Raguette, He Huang, Angela N. Migues, John Bickel, Yuzhang Wang, Jorge Pincay, Qin Wu, and Carlos Simmerling. Ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *Journal of Chemical Theory and Computation*, 16(1):528–552, 1 2020. 86

[296] Gabor Nagy, Maxim Igaev, Nykola C. Jones, Søren V. Hoffmann, and Helmut Grubmüller. SESCA: Predicting Circular Dichroism Spectra from Protein Molecular Structures. *Journal of Chemical Theory and Computation*, 15(9):5087–5102, 9 2019. 88, 91

[297] Federico Bernal, Mark Wade, Marina Godes, Tina N. Davis, David G. Whitehead, Andrew L. Kung, Geoffrey M. Wahl, and Loren D. Walensky. A Stapled p53 Helix Overcomes HDMX-Mediated Suppression of p53. *Cancer Cell*, 18(5):411–422, 11 2010. 89

[298] Woo Seok Byun, Sin Woo Heo, Gunhee Jo, Jae Won Kim, Sarang Kim, Sujie Lee, Hye Eun Park, and Jea Hyun Baek. Is coronavirus disease (COVID-19) seasonal? A critical analysis of empirical and epidemiological studies at global and local scales. *Environmental Research*, 196:110972, 5 2021. 93

[299] You Li, Xin Wang, and Harish Nair. Global Seasonality of Human Seasonal Coronaviruses: A Clue for Postpandemic Circulating Season of Severe Acute Respiratory Syndrome Coronavirus 2? *The Journal of Infectious Diseases*, 222(7):1090–1097, 9 2020.

[300] Cory Merow and Mark C. Urban. Seasonality and uncertainty in global COVID-19 growth rates. *Proceedings of the National Academy of Sciences of the United States of America*, 117(44):27456–27464, 11 2020. 93

[301] Neeraja Ravi, Dana L. Cortade, Elaine Ng, and Shan X. Wang. Diagnostics for SARS-CoV-2 detection: A comprehensive review of the FDA-EUA COVID-19 testing landscape. *Biosensors and Bioelectronics*, 165:112454, 10 2020. 93

[302] Danielle C. Morgan, Caroline Morris, Amit Mahindra, Connor M. Blair, Gonzalo Tejeda, Imogen Herbert, Matthew L. Turnbull, Gauthier Lieber, Brian J. Willett, Nicola Logan, Brian Smith, Andrew B. Tobin, David Bhella, George Baillie, and Andrew G. Jamieson. Stapled <scp>ACE2</scp> peptidomimetics designed to target the <scp>SARS-CoV</scp> -2 spike protein do not prevent virus internalization. *Peptide Science*, page e24217, 1 2021. 94, 114, 115, 141, 147

[303] G. Zhang, S. Pomplun, A. R. Loftis, X. Tan, A. Loas, and B. L. Pentelute. Investigation of ACE2 N-terminal fragments binding to SARS-CoV-2 Spike RBD. *bioRxiv*, page 2020.03.19.999318, 3 2020. 94, 106, 114, 115

[304] Christopher W Wood, Jack W Heal, Andrew R Thomson, Gail J Bartlett, Amaurys Á Ibarra, R Leo Brady, Richard B Sessions, and Derek N Woolfson. ISAMBARD: an open-source computational environment for biomolecular analysis, modelling and design. *Bioinformatics*, 33(19):3043–3050, 10 2017. 94, 133

[305] Shibo Jiang, Zhengli Shi, Yuelong Shu, Jingdong Song, George F. Gao, Wenjie Tan, and Deyin Guo. A distinct name is needed for the new coronavirus, 3 2020. 94

[306] Zhixing Zhu, Xihua Lian, Xiaoshan Su, Weijing Wu, Giuseppe A. Marraro, and Yiming Zeng. From SARS and MERS to COVID-19: A brief summary and comparison of severe acute respiratory infections caused by three highly pathogenic human coronaviruses, 8 2020. 94, 95, 100

[307] Roujian Lu, Xiang Zhao, Juan Li, Peihua Niu, Bo Yang, Honglong Wu, Wenling Wang, Hao Song, Baoying Huang, Na Zhu, Yuhai Bi, Xuejun Ma, Faxian Zhan, Liang Wang, Tao Hu, Hong Zhou, Zhenhong Hu, Weimin Zhou, Li Zhao, Jing Chen, Yao Meng, Ji Wang, Yang Lin, Jianying Yuan, Zhihao Xie, Jinmin Ma, William J. Liu, Dayan Wang, Wenbo Xu, Edward C. Holmes, George F. Gao, Guizhen Wu, Weijun Chen, Weifeng Shi, and Wenjie Tan. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet*, 395(10224):565–574, 2 2020. 95

[308] Abhigyan Choudhury and Suprabhat Mukherjee. In silico studies on the comparative characterization of the interactions of SARS-CoV-2 spike glycoprotein with ACE-2 receptor homologs and human TLRs. *Journal of Medical Virology*, 92(10):2105–2113, 10 2020. 95, 99

[309] Shutoku Matsuyama, Naganori Nao, Kazuya Shirato, Miyuki Kawase, Shinji Saito, Ikuyo Takayama, Noriyo Nagata, Tsuyoshi Sekizuka, Hiroshi Katoh, Fumihiro Kato, Masafumi Sakata, Maino Tahara, Satoshi Kutsuna, Norio Ohmagari, Makoto Kuroda, Tadaki Suzuki, Tsutomu Kageyama, and Makoto Takeda. Enhanced isolation of SARS-CoV-2 by TMPRSS2- expressing cells. *Proceedings of the National Academy of Sciences of the United States of America*, 117(13):7001–7003, 3 2020. 99

[310] Markus Hoffmann, Hannah Kleine-Weber, Simon Schroeder, Nadine Krüger, Tanja Herrler, Sandra Erichsen, Tobias S. Schiergens, Georg Herrler, Nai Huei Wu, Andreas Nitsche, Marcel A. Müller, Christian Drosten, and Stefan Pöhlmann. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*, 181(2):271–280, 4 2020. 95, 110

[311] I. Hamming, W. Timens, M. L.C. Bulthuis, A. T. Lely, G. J. Navis, and H. van Goor. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *Journal of Pathology*, 203(2):631–637, 6 2004. 96

[312] Josef M. Penninger, Maria B. Grant, and Joseph J.Y. Sung. The Role of Angiotensin Converting Enzyme 2 in Modulating Gut Microbiota, Intestinal Inflammation, and Coronavirus Infection, 1 2021. 96

[313] Evgeny Krissinel and Kim Henrick. Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology*, 372(3):774–797, 9 2007. 97, 160

[314] Yushun Wan, Jian Shang, Rachel Graham, Ralph S. Baric, and Fang Li. Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *Journal of Virology*, 94(7), 3 2020. 97

[315] Marijn N. Maas, Jordi C.J. Hintzen, Philipp M.G. Löffler, and Jasmin Mecinović. Targeting SARS-CoV-2 spike protein by stapled hACE2 peptides. *Chemical Communications*, 57(26):3283–3286, 4 2021. 97, 106

[316] Alexandra C. Walls, Young Jun Park, M. Alejandra Tortorici, Abigail Wall, Andrew T. McGuire, and David Veesler. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*, 181(2):281–292, 4 2020. 98, 101, 102

[317] Daniel Wrapp, Nianshuang Wang, Kizzmekia S. Corbett, Jory A. Goldsmith, Ching-Lin Hsieh, Olubukola Abiona, Barney S. Graham, and Jason S. McLellan. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, 367(6483):1260–1263, 3 2020. 98

[318] Puja Adhikari, Neng Li, Matthew Shin, Nicole F. Steinmetz, Reidun Twarock, Rudolf Podgornik, and Wai Yim Ching. Intra- And intermolecular atomic-scale interactions in the receptor binding domain of SARS-CoV-2 spike protein: Implication for ACE2 receptor binding. *Physical Chemistry Chemical Physics*, 22(33):18272–18283, 9 2020.

[319] Jiahua He, Huanyu Tao, Yumeng Yan, Sheng-You Huang, and Yi Xiao. Molecular Mechanism of Evolution and Human Infection with SARS-CoV-2. *Viruses*, 12(4):428, 4 2020.

[320] Yongfei Cai, Jun Zhang, Tianshu Xiao, Hanqin Peng, Sarah M Sterling, Richard M Walsh Jr, Shaun Rawson, Sophia Rits-Volloch, and Bing Chen. Distinct conformational states of SARS-CoV-2 spike protein. 2020.

[321] Han-ul Kim and Hyun Suk Jung. Cryo-EM as a powerful tool for drug discovery: recent structural based studies of SARS-CoV-2. *Applied Microscopy*, 51(1):13, 12 2021.

[322] Zunlong Ke, Joaquin Oton, Kun Qu, Mirko Cortese, Vojtech Zila, Lesley McKeane, Takanori Nakane, Jasenko Zivanov, Christopher J. Neufeldt, Berati Cerikan, John M. Lu, Julia Peukes, Xiaoli Xiong, Hans Georg Kräusslich, Sjors H.W. Scheres, Ralf Bartenschlager, and John A.G. Briggs. Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature*, 588(7838):498–502, 12 2020. 111

[323] Cong Xu, Yanxing Wang, Caixuan Liu, Chao Zhang, Wenyu Han, Xiaoyu Hong, Yifan Wang, Qin Hong, Shutian Wang, Qiaoyu Zhao, Yalei Wang, Yong Yang, Kaijian Chen, Wei Zheng, Liangliang Kong, Fangfang Wang, Qinyu Zuo, Zhong Huang, and Yao Cong. Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM. *Science Advances*, 7(1), 1 2021. 98

[324] Razie Amraie, Marc A. Napoleon, Wenqing Yin, Jacob Berrigan, Ellen Suder, Grace Zhao, Judith Olejnik, Suryaram Gummuluru, Elke Muhlberger, Vipul Chitalia, and Nader Rahimi. CD209L/L-SIGN and CD209/DC-SIGN act as receptors for SARS-CoV-2 and are differentially expressed in lung and kidney epithelial and endothelial cells. *bioRxiv*, page 2020.06.22.165803, 6 2020. 99

[325] Michel Thépaut, Joanna Luczkowiak, Corinne Vivès, Nuria Labiod, Isabelle Bally, Fátima Lasala, Yasmina Grimoire, Daphna Fenel, Sara Sattin, Nicole Thielens, Guy Schoehn, Anna Bernardi, Rafael Delgado, and Franck Fieschi. DC/L-SIGN recognition of spike glycoprotein promotes SARS-CoV-2 trans-infection and can be inhibited by a glycomimetic antagonist. *PLoS Pathogens*, 17(5):e1009576, 5 2021. 99

[326] Christian JA Sigrist, Alan Bridge, and Philippe Le Mercier. A potential role for integrins in host cell entry by SARS-CoV-2. *Antiviral Research*, 177:104759, 5 2020. 99

[327] Chao Gao, Junwei Zeng, Nan Jia, Kathrin Stavenhagen, Yasuyuki Matsumoto, Hua Zhang, Jiang Li, Adam J Hume, Elke Mühlberger, Irma van Die, Julian Kwan, Kelan Tantisira, Andrew Emili, and Richard D Cummings. SARS-CoV-2 Spike Protein Interacts with Multiple Innate Immune Receptors. *bioRxiv : the preprint server for biology*, 7 2020. 99

[328] Mihkel Örd, Ilona Faustova, and Mart Loog. The sequence at Spike S1/S2 site enables cleavage by furin and phospho-regulation in SARS-CoV2 but not in SARS-CoV1 or MERS-CoV. *Scientific Reports*, 10(1):1–10, 12 2020. 99

[329] Dorothea Bestle, Miriam Ruth Heindl, Hannah Limburg, Thuy van Lam van, Oliver Pilgram, Hong Moulton, David A. Stein, Kornelia Hardes, Markus Eickmann, Olga Dolnik, Cornelius Rohde, Hans Dieter Klenk, Wolfgang Garten, Torsten Steinmetzer, and Eva Böttcher-Friebertshäuser. TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. *Life Science Alliance*, 3(9), 2020. 99

[330] James L. Daly, Boris Simonetti, Katja Klein, Kai En Chen, Maia Kavanagh Williamson, Carlos Antón-Plágaro, Deborah K. Shoemark, Lorena Simón-Gracia,

Michael Bauer, Reka Hollandi, Urs F. Greber, Peter Horvath, Richard B. Sessions, Ari Helenius, Julian A. Hiscox, Tambet Teesalu, David A. Matthews, Andrew D. Davidson, Brett M. Collins, Peter J. Cullen, and Yohei Yamauchi. Neuropilin-1 is a host factor for SARS-CoV-2 infection. *Science*, 370(6518):861–865, 11 2020. 100

[331] Ioannis Kyrou, Harpal S. Randeva, Demetrios A. Spandidos, and Emmanouil Karteris. Not only ACE2—the quest for additional host cell mediators of SARS-CoV-2 infection: Neuropilin-1 (NRP1) as a novel SARS-CoV-2 host cell entry mediator implicated in COVID-19, 12 2021. 100

[332] Ludovico Cantuti-Castelvetri, Ravi Ojha, Liliana D. Pedro, Minou Djannatian, Jonas Franz, Suvi Kuivanen, Franziska van der Meer, Katri Kallio, Tuğberk Kaya, Maria Anastasina, Teemu Smura, Lev Levanov, Leonora Szirovicza, Allan Tobi, Hannimari Kallio-Kokko, Pamela Österlund, Merja Joensuu, Frédéric A. Meunier, Sarah J. Butcher, Martin Sebastian Winkler, Brit Mollenhauer, Ari Helenius, Ozgun Gokce, Tambet Teesalu, Jussi Hepojoki, Olli Vapalahti, Christine Stadelmann, Giuseppe Balistreri, and Mikael Simons. Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. *Science*, 370(6518), 11 2020. 100

[333] Brian Tripet, Megan W. Howard, Michael Jobling, Randall K. Holmes, Kathryn V. Holmes, and Robert S. Hodges. Structural characterization of the SARS-coronavirus spike S fusion protein core. *Journal of Biological Chemistry*, 279(20):20836–20849, 5 2004. 100

[334] B. Coutard, C. Valle, X. de Lamballerie, B. Canard, N. G. Seidah, and E. Decroly. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Research*, 176:104742, 4 2020. 100

[335] Tiffany Tang, Miya Bidon, Javier A. Jaimes, Gary R. Whittaker, and Susan Daniel. Coronavirus membrane fusion mechanism offers a potential target for antiviral development, 6 2020. 100

[336] Oliver C Grant, David Montgomery, Keigo Ito, and Robert J Woods. 3D Models of glycosylated SARS-CoV-2 spike protein suggest challenges and opportunities for vaccine development. *bioRxiv*, page 2020.04.07.030445, 4 2020. 101, 103, 150

[337] Yasunori Watanabe, Joel Allen, Daniel Wrapp, Jason McLellan, and Max Crispin. Site-specific analysis of the SARS-CoV-2 glycan shield. *bioRxiv : the preprint server for biology*, 2020. 101, 150

[338] Xiaohui Zhao, Huan Chen, and Hongliang Wang. Glycans of SARS-CoV-2 Spike Protein in Virus Infection and Antibody Production. *Frontiers in Molecular Biosciences*, 8:53, 4 2021. 101

[339] Maria Pia Lenza, Iker Oyenarte, Tammo Diercks, Jon Imanol Quintana, Ana Gimeno, Helena Coelho, Ana Diniz, Francesca Peccati, Sandra Delgado, Alexandre Bosch, Mikel Valle, Oscar Millet, Nicola G.A. Abrescia, Asís Palazón, Filipa Marcelo, Gonzalo Jiménez-Osés, Jesús Jiménez-Barbero, Ana Ardá, and June Ereño-Orbea. Structural Characterization of N-Linked Glycans in the Receptor Binding Domain of the SARS-CoV-2 Spike Protein and their Interactions with Human Lectins. *Angewandte Chemie - International Edition*, 59(52):23763–23771, 12 2020. 101

[340] Asif Shajahan, Nitin T Supekar, Anne S Gleinich, and Parastoo Azadi. Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. *Glycobiology*, 30(12):981–988, 12 2020. 101

[341] Miloslav Sanda, Lindsay Morrison, and Radoslav Goldman. N and O glycosylation of the SARS-CoV-2 spike protein. *bioRxiv*, page 2020.07.05.187344, 7 2020. 101, 103

[342] Hangping Yao, Yutong Song, Yong Chen, Nanping Wu, Jialu Xu, Chujie Sun, Jiaxing Zhang, Tianhao Weng, Zheyuan Zhang, Zhigang Wu, Linfang Cheng, Danrong Shi, Xiangyun Lu, Jianlin Lei, Max Crispin, Yigong Shi, Lanjuan Li, and Sai Li. Molecular Architecture of the SARS-CoV-2 Virus. *Cell*, 183(3):730–738, 10 2020. 101

[343] Paul Gale. Thermodynamic equilibrium dose-response models for MERS-CoV infection reveal a potential protective role of human lung mucus but not for SARS-CoV-2. *Microbial Risk Analysis*, 16:100140, 12 2020. 103

[344] Byeong Gwan Cho, Sakshi Gautam, Wenjing Peng, Yifan Huang, Mona Goli, and Yehia Mechref. Direct Comparison of N-Glycans and Their Isomers Derived from Spike Glycoprotein 1 of MERS-CoV, SARS-CoV-1, and SARS-CoV-2. *Journal of Proteome Research*, 20(9):4357–4365, 9 2021. 103

[345] Cheorl-Ho Kim. SARS-CoV-2 Evolutionary Adaptation toward Host Entry and Recognition of Receptor O-Acetyl Sialylation in Virus–Host Interaction. *International Journal of Molecular Sciences*, 21(12):4549, 6 2020. 103

[346] B. Robson. Bioinformatics studies on a function of the SARS-CoV-2 spike glycoprotein as the binding of host sialic acid glycans. *Computers in Biology and Medicine*, 122:103849, 7 2020.

[347] Chirag Dhar, Aniruddha Sasmal, Sandra Diaz, Andrea Verhagen, Hai Yu, Wanqing Li, Xi Chen, and Ajit Varki. Are sialic acids involved in COVID-19 pathogenesis? *Glycobiology*, 31(9):1068–1071, 9 2021. 103

[348] Sandrine Belouzard, Jean K. Millet, Beth N. Licitra, and Gary R. Whittaker. Mechanisms of Coronavirus Cell Entry Mediated by the Viral Spike Protein. *Viruses*, 4(6):1011–1033, 6 2012. 103

[349] Liping Zhang, Matthew Mann, Zulfeqhar Syed, Hayley M Reynolds, E Tian, Nadine L Samara, Darryl C Zeldin, Lawrence A Tabak, and Kelly G Ten Hagen. Furin cleavage of the SARS-CoV-2 spike is modulated by O-glycosylation. *bioRxiv*, page 2021.02.05.429982, 2 2021. 103

[350] Gwenaëlle Douaud, Soojin Lee, Fidel Alfaro-Almagro, Christoph Arthofer, Chaoyue Wang, Frederik Lange, Jesper L R Andersson, Ludovica Griffanti, Eugene Duff, Saad Jbabdi, Bernd Taschler, Anderson Winkler, Thomas E Nichols, Rory Collins, Paul M Matthews, Naomi Allen, Karla L Miller, and Stephen M Smith. Brain imaging before and after COVID-19 in UK Biobank. *medRxiv*, 2021. 103

[351] Toni M. Dando and Caroline M. Perry. Enfuvirtide, 9 2003. 104

[352] David Jarrom, Lauren Elston, Jennifer Washington, Matthew Prettyjohns, Kimberley Cann, Susan Myles, and Peter Groves. Effectiveness of tests to detect the presence of SARS- CoV-2 virus, and antibodies to SARS- CoV-2, to inform COVID-19 diagnosis: a rapid systematic review. *BMJ Evidence*, 2020. 104

[353] Jade E. Jones, Susheel Bhanu Busi, Jonathan B. Mitchem, James M. Amos-Landgraf, and Michael R. Lewis. Evaluation of a Tumor-Targeting, Near-Infrared Fluorescent Peptide for Early Detection and Endoscopic Resection of Polyps in a Rat Model of Colorectal Cancer. *Molecular Imaging*, 17, 7 2018. 104

[354] Ahmet M Ulusan, Praveen Rajendran, Wan Mohaiza Dashwood, Omer F Yavuz, Sabeeta Kapoor, Trace A Gustafson, Michelle I Savage, Powel H Brown, Shizuko Sei, Altaf Mohammed, Eduardo Vilar, and Roderick H Dashwood. Optimization of Erlotinib Plus Sulindac Dosing Regimens for Intestinal Cancer Prevention in an Apc-Mutant Model of Familial Adenomatous Polyposis (FAP). *Cancer prevention research (Philadelphia, Pa.)*, 14(3):325–336, 3 2021. 104

[355] Sumit Bhatnagar, Eshita Khera, Jianshan Liao, Victoria Eniola, Yongjun Hu, David E. Smith, and Greg M. Thurber. Oral and Subcutaneous Administration of a Near-Infrared Fluorescent Molecular Imaging Agent Detects Inflammation in a Mouse Model of Rheumatoid Arthritis. *Scientific Reports*, 9(1):1–11, 12 2019. 104

[356] Anh Thi Viet Nguyen, Tien Thi Thuy Trinh, Vui Thi Hoang, Tung Duy Dao, Hien Thi Tuong, Hak Sung Kim, Hyun Park, and Seon Ju Yeo. Peptide aptamer of complementarity-determining region to detect avian influenza virus. *Journal of Biomedical Nanotechnology*, 16(6):1185–1200, 2019. 104

[357] Duong Tuan Bao, Do Thi Hoang Kim, Hyun Park, Bui Thi Cuc, Nguyen Minh Ngoc, Nguyen Thi Phuong Linh, Nguyen Chien Huu, Trinh Thi Thuy Tien, Nguyen Thi Viet Anh, Tung Dao Duy, Chom Kyu Chong, Seung Taek Yu, Do Young Choi, and Seon Ju

Yeo. Rapid detection of avian influenza virus by fluorescent diagnostic assay using an epitope-derived peptide. *Theranostics*, 7(7):1835–1846, 2017. 104

[358] Ching-Hsuan Tung. Fluorescent peptide probes for in vivo diagnostic imaging. *Biopolymers*, 76(5):391–403, 1 2004. 104

[359] Cameron M. Lee, Christoph J. Engelbrecht, Timothy D. Soper, Fritjof Helmchen, and Eric J. Seibel. Scanning fiber endoscopy with highly flexible, 1 mm catheterscopes for wide-field, full-color imaging, 2010. 104

[360] Shuai Xia, Meiqin Liu, Chao Wang, Wei Xu, Qiaoshuai Lan, Siliang Feng, Feifei Qi, Linlin Bao, Lanying Du, Shuwen Liu, Chuan Qin, Fei Sun, Zhengli Shi, Yun Zhu, Shibo Jiang, and Lu Lu. Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. *Cell Research*, 30(4):343–355, 4 2020. 106, 107

[361] Rory D De Vries, Katharina S Schmitz 1@, Francesca T Bovier, Danny Noack, Bart L Haagmans, Sudipta Biswas, Barry Rockx, Samuel H Gellman, Christopher A Alabi, Rik L De Swart, Anne Moscona, and Matteo Porotto. Intranasal fusion inhibitory lipopeptide prevents direct contact SARS-CoV-2 transmission in ferrets One-sentence summary: A dimeric form of a SARS-CoV-2-derived lipopeptide is a potent inhibitor of fusion and infection in vitro and transmission in vivo. *bioRxiv*, page 2020.11.04.361154, 11 2020. 106, 107

[362] Yuanmei Zhu, Danwei Yu, Hongxia Yan, Huihui Chong, and Yuxian He. Design of Potent Membrane Fusion Inhibitors against SARS-CoV-2, an Emerging Coronavirus with High Fusogenic Activity. *Journal of Virology*, 94(14), 7 2020. 106, 107

[363] Victor K. Outlaw, Francesca T. Bovier, Megan C. Mears, Maria N. Cajimat, Yun Zhu, Michelle J. Lin, Amin Addetia, Nicole A.P. Lieberman, Vikas Peddu, Xuping Xie, Pei Yong Shi, Alexander L. Greninger, Samuel H. Gellman, Dennis A. Bente, Anne Moscona, and Matteo Porotto. Inhibition of coronavirus entry in vitro and ex vivo by a lipid-conjugated peptide derived from the sars-cov-2 spike glycoprotein hrc domain. *mBio*, 11(5):1–14, 9 2020. 106

[364] Francesca Curreli, Sofia M B Victor, Shahad Ahmed, Aleksandra Drelich, Xiaohe Tong, Chien-Te K Tseng, Christopher D Hillyer, and Asim K Debnath. Stapled peptides based on Human Angiotensin-Converting Enzyme 2 (ACE2) potently inhibit SARS-CoV-2 infection in vitro. *bioRxiv*, page 2020.08.25.266437, 11 2020. 106, 107, 122

[365] Philippe Karoyan, Vincent Vieillard, Estelle Odile, Alexis Denis, Amélie Guihot, Charles-Edouard Luyt, Luis Gómez-Morales, Pascal Grondin, and Olivier Lequin. Human ACE2 peptide mimics block SARS-CoV-2 Pulmonary Cells Infection. *bioRxiv*, page 2020.08.24.264077, 10 2020. 106, 122

[366] Longxing Cao, Inna Goreshnik, Brian Coventry, James Brett Case, Lauren Miller, Lisa Kozodoy, Rita E. Chen, Lauren Carter, Alexandra C. Walls, Young Jun Park, Eva Maria Strauch, Lance Stewart, Michael S. Diamond, David Veesler, and David Baker. De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. *Science*, 370(6515):426–431, 10 2020. 106

[367] Alexander Norman, Charlotte Franck, Mary Christie, § Paige, M E Hawkins, Pa-Tel Karishma, Anneliese S Ashhurst, Anupriya Aggarwal, Jason K K Low, Rezwan Siddiquee, Caro-Line L Ashley, Megan Steain, James A Triccas, Stuart Turville, Joel P Mackay, Toby Pas-Sioura, and Richard J Payne. Discovery of Cyclic Peptide Ligands to the SARS-CoV-2 Spike Pro-tein using mRNA Display. *bioRxiv*, page 2020.12.22.424069, 12 2020. 106

[368] Ross C. Larue, Enming Xing, Adam D. Kenney, Yuexiu Zhang, Jasmine A. Tuazon, Jianrong Li, Jacob S. Yount, Pui Kai Li, and Amit Sharma. Rationally Designed ACE2-Derived Peptides Inhibit SARS-CoV-2. *Bioconjugate Chemistry*, 32(1):215–223, 1 2021. 106

[369] Brandon J. Beddingfield, Naoki Iwanaga, Prem P. Chapagain, Wenshu Zheng, Chad J. Roy, Tony Y. Hu, Jay K. Kolls, and Gregory J. Bix. The Integrin Binding Peptide, ATN-161, as a Novel Therapy for SARS-CoV-2 Infection. *JACC: Basic to Translational Science*, 6(1):1–8, 1 2021. 106, 107

[370] Shuai Xia, Wei Xu, Qian Wang, Cong Wang, Chen Hua, Weihua Li, Lu Lu, and Shibo Jiang. Peptide-Based Membrane Fusion Inhibitors Targeting HCoV-229E Spike Protein HR1 and HR2 Domains. *International Journal of Molecular Sciences*, 19(2):487, 2 2018. 107

[371] Kehu Yuan, Ling Yi, Jian Chen, Xiuxia Qu, Tingting Qing, Xi Rao, Pengfei Jiang, Jianhe Hu, Zikai Xiong, Yuchun Nie, Xuanling Shi, Wei Wang, Chen Ling, Xiaolei Yin, Keqiang Fan, Luhua Lai, Mingxiao Ding, and Hongkui Deng. Suppression of SARS-CoV entry by peptides corresponding to heptad regions on spike glycoprotein. *Biochemical and Biophysical Research Communications*, 319(3):746–752, 7 2004. 107

[372] Berend Jan Bosch, Byron E.E. Martina, Ruurd Van Der Zee, Jean Lepault, Bert Jan Haijema, Cees Versluis, Albert J.R. Heck, Raoul De Groot, Albert D.M.E. Osterhaus, and Peter J.M. Rottier. Severe acute respiratory syndrome coronavirus (SARS-CoV)

infection inhibition using spike protein heptad repeat-derived peptides. *Proceedings of the National Academy of Sciences of the United States of America*, 101(22):8455–8460, 6 2004.

[373] Shuwen Liu, Gengfu Xiao, Yibang Chen, Yuxian He, Jinkui Niu, Carlos R. Escalante, Huabao Xiong, James Farmar, Asim K. Debnath, Po Tien, and Shibo Jiang. Interaction between heptad repeat 1 and 2 regions in spike protein of SARS-associated coronavirus: Implications for virus fusogenic mechanism and identification of fusion inhibitors. *Lancet*, 363(9413):938–947, 3 2004. 107

[374] Lu Lu, Qi Liu, Yun Zhu, Kwok Hung Chan, Lili Qin, Yuan Li, Qian Wang, Jasper Fuk Woo Chan, Lanying Du, Fei Yu, Cuiqing Ma, Sheng Ye, Kwok Yung Yuen, Rongguang Zhang, and Shibo Jiang. Structure-based discovery of Middle East respiratory syndrome coronavirus fusion inhibitor. *Nature Communications*, 5(1):1–12, 1 2014. 107

[375] Jing Gao, Guangwen Lu, Jianxun Qi, Yan Li, Ying Wu, Yao Deng, Heyuan Geng, Hongbin Li, Qihui Wang, Haixia Xiao, Wenjie Tan, Jinghua Yan, and George F Gao. Structure of the Fusion Core and Inhibition of Fusion by a Heptad Repeat Peptide Derived from the S Protein of Middle East Respiratory Syndrome Coronavirus. *Am Soc Microbiol*, 87(24):13134–13140, 12 2013. 107

[376] Shuai Xia, Lei Yan, Wei Xu, Anurodh Shankar Agrawal, Abdullah Algaissi, Chien Te K. Tseng, Qian Wang, Lanying Du, Wenjie Tan, Ian A. Wilson, Shibo Jiang, Bei Yang, and Lu Lu. A pan-coronavirus fusion inhibitor targeting the HR1 domain of human coronavirus spike. *Science Advances*, 5(4):4580, 2019. 107

[377] Bruno Sainz, Eric C. Mossel, William R. Gallaher, William C. Wimley, C. J. Peters, Russell B. Wilson, and Robert F. Garry. Inhibition of severe acute respiratory syndrome-associated coronavirus (SARS-CoV) infectivity by peptides analogous to the viral spike protein. *Virus Research*, 120(1-2):146–155, 9 2006. 108

[378] Xiaoqiang Huang, Robin Pearce, and Yang Zhang. De novo design of protein peptides to block association of the SARS-CoV-2 spike protein with human ACE2. *Aging*, 12(12):11263–11276, 6 2020. 108

[379] Yanxiao Han and Petr Král. Computational Design of ACE2-Based Peptide Inhibitors of SARS-CoV-2. *ACS nano*, 14(4):5143–5147, 4 2020. 108

[380] Xiang Li, W. David Tolbert, Hong-Gang Gang Hu, Neelakshi Gohain, Yan Zou, Fan Niu, Wang-Xiao Xiao He, Weirong Yuan, Jia-Can Can Su, Marzena Pazgier, and Wuyuan Lu. Dithiocarbamate-inspired side chain stapling chemistry for peptide drug design. *Chemical Science*, 10(5):1522–1530, 1 2019. 108, 194

[381] David E. Shaw, J. P. Grossman, Joseph A. Bank, Brannon Batson, J. Adam Butts, Jack C. Chao, Martin M. Deneroff, Ron O. Dror, Amos Even, Christopher H. Fenton, Anthony Forte, Joseph Gagliardo, Gennette Gill, Brian Greskamp, C. Richard Ho, Douglas J. Ierardi, Lev Iserovich, Jeffrey S. Kuskin, Richard H. Larson, Timothy Layman, Li Siang Lee, Adam K. Lerer, Chester Li, Daniel Killebrew, Kenneth M. Mackenzie, Shark Yeuk Hai Mok, Mark A. Moraes, Rolf Mueller, Lawrence J. Nociolo, Jon L. Peticolas, Terry Quan, Daniel Ramot, John K. Salmon, Daniele P. Scarpazza, U. Ben Schafer, Naseer Siddique, Christopher W. Snyder, Jochen Spengler, Ping Tak Peter Tang, Michael Theobald, Horia Toma, Brian Towles, Benjamin Vitale, Stanley C. Wang, and Cliff Young. Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer. In *International Conference for High Performance Computing, Networking, Storage and Analysis, SC*, volume 2015-January, pages 41–53. IEEE Computer Society, 1 2014. 109

[382] Lorenzo Casalino, Zied Gaieb, Jory A. Goldsmith, Christy K. Hjorth, Abigail C. Dommer, Aoife M. Harbison, Carl A. Fogarty, Emilia P. Barros, Bryn C. Taylor, Jason S. Mclellan, Elisa Fadda, and Rommie E. Amaro. Beyond shielding: The roles of glycans in the SARS-CoV-2 spike protein. *ACS Central Science*, 6(10):1722–1734, 10 2020. 109, 112, 131, 142, 149

[383] Tristan I. Croll, Kay Diederichs, Florens Fischer, Cameron D. Fyfe, Yunyun Gao, Sam Horrell, Agnel Praveen Joseph, Luise Kandler, Oliver Kippes, Ferdinand Kirsten, Konstantin Müller, Kristopher Nolte, Alexander M. Payne, Matthew Reeves, Jane S. Richardson, Gianluca Santoni, Sabrina Stäb, Dale E. Tronrud, Lea C. von Soosten, Christopher J. Williams, and Andrea Thorn. Making the invisible enemy visible, 5 2021. 109, 115, 149

[384] Oliver C Grant, David Montgomery, Keigo Ito, and Robert J Woods. Analysis of the SARS-CoV-2 spike protein glycan shield: implications for immune recognition. *bioRxiv*, page 2020.04.07.030445, 5 2020. 109, 112

[385] Reinis Danne, Chetan Poojari, Hector Martinez-Seara, Sami Rissanen, Fabio Lolicato, Tomasz Róg, and Ilpo Vattulainen. DoGlycans-Tools for Preparing Carbohydrate Structures for Atomistic Simulations of Glycoproteins, Glycolipids, and Carbohydrate Polymers for GROMACS. *Journal of Chemical Information and Modeling*, 57(10):2401–2406, 10 2017. 109

[386] Mateusz Sikora, Sören von Bülow, Florian E. C. Blanc, Michael Gecht, Roberto Covino, and Gerhard Hummer. Computational epitope map of SARS-CoV-2 spike protein. *PLOS Computational Biology*, 17(4):e1008790, 4 2021. 111

[387] Jiandong Huo, Yuguang Zhao, Jingshan Ren, Daming Zhou, Helen M.E. Duyvesteyn, Helen M. Ginn, Loic Carrique, Tomas Malinauskas, Reinis R. Ruza, Pranav N.M. Shah, Tiong Kit Tan, Pramila Rijal, Naomi Coombes, Kevin R. Bewley, Julia A. Tree, Julika Radecke, Neil G. Paterson, Piyada Supasa, Juthathip Mongkolsapaya, Gavin R. Screaton, Miles Carroll, Alain Townsend, Elizabeth E. Fry, Raymond J. Owens, and David I. Stuart. Neutralization of SARS-CoV-2 by Destruction of the Prefusion Spike. *Cell Host and Microbe*, 28(3):445–454, 9 2020. 116, 125

[388] Jiangdong Huo, Audrey Le Bas, Reinis R. Ruza, Helen M.E. Duyvesteyn, Halina Mikolajek, Tomas Malinauskas, Tiong Kit Tan, Pramila Rijal, Maud Dumoux, Philip N. Ward, Jingshan Ren, Daming Zhou, Peter J. Harrison, Miriam Weckener, Daniel K. Clare, Vinod K. Vogirala, Julika Radecke, Lucile Moynié, Yuguang Zhao, Javier Gilbert-Jaramillo, Michael L. Knight, Julia A. Tree, Karen R. Buttigieg, Naomi Coombes, Michael J. Elmore, Miles W. Carroll, Loic Carrique, Pranav N.M. Shah, William James, Alain R. Townsend, David I. Stuart, Raymond J. Owens, and James H. Naismith. Neutralizing nanobodies bind SARS-CoV-2 spike RBD and block interaction with ACE2. *Nature Structural and Molecular Biology*, 27(9):846–854, 9 2020. 116, 125

[389] Johanna Hansen, Alina Baum, Kristen E. Pascal, Vincenzo Russo, Stephanie Giordano, Elzbieta Wloga, Benjamin O. Fulton, Ying Yan, Katrina Koon, Krunal Patel, Kyung Min Chung, Aynur Hermann, Erica Ullman, Jonathan Cruz, Ashique Rafique, Tammy Huang, Jeanette Fairhurst, Christen Libertiny, Marine Malbec, Wen Yi Lee, Richard Welsh, Glen Farr, Seth Pennington, Dipali Deshpande, Jemmie Cheng, Anke Watty, Pascal Bouffard, Robert Babb, Natasha Levenkova, Calvin Chen, Bojie Zhang, Annabel Romero Hernandez, Kei Saotome, Yi Zhou, Matthew Franklin, Sumathi Sivapalasingam, David Chien Lye, Stuart Weston, James Logue, Robert Haupt, Matthew Frieman, Gang Chen, William Olson, Andrew J. Murphy, Neil Stahl, George D. Yancopoulos, and Christos A. Kyratsous. Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail. *Science*, 369(6506):1010–1014, 8 2020. 116, 125

[390] Yunlong Cao, Bin Su, Xianghua Guo, Wenjie Sun, Yongqiang Deng, Linlin Bao, Qinyu Zhu, Xu Zhang, Yinghui Zheng, Chenyang Geng, Xiaoran Chai, Runsheng He, Xiaofeng Li, Qi Lv, Hua Zhu, Wei Deng, Yanfeng Xu, Yanjun Wang, Luxin Qiao, Yafang Tan, Liyang Song, Guopeng Wang, Xiaoxia Du, Ning Gao, Jiangning Liu, Junyu Xiao, Xiao dong Su, Zongmin Du, Yingmei Feng, Chuan Qin, Chengfeng Qin, Ronghua Jin, and X. Sunney Xie. Potent Neutralizing Antibodies against SARS-CoV-2 Identified by High-Throughput Single-Cell Sequencing of Convalescent Patients' B Cells. *Cell*, 182(1):73–84, 7 2020. 116, 125

[391] Nicholas C. Wu, Meng Yuan, Sandhya Bangaru, Deli Huang, Xueyong Zhu, Chang-Chun D. Lee, Hannah L. Turner, Linghang Peng, Linlin Yang, Dennis R. Burton, David Nemazee, Andrew B. Ward, and Ian A. Wilson. A natural mutation between SARS-CoV-2 and SARS-CoV determines neutralization by a cross-reactive antibody. *PLOS Pathogens*, 16(12):e1009089, 12 2020. 116, 125

[392] Tingting Li, Hongmin Cai, Hebang Yao, Bingjie Zhou, Ning Zhang, Martje Fentener van Vlissingen, Thijs Kuiken, Wenyu Han, Corine H. GeurtsvanKessel, Yuhuan Gong, Yapei Zhao, Quan Shen, Wenming Qin, Xiao Xu Tian, Chao Peng, Yanling Lai, Yanxing Wang, Cedric A.J. Hutter, Shu Ming Kuo, Juan Bao, Caixuan Liu, Yifan Wang, Audrey S. Richard, Hervé Raoul, Jiaming Lan, Markus A. Seeger, Yao Cong, Barry Rockx, Gary Wong, Yuhai Bi, Dimitri Lavillette, and Dianfan Li. A synthetic nanobody targeting RBD protects hamsters from SARS-CoV-2 infection. *Nature Communications*, 12(1):1–13, 12 2021. 116, 125

[393] Dora Pinto, Young Jun Park, Martina Beltramello, Alexandra C. Walls, M. Alejandra Tortorici, Siro Bianchi, Stefano Jaconi, Katja Culap, Fabrizia Zatta, Anna De Marco, Alessia Peter, Barbara Guarino, Roberto Spreafico, Elisabetta Cameroni, James Brett Case, Rita E. Chen, Colin Havenar-Daughton, Gyorgy Snell, Amalio Telenti, Herbert W. Virgin, Antonio Lanzavecchia, Michael S. Diamond, Katja Fink, David Veesler, and Davide Corti. Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature*, 583(7815):290–295, 7 2020. 116, 125

[394] Rui Shi, Chao Shan, Xiaomin Duan, Zhihai Chen, Peipei Liu, Jinwen Song, Tao Song, Xiaoshan Bi, Chao Han, Lianao Wu, Ge Gao, Xue Hu, Yanan Zhang, Zhou Tong, Weijin Huang, William Jun Liu, Guizhen Wu, Bo Zhang, Lan Wang, Jianxun Qi, Hui Feng, Fu Sheng Wang, Qihui Wang, George Fu Gao, Zhiming Yuan, and Jinghua Yan. A human neutralizing antibody targets the receptor-binding site of SARS-CoV-2. *Nature*, 584(7819):120–124, 8 2020. 116, 125

[395] Daniel Wrapp, Dorien De Vlieger, Kizzmekia S. Corbett, Gretel M. Torres, Nianshuang Wang, Wander Van Breedam, Kenny Roose, Loes van Schie, Markus Hoffmann, Stefan Pöhlmann, Barney S. Graham, Nico Callewaert, Bert Schepens, Xavier Saelens, and Jason S. McLellan. Structural Basis for Potent Neutralization of Betacoronaviruses by Single-Domain Camelid Antibodies. *Cell*, 181(5):1004–1015, 5 2020. 116, 125

[396] Uroš Zavrtanik, Junoš Lukan, Remy Loris, Jurij Lah, and San Hadži. Structural Basis of Epitope Recognition by Heavy-Chain Camelid Antibodies. *Journal of Molecular Biology*, 430(21):4369–4386, 10 2018. 124

[397] Hung-Pin Peng, Kuo Hao Lee, Jhih-Wei Jian, and An-Suei Yang. Origins of specificity and affinity in antibody–protein interactions. *Proceedings of the National Academy of Sciences*, 111(26):E2656–E2665, 7 2014. 124

[398] Meng Yuan, Nicholas C. Wu, Xueyong Zhu, Chang Chun D. Lee, Ray T.Y. So, Huibin Lv, Chris K.P. Mok, and Ian A. Wilson. A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. *Science*, 368(6491):630–633, 5 2020. 125

[399] Bin Ju, Qi Zhang, Jiwan Ge, Ruoke Wang, Jing Sun, Xiangyang Ge, Jiazhen Yu, Sisi Shan, Bing Zhou, Shuo Song, Xian Tang, Jinfang Yu, Jun Lan, Jing Yuan, Haiyan Wang, Juanjuan Zhao, Shuye Zhang, Youchun Wang, Xuanling Shi, Lei Liu, Jincun Zhao, Xinquan Wang, Zheng Zhang, and Linqi Zhang. Human neutralizing antibodies elicited by SARS-CoV-2 infection. *Nature*, 584(7819):115–119, 8 2020. 125

[400] Zhixin Liu, Xiao Xiao, Xiuli Wei, Jian Li, Jing Yang, Huabing Tan, Jianyong Zhu, Qiwei Zhang, Jianguo Wu, and Long Liu. Composition and divergence of coronavirus spike proteins and host ACE2 receptors predict potential intermediate hosts of SARS-CoV-2. *Journal of Medical Virology*, 92(6):595–601, 6 2020. 125

[401] Georgii G. Krivov, Maxim V. Shapovalov, and Roland L. Dunbrack. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins: Structure, Function and Bioinformatics*, 77(4):778–795, 2009. 127

[402] Carl T. Wild, Diane C. Shugars, Teresa K. Greenwell, Charlene B. McDanal, and Thomas J. Matthews. Peptides corresponding to a predictive $\alpha$-helical domain of human immunodeficiency virus type 1 gp41 are potent inhibitors of virus infection. *Proceedings of the National Academy of Sciences of the United States of America*, 91(21):9770–9774, 10 1994. 130

[403] Shibo Jiang, Kang Lin, Nathan Strick, and A. Robert Neurath. HIV-1 inhibition by a peptide [3], 1993. 130

[404] Tom Matthews, Miklos Salgo, Michael Greenberg, Jain Chung, Ralph DeMasi, and Dani Bolognesi. Enfuvirtide: The first therapy to inhibit the entry of HIV-1 into host CD4 lymphocytes, 2004. 130

[405] Jack W Heal, Gail J Bartlett, Christopher W Wood, Andrew R Thomson, and Derek N Woolfson. Applying graph theory to protein structures: an Atlas of coiled coils. *Bioinformatics*, 34(19):3316–3323, 10 2018. 131

[406] Guto G. Rhys, Christopher W. Wood, Eric J.M. Lang, Adrian J. Mulholland, R. Leo Brady, Andrew R. Thomson, and Derek N. Woolfson. Maintaining and breaking sym-

metry in homomeric coiled-coil assemblies. *Nature Communications*, 9(1):1–12, 12 2018. 132

[407] Joseph Thomas Ortega, Maria Luisa Serrano, Flor Helene Pujol, and Hector Rafael Rangel. Role of changes in SARS-CoV-2 spike protein in the interaction with the human ACE2 receptor: An in silico analysis. *EXCLI Journal*, 19:410–417, 2020. 136

[408] Joanna Koszela, Nhan T. Pham, David Evans, Stefan Mann, Irene Perez-Pi, Steven Shave, Derek F.J. Ceccarelli, Frank Sicheri, Mike Tyers, and Manfred Auer. Real-time tracking of complex ubiquitination cascades using a fluorescent confocal on-bead assay. *BMC Biology*, 16(1):88, 8 2018. 136

[409] Irene Pérez-Pi, David A. Evans, Mathew H. Horrocks, Nhan T. Pham, Karamjit S. Dolt, Joanna Koszela, Tilo Kunath, and Manfred Auer. $\alpha$-Synuclein-Confocal Nanoscanning (ASYN-CONA), a Bead-Based Assay for Detecting Early-Stage $\alpha$-Synuclein Aggregation. *Analytical Chemistry*, 91(9):5582–5590, 5 2019. 136

[410] Jones Andrade, Paulo Fernando Bruno Gonçalves, and Paulo Augusto Netz. Why Does the Novel Coronavirus Spike Protein Interact so Strongly with the Human ACE2? A Thermodynamic Answer. *ChemBioChem*, 22(5):865–875, 3 2021. 142

[411] Kui K. Chan, Danielle Dorosky, Preeti Sharma, Shawn A. Abbasi, John M. Dye, David M. Kranz, Andrew S. Herbert, and Erik Procko. Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. *Science*, 369(6508):1261–1265, 9 2020. 142

[412] Kushal Suryamohan, Devan Diwanji, Eric W. Stawiski, Ravi Gupta, Shane Miersch, Jiang Liu, Chao Chen, Ying-Ping Jiang, Frederic A. Fellouse, J. Fah Sathirapongsasuti, Patrick K. Albers, Tanneeru Deepak, Reza Saberianfar, Aakrosh Ratan, Gavin Washburn, Monika Mis, Devi Santhosh, Sneha Somasekar, G. H. Hiranjith, Derek Vargas, Sangeetha Mohan, Sameer Phalke, Boney Kuriakose, Aju Antony, Mart Ustav Jr, Stephan C. Schuster, Sachdev Sidhu, Jagath R. Junutula, Natalia Jura, and Somasekar Seshagiri. Human ACE2 receptor polymorphisms and altered susceptibility to SARS-CoV-2. *Communications Biology*, 4(1):475, 12 2021. 142

[413] Ahmad Reza Mehdipour and Gerhard Hummer. Dual nature of human ACE2 glycosylation in binding to SARS-CoV-2 spike. *BIOPHYSICS AND COMPUTATIONAL BIOLOGY Downloaded at UCB PHARMA SA on*, 118(19), 2021. 142

[414] Shadi Rahnama, Maryam Azimzadeh Irani, Mehriar Amininasab, and Mohammad Reza Ejtehadi. S494 O-glycosylation site on the SARS-CoV-2 RBD affects the virus affinity to ACE2 and its infectivity; a molecular dynamics study. *Scientific Reports*, 11(1), 12 2021. 142

[415] Amaro Lab. COVID-19 Data Sets. 149

[416] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25, 9 2015. 150

[417] James A. Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E. Hauser, and Carlos Simmerling. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11(8):3696–3713, 8 2015. 153

[418] Karl N. Kirschner, Austin B. Yongye, Sarah M. Tschampel, Jorge González-Outeiriño, Charlisa R. Daniels, B. Lachele Foley, and Robert J. Woods. GLYCAM06: A generalizable biomolecular force field. carbohydrates. *Journal of Computational Chemistry*, 29(4):622–655, 3 2008. 153

[419] Richard J Gowers, Max Linke, Jonathan Barnoud, Tyler J E Reddy, Manuel N Melo, Sean L Seyler, Jan Domá nski, David L Dotson, Sébastien Buchoux, Ian M Kenney, and Oliver Beckstein. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. Technical report, 2016. 154

[420] Robert T. McGibbon, Kyle A. Beauchamp, Matthew P. Harrigan, Christoph Klein, Jason M. Swails, Carlos X. Hernández, Christian R. Schwantes, Lee Ping Wang, Thomas J. Lane, and Vijay S. Pande. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal*, 109(8):1528–1532, 10 2015. 154, 193, 194

[421] Fabian Pedregosa FABIANPEDREGOSA, Vincent Michel, Olivier Grisel OLIVIERGRISEL, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Jake Vanderplas, David Cournapeau, Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Bertrand Thirion, Olivier Grisel, Vincent Dubourg, Alexandre Passos, Matthieu Brucher, Matthieu Perrot andÉdouardand, andÉdouard Duchesnay, and FRÉdouard Duchesnay EDOUARDDUCHESNAY. Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot. Technical report, 2011. 154

[422] Luke A.J. O'Neill and Charles A. Dinarello. The IL-1 receptor/toll-like receptor superfamily: Crucial receptors for inflammation and host defense, 5 2000. 155, 156

[423] Feghali C. A. and T. M Wright. Cytokines in acute and chronic inflammation. *Frontiers in Bioscience*, 2(4):A171, 1997. 155

[424] Charles A. Dinarello and Robert C. Thompson. Blocking IL-1: interleukin 1 receptor antagonist in vivo and in vitro, 11 1991. 155

[425] Feifan Yu, Lindvi Gudmundsdotter, Anastassja Akal, Elin Gunneriusson, Fredrik Frejd, and Per Ake Nygren. An affibody-adalimumab hybrid blocks combined IL-6 and tnf-triggered serum amyloid a secretion in vivo. *mAbs*, 6(6):1598–1607, 2014. 155

[426] Wibke Schulte, Jürgen Bernhagen, and Richard Bucala. Cytokines in sepsis: Potent immunoregulators and potential therapeutic targets - An updated view. *Mediators of Inflammation*, 2013:16, 2013. 155

[427] David Grimaldi, Eduardo Wilfrido Goicoechea Turcott, and Fabio Silvio Taccone. IL-1 receptor antagonist in sepsis: New findings with old data?, 2016. 155

[428] Daniella M. Schwartz, Michael Bonelli, Massimo Gadina, and John J. O'Shea. Type I/II cytokines, JAKs, and new strategies for treating autoimmune diseases. *Nature Reviews Rheumatology*, 12(1):25–36, 1 2016. 156

[429] Luke A.J. O'Neill, Katherine A. Fitzgerald, and Andrew G. Bowie. The Toll-IL-1 receptor adaptor family grows to five members, 6 2003. 156

[430] Diana Boraschi and Aldo Tagliabue. The interleukin-1 receptor family, 12 2013. 157

[431] W P Arend, M Malyak, M F Smith, T D Whisenand, J L Slack, J E Sims, J G Giri, and S K Dower. Binding of IL-1 alpha, IL-1 beta, and IL-1 receptor antagonist by soluble IL-1 receptors and levels of soluble IL-1 receptors in synovial fluids. *The Journal of Immunology*, 153(10), 1994. 158

[432] Kenji Funami, Miwa Sasai, Hiroyuki Oshiumi, Tsukasa Seya, and Misako Matsumoto. Homo-oligomerization is essential for toll/interleukin-1 receptor domain-containing adaptor molecule-1-mediated NF-κB and interferon regulatory factor-3 activation. *Journal of Biological Chemistry*, 283(26):18283–18291, 6 2008. 159

[433] Ota Fekonja, Mojca Benčina, and Roman Jerala. Toll/interleukin-1 receptor domain dimers as the platform for activation and enhanced inhibition of toll-like receptor signaling. *Journal of Biological Chemistry*, 287(37):30993–31002, 9 2012. 159

[434] Javed A. Khan, Elizabeth K. Brint, Luke A.J. O'Neill, and Liang Tong. Crystal structure of the Toll/interleukin-1 receptor domain of human IL-1RAPL. *Journal of Biological Chemistry*, 279(30):31664–31670, 7 2004. 160

[435] PepCalc.com - Peptide calculator. 163

[436] **REDACTED**

[437] Chris Phillips, Lee R. Roberts, Markus Schade, Richard Bazin, Andrew Bent, Nichola L. Davies, Rob Moore, Andrew D. Pannifer, Andrew R. Pickford, Stephen H. Prior, Christopher M. Read, Andrew Scott, David G. Brown, Bin Xu, and Stephen L. Irving. Design and structure of stapled peptides binding to estrogen receptors. *Journal of the American Chemical Society*, 133(25):9696–9699, 6 2011. 177, 194

[438] Ya Qiu Long, Shao Xu Huang, Zahrah Zawahir, Zhong Liang Xu, Huiyuan Li, Tino W. Sanchez, Ying Zhi, Stephanie De Houwer, Frauke Christ, Zeger Debyser, and Nouri Neamati. Design of cell-permeable stapled peptides as HIV-1 integrase inhibitors. *Journal of Medicinal Chemistry*, 56(13):5601–5612, 7 2013. 177

[439] Benedict W.J. Irwin and David J. Huggins. Estimating Atomic Contributions to Hydration and Binding Using Free Energy Perturbation. *Journal of Chemical Theory and Computation*, 14(6):3218–3227, 6 2018. 178

[440] John G. Kirkwood. Statistical mechanics of fluid mixtures. *The Journal of Chemical Physics*, 3(5):300–313, 5 1935. 178

[441] Ursula Kahler, Julian E. Fuchs, Peter Goettig, and Klaus R. Liedl. An unexpected switch in peptide binding mode: from simulation to substrate specificity. *Journal of Biomolecular Structure and Dynamics*, 36(15):4072–4084, 11 2018. 179

[442] Shunzhou Wan, Bernhard Knapp, David W. Wright, Charlotte M. Deane, and Peter V. Coveney. Rapid, Precise, and Reproducible Prediction of Peptide-MHC Binding Affinities from Molecular Dynamics That Correlate Well with Experiment. *Journal of Chemical Theory and Computation*, 11(7):3346–3356, 7 2015.

[443] Veronica Salmaso, Mattia Sturlese, Alberto Cuzzolin, and Stefano Moro. Exploring Protein-Peptide Recognition Pathways Using a Supervised Molecular Dynamics Approach. *Structure*, 25(4):655–662, 4 2017. 179

[444] Huiyong Sun, Youyong Li, Mingyun Shen, Sheng Tian, Lei Xu, Peichen Pan, Yan Guan, and Tingjun Hou. Assessing the performance of MM/PBSA and MM/GBSA methods. 5. Improved docking performance using high solute dielectric constant MM/GBSA and MM/PBSA rescoring. *Physical Chemistry Chemical Physics*, 16(40):22035–22045, 9 2014. 180

[445] Yuchen Li, Yalong Cong, Guoqiang Feng, Susu Zhong, John Z.H. Zhang, Huiyong Sun, and Lili Duan. The impact of interior dielectric constant and entropic change on HIV-1 complex binding free energy prediction. *Structural Dynamics*, 5(6), 11 2018.

[446] Ercheng Wang, Gaoqi Weng, Huiyong Sun, Hongyan Du, Feng Zhu, Fu Chen, Zhe Wang, and Tingjun Hou. Assessing the performance of the MM/PBSA and MM/GBSA methods. 10. Impacts of enhanced sampling and variable dielectric model on protein-protein Interactions. *Physical Chemistry Chemical Physics*, 21(35):18958–18969, 9 2019. 180

[447] Ercheng Wang, Gaoqi Weng, Huiyong Sun, Hongyan Du, Feng Zhu, Fu Chen, Zhe Wang, and Tingjun Hou. Assessing the performance of the MM/PBSA and MM/GBSA methods. *Phys. Chem. Chem. Phys*, 21:18958, 2019. 181

[448] Srinivasaraghavan Kannan, Pietro G.A. Aronica, Yaw Sing Tan, and Chandra S Verma. Inhibiting S100B($\beta\beta$) for Activating Wild-Type p53: Design of Stapled Peptides. *ACS Omega*, 4(3):5335–5344, 2019. 194

[449] Srinivasaraghavan Kannan, Pietro G.A. Aronica, Simon Ng, Dawn Thean, Yuri Frosi, Sharon Chee, Jiang Shimin, Tsz Ying Yuen, Ahmad Sadruddin, Hung Yi Kristal Kaan, Arun Chandramohan, Jin Huei Wong, Yaw Sing Tan, Fernando J Ferrer, Prakash Arumugam, Yi Han, Shiying Chen, Christopher J Brown, Charles W Johannes, Brian Henry, David Lane, Tomi K Sawyer, Chandra Shekhar Verma, and Anthony Partridge. Macrocyclization of an all-D linear peptide improves target affinity and imparts cellular activity: A novel stapled $\alpha$-helical peptide modality. *bioRxiv*, page 767673, 9 2019. 194

[450] Sharon Min Qi Chee, Jantana Wongsantichon, Jiawei Siau, Dawn Thean, Fernando Ferrer, Robert C. Robinson, David P. Lane, Christopher J. Brown, and Farid J. Ghadessy. Structure-activity studies of Mdm2/Mdm4-binding stapled peptides comprising non-natural amino acids. *PLOS ONE*, 12(12):e0189379, 12 2017. 194

[451] Michelle L. Stewart, Emiko Fire, Amy E. Keating, and Loren D. Walensky. The MCL-1 BH3 helix is an exclusive MCL-1 inhibitor and apoptosis sensitizer. *Nature Chemical Biology*, 6(8):595–601, 6 2010. 194

[452] Jennifer A. Miles, David J. Yeo, Philip Rowell, Silvia Rodriguez-Marin, Christopher M. Pask, Stuart L. Warriner, Thomas A. Edwards, and Andrew J. Wilson. Hydrocarbon constrained peptides-understanding preorganisation and binding affinity. *Chemical Science*, 7(6):3694–3702, 5 2016. 194

[453] Edward P. Harvey, Hyuk Soo Seo, Rachel M. Guerra, Gregory H. Bird, Sirano Dhe-Paganon, and Loren D. Walensky. Crystal Structures of Anti-apoptotic BFL-1 and Its

Complex with a Covalent Stapled Peptide Inhibitor. *Structure*, 26(1):153–160, 1 2018. 194

[454] Thomas E. Speltz, Sean W. Fanning, Christopher G. Mayne, Colin Fowler, Emad Tajkhorshid, Geoffrey L. Greene, and Terry W. Moore. Stapled Peptides with $\gamma$-Methylated Hydrocarbon Chains for the Estrogen Receptor/Coactivator Interaction. *Angewandte Chemie International Edition*, 55(13):4252–4255, 3 2016. 194

[455] Jeffrey E Montgomery, Justin A Donnelly, Sean W Fanning, Thomas E Speltz, Xianghang Shangguan, John S Coukos, Geoffrey L Greene, and Raymond E Moellering. Versatile Peptide Macrocyclization with Diels−Alder Cycloadditions. *J. Am. Chem. Soc*, 141:59, 2019. 194

[456] Yana K. Rennie, Patrick J. McIntyre, Tito Akindele, Richard Bayliss, and Andrew G. Jamieson. A TPX2 Proteomimetic Has Enhanced Affinity for Aurora-A Due to Hydrocarbon Stapling of a Helix. *ACS Chemical Biology*, 11(12):3383–3390, 12 2016. 194

[457] Tom N Grossmann, Johannes T-H Yeh, Brian R Bowman, Qian Chu, Raymond E Moellering, and Gregory L Verdine. Inhibition of oncogenic Wnt signaling through direct targeting of $\beta$-catenin. *Proceedings of the National Academy of Sciences of the United States of America*, 109(44):17942–7, 10 2012. 195

[458] Daniel C. Scott, Julie K. Monda, Eric J. Bennett, J. Wade Harper, and Brenda A. Schulman. N-terminal acetylation acts as an avidity enhancer within an interconnected multiprotein complex. *Science*, 334(6056):674–678, 11 2011. 195

[459] Louise C. Briggs, A. W. Edith Chan, Christopher A. Davis, Nicholas Whitelock, Hajira A. Hotiana, Mehdi Baratchian, Claire Bagnéris, David L. Selwood, Mary K. Collins, and Tracey E. Barrett. IKK$\gamma$-Mimetic Peptides Block the Resistance to Apoptosis Associated with Kaposi's Sarcoma-Associated Herpesvirus Infection. *Journal of Virology*, 91(23), 12 2017. 195

[460] M Ultsch, A Braisted, H R Maun, and C Eigenbrot. 3-2-1: Structural insights from stepwise shrinkage of a three-helix Fc-binding domain to a single helix. *Protein Engineering, Design and Selection*, 30(9):619–625, 9 2017. 195

[461] Felix Findeisen, Marta Campiglio, Hyunil Jo, Fayal Abderemane-Ali, Christine H. Rumpf, Lianne Pope, Nathan D. Rossen, Bernhard E. Flucher, William F. DeGrado, and Daniel L. Minor. Stapled Voltage-Gated Calcium Channel (CaV) $\alpha$-Interaction Domain (AID) Peptides Act As Selective Protein-Protein Interaction Inhibitors of CaV Function. *ACS Chemical Neuroscience*, 8(6):1313–1326, 6 2017. 195

[462] Yaoying Wu and Joel H. Collier. $\alpha$-Helical coiled-coil peptide materials for biomedical applications. *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology*, 9(2):e1424, 3 2017. 195

[463] Gregory H. Bird, Sandhya Boyapalle, Terianne Wong, Kwadwo Opoku-Nsiah, Raminder Bedi, W. Christian Crannell, Alisa F. Perry, Huy Nguyen, Viviana Sampayo, Ankita Devareddy, Subhra Mohapatra, Shyam S. Mohapatra, and Loren D. Walensky. Mucosal delivery of a double-stapled RSV peptide prevents nasopulmonary infection. *Journal of Clinical Investigation*, 124(5):2113–2124, 2014. 195

[464] Shibani Bhattacharya, Hongtao Zhang, Asim K. Debnath, and David Cowburn. Solution structure of a hydrocarbon stapled peptide inhibitor in complex with monomeric C-terminal domain of HIV-1 capsid. *Journal of Biological Chemistry*, 283(24):16274–16278, 6 2008. 195

[465] Christopher H. Douse, Sabrina J. Maas, Jemima C. Thomas, James A. Garnett, Yunyun Sun, Ernesto Cota, and Edward W. Tate. Crystal Structures of Stapled and Hydrogen Bond Surrogate Peptides Targeting a Fully Buried Protein–Helix Interaction. *ACS Chemical Biology*, 9(10):2204–2209, 10 2014. 195

[466] Dilraj Lama, Anne Marie Liberatore, Yuri Frosi, Jessica Nakhle, Natia Tsomaia, Tarig Bashir, David P. Lane, Christopher J. Brown, Chandra S. Verma, and Serge Auvin. Structural insights reveal a recognition feature for tailoring hydrocarbon stapled-peptides against the eukaryotic translation initiation factor 4E protein. *Chemical Science*, 10(8):2489–2500, 2 2019. 195

[467] Jessica Iegre, Paul Brear, David J. Baker, Yaw Sing Tan, Eleanor L. Atkinson, Hannah F. Sore, Daniel H. O'Donovan, Chandra S. Verma, Marko Hyvönen, and David R. Spring. Efficient development of stable and highly functionalised peptides targeting the CK2$\alpha$/CK2$\beta$ protein-protein interaction. *Chemical Science*, 10(19):5056–5063, 5 2019. 195

[468] Ivan de Paola, Luciano Pirone, Maddalena Palmieri, Nicole Balasco, Luciana Esposito, Luigi Russo, Daniela Mazzà, Lucia Di Marcotullio, Sonia Di Gaetano, Gaetano Malgieri, Luigi Vitagliano, Emilia Pedone, and Laura Zaccaro. Cullin3 - BTB Interface: A Novel Target for Stapled Peptides. *PLOS ONE*, 10(4):e0121149, 4 2015. 195