



Politechnika Wroclawska

Field of Exact and Natural Sciences

Discipline of Chemical Sciences

PhD dissertation

**Design, synthesis, and structural
characterization of miniproteins, and
incorporation of protein-protein interaction
inhibition activity**

Juan Lizandra Pérez

Supervisor:

Prof. dr hab. Łukasz Berlicki

Keywords: peptide foldamers, β -amino acids, thermodynamics, inhibitors, PD-1/PD-L1.

WROCŁAW 2023

TABLE OF CONTENTS

Abstract	v
List of Abbreviations.....	ix
Acknowledgments.....	xi
1. Introduction	1
1.1. The protein fold problem.....	1
1.1.1. Information encoded in the amino acid sequence	1
1.1.2. Protein folding pathway	2
1.1.3. Structure prediction	4
1.2. Protein-protein interactions	8
1.2.1. PD-1/PD-L1 immune checkpoint.....	9
1.3. Miniproteins	12
1.3.1. Miniproteins in nature	13
1.3.2. Engineered miniproteins	17
1.3.3. <i>De novo</i> designed miniproteins	22
1.3.4. Backbone alteration.....	27
1.4. Biological activity modulation	32
1.4.1. Native and modulated activity.....	32
1.4.2. Molecular grafting.....	35
1.4.3. <i>De novo</i> activity design	38
2. Goals	41
3. Results and discussion.....	43
3.1. β -amino acid-containing miniproteins	43
3.1.1. From EEE to HEEE topology by fragment assembling	43
3.1.2. Circular permutation of HEEE to EHEE.....	57
3.1.3. Optimisation and thermodynamics of the EHEE fold.....	62
3.1.4. Reversing the EHEE fold to HEEE.....	81

3.1.5. Conclusions	85
3.2. PD-1/PD-L1 interaction inhibitors	87
3.2.1. WW-domain-based inhibitors	87
3.2.2. Optimized inhibitors grafted into HEEE scaffold	107
3.2.3. Inhibitor modification to the EHEE fold	113
3.2.4. Hybrid design over optimized EHEE	120
3.2.5. Conclusions	139
4. Summary	141
5. Experimental	143
5.1. Miniprotein design	143
5.1.1. Rosetta FastDesign Protocol	143
5.1.2. Structure predictors	143
5.1.3. Minimization and molecular dynamics	144
5.2. Microwave assisted solid phase peptide synthesis	146
5.3. Miniprotein purification and characterization	147
5.4. Circular Dichroism	154
5.4.1. Circular dichroism scan at 25 °C	154
5.4.2. Thermal denaturation measurements	154
5.4.3. Thermodynamic studies	156
5.5. NanoDSF	160
5.6. BioLayer Interferometry	160
5.7. Homogeneous time resolved fluorescence	161
5.8. Cell-based assay	162
5.9. Small angle X-ray scattering	163
5.10. NMR studies and calculations	164
5.10.1 Chemical shifts tables from the NMR experimental results	164
5.11. Crystallography	172
6. References	172

Abstract

The vast array of molecular functions carried out by naturally occurring proteins is made possible by their precisely folded structures, which are encoded in their amino acid sequences. Protein folds have undergone evolutionary changes to display diverse and distinct functional properties essential for biological processes. Moreover, interaction between proteins, protein-protein interactions (PPI), play a pivotal role in most of the biological processes. Misfolding of proteins or aberrant PPIs is associated with many diseases such as cancer, infections, or neurodegenerative diseases. Hence, understanding the fold of proteins and its relationship to functionality is crucial for the efficient design of protein or PPIs modulators.

In recent decades, miniproteins have emerged as an excellent model system for studying protein fold. Miniproteins, which are polypeptides that weight less than 10 KDa, can fold into well-defined structures capable of retaining structural and functional elements of full-length proteins. Because of their relatively small size, miniproteins are synthetically available, allowing for investigation of the effects of single mutations on the fold, solvent conditions, and/or activity. In addition to enriching our knowledge of native protein folds, miniproteins have emerged as valuable tools in the design of *de novo* proteins and peptide-based therapeutics. By manipulating the amino acid sequence, miniproteins can be engineered with customised folds, opening new avenues for rational drug design.

Since the introduction of foldamers in the 1990s by Seebach and Gellman, α/β peptides have been extensively studied. The use of α -amino acids enables the introduction of specific functionalities by available side-chains, while β -amino acids are commonly used to control the overall shape of the molecule. The combination of peptide foldamers with miniprotein design, can derive in a wide range of new folds and functionalities not found in nature. Yet, there is no established methodology for the *de novo* design of β -amino acid containing miniproteins. In the present dissertation, our aim is to develop a methodology that will efficiently allow the design of a complex tertiary structure that incorporates cyclic β -amino acids. The miniproteins designed will be used as scaffolds for the design of inhibitors of complex targets involved in PPIs.

Streszczenie

Szeroki wachlarz funkcji molekularnych pełnionych przez naturalnie występujące białka jest możliwy dzięki ich precyzyjnie ukształtowanym strukturom, które zakodowane są w ich sekwencjach aminokwasowych. Struktury białkowe przeszły zmiany ewolucyjne tak, aby wykazywały różnorodne i odrębne funkcje istotne dla procesów biologicznych. Co więcej, interakcje między białkami (ang. *protein-protein interactions*, PPI), odgrywają kluczową rolę w większości procesów biologicznych. Nieprawidłowe fałdowanie białek lub nieprawidłowe PPI wiąże się z wieloma chorobami, takimi jak nowotwory, infekcje lub choroby neurodegeneracyjne. Dlatego zrozumienie zwijania się białek i jego związku z funkcjonalnością ma kluczowe znaczenie dla efektywnego projektowania białek lub modulatorów PPI.

W ostatnich dziesięcioleciach minibiałka stały się doskonałym systemem modelowym do badania fałdowania białek. Minibiałka, które są polipeptydami o masie mniejszej niż 10 kDa, mogą tworzyć dobrze zdefiniowane struktury, które zdolne są do naśladowania elementów strukturalnych i funkcjonalnych białek pełnej długości. Ze względu na ich stosunkowo mały rozmiar minibiałka są dostępne syntetycznie, co pozwala na badanie wpływu pojedynczych mutacji na fałd, warunki rozpuszczalnika i/lub aktywność. Oprócz wzbogacenia naszej wiedzy na temat natywnych struktur białkowych, minibiałka okazały się cennymi narzędziami w projektowaniu białek *de novo* oraz leków. Manipulując sekwencją aminokwasów, można konstruować minibiałka o wybranych strukturach, otwierając nowe możliwości racjonalnego projektowania związków biologicznie czynnych.

Od czasu wprowadzenia folderamerów w latach 90. XX wieku przez Seebacha i Gellmana, badania α/β -peptydów były szeroko rozwijane. Zastosowanie α -aminokwasów umożliwia wprowadzenie określonych funkcjonalności poprzez dostępne łańcuchy boczne, natomiast β -aminokwasy są powszechnie stosowane do kontrolowania ogólnego kształtu cząsteczki. Połączenie folderamerów peptydowych z konstrukcją minibiałek może skutkować powstaniem wielu nowych struktur i funkcjonalności niespotykanych w naturze. Nie ma jednak ustalonej metodologii projektowania *de novo* minibiałek zawierających β -aminokwasy. W niniejszej rozprawie naszym celem jest opracowanie metodologii, która skutecznie umożliwi zaprojektowanie złożonej struktury

trzeciorzędowej zawierającej cykliczne β -aminokwasy. Zaprojektowane minibiłka zostaną wykorzystane jako rusztowania do projektowania złożonych inhibitorów PPI.

List of Abbreviations

ACN	Acetonitrile
Ala	Alanine
Asp	Aspartic acid
Asn	Asparagine
BLI	BioLayer Interferometry
BSA	Bovine Serum Albumin
CD	Circular Dichroism
Cys	Cysteine
DIC	N,N'-Diisopropylcarbodiimide
DMF	N,N-Dimethylformamide
DSF	Differential Scan Fluorimetry
EC ₅₀	Half maximal effective concentration
E	Extended structure
ESI-MS	Electrospray Ionization Mass Spectrometry
Fmoc	Fluorenylmethyloxycarbonyl
Glu	Glutamic acid
Gln	Glutamine
Gly	Glycine
H	Helical structure
His	Histidine
HPLC	High Performance Liquid Chromatography
HTRF	Homogeneous Time Resolved Fluorescence
IC ₅₀	Half maximal Inhibitory Concentration
Ile	Isoleucine
K _a	Association constant
K _d	Dissociation constant
K _D	Affinity constant
KPB	Potassium Phosphate Buffer
Leu	Leucine
Lys	Lysine
Met	Methionine

NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
Oxyma	Ethyl cyanohydroxyiminoacetate
PBS	Phosphate Buffer Saline
PD-1	Programmed Death cell protein
PD-L1	Programmed Death cell protein Ligand 1
PPI	Protein-Protein Interaction
Pro	Proline
Phe	Phenylalanine
R _t	Retention Time
Ser	Serine
SAXS	Small Angle X-ray Scattering
SPPS	Solid Phase Peptide Synthesis
Thr	Threonine
TFA	Trifluoro Acetic acid
T _m	Melting Temperature
TOCSY	Total Correlation Spectroscopy
<i>Trans</i> -ACPC (◆)	<i>Trans</i> -(1 <i>S</i> ,2 <i>S</i>)-2-aminocyclopentanecarboxylic acid
<i>Trans</i> -APC	<i>Trans</i> -(3 <i>R</i> ,4 <i>S</i>)-4-aminopyrrolidine-3-carboxylic acid
Trp	Tryptophan
Tyr	Tyrosine
Val	Valine

Acknowledgments

The PhD journey is finished after four years, and what four years. Nothing of this would have been possible without the help of many people inside and outside the laboratory. Firstly, I would like to thank my supervisor, Professor Łukasz Berlicki for his patience in my endless interventions to discuss work. Also, I would like to thank him for the opportunity to participate in such a stimulating project as this dissertation. It has been extremely rewarding and has helped me develop as a scientist and individual.

I would also like to thank my research group, the ones that stayed and the ones that left, we had invaluable moments of discussion and fun inside and outside the laboratory either in Remont, doing barbecues or just by the bulwar of Politechnika. They have been a great source of support as they have always strongly showed their belief in my abilities. Special thanks to Agnieszka Ciesiolkiewicz for her irreplaceable support since day one. For her efforts to integrate me with the university, the shared discussions and friendship developed over the past four years.

Special thanks to the people with whom I had the privilege to cooperate. With Dr Ewa Rudzińska-Szostak and the endless efforts and patience with NMR studies and myself. I had the opportunity to learn beyond my expectations and it has been a great support for the finalisation of my work. I also thank Dr Magdalena Bejger for being always so generous and enthusiastic during my visits to the Polish Academy of Science in Poznan and the great dedication to my work and precious crystals. Likewise, I would like to thank Dr Łukasz Skalniak for his insight in cell assays, for always welcoming me into his laboratory and providing me with the freedom to work as if I was one more member of the group. Thank you for trusting me and thank you for the experience of working together. It has been a very exciting and, allow me to quote you, a fruitful journey through the narrow path between the hills of solubility and the abyss of toxicity.

Above all, I want to say thank you to my family and friends. To have always been there for me, in the good and the bad. I wish to thank my mother Concha and my father Juanvi. Also, it makes me very proud to dedicate this work to my Iaia Enriqueta, I wish she could be here to see me, I am sure she would be very happy. Thank you to my sisters, María, Marta, and Conchita. Thank you for always believing in me. For the visits, arrocitos, talks, for listening and for never slipping away in your enthusiasm regardless of how long

I stay out of sight, I love you. You have been able to escape my presentations during the past four years, but not the last. To my friend Borja Espert because despite the distance and time you are always there for me. Special thanks to my love Milena Kwiatkowska, none of this would have been possible without you. You are the best companion I could ever wish for. I am so grateful for your patience, for withstanding my moods and eternal peptide talks. But above all, for always believing in me. This PhD also belongs to you. I cannot wait to see what the future has prepared for us and what we are going to do together, I love you.

To my mother and father, I want to express my gratitude again as you have helped me become the person I am today. You have never flickered to help and support me, and not a single blink of doubt came out of you. I am very proud to dedicate this work to you all.

1. Introduction

1.1. The protein fold problem

Proteins mediate most of the biological processes in living organisms. The capacity and specificity to carry out their biological function are determined by their ability to fold into precise three-dimensional structures. Moreover, it is known that misfolding of proteins is the cause of many diseases. Considering this, understanding the protein fold is crucial for the development of molecules that regulate protein-mediated processes. To date, more than 200,000 high-resolution structures of proteins have been deposited in the Protein Data Bank¹ (PDB - RCSB.org). However, the intrinsic capacity encoded in its amino acid sequence to fold rapidly into precise and reproducible structures is yet to be fully understood. This is known as the protein fold problem. Since the resolution of the first protein structure of myoglobin in 1958², three major questions have arisen: I) what is the information that is contained in the sequence that determines the final fold, II) how do protein fold so quickly and what is the mechanism, and III) will it be possible to predict protein fold.

1.1.1. Information encoded in the amino acid sequence

The study of contributions to the fold from the deposited PDB of proteins sheds light on the encoded requirements for a protein to fold, (Figure 1). Six types of contributions were described, A) The hydrogen bonds, L. Pauling in 1933³ already described the formation of the secondary structures by intramolecular hydrogen bonding of the residues in the sequence, B) The Van der Waals interactions, which are weak interactions; however, considering the globular and compact nature of the fold, they are counted in thousands and have a great influence on protein stability⁴, C) Proteins cannot adopt any backbone angle. The dihedral angles of a protein backbone have determined preferences, as was described by Ramachandran in 1962⁵, D) Chain entropy and thermodynamics⁶⁻⁸, contrary to the folding process, there is an entropic penalty promoted by the higher degree of structural organisation of the protein when folded, E) The hydrophobic interactions^{7,9-14}, and F) electrostatic interactions¹³, although one of the less relevant interactions on protein fold, due to constant exchange with solvent. Yet, they control conformation.

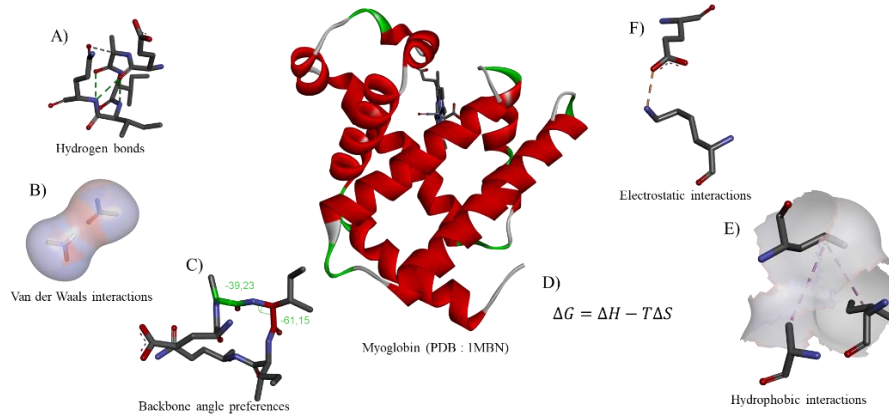


Figure 1. Information encoded into the amino acid sequence, A) Hydrogen bonds, B) Van der Waals interactions, C) Backbone angle preferences, D) Chain entropy, E) Hydrophobic interactions, and F) Electrostatic interactions.

1.1.2. Protein folding pathway

In 1960, Cyrus Levinthal proposed what later became known as Levinthal's paradox. The argument stated that there are too many possible conformations for a polypeptide chain to explore through a random conformational search. For instance, considering a 101 amino acid-long chain with three degrees of freedom per residue, it would have 3^{100} possible configurations. If the chain was to explore 10^{13} configurations per second, it would take approximately 10^{27} years to exhaustively search all possibilities. Nevertheless, experimental observations demonstrated that proteins fold spontaneously in a matter of seconds or less. Levinthal concluded that a specific folding pathway, characterised by a well-defined sequence of events, must exist. He proposed that achieving the native structure and folding speed were two distinct factors: one governed by thermodynamics (achieving the stable minimum energy state) and the other by kinetics (reaching the folded state within a reasonable time frame).

The protein folding pathway involves several stages and mechanisms¹⁵. It begins with the linear sequence of amino acids, known as the protein's primary structure, which ultimately dictates its folding pattern and stability¹⁶. Secondary structure formation, such as the folding of α -helices and β -sheets, occurs through hydrogen-bonding interactions in the protein backbone. Yet, the formation of intramolecular hydrogen-bonds is not spontaneous due to an energetic barrier¹⁷, (Figure 2).

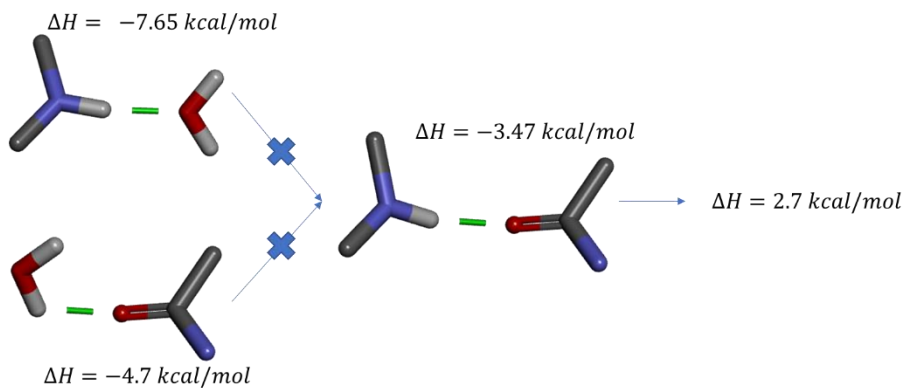
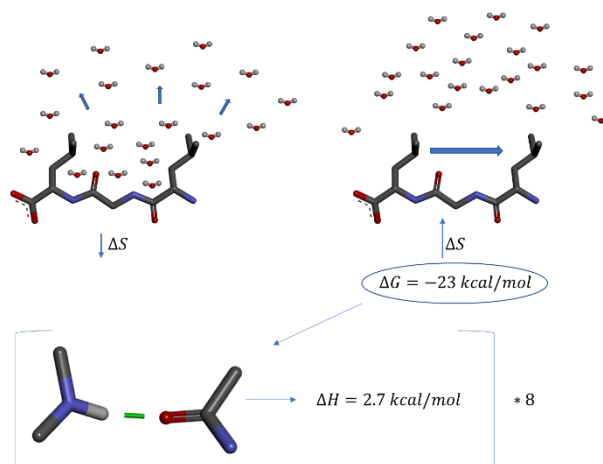


Figure 2. Energy barrier for intramolecular hydrogen bond formation.

The folding process often involves the formation of nucleation sites, where a small region or a few amino acids initiate the formation of the protein's core structure. The reorganisation of water molecules compensates entropically the enthalpic penalty of forming folding intermediates and intramolecular hydrogen bonds¹⁸, (Figure 3). Partially folded structures are frequently populated during the folding process and can influence the folding kinetics and stability¹⁹ of the protein.

A)



B)

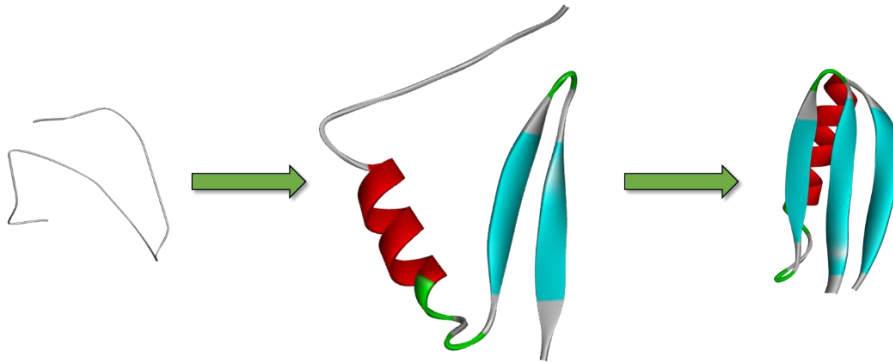


Figure 3. A) Entropic compensation to hydrogen bond formation from buried hydrophobic regions, and B) schematic representation of the protein folding pathway.

The protein folding pathway can be imagined as a funnel-shaped energy landscape^{16,20}, where proteins explore a vast conformational space and converge toward the native structure due to its favourable energy^{21,22}. Molecular chaperones play a crucial role in assisting protein folding by preventing misfolding and aggregation, as well as guiding proteins along the correct folding pathways²³. Experimental techniques such as NMR spectroscopy²⁴ and single-molecule fluorescence spectroscopy²⁵ have provided invaluable insights into the folding pathway, allowing the study of folding kinetics, transient intermediates, and energy landscapes. By unravelling the intricacies of the protein folding pathway²⁶, it is aimed to understand how proteins achieve their functional three-dimensional structures and apply this knowledge in fields such as drug design and protein engineering.

1.1.3. Structure prediction

Prediction of protein fold²⁷ is a topic of great interest in the scientific community due to its significant implications in various fields such as biology, medicine, and biotechnology. The field of protein structure methods has gained great advances thanks to the biennial critical assessment of structure prediction (CASP) meetings^{28,29}. The CASP meeting was first introduced by Moulton and colleagues in 1994, where over 150 groups all over the world were assigned with 100 sequences of unpublished but resolved structures and challenged to find an algorithm that could predict the structure with great precision. Initially, most of the methodologies were approached by homology of published structures

of related sequences. However, the progress in protein sequencing pushed the development of algorithms that could predict protein fold from the sequence in the absence of similar structures. These new methodologies were called *Ab initio*³⁰. *Ab initio* methodologies are usually a combination of approaches, e.g., fragment assembling, template-based, Monte Carlo, all-atom refinement, neural networks, molecular dynamics, etc. Here, the most representative *ab initio* methodologies are presented, which are related to the present dissertation.

1.1.3.1. Stochastic methods (Rosetta)

Stochastic methods³¹ usually use coarse-grained lattice models for the initial stages of the fold and a finer lattice for the later stages. It applies random rotational and translational movements in a 3D lattice and validates the movements by either estimating the probability of this movement to happen (Boltzmann distribution) or directly by energy decrease. This approach is the Monte Carlo method³², (Figure 4). By reproducing the above-mentioned steps, the algorithm samples a range of conformations, gradually converging towards the conformations with the lowest energy. Monte Carlo methods offer computational efficiency but may struggle with accurately capturing the lowest energy conformation.

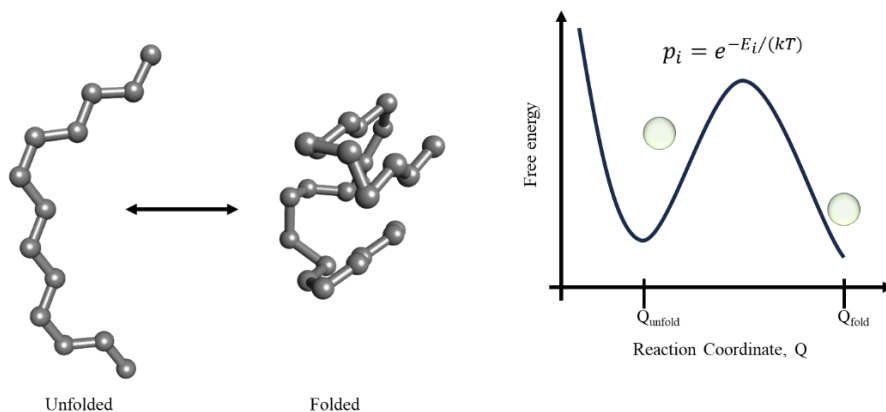


Figure 4. Representation of the coarse-grained lattice model used in Monte Carlo simulations (left) and Boltzmann distribution of probability (right).

Rosetta software was presented in 1997³³ as an *Ab initio* protein-fold prediction³⁴ programme. Rosetta begins searching for, similar nine-residue length fragments in the PDB. Using the Monte Carlo approach, it randomises the position of approximately 30000 of these fragments and the torsion angles, over the initial extended chain. The intermediate folded structure is then reprocessed with a search of three-residue fragments. This approach quickly collapses the linear sequence into a folded structure³⁵. Subsequently, an all-atom refinement^{36,37} of the lowest energy models is performed to generate the final structure.

1.1.3.2. Deep learning (AlphaFold2)

Deep neural networks have emerged as powerful tools for prediction of protein folding. In recent years, deep learning approaches, including convolutional neural networks³⁸ (CNNs) and recurrent neural networks³⁹ (RNNs), have shown promising results in predicting protein structures accurately from their amino acid sequences. In 2018 in the CASP13⁴⁰, AlphaFold⁴¹ was presented as a revolutionary algorithm that applies deep learning to predict protein fold from its amino acid sequence. Two years later, in the CASP14, DeepMind presented an improved version called AlphaFold2⁴².

AlphaFold2 is a deep learning model that was trained on a large database of protein structures from the PDB and other sources. The learning process combined the high-resolution structures and their corresponding amino acid sequences. Once an input sequence is provided, the neural network takes the sequence and convolutes it continuously. The amino acids of the sequence are paired between them, and, through a search in the database, it estimates the relation between amino acid pairs. Simultaneously, it generates multi-sequence alignments (MSA) of similar sequences from different organisms to recognise patterns and conserved regions related to relevant mutations of the structure. AlphaFold2 then predicts the three-dimensional structure of the protein by estimating the spatial coordinates of the atoms through an iterative process that occurs three times. The generated model is based on the estimated distances and angles between pairs of residues. A final refinement of the model is performed by energy minimisation using the AMBER99SB⁴³ force field, to remove violations and clashes⁴⁴, in the presence of harmonic restraints in heavy atoms. The process is repeated if violations remain without the harmonic restraints over the heavy atoms intervening in the clash.

1.1.3.3. Molecular dynamics

Molecular dynamics (MD) simulations are computational techniques that simulate the movements and interactions of atoms and molecules over time. In the context of protein structure prediction, MD simulations can be used to model the dynamics of a protein in an aqueous environment⁴⁵. MD simulations solve Newton's equations of motion for each atom in the protein system, considering the forces described in Section 1.1.1. of the present dissertation. These forces are defined by potential energy models compiled in the force fields⁴⁶. By integrating these equations over time, the simulation generates a trajectory depicting the protein's motion. MD simulations enable the exploration of the conformational space of a protein, facilitating folding studies and analysis of structural dynamics. Through examination of the sampled conformational ensemble during the simulation, insights into the structure and behaviour of the protein are gained. Even though MD simulations have succeeded in the prediction of several folds^{47,48}, they lacked accuracy on protein stability and thermodynamics. More recently, *ab initio* MD^{49,50} has a growing importance in the field of structure prediction, reflecting more accurate results than classical MD⁵¹⁻⁵³. These simulations utilise electronic structure methods, such as density functional theory (DFT) or wave function-based methods, to determine the forces that act on the atoms. This allows for an explicit treatment of quantum effects, including chemical bonding, electron delocalisation, and charge transfer. *Ab initio* MD simulations provide a more accurate depiction of electronic and atomic dynamics. Nevertheless, they require substantially more computational resources than classical MD.

1.2. Protein-protein interactions

Protein-protein interactions (PPIs) are fundamental biological processes that involve physical associations between two or more proteins⁵⁴. PPIs play a crucial role in regulating various cellular functions, such as cell signalling, enzymatic activity, cell proliferation growth, apoptosis, etc⁵⁵⁻⁵⁷. The intricate system of PPI in cells is called interactome⁵⁸. Therefore, understanding PPIs provides valuable information on the complexities of biological pathways and cellular processes⁵⁹.

In the context of the immune system, immune checkpoints are specific regulatory mechanisms controlled by PPIs. Immune checkpoints are essential to maintain immune balance and prevent excessive immune responses that could potentially harm healthy tissues⁶⁰. An important immune checkpoint involves the interaction between the programmed cell death protein 1⁶¹ (PD-1) on T cells and its ligand, programmed death-ligand 1 protein (PD-L1), on target cells, including cancer cells. When PD-1 binds to PD-L1, T cell activity^{62,63} is inhibited, leading to immune suppression and enabling cancer cells to evade immune surveillance. The field of immunotherapy has revolutionised cancer treatment by harnessing the power of the immune system⁶⁴⁻⁶⁶. Immunotherapy encompasses checkpoint inhibitors, which are drugs designed to block interactions between immune checkpoint molecules, e.g., PD-1/PD-L1. By doing so, these inhibitors release the "brakes" on the immune system, reactivating T cells to effectively target and attack cancer cells⁶⁷. The remarkable success of checkpoint inhibitors highlights the importance of understanding protein-protein interactions in immune regulation. By specifically targeting PPIs through immunotherapy, promising results have been achieved in the treatment of various cancers and other diseases, improving the immune response against malignant cells⁶⁸. The interaction between PD-1 and PD-L1 has undergone extensive clinical validation, and several monoclonal antibodies (mAbs), including atezolizumab, durvalumab, and avelumab, have been approved for the treatment of various cancers, such as malignant melanoma^{69,70}. The research focus on miniproteins, is incited by the potential to provide affinities and specificities comparable to mAb⁷¹, but with improved bioavailability, stability, and cost efficiency⁷²⁻⁷⁴. Moreover, the introduction of noncanonical amino acids can expand the binding modes and affinities by providing new folds and side-chain availability⁷⁵.

1.2.1. PD-1/PD-L1 immune checkpoint

The PD-1 protein is a transmembrane glycoprotein with a sequence length of 288 amino acids. It belongs to the immunoglobulin (Ig) superfamily and is divided into an extracellular N-terminal IgV-like domain, which assumes a β -sandwich by the formation of a hydrophobic core between two large β -sheets, a transmembrane domain and a cytoplasmic tail⁷⁶. The PD-1 ligand, PD-L1, is formed by two IgV domains, the N-terminal domain is known to be responsible for the interaction with PD-1. However, the function of the C-terminal domain has not yet been elucidated. Potentially, it could exist as a spacer to minimise steric hindrance when interacting with PD-1⁷⁷.

The structure of the complex between PD-1 and PD-L1 has been resolved by X-ray crystallography. The crystal structure of the PD-1/PD-L1 complex highlighted the notorious difficulty of addressing this interaction. Nevertheless, it has been used to widely study the basis of this interaction for the design of effective binders⁷⁸. It was found that the interaction is constructed around a central hydrophobic core, dominated by CH-CH and CH- π interactions, generated between the side chains of PD-1 (Val64, Ile126, Leu128, Ala132 and Ile134) and PD-L1 (Ile54, Tyr56, Met115, Ala121 and Tyr123), (Figure 5), PDB id: 4ZQK⁷⁹. Supporting the hydrophobic binding there are two major regions of interactions that contribute to the formation of the complex.

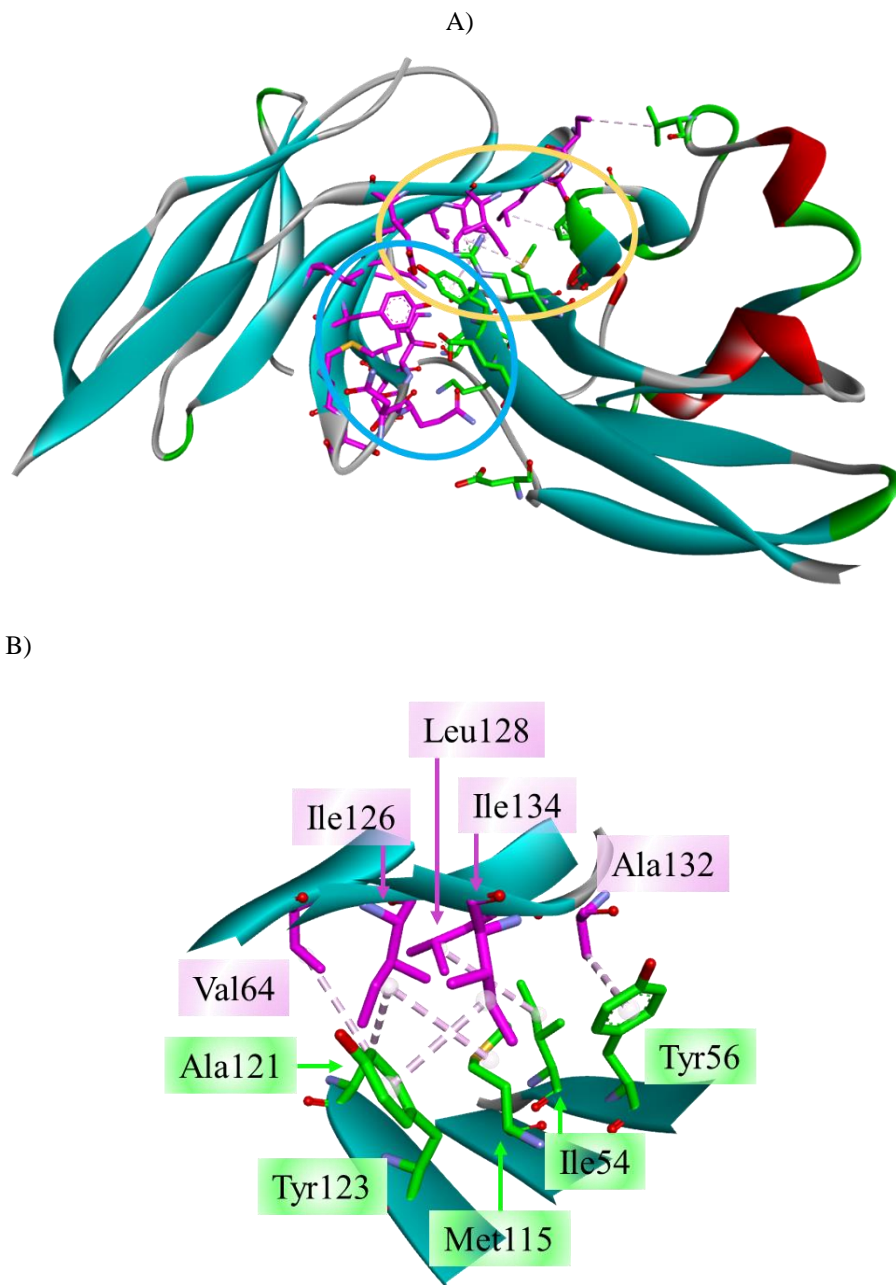


Figure 5. A) Crystal structure of human PD-1 in complex with human PD-L1 (PDB id: 4ZQK). PD-1 interactive residues are represented in magenta while PD-L1 residues are represented in green. The central hydrophobic interaction is indicated with a yellow circle and the polar and nonpolar region in blue, and B) Central hydrophobic cluster of the PD-1/PD-L1 interaction.

A buried region of polar and nonpolar interactions with a pronounced π - π stacking interaction between the side-chains of Tyr68 and Tyr123 (to facilitate understanding of the precedence of the residues, the PD-1 interacting residue will always be described first while the PD-L1 interacting residues will be named after). Additionally, the hydroxyl group within the side chain of Tyr68 forms a hydrogen bond with the side-chain carboxyl of Asp122, and a similar contact is observed between the side-chain carboxyl of Glu136 and the side-chain hydroxyl of Tyr123. Furthermore, a buried hydrogen bond between the side chain of Asn66 and the main-chain carbonyl oxygen of Ala121 also contributes to the hPD-1/hPD-L1 interaction, (Figure 6).

A peripheral network of solvent-exposed interactions, such as the backbone amide or salt bridge formations, are less relevant to the inhibitor design because solvent-exposed hydrogen bonding and electrostatic interactions are in constant exchange with the solvent, minimising their contribution to the binding.

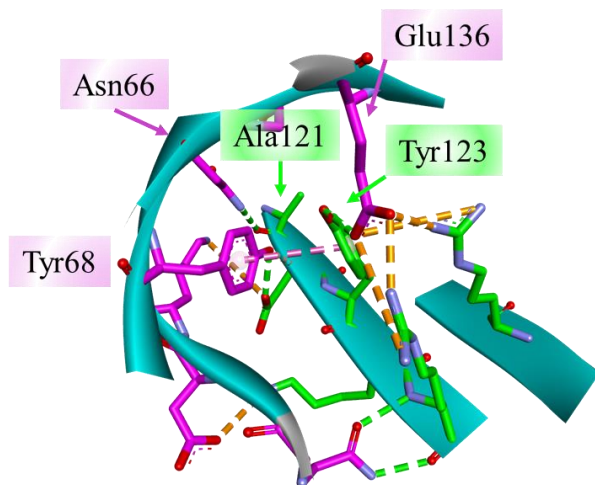


Figure 6. Buried region of polar and nonpolar interaction in the PD-1/PD-L1 complex. PD-1 interacting residues are highlighted in magenta while PD-L1 are highlighted in green. π - π stacking is indicated with a dashed magenta line and hydrogen bonds in green. Orange dashed lines indicates electrostatic interactions.

In conclusion, the essential components of this system are large, smooth, and lipophilic surfaces with no notable pockets, leading to numerous hydrophobic interactions. This specific mode of interaction presents challenges in targeting the PD-1/PD-L1 system⁸⁰.

1.3. Miniproteins

Miniproteins have Mw below 10 KDa and fold into well-defined and diverse tertiary structures. The miniprotein fold is driven by one of these possible forces, A) cysteine bridges (kunitz domain), B) metal binding (Zinc fingers), or C) hydrophobic core (Villin headpiece), (Figure 7).

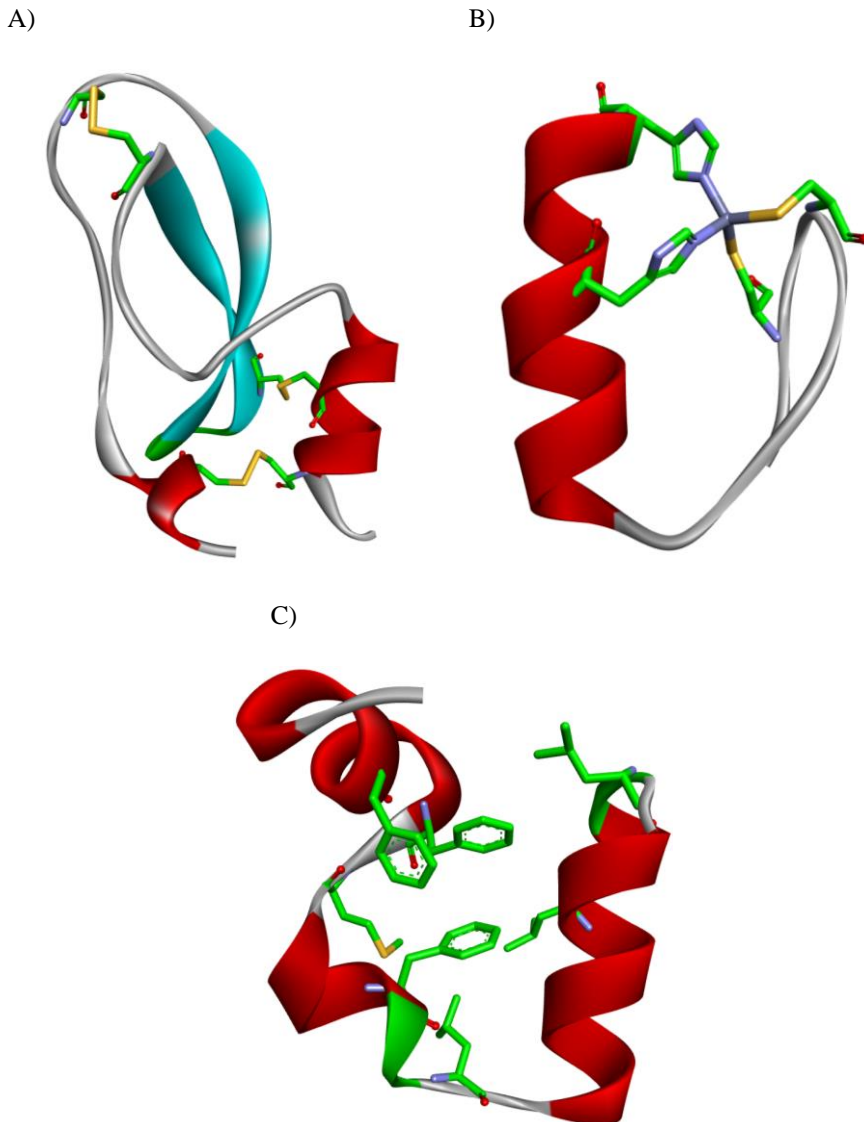


Figure 7. A) Kunitz domain (PDB id: 4BQD)⁸¹, B) Zinc finger (PDB id: 1ZNF)⁸², and C) Villin headpiece (PDB id: 1YRF)⁸³, Highlighted in green the residues responsible for the fold stabilisation.

Miniproteins, which are found in nature on the order of thousands⁸⁴, play important roles in living organisms, primarily exhibiting regulatory functions. Miniproteins possess the ability to bind targets specifically and with high affinity. Moreover, miniproteins are more convenient to produce, via chemical synthesis, and administer compared to antibody-based therapeutics. Therefore, miniproteins can provide the qualities necessary to cover the space between small-molecule drugs and biologic drugs, adapting the best of both worlds⁸⁵. From the design point of view, miniproteins small size and protein-like fold, have attracted interest in these systems for the study of protein fold⁸⁶. Furthermore, miniproteins can be used as scaffolds for mimicking or designing new functionalities^{87,88} as enzyme activity⁸⁹, protein-protein interaction (PPI) inhibitors⁹⁰⁻⁹² and receptor agonists/antagonists^{93,94}.

1.3.1. Miniproteins in nature

Miniproteins are abundant in nature, and they can be found showing diverse topologies and functionalities. The most predominant natural miniproteins are the cysteine-rich miniproteins (CRMPs), linear (knottins), or cyclic (cyclotides) miniproteins with several disulphide bridges. However, since the discovery of the avian pancreatic polypeptide (aPP) in 1981⁹⁵, a diverse group of folds, stabilised by hydrophobic core or metal binding, have been identified and/or optimised.

1.3.1.1. Cysteine-rich miniproteins

CRMPs⁹⁶ are a diverse family of small proteins that possess well-defined folded structures and exceptional stability, all conferred by their constrained cysteine knot. The cysteine knot, unique to this family of miniproteins, is formed by six cysteines in a span of approximately 30 sequential amino acids, which form three disulphide bridges⁹⁷. The majority of CRMPs families exhibit a high degree of homology on the arrangement of these bridges⁹⁸, (Figure 8), where Cis(I)-Cys(IV), Cys(II)-Cys(V) and Cys(III)-Cys(VI) bridges are observed. The knot⁹⁹ terminology refers to the formation of the bridge connecting Cys(III) with (VI) crossing through the stabilised structure generated by the other two bridges. Yet, common motif among all CRMPs, is the cysteine-stabilized β -sheet motif^{100,101}, formed by Cys(II), (III), (V) and (VI), that is most conserved for all the CRMPs families¹⁰², and it is considered the primary motif.

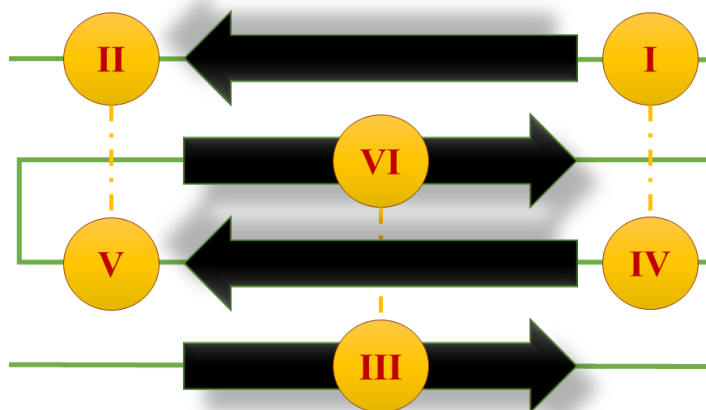
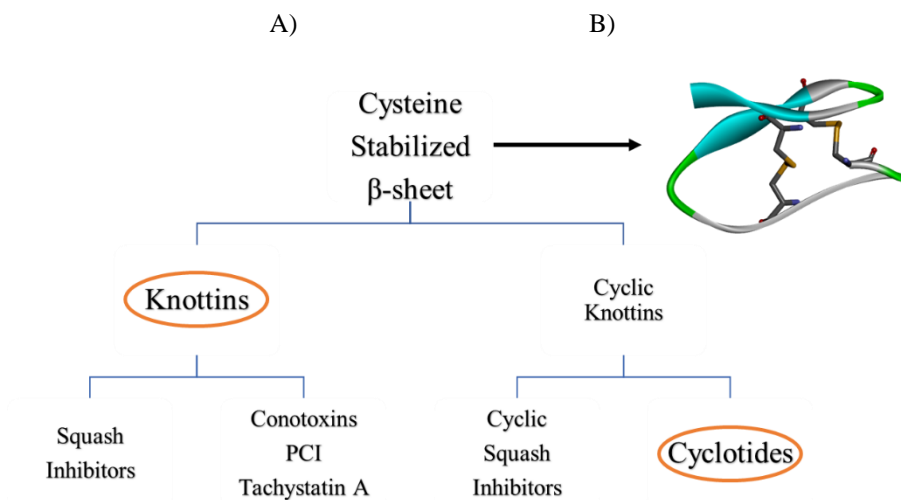


Figure 8. Cysteine knot representation.

From this primary motif two main CRMPs families can be classified, the Knottins¹⁰³, linear miniproteins with the cysteine knot, and the cyclic Knottins, which exists in cyclic form. Within the cyclic knottins, the most widely studied subfamily and used, as scaffold for the design of active compounds, are the Cyclotides^{104,105}, Scheme 1. The two major differences between these categories of CRMPs are, the head-to-tail cyclization of the cyclotides¹⁰⁶ and the origin of these miniproteins. Although knottins can be found in all living organisms, cyclotides are found exclusively in plants^{107,108}.



Scheme 1. A) distribution of the CRMPs families and subfamilies, and B) Cysteine-stabilised β -sheet motif (CSB) from Kalata B1.

Nowadays there are over 3300 sequences of knottins in the KNOTTIN database and to date there are over 1300 cyclotide sequences in the Cybase. However, the two most representative miniproteins for each subfamily are the *Ecballium elaterium* Trypsin Inhibitor II (EETI II)^{109,110} and Kalata B1^{111,112}, respectively, (Figure 9). Both miniproteins have proven to be ready-to-use scaffolds for therapeutic design. Furthermore, the NMR studies¹¹³ of EETI-II, and thermochemical and enzymatic denaturation¹¹⁴ of Kalata B1, in comparison to analogue sequences in the absence of the cysteine knot, confirmed the crucial role of the cysteine knot in the fold and stability. Additionally, these experiments indicated a low contribution of the rest of the sequence to the stability and fold. These studies confirmed the potential of knottins and cyclotides for the design or graft of active sequences into their loops without the loss of the three-dimensional arrangement^{115,116}.

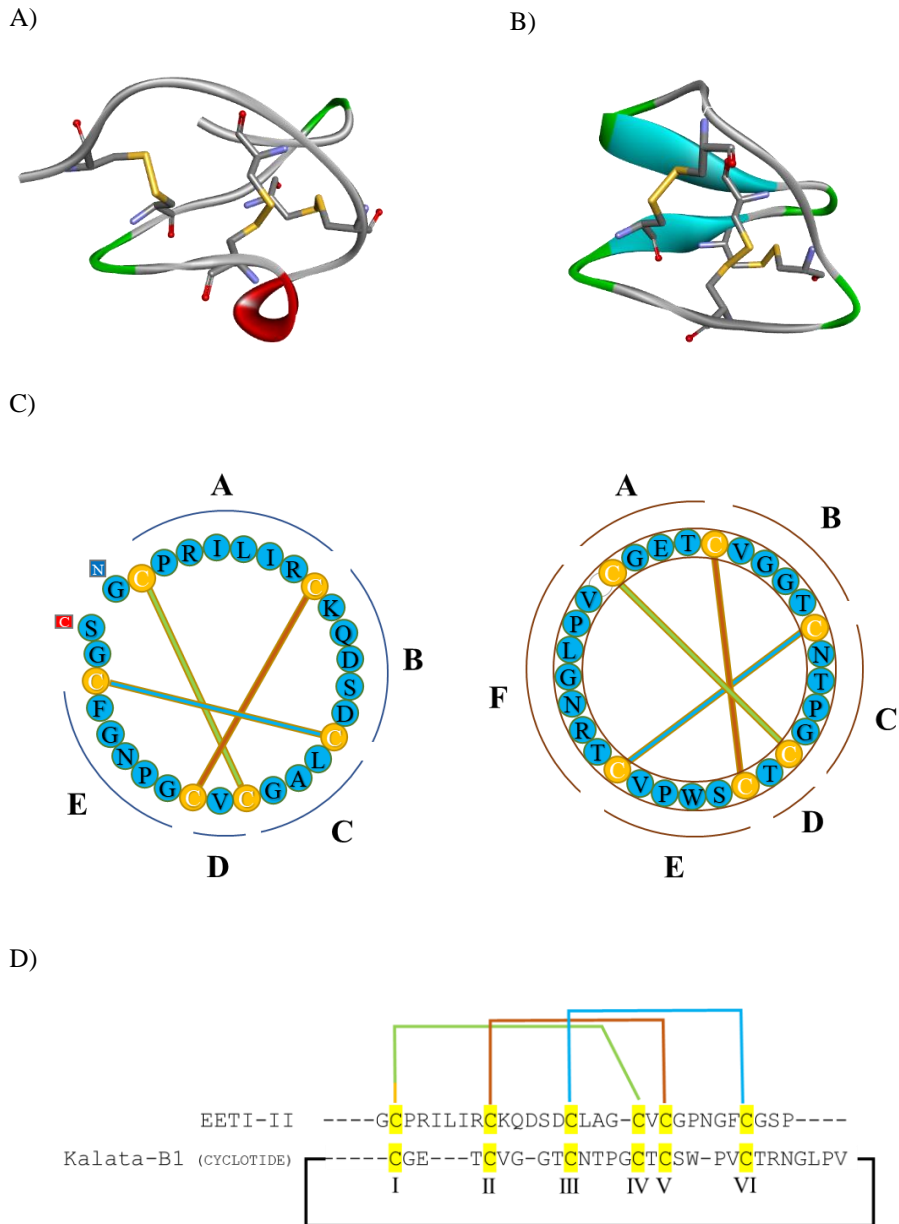


Figure 9. A) Crystal structure of *Ecballium elaterium* Trypsin Inhibitor II EETI-II (PDB id: 2IT7), B) KalataB1 (PDB id: 1NB1), C) Schematic representation of the fold for EETI_II and Kalata B1, and D) Linear representation of the highly conserved cysteine residues.

1.3.2. Engineered miniproteins

Miniproteins stabilised by hydrophobic core are of great importance for the design and understanding of the protein fold. Most of the cases found in nature are fragments of proteins that fold in solution in the absence of the rest of the protein or have been engineered from motifs in nature to do so. Additionally, miniproteins stabilised with non-covalent interactions manifest an immense variety of secondary structure combinations^{86,87}.

Among these motifs, the β -sheet is highly intriguing for the research community. With its extensive flat surfaces, it holds significant relevance as a candidate for developing protein-protein interaction (PPI) inhibitors, as it can effectively mimic protein surfaces, but also for the design of materials as amyloid fibrils. In addition to its medicinal applications, the design of β -sheets remains an expanding field¹¹⁷. This is because isolated β -sheet units often face challenges such as aggregation¹¹⁸ or inadequate folding.

β -Hairpins are one of the most repeated motifs in proteins¹¹⁹. They are composed by two β -strands connected by a loop that orients the strands to form the hairpin. The first isolated β -hairpin, from the N-terminal domain of Ubiquitin, (Figure 10A) showed adequate folding in aqueous methanol at 30-60 %. The use as building blocks of these motifs with other secondary structures can derived in the stabilisation of greater and more complex structures.

WW-domains are one of the smallest protein domains that fold in solution when cleaved from the rest of the protein. In solution, WW-domains form a right twisted, triple stranded, antiparallel β -sheet¹²⁰. Their names originate from the presence of two conserved Trp residues close to the N- and C-termini of the sequence¹²¹. In 2001, the WW-prototype was designed¹²², (Figure 10B). This new sequence, with a T_m value of 44.2 °C, showed by NMR the propensity to fold as WW-domains in solution. The design of this miniprotein involved identifying key elements for folding, namely the conserved N-terminal Trp and C-terminal Pro residues. Furthermore, a cluster of aromatic/hydrophobic interactions between the C-terminal Trp residue and two Tyr residues from the second strand was found, contributing to the enhanced stability of the miniprotein. Moreover, native WW-

domains have been used as scaffolds for the construction of biologically active molecules¹²³.

Avian pancreatic polypeptide (aPP) is a peptide hormone released by the pancreas¹²⁴. The structure of this 36 residue miniprotein, is composed of a poly proline type II helix (PPII) connected by a loop to an α -helix⁹⁵. The fold is stabilised through a network of CH- π interactions between the Pro, from the PPII, and aromatic residues of the helix. The compact fold of this miniprotein is mediated through a hydrophobic core. Considering the folding forces of the aPP, (CH- π interactions), the PP α -Tyr was designed¹²⁵. The design fragment assembled a Tyr rich helix, from the surface adhesin and antigen (AgI/II) of *Streptococcus mutants*, and the bovine pancreatic polypeptide hormone. The designed PP α -Tyr had a reasonable thermal stability with a T_m of 39 °C. Optimisation of the scaffold increasing CH- π and electrostatic contacts, improved the T_m values by 12 °C¹²⁶, (Figure 10C).

The Trp-Plexus motif was extracted from the Fibronectin type III domain of Interleukin 3 Receptor Beta (IL3RB). The fragment was cut off from the protein composed of a β -strand rich in Arg residues, forming cation- π -cation interactions¹²⁷ with a polyproline-free type II helix rich in Trp residues. This polyproline-free type II helix can be found in proteins with WSXWS motifs. To form the final fold of the miniprotein, these two secondary structures were connected with a D-Pro-Gly loop to promote the proper orientation¹²⁸, (Figure 10D). The mimicry of the PPII helix by this scaffold allows the generation of therapeutic compounds that imitate PPII-protein interactions¹²⁹.

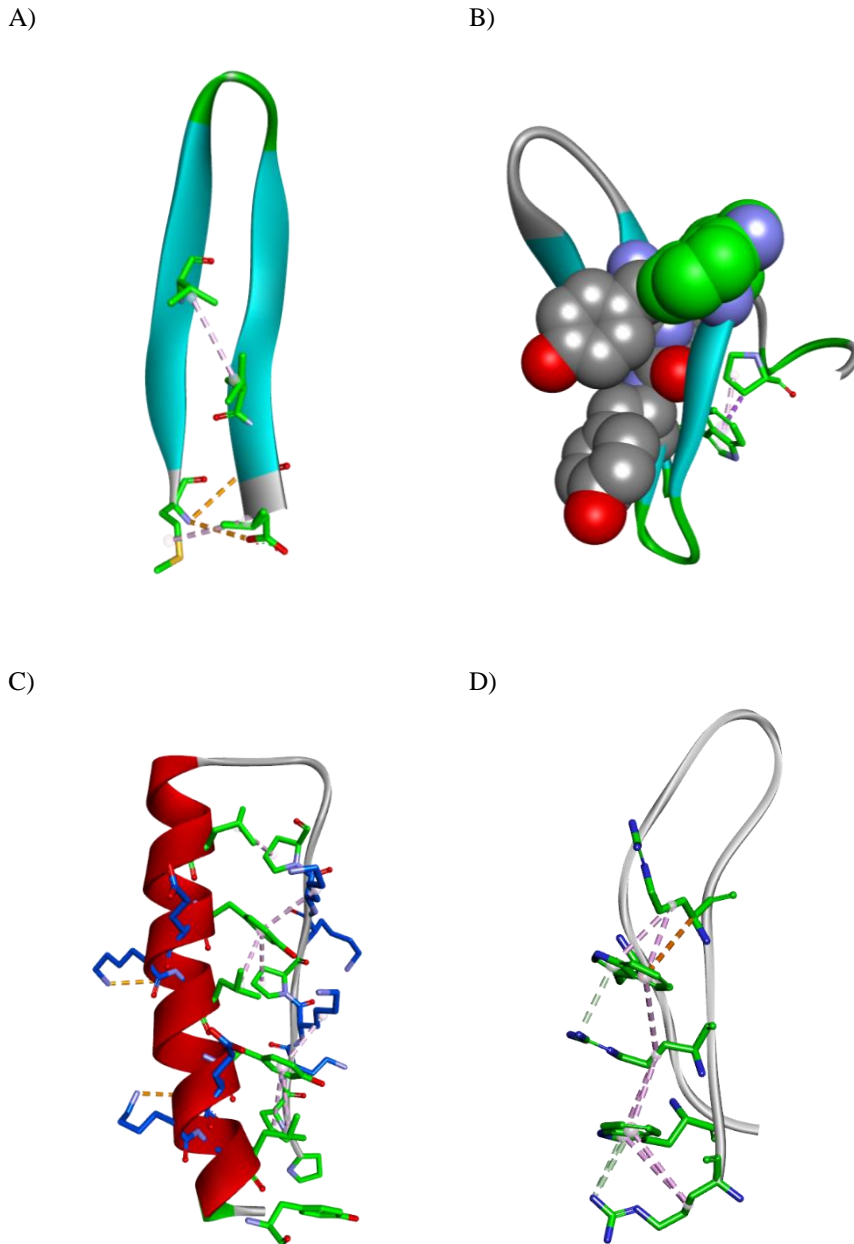


Figure 10. Engineered miniproteins. A) β -hairpin (PDB id: 1E0Q), B) WW-domain (PDB id: 1E0M), C) PP α (PDB id: 6GWX), and D) Trp_Plexus (structure available in the supplementary information of reference 127). Residues crucial for the folding are shown as sticks in green, conserved aromatic contacts of the WW-domain are presented as van der Waals sphere, and key mutations for indirect fold are highlighted in dark blue. The most significant interactions are indicated with dashed lines.

TrpCage is to date the shortest designed sequence that folds through a hydrophobic core. Though, TC5b¹³⁰, cut off from the peptide Exendin-4 from Gila monster saliva, showed a T_m value of 42 °C, complete fold was only observable below 10 °C. This observation indicates that the protein was only partially folded at room temperature. TC10b¹³¹, a TC5b mutant with an enhanced helical fold and a T_m value of 64 °C, (Figure 11A), exhibited the desired fold by NMR. Additionally, it was possible to elucidate the importance of the interaction between Try3 and Pr19 in the overall stability. TrpCage is one of the most studied examples of protein fold experimentally and computationally¹³².

Zinc fingers¹³³ are well known for folding in the presence of a metal ion, which induces the conformation of these domains¹³⁴. The organised structure of a zinc finger is composed of a β -hairpin bounded to an α -helix through the coordination of Zn^{+2} ion by two His and two Cys residues⁸². However, engineering efforts were made towards the design of a zinc finger-like fold in the absence of metal binding, (Figure 11B). The design intended to replace the metal dependency of the fold by a hydrophobic core. To that extent, hydrophobic conserved residues were identified and preserved, by sequence alignment of several naturally occurring zinc fingers. Furthermore, the Cys at the binding site was replaced by 3-(1,10-phenanthrol-2-yl)-L-alanine to enhance CH- π contacts. Optimisation of solvent-exposed residues in combination with a type II' β -turn (D-Pro-Ser) led to a folded metal-free zinc finger analogue with two orientations. The final substitution of His3 for Val reduced the isomerisation of the type II' β -turn and led to the first metal-free zinc finger¹³⁵.

The Villin headpiece is a fast-folding¹³⁶ triple helix motif from the chicken protein⁸³, fragment 41-76, (Figure 11C). It is composed of 35 residues that generate a hydrophobic core between the three helices and show remarkable thermal stability with a T_m of 70 °C. Regardless of the intrinsic stability of this miniprotein, the rational modification of HP36, a truncation of one residue longer of the VHP, led to an improvement of the thermal stability with a T_m of 17 °C¹³⁷. Two mutations were incorporated into the sequence, Ala68 to enhance the helical fold and M70 to reduce adverse electrostatic interactions. This research showed that there is still a window for the improvement of naturally occurring miniproteins to further enhance stability and/or activity.

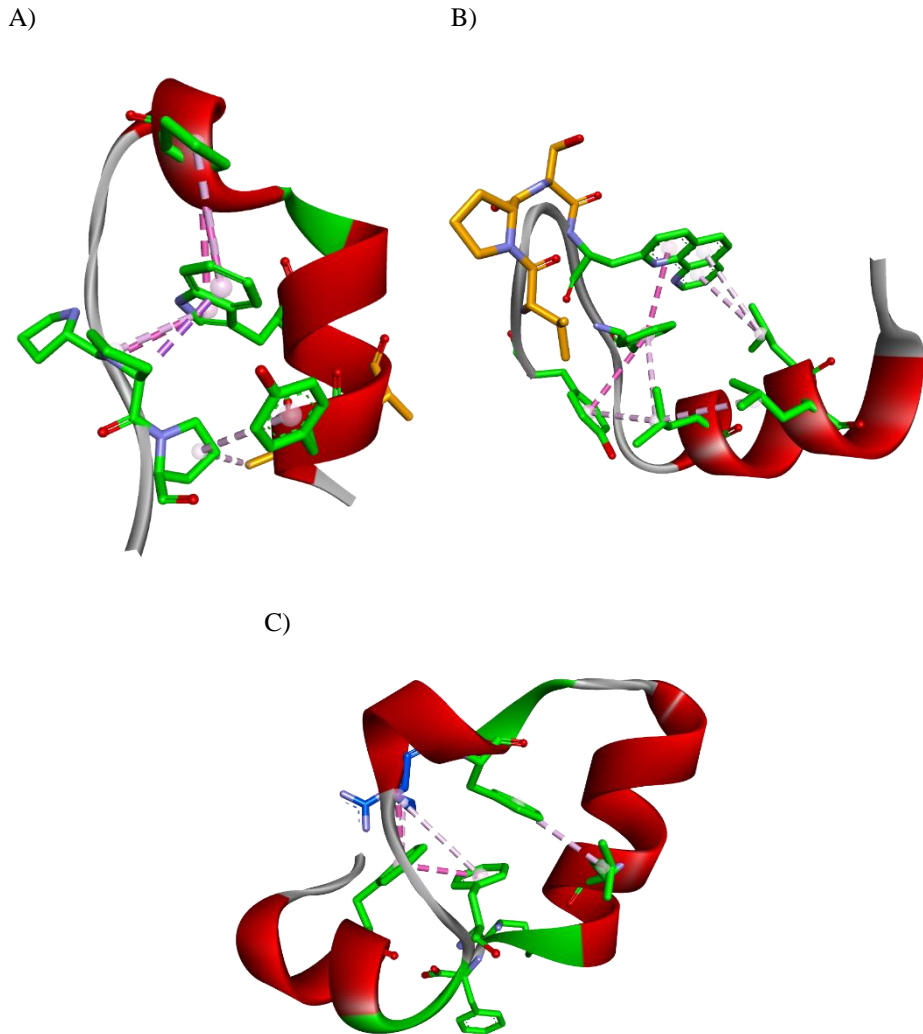


Figure 11. Engineered mini-proteins A) Trp_cage (PDB id: 2JOF), B) Metal free Zinc finger analogue (PDB id: 1HCW), and C) Villin HP35 (PDB id: 1VII). Residues crucial for the folding are shown as sticks in green, and key mutations for indirect fold are highlighted in orange. The most significant interactions are indicated with dashed lines.

1.3.3. *De novo* designed miniproteins

The increased complexity and understanding of biological processes has pushed the research community to develop new methodologies for the design of highly active molecules^{138–140}. Miniprotein design has significantly evolved from initial designs based on amino acid propensities and low access to high-resolution structures to the design of entirely new sequences and topologies with atomic accuracy, so-called the *de novo* design. *De novo* miniprotein design¹⁴¹ allows to generate new topologies never found in nature with completely new sequences. The design explores all possible combinations of amino acids led by the physical principles that dominate the protein fold. In the *de novo* design, initially the desired backbone conformation is generated. To fill the sequence that will form the target fold, a set of rules is applied: i) energetically favourable contacts between side chains backbone side chains, ii) classification of amino acid residues for favourable regions (solvent exposed, hydrophobic core, or non-preferential), iii) identification of rotamers that lower the energy, and iv) reduction of accessible conformations implementing negative design.

Early work on *de novo* design led to a cooperatively folded miniprotein with the $\beta\beta\alpha$ motif. The full sequence design (FSD)¹⁴², (Figure 12A). In this early work, by Dahiyat and Mayo, residues 33-60 of the zinc finger Zif268 (PDB: 1ZAA) were used as template for the backbone. Following the backbone selection, the sequence design was based on the estimated potentials of interactions between the side-chains and backbone-side-chains, using DREIDING force field¹⁴³. Amino acids were further classified into: A) core position, if the distance from the $C\alpha$ (along the vector with $C\beta$) to the solvent accessible surface was greater than 5 Å and the distance from $C\beta$ to the surface greater than 2 Å, B) surface positions, if the sum the distances mentioned above was less than 2.7 Å, C) all remaining residues were classified as boundary. Yet, the number of sequences generated was too great to be synthesised. To reject unviable designs, the sequences were optimised with pairwise rotamer energy calculations and dead-end elimination (DEE) searching¹⁴⁴. The resulting sequence led to the FSD miniprotein with a T_m value of 39 °C.

Trp-zipper¹⁴⁵ was a *de novo* designed miniprotein obtained by iterative modification of the non-hydrogen bonded (NHB) strand positions of a previously reported disulphide cyclized β -hairpin¹⁴⁶. During the modifications, the thermodynamic repercussions of the mutants was studied. It was concluded that the presence of cross-stranded Trp enhanced the stability to levels comparable to large protein domains, $\Delta G^0 = 90 \text{ cal} \times \text{mol}^{-1} \times \text{residue}^{-1}$. Trpzip1, (Figure 12B), showed cooperative unfolding with T_m of approximately 50 °C¹⁴⁷, that is remarkable for a 12 residue miniprotein, and the desired fold confirmed by NMR. Other early examples of *de novo* β -hairpin¹⁴⁸ and triple stranded β sheet¹⁴⁹ were reported.

The DS119 was the first $\beta\alpha\beta$ motif successfully designed *de novo*. As the dependence on the length of the secondary structure was described^{150,151}, the design of the backbone was performed by analysing the length of the helices and strands from α/β protein structures. The analysis concluded that for a 5 residues β -strand, a 12-residue helix was appropriate. The backbone was built using secondary structure geometry constraints. For the sequence design, the established *de novo* design rules were applied. A hydrophobic core was created¹⁵² by linking an amphiphilic α -helix¹⁵³ to a β -sheet fragment. The helix contained leucine and alanine residues¹⁵⁴ on the hydrophobic side, and glutamic acid and lysine¹⁵⁵ residues, forming salt bridges, on the solvent exposed region. This hydrophobic core was composed of isoleucine and valine residues¹⁵⁶. The initial design was finalised with a TPEE N-terminus cap¹⁵⁷ for the helix. The resulting miniprotein design showed poor tertiary fold and aggregation by CD. The nine-position core was computationally reconstructed, and the desired stable $\beta\alpha\beta$ motif was achieved by introducing a Trp zipper into the β -sheet. To prevent aggregation, lysine and arginine residues were strategically placed in noncritical regions. Their positively charged nature and long hydrophobic side chains were utilised to cover the hydrophobic core and mitigate potential aggregation. The resulting sequence was resolved by NMR, (Figure 12C), which confirms the successful design of a $\beta\alpha\beta$ motif.

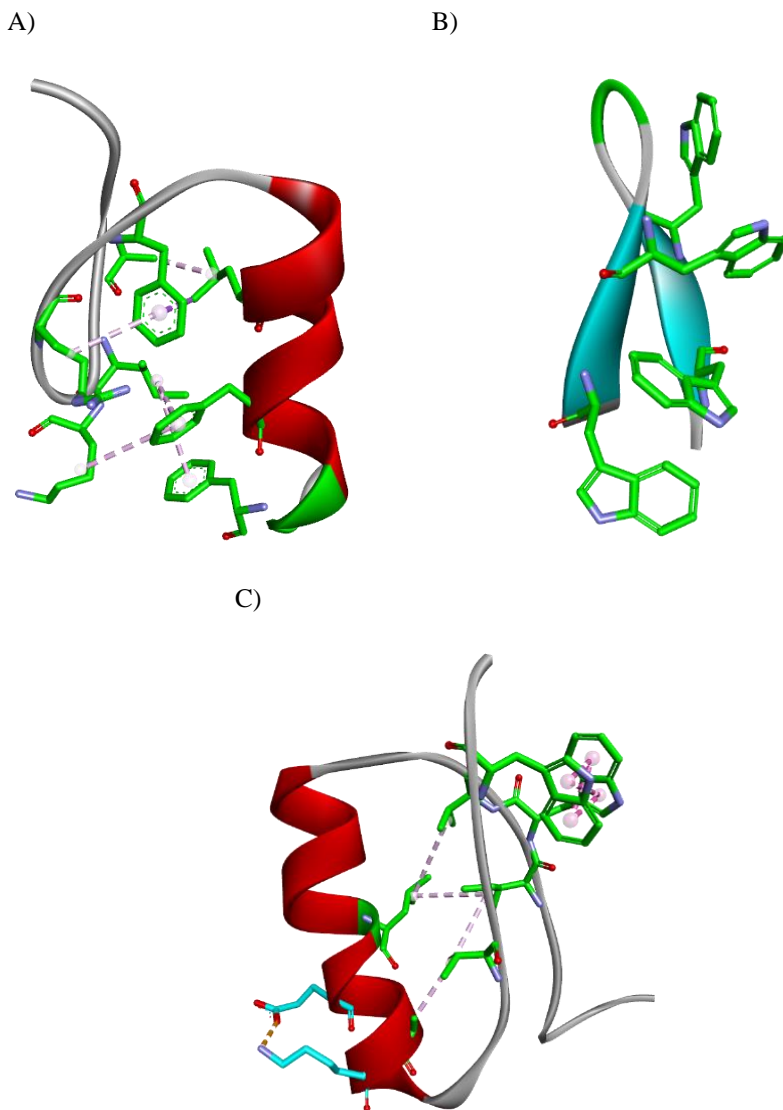
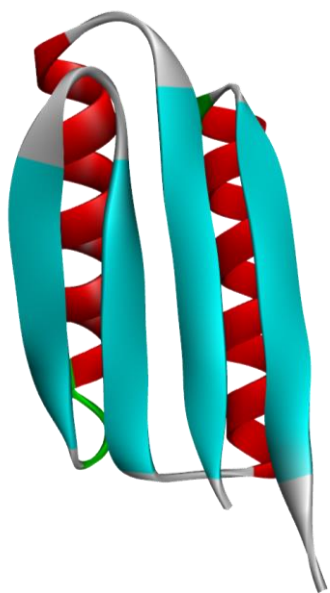


Figure 12. *De novo* designed miniproteins A) FSD miniprotein (PDB id: 1FSD), B) Trp Zipper (PDB id: 1LE1), and C) DS119 miniprotein (PDB id: 2KI0), Critical interacting residues for folding are shown as sticks in green and key folding mutations are highlighted in cyan. The most significant interactions are indicated by dashed lines.

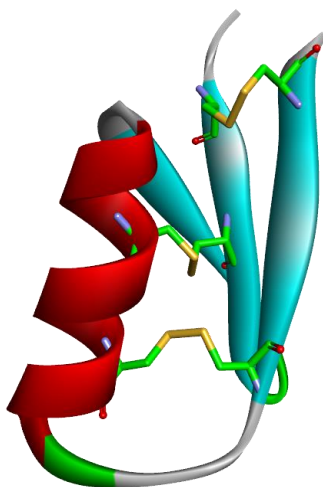
In 2012, Baker and co-workers introduced several new principles for the precise *de novo* design of proteins¹⁵⁸. It was found a set of rules that defined the connectivity between the secondary structures of α/β proteins, with variations in the length of the secondary structure. The explored connectivity was $\alpha\beta$, $\beta\alpha$ and $\beta\beta$. The $\alpha\beta$ -rule states that this motif had a parallel preference, where the side-chain orientation of the last residue of the strand has the same orientation as the vector helix-strand, independently of the loop length. However, longer loops than two residues are required to provide a hydrogen bonds to the helix without elongation of the strand. The $\beta\alpha$ -rule shows that this motif prefers the orientation based on the loop length and the side-chain orientation of the last residue of the strand. While the $\beta\beta$ -rule reflected a direct relationship between chirality and loop length, being left-handed with 2-3 residue loops, no apparent preference for 4 residues and right-handed for 5 residues. The rules on secondary structure connectivity were used as the bases of the funnel-shaped energy landscape design or negative design. By applying the above-mentioned rules, it is possible to incline the fold towards the desired topology by generating an energy gap with undesired folds. With the support of RosettaDesign^{140,159} algorithm, a set of sequences can be generated that satisfy the previously stated *de novo* design rules. To test the rules, a set of five complex topologies with adequate secondary structure length was generated. For each backbone, thousands of Rosetta fold simulations were performed, leading to a highly populated correct conformation. Following an iterative process of folding and redesign of the sequences, coupled with Rosetta all-atom relaxation¹⁶⁰, they scored and filtered the sequences by low energy and packing quality¹⁶¹. The resulting sequences were then folded hundreds of thousands of times with Rosetta ab initio³⁵ leading to a 10 % of sequences with funnel-shaped energy landscape. Compared to previous *de novo* designs¹⁴⁰, they could confirm that the newly defined rules were also satisfied. By expressing the sequences and filtering them by solubility, CD profile, thermal stability, and oligomerization state, it was possible to resolve at least one of each design, using NMR, (Figure 13A), what confirmed the success of this approach.

With an improved backbone design algorithm, so-called fragment assembling, CRMPs have been also *de novo* designed¹⁶². Initially, a blueprint of the scaffold, containing length and connectivity of secondary structures, was described. With Monte Carlo simulations, fragments obtained from databases of crystal structures are assembled to satisfy the blueprint^{163,164}. Prior to sequence design, the scaffolds were screened to select positions for the disulphide bridges, by constraining the geometry to ideal bonds. By means of iterative design and minimisation, the sequences were generated and characterised by CD, HPLC, NMR, and X-ray crystallography, confirming the atomic accuracy of the designs, (Figure 13B). Furthermore, designing heterochiral CRMPs was also successful, (Figure 13C), extending the Rosetta energy function to support D-amino acids and cyclic miniproteins, (Figure 13D), by implementing a generalised kinematic loop closure method^{165,166} GenKIC. Moreover, a GenKIC-based *ab initio* structure prediction protocol allowed the design of a heterochiral cyclized combination of a left- and a right-handed helix (the structure is not available).

A)



B)



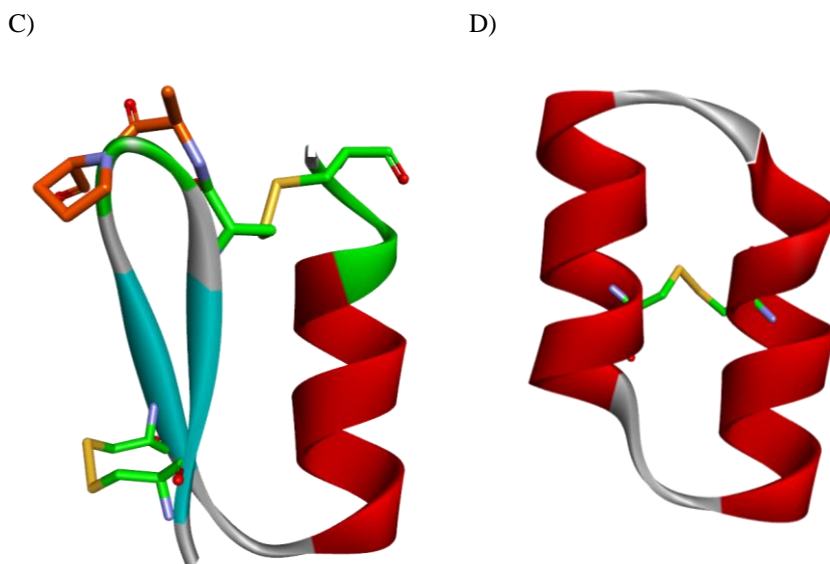


Figure 13. *De novo* mini-proteins A) α/β mini-protein (PDB id: 2KL8), B) trisulfide mini-protein (PDB id: 5JI4), C) Heterochiral mini-protein (PDB id: 5KWP), and D) Cyclic mini-protein (PDB id: 5KWZ), Residues crucial for the folding are shown as sticks in green, and key mutations for folding are highlighted in orange.

1.3.4. Backbone alteration

With advances in chemical synthesis of peptides¹⁶⁷, the protein fold was expanded. The synthetic approaches allowed for the incorporation of noncanonical amino acids into the peptide sequences. New building blocks derived on new backbone conformations, spatial arrangements, and side chain diversity, expanding the possibility to design active molecules. An important concept was introduced under the name foldamers¹⁶⁸, which was a comprehensive way of describing any polymer with the ability to fold in solution into well-defined structures. Since the incorporation of noncanonical amino acids into the peptide sequences, a great number of studies have focused on the formation of isolated secondary structures with new shapes^{169,170}. These new building blocks, in particular β -amino acids, have been well-known since the early stages of this research^{171,172}. The folding propensities of α - β peptides, have been profoundly researched over the past decades, deciphering the stereochemical patterning¹⁷³⁻¹⁷⁷, used to predict the secondary structure formed by different arrangements of α and β amino acids. For the scope of this doctoral dissertation, we will focus on the construction of higher-order structural

mimicry with β -amino acids, with emphasis on β -constrained amino acids. The early works on heterogeneous tertiary constructions¹⁷⁸ were successful, however, precluded to the connectivity^{179–181} or assembling^{182,183} of foldameric helices. Protein backbone engineering¹⁸⁴ was the next step in the development of complex folds incorporating non-canonical amino acids.

The engineering intended to modify residues of an existing miniproteins by non-canonical ones to evaluate their impact on stability, conformation, and function. The impact of these modifications has been explored in several systems. Villin headpiece was single point modified in four solvent exposed positions of the third helix, (Figure 14A)¹⁸⁵. Modifications were carried out with β^3 -amino acids, homologous to the substituted, *trans*-(1*S*,2*S*)-2-aminocyclopentanecarboxylic acid (*trans*-ACPC) and *trans*-(3*R*,4*S*)-4-aminopyrrolidine-3-carboxylic acid (*trans*-APC). It was concluded that β -constrained amino acid substitutions better sustained the native conformation. Another example of Villin headpiece β -constrained amino acid substitution explored the possibility to generate a new sequence which will preserve the native fold¹⁸⁶. The design was performed by *de novo* approaches such as Rosetta FastDesign¹⁵⁹, from which a total of five sequences were synthesised and four showed a cooperative fold by CD. The crystal structure revealed a novel fold that preserved the three characteristic helices of the Villin headpiece, (Figure 14B), and the formation of a quaternary structure in solution and crystal by dimerisation through the foldameric helix.

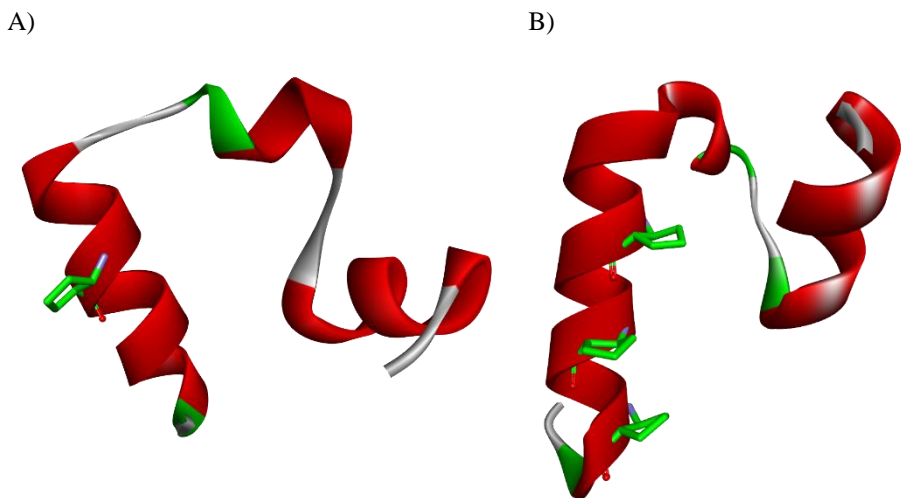


Figure 14 Villin headpiece backbone alteration A) Single backbone modifications of the Villin headpiece (PDB id: 5I1O), and B) Engineered Villin headpiece (PDB id: 7AAR). The *trans*-ACPC positions in the sequence are highlighted in green.

β -Substitutions in hairpins¹⁸⁷ and loops of the Pin1 WW-domain¹⁸⁸ have also been explored (Figure 15A). Replacements of loop 1 positions in the Pin1 WW domain and two variants with modified loop 1 length were performed. The results illustrated that loop modifications into β -constrained amino acids had a strong dependence on the stereochemistry of the *trans*-ACPC. Moreover, while the tendency of these substitutions is to retain or decrease the stability of the native structure, some of the replacements had a significantly enhanced the thermal stability. The possibility of replacing loop positions, in existing or *de novo* miniproteins, opens the gates to modify other regions than secondary structures to diminish proteolytic susceptibility without influencing the overall fold and stability.

The impact of β -substitutions in more complex systems was explored in the previously described miniproteins¹⁸⁹, the Trp-cage¹³⁰ and the FSD-1¹⁴² miniprotein. The initial strategy consisted of a β -scan, where all possible positions, of the corresponding helices, were single-point mutated to *trans*-ACPC. For the Trp cage, substitutions on the N-terminus had minimal effects on the helical conformation while in central positions the effects were more pronounced. As expected, substitutions of crucial residues for the fold led to loss of conformational stability. However, the addition of *trans*-ACPC to the N-terminus of the helix significantly improved stability by 10 °C. Considering that the CD

spectrum was close to the native, it was assumed that *trans*-ACPC in terminal positions, promotes structuration of the helix. A similar effect was observed for FSD; however, due to the reversed topology, the enhancement was observed when the C-terminus was modified. With the knowledge acquired, a stereochemical pattern of $\alpha\alpha\beta\alpha\alpha\beta$ led to a series of mutants for which one had the ability to adopt the native fold and preserve stability.

Most of the work on the engineering of protein backbones has been led by Horne and co-workers¹⁹⁰. Their efforts were focused on understanding to which extent a native protein could withstand backbone alterations and the impact of the modifications on the thermodynamic properties of folding. By systematically modifying sections of the GB1 miniprotein, effects on stability and fold were explored¹⁹¹. For the helix, β^3 (Figure 15B); β^2 ; β^{cyc} Aib and C α -Me- α substitutions were implemented. While for β -sheet, N-Me α -amino acids; $\beta^{2,3}$ and γ^{cyc} substitutions were explored. On the turns D-amino acids; C α -Me- α (Aib) and d² (Orn). As for the loops PEG; β^3 ; and β^{cyc} substitutions. With their understanding of the effects on stability and conformation, they were able to synthesise a GB1 mutant with 20 % backbone alteration¹⁹². This mutant could retain, if not improve, the native stability and conformation of GB1. Furthermore, their work has shed light on the design of heterogeneous beta-sheet motifs¹⁸⁷ or zinc fingers¹⁹³ that retain activity. Regarding the zinc finger, a work worth mentioning was performed by Guichard and co-workers. In their work, they completely substituted the helical fragment of Zif268 with an oligourea mimetic¹⁹⁴, which showed native fold and activity.

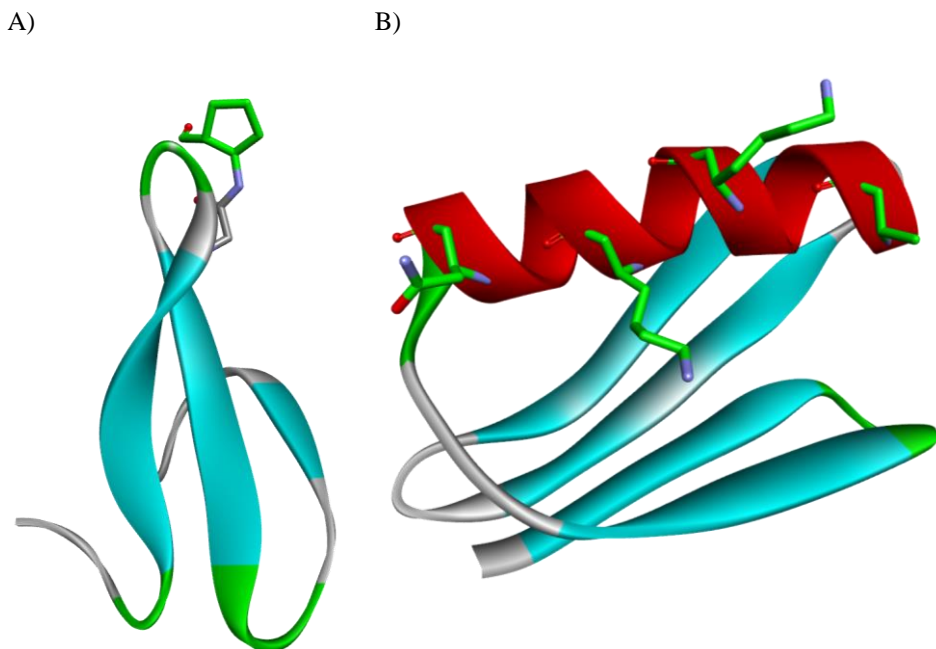


Figure 15. Backbone alteration, A) WW-domain with modified loops (PDB id: 5VTI), and B) GB1 with β^3 substitutions on the helix (PDB id: 4KGR). The altered positions of the backbone are highlighted in green.

The engineering of protein domains has provided valuable insights into backbone alterations. We now have a better understanding how noncanonical amino acids can be accommodated into complex folds. However, the complete design of miniproteins bearing heterogeneous backbones remains unexplored, requiring the development of a proper approach. Additionally, the use of noncanonical amino acids as tertiary structure stabilisers, such as hydrophobic core packing, has received little to no exploration.

1.4. Biological activity modulation

To succeed in the design of an active miniprotein, two factors must be effectively combined, precise fold and interactive side-chains. Hence, the selection of an adequate scaffold is crucial for the optimal activity of the compound. Scaffolds can be either found in nature or designed computationally. After selecting or designing the scaffold, the specific activity should be incorporated into the sequence. To date, three major approaches can be used. Direct application of active miniproteins found in nature^{195–197}. Grafting the functional groups of existing active motifs on a suitable scaffold^{198,199}, where the side chain orientation will prevail. Computational design of a new active motif on the miniprotein surface^{200–202}, or the design of entirely new active miniproteins.

1.4.1. Native and modulated activity

Naturally occurring miniproteins have attracted significant attention from the scientific community because of their intrinsic capacity to modulate biological processes. More specifically, active CRMPs extracted from plants²⁰³ (cyclotides) or venoms²⁰⁴ (knottins).

Kalata B1 was the first bioactive natural compound identified. It was used as a childbirth accelerator in Africa due to its uterotonic activity²⁰⁵. It is extracted from an *Oldenlandia affinis* herbal tea. Since then, cyclotides have been isolated from the *Cucurbitaceae*, *Fabaceae*, *Solanaceae*, and *Apocynaceae* families, in addition to *Rubiaceae* and *Violaceae*²⁰⁶. Moreover, they have proven to be effective in a broad range of biological activities such as antimicrobial²⁰⁷, anticancer²⁰⁸ or anti-HIV²⁰⁹ amongst others. Additionally, their physicochemical properties, such as cell intake²¹⁰ and hyperstability¹¹⁴ grant them great potential for therapeutic development.

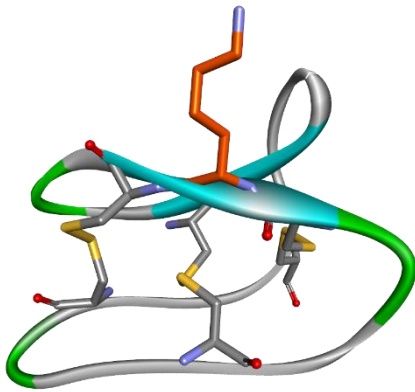
Kalata B1 has shown antiproliferative effects on activated primary human lymphocytes. By using the crude extract of the *Oldenlandia affinis*, flow cytometric-based forward-side-scatter analysis was performed to study antiproliferative effects of the Kalata B1²¹¹. The experiment showed a reduction of peripheral blood mononuclear cells (PBMC) proliferation in the presence of Kalata B1 with an IC₅₀ of 3.9 μM. It was described that, as long as Kalata B1 concentration was below 14 μM, the antiproliferation was derived from

a cytostatic effect (slow or growth stop) rather than apoptosis or necrosis. This was consistent with the reported capacity of Kalata B1 to cause membrane disruption and hemolysis at concentrations above $\sim 50 \mu\text{M}$ ^{212,213}.

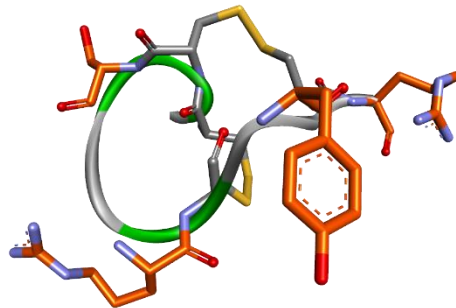
An example of activity modulation of native miniprotein is the Kalata B1 mutant T20K. The antiproliferation activity of Kalata B1 can be enhanced by modification of residues from the region interacting with the target. T20K, (Figure 16A), was able to inhibit interleukin-2 (IL2) biology in T-cells²¹⁴, with an IC_{50} of $1.9 \mu\text{M}$. Oral administration of T20K in mouse model of experimental autoimmune encephalomyelitis²¹⁵ (EAE) showed remarkable results for the treatment of multiple sclerosis²¹⁶ and the compound is currently in Phase I clinical trials¹⁹⁶. Due to the constrained nature of CRMPs, oral administration of these drugs is possible. Moreover, the cyclotide structure can maintain the appropriate conformation, enabling it to penetrate cell membranes effectively. The significance of two surfaces of Mobius cyclotides, such as kalata B1, has been recognised as crucial in facilitating membrane permeability. Specifically, the "bioactive face" and "hydrophobic face" play an essential role in binding to phosphatidylethanolamine phospholipids and cell membrane penetration²¹⁷. Such knowledge has allowed to successfully modify biologically active miniproteins that lacked cell permeability. Moreover, there are useful strategies that leverage naturally occurring mechanisms for delivery. Among these, a bacterial machine consisting of components from anthrax toxin-protective antigen²¹⁸ (PA) and lethal factor (LF)²¹⁹ has been demonstrated to be an excellent approach to effectively transporting commonly used miniprotein-based antibody mimics.

The KPC-4000 is one example of activity modulation of a naturally occurring venom knottin. KPC-400¹⁹⁷ is an analgesic drug derived from the α -conotoxin RgIA from *Conus regius*, (Figure 16B). The native RgIA was identified as a potential analgesic. However, its capacity to block human $\alpha 9\alpha 10$ nicotinic acetylcholine receptor (nAChR) was 4200-fold weaker than that of mice, IC_{50} 10000 nM and 2.4 nM, respectively. The screening of mutants with modifications that kept the polar-nonpolar character of the native residues led to RgI4 (KPC-400). This mutant showed a ~ 7000 -fold enhancement of the affinity for $\alpha 9\alpha 10$ -nAChRs with an IC_{50} of 1.5 nM and high selectivity. Currently, the mutant KPC506, is in phase I clinical trials²²⁰.

A)



B)



C)



D)



Figure 16. A) Structure of T20K immunomodulator (PDB id: 7HLC), highlighted in orange mutated residue to enhance activity B) α -conotoxin RgIA from *Conus regius* (PDB id: 2JUT), modified residues to generate KPC-400 miniprotein are highlighted in orange., C) Sequence representation of Kalata B1 and modified T20K, the modification is highlighted in orange and the disulfide bridges in yellow, and D) Sequence representation of RgIa and modified KPC-400, the modifications where Cit = Citrulline; 3-I-Tyr = 3-iodo-Tyrosine are highlighted in orange, and the disulfide bridges in yellow.

1.4.2. Molecular grafting

For molecular grafting to succeed, it is extremely important to select a robust scaffold that will be able to accommodate the active epitope while retaining the fold. Following these requirements, CRMPs have become the main source of scaffolds for molecular grafting^{221–223}. This approach has proven to be successful in a wide range of targets, from cell membrane²²⁴, intracellular targets²²⁵, cell surface receptors²²⁶ to cell extracellular targets²²⁷.

MOG3¹⁹⁸, a Kalata B1 grafted with fragment 41–47, (Figure 17A), of the myelin oligodendrocyte glycoprotein (MOG), has shown positive results for the treatment of multiple sclerosis in mouse models. The initiative to investigate this protein emerged from the identification of MOG as a potential drug for the treatment for multiple sclerosis. However, its capacity to provide structural integrity to the myelin sheath is jeopardised by its low stability and bioavailability. The approach tested several grafts on loops 5 and 6 of Kalata B1 of fragments from the MOG epitope, residues 35–55. The resulting miniprotein, MOG3, may be potentially used for the treatment of multiple sclerosis.

The helical GCN4 leucine zipper is an example of a linear scaffold used for grafting the binding epitope of the HIV C-peptide, C34, (Figure 17B). Regardless of a sequence modification of approximately 45%, the obtained miniprotein C34coil had potent antiviral activity competing with HIV-1. The grafted molecule showed IC₅₀ 16 nM by viral infectivity assay and IC₅₀ 3 nM in cell-cell fusion assay¹⁹⁹. Moreover, the C34coil had high conformational stability and increased proteolytic resistance.

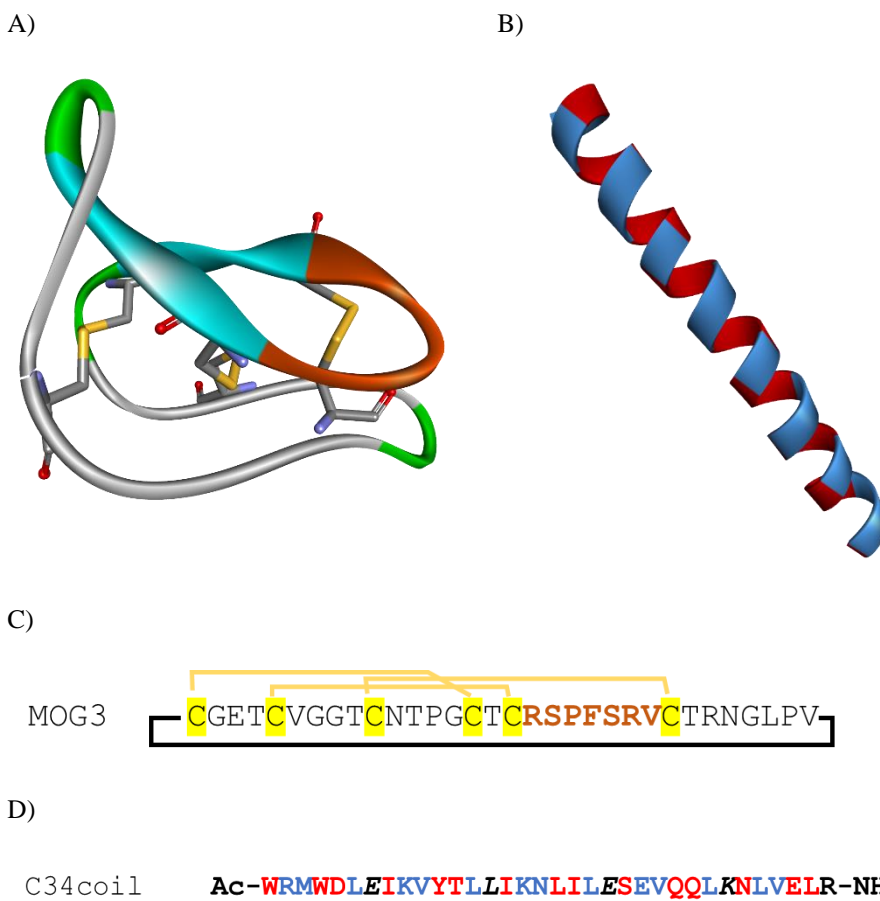


Figure 17. A) Structure of Kalata B1 (PDB id: 1NB1), the grafted loop is highlighted in orange, B) GCN4 leucine zipper (PDB id: 2ZTA) modified positions are highlighted in red and conserved ones in blue, C) Sequence representation of MOG3 with the grafted loop of Kalata B1 highlighted in orange and disulfide bridges in yellow, and D) Sequence representation of the C34coil helix where in blue are the conserved residues from the Leucine zipper and in red the insertion points of the interactive region of C34.

Furthermore, it is worth mentioning that hybrid molecular grafting has been also successful when non-sequential epitopes have been placed on *de novo* designed scaffolds. This approach is of great interest for binding complex targets where the structure plays an important role on the interaction. Particullary for targets considered undruggable, as the epitope has a discontinuous character due to the large interactive surface. MOPD is an inhibitor of the PD-1/PD-L1 immune checkpoint²²⁸. The scaffold used for the inhibitor design was selected from a previously published *de novo* design of Baker and coworkers¹⁶²,

(Figure 18A). After superimposition of the scaffold, over the optimised PD-1 crystal structure in complex with PD-L1²²⁹, the residues of the scaffold were modified to the overlapped interactive residues of PD-1. The resulting mimetic had a K_D 300 nM affinity for PD-L1, 28-fold higher than the affinity of native PD-1, and an IC_{50} of 4.84 μ M. Furthermore, it showed activity *in vivo* in mouse models and hyperstability in serum. Due to the constrained conformations, the modifications had nearly to no impact on the tertiary fold, as confirmed by X-ray crystallography, (Figure 18B).

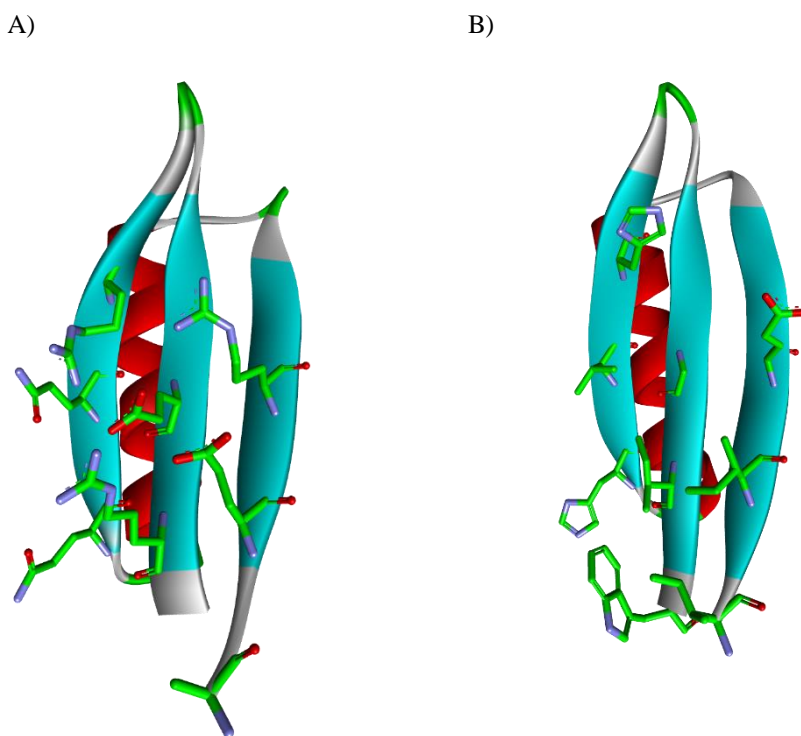


Figure 18. Crystal structures of A) *De novo* designed scaffold (PDB id: 5JG9), and B) The L47W mutant of MOPD inhibitor (PDB id: 7RJF). The interactive residues are highlighted in green before (A) and after (B) modification to enhance affinity for PD-L1.

1.4.3. *De novo* activity design

The *de novo* design of miniproteins has also proven to be a resourceful tool to obtain active molecules²³⁰. The design allows us to not only select an optimal distribution of the interactive residues but also to simultaneously sample the overall conformation. The *de novo* design of active miniproteins has opened the doors to address more complex targets such as PPI mediated by flat surfaces.

Applying massively parallel *de novo* protein design, Baker and co-workers designed highly active miniproteins against influenza. This approach combines techniques such as large-scale computational design and yeast display screening. With the massive production and screening, over 22000 sequences, 2618 high-affinity binders, against influenza haemagglutinin and botulinum neurotoxin B, were obtained²⁰⁰. The designs evoked minimal to no immune response when administered nasally to mice. This massive approach showed remarkable results in the design of therapeutics against influenza.

LCB1⁷¹ was another success in the *de novo* design of active miniproteins. In this work, inhibitors of the SARS-CoV-2 spike receptor binding protein (RBD) that compete with ACE2 protein were designed. The problem was approached in two ways, one hybrid way as described in the previous section, where the epitope of the ACE2 was incorporated into a *de novo* designed scaffold. A second approach, a *de novo* design of a miniprotein with new binding mode, (Figure 19A). For the first approach, standard *de novo* protocols were implemented. However, for the second approach, a rotamer interaction field docking was used. AHB2 resulting from the first approach, had an IC₅₀ of 15.5 nM and T_m of 82.3 °C, while for LCB1, obtained with the second approach had an IC₅₀ of 23.54 pM with T_m higher than 95 °C. These results probed the potential of miniproteins to surpass any other active compound in terms of stability bioavailability and affinity.

De novo design coupled with grafting went beyond inhibition by designing a PD-1 agonist. The work identified a five-edge residue of hPD-L1 (ADYKR), mPD-L2 (WDYKY) with the largest contribution to the binding energy to PD-1. These two patterns and a hybrid WDYKR were used to form a strand in more than 34.000 *de novo* disulphide stabilised scaffolds. The assessment of the activity of the initial designs, (Figure 19B), reflected micromolar affinities for the target. Combining and extension with a redesign of one of the loops led to PD-MP1, with 100 nM affinity for both human and murine PD-1.

The trimeric form of the miniprotein showed potential for the treatment of autoimmune and inflammatory diseases. The authors required to include cysteine bridges to increase the funnel-shaped energy landscape of the design, to target a complex PPI as is the PD-1/PD-L1 where dominant forces are flat large hydrophobic surfaces. This is due to the loss of hydrophobic/hydrophilic patterning because of the core and the solvent exposed region of the β -sheet are hydrophobic. Yet again.

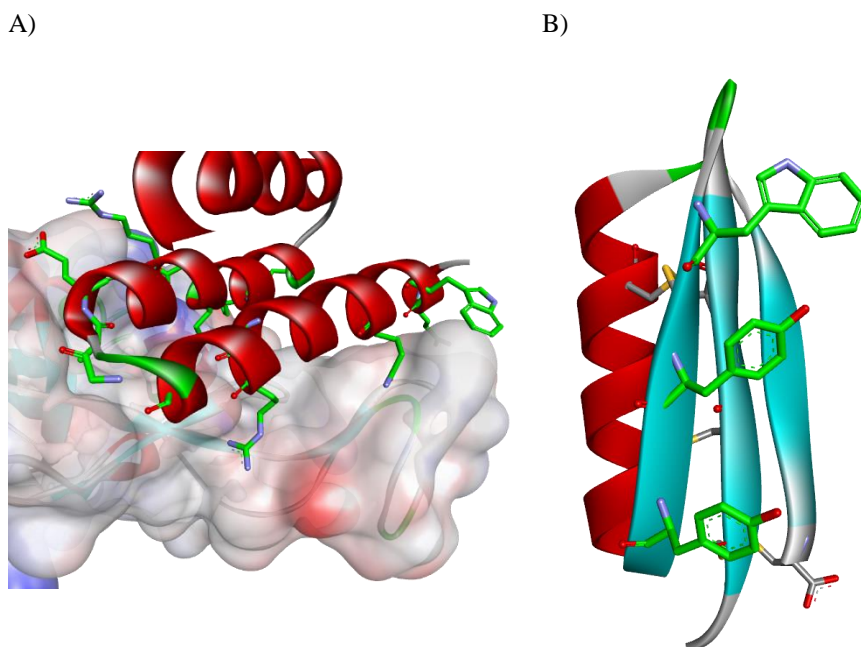


Figure 19. *De novo* designed protein binders, A) Complex of LCB1 (red ribbon) and SARS-CoV-2 spike protein RBD (surface display) (PDB id: 7JZU), and B) PD-1 agonist GR-918.2 (PDB id: 6V67). The designed interface for the effective binding to the target proteins is highlighted in green.

2. Goals

In the present dissertation, we set out to investigate three major aims on the design of active β -amino acid-containing miniproteins.

The first aim was to develop a methodology for the *de novo* design of β -amino acid-containing miniproteins with various topologies. The new sequences, combining different secondary elements were expected to fold into well-defined structures with native-like thermodynamic properties. Additionally, a novelty of the design, was to incorporate β -amino acid residues in core positions to contribute to the tertiary fold.

The second aim was to design new binders based on the WW-Prototype, to PD-L1, with new binding modes. By optimising computationally, the solvent-exposed interface of the domain, we intended to generate low-energy mutants in complex with the target protein. The high affinity binders were grafted into the scaffolds generated during the first aim.

The third aim was to design new inhibitors of the PD-1/PD-L1 immune checkpoint using the miniprotein with best thermodynamic properties obtained on the first part of the study. The aim of the work was to prove that the accommodation of a high number of modifications in a sequence requires an accurate stabilisation of the fold. Moreover, we explored the effects of physicochemical properties of the inhibitors, e.g., net charges, on the binding affinity to the target.

Successful implementation of these three aims would open the possibility to accurately design new and complex folds. It would also open the doors to targets considered undruggable by small molecules. Furthermore, we would be able to access molecules with protein-like properties not found in nature.

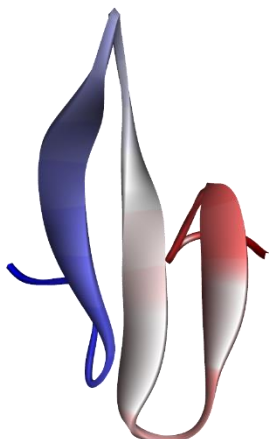
3. Results and discussion

3.1. β -amino acid-containing miniproteins

3.1.1. From EEE to HEEE topology by fragment assembling

Initially, we explored the possibility of generating an α/β tertiary structure that could later be used as scaffold for the design of PD-1/PD-L1 immune checkpoint inhibitors. Due to the extended nature of the PD-1 interacting region, the desired new miniprotein was aimed to contain a β -sheet. Furthermore, to accommodate modifications without the loss of the tertiary fold, we hypothesised that the incorporation of a helix could potentially stabilise the β -sheet. Furthermore, a β -amino acid-containing helix could enhance the proteolytic stability of the miniprotein. To generate the initial backbone two independent subunits were assembled, a β -amino acid-containing helix and the WW-prototype¹²² (**1**), a triple-stranded antiparallel β -sheet, (Figure 20A). Given the folded nature of these two subunits, we hypothesised that the adequate optimisation of a hydrophobic core will stabilise a tertiary fold. On the one hand, the selection of a WW-domain, as the initial template for the β -sheet, was based on, a) their capacity to withstand mutations without loss of its tertiary fold, b) the existence of high-resolution structures in the PDB, c) adequate length for solid peptide synthesis, d) potential surface for the later development of PD-1/PD-L1 interaction inhibitors. On the other hand, combining the need to introduce backbone modifications and that the vast majority of foldamer research has been focused to mimic α -helical fold, a fragment of the gp41 subunit mimetic²³¹, (Figure 20B), was used to assemble to miniprotein **1**. Several unnatural amino acids have been shown to induce the folding of the secondary structure, with a special emphasis on constrained β -amino acids^{169,170}, namely (1*S*,2*S*)-aminocyclopentanecarboxylic acid (*trans*-ACPC), which has been implemented in the present work. *Trans*-ACPC has been extensively studied and its capacity to form helical mimics is unquestionable. For the fragment assembling, a stereochemical patterning of $\alpha\beta\alpha\alpha\beta\alpha\beta$ was used. Due to the extension of the backbone with the C_β of the *trans*-ACPC, this pattern allows placing all β -amino acids on the same side of the helix closely reproducing the native α -helix turn ($i-i+3$)²³².

A)



B)

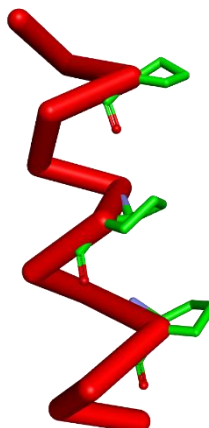
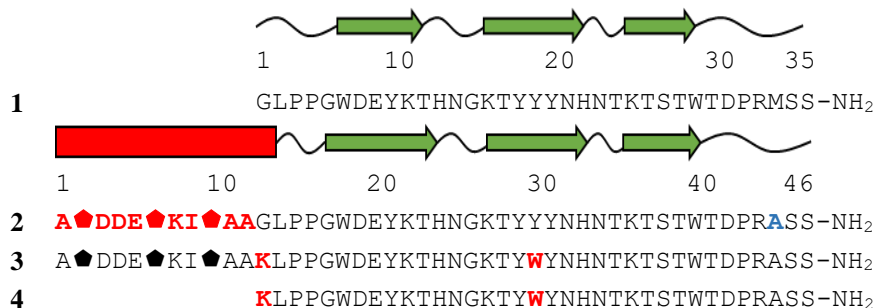


Figure 20. Ribbon representation of the crystal structure of A) the WW-Prototype (PDB: 1E0M), and B) Fragment of the gp41 subunit mimetic residues 13-24 (PDB: 3F50). The WW-Prototype is coloured from the N-terminus (blue) to C-terminus (red), while the *trans* ACPC is highlighted in green.

The model miniprotein **2** generated by fragment assembling was processed with the Rosetta FastDesign protocol to generate mutants that will fold through a hydrophobic core. Only residues with the potential to generate interactions between the β -sheet and the helix were mutated. However, prior to Rosetta and because the design intended to place the *trans*-acpc into the hydrophobic core, Lys12 was introduced to promote the adequate orientation of the helix. Additionally, Trp29 residue was incorporated to increase contacts and as a basis for the generation of the hydrophobic core. Together with miniproteins **1-3**, miniprotein **4**, an EEE analogue of **3**, was synthesised to study the initial effects of core optimisation, (Table 1).

Table 1. Entry, topology, and sequences of miniproteins **1-4**. Highlighted in red are the modifications incorporated and in blue the Met substitution to facilitate synthesis. (● = *trans*-ACPC)



The structural study of miniproteins by circular dichroism (CD) is crucial for the assessment of secondary structures present in solution²³³. CD spectra of several secondary structures, α -helix²³⁴, β -sheets with different degrees of torsion²³⁵, and β -amino acid-containing helices¹⁸⁶, have been widely studied and described. For the case study, it is characteristic of WW-domains to reflect a maximum at 230 nm, originated from the aromatic contributions to the fold. Additionally, there are two minima one at 217 nm and another at 204 nm, corresponding to the β -sheet fold. However, the signal at 217 nm could be diminished due to the predominant aromatic side chain-backbone interactions. CD spectra of α/β shows minima at 209 nm.

The CD spectra of **2** and **3** are consistent with the presence of both the WW-domain and the α/β helix, (Figure 21). The slight shift of the minima by ~ 2 nm may indicate that the introduction of Lys12 enhances the helical character in **3** compared to **2**. Moreover, the reduced ellipticity of the aromatic contributions may suggest a globular packing in the absence of cooperative stability. This observation is confirmed by the absence of any thermal stability enhancement between the HEEE miniproteins and their EEE analogues. However, incorporation of Trp29, enhances the stability of the β -sheet by 5 °C, when comparing **3** to **1**. Yet, Trp29 induces conformational changes in the β -sheet, with lower ellipticity at wavelengths below 220 nm. The structural variation is impossible to accurately assess in the absence of a high-resolution structure.

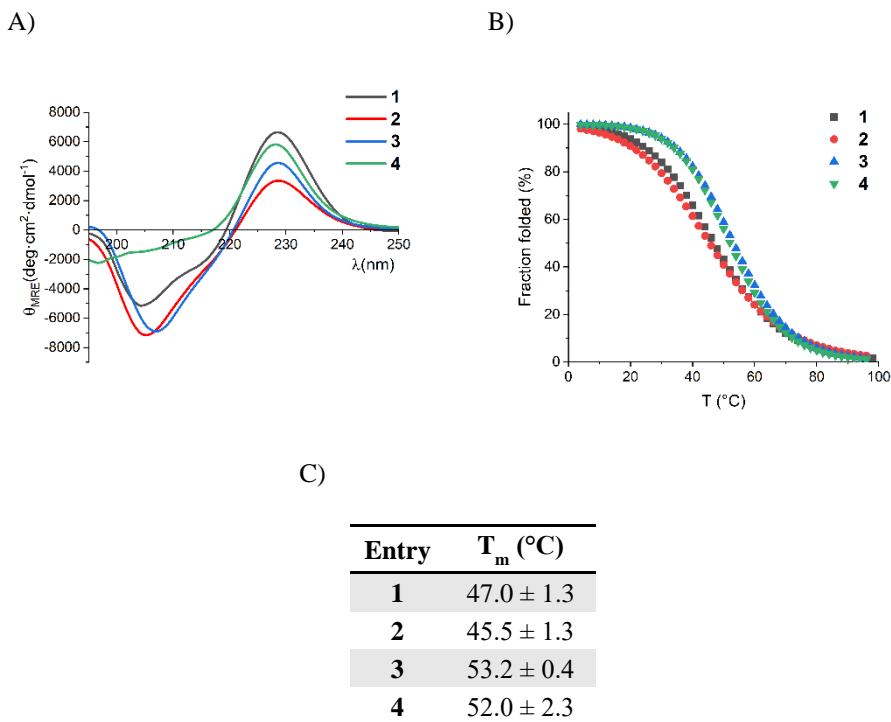


Figure 21. A) CD spectra, B) Normalised thermal denaturation, and C) Estimated T_m values of miniproteins **1-4**.

Miniprotein **3** was submitted for small angle X-ray scattering (SAXS), in the synchrotron in Hamburg. The experiments were carried out at two different concentrations of the miniprotein, 2 and 5 mg/ml, both in 50 mM potassium phosphate buffer pH 7.5. The experimental mass, of 5.6 KDa, versus the theoretical mass, 5.3 KDa, indicated a monomeric oligomerization state. Gnom real space distribution reflected a globular miniprotein with peak at 17 Å and d_{max} of 54 Å. However, the miniprotein was in a monomeric state and the radius of gyration (R_g) was in agreement with the observed from the theoretical model, and the Kratky plot suggested a partially folded state. This result validated the hypothesis of an uncooperative globular fold observed by CD. The electron density map generated from the experimental data, processed with DENSS (Figure 22), shows the potential movement of the helix between the folded state and the solvent.

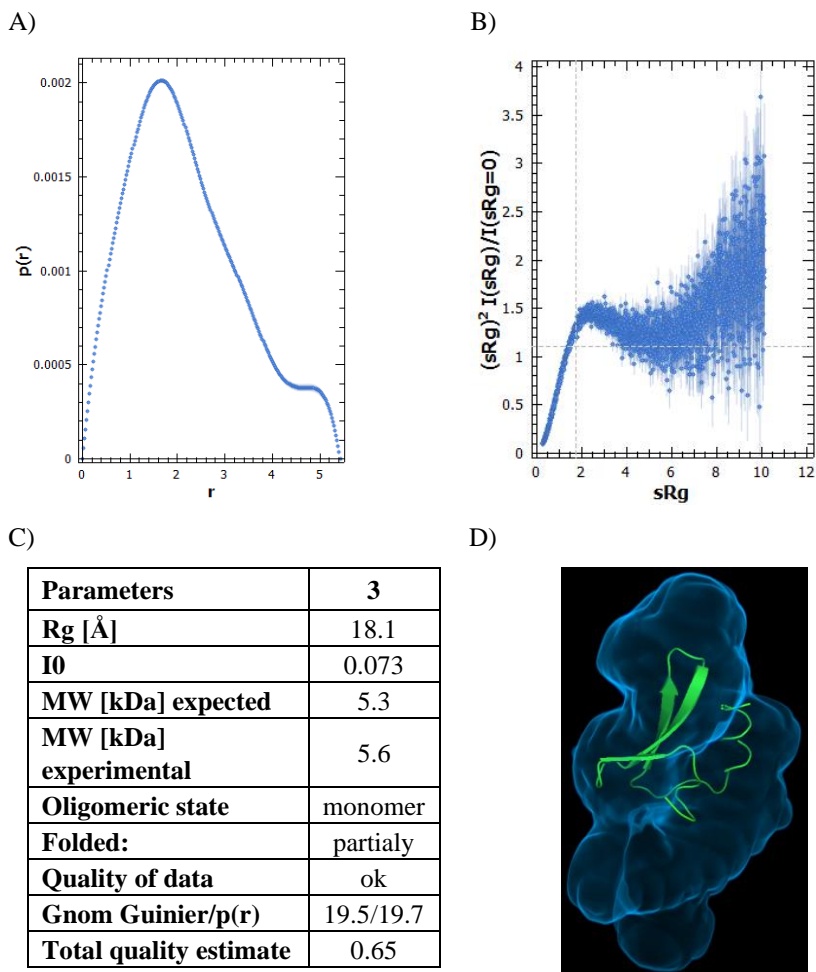
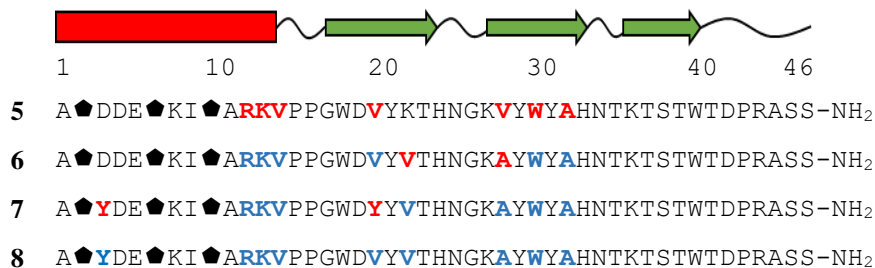


Figure 22. SAXS experimental results, processed with ATSAS software, of miniprotien **3**, A) gnom real space distribution, B) Kratky plot, C) Experimentally determined parameters, and D) electron density map of **3** generated with DENSS from the experimental data.

The in-depth study of these initial mutants was followed by the synthesis of the lowest energy scored designs of Rosetta FastDesign, leading to miniproteins **5-8**, (Table 2). The mutations were placed on the residues of the first and second strands of the β -sheet facing the hydrophobic core and the newly generated connection. The systematic mutation done in Rosetta intended to increase the number of hydrophobic contacts between the β -sheet and the three *trans*-ACPC residues of the helix. The modification of the connection was intended to increase the helical fold and the contacts with the β -sheet.

Table 2. Sequences of miniproteins **5-8**. Modifications already explored are highlighted in blue and new mutations are highlighted in red.



Mutations in miniprotein **3** had drastic effects on CD spectra, (Figure 23). The ellipticity of the aromatic contributions dropped to zero for mutants **5** and **6**. These contributions are partially recovered after insertion of Tyr19 for miniprotein **7**, and Tyr3 for miniproteins **7** and **8**. Despite the change in CD spectra, it is worth highlighting that the notably larger molar ellipticity at the α/β helix region may be indicative of greater helical character. Interestingly, miniprotein **8**, the only miniprotein with slightly more pronounced lower maxima at 215 nm (characteristic of β -sheet fold), had a small T_m of 26.1 °C.

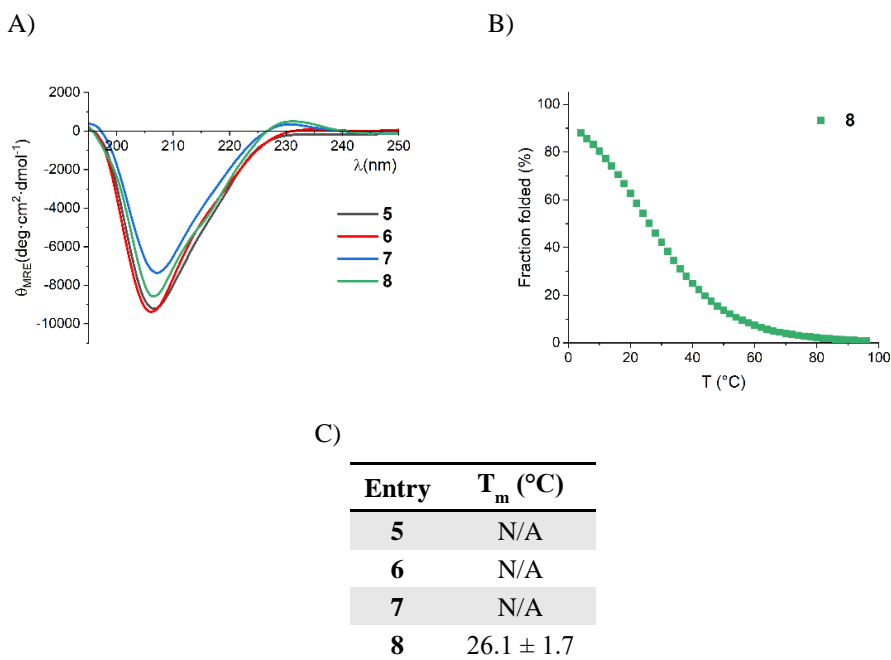
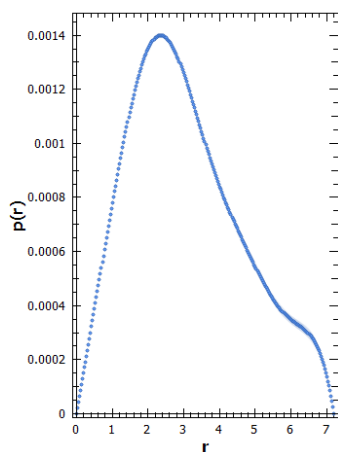


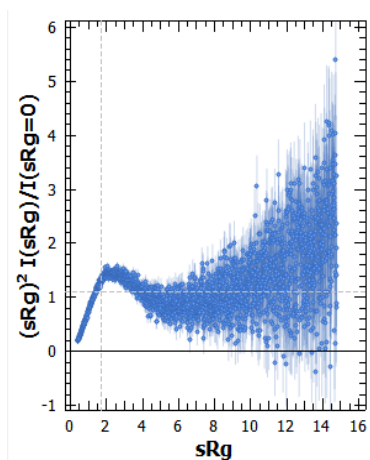
Figure 23. A) CD spectra, B) Normalised thermal denaturation, and C) Estimated T_m values of miniproteins **5-6**.

SAXS data was obtained for miniprotein **8**, the raw data indicated aggregation of the miniprotein at a concentration of 5 mg/mL in 50 mM potassium phosphate buffer pH 7.5. However, the quality of the data collected at a lower concentration, 2.5 mg/ml, showed no signs of aggregation and was used for data processing. The analysis indicated that miniprotein **8**, which exists as a tetramer in solution, has a globular shape with a peak at 24 and a D_{\max} of 83. The Kratky plot indicates a partially folded state in solution, (Figure 24B). Still, the higher degree of curvature compared to **3**, and the recovery of the lost thermal stability, suggest the formation of a more compact tertiary fold with an elevated degree of flexibility. From the experimental data, an electron density map was generated and fitted to the theoretical model with the use of DENSS software. The results of the fitting, (Figure 24D), agreed with the estimated oligomerization state, most likely occurring from the flexible region of the miniprotein.

A)



B)



C)

Parameters	8
Rg [Å]	26.5
I0	0.07
MW [kDa] expected	5.3
MW [kDa] experimental	19.9
Oligomeric state	tetramer
Folded:	partially
Quality of data	ok
Gnom Guinier/p(r)	24.4/24.5
Total quality estimate	0.84

D)

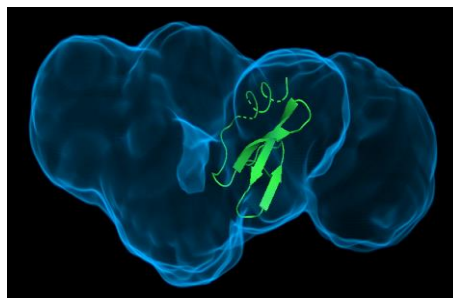
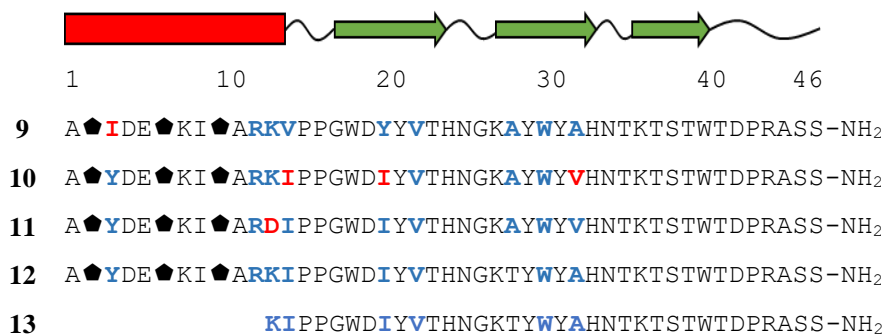


Figure 24. SAXS experimental results of mini-protein **8**, processed with ATSAS software, A) gnom real space distribution, B) Kratky plot, C) Experimentally determined parameters, and D) electron density map of **8** generated with DENSS from the experimental data.

Promising progress from mutants **2-8** was followed by synthesis of mini-proteins **9-13**, (Table 3). Ile3 residue was introduced to explore the effects of hydrophobic residues instead of aromatics, as it is more common to be found in the core of folded mini-proteins than Tyr. The incorporation of Ile13 and Ile19 residues was expected to increase the necessary contacts between the helix and the first strand by insertion of branched hydrophobic residues, consolidating the HEEE fold. Although Asp12 and Val31 residues were a product of Rosetta FastDesign, they helped to explore the consequences of already introduced mutations.

Table 3. Sequences of mini-proteins **9-13**. New mutations are highlighted in red, and modifications already explored in blue.



The CD experiments were consistent with the hypothesised contribution of Ile13 and 19, (Figure 25). For miniproteins **10-13**, which had the double mutation, increased the thermal stability with a T_m of ~ 47 °C. The CD spectra of **9** was similar to those of the previous miniproteins. However, for **10-13**, the profiles were clearly changed, and the aromatic contributions were partially restored in reference to **1**, suggesting an increase in the interactions. While for the helical regions the minima was shifted to higher λ values indicating a more helical character of the fold. The distinctive β -sheet signal at 215 nm was more pronounced, specially for mutant **12**. The results also suggested that the Asp12 and Val31 mutations had no structural or stabilisation effects. However, restoration of Thr27 residue appeared to promote the β -sheet character. The synthesis of **13**, the EEE analogue of **12**, confirmed the presence of cooperative folding of the HEEE topology by reducing T_m in the absence of the helix by ~ 20 °C

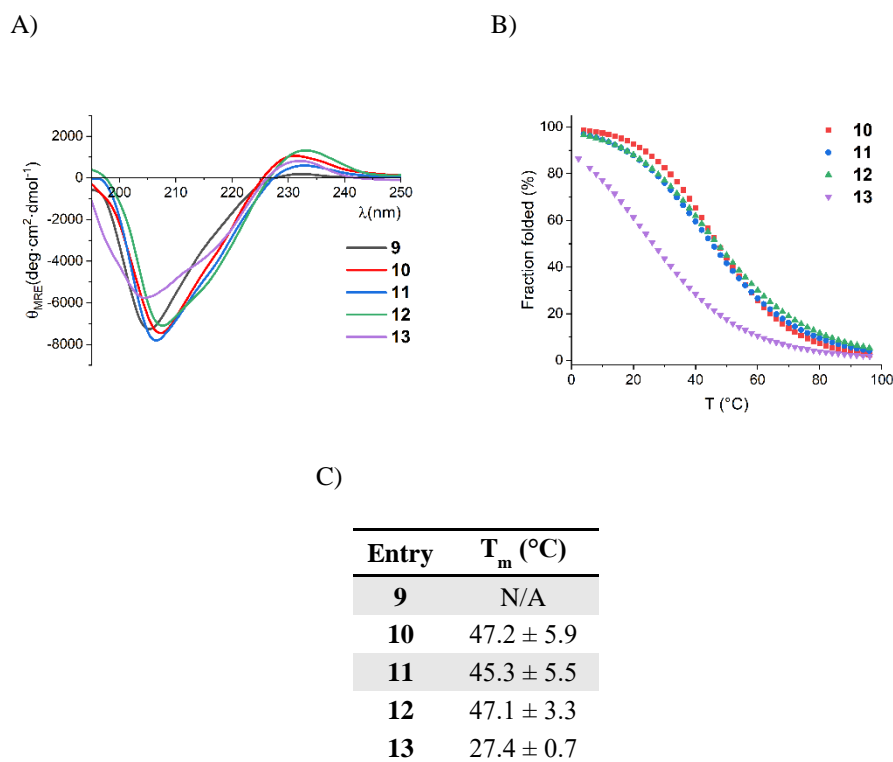


Figure 25. A) CD spectra, B) Normalised thermal denaturation, and C) Estimated T_m values of miniproteins **9-13**.

To confirm the experimentally observed results with a high resolution model, crystallisation sets of **10-12** were prepared. The plates were prepared with structure screens 1 and 2 (SS1, SS2) of molecular dimensions, by hanging-drop vapour diffusion. The initial stock solutions were 15 mg/ml of miniproteins in water, and three drops per condition were prepared in ratios 1:2; 1:1; 2:1 (miniprotein:condition). **12** showed needle-shaped crystals for condition 40 of SS1, (0.4 M sodium/potassium tartrate tetrahydrate) , (Figure 26). The needles were analysed in an XtaLAB Synergy-R from RIKAGU*. The analysis confirmed that the crystals were protein and not salts. Further efforts, to obtain better quality crystals, were done by combining condition 40 of SS1 with the additive screen, HR2-138, from Hampton Research, seeding of the needles and optimisation of the SS1-40 conditions. These efforts led only to the formation of needle-shaped crystals that, due to their small size, showed no diffraction when measured in the BESSY synchrotron in Berlin.

A)



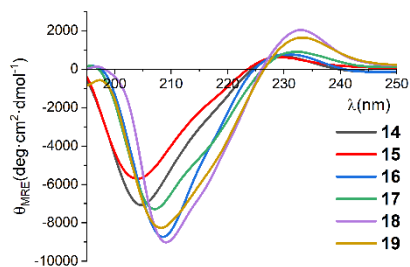
B)

Screen	Condition
40/SS1	0.4 M sodium/potassium tartrate tetrahydrate

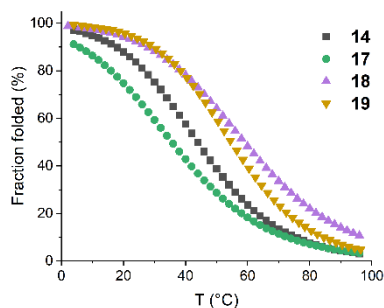
Figure 26. Results of the crystallisation of miniprotein **12**, A) Needle shaped crystals, and B) Crystallisation condition.

*At the Institute of Bioorganic Chemistry of the Polish Academy of Sciences in cooperation with Dr Magdalena Bejger and Prof. Wojciech Rypniewski.

A)



B)



C)

Entry	T_m (°C)
14	43.9 ± 3.5
15	N/A
16	N/A
17	36.2 ± 1.6
18	58.9 ± 8.6
19	54.7 ± 1.9

Figure 27. A) CD spectra, B) Normalised thermal denaturation, and C) Estimated T_m values of mini-proteins **14-19**.

The crystallisation of **18** and **19**, by hanging-drop vapour diffusion, generated crystal needles of **18**. Further optimisation and exploration of crystallisation screens led to the formation of larger size crystals, from the index screen, (Table 5). Disappointingly, these crystals (Figure 28) showed no diffraction, powder diffraction, or weak diffraction. One crystal diffracted at 3 Å, however, the crystal was damaged reducing the quality of the data and precluding us from resolving the structure.

Table 4. Crystallisation conditions for mini-protein **18**.

Screen	Buffer	Precipitants
39	100 mM HEPES pH 7.0	30% Jeffamine ED-2001
70	100 mM BIS-TRIS pH 5.5	200 mM NaCl, 25% PEG 3350
71	100 mM BIS-TRIS pH 6.5	200 mM NaCl, 25% PEG 3350
83	100 mM BIS-TRIS pH 6.5	200 mM MgCl ₂ , 25% PEG 3350
84	100 mM HEPES pH 7.5	200 mM MgCl ₂ , 25% PEG 3350

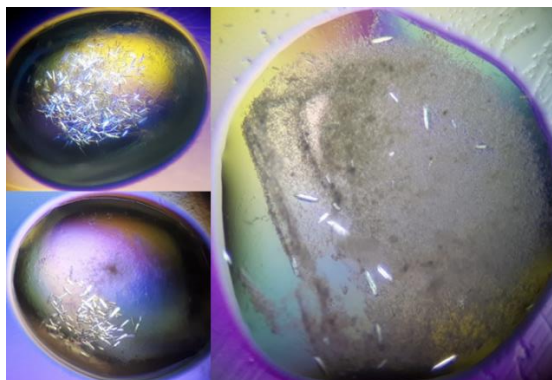
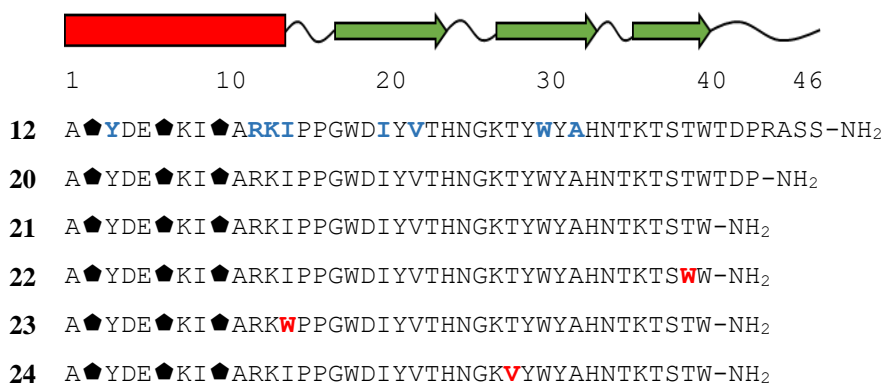


Figure 28. Crystal formation of **18** under different conditions.

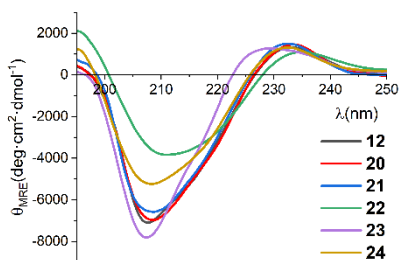
Considering the difficulties faced during the crystallisation and in need of a high-resolution structure, we decided to explore the possibility to resolve it through 2D NMR. However, the study of sequences longer than 30 amino acids represents a challenge for this technique in the absence of N-isotope labelled amino acids. Therefore, we decided to shorten the sequence of miniprotein **12** in two steps, one with 4 amino acids less from the C-terminus, **20**, and another one with 7 amino acids less, **21**. Furthermore, seeing the results of **21**, mutants **22-24** (Table 5) were designed to explore the potential enhancement of stability.

Table 5. Sequences of miniproteins **12** and **20-24**. New mutations are highlighted in red, and modifications already explored in blue.

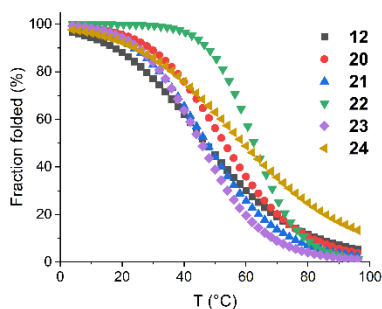


As expected, the CD spectra of miniprotein **12**, and its truncated analogues, **20-21**, had virtually no difference in the secondary structure content, (Figure 29). Furthermore, its thermal stability was similar to **12**. However, for structural elucidation purposes, **21** was more adequate. In contrast, **22-24** exhibited changes on their CD spectra. The lower ellipticity at 209 nm for **22** and **24** could suggest a lower content of helical fold. While for **23**, the lower ellipticity at 215 nm may indicate a disruption of the β -sheet. Moreover, the lower quality of the data collected for **22** and **24**, (Figure 29C) was the determinant of using **21** in the subsequent research of the present dissertation.

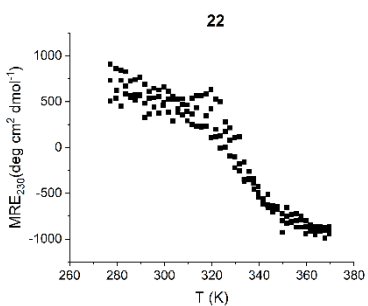
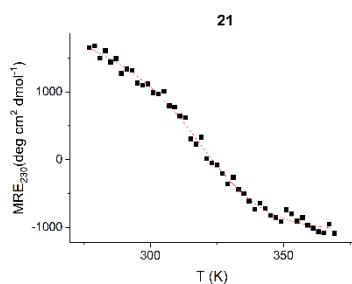
A)



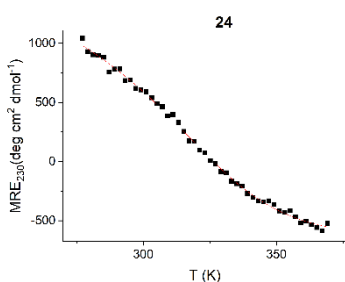
B)



C)



D)



Entry	T_m (°C)
12	47.1 ± 3.3
20	53.0 ± 3.9
21	47.2 ± 4.6
22	62.6 ± 1.0
23	45.5 ± 16.9
24	59.2 ± 6.4

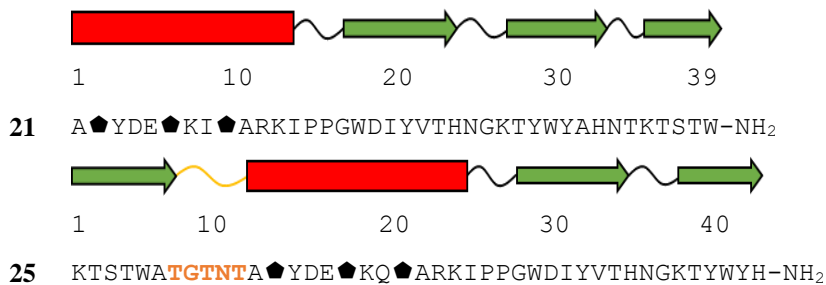
Figure 29. A) CD spectra, B) Normalized thermal denaturation, C) Recorded ellipticity vs T (K) for miniproteins **21-24**, and D) Estimated T_m values of miniproteins **14-19**.

Disappointingly, the structural elucidation efforts for **21**, were unsuccessful. Despite the crystallisation of this miniprotein, needle crystals formed, as **12**, they were too small to diffract. Furthermore, the high level of flexibility of the HEEE fold led to NMR results that prevented us from resolving the structure.

3.1.2. Circular permutation of HEEE to EHEE

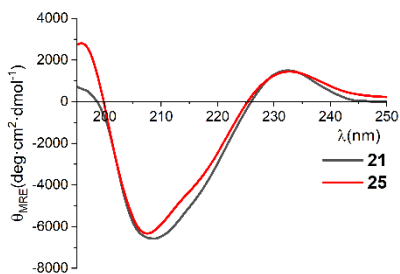
To achieve modification of the HEEE fold, and to explore the mechanism of folding for these miniproteins, we made use of the circular permutation²³⁶. Circular permutation can only be performed on those proteins whose N and C termini are in proximity in the three-dimensional space. This modification is achieved by the direct connection of the N- and C-termini or through a short peptide, and by the cleave of the sequence on the other point to generate new N- and C-termini. This approach has been widely used on large proteins²³⁷, to study protein fold²³⁸, enhance proteolytic stability²³⁹, improve activity²⁴⁰ and modify ligand specificity²⁴¹. However, for some small proteins, such as Trp-Cage²⁴², this approach has been successfully applied to study the dependencies of the sequences and the fold. We envisioned that the use of this technique in newly generated miniproteins could not only provide us information on the fold pathways but also enhance the folding stability by generating a more compact structure. To that end, miniprotein **21** was circularly permuted to generate miniprotein **25**, expected to form an EHEE fold. The linker used to connect both termini was designed to minimize aggregation by using Thr residues and to have a β -preference, (Table 6).

Table 6. Sequences of miniproteins **21** and **25**. The linker generated during the circular permutation is highlighted in orange.

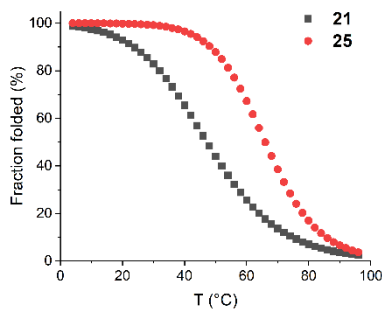


As expected, CD spectra of **25** show a secondary structure content similar to that of **21**, indicating the prevalence of the tertiary fold, (Figure 30). The thermal stability increased with a T_m value of ~ 64 °C for **25**, which may indicate an improved compactness of the EHEE fold.

A)



B)



C)

Entry	T_m (°C)
21	47.2 ± 4.6
25	63.7 ± 1.0

Figure 30. A) CD spectra, B) Normalised thermal denaturation, and C) Estimated T_m values of miniproteins **21** and **25**.

The cooperative folding of **25** was confirmed by CD. By monitoring the dependence of the ellipticity versus temperature of the miniprotein at selected wavelengths, we were able to track the stability of the different secondary structures (208 nm for helix, 215-220 nm for β -sheet) as well as the hydrophobic core (230 nm for aromatic contributions), (Figure 31). Due to data quality, **20** was used for the comparison. As stated above, miniprotein **20** has 3 amino acid longer sequence than **21** but the same CD profile. The resulting experiment reflected that circular permutation not only improved the fold and stability but generated a compact structure, which unfolds almost cooperatively. These kinds of experiments can provide us hints about potential unfolding pathways. The thermal CD data show that the T_m value of **20** is determined by the unfolding of the β -sheet, stabilise by the presence of the helix as seen for mutant **13**. However, the helix denatures at a later temperature. While for **25**, the EHEE topology induces a dependency of the unfolding between the present secondary structures. From the collected data we can hypothesise the unfolding pathway of **25**. Initially the first strand and the helix starts the unfolding, followed by the β -hairpin and the final disruption of the hydrophobic core, indicating a fully cooperative unfolding.

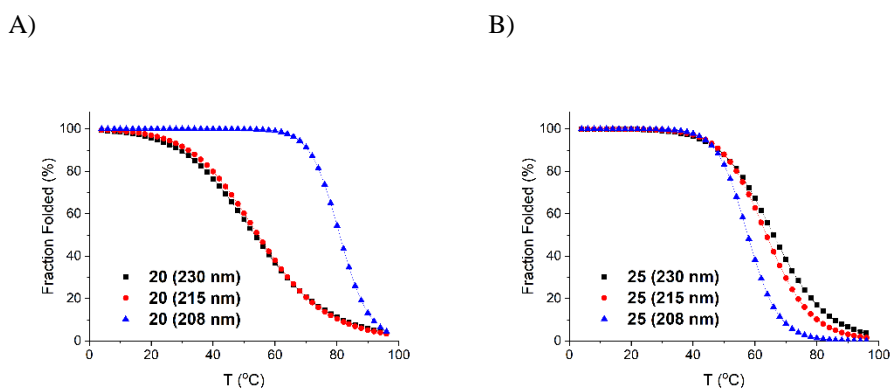



Figure 31. Comparative CD analysis of secondary structure stability between A) miniprotein **20**, and B) miniprotein **25**.

Following the successful permutation of the topology, we explored the possibility of enhancing the stability of the EHEE fold by modification on the residues of the β -sheet. We theorised that the generation of long range π -cation, π - π and π -CH inter-strand interactions could stabilise miniprotein **25**. To satisfy this class of interactions, branched

hydrophobic residues, aromatics, and histidine were placed along the first, second and third strands, focussing efforts to generate contacts with mutation Trp41, (Table 7).

Table 7. Sequences of miniproteins **25** and **34**. The new mutations are highlighted in red.



	1	10	20	30	40
25	KTSTWAT	TGNTA	YDE	KQ	ARKIPP
26	HTSTWAT	TGNTA	YDE	KQ	ARKIPP
27	KTSTWAT	TGNTA	YDE	KQ	ARKIPP
28	KTSTWAT	TGNTA	YDE	KQ	ARKIPP
29	KTSTWAT	TGNTA	YDE	KN	ARKIPP
30	HTSTL	ATGNTA	YDE	KQ	ARKIPP
31	HTSTWAT	TGNTA	YDE	KQ	ARKIPP
32	KTVTWAT	TGNTA	YDE	KQ	ARKIPP
33	KTVTWAT	TGNTA	YDE	KQ	ARKIPP
34	HTVTWAT	TGNTA	YDE	KQ	ARKIPP

The sequences are shown with mutations highlighted in red. The full sequences are:

25 KTSTWAT**TGNTA**YDEKQARKIPPGWDIYVTHNGKTYWYH-NH₂

26 **HT**STWATGTNTAYDEKQARKIPPGWDIYVTHNGKTY**W**WA-NH₂

27 KTSTWATGTNTAYDEKQARKIPPGWDIYV**L**HNGKTYWYH-NH₂

28 KTSTWATGTNTAYDEKQARKIPPGWDIYVTHNGK**V**WYH-NH₂

29 KTSTWATGTNTAYDE**KN**ARKIPPGW**HIF**VTHNGKTY**W**WA-NH₂

30 **HT**ST**L**ATGNTAYDEKQARKIPPGW**DI**L**V**THNGK**T**L**W**WA-NH₂

31 **HT**STWATGTNTAYDEKQARKIPPGW**HIF**VTHNGKTY**W**WA-NH₂

32 KT**V**TWATGTNTAYDEKQARKIPPGW**L**IYV**L**HNGKTYWYH-NH₂

33 KT**V**TWATGTNTAYDEKQARKIPPGW**LIF**V**L**HNGKTY**W**WA-NH₂

34 **HT****V**TWATGTNTAYDEKQARKIPPGWDIYVTHNGK**V**Y**W**WA-NH₂

The CD experimental results, provided a disparate palette of profiles, (Figure 32). Besides the enhancement of the stability for some of these mutants, the overall control of the fold was considered unsatisfactory. Supporte by the lack of agreement between the minima corresponding to the β -sheet, we concluded that the precise control of the fold was deficient. However, it was impossible to confirm this in the absence of a high-resolution structure. Nevertheless, it is worth highlighting the change on the ellipticity sign, for **30** and **33**, at 230 nm. Such an effect, observed later in other miniproteins of the present dissertation, could be explained by the nature of the surroundings of the aromatic residues. We hypothesized that when the aromatic residues are placed in a solvent exposed region, a positive band at 230 nm is observed. But, when they are in a hydrophobic environment, the signal changes sign to negative values. This could be explained for **30**, where the majority of aromatic residues, from the solvent exposed face of the β -sheet, were substituted by hydrophobic residues, leaving three of the four remaining aromatic residues within the hydrophobic core. Concerning **33**, a series of polar residues from the outer face of the β -sheet were modified to branched hydrophobic residues. We believe that these mutations generated a network of hydrophobic contacts along the β -sheet, inducing a

second hydrophobic environment and, consequently, placing all aromatic residues in solvent-free regions.

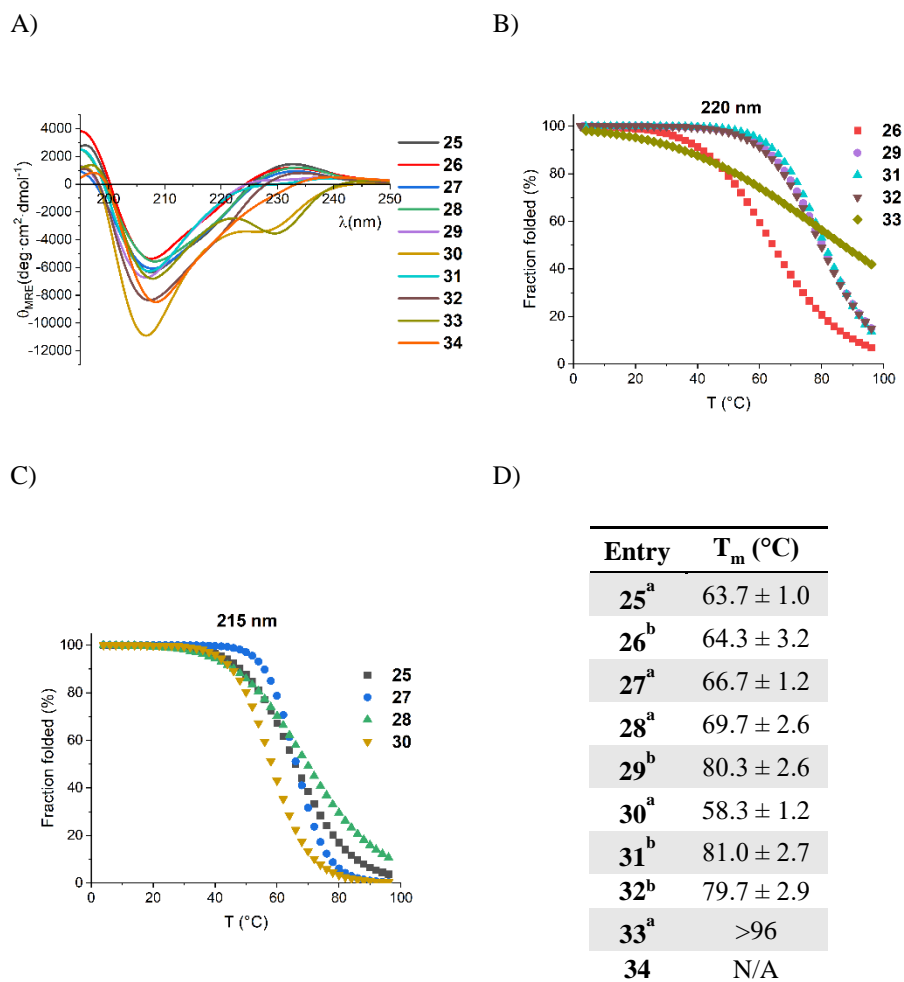


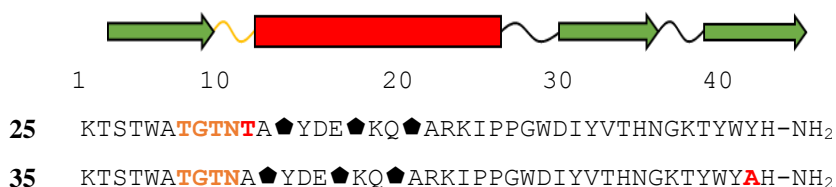
Figure 32. A) CD spectra of miniproteins **25-34**, B) Normalised thermal denaturation of miniproteins at 220 nm, C) Normalised thermal denaturation of miniproteins at 215 nm, and D) estimated T_m values of miniproteins **25-34**, where a stands for estimated T_m values at 220nm and b for estimated T_m values at 215 nm.

Further structural analysis of miniproteins **25-34**, by crystallography and 2D NMR, was unsatisfactory.

3.1.3. Optimisation and thermodynamics of the EHEE fold

Motivated by the successful design of the EHEE fold, and on sight of the stability question of miniproteins **25-34**, we decided to optimise **25**. The optimisation began with shortening the loop generated from the circular permutation and extend the third strand by one residue. Such modifications were introduced to allow the fold to be more compact. This led to miniprotein **35**, (Table 8).

Table 8. Sequences of miniproteins **25** and **35**.



The CD data analysis of miniprotein **35** displayed almost no changes in secondary structure content in comparison to **25**, (Figure 33). In addition, the thermal stability of this new miniprotein was slightly lower, indicating that the mutations were not enough to improve the overall stability.

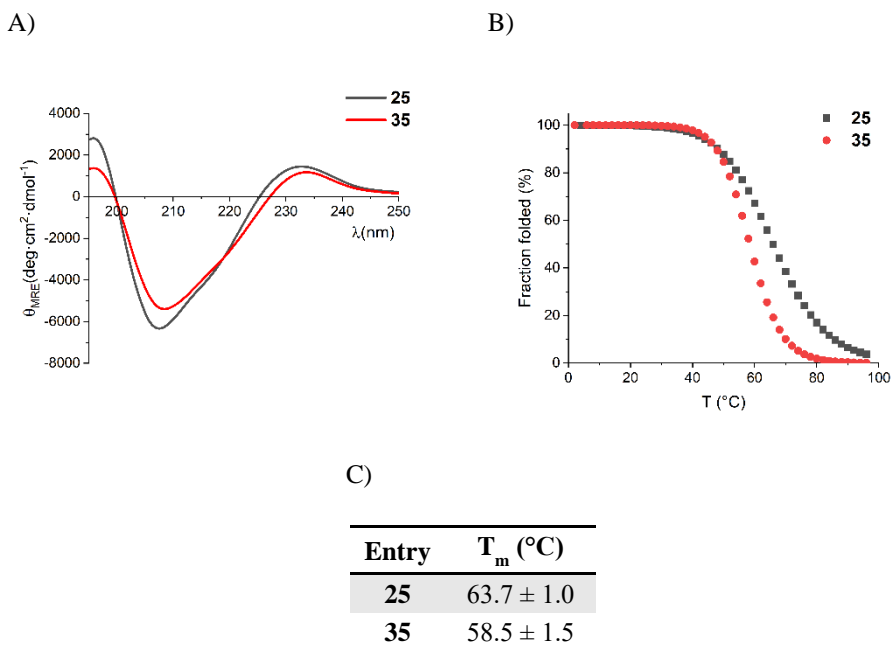


Figure 33. A) CD spectra, B) Normalised thermal denaturation, and C) Estimated T_m values of miniproteins **25** and **35**.

Thermochemical denaturation experiments, monitored by CD, were performed to obtain folding thermodynamic data of these new miniproteins, (Figure 34). The thermal stability of a protein in solution at different concentrations of denaturant, guanidine hydrochloride (GuHCl) is measured. Measurements are carried out at a wavelength of interest e.g., wavelength 220 nm was used to track the stability of the β -sheet. The resulting 2D plots are combined to generate a 3D plot of mean residue ellipticity (MRE) in function of guanidine concentration and temperature. Subsequently these data are fitted to obtain changes of unfolding enthalpy (ΔH°), entropy (ΔS°), heat capacity (ΔC_p°) and Gibbs free energy (ΔG°). The estimation of these values provides a deep understanding of the protein-fold process.

The thermodynamic data collected for miniprotein **35**, supported the hypothesis that the sequence folds into a stable tertiary structure, however its flexibility is translated into low-quality experimental data, poor data fit, and high susceptibility to the denaturant (Figure 34B). The results indicated a favourable fold of the miniprotein ($\Delta G^\circ = -0.8 \pm 0.4$ kcal/mol), with a low energy of inter-residue contacts (enthalpic contribution to the fold, $\Delta H^\circ = -3.6 \pm 6.1$ kcal/mol) barely enough to compensate for the entropic penalty ($-T\Delta S^\circ = +2.8 \pm 6.5$ kcal/mol), inherent in the protein fold mechanism. Additionally, the negativity of the heat capacity indicated that, as intended by the design, the fold is dominated by the burial of hydrophobic side chains.

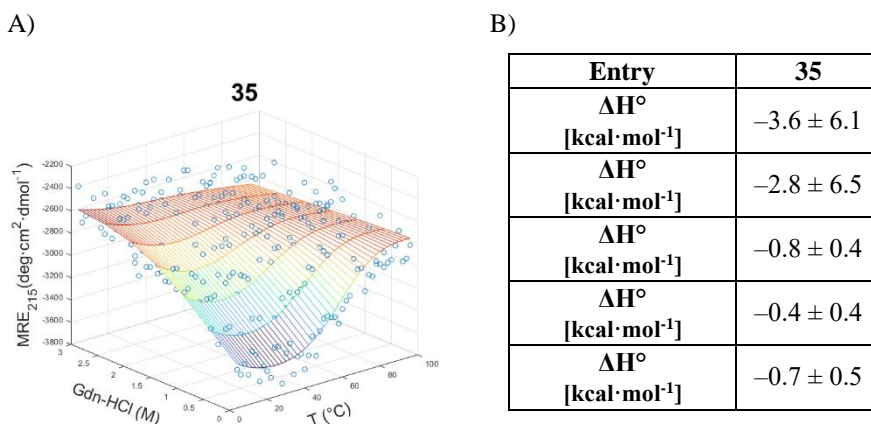



Figure 34. A) Global fit of thermal and guanidinium chloride denaturation of **35**. B) Thermodynamic parameters for the folding transitions of **35**.

The thermodynamics of folding for **35**, shed light on the difficulties in optimising the EHEE fold. The fold required of an increase in the enthalpic contribution to stabilise the fold, and consequently the increase on the number of contacts. To overcome this problem, we explored the possibility of enhancing the stability and contacts between the first and third strands resulting in miniprotein **36**. A total of 11 mutations were inserted in **36** for different purposes. Glu1 and Arg42 mutations were included to promote contacts between the N- and C-terminus and Asp28 residue of the second strand. To avoid undesirable conformations and reduce the contributions of aromatic residues to the fold, Tyr30 and Tyr40 residues were substituted for Ser30 and Thr40 respectively, which are β -strand inducing residues that improve solubility and reduce aggregation. Trp3 and 38 mutations were expected to form π - π interactions that will stabilise the first strand against the third and form the desired triple-stranded antiparallel beta sheet. Toward stabilisation and orientation of the first strand, Trp5 was modified for Glu5 to form a salt bridge with Lys36. Moreover, we introduced branched residues Ile4, Val6, and Val37 mutations, to increase contacts in the hydrophobic core and secure the packing. Two more design were synthesised to explore the repercussions of the contacts introduced in sequence **36** on the thermodynamics of folding. Concerning **37**, Trp3 and 38 mutations, expected to form π - π interactions, were removed as well as Asp28 was mutated to Gln to study the thermodynamic impact of this contacts in solvent exposed regions. While for **38** the modification of Trp39 to Leu was designed to study the thermodynamic impact of the interactions in the hydrophobic core. Finally, **39** was designed to study the influence of Gly as a secondary structure disruptor in another position (from 8 to 9) and the use of Ala8 as a guide of the first strand towards the core, (Table 9).

Table 9. Sequences of miniproteins **36-39**.



	1	10	20	30	40
36	E TWIEV T GTNK \blacklozenge YDE \blacklozenge KQ \blacklozenge ARKIPPGW D ISVTHNGK V W W TAR-NH ₂				
37	E L TIE A TGTNK \blacklozenge YDE \blacklozenge K E \blacklozenge ARKIPPGW Q ISVTHNGK V T W TAR-NH ₂				
38	ETWIEV T GTNK \blacklozenge YDE \blacklozenge KQ \blacklozenge ARKIPPGW D ISVTHNGK V T L W AR-NH ₂				
39	ETWIEV T A GNK \blacklozenge YDE \blacklozenge KQ \blacklozenge ARKIPPGW D ISVTHNGK V W W TAR-NH ₂				

The resulting miniproteins were studied by CD where miniproteins **36** and **39** showed an enhancement on β -sheet content as observable with an increase in ellipticity at 220 nm. (Figure 35). The removal of Trp39 from the hydrophobic core in miniprotein **38** proved the relevance of this residue in generating the β -sheet and a thermostable packing. A negative band at 230 nm was observed for miniprotein **37**. Considering that aromatic residues were removed from the solvent-exposed regions of the fold, it supported the hypothesis that the hydrophobic environment around aromatic residues induces a change in ellipticity at 230 nm. Except for **38**, all other mutants did not show apparent changes in T_m values.

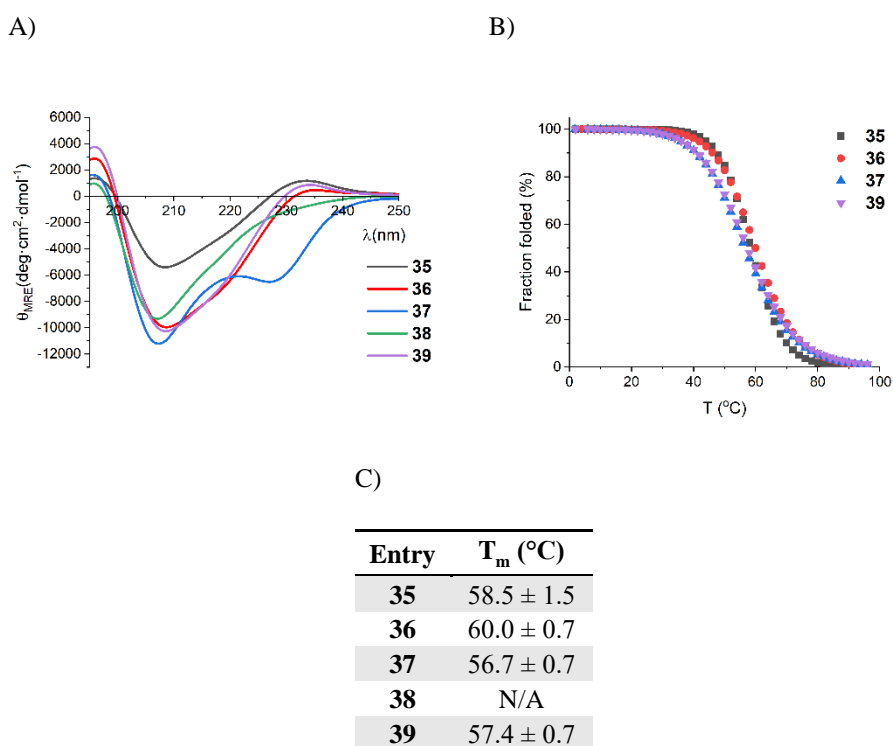


Figure 35. A) CD spectra, B) Normalised thermal denaturation, and C) Estimated T_m values of miniproteins **35-39**.

The significant variation on MRE signal, during unfolding studies, allowed us to estimate the folding thermodynamics by temperature and guanidine hydrochloride-dependent denaturation. As expected, the mutagenesis implemented at **35** enhanced the quality of the data and fitting, allowing the adequate estimation of the thermodynamic contributions to the fold, (Figure 36). The applied modifications satisfied the desired

increase in the number of contacts, as indicated by an increase in the folding enthalpy, (Table 10). However, the higher structural order was accompanied by an increase in the entropic penalty. This result is consistent with the general concept of protein folding, a process that is enthalpically favourable and entropically opposed.

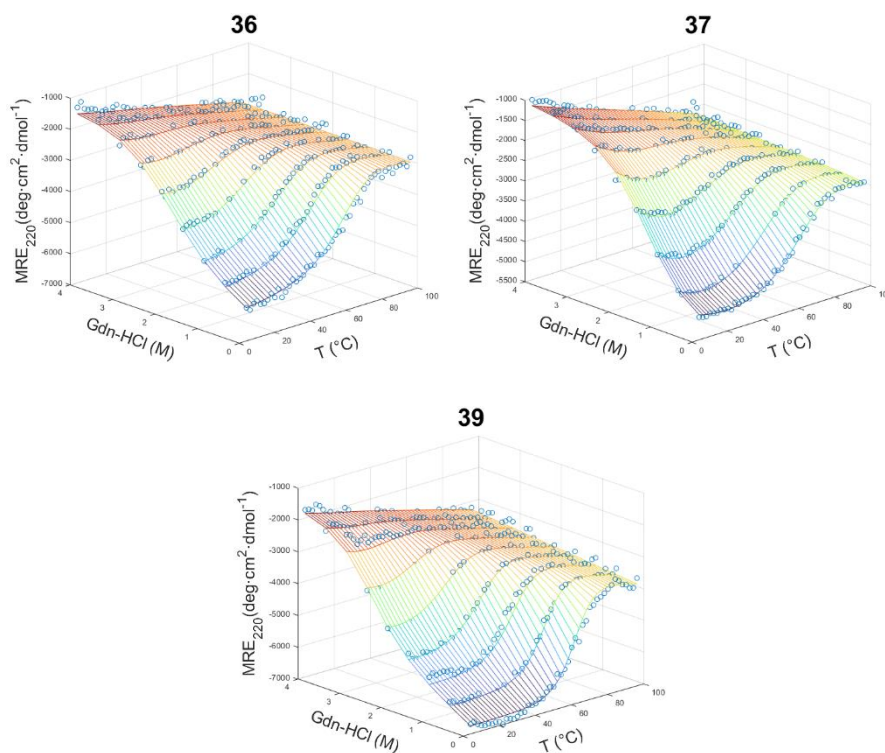


Figure 36. Global fit of thermal and guanidinium chloride denaturation of **36**, **37** and **39**.

The thermodynamic data provided great insight into the effects of the incorporated mutations. The comparison of mini-proteins **36** and **37**, revealed that the incorporation of stapled aromatic residues, π - π interactions, stabilised enthalpically the fold for **36**, by $\Delta\Delta H^\circ -4.2$ kcal/mol, a considerable contribution considering that the average ΔG° of folding is 5-15 kcal/mol. However, the entropic penalty generated from the hydrophobic side chain of Trp was $\Delta\Delta S^\circ -3.6$ kcal/mol. The overall contribution of these mutations to the stability was $\Delta\Delta G^\circ -1$ kcal/mol in reference to **37**. Additionally, the absence of this π - π contact on **37**, allowed us to approximate the influence of the introduced electrostatic interactions in reference to **35**. As expected, electrostatic contributions to the fold are significantly weaker, with a $\Delta\Delta G^\circ -0.4$ kcal/mol. The reduced energy contribution of

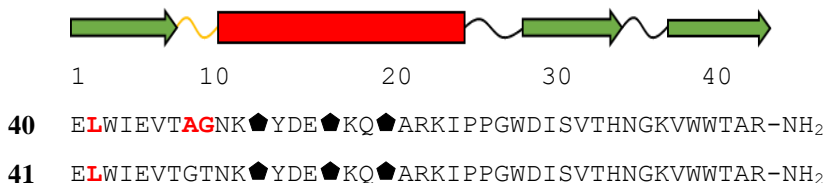
electrostatic interactions to the fold is promoted by the constant formation of hydrogen bonds between the charged residues and water molecules. Although their contribution to the fold is weak, they do not have a strong influence on the entropic penalty. The relocation of Gly8 to position 9 and incorporation of Ala8 instead increased the number of contacts and enthalpic penalty for miniprotein **39**. We hypothesise that the incorporation of Ala8 may cause a better packing of the first strand towards the hydrophobic core. This is supported by the significant increase in the number of contacts, $\Delta\Delta H^\circ \sim 4.5$ kcal/mol in reference to **36**, and the heat capacity, ~ 0.2 kcal/mol, which depends, between other factors, on changes in the solvent-accessible surface area (SASA).

Table 10. Thermodynamic parameters for the folding transitions of **35-37** and **39**.

Entry	ΔH° [kcal·mol ⁻¹]	$T \cdot \Delta S^\circ$ [kcal·mol ⁻¹]	ΔG° [kcal·mol ⁻¹]	ΔC_p° [kcal·mol ⁻¹ ·K ⁻¹]	m [kcal·mol ⁻¹ ·M ⁻¹]
35	-3.6 ± 6.1	-2.8 ± 6.5	-0.8 ± 0.4	-0.4 ± 0.4	-0.7 ± 0.5
36	-9.2 ± 1.0	-7.4 ± 1.0	-1.8 ± 0.09	-0.5 ± 0.09	-0.9 ± 0.09
37	-5.0 ± 0.4	-3.8 ± 0.4	-1.2 ± 0.04	-0.4 ± 0.05	-1.0 ± 0.05
39	-13.7 ± 1.1	-10.7 ± 0.9	-3.0 ± 0.2	-0.7 ± 0.07	-1.2 ± 0.09

The previously reported results suggested that miniprotein **39** could have the greatest thermodynamic properties of the set. However, the sensitivity of the protein backbone to denaturation (m) was greater than for **36**. The increase of m values in experimental error, though slight, could be explained by greater flexibility of the backbone. Consequently, to determine which sequence will be optimised, we generated mutants **40** and **41**, for **37** and **36** respectively, where each sequence had Leu2 instead of Thr2, (Table 11). The incorporation of Leu residue was expected to induce contacts in the hydrophobic core between the first strand and Trp39 of the third strand, and consequently generate an increase in enthalpic contributions.

Table 11. Sequences of miniproteins **40-41**.



The CD spectra of miniproteins **40** and **41**, displayed an almost identical secondary structure content in comparison to **39** and **36** respectively, (Figure 37). Thermal denaturation experiments indicated that not only fold but also stability prevailed, with T_m value of ~ 60 °C.

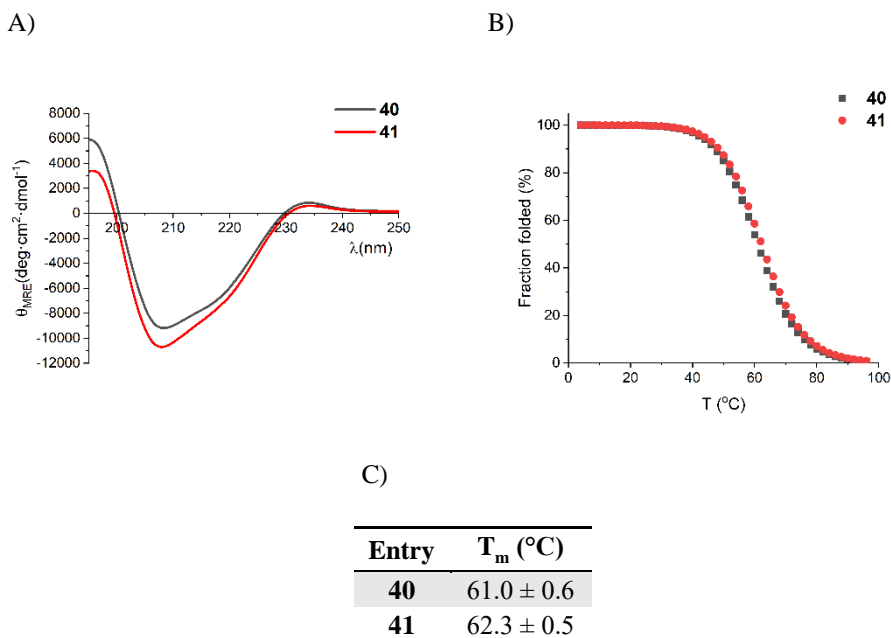


Figure 37. A) CD spectra, B) Normalised thermal denaturation, and C) Estimated T_m values of miniproteins **40** and **41**.

Contrary to our hypothesis, during the thermodynamic studies, (Figure 38) showed that the incorporation of Leu2 significantly lowered the enthalpy and entropy of folding for both mutants, (Table 12). We speculate that possibly the presence of Thr2 may modify the orientation of the first strand, in reference to the model, locating Trp3 in the direction of the hydrophobic core and, consequently, increasing the number of contacts. However, this hypothesis is far from being justified in the absence of a high-resolution model. Overall, the insertion of Leu2 did not significantly alter the free folding energy for **41**, while for **40**, a destabilisation with $\Delta\Delta G^\circ +0.9$ kcal/mol was observed. The greater capacity of **36** to withstand modifications with a slight improvement in the free energy of folding, $\Delta\Delta G^\circ -0.1$ kcal/mol, was the reason why it was chosen to continue the optimisation process.

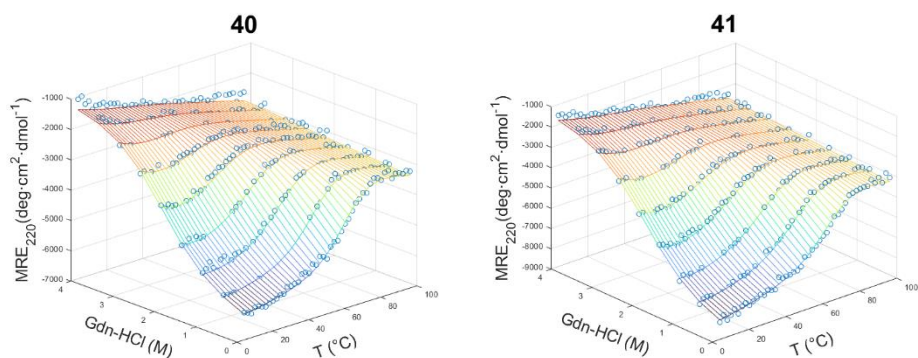


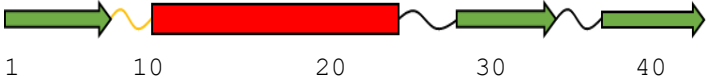
Figure 38. Global fit of thermal and guanidinium chloride denaturation of **40** and **41**.

Table 12. Thermodynamic parameters for the folding transitions of **40** and **41**.

Entry	ΔH° [kcal·mol ⁻¹]	$T \cdot \Delta S^\circ$ [kcal·mol ⁻¹]	ΔG° [kcal·mol ⁻¹]	ΔC_p° [kcal·mol ⁻¹ ·K ⁻¹]	m [kcal·mol ⁻¹ ·M ⁻¹]
40	-9.3 ± 0.8	-7.2 ± 0.002	-2.1 ± 0.1	-0.6 ± 0.08	-1.0 ± 0.1
41	-7.0 ± 0.6	-5.1 ± 0.6	-1.9 ± 0.01	-0.5 ± 0.07	-0.9 ± 0.07

The optimisation process was designed to rationally improve the folding energies and stability of miniprotein **36**. Single point mutations were introduced into the hydrophobic core to produce miniproteins **42-45**, except for miniprotein **44**. Miniprotein **44** was designed to explore the possibility of stabilising the helix against the β -sheet by polar interactions that cover possible solvent-accessible areas of the core, (Table 13). Also, to prove the crucial role of Ile23 residue in the overall stability of the EHEE fold was modified to Ala23 for miniprotein **42**. Ile23 would correspond to Ile13 of the HEEE fold, which was critical for the successful stabilisation of the tertiary structure. Given that phenylalanine is a hydrophobic amino acid with the capacity to generate strong aromatic interactions, it was introduced at positions 31 and 6 for miniproteins **43** and **45**, respectively. Ultimately, we proposed a sequence that contains all beneficial mutations, **46**, to study the overall contributions of the individual optimisations.

Table 13. Sequences of miniproteins **42-46**.



	1	10	20	30	40						
42	ETWIEVTGTNK	YDE	KQ	ARK	APPGWDISVTHNGKVWWTAR-NH ₂						
43	ETWIEVTGTNK	YDE	KQ	ARKI	PPGWDISFTHNGKVWWTAR-NH ₂						
44	ETWIEVTGTNK	YDE	RQ	ARKI	PPGWDISDTHNGKVWWTAR-NH ₂						
45	ETWIE	F	TGTNK	YDE	KQ	ARKI	PPGWDISVTHNGKVWWTAR-NH ₂				
46	E	L	WIE	F	TGTNK	YDE	KQ	ARKI	PPGWDIS	F	THNGKVWWTAR-NH ₂

The structural analysis of the mutants by CD indicated the prevalence of the desired fold except for **44**, (Figure 39). Inserting Arg17 and Asp31, to generate a salt bridge between the sheet and the helix, resulted in a CD spectrum distinctively different from the other analysed miniproteins and a complete fold disruption. Mutagenesis helped confirm the hypothesis of the crucial role of Ile23 in the fold, as a loss in thermal stability was observed for miniprotein **42**. Regarding the Phe mutants, the incorporation of bulked aromatic/hydrophobic residues in the vicinity of the core increased the T_m value of **43** in comparison to **36** by ~ 10 °C, what is a considerable improvement for a single mutation. However, the combination of all the favourable mutations on the all-mutant **46**, did not show a cumulative improvement of the thermal stability yet it provided the greatest stabilisation.

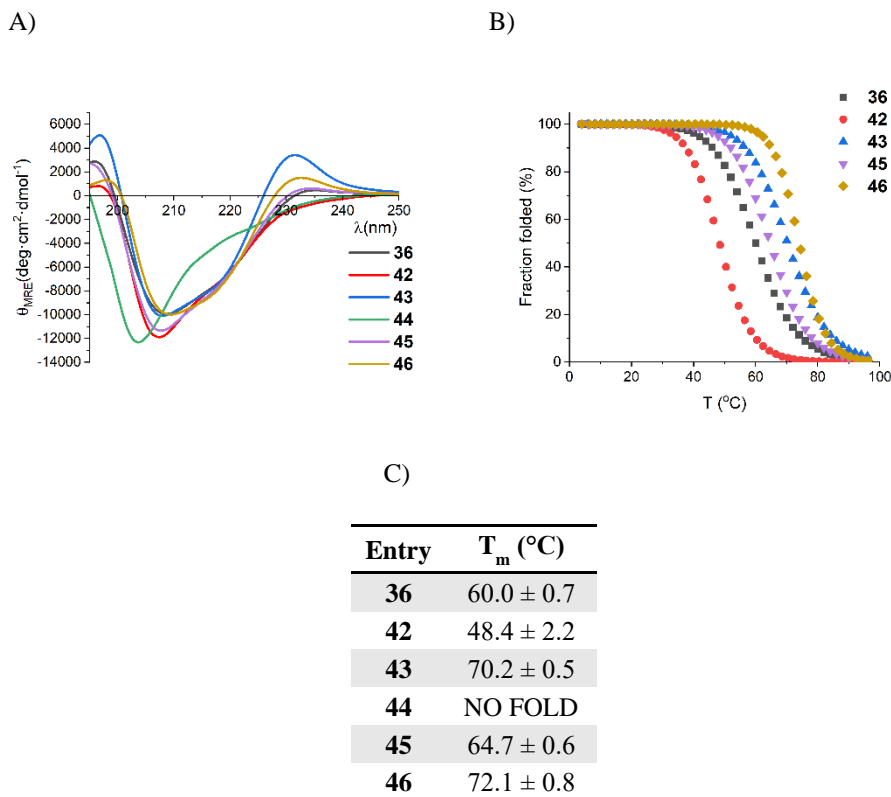
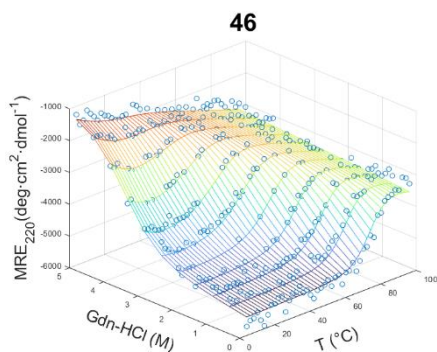


Figure 39. A) CD spectra, B) Normalised thermal denaturation, and C) Estimated T_m values of miniproteins **36** and **42-46**.

The thermodynamic studies, (Figure 40), of the optimised miniproteins agreed with the hypothesised idea that the increased size of the side-chain between valine and phenylalanine, could have a notable improvement in the folding parameters (mini-protein **43**, $\Delta\Delta G^\circ = -1.4$ kcal/mol and mini-protein **45**, $\Delta\Delta G^\circ = -1$ kcal/mol in comparison to **36**), (Table 14). The analysis of the contributions to the fold for each mutation can be done on the basis of the changes for the individual thermodynamic parameters. The observed increased ellipticity at 230 nm of **43**, in combination with the greater energetical stabilisation of the fold suggest the formation of contacts between residues Phe31 and Tyr13 from the helix. In addition, the lower increase of the heat capacity and no modifications of the entropy may suggest a residue more solvent exposed than at position 6 of miniprotein **45**. These characteristics are enhanced more for the all-mutant, a miniprotein with a $\Delta G^\circ = -4.2 \pm 0.3$ kcal/mol. The free energy of folding for **46** is comparable to that of native proteins.

A)



B)

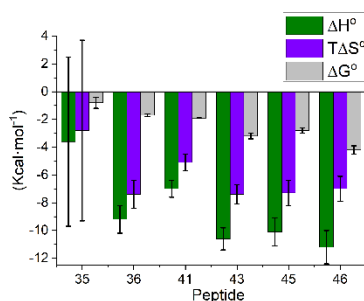


Figure 40. A) Global fit of thermal and guanidinium chloride denaturation of mini-protein **46**, and B) Contributions of thermodynamic parameters to the free energy of folding.

Table 14. Thermodynamic parameters for the folding transitions of **36**, **43**, **45**, and **46**.

Entry	ΔH°	$T \cdot \Delta S^\circ$	ΔG°	ΔC_p°	m
	[kcal·mol ⁻¹]	[kcal·mol ⁻¹]	[kcal·mol ⁻¹]	[kcal·mol ⁻¹ ·K ⁻¹]	[kcal·mol ⁻¹ ·M ⁻¹]
36	-9.2 ± 1.0	-7.4 ± 1.0	-1.8 ± 0.09	-0.5 ± 0.09	-0.9 ± 0.09
43	-10.6 ± 0.8	-7.4 ± 0.7	-3.2 ± 0.2	-0.6 ± 0.07	-1.0 ± 0.08
45	-10.1 ± 1.0	-7.3 ± 0.9	-2.8 ± 0.2	-0.7 ± 0.1	-1.1 ± 0.1
46	-11.2 ± 1.2	-7.0 ± 0.9	-4.2 ± 0.3	-0.75 ± 0.1	-1.1 ± 0.1

Miniproteins **36-41**, **43**, **45** and **46**, were analysed by 2D NMR. The samples were prepared at concentrations of 2-9 mg/ml in potassium phosphate buffer pH 6, 25 mM at 10 % D₂O (v/v). ¹H-NMR, total correlation spectroscopy (TOCSY), and nuclear Overhauser effect spectroscopy (NOESY) spectra were recorded. Analysis of the collected data suggested the presence of a well-defined β -hairpin as well as the helix and first strand. However, it was possible to observe a general loss of signals after Thr7. These observations, common among the entire set of miniproteins, may indicate a high degree of flexibility of the connectivity between the first strand and the helix. The flexibility of the loop was explored using molecular dynamics.

Using GROMACS molecular dynamics simulation software, simulations of 150 ns, (Figure 41), were performed for miniproteins **39** and **43**. Simulations were carried out in a cubic box with 1.0 nm distance from any box edge using explicit solvent. The solvated complex was neutralized by replacing water molecules for the corresponding ions to attain neutral charge of the system by the corresponding ions to achieve neutral charge of the box.

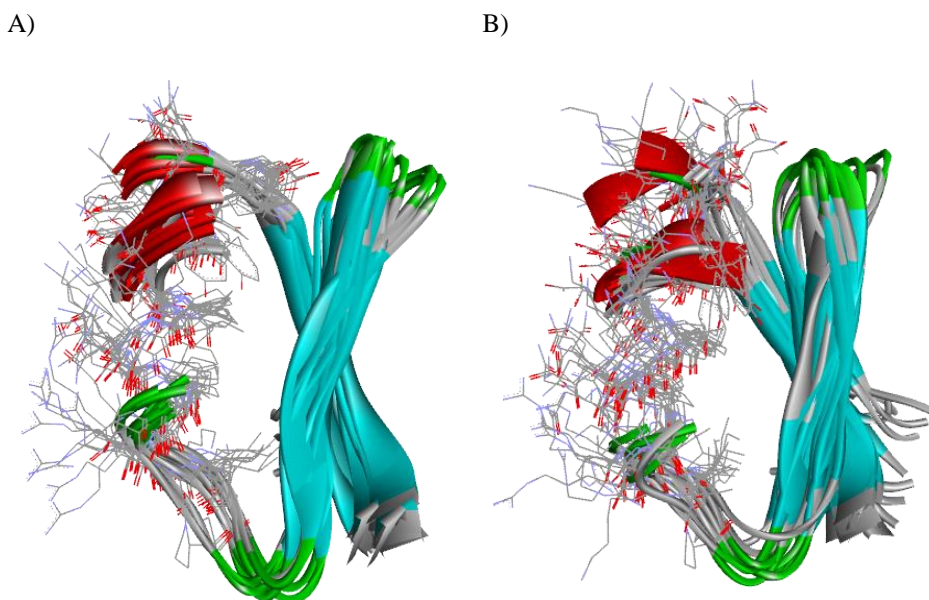
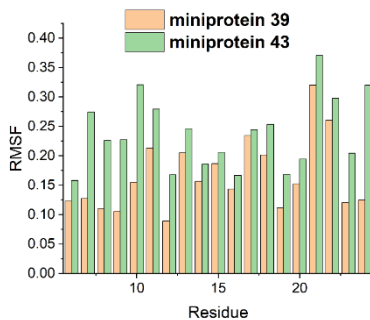


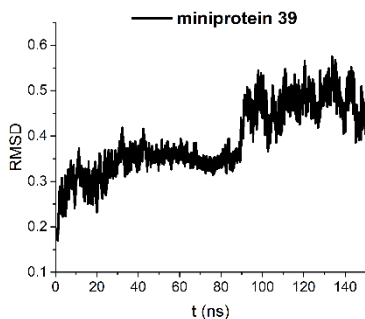
Figure 41. Snapshots from MD simulations of miniproteins A) **39**, and B) **43**.

The results were consistent with the theorised flexible loop. The root mean square fluctuation, (Figure 42A), of loop 1 and the helix was lower for miniprotein **39**, where Gly was shifted to position 9 and Val was introduced instead, than for **43**. These results were consistent with the NMR analysis where the loop of **39** was assignable, however, not the helix. The plot of the root mean square deviation, (Figure 42B), along the simulations, showed that miniprotein **43**, was able to reach an equilibrated conformation faster than miniprotein **39**. This observation agreed with the better thermodynamic properties of the folding derived from a better packed hydrophobic core, suggesting that the **43** mutant was a better candidate for loop optimisation.

A)



B)



C)

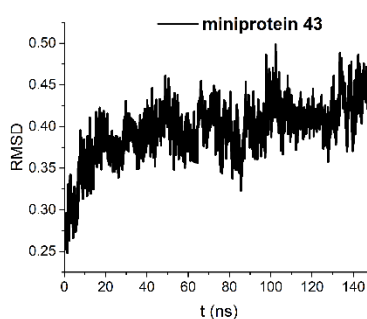


Figure 42. MD simulations of miniproteins **39** and **43**. A) Root mean square fluctuation (RMSF) of residues 6-24, corresponding to the flexible loop (residues 7-9) and the helix (residues 10-23), of miniprotein **39** (sand) and **43** (light green), B) Root mean square deviation (RMSD) of **39** along 150 ns of simulation, and C) Root mean square deviation (RMSD) of **43** along 150 ns simulation.

To explore this hypothesis, we designed the sequence of miniprotein **47**, (Table 15). In this mutant, Gly8 was modified to Val8 to decrease flexibility and to promote the orientation of the connection towards the hydrophobic core and generate contacts with residues Phe6 and Val37. Lys10, Glu11 and Lys15 mutations were introduced to generate electrostatic contacts that will minimise the flexibility. Finally, His33 was modified to Ser33 as a residue more suitable for a loop, (Figure 43). Phe6 and Phe31 from miniprotein **46** were incorporated and Thr2 left to minimise the risk of aggregation during thermal stability studies.

Table 15. Sequence of miniprotein **47**.

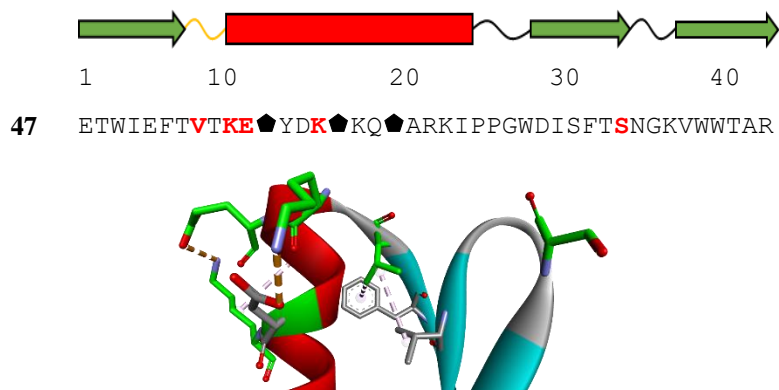


Figure 43. Highlighted in green are the modification to the loops in miniprotein **47**.

Following the design of miniprotein **47**, we sought to analyse the new model using MD, (Figure 44A). The results after a simulation of 500 ns revealed that **47** was reaching a stable conformation, in very close agreement with the model, much earlier than the miniproteins analysed previously, (Figure 44B).

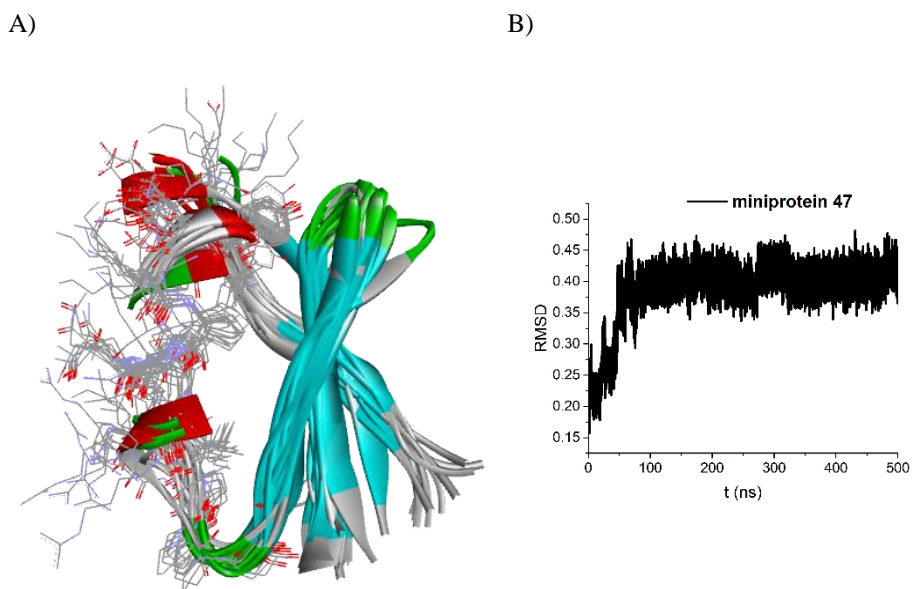


Figure 44. MD simulations of miniprotein **47**, A) Snapshots of MD simulation and B) Root mean square deviation (RMSD) of miniprotein **47** for 500 ns simulation.

The **47** miniprotein displayed a CD spectrum very similar to previously reported results within this section, (Figure 45). The secondary structure content prevailed after the modifications. Moreover, miniprotein **47** showed a high degree of thermal stability with a T_m value of ~ 70 °C.

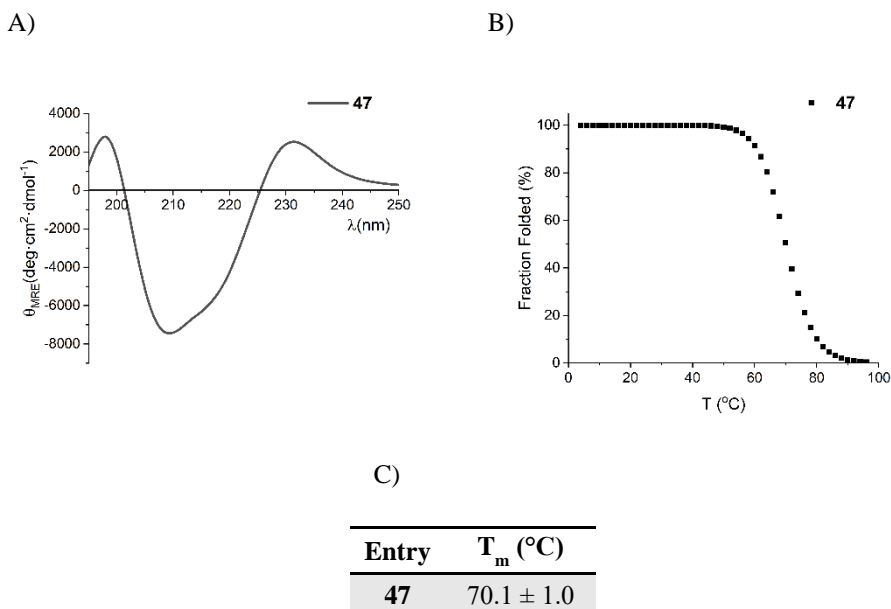


Figure 45. A) CD spectra, B) Normalised thermal denaturation, and C) Estimated T_m values of miniprotein **47**.

Furthermore, thermodynamic studies by thermal and guanidine hydrochloride denaturation, (Figure 46), supported the design approach with an increase of 1.1 kcal/mol in ΔG° of folding compared to **43**, (Table 16).

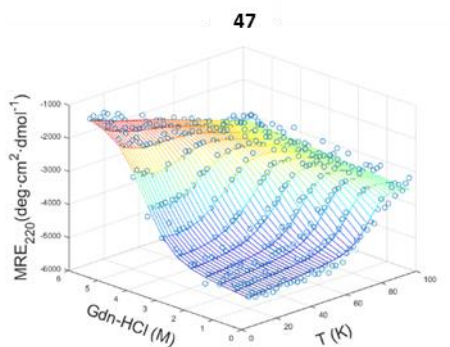


Figure 46. Global fit of thermal and guanidinium chloride denaturation of miniprotein **47**.

Table 16. Thermodynamic parameters for the folding transitions of **47**.

Entry	ΔH° [kcal·mol ⁻¹]	$T \cdot \Delta S^\circ$ [kcal·mol ⁻¹]	ΔG° [kcal·mol ⁻¹]	ΔC_p° [kcal·mol ⁻¹ ·K ⁻¹]	m [kcal·mol ⁻¹ ·M ⁻¹]
47	-12.8 ± 1.1	-8.5 ± 1.1	-4.3 ± 0.2	-0.72 ± 0.07	-1.2 ± 0.09

Miniprotein **47** was characterized using NMR spectroscopy*. The secondary shift of the α -protons indicated the presence of a well-defined helix (residues 10-23) with chemical shifts above +0.1 ppm, and the β -hairpin (residues 27-33 and 37-42) with chemical shifts below -0.1 ppm, (Figure 47), while the first strand (residues 1-8) remained undefined. However, the long-range contacts found in the 2D spectra support the packing of the first strand towards the hydrophobic core.

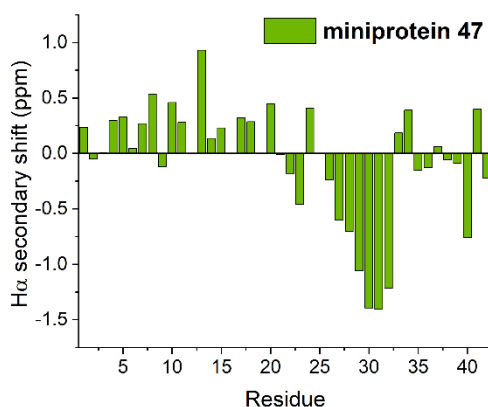
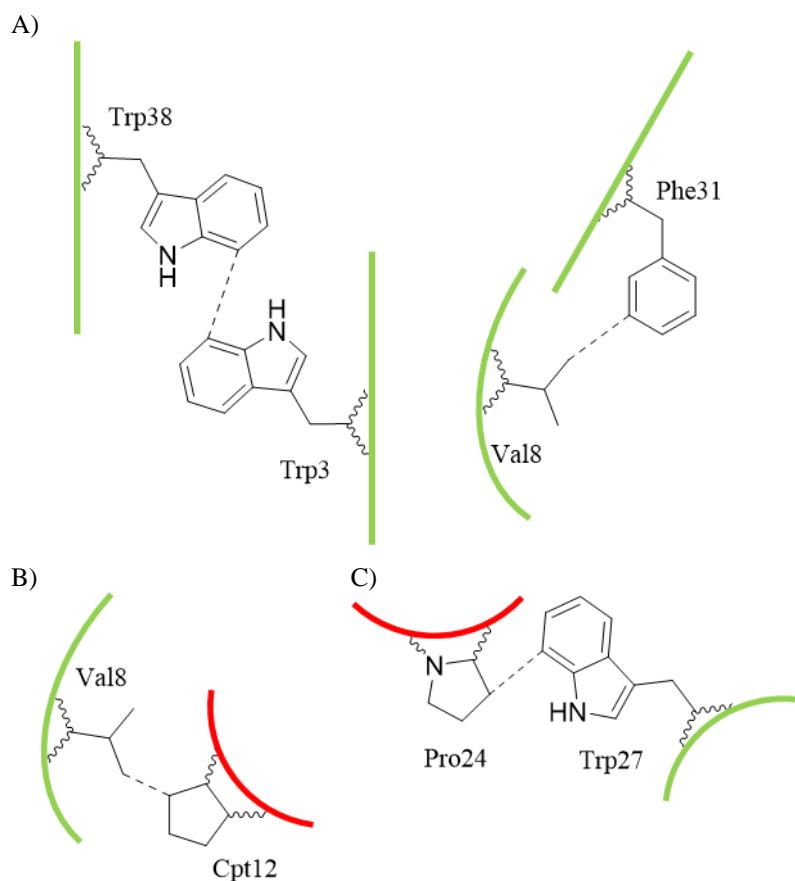


Figure 47. Secondary H_α chemical shift analysis of miniprotein **47**.

The analysis of the long-range contacts found in the NMR spectra also indicates the presence of well-defined regions of the miniprotein with different secondary structures and a compact fold through a hydrophobic core. Contacts found between Trp3 residue with Trp38 and Val8 with Phe31 suggests the presence of a compact first strand, (Figure 48A). Moreover, contacts found between Val8 residue and Cpt12 (*trans*-ACPC) suggest that the desired orientation of the first loop and contribution of the β -amino acids to the fold was achieved, (Figure 48B).

* In cooperation with Dr Ewa Rudzinska-Szostak, from the Department of Bioorganic Chemistry at the Faculty of Chemistry from Wrocław University of Science and Technology.

Additionally, the high number of $i - i+3$ and $i -i+4$ contacts between residues, found in the range Thr9-Ile23, supports the formation of the desired β -amino acid containing helix. Contacts found between Pro24 residue and Trp17 indicates that the desired orientation of the loop connecting the helix with the β -hairpin was successful, allowing for the formation of a hydrophobic core, (Figure 48C). The presence of a hydrophobic core is supported by contacts found between Trp27 residue and Trp39, Phe31 and Val37, Ile29 and Val37, (Figure 48D) in addition to contacts between the helix and the β -sheet as Tyr13 residue and Phe31, Phe31 and Cpt16, Ala20 and Ile29, (Figure 48E). This observations are in agreement with the described design of the core and the contributions of the different residues to the overall stability. The most defined region of the miniprotein is the β -hairpin formed between the second and third strand. The loop is well defined as can be found $i - i+3$ contacts between residues Thr32 with Gly35 and Ser33 with Lys36 supported by long range contacts between both strands.



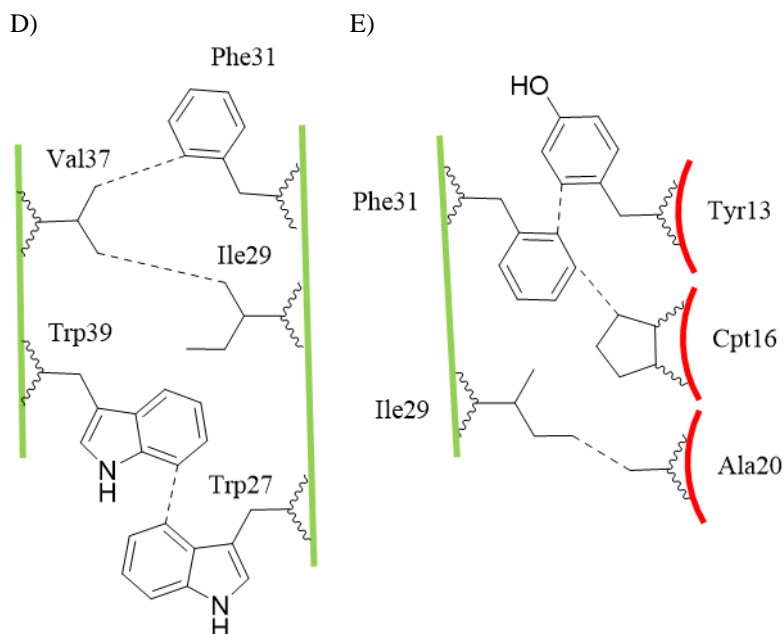


Figure 48. Relevant contacts found in the long range assignments of the 2D NMR data. Side chains of the interactive residues are represented connected to the structure to a strand (green lines) or to the helix (red lines). Interactive residues are connected with dashed lines.

From the data collected from the NMR experiments we can confirm the presence of a compact β -amino acid-containing miniprotein which fold is stabilized by the presence of a hydrophobic core. The hydrophobic core is generated between the helix and the β -hairpin as predicted on the design, where the interactive residues at the core are in contact with the *trans*-ACPC of the helix. This observation confirms the contribution of *trans*-ACPC to form the helical structure and to the stability of the overall tertiary fold. The analysis of the NMR spectra allowed for the generation of a preliminary model, using XPLOR-NIH software²⁴³, (Figure 49). The model supports the presence of a hydrophobic core between the desired secondary structures and the residues intended to interact during the design and optimization process. Also, the model converges into a common structure where, the β -hairpin, adopts the adequate orientation of the residues towards the core. However, the loop is left-handed and not right-handed as expected. This could be promoted by a similar energy between both conformations. The lack of convergence of the first strand may be caused by the high temperature, 306 K, required during the experiments to obtain well resolved signals for the processing.

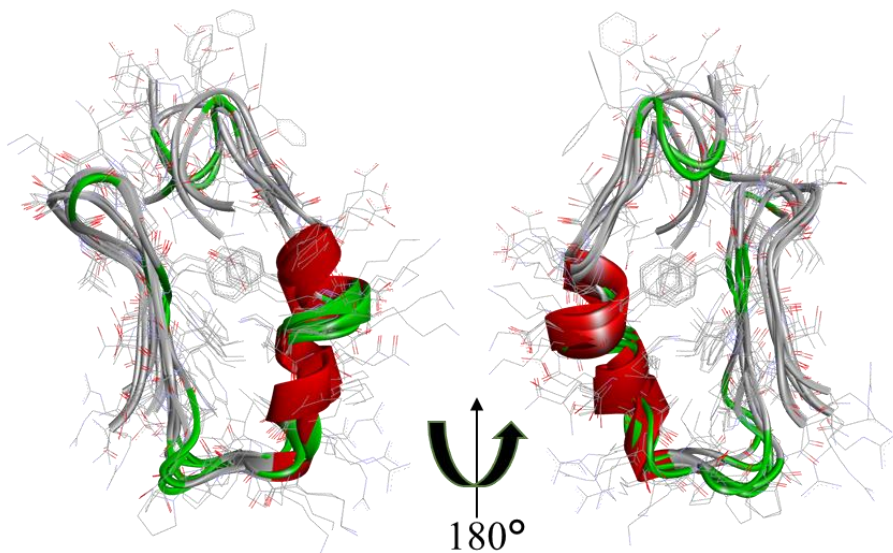


Figure 49. Preliminary NMR model of miniprotein **47**, where the five lowest energy models are superimposed. Backbone is represented with solid ribbons and the side chains are represented with lines.

In addition, to the structural analysis performed by CD, NanoDSF studies of denaturation and renaturation, were carried out for those miniproteins that were thermally stable, (Table 17). The results indicated that these miniproteins also refold spontaneously with estimated T_m values in agreement with the reported by CD.

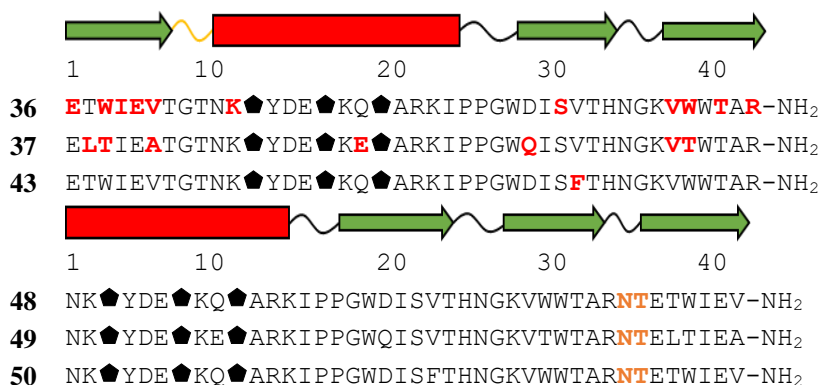
Table 17. Estimated T_m values of folding and unfolding using NanoDSF for miniproteins **35-37**, **39**, **41-43**, **45**, **46** and **47**.

Entry	T_m (°C)	T_m (°C)
	Unfolding	Refolding
35	60.4	58.8
36	65.1	59.6
37	62.8	75.9
39	65.4	58.2
41	65.7	50.5
42	49.4	34.7
43	72.2	67.4
45	65.6	60.2
46	76.9	71.2
47	77.3	74.8

3.1.4. Reversing the EHEE fold to HEEE

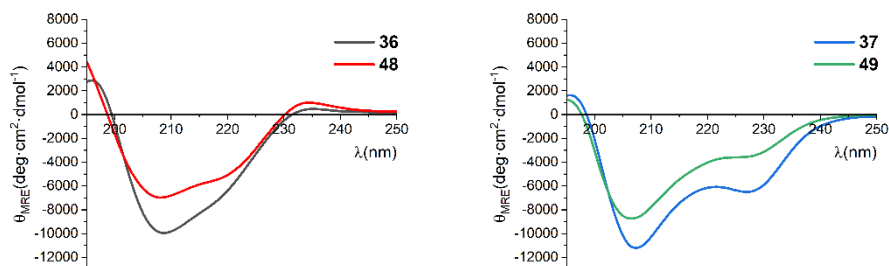
Successful optimisation of the EHEE was followed by reversion of the topology back to the HEEE fold by applying, once again, circular permutation. The purpose of this section was to explore the effects on the stability, and unfolding pathways, of the different topologies for sequences that have protein-like thermodynamic stability. To that end, miniproteins **36-37** and **43**, were permuted by removing the Thr-Gly-Thr linker between the first strand and the helix, and generating a new connectivity between the N- and C- terminus with the Asn-Thr dipeptide, (Table 18).

Table 18. Sequences of miniproteins **36**, **37** and **43** versus their circularly permuted analogues **48-50**. The mutations introduced in the previous sections are highlighted in red and the new linker for the circular permutation is highlighted in orange.

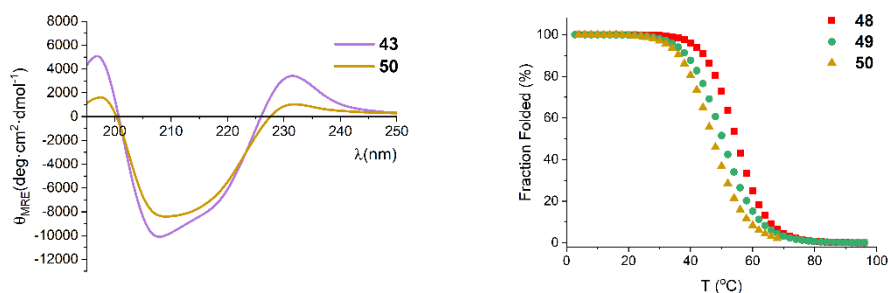


The structural analysis of **48-50** indicated the prevalence of secondary structures present in their EHEE analogues, suggesting that the formation of the antiparallel triple stranded β -sheet and the helix occurred in a similar fashion, (Figure 50). However, the lower ellipticity of the signal may indicate a more flexible fold with a lower number of contacts. The aggregation of **50** above 70 °C precluded us from the adequate estimate of T_m . However, the data was fitted up to that temperature allowing us to obtain an approximate T_m value of this miniprotein.

A)



B)



C)

Entry	T_m (°C)
48	54.5 ± 0.9
49	50.3 ± 1.2
50	$\sim 50^*$

Figure 50. A) Superimposed CD spectra of mini-proteins **48-50** with their EHEE analogues, B) Normalised thermal denaturation, and C) Estimated T_m values of mini-proteins **48-50**. (*) Indicates that the T_m value is an approximation derived from the data collected prior to precipitation of the mini-protein.

The thermal denaturation of the helix, β -sheet and aromatic contributions was simultaneously monitored by CD for mini-proteins **48** and **49**. As expected, and despite the thermal stability of their EHEE analogues, the cooperativeness of the fold is modified with a less compact topology. Observing the results, we hypothesise that, for the EHEE fold, initially the first strand begins to unpack from the core, which destabilises the helix and finally breaks the hydrophobic core, almost simultaneously. While the HEEE fold disrupts

first the hydrophobic core, followed by the denaturation of the beta sheet and finally the helix, (Figure 51).

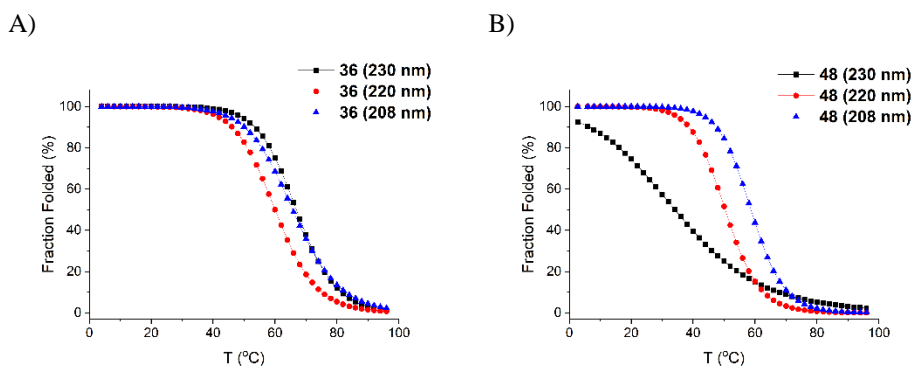


Figure 51. Comparative analysis of secondary structure stability between A) miniprotein **36**, and B) miniprotein **48**.

The observed thermal stabilities were in agreement with the data obtained during the thermodynamic studies. The susceptibility of the HEEE fold to the denaturant was greater than that of the more compact EHEE fold, as indicated by the estimated m value, (Table 19). The presence of low levels of GuHCl for **48**, induced a two-stage denaturation of the protein without precipitation, indicating disruption of the core. The consequence of such effects was the poor data fit and low confidence in the thermodynamic parameters. Concerning **49**, the thermodynamic data showed that the modification of the topology had almost no effect on the entropic penalty of folding with an increase of 0.1 kcal/mol. This result suggests a similar prearrangement of the sequence in the unfolded state as for **37**. However, the less compact nature of the fold decreased by 1 kcal/mol the folding enthalpy, leading to a barely energetically favourable fold ($\Delta G^\circ = -0.1$ kcal/mol). Moreover, the decrease in the heat capacity validated the hypothesis that the hydrophobic core was less relevant to the folding stability of **49**.

Table 19. Thermodynamic parameters for the folding transitions of **48** and **49**.

Entry	ΔH° [kcal·mol ⁻¹]	$T \cdot \Delta S^\circ$ [kcal·mol ⁻¹]	ΔG° [kcal·mol ⁻¹]	ΔC_p° [kcal·mol ⁻¹ ·K ⁻¹]	m [kcal·mol ⁻¹ ·M ⁻¹]
48	-30.0 ± 9.3	-27.4 ± 8.4	-2.6 ± 0.9	-0.2 ± 0.2	-2.2 ± 0.7
49	-4.0 ± 1.1	-3.9 ± 1.1	-0.1 ± 0.06	-0.3 ± 0.1	-2.9 ± 0.5

3.1.5. Conclusions

In conclusion, our aim in this section of the thesis was the rational design of β -amino acid containing miniproteins and topology exploration, in order to generate new, complex, and controlled protein-like tertiary structures. A significant aspect of our work involved implementing interconnectivity modifications to enhance the stability of the same sequences, resulting in more robust structures. Moreover, we conducted an extensive study of the secondary structures present on each scaffold, using data analysis of circular dichroism results. By assigning specific wavelengths to each secondary structure, we could monitor changes in the structure and predict the unfolding pathway of the miniprotein models presented. From the analysis of miniproteins **1-21**, we could recognise the importance of Tyr3, Ile13, Ile19 and Trp 29 mutations (Tyr13, Ile23, Ile29 and Trp39 in the EHEE fold) in combination with the three *trans*-ACPC of the helix for the formation of a stable hydrophobic core. These modifications influenced the overall fold and stability of the HEEE miniproteins, which was preserved after the topology modification. Circular permutation to the EHEE topology improved fold packing, increasing T_m by ~ 16 °C, and generated a cooperatively folded structure. We assumed that the stability enhancement was promoted by a reduced solvent-accessible core. The sequence exploration and optimisation of the EHEE fold provided strong evidence of the influence of the modifications on the stability. The results obtained in the thermodynamic studies of miniproteins **36-38** showed that the removal of relevant interactions, e.g., π - π stapling of Trp-Trp, from the solvent-exposed residues of the β -sheet was having a small negative impact on the folding thermodynamics of folding ($\Delta\Delta G^\circ = +0.6$ kcal/mol). However, removal of crucial residues as Trp39, was deriving in complete destabilisation of the fold. Proving, as intended at the beginning of the work, that the hydrophobic core around Trp39 was the main driving force of the miniprotein fold. The single-point mutations of miniprotein **36**, also revealed the impact of modifying other crucial residues as Ile23, which modification to Ala in miniprotein **42**, led to a loss in T_m of ~ 12 °C. Three mutations were identified that enhanced the thermodynamic stability of miniprotein **46** by -2.4 kcal/mol, to a total of -4.2 kcal/mol, more than two times higher than **36**, and a T_m of 72.1 °C. Furthermore, the constant structure analysis by NMR, CD, NanoDSF and MD allowed us to understand the dynamics of the fold at the atomic level, leading to miniprotein **47** with a ΔG° of folding of -4.3 kcal/mol and a T_m of 70 °C.

Notably, these miniproteins displayed a remarkable ability to fold into various topologies through loop replacements while retaining their stability and conformation. Additionally, the design allowed us to incorporate *trans*-ACPC, not only as a helical structure inducer but also as a contributor to the tertiary structure stability. Moreover, these miniproteins exhibit a strong dependence of the fold in the sequence directed toward the hydrophobic core. This characteristic confers them with the potential to be redesigned into highly active inhibitors of the PD-1/PD-L1 immune checkpoint without jeopardizing their three-dimensional conformation. The findings of this research contribute to the advancement of designing miniproteins with customised tertiary structures and heterogeneous backbones, broadening the scope of miniprotein design.

3.2. PD-1/PD-L1 interaction inhibitors

3.2.1. WW-domain-based inhibitors

Protein-protein interactions remain a challenging target in the design of active compounds due to the intricacy of the involved binding sites and the diverse conformations adopted by proteins. Additionally, the large and dynamic interfaces between interacting proteins make it difficult to identify small molecules that can effectively disrupt or modulate these interactions with high specificity and potency. To target the extended hydrophobic surface of the PD-L1 protein, it was crucial to look for a proper initial scaffold for the design. The search for a suitable miniprotein followed a series of criteria, a) a well-folded miniprotein with an extended surface for an adequate interaction with PD-L1, b) adequate sequence length for chemical synthesis by SPPS, c) a high-resolution structure should be available for modelling, d) the sequence should be tolerant to modifications and e) the topology or backbone should resemble the surface of PD-1. Therefore, it is crucial that the fold is driven mostly by residues that are not designed to interact with PD-L1. The search ended with WW-domains, protein domains composed of a right-twisted, triple-stranded, antiparallel β -sheet, with an average length of 40 amino acids. WW-domains are characteristic functional protein domains that have the capacity to fold in solution independently of the whole protein. This fold and its stability, are driven by the presence of conserved residues, in the absence of covalent bonds or metal ions, (Figure 52). Furthermore, the residues responsible for the fold are not present at the positions necessary for the design of the inhibitors.

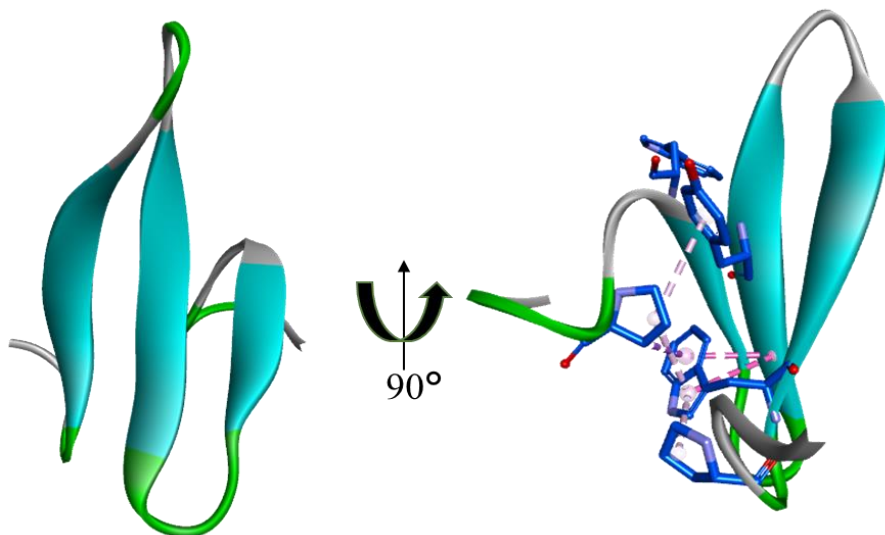
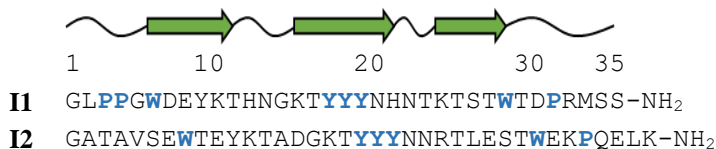


Figure 52. Ribbon representation of the high-resolution structure of the WW-Prototype (PDB id: 1E0M), The most relevant residues to the fold are highlighted in blue, and the hydrophobic contacts are highlighted in dashed lines.

Initially, to expand the search for a starting point for the development of inhibitors, two WW-domains were selected. A designed WW-domain, the WW-Prototype¹²² (**I1**), and a well-known native domain, the FBP28WW domain¹²² (**I2**) (Table20).

Table 20. Sequences of miniproteins **I1** and **I2**.



At first, stability and structural characteristics for both miniproteins were analysed by circular dichroism. For example, since the WW-domains fold in solution is mostly driven by tryptophan residues, monitoring thermal denaturation at 230 nm can provide us with the necessary information about structural stability. CD scans were recorded at 100 μ M concentrations of the miniproteins in potassium phosphate buffer (KPB) pH 7.5 50 mM, at 25°C, (Figure 53). For both WW-domains, the profiles observed in the CD spectrum agreed with those previously reported in the literature. Additionally, thermal denaturation performed also at a concentration of 100 μ M of the miniprotein, in KPB

pH 7.5 50 mM, and recorded at a wavelength of 230 nm from 4 to 96°C, at a speed of 5 °C/min, provided T_m values consistent with those reported in the literature¹²².

The thermal denaturation curves of these model scaffolds were fitted to a two-state unfolding curve, as visible by the sigmoidal shape of the experimental data. Such cooperative unfolding is characteristic of well-defined structures. Stability data showed that **I1** had a $\sim 20^\circ\text{C}$ lower T_m value than **I2** but similar CD spectrum. This analysis indicates, as is known by the NMR structure, that the three-dimensional arrangement of these two scaffolds is similar. Moreover, similar CD spectra indicate the presence of analogous structures.

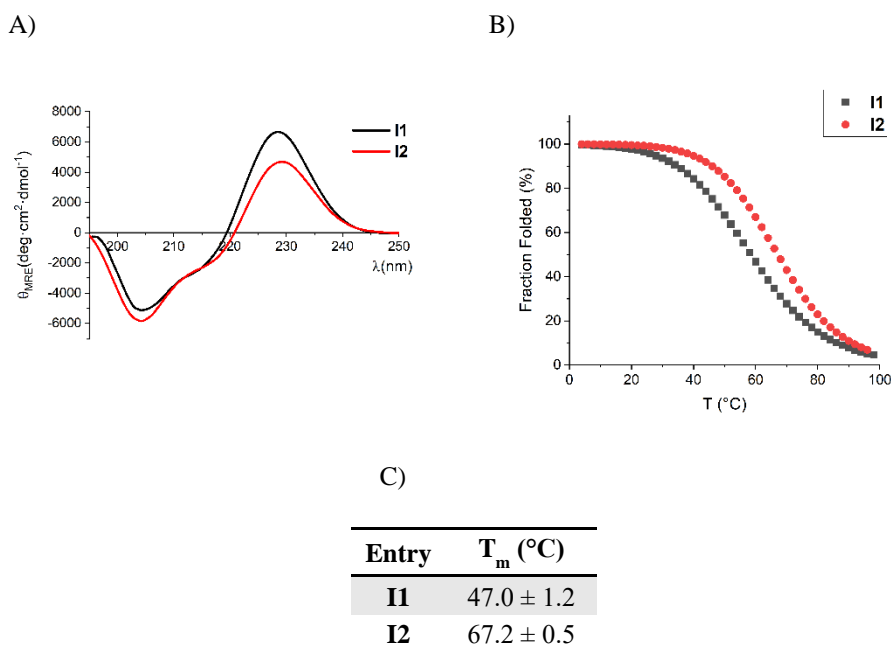
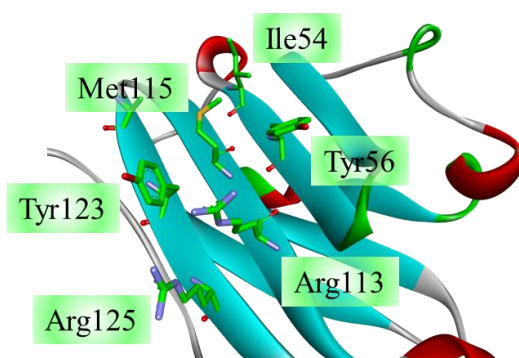


Figure 53. A) CD spectra, B) Normalised thermal denaturation, and C) Estimated T_m values of miniproteins **I1** and **I2**.

The inhibitor design was intended to generate a new binding mode of the inhibitors, by generating low-energy complexes with the PD-L1 protein. The applied approach was a residue-to-residue design, in which selected positions of the scaffolds were mutated to attain a higher affinity toward selected residues of the target protein. The crystal structure of the PD-1/PD-L1 complex⁷⁹, (PDB id: 4ZQK), was used for reference. In a first step, toward the design of potent PD-1/PD-L1 inhibitors, **I1** and **I2** were superimposed on the interactive surface of PD-L1. After the analysis of the native complex, the scaffolds

were placed on PD-L1 with the intention of covering the maximum the PD-L1 surface which interacts with PD-1. Considering that WW-domains have a smaller interactive surface than the PD-1 protein, there was the need to select a reduced number of targeted residues from the PD-L1 protein. We identified two hotspots for the design of inhibitors, a) a cavity between Arg113, Met115, Tyr123 and Arg 125; b) hydrophobic cavity between Ile54, Tyr56 and Met115, (Figure 54A). Once a suitable model of the complex was generated, the structures and interactive side chains were minimised by AMBER software, (Figure 54B).

A)



B)

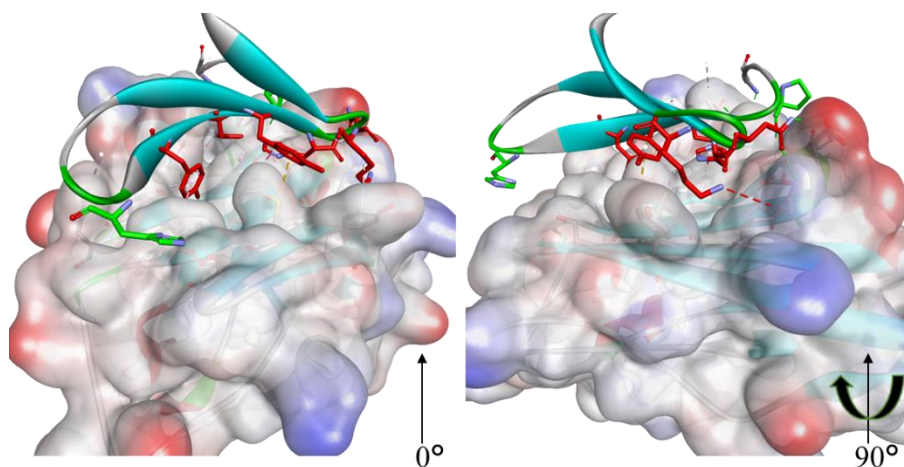
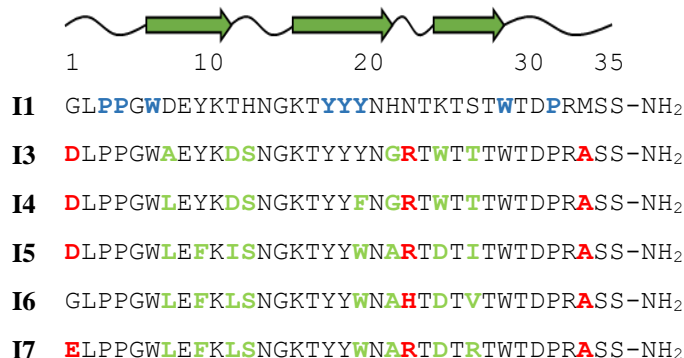


Figure 54. A) PD-L1 binding residues targeted for inhibitor design are highlighted in green and B) Superimposition of **II** over PD-L1.

Minimisation was followed by inhibitor design using the Rosetta FastDesign protocol. By providing an initial PDB file and a resfile that indicates points of mutation, Rosetta generates a selected number of mutants. The models were then scored by global energy and creating a ranking list. For the **I1** scaffold, inhibitors **I3-I7** were selected for synthesis and analysis, (Table 21). Design efforts were focused on positions 7, 11, 12, 19, 21, 24 and 26. In addition, post-design modifications were introduced. Asp/Glu1 and Arg/His22 were added to interact with PD-L1 residues Lys75 and Asp61 respectively. The interaction and specificity of inhibitors **I3** and **I4**, was intended to be driven by the side-chain of Trp24, expected to fit within the cavity generated between residues Arg113/Tyr123/Arg125 of PD-L1. Furthermore, Gly21 was hypothesised to promote the interaction of Trp24 interaction with PD-L1 by reducing steric hindrance in the absence of side chains. To achieve the same objective and to increase the hydrophobic interacting surface of the inhibitor, Ala or Leu residues were included in position 7. Although in the case of inhibitor **I3** Tyr19 was kept not to disrupt the stabilisation effects of the Tyr triad, found in native scaffolds, Tyr19 was modified to Phe19, in inhibitor **I4**, to study the effects of increasing the hydrophobic character of the β -sheet. Finally, Asp11 and Ser12 mutations were included to generate potential hydrogen bonds with Ser117. On the contrary, the **I5-I7** designs were modelled assuming that Trp19 and Asp24 are key mutations. These modifications were intended to interact through CH- π , π -cation and salt bridges with Arg113, Tyr123, and Arg125 of PD-L1. Furthermore, the other mutations were expected to favour the interaction through larger hydrophobic-aromatic surfaces of the inhibitors, except for **I7** where Arg26 was hypothesised to cover solvent-accessible areas at the edge of the interaction. To minimise synthetical difficulties, Met33 residue was mutated into Ala33.

Table 21. Sequences of inhibitors **I1** and **I3-I7**. The new mutations are highlighted in green while the post-design modifications are highlighted in red and crucial residues to the fold are in blue.



The synthesis and purification of the inhibitors was followed by structural analysis by CD, (Figure 55). The experimental results were then compared to those obtained from the original scaffold. Miniprotein **I3** showed a CD spectrum similar to **I1**, but a lower ellipticity at 230 nm and a higher ellipticity at 217 nm was observed. These observations, together with an increase in T_m values of ~ 10 °C suggests a higher β -sheet order, yet with lower aromatic contributions to the fold. On the other hand, the CD spectra of **I4**, shows an almost complete loss of the aromatic contribution, at 230 nm, but with a tendency to form an overall similar fold to **I1**, as the spectra was very similar in the range of 195-220 nm. Considering that the major difference in sequence of inhibitors **I3** and **I4** is the mutation Tyr19-Phe19, it seems reasonable to consider that the mutation has a great impact on the aromatic stabilisation of the native Tyr triad, characteristic of this type of scaffolds. This destabilising effect is reflected in the lack of fitting for the T_m values for **I4**. Although inhibitors **I5-I7** manifest similarities between their CD spectra, they differ the most from the native fold. The aromatic contributions remain almost unchanged for **I5** and **I6** but with a noticeable loss of the minima at 205 nm. It is possible to assume that modifications performed over the original sequence may have affected the shape of the β -sheet fold but increased the T_m value by 7 and 14°C, respectively for **I5**, and **I6**. A similar profile to **I5** and **I6** is observed for **I7** but with a clear decrease in aromatic contributions. This decrease in ellipticity at 230 nm may arise from the incorporation of Arg26. The presence of the long polar side chain of Arg may generate a π -cation interaction with Trp19 that disrupts aromatic interactions between Phe9 and Trp19.

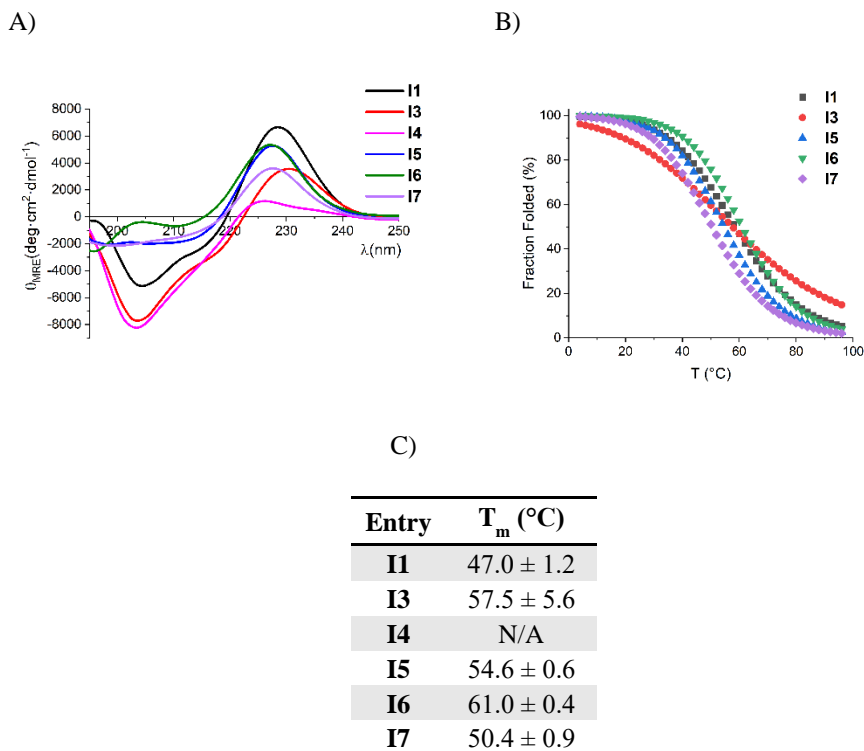


Figure 55. A) CD spectra, B) Normalised thermal denaturation, and C) Estimated T_m values of inhibitors **I1** and **I3-I7**.

In the present work, BioLayer Interferometry (BLI) was used to study the binding kinetics of the inhibitors to the target, PD-L1. Biolayer interferometry is a real-time and label-free technique used to analyse biomolecular interactions by measuring changes in the interference pattern of light as molecules bind to a surface-immobilised target. During the measurements, a biosensor is loaded with the target protein or ligand and dipped into a solution with the analyte, the association step, and the light interference is measured. The association step is followed by a dissociation step by dipping the sensor into an analyte-free buffer. The rates of association (k_a) and dissociation (k_d) are estimated by fitting the experimental data to obtain affinity (K_D) values. BLI studies were performed using 10 mM HEPES buffer with 150 mM NaCl, 3 mM EDTA, 0.05% (v/v) tween 20 and 1% (w/v) of BSA, at pH 7.4 for all sample preparations. During the loading step, 1 μ g/mL of biotinylated PD-L1 (biot-PD-L1) was immobilised at the tip of a streptavidin biosensor (SAX), while the other sensor was left unloaded to measure nonspecific binding (NSB) of the analytes. The subtraction of the NSB signal was performed after the measurements. Once the ligand was immobilised on the sensor, a quenching step was added by dipping

both sensors, loaded and unloaded, in a Biocytin solution at a concentration of 1 $\mu\text{g/mL}$ concentration. After a baseline measurement, the quenching was followed by the association step (300 s), where both sensors were dipped into wells containing 50 μM solutions of the inhibitors. The dissociation step (500 s), which followed the association, was performed by dipping the sensor into the buffer and measuring the changes in light interference.

The **I4** inhibitor was excluded for BLI studies due to its lack of cooperative unfolding observed by circular dichroism. The analysis of the results showed that inhibitors **I3**, **I5-I7** had a moderate binding capacity to PD-L1 at high concentrations. The K_D values ranged between 31-270 μM with high fit errors, e.g., >100% for **I5**, and low shifts in the sensograms, (Figure 56).

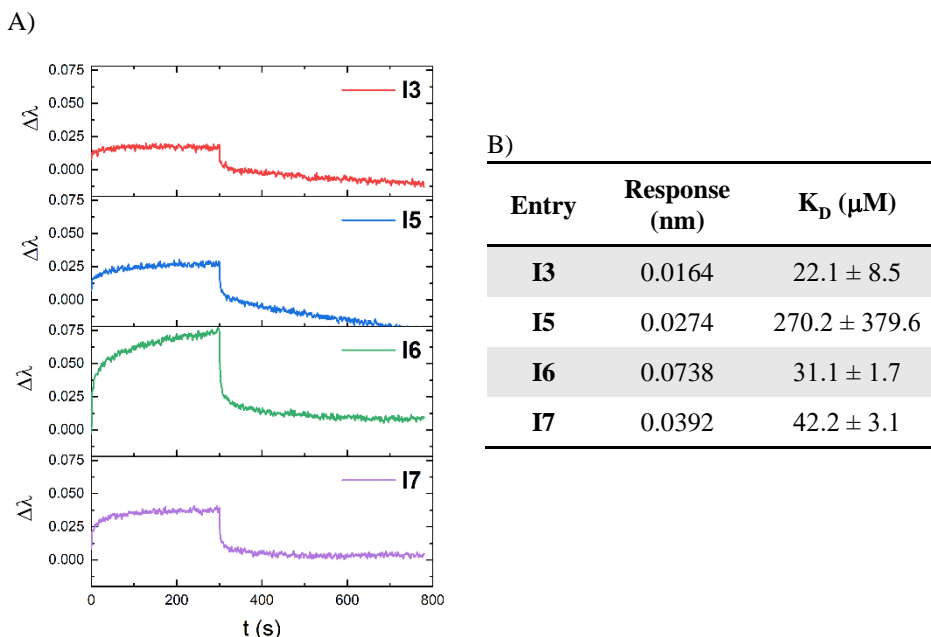


Figure 56. BLI results for inhibitors **I3**, and **I5-I7**. A) BLI plotted sensograms, and B) Highest BLI response and estimated K_D values.

The affinity studies were followed by inhibition studies by Homogeneous Time Resolved Fluorescence (HTRF). This protein-based assay provides information on the ability of an inhibitor to disrupt the PD-1/PD-L1 complex. HTRF analysis was performed at two concentrations, 50 and 0.5 μM . The stock solutions, with a 10-fold final concentration in the well, were prepared using the assay buffer, as indicated in the protocol, and incubated for two hours before measuring the fluorescence. The use of these two concentrations was intended to monitor the behaviour of the inhibitors in the assay.

The results indicated that the three studied inhibitors had an approximate IC_{50} of 50 μM , (Figure 57). However, **I7** reflected that it could potentially interfere with the basis of the assay. This observation was based on the results obtained for the lower concentration of the inhibitor, where with 100 times less concentration of the inhibitor the difference of undissociated complex was nearly 18% more.

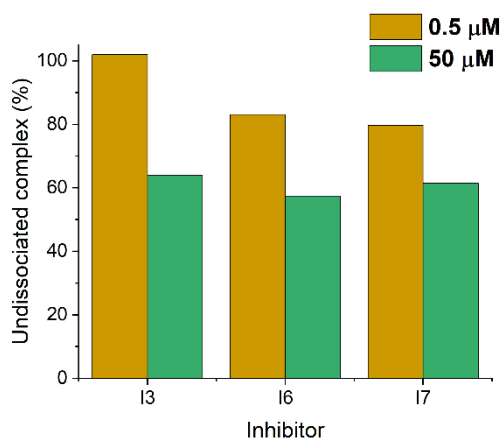


Figure 57. HTRF results for inhibitors **I3**, **I6**, and **I7**.

The general analysis of the data, collected during this section of the thesis, indicated that **I6** was the best candidate for further optimisation of the interactive surface. The **I6** inhibitor showed the highest thermal stability, with a T_m of 61 $^{\circ}\text{C}$, as well as the highest affinity for PD-L1 with a K_D of $31.1 \pm 1.7 \mu\text{M}$ and estimated IC_{50} of $\sim 50 \mu\text{M}$.

The Scaffold **I2** was located on the interacting surface of PD-L1, in the same manner as **I1**, and was minimised prior to the design of inhibitors with affinity for the PD-L1 protein, (Figure 58).

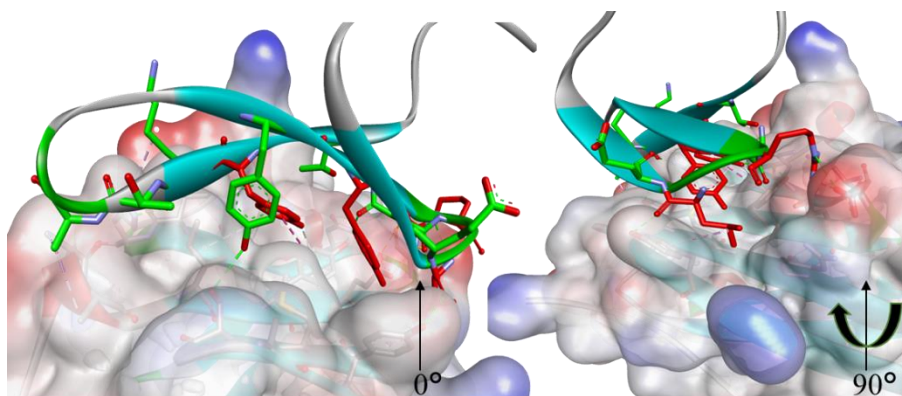
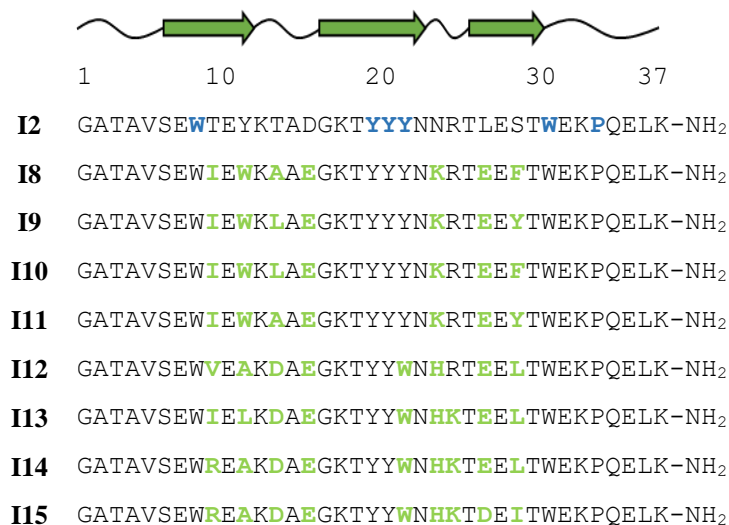


Figure 58. Superimposition of **I2** on the interactive region of PD-L1.

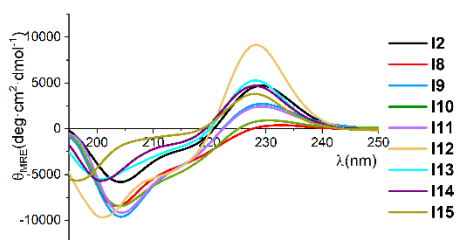
For the design, it was taken under consideration that aromatic amino acids are optimal residues for miniprotein-based PPI inhibitors, for two major reasons. First, they are great contributors to the stability of miniproteins. Second, the aromatic side chains provide excellent hydrophobic surfaces for interactions with the target protein. Following this assumption, two sets of inhibitors were obtained, (Table 22). The first group, inhibitors **I8-I11**, were designed around the periphery of four aromatic residues to increase binding affinity. Tyr19 and Tyr21, already present on the scaffold, together with Trp11 were kept in all four sequences. Positions 9 and 13 were mutated to increase favourable hydrophobic interactions (Ile, Leu, and/or Ala). Mutations Lys23 and Glu26 were intended to form salt bridges with PD-L1 Asp61 and Arg113 respectively. For the second group, inhibitors **I12-I15**, the design relied on two central aromatic residues Tyr19 and Trp21 flanked by three branched hydrophobic amino acids for **I12** and **I13** and two for **I14** and **I15**. Moreover, a set of polar residues to surround the central interactive surface by salt bridge formation with PD-L1 and to increase solubility was added.

Table 22. Sequences of inhibitors **I2** and **I8-I15**. The new mutations are highlighted in green, and crucial residues to the fold are in blue.

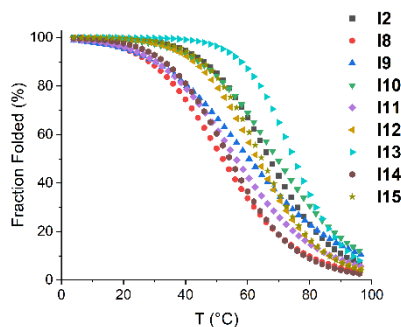


CD studies of inhibitors **I8-I15** reflected high similarity in CD spectra compared to the native scaffold. The maximum was present at 230 nm and two minima at 217 and 205 nm, (Figure 59). Major differences to the native fold could be found in **I15** with a tendency towards minima corresponding to a random coil. **I12-I14** had a shift on the minima, but smaller than **I15**. On the other hand, **I8** and **I10** had a clear loss of aromatic contributions, i.e., maxima at 230 nm. When comparing the sequences between **I8** and **I10** with **I9** and **I11**, we can assume that the changes on aromatic contributions are due to mutations at position 28. When Phe is placed at that position, the aromatic contributions drastically drop. However, in the presence of Tyr in position 28, those contributions remain close to the native structure.

A)



B)



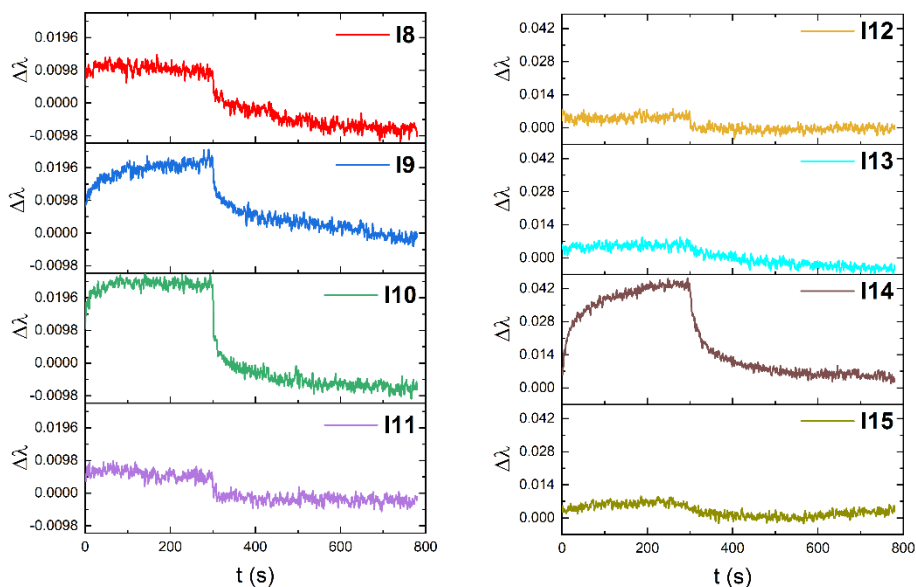
C)

Entry	T_m (°C)
I2	67.2 ± 0.5
I8	52.5 ± 1.6
I9	60.5 ± 2.1
I10	69.5 ± 3.3
I11	56.3 ± 1.6
I12	63.2 ± 0.5
I13	75.0 ± 0.3
I14	54.3 ± 0.6
I15	64.6 ± 0.4

Figure 59. A) CD spectra, B) Normalised thermal denaturation, and C) Estimated T_m values of inhibitors **I2** and **I8-I15**.

BLI measurements, (Figure 60), indicated that from the entire set, inhibitors **I8**, **I11-I13**, and **I15** had neglectable changes on the sensogram, for that reason the affinity values of the fitting were not considered during the analysis of results. For **I9-I10** and **I14**, the results of the experiment showed more promising profiles, with K_D values range in between 12-96 μM . When comparing **I9** and **I10**, it could be assumed that the presence of Tyr28, in **I9** had a significant impact on affinity, improving it close to 10-fold compared to **I10** which contained Phe28 instead. By comparing **I14** with the other set of mutants and its BLI results, it could be assumed that Arg9 in combination with Lys24 and Glu24, enhances the affinity of the inhibitor towards PD-L1.

A)



B)

Entry	Response (nm)	K_D (μM)
I8	0.0094	N/A
I9	0.0222	11.9 ± 0.7
I10	0.0247	96.5 ± 33.1
I11	0.0057	N/A
I12	0.0054	N/A
I13	0.0071	N/A
I14	0.0441	19.4 ± 0.6
I15	0.0054	N/A

Figure 60. BLI results for inhibitors **I8-I15**, A) BLI plotted sensograms, and B) Highest BLI response and estimated K_D values.

Using the HTRF inhibition assay, the best inhibitors from BLI were analysed at two concentrations, 50 and 0.5 μM . During the assay **I8** and **I13** were used as references to relate the results between BLI and HTRF assays. (Figure 61). The HTRF assay results did not agree with those observed for these inhibitors by BLI. Kinetic studies indicated that

I13 had no affinity for PD-L1, while by the HTRF assay the results were comparable between the 5 inhibitors studied, independently of the affinity values obtained experimentally. However, when focussing on the results of those inhibitors that had sufficient BLI response to obtain an adequate fit, **I9** showed the highest inhibition levels at 50 μM as it showed the lowest K_D .

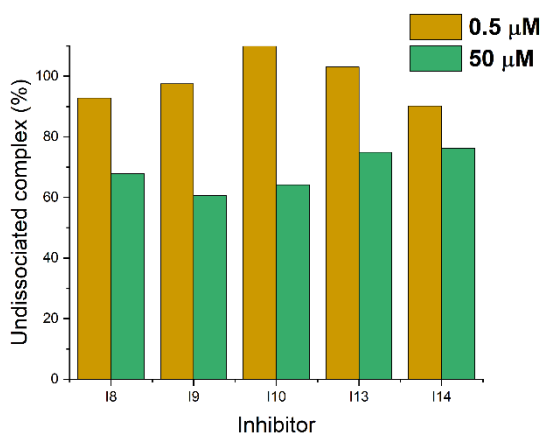


Figure 61. HTRF results for inhibitors **I8-I10**, **I13**, and **I14**.

Complete analysis of the experimental data in this thesis section led to the selection of **I9** for further optimisation. **I9** showed cooperative unfolding with a T_m value of 60°C, the highest affinity for PD-L1 with a K_D of 12 μM and an estimated $\text{IC}_{50} > 50 \mu\text{M}$.

3.2.1.1. Inhibitor optimisation based on optimised PD-1

Mimicry of binding epitopes is a common approach for inhibitor design. However, as mentioned in the previous section, we sought to design inhibitors with new binding modes. For that purpose, a single mutation of a surface-optimised PD-1, (Figure 62), was introduced in the sequences at different locations. The reference crystal structure used in this section was the optimised PD-1 in complex with PD-L1²²⁹, (PDB id: 5IUS). This optimised mutant of PD-1 displayed high affinity towards PD-L1, increasing the K_D by 35000-fold, from 8.4 μM to 120 pM. From the binding region of opt-PD-1, in contact with the targeted surface of PD-L1 by the inhibitors, mutations His64 and His68 were identified that could be of useful for optimising inhibitors **I6** and **I9**.

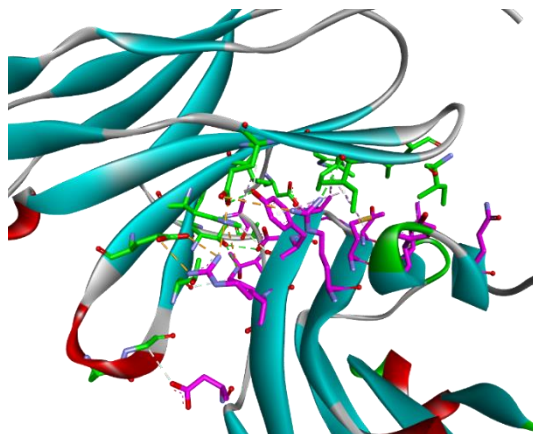


Figure 62. Optimised hPD-1 in complex with hPD-L1, (PDB: 5IUS), Highlighted in green interactive residues on the surface of hPD-1 and in magenta, interactive residues of hPD-L1.

His residue was tested as an affinity enhancer at different locations of the **I6** sequence, generating a total of six new sequences, **I16-I21**, (Table 23). The reason behind the selection of His, as the residue to include in the sequence was to minimise the hydrophobic nature of the interactive surface. Also, it was expected to minimise solubility problems and aggregation of the inhibitors. Moreover, for inhibitors **I17** and **I18**, residue Val26 was modified to Tyr26 to potentially cover solvent-accessible areas at the edge of the interaction. While Arg22 was included in inhibitors **I18-I21** for the formation of peripheral salt bridges with Glu60 and Asp61 of PD-L1.

Table 23. Entry, topology, and sequences of inhibitors **I6** and **I16-I21**. The mutations are highlighted in red, and charged mutations are in blue.

	1 10 20 30 35
I6	GLPPGWLEFKLSNGKTYYYWNAHTDTVTWTDPRASS-NH ₂
I16	GLPPGWLEFKLSNGKTYYYWNAHTHTVTWTDPRASS-NH ₂
I17	GLPPGWHEFKLSNGKTYYYWNAHTDTYTWTDPRASS-NH ₂
I18	GLPPGWLEFKLSNGKTYYYWNAR ⁺ TD ⁺ YTWTDPRASS-NH ₂
I19	GLPPGWLEFKLSNGKTYYYWNAR ⁺ HTVTWTDPRASS-NH ₂
I20	GLPPGWLEFKLSNGKTYYYHNAR ⁺ TD ⁺ VTWTDPRASS-NH ₂
I21	GLPPGWLEHKLSNGKTYYYWNAR ⁺ TD ⁺ VTWTDPRASS-NH ₂

The analysis of the CD spectra of inhibitors **I16-I21**, (Figure 63), reflected a high diversity at wavelengths below 220 nm. Such deviations in the CD spectrum could be related to the changes in the torsional angle of the β -sheet. It is known that β -sheets do not have a unique profile in CD and the response is directly related to the direction and torsional degree of the β -sheet²³⁵. The effects of the mutations over the ellipticity at 230 nm can be individually analysed and grouped by mutation. Introduction of Tyr28 shows a shift of the maxima for **I17** and **I18** when compared to **I6**. However, they do have a greater ellipticity suggesting an increase in aromatic contacts. Furthermore, the incorporation of His7 into inhibitor **I7** led to an increased T_m value of 3°C with reference to **I6**. This observation is supported by the negligible effects on stability observed for inhibitor **I8** in absence of His7. The magnitude of ellipticity at 230 nm was drastically affected by removal of aromatic residues of the β -sheet, as observed for the case of **I20** and **I21**, which modification not only led to a decrease in signal intensity, but a decrease in T_m value of 20°C. While it was possible to support most of the observed experimental results based on the models, the analysis of the data of inhibitors **I16** and **I19** was limited in the absence of a high-resolution structure.

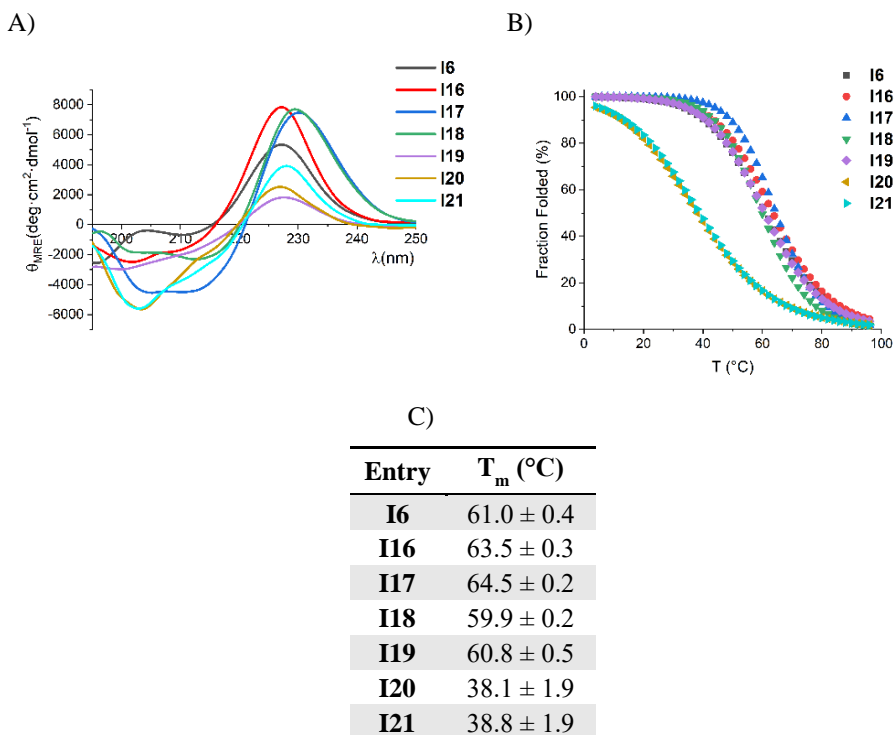
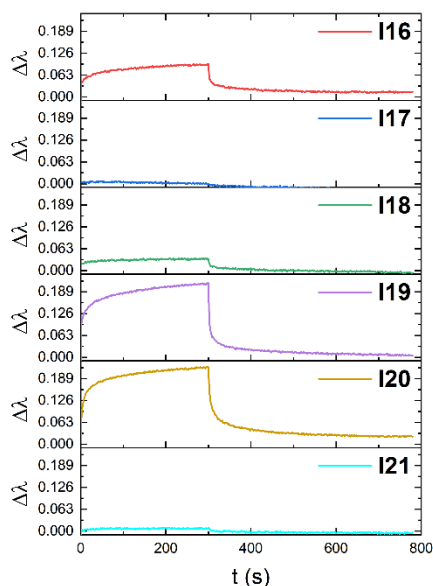


Figure 63. A) CD spectra, B) Normalised thermal denaturation, and C) Estimated T_m values of inhibitors **I6** and **I6-I21**.

BLI data showed that, as hypothesised, inclusion of the His residue can enhance, the affinity for PD-L1, (Figure 64). A 3-fold improvement was observed for inhibitor **I16**, with a K_D of 9 μM . However, in other cases, for instance inhibitors **I17**, **I18** and **I21**, it led to a lack of affinity. It is perhaps due to the removal of interacting hydrophobic residues. The binding energy between PD-1 and PD-L1 is dominated by a central hydrophobic cluster, so it is reasonable to assume that a minimal number of these residues should be present on the inhibitor surface. The comparison of **I16** and **I19** indicates that the presence of Arg24 enhances the sensogram response, but greatly decreases the affinity for PD-L1, from K_D of 9 to K_D of 620 μM . Additionally, the modification of Trp21 residue into His21, in **I20**, busted both affinity and response, which could be interpreted as a modification of the binding site of the inhibitor in reference to the theoretical model where Trp21 plays a key role in the interactions.

A)



B)

Entry	Response (nm)	K_D (μM)
I6	0.0738	31.1 ± 1.7
I16	0.0918	9.45 ± 0.5
I17	0.0009	N/A
I18	0.0344	N/A
I19	0.2113	620 ± 250
I20	0.2201	13.9 ± 0.6
I21	0.0074	N/A

Figure 64. BLI results for inhibitors **I6**, and **I16-I21**, A) BLI plotted sensograms, and B) Highest BLI response and estimated K_D values.

Only inhibitors with higher affinity for PD-L1 than **I6** were measured by HTRF, (Figure 65). The results in the assay had the same inhibition tendencies as those observed by BLI in terms of affinity. **I16** showed the strongest inhibition by HTRF, with IC_{50} of $\sim 50\mu\text{M}$, followed by **I20** and finally **I18**. Furthermore, the inhibitory capacity of **I16**

was increased with respect to **I6**, which confirmed that the use of optimised native proteins is a good approach for the design of inhibitors.

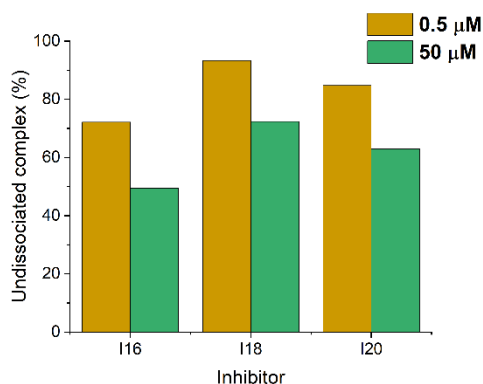
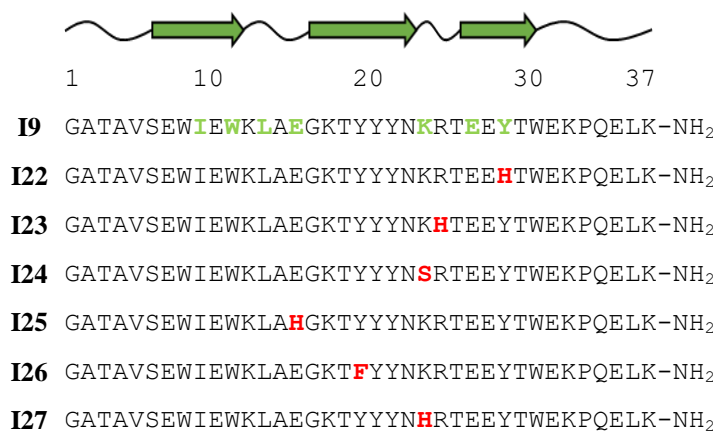


Figure 65. HTRF results for inhibitors **I16**, **I18**, and **I20**.

Following the same basis as for **I6**, a new set of sequences analogous to **I9** was obtained, generating inhibitors **I22-I27**, (Table 24). Except for **I24** and **I26**, which were used to provide information on crucial residues for the interaction.

Table 24. Sequences of inhibitors **I9** and **I22-I27**. The new mutations are highlighted red.



Analysis of CD data of mutants **I22-I27**, (Figure 66), reflected a general loss in aromatic contributions to the fold. The lower contribution had a great impact on the thermal stability of the inhibitors. From all data studied, **I22** and **I25** showed the closest spectra to the original sequence. However, in the absence of a high-resolution model, we must limit the analysis to assumptions if inhibitors **I23-I27** retain the β -sheet fold, since the minima

are above 200 nm, but with different β -sheet conformation in solution. The detailed analysis of the thermal stability versus the mutations indicates that from a structural point of view, positions 23 and 24 of the sequence, should contain a positively charged residue. The change of Lys23 residue to Ser23, for inhibitor **I24**, led to a sequence with no thermal stability and a drop on T_m value of ~ 6 °C when modified to His23 for inhibitor **I27**. The reduction in thermal stability was also pronounced for **I23**, with the modification of Arg24 residue to His24 and a T_m value lowered by ~ 25 °C. Based on the models, the presence of positively charged residues at positions 23 and 24 of the sequence may induce stabilisation of the β -sheet by generating electrostatic interactions with Glu26 of the third strand and Glu7 of the first strand, respectively. Modification of the Tyr triad in the middle strand, for instance, in **I26**, disrupted the stability of the entire β -sheet. Despite the similarity in the profile of **I22** and **I25** with **I9**, it seems possible to assume that for the first, the change of Tyr28 residue for His28, lowered the contacts between strands generating a ~ 15 °C penalty in the T_m value. Additionally, the removal of the electrostatic interaction on the second between Glu15 and Lys17 reflected a ~ 14 °C loss in the T_m value of the miniprotein.

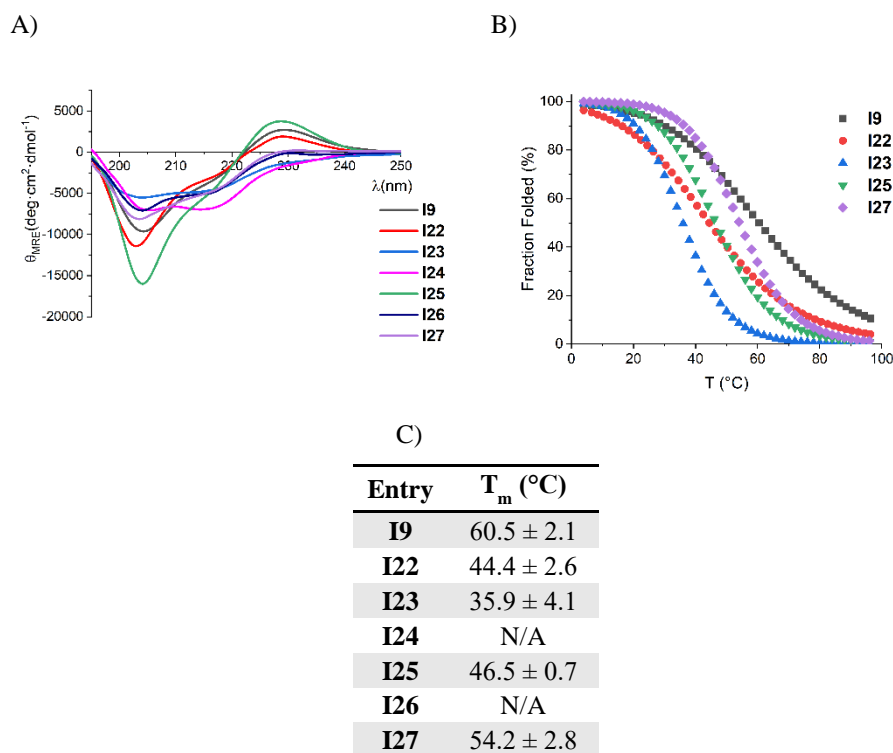
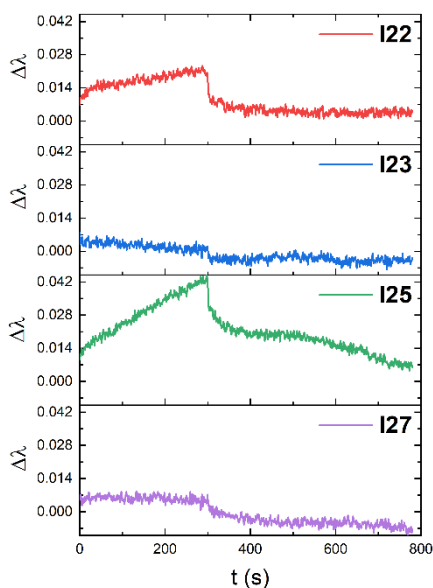


Figure 66. A) CD spectra, B) Normalised thermal denaturation, and C) Estimated T_m values of inhibitors **I9** and **I22-I27**.

BLI affinity studies of those inhibitors, that showed cooperative unfolding, revealed a low or no improvement in relation to **I9**, (Figure 67). Inhibitor **I22**, with a 2-fold enhancement of affinity, had a low response at 50 μM concentrations. These results confirm that a well-defined, cooperatively folded, three-dimensional structure is required for the adequate effectiveness of protein-protein interaction inhibitors.

A)



B)

Entry	Response (nm)	K_D (μM)
I9	0.0222	11.9 ± 0.7
I22	0.0209	6.9 ± 0.4
I23	0.0008	N/A
I25	0.042	13.3 ± 0.3
I27	0.005	N/A

Figure 67. BLI results for inhibitors **I9**, and **I22**, **I23**, **I25**, and **I27**, A) BLI plotted sensograms, and B) Highest BLI response and estimated K_D values.

From the analysed inhibitors, **I22** and **I25** were further studied by HTRF, (Figure 68). While for **I25** the affinity and inhibition levels were comparable to **I9**, showing no enhancement of the affinity by including the His residue in the sequence, **I22** indicate to have an inhibition capacity below 50 μM , a great improvement when compared to **I9**. When we come back to the model of the inhibitor complex with PD-L1 and the crystal structure of opt-PD-1 in complex with PD-L1, we can observe that the His mutation in **I22** is in close spatial agreement with His68 of opt-PD-1. Assuming that the modelling of the inhibitor is correct with the mode of binding, this could explain the improvement in inhibition observed by HTRF.

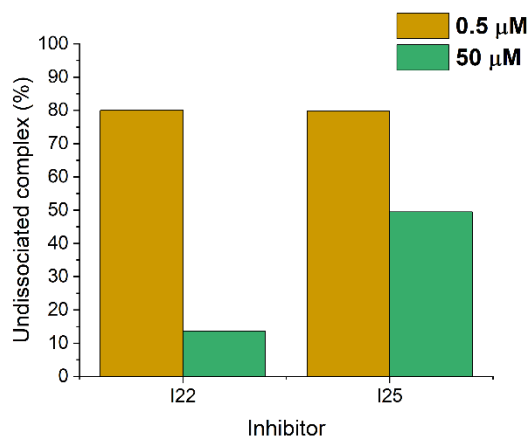


Figure 68. HTRF results for inhibitors **I22** and **I25**.

The general analysis of the results regarding mutants **I22-I27** concluded that the optimisation approach was adequate, as seen by HTRF with inhibitor **I22**. However, compared to **I6** mutants, **I9** displayed a poor capacity to integrate new mutations without effects on the miniprotein fold, limiting the possibility for further optimisation. These results, combined with the progress achieved in scaffold development with **I1**, led to discontinuation of research on **I2**.

3.2.2. Optimized inhibitors grafted into HEEE scaffold

Following the successful assembling of the WW-Prototype with a β -amino acid containing helix (section 3.1.1. of the present thesis) to obtain a novel HEEE scaffold, a set of inhibitors based on **I1**, were grafted into miniprotein **21**, (Figure 69).

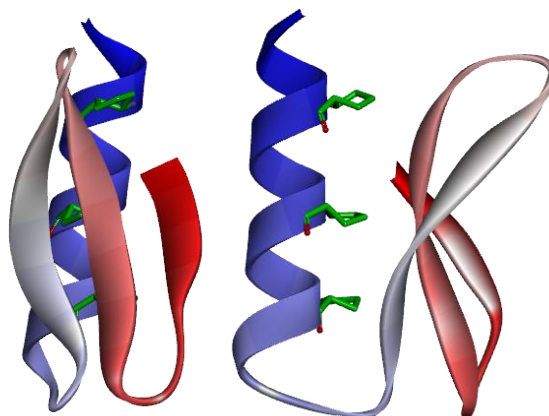
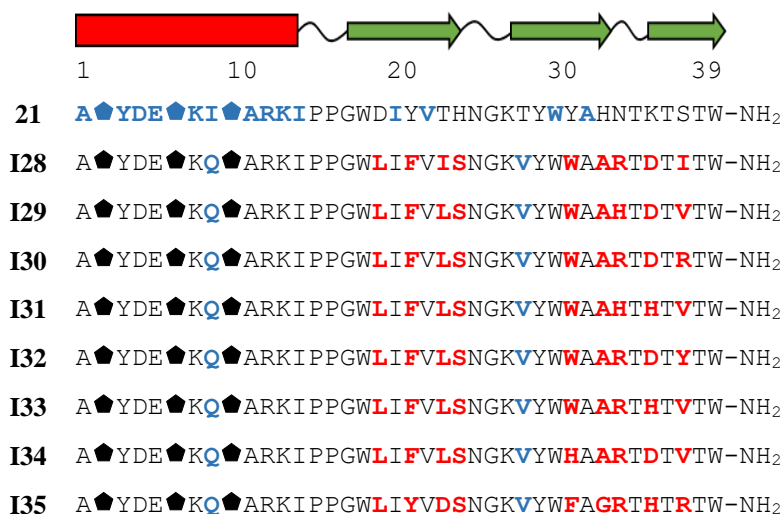


Figure 69. Scaffold topology, of miniprotein **21**, used for the drafting of the optimized inhibitors.

The control of the fold for miniprotein **21**, has been optimised through a hydrophobic core between the triple-stranded antiparallel β -sheet and the helix. For that reason, it is hypothesised that the inclusion of mutations in residues placed on the solvent exposed side of the β -sheet, may not significantly influence the stability of the miniproteins. We generated inhibitors **I28-I34** by grafting the active site of inhibitors **I5-I7**, **I16** and **I18-I20** in the sequence of miniprotein **21**. Additionally, inhibitor **I35** was designed using the combination of **I4**, **I7** and **I16**, (Table 25). The new inhibitors included Gln8 and Val21 to induce the proper orientation of the helix and increase the hydrophobic packing, respectively.

Table 25. Sequences of inhibitors **21** and **I28-35**. The grafted epitope is highlighted in red while structural enhancers are highlighted in blue.



Extension of the structure, insertion of constrained β -amino acids and modification of the topology had strong effects on the CD spectra and thermal stabilities of these new inhibitors. As observed by CD, the spectra of these new inhibitors differed from those of their analogous EEE inhibitors, (Figure 70). The profiles were more homogeneous between mutants now than before topology modification. This may be indicative that the scaffold utilised for this design has a well-defined three-dimensional structure and is preserved after the modification of the sequence. It is well known that a constrained β -amino acid-containing helix shows local minima in the CD spectrum, at approximately 210 nm. This characteristic minimum is found in all the inhibitors studied except for **I35**, which together with **I28**, did not show thermal stability. The largest

difference in shape and magnitude of the spectra of inhibitors **I28-I35** is on the signal at 230 nm, corresponding to aromatic contributions. The diminished ellipticity and presence of negative bands at 230 nm is theorised to be induced by the hydrophobic environment of the aromatic residues, as explained in Section 3.1. of the present thesis.

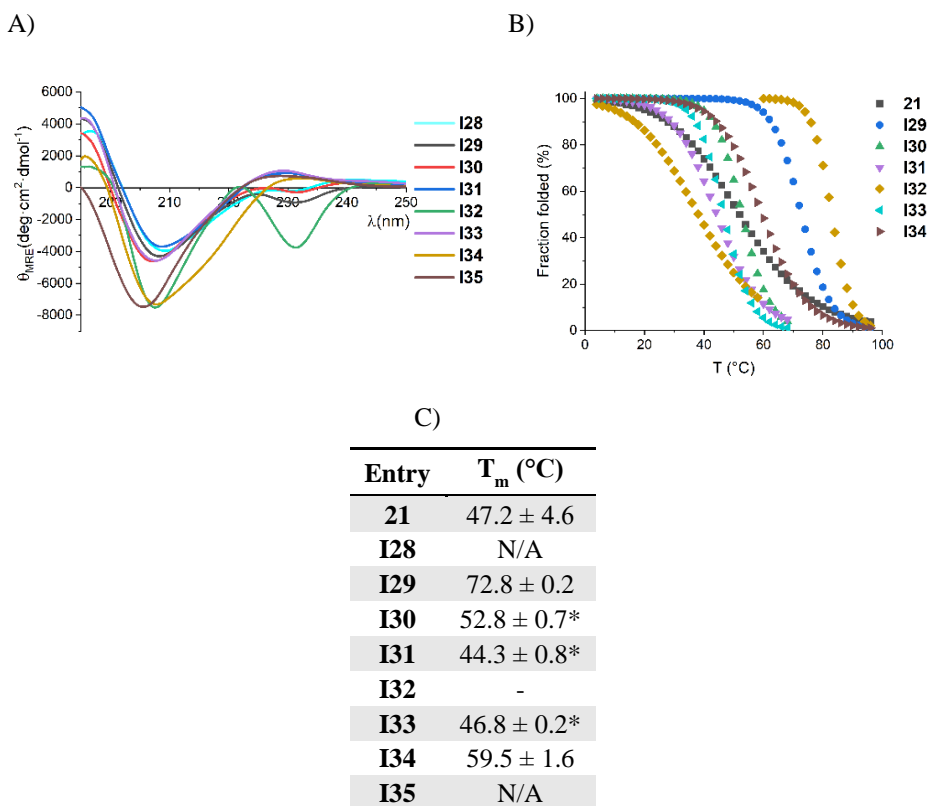


Figure 70. A) CD spectra, B) Normalised thermal denaturation, and C) Estimated T_m values of miniprotein **21** and inhibitors **I28-I35**. (*) Indicates that the T_m value is an approximation derived from the data collected prior to precipitation of the miniprotein.

Despite the exhaustive analysis of the CD data, these inhibitors showed a general tendency to aggregate during thermal denaturation studies, precluding an accurate determination of T_m values for inhibitors **I28** and **I30-I33**, (Figure 71A). The complexity in the design of extended inhibitors, for the PD-1/PD-L1 interaction, is the conflicting percentage of hydrophobic residues required for both the stabilisation of the fold and the activity toward PD-L1. Furthermore, the HEEE fold indicated that two stages of unfolding could occur. Such effect only could be observed on inhibitor **I32**, where two transitions were measurable and the peptide did not precipitate during studies, (Figure 71B). This

(as explained in Section 3.1.2) could be promoted by an independent unfolding of the present secondary structures.

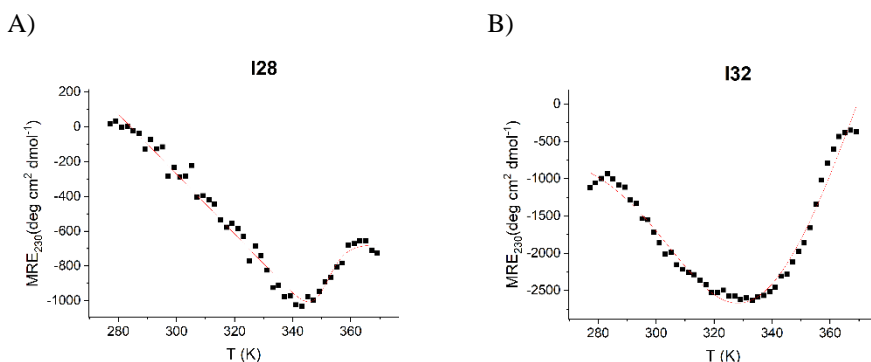


Figure 71. Recorded ellipticity versus T (K) for inhibitors **I28** and **I32**.

Affinity studies were performed for the set of inhibitors except for inhibitors **I28** and **I35**, which did not show thermal stability. In particular, the extension of the structure, greatly boosted the affinity and sensogram responses of these inhibitors. It is clearly visible, when comparing the sensograms of the different optimization steps of **I1** following the design of inhibitor **I6**, that the affinity towards PD-L1 increases with each optimisation step, (Figure 72). It can be assumed that the incorporation of the helical fragment may be beneficial in two ways: the increased structural order of the β -sheet required for adequate binding and the improvement of specificity when blocking the back side of the β -sheet.

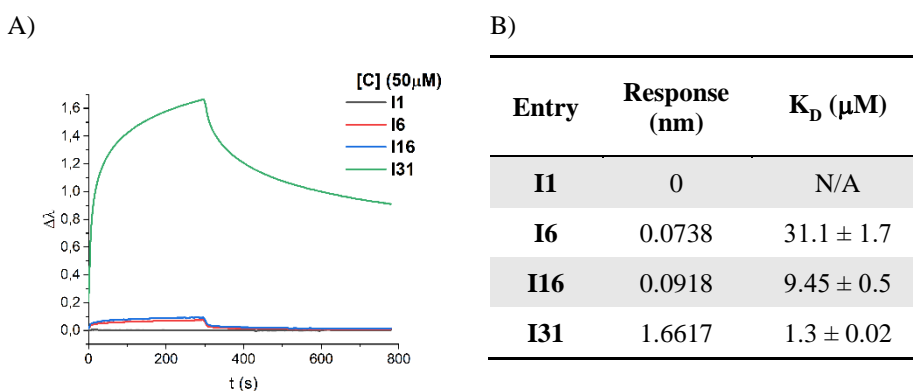


Figure 72. Comparative BLI results for miniprotein **I1**, **I6**, **I16**, and **I31**, A) BLI plotted sensograms, and B) Highest BLI response and estimated K_D values.

The improved response in BLI measurements made it possible to study the binding kinetics in a dose-response manner, and to obtain a global fitting for a more accurate estimate of the K_D , (Figure 73). The extension of the topology showed an enhancement of 2-fold for **I31** and up to 400-fold for **I33** in reference to their EEE analogues **I16** and **I19** respectively. These results support the requirement of a well-defined three-dimensional structure for adequate performance of PPI inhibitors. The interactive epitopes of these inhibitors differ in a maximum of three residues between sequences. Such similarities allow for the analysis of the individual contributions of the mutations based on their K_D values. Taking inhibitor **I30** as the reference, due to its lowest $K_D = 1.4 \pm 0.01 \mu\text{M}$, we can observe that those other two inhibitors with K_D value below $5 \mu\text{M}$ contain Arg33 mutation, suggesting that this residue plays an important role on the binding capacity to PD-L1. The other inhibitor with Arg33 mutant, inhibitor **I32**, differs from **I30** by one residue, where Arg37 is modified to Tyr37 and has a four-fold less affinity for the target, suggesting that Arg37 is a suitable residue for binding in that position. This is supported by the other sequences that contain Val37 instead and show no major differences on the binding. Moreover, comparison of inhibitors **I30** and **I34**, suggested that Trp30 enhances up to two-fold the affinity for PD-L1 as K_D drop when modifying it to His30 in inhibitor **I34**. However, it does not represent a crucial interactive residue as its absence retains the affinity at micromolar levels.

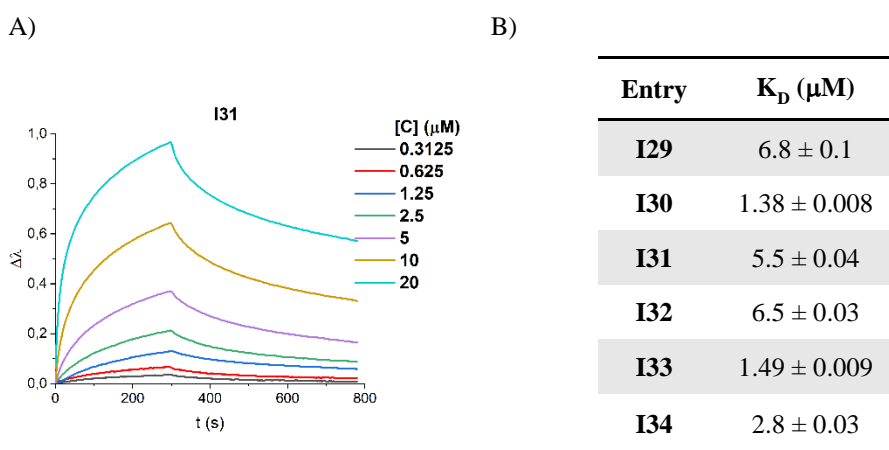
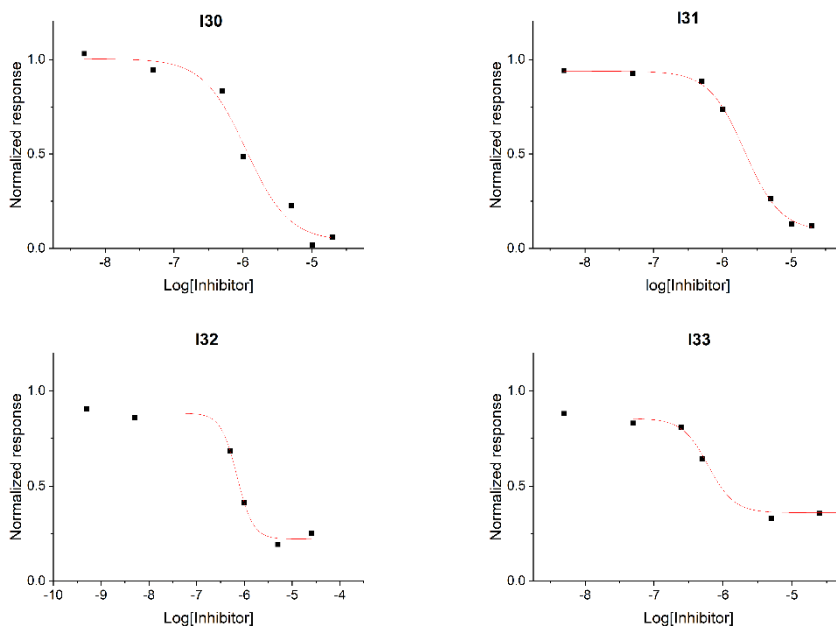


Figure 73. BLI results, A) Dose-dependent biosensorgrams of inhibitor **I31**, and B) Estimated global K_D values for inhibitors **I29-I34**.

We sought to obtain the IC_{50} of the five best binders, **I30-I34** using HTRF assay.. The inhibitors were studied at concentrations ranging from 50 μ M to 5nM. Interestingly, in most of the inhibitors studied, PD-L1 precipitated at higher concentrations of inhibitors, preventing adequate estimation of the IC_{50} . Such a situation is a common false positive for HTRF studies. Such phenomena are possible to detect when the fluorescence intensity for 620 nm decreases. Fluorescence at 620 nm corresponds to the donor whose fluorescence capacity is independent of the acceptor; for that reason, the fluorescence intensity of the donor should not change along the experiment. Although three of the five inhibitors studied induced precipitation of PD-L1, **I30** and **I31** showed the complete inhibition curve. Interestingly, these two inhibitors showed an IC_{50} of 1-2 μ M, stronger than that of the PD-1/PD-L1 complex, estimated to be approximately 4 μ M²²⁸. Such results indicate that the presented inhibitors not only inhibit the PD-1/PD-L1 complex, but competitively antagonise PD-1, (Figure 74).

A)



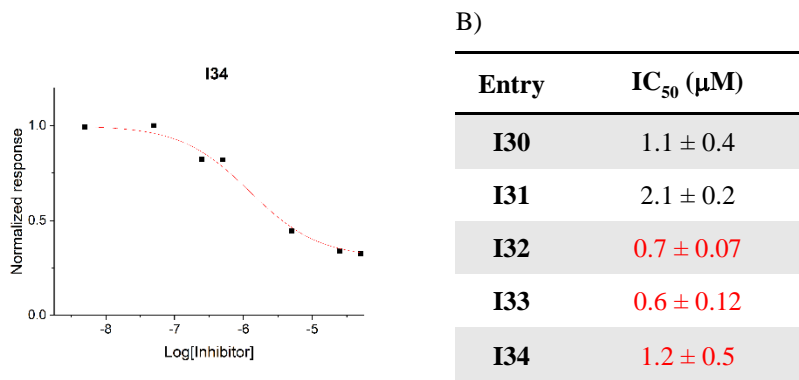


Figure 74. HTRF results for inhibitors **I30-I34**. A) Inhibition curves from HTRF experiments, and B) IC₅₀ estimations based on HTRF results, highlighted in red are those experiments where the results are affected by false positives.

To further research the inhibitory properties of these HEEE inhibitors, our collaborators (Katarzyna Magiera-Mularz, Bogdan Musielak, and Lukasz Skalniak, Jagiellonian University) studied inhibitor **I33** affinity for PD-L1 by NMR and inhibitory capacity in cells with the PD-1/PD-L1 Blockade Bioassay (Promega). However, we were unable to obtain any reasonable data from these experiments. The experimental conditions required for the protein were not suitable for the inhibitor, causing precipitation of the protein and the inhibitor when they dissolve in the NMR buffer. On the other hand, the maximum solubility of **I33**, for the cell-based assay, was too poor to reach concentrations capable of triggering T cell activation.

3.2.3. Inhibitor modification to the EHEE fold

Compactness and stability of a miniprotein-based inhibitor is crucial for the activity. Specially in those cases where the miniprotein fold is directed by noncovalent interactions and a hydrophobic core. Following the systemic exploration of sequence-fold relationship with the activity, we decided to implement the scaffold obtained by circular permutation (of the section 3.1.2 of this thesis), where the HEEE fold of miniprotein **21** was modified to the EHEE fold of miniprotein **25**, (Figure 75).

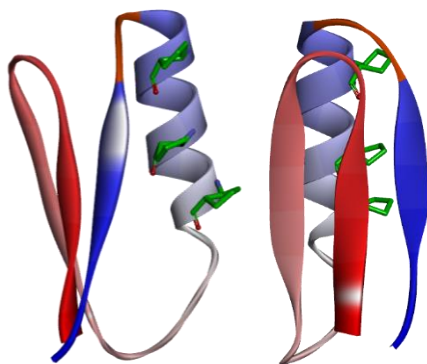


Figure 75. Scaffold topology, of mini-protein **25**, used for the grafting of the optimized inhibitors.

Likewise, as in section 3.2.2, some of the most promising active sites were selected to be grafted into the outer face of the newly designed mini-protein **25**. This approach led to the inhibitors **I36-I39**, which are based on **I30**, **I32-I34**, respectively, (Table 26).

Table 26. Sequences of mini-protein **25** and inhibitors **I36-I39**. The grafted epitope is highlighted in red while structural enhancers are highlighted in blue.

	1 10 20 30 40
25	KTSTWATGTNTA●YDE●KQ●ARKIPPGWDIYVTHNGKTYWYH-NH ₂
I36	D RT WATGTNTA●YDE●KQ●ARKIPPGW LIFVLS NGK VYWWA -NH ₂
I37	D TY WATGTNTA●YDE●KQ●ARKIPPGW LIFVLS NGK VYWWA -NH ₂
I38	H TV WATGTNTA●YDE●KQ●ARKIPPGW LIFVLS NGK VYWWA -NH ₂
I39	D TV WATGTNTA●YDE●KQ●ARKIPPGW LIFVLS NGK VYWHA -NH ₂

CD spectra of the EHEE inhibitors show a tendency to change the ellipticity to negative values at 230 nm, except for **I39**, (Figure 76A). We hypothesised that as the ellipticity at 230 nm is induced by aromatic contributions, the absence of Trp41 in inhibitor **I39**, may be the responsible of the absence of ellipticity at 230 nm and lack of thermal stability. The inhibitors analysed, by thermal denaturation, showed that topology modification improved thermal stability, with T_m values above 96 °C. Moreover, the structure showed no changes on the CD spectra when measured at 25 °C, at 96 °C and then again at 25 °C, (Figure 76B). We presume that the combination of aromatic side-chains,

located in the solvent exposed region of the β -sheet, with a network of hydrophobic interactions generates a hydrophobic region that acts as a second hydrophobic core. The high hydrophobic contribution of the fold could explain the strong thermal stability of this inhibitors, where a high entropic penalty exists for the unfolded state. This hydrophobic solvent-exposed region may be responsible for the low susceptibility, of this inhibitor, to guanidine hydrochloride. Guanidine destabilises protein fold by interrupting intra-H bonds. For the guanidine studies, stock solutions of 100 μ M of each inhibitor were prepared in potassium phosphate buffer pH 6 and incubated, for half an hour, at different concentrations of GuHCl from 0 to 7 M. The incubated samples were then thermally denatured using NanoDSF from 15 to 110 $^{\circ}$ C. The results reflected excellent stability properties of the inhibitors with T_m values up to 75 $^{\circ}$ C at 7 M guanidine hydrochloride, (Figure 76C).

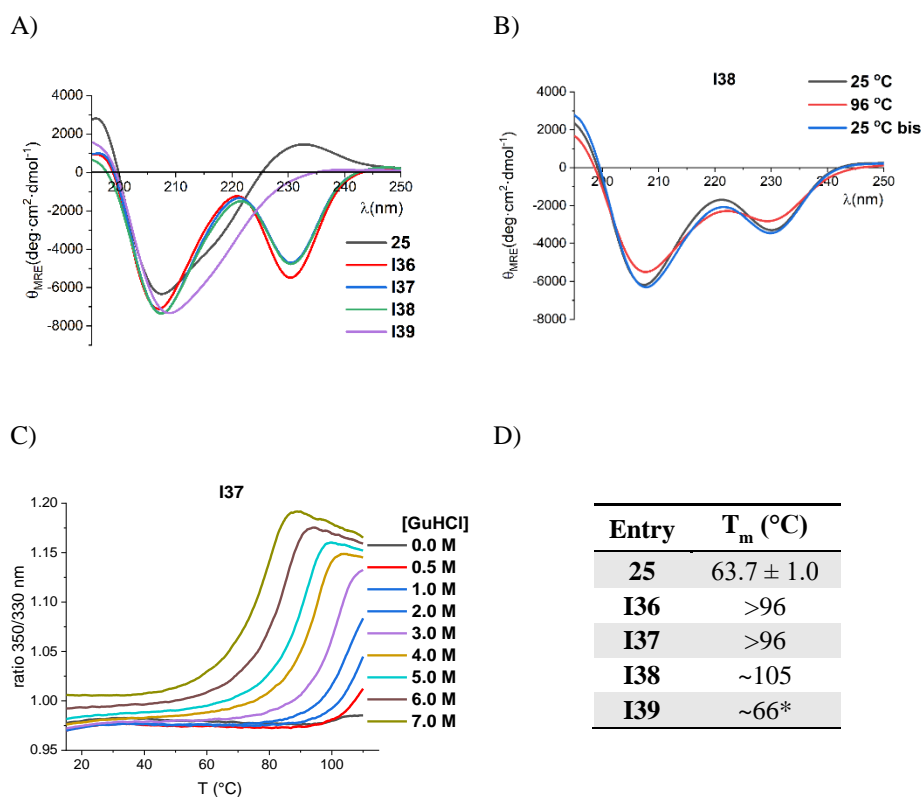


Figure 76. A) CD spectra of miniprotein **25** and inhibitors **I36-I39**, B) CD scans of **I38** at 25, 96 and again at 25 $^{\circ}$ C, C) Thermal denaturation curves by NanoDSF at different concentrations of Guanidine Hydrochloride from inhibitor **I37**, and D) Estimated T_m values of inhibitors **I1** and **I3-17**.

The topology modification had virtually no improvement on the binding capacity of the optimised surfaces, (Figure 77). From this information, we can deduce that the structure of the β -sheet prevails after the permutation to the EHEE fold. Moreover, in addition to the retained affinity for the target and higher stability, the new topology drastically improves the solubility of the sequences. When comparing the sequences, we can observe that the differences in interactive residues are in the first strand of the inhibitors, except for **I39**, which has His41 instead of Trp. Regardless of this, it is important to note that the lower estimated K_D are for those two inhibitors with Val3, suggesting that this residue plays an important role in the interaction. Also, for Trp41, one will expect to have a great impact on the binding capacity since it is buried in a hydrophobic pocket between Tyr39 and Met98 from PD-L1, (Figure 78).

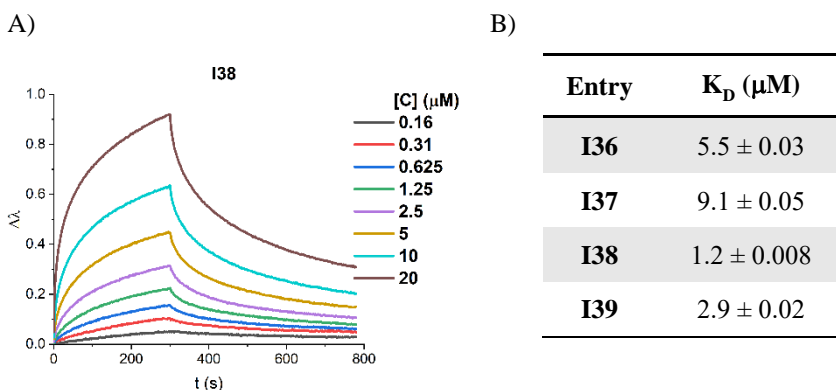


Figure 77. BLI results, A) Dose-dependent sensograms of inhibitor **I38**, and B) Estimated global K_D values for inhibitors **I36-I39**.

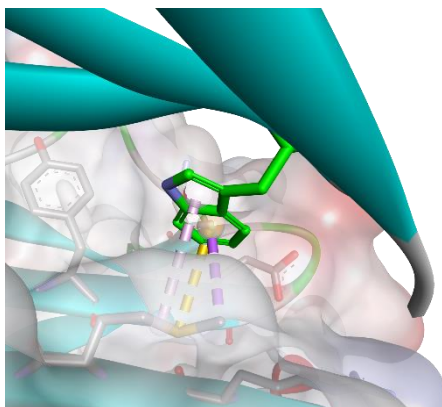


Figure 78. Binding mode of Trp41 to PD-L1 from the theoretical model of **I38**.

From this series, inhibitors **I38** and **I39**, were tested in the HTRF assay at two concentrations, 5 and 10 μM , (Figure 79). The inhibitory capacity of these inhibitors was in agreement with the results observed by BLI, where **I38** had a higher inhibitory capacity than **I39**. Moreover, no false positive or protein precipitation was observed at higher concentrations.

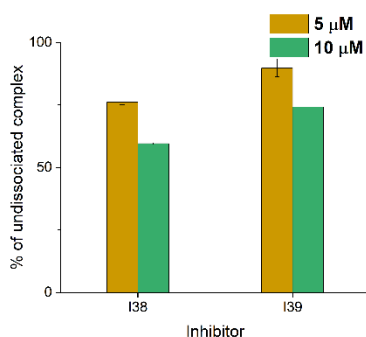


Figure 79. HTRF results for inhibitors **I38** and **I39**.

Due to the better physicochemical properties of **I38**, with reference to previously reported inhibitors (Higher solubility, thermal hyperstability and binding affinity), we were able to perform an immune checkpoint blockade cell-based assay*. Prior to the experiment, a 2 mM stock solution in water of the inhibitor was prepared. For late adjustments of the concentration in the assay, an aliquot of 100 μM concentration was prepared by diluting the stock in 8M GuHCl. Following an incubation of 30 minutes, the absorbance of the sample was measured at 280 nm and the concentration of the stock adjusted if necessary. Considering the presence of water in the inhibitor dilutions with the cell culture medium, a negative control, a blank and a positive control with 2.5 $\mu\text{g/mL}$ of Durvalumab were prepared for each experiment with the same amounts of water as present in the final wells of the inhibitor. For the first experiment, two concentrations of the inhibitor, 10 and 50 μM , were tested. The procedure indicates to premix 40 μL of the inhibitor with the PD-L1 aAPC/CHO-K1 Cells and immediately after adding 40 μL of suspended Jurkat T cells. The mixture is then incubated for 6 h at 37 $^{\circ}\text{C}$ and 5 % CO_2 . After incubation, the plate was equilibrated at room temperature for 10 minutes prior to the addition of Bio-Glo reagent

* In cooperation with Dr Łukasz Skalniak from the Department of Organic Chemistry at the Faculty of Chemistry from Jagiellonian University.

(Promega). The luminescence was determined after another 20 min of incubation. For the first assay, we could observe the fold induction of **I38**. However, the promoted induction was very low, (Figure 80).

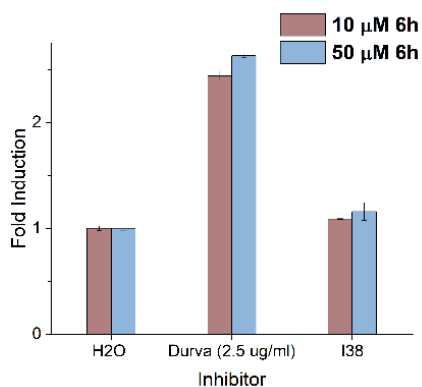


Figure 80. Immune checkpoint blockade cell-based assay results for inhibitors **I38** versus water and Durvalumab as negative and positive control.

Knowing the data from BLI and HTRF, we knew that the measurement readings were far from the real capacity of **I38**. In the second experiment, we decided to allow the inhibitor to bind to PD-L1 by incubating the inhibitor at a concentration of 50 M for 24 h before adding Jurkat T cells and later incubation for 6 h. As a control, we simultaneously reproduced the first experiment to be able to compare whether the longer incubation had an effect and if the initial positive result could be reproduced, (Figure 81). As expected, a longer incubation time resulted in increased induction, suggesting a possible slow kinetics of binding of the inhibitor to the PD-L1 surface.

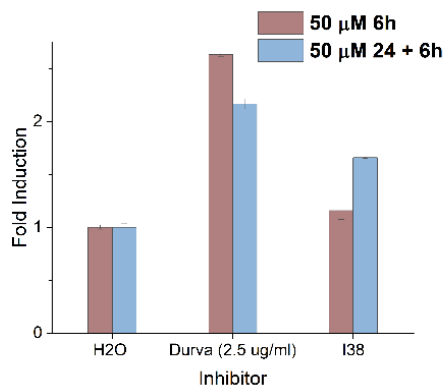


Figure 81. Immune checkpoint blockade cell-based assay results for inhibitors **I38** versus water and Durvalumab was negative and positive control. Comparison of incubation of 6 and 24 + 6 h.

With adequate conditions in hand, we prepared an inhibition curve with seven-point concentrations of **I38** by double dilutions with final concentrations ranging from 0.625-40 μM. The inhibitor was premixed 24 h with APC cells before adding Jurkat T cells for another 6 h of incubation. To estimate the maximum induction for the adjustment, wells containing 2.5 μg/mL Durvalumab were prepared for each inhibitor concentration by adjusting the final water concentration in the wells. The curve was fitted in the range of 0.1-700 μM for better visualisation, (Figure 82). The results with **I38** indicated that the design of an effective binder with a novel interactive surface and fold was successful with an EC_{50} of 27.4 ± 1.1 μM.

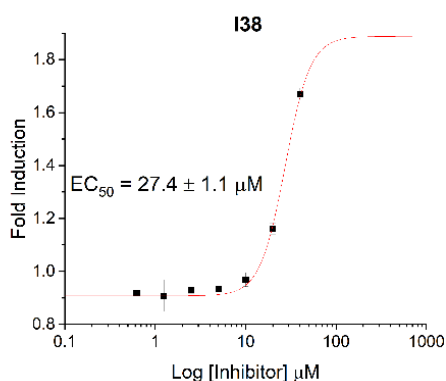


Figure 82. Results of the immune checkpoint blockade cell-based assay of inhibitors **I38** at different concentrations for the estimation of EC_{50} .

3.2.4. Hybrid design over optimized EHEE

Mimicry has been proven to be one of the most successful approaches for inhibitor design^{202,228}. One major requirement for the mimicry of interactive surfaces is the inherent folding stability of the selected scaffold. For those miniproteins which fold are driven by noncovalent interactions, the direct mimicry of surfaces into the sequence may lead to unstable conformational stability and, consequently, the loss of any potential activity towards the target. For that reason, in the present dissertation, the complete mimicry of optimised PD-1 was not possible without the introduction of modifications intended to first generate affinity for the target, PD-L1, and second help the prevalence of the EHEE fold. However, the design of the EHEE miniproteins was focused on the stabilisation of the topology through a network of hydrophobic interactions in the core. Such approach, intended to minimise the dependence of the stability on solvent exposed residues of the β -sheet, minimising the potential loss of thermodynamic stability of the inhibitors, (Figure 83).

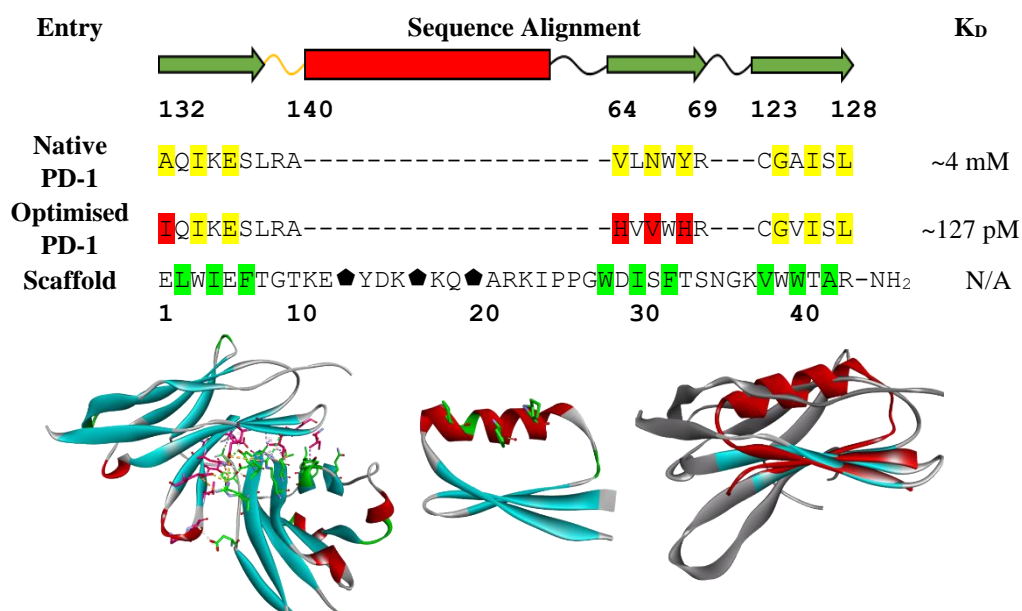
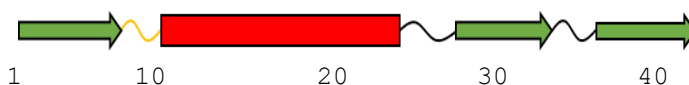


Figure 83. Sequence alignments of hPD1, optimised hPD-1 and the Scaffolds miniprotein **46**. The interactive residues of the PD-1 region of interest are highlighted in yellow, the mutated residues of the optimized PD-1 corresponding to the region of interest are in red, while the residues responsible for the fold of the scaffold are in green.

The resulting miniprotein, **O_I1**, was able to accommodate 60% of the optimised interactive residues of PD-1 extracted from the superimposition of both proteins. Together with mutants **I2-I8**, (Table 27). Trp38 and Tyr42 residues were necessary for the generation of contacts between the second strand and the first and third. We hypothesised that these hydrophobic aromatic residues could improve the stability and affinity of the scaffold towards PD-L1. Glu136 residue of Opt-PD-1 was placed on the third strand of the scaffold, mutation Glu16. This location was in proximity in the model to Arg125 residue of PD-L1, implying that it could have the same contribution as Glu136. To remove potential electrostatic repulsion between first and third strand, Glu5 residue was modified to Thr5, yet it was recovered with **O_I3** to confirm the hypothesis. Additionally, His7 mutation was introduced to provide a contact with Glu36 residue. Following this design, a series of mutants were generated based on the theoretical model to explore the possibility of enhancing inhibition by generating new contacts.

Table 27. Sequences of miniprotein **46** and inhibitors **O_I1-O_I8**. New mutations are highlighted in red.



46	ELWIEFTGTNK●YDE●KQ●ARKIPPGWDISFTHNGKVWWTAR-NH ₂
O_I1	ILVIT TFH GTNK●YDE●KQ●ARKIPPGW HIVFHSNGEVVW I AY -NH ₂
O_I2	ILVIT FH GTNK●YDE●KQ●ARKIPPGWHIVFHSNGEVVW IAF -NH ₂
O_I3	NL VIT EFH GTNK●YDE●KQ●ARKIPPGWHIVFHSNGEVVW IA Y-NH ₂
O_I4	NL VIT FH GTNK●YDE●KQ●ARKIPPGWHIVFHSNGEVVW IA Y-NH ₂
O_I5	NL VIT FH GTNK●YDE●KQ●ARKIPPGWHIV FYS NGEVVW IA Y-NH ₂
O_I6	ILVIT FH GTNK●YDE●KQ●ARKIPPGWHIV FPS NGEVVW IA Y-NH ₂
O_I7	ILVIT FH GTNK●YDE●KQ●ARKIPPGWHIVFHSNGEVVW IAW -NH ₂
O_I8	I TV IT FH GTNK●YDE●KQ●ARKIPPGWHIVFHSNGEV TW IA Y -NH ₂

The modification of an elevated number of residues from the β -sheet was reflected a change on the CD spectra of these inhibitors, (Figure 84). The lower molar ellipticity at 209 and 220 nm suggests a less native-like fold. Despite the differences in CD spectra in comparison with **46**, they are similar within the set of inhibitors the set of inhibitors, indicating a similar fold. Major changes of the profile are observable for **O_I5** and **O_I8**, and it is hypothesised that the addition of Tyr32 mutation close to the β -turn of the

β -hairpin. Also, the removal of Trp38 residue from the third strand may destabilise the entire fold. The thermal denaturation studies of these inhibitors proved to be challenging because of the aggregation as a consequence of the unfolding, limiting the capacity to estimate T_m values.

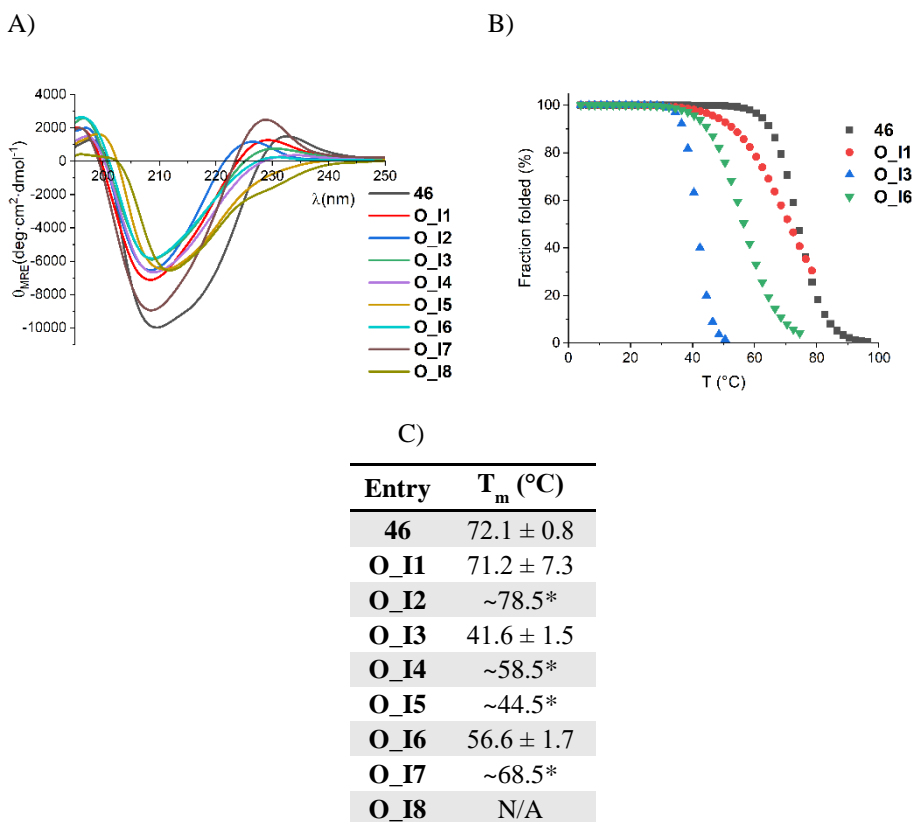


Figure 84. A) CD spectra of miniprotein **46** and inhibitors **O_I1-O_I8**, B) Normalised thermal denaturation plots of miniprotein **46** and inhibitors **O_I1**, **O_I3**, and **O_I16**, and C) Estimated T_m values of miniprotein **46** and inhibitors **O_I1-O_I8**. (*) Indicates that the T_m value is an approximation derived from the data collected prior to precipitation of the inhibitors.

In the same way as section 3.2.2 of the present dissertation, the combination of a hydrophobic core with a hydrophobic surface led to the aggregation of the inhibitors during thermal denaturation studies, (Figure 85). However, the presence of secondary structures, observable on the CD profile, together with the observed aggregation at higher temperatures, due to structural changes, are evidence of a folded state. For that reason, the T_m values were determined in two ways, a) For those inhibitors where partial transition was observable, the data was fitted up to that point, e.g. **O_I16**, (Figure 85D). b) For those

inhibitors that the transition was not possible to discern, the T_m was estimated to be at the temperature where the aggregation process started, indicating partial unfolding, e.g., **O_I7**, (Figure 85B).

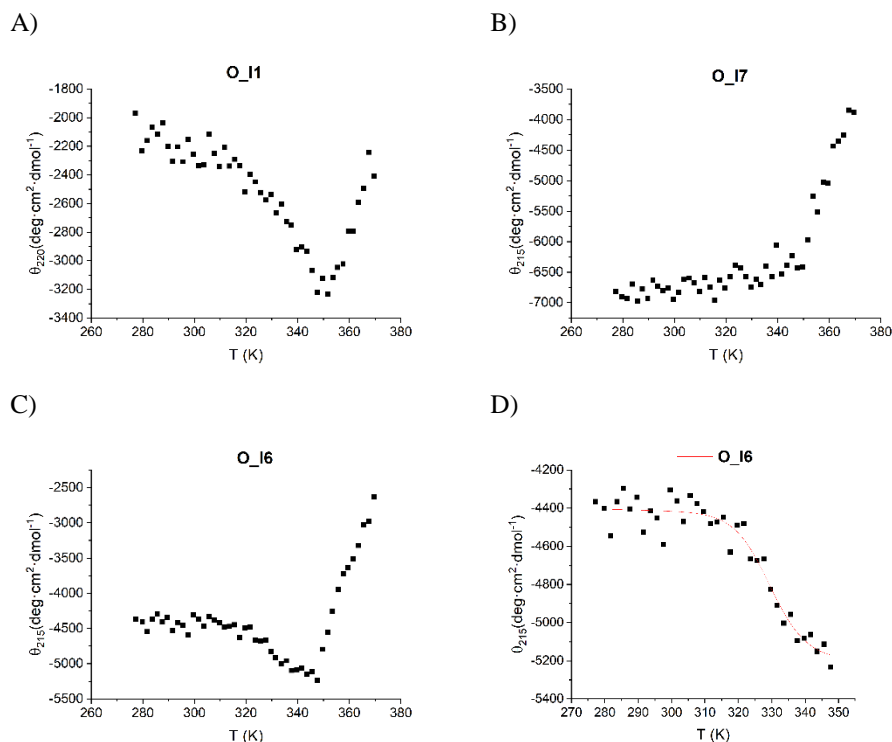
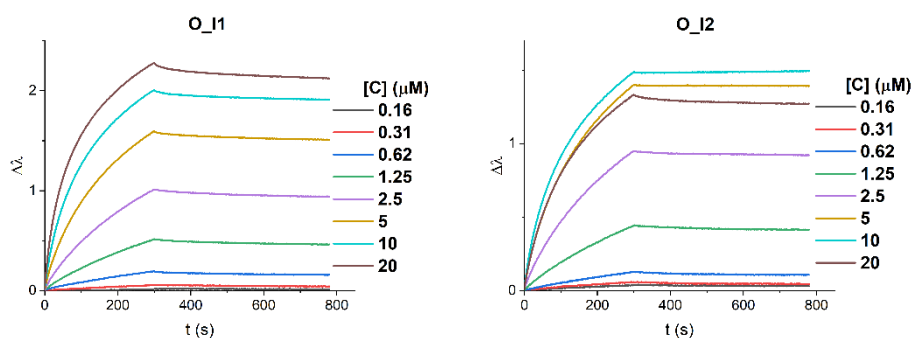


Figure 85. Recorded ellipticity versus T (K) for inhibitors, A) Inhibitor **O_I1**, B) Inhibitor **O_I7**, C) Inhibitor **O_I6**, and D) Fitting of the thermal denaturation data of inhibitor **O_I6** collected below the aggregation point.

BLI studies were performed in a concentration range of 0.156 to 20 μM , in the kinetic buffer. As expected, mimicry of optimised PD-1 led to inhibitors with nanomolar levels of affinity towards PD-L1. Interestingly, those inhibitors that were active during the assay showed almost irreversible association of the complex with PD-L1, (Figure 86). This behaviour could be explained by the hydrophobic character of the interaction, as when both surfaces are in contact with each other, they bury a large solvent-accessible area. Such an effect, as the stabilising hydrophobic core of the inhibitors, may have an entropic penalty above the energy required to dissociate them. Additionally, the affinity studies shed light on crucial residues for the interaction, reinforcing the concept of a well-optimised surface for the adequate binding between inhibitor and target. Substituting Ile1 residue by Asn1 in mutants **O_I3**, **O_I4**, and **O_I5**, led to complete loss of affinity for the target,

independent of other mutations. The same effect was observed after the removal of Trp38 residue, in mutant **O_I8**. As for **O_I6** the substitution of His32 residue, mimic of His68 of optimised PD-1, considerably reduced the affinity for PD-L1. Moreover, during the measurements it was possible to observe a tendency of the inhibitors to form aggregates with time. The aggregation was translated in a lower response of the sensogram for concentrations twice or higher than the previous measured steps of the experiment, e.g., **O_I2**, (Figure 86). Such behaviour of the inhibitors led to inadequate estimations of binding kinetics for some of the miniproteins. Such aggregation was more pronounced, on mutants **O_I2** and **O_I7**, where Tyr42 residue was substituted by more hydrophobic aromatic residues, Phe42 and Trp42, respectively. Also, the aggregation effect was more in mutant **O_I6**, where the modification of His32 residue into Pro32 may induce a higher order of flexibility of the fold and increase its susceptibility to aggregate.

A)



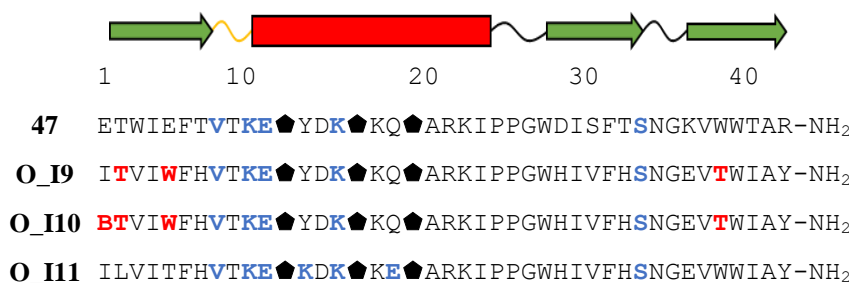
B)

Entry	Max [C] (μM)	K_D (nM)
O_I1	20	122 ± 2.8
O_I2	5	404 ± 9.2
O_I3	-	N/A
O_I4	-	N/A
O_I5	-	N/A
O_I6	10	842 ± 19
O_I7	5	381 ± 13
O_I8	-	N/A

Figure 86. BLI results for inhibitors **O_I1-O_I8**, A) Dose-dependent sensograms, and B) Maximum measurable concentration and estimated global K_D values.

To complete the set of inhibitors, we explored further possible interactive surfaces and effects of folding stabilities over the affinity by designing inhibitors **O_I9-O_I11**, based on **47**, (Table 28). For inhibitors **O_I9** and **O_I10**, Trp38 residue was shifted to position 5, transferring it from the first to the third strand, (Figure 87), and norleucine was used instead of isoleucine for **O_I10** to explore aggregation effects. When the aromatic residue was moved, it was expected, as observed from the theoretical model, that the binding will be more compact, allowing for an increase in the number of interactions with the target. While **O_I11**, was a mutant of **O_I1** with the structural features of **47** and an extra electrostatic interaction in the helix.

Table 28. Sequences of miniprotein **47** and inhibitors **O_I9-O_I11**. New interactive residues are highlighted in red, and structural enhancers are in blue.



B = Norleucine

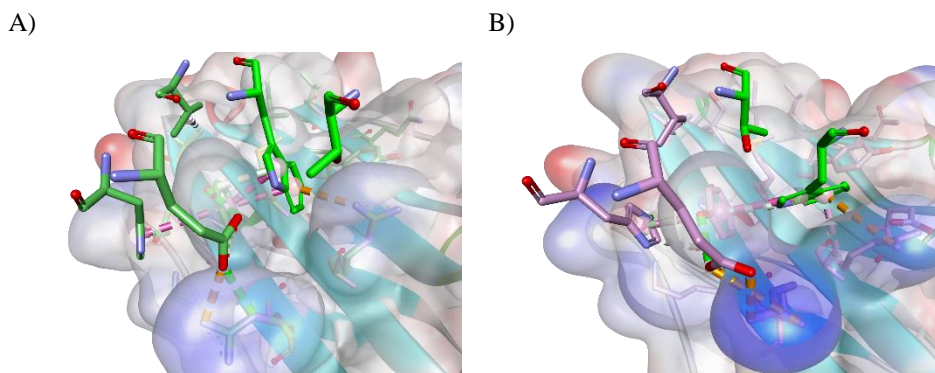


Figure 87. Comparison of the binding mode between **O_I9** and **O_I11**.

The CD spectra showed a similar secondary structure content of the inhibitor when compared to miniprotein **47**, (Figure 88). The modification of the Trp38 to position 5, from the third to the first strand for inhibitors **O_I9** and **O_I10**, had an impact on aromatic contributions to the fold. This was also reflected in the lower thermal stability of the fold, as precipitation was observed earlier than for **O_I1**. The modifications did not show any improvement in the aggregation derived from the thermal denaturation of these inhibitors.

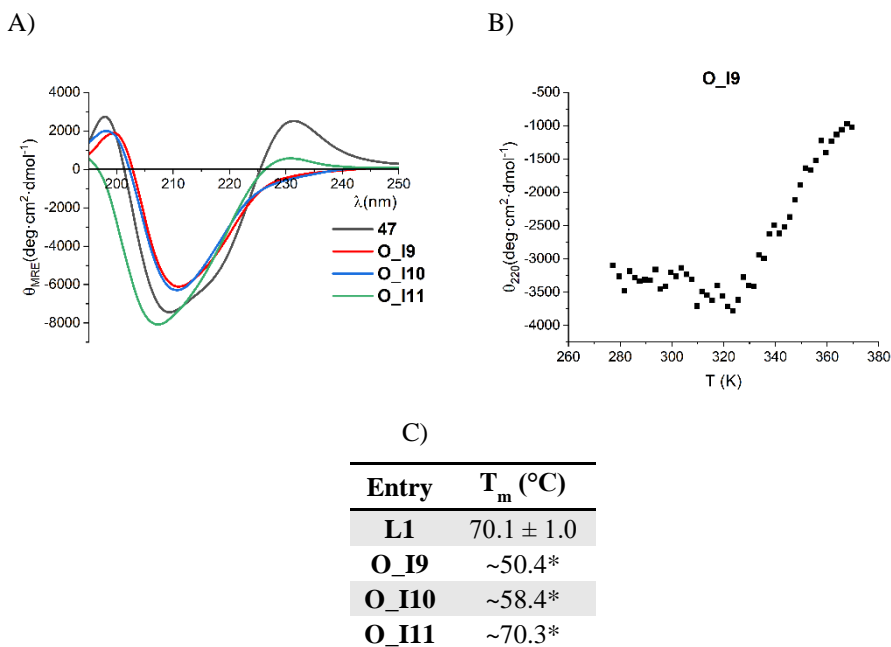
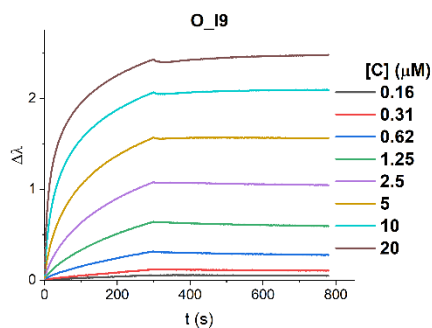


Figure 88. A) CD spectra, B) Recorded ellipticity versus T (K) for inhibitor **O_I9**, and C) Estimated T_m values of miniprotein **47** and inhibitors **O_I9-O_I11**. (*) Indicates that the T_m value is an approximation derived from the data collected prior to precipitation of the inhibitors.

The affinity studies performed by BLI showed an improvement on the aggregation, during the measurements, at the higher concentrations, (Figure 89). Yet, the binding affinities were at the same range independently of the modifications of the β -sheet residues. This may indicate that the inhibitors have a similar organised interactive surface and have similar effectivity despite the solubility and aggregation drawbacks.

A)



B)

Entry	K_D (nM)
O_I9	149 ± 3.3
O_I10	176 ± 4.5
O_I11	175 ± 5.5

Figure 89. BLI results, A) Dose-dependent sensograms of inhibitor **O_I9**, and B) Estimated global K_D values for inhibitors **O_I9-O_I11**.

To further explore the capacity of these inhibitors in more complex systems, we tested inhibitors **O_I1**, **O_I2**, and **O_I9-O_I11** and the scaffold **47**, with the HTRF assay at two concentrations, 1 and 3 μM , (Figure 90). The results reflected a successful modulation of the activity toward the inhibition of the PD-1/PDL-1 immune checkpoint. Although the scaffold did not show inhibition towards the complex, the designed inhibitors had an estimated IC_{50} between 2 and 3 μM .

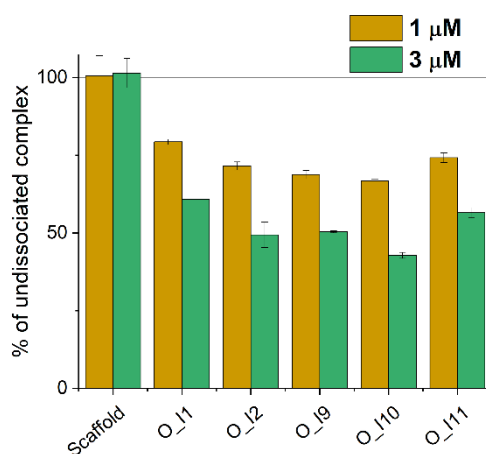



Figure 90. HTRF results for inhibitors **O_I1**, **O_I2**, **O_I9-O_I11**.

Simultaneously to the above-mentioned research, we decided to explore what the effects were on the charged residues content and net charges on the inhibitor behaviour. To that end, we generated six mutants of **O_II**, where residues from non-interactive regions of the inhibitor, were modified to charged ones (Arg, Lys, and/or Glu) to modulate the net charges at physiological pH (7.4).

The resulting mutants were modified to have a total net charge matrix of -4.3 to +4.6, (Table 29). Moreover, the total percentage of charged residues in the sequence was taken into consideration during the design. This way, we were able to discern which of these properties had a greater impact on solubility and binding. It is important to note that, due to the chemical synthesis with H-Rink amide resin, the C-terminus of the inhibitors are amidated. Therefore, the modification of net charges to negative values required a higher number of Glu and Asp residues, hindering the adequate distribution of electrostatic interactions.

Table 29. Sequences and net charges at pH 7.4 of inhibitors **O_II** and **O_C1-O_C6**. New mutations are highlighted in red, and modifications already explored in blue.



The diagram shows a protein sequence from residue 1 to 40. Residues 1-10 are in a green arrow, 11-28 are in a red box, 29-32 are in a blue arrow, and 33-40 are in a green arrow. Mutations are indicated by colored diamonds: red for new mutations and blue for previously explored ones.

Inhibitor	Sequence	Net Charge
O_II	ILVIT TF HG TN K◆ YDE ◆ KQ ◆ ARK I PPG W HIV F HS NG EV W WI A Y -NH ₂	+1,6
O_C1	ILVIT FHG EER ◆ YRE ◆ KE ◆ AER I PE GW HIV F HK NG EV W WI A Y -NH ₂	-1,3
O_C2	ILVIT FHG EER ◆ ERE ◆ EE ◆ AER I PE GW HIV R HK NG EV W WI A Y -NH ₂	-3,3
O_C3	ILVIT FHG T EK ◆ YDE ◆ KR ◆ AEE I PE GW HIV F HE NG EV W WI A Y -NH ₂	-4,3
O_C4	ILVIT FHG T KK ◆ YDE ◆ KE ◆ ARK I PK GW HIV F HE NG EV W WI A Y -NH ₂	+1,6
O_C5	ILVIT FHG T RE ◆ YEK ◆ KR ◆ ARK I PPG W HIV F HS NG EV W WI A Y -NH ₂	+3,6
O_C6	ILVIT FHG T RE ◆ YEK ◆ KR ◆ ARK I PK GW HIV F HS NG EV W WI A Y -NH ₂	+4,6

The structural analysis of the above-mentioned mutants provided an initial understanding of the solubility effects of the modifications. For the studies, the inhibitors were prepared at a concentration of 1 mM in water and diluted at 100 μM in three different buffers, I) potassium phosphate buffer pH 6.5 50 mM, II) potassium phosphate buffer pH 7.5 50 mM and III) phosphate buffer saline pH 7.4. The negatively charged inhibitors were poorly (**O_C2** and **O_C3**) or non-soluble (**O_C1**) in water or any buffer. However, inhibitors **O_C2** and **O_C3**, showed improved solubility, with no precipitate, after heating

up the solutions, (Figure 91). We believe that the high number of negative charges for these two inhibitors is responsible for generating repulsions that prevent them from aggregating. Yet, these repulsions and the lack of proper distribution of electrostatic interactions, may generate an energetical barrier of folding only overcome when heated.

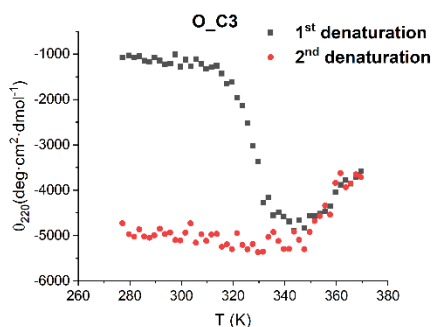


Figure 91. Recorded ellipticity vs T (K) for inhibitors **O_C3** during first (black) and second (red) denaturation.

Accordingly to the CD analysis, we can assume that the general fold of the inhibitor prevails except for **O_C2**. For which, the local minima for the helical region of the CD spectrum, is shifted to lower values, (Figure 92). It is possible to predict that the lack of a proper electrostatic distribution on the helical region may induce a loss of the native fold and, in consequence, a non-stable fold that translate on the lack of T_m . However, for inhibitor **O_C3**, a better distribution of positive and negative residues enhances the helical stability. Moreover, the lack of aggregation during thermal denaturation studies allowed for the estimation of the T_m of unfolding. In contrast, positively charged inhibitors showed greater solubility in water, but lower capacity to be solublized in buffer. The three inhibitors were soluble in KPB at pH 6 but not at pH 7.5, as for PBS buffer pH 7.4, only those with a positive charges above +3, **O_C5** and **O_C6**, were soluble. The CD spectrum indicated the prevalence of the secondary structure, with a slight enhancement of the helical fold, possibly related to the greater number of salt bridges. Yet, the solubility was improved in reference to **O_II**, and the aggregation during the thermal denaturation prevailed, precluding us from the proper estimation of the T_m values.

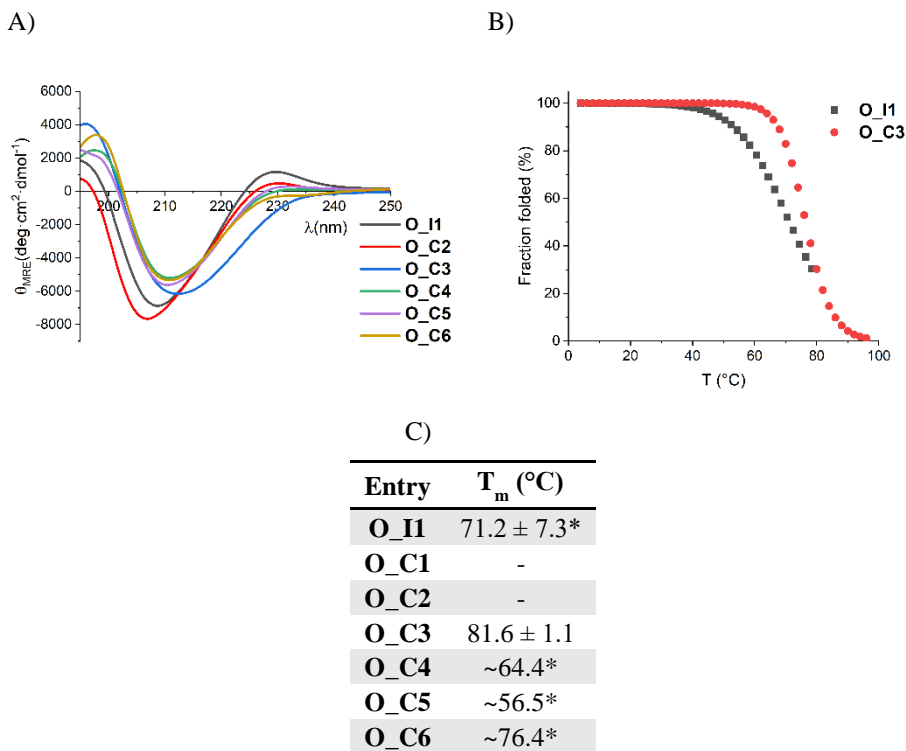


Figure 92. A) CD spectra of inhibitors **O_I1** and **O_C2-O_C6**, B) Normalized thermal denaturation plots of inhibitors **O_I1** and **O_C3**, and C) Estimated T_m values of inhibitor **O_I1** and **O_C1-O_C6**. (*) Indicates that the T_m value is an approximation derived from the data collected prior to precipitation of the inhibitors.

The most interesting characteristics of this set of inhibitors occurred when analysing their binding to PD-L1 by BLI. Although positively charged inhibitors retained affinity towards PD-L1, those analysed inhibitors with negative net charges showed almost no affinity toward the target. When comparing the sensogram of **O_C3** (-4.3 charge) with **O_C6** (+4.6 charge) we can observe a more than 40-fold loss of response under the same assay conditions, (Figure 93). Furthermore, against all predictions, it seemed that the greater solubility was promoting a loss in affinity since the estimated K_D was changed from ~300 nM for **O_C4**, to ~2000 nM for **O_C6**. We hypothesised that the results for those inhibitors with lower solubility were deviated by potential aggregation effects, while for **O_C6**, these effects were minimised and the estimated value more accurate.

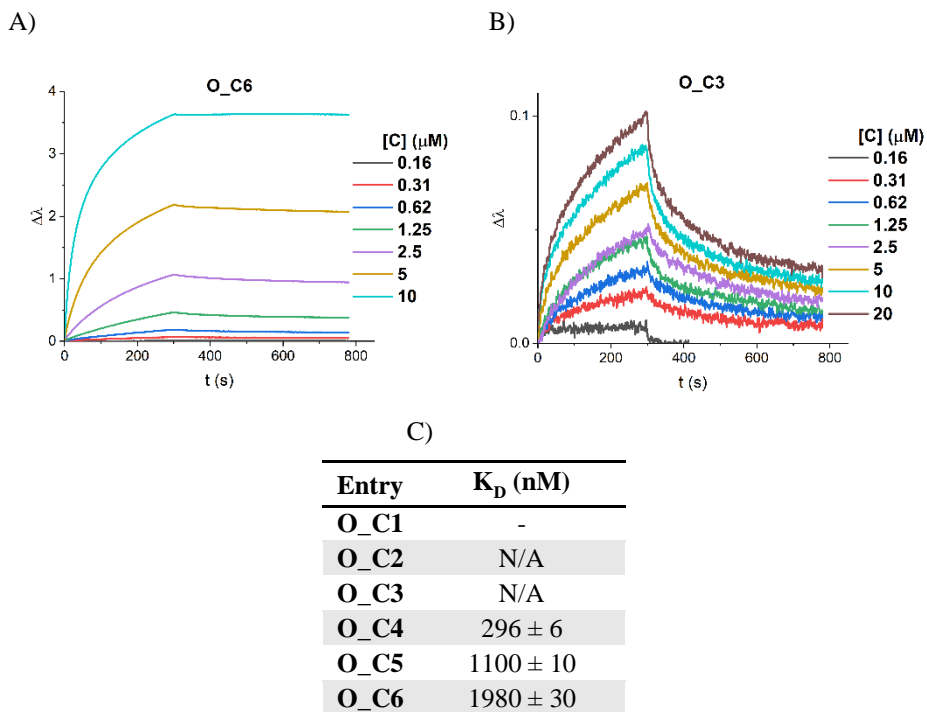


Figure 93. BLI results for inhibitors **O_C1-O_C6**, A) Dose-dependent sensograms of inhibitor **O_C6**, and B) Dose-dependent sensograms of inhibitor **O_C3**, and C) Estimated global K_D values for inhibitors **O_C2-O_C6**.

The observed results from BLI were then confirmed by HTRF assay. Inhibitors **O_C3-O_C6** were studied at 1 and 3 μM concentrations and the fluorescence was measured after 1 and 2 h of incubation, (Figure 94). The longer incubation had a slight improvement in inhibition for those inhibitors with positive charge. However, for the scaffold and the negatively charged inhibitor **O_C3**, no inhibition was observed. This observation reinforced the theory that positively charged inhibitors are required for the efficient disruption of the PD-1/PD-L1 interaction. Additionally, the charge studies not only enhanced the solubility of **O_I1**, but were able to improve the estimated IC_{50} to values below 3 μM .

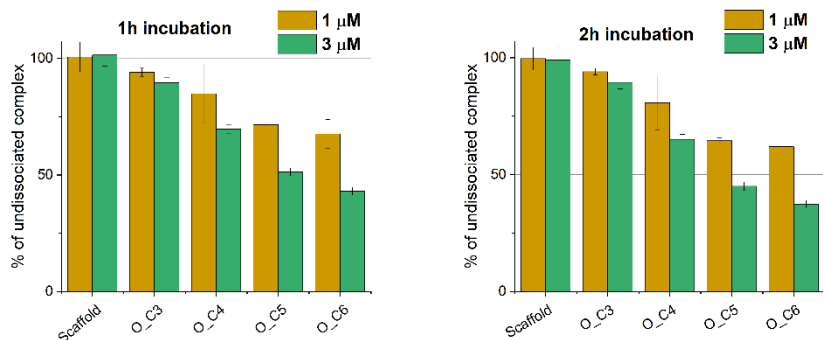


Figure 94. HTRF results for inhibitors **O_C3-O_C6**.

We were able to perform immune checkpoint blockade cell-based assay*. The initial experiment consisted of the measurement of two inhibitor concentrations, 10 and 50 μM, that were incubated for 6h with APC and Jurkat T-cells before the addition of Bio-Glo reagent and measurement of the luminescence. The results, (Figure 95), indicated that the inhibitors were not active at the lower concentration under the assay conditions. Furthermore, the poor solubility in the media and consequent precipitation at the higher concentration was harming the cells, as visible for a luminescence below the water control.

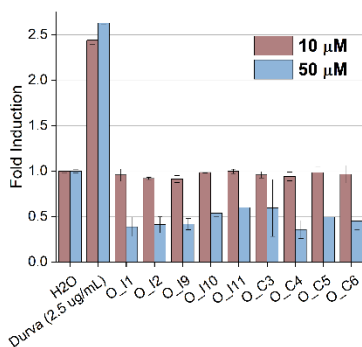


Figure 95. Immune checkpoint blockade cell-based assay results for inhibitors **O_I1, O_I2, O_I9-O-I11** and **O_C3-O_C6** versus water and Durvalumab was negative and positive controls.

* In cooperation with Dr Łukasz Skalniak from the Department of Organic Chemistry at the Faculty of Chemistry from Jagiellonian University

Since we had evidence from BLI and HTRF that these inhibitors may be able to inhibit the PD-1/PD-L1 complex in cells, we set out to optimise the conditions of the assay. Initially, we started by screening the pH / solubility / cell survival relationship. It is important to note that the RPMI 1640 medium with l-glutamate is a sodium bicarbonate buffer system. This kind of buffering, when stored in a 5-10 % CO₂ environment, has a pH of ~7.4. However, when exposed to oxygen in the air, the pH changes to pH ~8.4. Such conditions are not suitable for the adequate solubility of the inhibitors. Supported by previous works of Dr Łukasz Skalniak laboratory we had proof that the cell survival was not possible at pH below 6.8. Following a pH modification study (described in section 5.8. of the experimental work) we found the optimal conditions for the assays. The lack of precipitate for stocks at 20 μM, with final concentration 10 μM and pH ~6.9 was used for the preliminary experiment, together with a concentration of 5 μM of the inhibitors in unmodified pH medium. The results of the measurement after 6 h of incubation were very revealing, (Figure 96). The plate section containing nonmodified media at a concentration of 5 μM inhibitor showed almost no activation. However, the second half of the plate at 10 μM and pH ~6.9 reflected fold induction for the inhibitors **O_I9-O_I11**, **O_C4**, and **O_C5**. The lack of results for **O_C6** was due to an error in sample preparation. Furthermore, the negatively charged inhibitor, **O_C3**, showed no activity under any of the conditions studied.

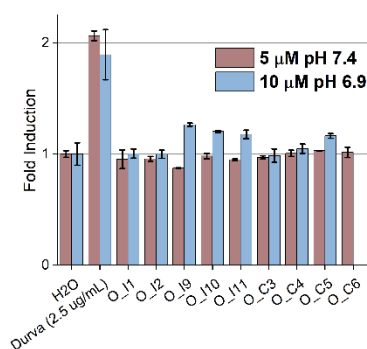


Figure 96. Immune checkpoint blockade cell-based assay results for inhibitors **O_I1**, **O_I2**, **O_I9-O-I11** and **O_C3-O_C6** versus water and Durvalumab was negative and positive control in two pH.

To verify the enhancement of activity at lower pH, two new plates were prepared with 5 and 10 μM concentrations of the inhibitors, each. The first plate was incubated for 6 h with APC, inhibitors, and Jurkat T cells prior to the addition of Bio-Glo, while the

second plate was incubated for 24 h with inhibitors and APC cells prior to a second 6 h incubation with Jurkat T cells. For the second plate, to avoid APC cell death for the lower pH, the inhibitors were diluted, to final concentration with the unmodified medium, and after 24 h, 10 μ L of solution from each well was removed to add Jurkat T cells. Equally to inhibitor **I38** (from section 3.2.3 of the present dissertation) the longer incubation time allowed for a higher response of the inhibitors, in particular **O_C5** and **O_C6**, (Figure 97). These two later inhibitors showed T cell activation in the assay at 10 μ M comparable to the maximum activation induced by the positive control Durvalumab.

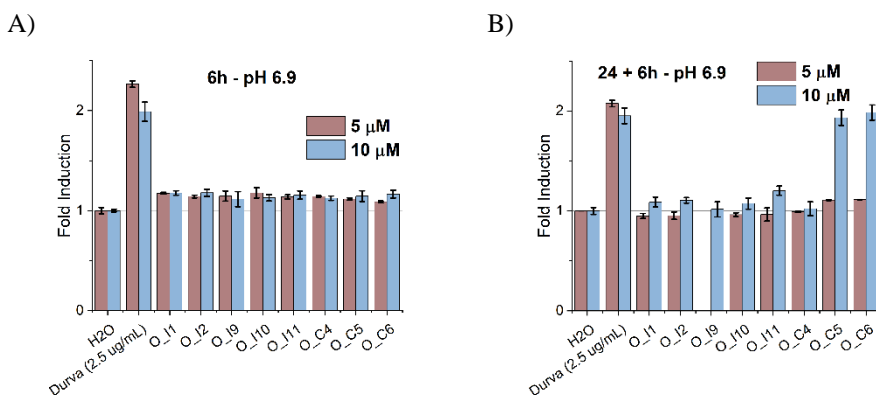
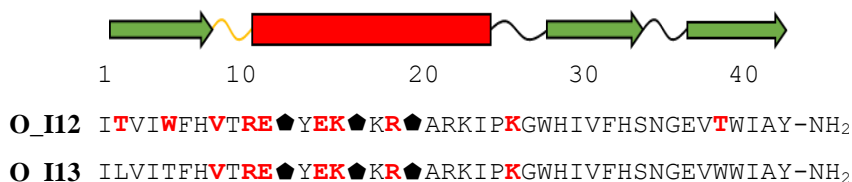


Figure 97. Immune checkpoint blockade cell-based assay results for the inhibitors **O_I1**, **O_I2**, **O_I9-O_I11** and **O_C3-O_C6** versus water and Durvalumab as negative and positive control, A) Lower pH for 6, and B) 24 + 6 h incubation.

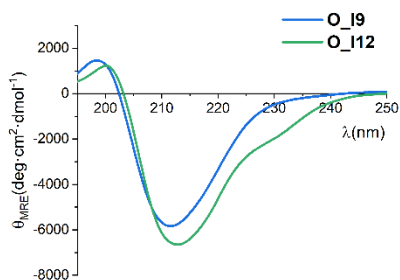
Taking into consideration that the final conditions for cell-based assay were determined, we decided to synthesise two more inhibitors that will combine the best characteristics found during the work, (Table 30). To generate inhibitor **O_I12**, the interactive surface of **O_I9** was merged with the charge distribution of **O_C6**. While for inhibitor **O_I13**, we modified Gly8 residue for Val8 in **O_C6** to explore stability enhancement effects on binding.

Table 30. Sequences of the inhibitors **O_I12** and **O_I13**. The new mutations are highlighted in red.

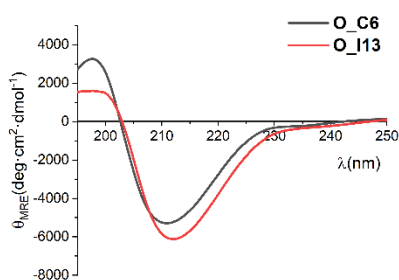


Structural analysis in solution of both inhibitors showed that the modifications performed had a great impact on the conformation of **O_I9**, as seen by differences in the spectra in the region of aromatic contributions of the inhibitor **O_I12**. However, the shift of the helical minima to higher λ may be an indicative of a more pronounced helical fold derived from the increase in electrostatic contacts along the helix. The CD spectrum of **O_I13** had virtually no changes, as expected from a single modification, (Figure 98). However, we could again observe a shift to higher λ of the minima corresponding to the β -amino acid-containing helix. This could be the consequence of a more constrained conformation after modification of Gly8 residue, a secondary structure disruptor, to Val8, which induces a more rigid fold. Disappointingly, modifications of these new two inhibitors did not improve the tendency to aggregate during the thermal denaturation studies. This preventing us from accurately estimating the T_m values.

A)



B)



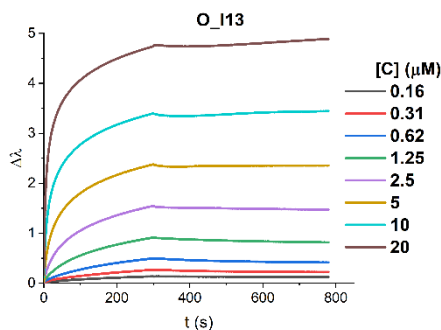
C)

Entry	T _m (°C)
O_I9	~50.4*
O_I12	~60.4*
O_C6	~76.4*
O_I13	~66.6*

Figure 98. A) CD spectra of inhibitors **O_I9** and **O_I12**, B) CD spectra of inhibitors **O_C6** and **O_I13**, and C) Estimated T_m values of inhibitor **O_I9**, **O_I12**, **O_C6** and **O_I13**. (*) Indicates that the T_m value is an approximation derived from the data collected prior to precipitation of the inhibitors.

BLI studies of the new inhibitors revealed, on the one hand, that **O_I12** had a loss of affinity compared to **O_I9**. This could be the consequence of a conformational change that results in an undesired distribution of the interactive residues. On the other hand, **O_I13** had not only an improved binding to PD-L1 with a K_D of 621 nM, but greater solubility properties and decreased aggregation, (Figure 99). The lower aggregating effect was visible as we could measure binding at concentrations up to 20 μM. Those concentration levels were precluded in previous sections due to low solubility and high aggregation tendencies of the inhibitors.

A)



B)

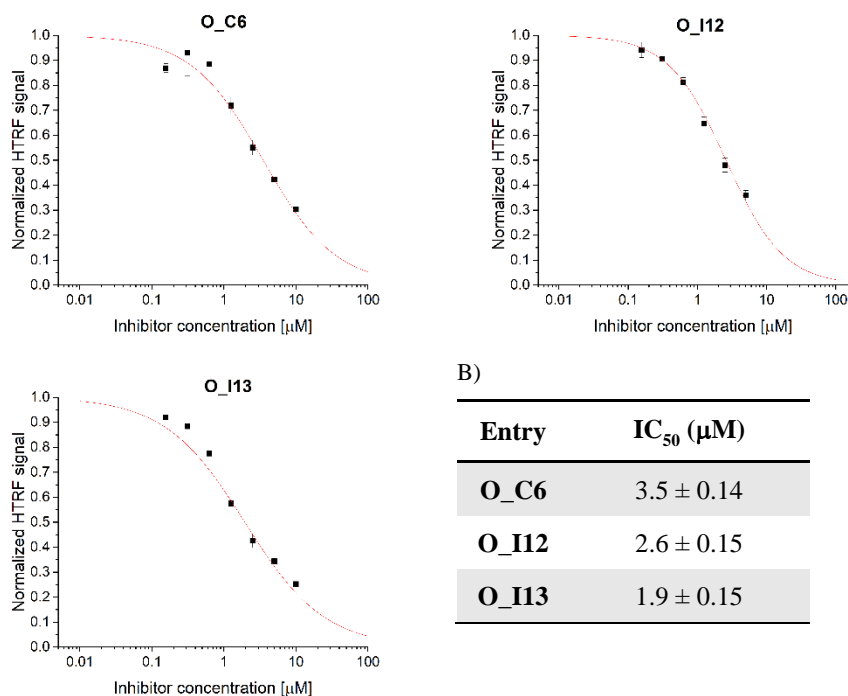
Entry	K _D (nM)
O_I12	922 ± 19.8
O_I13	621 ± 9.6

Figure 99. BLI results, A) Dose-dependent sensograms of inhibitor **O_I13**, and B) Estimated global K_D values for inhibitors **O_I12** and **O_I13**.

The inhibition capacity of **O_C6**, **O_I12** and **O_I13** was further explored by HTRF assay. A dose response experiment was performed at variable concentrations of the inhibitor in the range of 0.15-10 μM, (Figure 100A). All three inhibitors could block the interaction with IC₅₀ of 2-3.5 μM, (Figure 100B), displaying the same tendency in their

activity as observed by BLI. Interestingly, the structural optimisation incorporated for **O_I13** enhanced the inhibitory capacity by almost 2-fold, a considerable improvement for a single mutation independent of the interactive side. Moreover, the inhibitory capacity of the latest design is stronger than that of PD-1, IC_{50} 3.3 μ M, as reported previously²²⁸.

A)



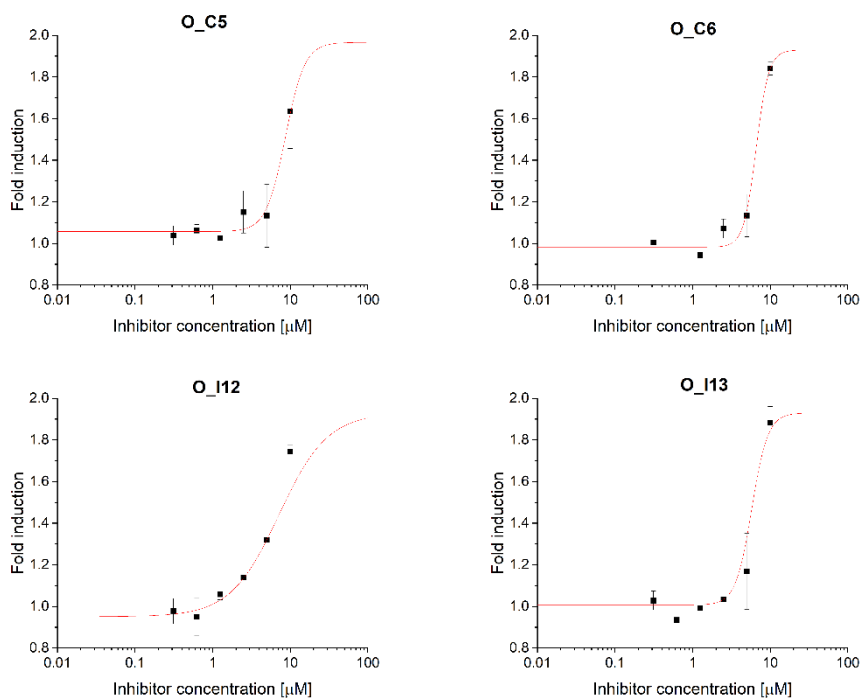
B)

Figure 100. HTRF results for inhibitors **O_C6**, **O_I12** and **O_I13**. A) Inhibition curves from HTRF experiments, and B) IC_{50} estimations based on HTRF results.

For the EC_{50} estimation by cell-based assay, inhibitors **O_C5**, **O_C6**, **O_I12**, and **O_I13** were selected. The preparation of the conditions was as explained previously where the medium was adjusted for a final pH of 6.9. The inhibitors were prepared in stocks of 20 μ M and incubated for 24 h with APC cells at six concentrations with two-fold serial dilutions. Following 24 h of incubation, Jurkat T cells were added and incubated for another 6 h prior to adding Bio-Glo and measuring luminescence. The results supported the successful incorporation of the activity into the scaffold with estimated EC_{50} of 5.8-8.7 μ M, (Figure 101). The presence of greater positively charged residues enhanced the inhibitory effect as observed by the lower EC_{50} of **O_C6** when compared to **O_C5**, yet the initially designed interface was optimal for the binding of the inhibitors into PD-L1 as

observed by a lower EC_{50} of **O_I12** in comparison with **O_I13**. The final design, **O_I13**, incorporating all the optimization along this section was indeed the most active inhibitor in cell assays. Moreover, it is to date, the most potent β -amino acid-containing miniprotein-based inhibitor, stabilized by a hydrophobic core, of the PD-1/PD-L1 immune checkpoint.

A)



B)

Entry	EC_{50} (μM)
O_C5	8.7 ± 1.1
O_C6	6.6 ± 1.4
O_I12	7.1 ± 0.2
O_I13	5.8 ± 0.6

Figure 101. Results of the immune checkpoint blockade cell-based assay inhibition curve of inhibitors **O_C5**, **O_C6**, **O_I12** and **O_I13**. A) Fold induction versus inhibitor concentration, and B) Estimated EC_{50} from the inhibitory curve.

3.2.5. Conclusions

In conclusion, two aims of the present dissertation were successfully explored. The first aim was to generate new interactive and binding modes to PD-L1 target proteins. Initially to model system were explored, the WW-prototype and the FBP28WW. Two design stages were implemented. First, novel interactive surfaces were computationally generated with Rosetta FastDesign, to provide inhibitors **I3-I15**, and the miniproteins were synthesised and studied by CD, BLI, and HTRF, selecting the best inhibitor of each model, **I6** and **I9**. Second, after analysis of the mutations of an optimised PD-1, the His residue was incorporated into the inhibitors **I6** and **I9** that generated the inhibitors **I16-I27**. Regarding mutants **I16-I27**, it was concluded that the optimisation approach was adequate, as seen by a remarkable improvement of the inhibition by **I22**. Nevertheless, when compared to inhibitor **I6**, **I9** demonstrated a limited ability to incorporate new mutations without affecting the miniprotein fold, thus restricting its potential for further optimisation. For that reason, this research line was ended. Incorporation of the optimised epitope of several **I1**-derived inhibitors into miniprotein **21** (described in section 3.1.1.), led to inhibitors **I28-I35**. These new inhibitors displayed an enhancement in binding, with an increase in sensogram response of ~24-fold, and specificity towards PD-L1, with K_D and IC_{50} values of low micromolar levels. We concluded that the higher structural order of the β -sheet in the presence of the stabilising helix and hydrophobic core was promoting the necessary structuration to maintain the epitope in the desired conformation. However, because of the low cooperativeness of the fold and poor solubility, the success of these inhibitors in complex systems, such as cell-based assays, was precluded. In a final step of the design, a selected group of HEEE inhibitors was circularly permuted to the EHEE fold, to give inhibitors **I36-I39**. The permutation of the topology significantly enhanced the thermal stability, in some cases with T_m values above 96 °C. Moreover, these miniproteins, showed outstanding resistance to guanidine denaturation. Moreover, the improved physicochemical properties of these inhibitors allow the estimation of EC_{50} in the immune checkpoint cell-based assay with an EC_{50} of 27 μ M for inhibitor **I38**. In the context of designing new binding modes to PD-L1, the research successfully generated new interactive surfaces for PD-L1 target proteins and identified promising inhibitors, **I38**. However, while optimisation efforts showed significant improvements in certain inhibitors, limitations on stability, solubility, and aggregation were observed, restricting us from the full characterisation of these compounds.

The second aim of this section was to follow the successful design of β -amino acids containing miniproteins, with EHEE topology, whose fold was dominated by an optimised network of hydrophobic interactions in the core. We used a hybrid approach for the design of PD-1/PD-L1 immune checkpoint inhibitors. The method combined mimicry of the interactive surface of optimised PD-1 with mutations that will increase the affinity for the target, while preserving the EHEE fold. In the initial step, we incorporated the desired activity into miniprotein **46**, by generating inhibitor **O_I1**. The BLI and HTRF study of the scaffold and the designs confirmed the successful incorporation of activity, as some of the designs showed nanomolar affinities and IC_{50} of approximately 3 μ M while the scaffold was not active in the assay. The single- or double-point modifications of **O_I1** enabled us to investigate the possibility of optimising the surface of the inhibitor. Surprisingly, these inhibitors displayed close to no dissociation from PD-L1 when studied by BLI. We hypothesised that the entropic penalty required to break the hydrophobic cluster generated from the interaction was generating an irreversible binding. However, the poor solubility of these inhibitors and the tendency to aggregate during denaturation studies was problematic. To examine the effects of net charges on solubility and inhibition, we designed six inhibitors with net charges spanning from -4.3 to +4.6, creating inhibitors **O_C1-O_C6**. To prevent from interfering on the binding mode of **O_I1**, only residues from noninteractive regions were modified to charged ones. The first results from BLI revealed that, an inhibitor had to be positively charged. To bind to PD-L1. Further studies by HTRF demonstrated that the higher the positive charge, the higher the inhibition. Moreover, the improved solubility of the designs allowed for the estimation of EC_{50} by cell-based assay. We found that it was necessary to adjust the pH of the media to 6.8-6.9 to measure activity of this set of miniproteins. This was of the outmost interest for our investigation, as cancer cells display a slightly more acidic extracellular pH than healthy cells, suggesting that these new inhibitors may have specificity for cancer cells over healthy cells. The assembling of sequences from **O_C6**, the most soluble inhibitor, with **O_I1**, the most active inhibitor and the structural enhancement of miniprotein **47** led to the inhibitor **O_I13**. This last inhibitor had a nanomolar affinity towards PD-L1, with a K_D of 621 nM studied by BLI, IC_{50} of 1.9 μ M by HTRF and an EC_{50} of 5.8 μ M by immune checkpoint blockade cell-based assay. Furthermore, up until now, it stands as the most powerful miniprotein-based inhibitor containing β -amino acids, supported by a hydrophobic core, targeting the PD-1/PD-L1 immune checkpoint.

4. Summary

During the present dissertation three major aims were addressed, the design of β -amino acid containing miniproteins, the design of novel binding modes for the PD-1/PD-L1 immune checkpoint and the incorporation of activity against PD-1/PD-L1 immune checkpoint on the designed scaffolds.

To sum up, our research focused on the evolutionary design of miniproteins containing β -amino acids, exploring different topologies to create sophisticated and controlled protein-like tertiary structures. We utilized native and β -amino acid-containing structures for backbone generation and as a guide for the development of more stable miniproteins with potential as scaffolds for the design of Protein-Protein Interaction (PPI) inhibitors. Introducing interconnectivity modifications further reinforced the stability of the sequences, resulting in structurally robust miniproteins. The profound structural analysis was of great support for the full understanding of the designs and the implications of the modifications. As a result, we successfully designed novel β -amino acid containing miniproteins with diverse tertiary folds and heterogeneous backbones. Additionally, our design allowed for the incorporation of *trans*-ACPC, which not only induced helical structures but also contributed to the overall stability of the tertiary structure. The designed β -amino acid containing miniproteins were applied as scaffolds for the development of novel PD-1/PD-L1 immune checkpoint inhibitors. Two approaches were used: first designing interactive surfaces over native domains and grafting binding epitopes onto different scaffolds with various topologies. Second, use of optimized miniproteins with an EHEE topology and combining mimicry of optimized PD-1 with mutations aimed to increase affinity for the target while preserving the EHEE fold.

The study resulted in the first designed β -amino acid containing miniproteins encompassing a complex α - β fold incorporating *trans*-ACPC in the hydrophobic core. Also, the work has led to the successful design of the first β -amino acid containing inhibitors of the PD-1/PD-L1 immune checkpoint, stabilized by a hydrophobic core. The findings of this research contribute to the advancement of designing miniproteins with tailored tertiary structures, broadening the scope of potential applications in medicinal chemistry, particularly for targeting protein-protein interactions. These insights may open new avenues for drug development and therapeutic interventions based on stable and versatile miniprotein scaffolds. The research highlights the potential of miniproteins and

foldamers as valuable tools in medicinal chemistry and the promising prospects they offer for future drug development and precision medicine.

5. Experimental

5.1. Miniprotein design

5.1.1. Rosetta FastDesign Protocol

Rosetta software²⁴⁴ is a molecular modelling package of algorithms that are used for understanding protein structures, protein design, protein docking, protein-DNA, and protein-protein interactions. These options are available through a series of different modules. For protein design the Rosetta Design protocol, it applies modifications on the indicated positions of the sequence by the resfile and applies Monte Carlo search over fragments of the miniprotein producing structures of low free energy^{245,246}. The free energy calculated is a combination of weighted score terms which describe:

- I) Hydrogen bonds
- II) Electrostatic interactions
- III) Van der Waals interactions
- IV) Solvation
- V) Backbone torsion angles
- VI) Side chain rotamer energies
- VII) Disulfide bonds
- VIII) Unfolded reference energy

This energy scores the quality of the generated miniprotein and it is used to select the sequences to synthesize.

5.1.2. Structure predictors

The structure predictors used to support the miniprotein design were TrRosetta^{247,248} and AlphaFold^{41,42}. Both software can be found in servers online without

the need to download and install them. Considering that these fold predictors do not recognize, yet, non-canonical amino acids, the fetched sequences were modified to contain alanine residue instead of *trans*-ACPC, due to its similar size and physicochemical properties. For both software the predictions were performed in absence of templates and with single sequence folding.

5.1.3. Minimization and molecular dynamics

Preparation of the structures:

Initial structures, for energy minimization and/or molecular dynamic simulations, were either fetched directly from the PDB files obtained with Rosetta FastDesign protocol, or by modification of the predicted structures in either TrRosetta or AlphaFold2. When using the PDB models obtained from structure predictors, a helical fragment of the HIV gp41 CHR domain α/β -peptide analogue (residues 13-24), was superimposed over the predicted helix using DiscoveryStudio Visualizer (v17.2.0). The superimposed α/β -helix was then linked to the structure, followed by the modification of Nitrogen 4 from *trans*-(3*S*,4*R*)-4-aminopyrrolidine-3-carboxylic acid (*trans*-APC) into Carbon to convert it into *trans*-ACPC. The PDB files were then modified in text mode to locate the helix in the corresponding atomic position and the *trans*-ACPC atoms relabelled based on the parametrization of the corresponding software.

Amber minimization and simulations.

Calculations were performed using AMBER14²⁴⁹ software available at Wrocław Centre for Networking and Supercomputing (WCSS).

All the systems were processed with `pdb4amber` in AmberTools21²⁵⁰ to remove any water from the crystal structures, addition of missing hydrogens and modify the nomenclature based on Amber.

The clean PDB was then processed with `tleap` in AmberTools21 for the preparation of the system for MD, a truncated octahedron solvent of TIP3P with a 5 Å padding a closeness of 0.75 Å was implemented, followed by a charge neutralization with

NaCl. The protein was defined with leaprc.mimetic.ff15ipq²⁵¹, a force field that includes parametrization of trans-ACPC (BCY).

Energy minimization was carried out in four stages, all the stages were done with 500 steps of steepest descent and 9500 steps of conjugate gradient, and the coordinates of the previous minimization was used for the following step:

1st → Water molecules, ions and all hydrogen atoms of the system were minimized applying on all the other heavy atoms Cartesian restraints of 25 kcal/mol.

2nd → Side chains were then minimized by applying a gradient restraint over backbone atoms, from 25-5 kcal/mol.

3rd → Minimization with restraints over alpha carbons of 1 kcal/mol.

4th → Unrestrained minimization of the entire system.

Heat equilibration phase, the temperature was increased linearly from 0 to 300 K over 18ps and then kept at 300 K over 2ps. This equilibration was carried out with a 5 kcal/mol restraint on backbone and ligand atoms. Using Langevin dynamics with a $\gamma_{ln} = 5 \text{ ps}^{-1}$.

NPT equilibration, ensemble using Langevin dynamics with a $\gamma_{ln} = 5 \text{ ps}^{-1}$. Berendsen barostat was chosen to maintain the pressure at 1.01325 bar with a $\tau_{p} = 1 \text{ ps}$. Restraint on the backbone were imposed and decrease along the simulation from 5-1 kcal/mol over 1ns simulation.

Finally, the system was further equilibrated over 2ns without any restraint.

MD production, once the system was equilibrated the simulation was started from the coordinates of the equilibrated system for as long as necessary for the experiment.

Gromacs minimization and simulations.

Calculations were performed using GROMACS²⁵² software available at Wrocław Centre for Networking and Supercomputing (WCSS).

The PDB file, was processed with pdb2gmx to clean the file and to generate the topology file using Amber03ff^{253,254}, which includes parametrization of 1S-2S, aminocyclopentanecarboxylic acid (Trans-ACPC) under the three-letter code SS5, as the forcefield.

The protein was placed in a cubic box with 1.0 nm distance from any box edge, and the box was filled with water molecules using spc216.gro, a 3-point solvent model. The solvated complex was neutralized substituting water molecules by the corresponding ions to achieve neutral charge of the box.

Energy minimization: was carried out using steepest descent minimization with maximum 50000 steps and maximum force of < 1000 Kj/mol/nm. Long range electrostatic interactions were treated with Particle Mesh Ewald (PME).

Heat equilibration phase (NVT equilibration), the temperature was set at 300 K, over 100 ps simulation in 50000 steps at 2 fs per step, applying Periodic Boundary Conditions and modified Berendsen thermostat, with position restrains on the protein complex.

Pressure equilibration phase (NPT equilibration), the pressure equilibration used the same molecular dynamic parameters as the NVT phase using pressure coupling, controlled by Parrinello-Rahman with a reference pressure of 1.0 bar and tau_p 2ps. The equilibration was runed over 100 ps.

MD production, once the system was equilibrated the simulation was started from the coordinates of the equilibrated system for as long as necessary for the experiment.

5.2. Microwave assisted solid phase peptide synthesis

All commercially available reagents and solvents were purchased from Sigma-Aldrich, Merck, Iris Biotech, Lipopharm or Trimen and used without further purification. Fmoc-(1S-2S)-trans-aminocyclopentanecarboxylic acid (trans-ACPC) with a >95% purity was purchased from Synnovator. The mini-protein were synthesized by microwaved assisted solid-phase peptide synthesis (SPPS), in a Biotage® Initiator+Alastra or Liberty Blue CEM. The solid support employed was H-rink amide ChemMatrix®, loading capacity 0.59 meq/g at a 0.06 mmol scale. Fmoc- α -amino acids were used at a 0.1M concentration

and 5 equivalents, while Fmoc-trans-ACPC was used at a 0.1M concentration and 3 equivalents and, Oxyma and DIC at 0.5M concentration in DMF.

A mixture of single and double coupling procedures was implemented to maximize the synthesis yields. Fmoc deprotection was performed using 20 % piperidine solution in DMF at 75°C/1 min for single coupling and r.t./13 min for double coupling. Fmoc- α -amino acids, except His, are mixed with Oxyma and DIC and reacted 75°C/1 min 50 sec followed by 90°C/3 min for single coupling, and for double coupling the mixture is heated up to 75°C/15 min x2 before deprotection. To avoid its racemization, His is heated to 50°C/10 min for single coupling and r.t./90 min x2 for double coupling. For trans-ACPC coupling the method used is 75°C/30 min followed by r.t./30 min. Following every deprotection and coupling the resin was washed 3 times with DMF. Upon completion of the synthesis the resin was washed with 4.5 mL of DCM (x6) to remove DMF and dry the resin.

Synthesis products were cleaved from the resin, and side chain deprotected, using a TFA/TIS/thioanisole/H₂O cleaving cocktail (ratio 85:5:5:5), and mixing it in a shaker for 3.5 hours. 4.6 mL of cleaving cocktail was used for 100mg of resin, combined in a fritted syringe. After filtration of the cocktail mixture through the frit, and separating the resin, into a 50 mL falcon, the crude products were precipitated with ice-cold diethyl ether and centrifuged at 6000 rpm for 6 min at 2°C. The resulting products of the SPPS, with H-rink amide ChemMatrix®, derives in C terminus amidated products.

5.3. Miniprotein purification and characterization

HPLC purification and characterization.

The pellet obtained from the centrifugation was dissolved in a mixture of 40:60 H₂O + 0.05% TFA and MeCN + 0.05% TFA and purified by reverse phase HPLC with a Dionex Ultimate 3000, in a C18 Thermo Scientific preparative column Hypersil, 5 μ m, 250 x 20 mm. The binary solvent used for the purification, consisted of Solvent A: 99.95 % H₂O / 0.05% TFA and Solvent B: 99.95 % MeCN / 0.05% TFA.

Preparative HPLC gradient:

Total flow: 10 mL/min, t = 0 min – 90% A, t = 5 min – 90% A, t = 10 min – 70% A, t = 35 min – 45% A, t = 40 – 10% A, t = 45 min – 10% A, t = 50 min – 90% A, t = 55 min – 90% A.

(1) Purity of the product was monitored by analytical HPLC in a Shimadzu using a Dr Maisch Reprosil Saphir 100 C18, 5 μ m, 150 x 4.6 mm. The binary solvent used for the purification, consisted of Solvent A: 99.95 % H₂O / 0.05% TFA and Solvent B: 99.95 % MeCN / 0.05% TFA. Pure fractions (>95 %) were then concentrated using a rotary evaporator. The concentrated samples were flash frozen using liquid nitrogen and lyophilized for 48 h at 0.63 bar and -25 °C.

(2) Purity of the product was monitored by analytical HPLC in a Shimadzu using a CHROMSERVIS® CromShell® C18-XB, 2.6 μ m, 75 x 4.6 mm. The binary solvent used for the purification, consisted of Solvent A: 99.95 % H₂O / 0.05% TFA and Solvent B: 99.95 % MeCN / 0.05% TFA. Pure fractions (>95 %) were then concentrated using a rotary evaporator. The concentrated samples were flash frozen using liquid nitrogen and lyophilized for 48 h at 0.63 bar and -25 °C.

Analytical HPLC gradient:

(a) Total flow: 0.9 mL/min, t = 0 min – 90% A, t = 2 min – 90% A, t = 20 min – 10% A, t = 23 min – 10% A, t = 26 – 90% A, t = 30 min – 90% A.

(b) Total flow: 0.9 mL/min, t = 0 min – 90% A, t = 1 min – 90% A, t = 12.5 min – 10% A, t = 15 min – 10% A, t = 17 – 90% A, t = 18 min – 90% A.

Mass spectrometry.

The product of the synthesis and purification were identified by mass spectrometry with a WATERS LCT Premier XE System by electrospray ionization and time of flight.

Experimental [M + XH], together with analytical HPLC retention times, are indicated in Table 31.

Table 31. Mass spectrometry and analytical HPLC experimental data of mini-proteins **1-50**, **I1-I39**, **O_I1-O_I13** and **O_C1-O_C6**.

Peptide	MS Theoretical [M/Z]	MS Experimental [M/Z]	Analytical HPLC	
			Gradient	R _t
1	1034.2334 [M + 4H]	1034.3007 [M + 4H]	1a	9.22
2	1390.004 [M + 3H]; 1042.7550 [M + 4H]	1390.0294 [M + 3H]; 1042.7567 [M + 4H]	1a	8.05
3	1360.6738 [M + 3H]; 1020.7573 [M + 4H]	1360.6453 [M + 3H]; 1020.7567 [M + 4H]	1a	8.71
4	1306.1675 [M + 4H]; 1045.1340 [M + 5H]	1306.1638 [M + 4H]; 1045.0917 [M + 5H]	1a	9.02
5	886.7718 [M + 6H]; 760.2341 [M + 7H]	886.7931 [M + 6H]; 760.2465 [M + 7H]	1a	8.97
6	1328.676 [M + 4H]; 1063.1423 [M + 5H]	1328.7520 [M + 4H]; 1063.144 [M + 5H]	1a	8.87
7	1314.4116 [M + 4H]; 1051.7308 [M + 5H]	1314.4801 [M + 4H]; 1051.9453 [M + 5H]	1a	9.11
8	1074.1371 [M + 5H]; 895.2822 [M + 6H]	1074.1543 [M + 5H]; 895.2947 [M + 6H]	1a	9.53
9	1326.4207 [M + 4H]; 1061.3381 [M + 5H]	1326.4514 [M + 4H]; 1061.3040 [M + 5H]	1a	9.44
10	1329.9246 [M + 4H]; 1064.1412 [M + 5H]	1329.9619 [M + 4H]; 1064.1482 [M + 5H]	1a	9.73
11	1340.4363 [M + 4H]; 1072.5507 [M + 5H]	1340.4299 [M + 4H]; 1072.5513 [M + 5H]	1a	7.71
12	1337.1693 [M + 4H]; 1069.9370 [M + 5H]	1337.0691 [M + 4H]; 1069.9358 [M + 5H]	1a	10.09
13	1340.9321 [M + 4H]; 1072.9465 [M + 5H]	1341.0039 [M + 4H]; 1072.9656 [M + 5H]	1a	9.65
14	1317.9246 [M + 4H]; 1054.5413 [M + 5H]	1317.7075 [M + 4H]; 1054.5396 [M + 5H]	1a	12.13
15	1316.3918 [M + 4H]; 1053.3151 [M + 5H]	1316.4668 [M + 4H]; 1053.3162 [M + 5H]	1a	9.41
16	1319.6526 [M + 4H]; 1055.9236 [M + 5H]	1319.7173 [M + 4H]; 1055.9243 [M + 5H]	1a	9.10
17	1337.1721 [M + 4H]; 1069.9392 [M + 5H]	1337.0210 [M + 4H]; 1069.9399 [M + 5H]	1a	9.72
18	1340.4363 [M + 4H]; 1072.5507 [M + 5H]	1340.4373 [M + 4H]; 1072.5505 [M + 5H]	1a	9.80
19	1344.6748 [M + 4H]; 1075.9414 [M + 5H]	1344.6599 [M + 4H]; 1075.9395 [M + 5H]	1a	9.71
20	1240.6306 [M + 4H]; 992.7061 [M + 5H]	1240.6321 [M + 4H]; 992.4860 [M + 5H]	1a	9.81
21	1162.0980 [M + 4H]; 929.8800 [M + 5H]	1162.0996 [M + 4H]; 929.8727 [M + 5H]	1a	9.92
22	1183.3560 [M + 4H]; 946.8863 [M + 5H]	1183.5510 [M + 4H]; 947.0211 [M + 5H]	1a	10.17

23	1180.3468 [M + 4H]; 944.4790 [M + 5H]	1180.4749 [M + 4H]; 944.5945 [M + 5H]	1a	9.85
24	1165.3468 [M + 4H]; 932.4790 [M + 5H]	1165.5957 [M + 4H]; 932.6770 [M + 5H]	1a	9.90
25	1230.8716 [M + 4H]; 984.6982 [M + 5H]	1230.8721 [M + 4H]; 984.6609 [M + 5H]	1a	10.02
26	1222.1104 [M + 4H]; 977.8899 [M + 5H]	1222.1084 [M + 4H]; 977.7218 [M + 5H]	1a	10.47
27	1233.8806 [M + 4H]; 987.3061 [M + 5H]	1233.8804 [M + 4H]; 987.2400 [M + 5H]	1a	10.20
28	1230.3767 [M + 4H]; 984.3024 [M + 5H]	1230.4199 [M + 4H]; 984.3030 [M + 5H]	1a	10.07
29	1217.1275 [M + 4H]; 974.7602 [M + 5H]	1217.3970 [M + 4H]; 974.7932 [M + 5H]	1a	10.34
30	1178.8719 [M + 4H]; 943.2991 [M + 5H]	1178.8708 [M + 4H]; 943.2878 [M + 5H]	1a	10.74
31	1223.6196 [M + 4H]; 979.0973 [M + 5H]	1223.8442 [M + 4H]; 979.2908 [M + 5H]	1a	10.34
32	1236.4041 [M + 4H]; 989.3248 [M + 5H]	1236.3838 [M + 4H]; 989.3302 [M + 5H]	1a	10.88
33	1221.4032 [M + 4H]; 977.3241 [M + 5H]	1221.2402 [M + 4H]; 977.1931 [M + 5H]	1a	11.36
34	1224.6246 [M + 4H]; 979.9013 [M + 5H]	1224.6248 [M + 4H]; 979.8995 [M + 5H]	1a	10.53
35	1223.1182 [M + 4H]; 978.6961 [M + 5H]	1223.1611 [M + 4H]; 978.6816 [M + 5H]	1a	10.14
36	987.1193 [M + 5H]; 822.7673 [M + 6H]	987.1331 [M + 5H]; 822.7783 [M + 6H]	1a	10.43
37	1190.6365 [M + 4H]; 952.7108 [M + 5H]	1190.6467 [M + 4H]; 952.7001 [M + 5H]	1a	10.11
38	1215.3983 [M + 4H]; 972.5202 [M + 5H]	1215.3799 [M + 4H]; 972.5223 [M + 5H]	1a	11.47
39	1226.1445 [M + 4H]; 981.1171 [M + 5H]	1226.1645 [M + 4H]; 981.1173 [M + 5H]	1a	10.58
40	1229.1536 [M + 4H]; 983.5244 [M + 5H]	1229.1589 [M + 4H]; 983.5260 [M + 5H]	1a	10.61
41	1236.6563 [M + 4H]; 989.5266 [M + 5H]	1236.7014 [M + 4H]; 989.5297 [M + 5H]	1a	10.47
42	1223.1354 [M + 4H]; 978.9105 [M + 5H]	1223.2097 [M + 4H]; 978.9073 [M + 5H]	1a	10.00
43	1245.8978 [M + 4H]; 996.9199 [M + 5H]	1245.9663 [M + 4H]; 996.9360 [M + 5H]	1a	10.54
44	1244.6383 [M + 4H]; 995.9122 [M + 5H]	1244.6631 [M + 4H]; 995.9282 [M + 5H]	1a	10.21
45	1245.8978 [M + 4H]; 996.9199 [M + 5H]	1245.8113 [M + 4H]; 996.9198 [M + 5H]	1a	10.47

46	1260.9070 [M + 4H]; 1008.9271 [M + 5H]	1260.8708 [M + 4H]; 1008.9264 [M + 5H]	1a	10.79
47	1007.7343 [M + 5H]; 839.9465 [M + 6H]	1007.5322 [M + 5H]; 389.9507 [M + 6H]	2b	7.07
48	1222.6406 [M + 4H]; 978.3140 [M + 5H]	1222.6379 [M + 4H]; 978.3124 [M + 5H]	1a	10.32
49	1179.6296 [M + 4H]; 944.1061 [M + 5H]	1179.6296 [M + 4H]; 944.1140 [M + 5H]	1a	9.92
50	823.4296 [M + 6H]	823.4290 [H + 6H]	1a	10.32
I1	1034.2334 [M + 4H]	1034.3007 [M + 4H]	1a	9.22
I2	1090.5337 [M + 4H]	1090.5325 [M + 4H]	1a	8.56
I3	1022.2299 [M + 4H]; 817.9855 [M + 5H]	1022.2297 [M + 4H]; 817.9996 [M + 5H]	1a	8.69
I4	1371.3212 [M + 3H]; 1028.7429 [M + 4H]	1371.3408 [M + 3H]; 1028.7417 [M + 4H]	1a	9.04
I5	1363.3455 [M + 3H]; 1022.761 [M + 4H]	1363.3431 [M + 3H]; 1022.7612 [M + 4H]	1a	9.69
I6	1332.9911 [M + 3H]; 999.9952 [M + 4H]	1332.9893 [M + 3H]; 999.9955 [M + 4H]	1a	9.49
I7	1382.3563 [M + 3H]; 1037.0192 [M + 4H]	1382.3538 [M + 3H]; 1037.019 [M + 4H]	1a	9.28
I8	1117.8098 [M + 4H]; 894.4494 [M + 5H]	1117.8076 [M + 4H]; 894.4485 [M + 5H]	1a	9.43
I9	1132.3202 [M + 4H]; 906.0577 [M + 5H]	1132.5868 [M + 4H]; 906.0574 [M + 5H]	1a	9.82
I10	902.8588 [M + 5H]; 752.5503 [M + 6H]	902.8764 [M + 5H]; 752.5649 [M + 6H]	1a	9.80
I11	1121.8085 [M + 4H]; 897.6484 [M + 5H]	1121.8032 [M + 4H]; 897.6495 [M + 5H]	1a	9.39
I12	1461.0529 [M + 3H]; 1096.0416 [M + 4H]	1461.0526 [M + 3H]; 1096.0400 [M + 4H]	1a	9.03
I13	1470.4050 [M + 3H]; 1103.0558 [M + 4H]	1470.3911 [M + 3H]; 1103.0583 [M + 4H]	1a	9.34
I14	882.8401 [M + 5H]; 735.8681 [M + 6H]	882.8597 [M + 5H]; 735.8854 [M + 6H]	1a	8.54
I15	1466.0565 [M + 3H]; 1099.7943 [M + 4H]	1466.4026 [M + 3H]; 1099.7953 [M + 4H]	1a	8.5
I16	1340.335 [M + 3H]; 1005.5032 [M + 4H]	1340.3059 [M + 3H]; 1005.5063 [M + 4H]	1a	9.26
I17	1362.3142 [M + 3H]; 1021.9877 [M + 4H]	1362.5483 [M + 3H]; 1021.8204 [M + 4H]	1a	8.70
I18	1360.6700 [M + 3H]; 1020.7545 [M + 4H]	1360.6836 [M + 3H]; 1020.7534 [M + 4H]	1a	9.33
I19	1346.6824 [M + 3H]; 1010.2637 [M + 4H]	1346.6736 [M + 3H]; 1010.2637 [M + 4H]	1a	9.26

I20	1323.9983 [M + 3H]; 992.5007 [M + 4H]	1323.0339 [M + 3H]; 992.5002 [M + 4H]	1a	9.23
I21	1336.0020 [M + 3H]; 1002.2534 [M + 5H]	1335.8872 [M + 3H]; 1002.2549 [M + 4H]	1a	9.08
I22	900.8569 [M + 5H]; 750.8820 [M + 6H]	900.8657 [M + 5H]; 750.8822 [M + 6H]	1a	9.59
I23	1127.5597 [M + 4H]; 902.2493 [M + 5H]	1127.5278 [M + 4H]; 902.2489 [M + 5H]	1a	9.73
I24	1122.0544 [M + 4H]; 897.8452 [M + 5H]	1122.0796 [M + 4H]; 897.8453 [M + 5H]	1a	9.99
I25	1134.3243 [M + 4H]; 907.6610 [M + 5H]	1134.6296 [M + 4H]; 907.6615 [M + 5H]	1a	9.32
I26	1128.3215 [M + 4H]; 902.8588 [M + 5H]	1128.2800 [M + 4H]; 902.8578 [M + 5H]	1a	9.90
I27	1134.5612 [M + 4H]; 907.8505 [M + 5H]	1134.6018 [M + 4H]; 907.8507 [M + 5H]	1a	9.89
I28	923.5001 [M + 5H]; 769.7531 [M + 6H]	923.4941 [M + 5H]; 769.7401 [M + 6H]	1a	12.03
I29	1146.1094 [M + 4H]; 917.0891 [M + 5H]	1146.1101 [M + 4H]; 917.0784 [M + 5H]	1a	11.97
I30	1165.1282 [M + 4H]; 932.3041 [M + 5H]	1165.2490 [M + 4H]; 932.4043 [M + 5H]	1a	11.71
I31	1151.6174 [M + 4H]; 921.4955 [M + 5H]	1151.5132 [M + 4H]; 921.4021 [M + 5H]	1a	11.77
I32	1166.8687 [M + 4H]; 933.6965 [M + 5H]	116.9883 [M + 4H]; 933.5896 [M + 5H]	1a	11.91
I33	1156.3779 [M + 4H]; 925.3039 [M + 5H]	1156.4565 [M + 4H]; 925.3229 [M + 5H]	1a	11.80
I34	1138.6149 [M + 4H]; 911.0934 [M + 5H]	1138.7917 [M + 4H]; 911.2264 [M + 5H]	1a	11.54
I35	1161.8639 [M + 4H]; 929.6927 [M + 5H]	1161.8643 [M + 4H]; 929.8517 [M + 5H]	1a	10.65
I36	1219.3928 [M + 4H]; 975.7158 [M + 5H]	1219.3989 [M + 4H]; 975.7205 [M + 5H]	2b	7.57 (b)
I37	1221.1333 [M + 4H]; 977.1082 [M + 5H]	1221.3809 [M + 4H]; 977.1113 [M + 5H]	2b	7.74 (b)
I38	1210.6462 [M + 4H]; 968.7156 [M + 5H]	1210.8979 [M + 4H]; 968.7178 [M + 5H]	1a	11.84
I39	954.5052 [M + 5H]; 795.5889 [M + 6H]	954.5040 [M + 5H]; 795.5906 [M + 6H]	1a	7.41 (b)
O_I1	1247.1685 [M + 4H]; 997.9363 [M + 5H]	1247.4038 [M + 4H]; 997.9380 [M + 5H]	1a	10.84
O_I2	1243.1697 [M + 4H]; 994.7373 [M + 5H]	1243.1697 [M + 4H]; 994.7386 [M + 5H]	1a	11.21
O_I3	1254.1561 [M + 4H]; 1003.7271 [M + 5H]	1254.1621 [M + 4H]; 1003.7284 [M + 5H]	1a	10.74

O_I4	1247.4081 [M + 4H]; 998.1281 [M + 5H]	1247.4248 [M + 4H]; 998.1267 [M + 5H]	1a	10.74
O_I5	1253.9092 [M + 4H]; 1003.3290 [M + 5H]	1253.7849 [M + 4H]; 1003.3286 [M + 5H]	1a	10.88
O_I6	1237.1669 [M + 4H]; 989.9351 [M + 5H]	1237.2512 [M + 4H]; 989.9353 [M + 5H]	1a	11.10
O_I7	1252.9224 [M + 4H]; 1002.5395 [M + 5H]	1252.9207 [M + 4H]; 1002.5181 [M + 5H]	1a	11.11
O_I8	1222.6506 [M + 4H]; 978.3221 [M + 5H]	1222.6470 [M + 4H]; 978.5169 [M + 5H]	2b	6.58 (b)
O_I9	1006.7488 [M + 5H]; 839.1253 [M + 6H]	1006.7567 [M + 5H]; 839.1288 [M + 6H]	2b	7.18 (b)
O_I10	1006.7488 [M + 5H]; 839.1253 [M + 6H]	1006.7524 [M + 5H]; 839.1281 [M + 6H]	2b	7.2 (b)
O_I11	1002.3592 [M + 5H]; 835.4673 [M + 6H]	1002.3583 [M + 5H]; 835.4669 [M + 6H]	2b	6.86 (b)
O_I12	1026.9701 [M + 5H]; 855.9764 [M + 6H]	1026.7664 [M + 5H]; 855.9749 [M + 6H]	2b	7.01 (b)
O_I13	1029.3773 [M + 5H]; 857.9824 [M + 6H]	1029.3829 [M + 5H]; 857.9806 [M + 6H]	2b	6.97 (b)
O_C1	1293.9332 [M + 4H]; 1035.3481 [M + 5H]	1293.9041 [M + 4H]; 1035.3479 [M + 5H]	1a	11.12
O_C2	1284.4192 [M + 4H]; 1027.7369 [M + 5H]	1284.4553 [M + 4H]; 1027.8960 [M + 6H]	1a	11.18
O_C3	1016.1226 [M + 5H]; 846.9368 [M + 6H]	1016.1224 [M + 5H]; 846.9594 [M + 6H]	1a	11.79
O_C4	1269.1907 [M + 4H]; 1015.5541 [M + 5H]	1269.0168 [M + 4H]; 1015.5673 [M + 5H]	1a	10.71
O_C5	1268.1975 [M + 4H]; 1014.7595 [M + 5H]	1268.1962 [M + 4H]; 1014.7571 [M + 5H]	2b	6.93 (b)
O_C6	1275.9580 [M + 4H]; 1020.9680 [M + 5H]	1275.9554 [M + 4H]; 1020.9645 [M + 5H]	2b	6.90 (b)

5.4. Circular Dichroism

5.4.1. Circular dichroism scan at 25 °C

Circular dichroism. CD studies were performed in a JASCO J1500 at a 0.2nm resolution, 1.0 nm band width, 200 mdeg sensitivity, 4 s response time, 50nm/min scanning speed and, 25°C. Peptides were dissolved in water at 1mM concentrations and diluted to 100 µM in 50mM potassium phosphate buffer pH 7.5 for the measurements. The CD spectra of the solvent was recorded and subtracted from the raw data. Measurements were done in a 0.1 cm path length quartz cuvette, in the range of 250 to 190nm, for a total of 5 scans, the average signal is presented in mean residue ellipticity, MRE (deg·cm²·dmol⁻¹).

5.4.2. Thermal denaturation measurements

For the study of thermal stability/unfolding of the peptides, 100 µM solutions in 50 mM potassium phosphate buffer pH 7.5 were used. In most of the cases full thermal denaturation scans were recorded, where CD scans from 240-195 nm are measured in a 0.1cm cuvette, for a temperature range between 2-96°C with a slope increase of 5°C/min in intervals of 2°C and a hold time of 5 min.

The data was then converted into MRE and plotted versus T(K) to determine T_m, based on previously reported non liner fittings¹⁸⁵, assuming that the measured ellipticity is a linear combination of the signal coming from the folded and unfolded conformation:

$$\theta_{obs} = \alpha \cdot \theta_f + (1 - \alpha) \cdot \theta_u \quad (1)$$

θ_{obs} → corresponds to the observed/measured ellipticity, θ_f → corresponds to the ellipticity of the folded mini-protein, θ_u → is the ellipticity of the unfolded mini-protein and α → is the degree of folding.

Both contributions to the ellipticity are modelled as linear functions of temperature. Slope and y-intercept are determined by linear regression of the number of points belonging to the folded and unfolded regions. Equations 2-3

$$\theta_f = m_f \cdot T + b_f \quad (2)$$

$$\theta_u = m_u \cdot T + b_u \quad (3)$$

Where $m_f \rightarrow$ is the slope of folded baseline, $b_f \rightarrow$ is the y-intercept of folded baseline, $m_u \rightarrow$ is the slope of the unfolded baseline and $b_u \rightarrow$ is the y-intercept of the unfolded baseline.

The fraction of folded conformation, α , is directly dependent of the equilibrium constant of folding, K , (Equations 4-5) and at the same time the equilibrium constant is a function of the Gibbs free energy of unfolding, ΔG . Which can be obtained by combining equations 6-8.

$$\alpha = \frac{1}{1 + K} \quad (4)$$

$$K = e^{-\Delta G/RT} \quad (5)$$

$$\Delta H = \Delta H_m + C_p \cdot (T - T_m) \quad (6)$$

$$\Delta S = \Delta S_m + C_p \cdot \ln\left(\frac{T}{T_m}\right) \quad (7)$$

$$\Delta G = \Delta H - T \cdot \Delta S = \Delta H_m \cdot \left(1 - \frac{T}{T_m}\right) \quad (8)$$

Where $R \rightarrow$ is the constant of the ideal gases, $\Delta H_m \rightarrow$ is the enthalpy of folding, $T \rightarrow$ is the experimental values of temperature, $T_m \rightarrow$ is the melting point of the miniprotein, $\Delta S_m \rightarrow$ is the entropy of folding.

By combining all the previous equations, the relation between θ_{obs} and T can be obtained with the following equation:

$$\theta_{obs} = \frac{1}{1 + e^{\frac{-\Delta H \cdot (1 - \frac{T}{T_m})}{R \cdot T}}} \cdot (b_f - b_u - m_u \cdot T + m_f \cdot T) + b_u + m_u \cdot T \quad (9)$$

The above-described equation was introduced in OriginPro9.0 software as non-linear fit to fit the experimental data and obtain T_m and ΔH_m .

5.4.3. Thermodynamic studies

Thermodynamic studies by Guanidine-dependent denaturation.

Prior to the preparation of the mini-protein aliquots, an initial stock of 1mM concentration in distilled water was prepared. From this stock, 100 μ M solutions in 50 mM potassium phosphate buffer pH 7.5 in presence of GuHCl (pH 7.5) in different concentrations, with steps of 0.5M GuHCl, until complete denaturation (9-12 steps of GuHCl).

Each sample was incubated at room temperature for 30 min, upon completion of the incubation, the solutions were measured at 220 nm, using a 1mm optical length quartz cuvette. The temperature was increased from 4-96°C, in intervals of 2°C and a slope of 5°C/min and a hold time of 5 min.

All data collected from the experiments were converted into MRE units and analyzed using MatLab R2021b (The MathWorks, Inc) for the global fit from the data generated on the unfolding studies. The thermodynamic equations for the fitting assume that the θ_{obs} is dependent of K and the contributions of the folded and unfolded state.

$$\theta_{obs} = \frac{1}{1 + K} (\theta_f + K \cdot \theta_u) \quad (10)$$

The dependence of θ_f and θ_u to temperature and GuHCl concentration is defined by:

$$\theta_f = a + b \cdot T + c[GuHCl] \quad (11)$$

$$\theta_u = d + e \cdot T + f[GuHCl] \quad (12)$$

Where a, b, c, d, e and f are coefficients of these linear dependences. The equilibrium constant is defined as equation 5 and does not depend on GuHCl concentration. On the other hand, ΔG does depend on GuHCl, and it is defined by equation 13.

$$\Delta G^\circ = \Delta H^\circ - T \cdot \Delta S^\circ + \Delta C_p \cdot \left(T - T_0 + T \ln \left(\frac{T_0}{T} \right) \right) - m[GuHCl] \quad (13)$$

The following script was used to fit the experimental data as function of T and [GuHCl] to the equations described previously^{255,256}.

```
R = 8.3145; %Ideal Gas Constant in SI units
P = 1e-4; %Free Peptide Concentration in M
T0 = 298.15; %Reference Temperature in K
% 1 kcal = 4184 J exactly
%Make sure data is saved in three columns:
%Temp (Celcius), Theta, [Den] (Molar)
gu_max=5; %max concentration of guanidine [M]
gu_step = 0.5; % change of guanidine concentration in each step [M]
temp_min = 4; % min temperature [deg C]
temp_max = 96; %max temperature [deg C]
temp_step = 2; % temperature step [deg C]

nt=(temp_max-temp_min)/temp_step+1; % number of temperature points
ng=11; % number of guanidine points
xx = [ 0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5 ];
xx=transpose(xx)*ones(1,nt);
y=temp_min:temp_step:temp_max;
y=ones(ng,1)*y;
data11 = ones(ng*nt,3);
for i = 1:ng
    for j = 1:nt
        data11(j*ng-ng+i,1)=y(i,j);
        data11(j*ng-ng+i,2)=EHEEO11GuHCl220nm(i,j);
        data11(j*ng-ng+i,3)=xx(i,j);
    end
end
end
```

```

data=data11;
Temp = data(:,1) + 273.15; %Converts to Kelvin
Theta_Obsd = data(:,2);
Den = data(:,3);
Cond = [Temp Den];
%Cond is the n x 2 matrix of Temp and Den, and Theta_Obsd is output
%Define additional functions K, U, and f
K = @(b,Cond)exp(-(b(3)-Cond(:,1))*b(4)+b(5)*(Cond(:,1)-T0 ...
+Cond(:,1)).*log(T0/Cond(:,1)))-Cond(:,2)*b(6))./(R*Cond(:,1)));
U = @(b,Cond) arrayfun(@(k) fzero(@(x) (1+k)*x-k*P,[0 P]),K(b,Cond));
%U = @(b,Cond) K(b,Cond)*P/(1+K(b,Cond));
f = @(b,Cond) 1/P*((b(1)*ones(length(Cond(:,1)),1)+b(7).*Cond(:,1)+...
b(8).*Cond(:,2)).*U(b,Cond)...
+(b(2)*ones(length(Cond(:,1)),1)+b(9).*Cond(:,1)+b(10).*...
(Cond(:,2))).*(P*ones(length(Cond(:,1)),1)-U(b,Cond)));
%Initial Values Module
beta0 = zeros(10,1);
beta0(1) = max(Theta_Obsd)+1;
beta0(2) = min(Theta_Obsd)-1;
Uest = P*(Theta_Obsd-beta0(2)*ones(length(Theta_Obsd),1))/...
(beta0(1)-beta0(2));
Kest = Uest./(P*ones(length(Temp),1)-Uest);
DGest = -R*Temp.*log(Kest);
TestMat = ones(length(Temp),4);
for i = 1:length(Temp)
TestMat(i,1) = 1;
TestMat(i,2) = -Temp(i);
TestMat(i,3) = Temp(i)-T0+Temp(i)*log(T0/Temp(i));
TestMat(i,4) = -Den(i);
end

```

```

ParaEst = linsolve(TestMat,DGest);
beta0 = [max(Theta_Obsd);min(Theta_Obsd); ParaEst(1); ParaEst(2);...
ParaEst(3);ParaEst(4);0;0;0;0];
% Actual Curve Fitting Portion
[beta, r, J, COVB, mse] = nlinfit(Cond, Theta_Obsd, f, beta0);
ci = nlparci(beta,r,'covar', COVB);

t=temp_min:temp_step:temp_max;
t=t+273.15;
dG=ones(length(t),3);
for i=1:length(t)
    dG(i,1)=t(i);
    dG(i,2)=beta(3)-t(i)*beta(4)+ beta(5)*(t(i)-T0+t(i)*log(T0/t(i)));
    dG(i,3)=exp(-dG(i,2)/(R*t(i)));
end

fitQ = ones(ng,nt);
for i = 1:ng
    for j = 1:nt
        %          g=ParaEst(1)-t(j)*ParaEst(2)+ParaEst(3)*(t(j)-T0+t(j)*log(T0/t(j)))-
ParaEst(4)*xx(i);
        g=beta(3)-t(j)*beta(4)+ beta(5)*(t(j)-T0+t(j)*log(T0/t(j)))-beta(6)*xx(i);
        k=exp(-g/(R*t(j)));

fitQ(i,j)=1/(k+1)*((beta(1)+beta(7)*t(j)+beta(8)*xx(i,j))*k+(beta(2)+beta(9)*t(j)+beta(10)
)*xx(i,j));
    end
end

```

5.5. NanoDSF

For the denaturation and renaturation studies, nano differential scanning fluorimetry (NanoDSF) was performed. NanoDSF measures changes of fluorescence at 350 and 330 nm versus temperature.

Data was collected with a Prometheus (NanoTemper), from 15-110 °C, at a ramp rate of 2 °C/min, and a measurement interval of 0.02 °C, for the unfolding and folding stages. Samples were prepared at final concentration of 100 µM, in potassium phosphate buffer pH 7.5, 50 mM. The first derivative of the ration between 350 and 330 nm was plotted as a function of the temperature and T_m was determined as the peak of the derivative curve.

5.6. BioLayer Interferometry

Biolayer interferometry kinetic assay was performed on an Octet K2 (Sartorius), by immobilising the target protein on a biosensor, and leaving a second sensor unloaded to measure non-specific bindings (NSB), we measured changes on light during the association and dissociation stages of the experiment allowing to estimate the binding kinetics of the inhibitors. All binding studies were performed in the kinetic buffer (KB) consisting of 10 mM HEPES, 150 mM NaCl and 3mM EDTA, pH 7.4, supplemented with 0.05 % (v/v) Tween-20 and 1% (w/v) bovine serum albumin. Biotinylated human PD-L1 (ECD, His Tag) (1 µg/mL), was loaded over a high precision streptavidin biosensor (SAX), for 300 seconds, while for the reference sensor the process was carried out in plain KB. Both sensors were then stabilised for 120 seconds in KB. Following the baseline, both sensors were dipped in wells containing Biocytin quencher (1 µg/mL) for 60 seconds to block the active sites of the streptavidin. Quenching step was succeeded by a 120 second baseline. Following the loading and quenching, the sensors were then dipped in solutions of the inhibitors to measure association for 300 seconds and then dipped in wells containing KB for the dissociation for 480 seconds. The sensors were regenerated between measurements by dipping them in 10 mM glycine solution pH 1.7 for 60 seconds followed by 60 seconds in KB, this action was repeated two times. The measurements were carried out in a dose-dependent manner (0.156-20 µM) of the inhibitors.

Recombinant protein Biotinylated PD-L1 (ECD, His Tag) N-terminal segment of the extracellular domain (Met1-Thr239) was purchased from Sinobiological.

The data collected for the unloaded sensor was then subtracted from the loaded sensor data to remove NSB effects on the affinity calculations. The data was processed with the Data Analysis HT software from Octet and adjusted to the global fitting model.

5.7. Homogeneous time resolved fluorescence

Homogeneous Time Resolved Fluorescence (HTRF) assay is a cell-free assay that enables the measurement of the PD-1/PD-L1 interaction. The Interaction between PD-1 and PD-L1 generates an energy transfer between anti-Tag1 labelled with Europium (HTRF donor) and anti-Tag2 labelled with XL665 (HTRF acceptor). When these two antibodies are in spatial proximity, due to the binding between PD-1 and PD-L1, the donor Europium excitation generates a fluorescence resonance energy transfer (FRET) towards the XL665 acceptor, whose emission is at 665 nm. The ratio between these signals is directly proportional to the percentage of PD-1/PD-L1 complexes. The disruption of the interaction decreases the fluorescence intensity of the acceptor causing a reduction of the HTRF signal.

All binding studies were performed in the HTRF assay buffer using a Cisbio HTRF 96-well low volume white plate, and the reconstitution of the proteins and anti-Tags were performed following the assay's protocol. Initially, 4 μ L of Tag1-PD-L1 were placed in the well, followed by 2 μ L of the inhibitor at x10 the final concentration. To the premixture of Tag1-PD-L1 with the inhibitor was added 4 μ L of Tag2-PD-1 and incubated for 15 minutes before the addition of the anti-Tags. The anti-Tags (Anti-Tag1 Eu Cryptate and Anti-Tag2 XL665) were premixed prior to addition of 10 μ L of the mixture into the wells. The final mixture of reagents was incubated for one hour at room temperature before the measurement of the fluorescence. The measurement of the signal ($\lambda_{\text{ex}}=320$ nm; $\lambda_{\text{em}}=665$ nm/620 nm) was performed in a Clariostar Plus fluorometer. The integration delay was 50 μ s with an integration time of 200 μ s, with 200 flashes and 0.5 s of settling time.

The assay buffer was used as blank. Tag1-PD-L1 in absence of inhibitor and Tag2-PD-1 was used as negative control. Mixture of proteins with anti-Tags in absence of inhibitor was used as a positive control. In addition, BMS1166 at 5 nM concentration was

used as a second positive control (IC₅₀=1.7 nM). All measurements were performed in triplicate.

5.8. Cell-based assay

The cell-based assay at physiological pH led to poor solubility of the inhibitors. To explore to what extent the pH could be modified and to determine the conditions to control the final pH of the medium, a series of double dilutions, in water, of 37 % HCl (or 12M) were mixed with the medium at 6.6 µL/mL. The acidified medium was mixed in a 1:1 ratio with the unmodified medium to achieve the final conditions of the wells and the measured pH. The results indicated that the ideal condition for the assay was the addition of 5.5 µL of 6M HCl per 1 ml of medium. Still, as mentioned before, the buffer system of the RPMI 1460 media tends to alkalise in the presence of oxygen. This was confirmed by measuring the pH of fresh media with 1% FBS (pH ~7.6), and the medium used for the above-mentioned experiments which was already exposed to oxygen (pH ~8.4), as well as fresh media in the absence of FBS (pH ~7.5). Following the previous results, we decided to use alkalised medium for the assay, since the pH range was within the range allowed by the cells (>6.8).

The assay was performed using Promega cell-based assay. The PD-1/PD-L1 blockade bioassay, is a bioluminescent cell-based assay. For the assay, two genetically engineered cell lines were purchased and used:

- I) PD-1 Effector Cells: Jurkat T cells overexpressing human PD-1 and a luciferase reporter driven by an NFAT response element (NFAT-RE)
- II) PD-L1 aAPC/CHO-K1 Cells: CHO-K1 cells overexpressing human PD-L1, and an engineered cell surface protein designed to activate cognate TCRs, further referred as APC cells.

The cells were cultured in RPMI 1640 medium (Lonza) supplemented with 10% Fetal Bovine Serum (BioWest), 100 U/ml Penicillin and 100 U/ml Streptomycin. Before the experiments both cell lines were propagated in a constant presence of Hygromycin B (50 µg/ml) and G418 (250 µg/ml) to provide a stable presence of the introduced genetic constructs. The overexpression of PD-1 and PD-L1 was confirmed by flow cytometry. The

cells were periodically tested and found negative for Mycoplasma contamination using PCR-based method. 24h prior to the experiment, 10000 APC cells, per well, were seeded. Modifications of the assay conditions were required for the adequate performance of the inhibitors. On the day of the assay, the required volume of culture media to prepare the inhibitor stocks, was pH adjusted with 5,5 $\mu\text{L}/\text{mL}$ of 6M HCl. With the pH modified media, stocks of inhibitors at two-fold final concentrations were prepared. 40 μL of inhibitor stocks were added to the wells containing APC cells and diluted to final assay concentrations with 40 μL of unmodified pH culture media. The mixture of medias was adjusting the pH to 6.8, suitable for cell line survival. The inhibitors were then incubated for 24h at 37 °C with 5% CO₂. Since the stocks were prepared in water, all preparations were done to keep the final concentration of water constant. Following the 24h incubation of APC with the inhibitors, 10 μL of media was removed and immediately after 20000 PD-1 Effector Cells in 10 μL of culture media were added. The cells were incubated for another 6 hours and equilibrated at room temperature for 10 minutes prior to addition of Bio-Glo reagent (Promega). The luminescence was determined after further 20 min incubation. Half maximal effective concentrations (EC50) and maximal luminescence values (RLU_{max}) were determined by fitting the Hill equation to the experimental data.

5.9. Small angle X-ray scattering

SAXS measurements were performed in batch mode at 2.5 and 5 mg/mL concentrations in Postassium phosphate buffer pH 7.5 50 mM. at DESY (Deutsches Elektronen-Synchrotron), Hamburg, Germany, beamline P12. Scattering data was recorded during an exposure time of 0.095 s using a 2D Pilatus 6M at a photon energy of 10.0 keV ($\lambda=1.24 \text{ \AA}$) at 3.0 m sample-to-detector distance. Four frames were merged, integrated and background subtracted on the matching buffer.

When samples at higher concentration were not having indications of aggregation, they were merged with the lower concentration data to improve the signal-to-noise ratio, to do so Primus software²⁵⁷ was used. The pre-processed data was analysed using ATSAS package software²⁵⁸.

5.10. NMR studies and calculations

NMR experiments were performed on a Bruker AdvanceTM III 600.58 MHz spectrometer using a 5 mm PA BBO probe. NMR was collected for samples in the range of 2-12 mg/mL dissolved in 25 mM potassium phosphate buffer pH 6 with 10 % D₂O at temperature ranging between 298-311 ± 0.1 K. ¹H NMR spectra were recorded using 3-9-19 pulse sequence water suppression with gradients. TOCSY and NOESY experiments were recorded at 90 and 250 ms mixing times respectively. The spectra had a width of 8620 Hz in both dimensions. While for HSQC, a mixing time of 90 ms and spectral width F2: 8620 Hz and F1: 30205 Hz.

The data were acquired using Topspin 4.1.3 and processed using Topspin 3.1 (BrukerBioSpin, Rheinstetten, Germany). The processed spectra were assigned with the help of the SPARKY program²⁵⁹.

5.10.1 Chemical shifts tables from the NMR experimental results.

Table 32. Chemical shifts and contacts of miniprotein 47.

Residue	Proton	Chemical shift (ppm)	
Glu1	HA	4.113	
	HB1	2.029	
	HB2	1.982	
	HG	2.328	
	HN	7.719	
Thr2	HA	4.402	
	HB	4.184	
	HG	1.200	
	HN	8.550	
Trp3	HA	4.656	
	HB	3.251	
	HD1	7.232	
	HE1	10.017	
	HE3	7.554	
	HH2	7.100	
	HN	8.407	
	HZ2	7.391	
	HZ3	7.038	
	Ile4	GB	1.610
HA		3.873	
HB		1.613	
	HD	0.718	
	HG11	1.171	
	HG12	0.937	
	HG2	0.644	
	HN	7.806	
Glu5	HA	4.023	
	HB1	1.843	
	HB2	1.789	
	HG1	2.194	
	HG2	2.132	
	HN	7.859	
	Phe6	HA	4.573
		HB1	3.103
HB2		2.913	
HN		8.060	
Hm-Ar		7.189	
Ho-Ar		7.142	
HoAr		7.140	
Hp-Ar		7.136	
Thr7	HA	4.082	
	HB	3.932	
	HG	1.004	
Val8	HN	7.721	
	HA	3.587	

	HB	1.248
	HG1	0.020
	HG2	-0.156
	HN	7.162
Thr9	HA	4.474
	HB	4.072
	HG	1.198
	HN	7.651
Lys10	HA	3.861
	HB1	1.823
	HB2	1.759
	HD1	1.690
	HD2	1.639
	HE	3.016
	HG	1.420
	HN	8.361
Glu11	HA	4.069
	HB1	2.037
	HB2	1.931
	HG	2.424
	HN	8.165
Cpt12	CH21	1.890
	CH22	1.704
	CH23	1.607
	CH24	1.492
	HA	2.583
	HB	4.202
	HN	7.161
Tyr13	HA	3.619
	HB	2.662
	HD	5.921
	HE	6.129
	HN	7.297
Asp14	HA	4.506
	HB1	3.079
	HB2	2.819
	HN	8.673
Lys15	HA	4.091
	HB1	1.846
	HB2	1.790
	HD1	1.684
	HD2	1.616
	HE	2.975
	HG	1.551
	HN	7.922
Cpt16	CH21	1.918
	CH22	1.846
	CH23	1.778

	CH24	1.670
	HA	3.182
	HB	4.358
	HN	7.965
Lys17	HA	3.997
	HB1	2.251
	HB2	2.058
	HD1	1.902
	HD2	1.779
	HE	3.124
	HG	1.635
	HN	9.087
	HX	2.142
Gln18	HA	4.055
	HB1	2.227
	HB2	2.169
	HG1	2.527
	HG2	2.416
	HN	7.978
Cp19	CH21	2.075
	CH22	1.967
	CH23	1.743
	CH24	1.635
	HA	2.674
	HB	4.197
	HN	7.461
Ala20	HA	3.874
	HB	1.488
	HN	8.995
Arg21	HA	4.353
	HB	2.056
	HD	3.292
	HE	7.303
	HG1	2.059
	HG2	1.808
	HN	7.383
Lys22	HA	4.501
	HB	2.072
	HD1	1.727
	HD2	1.675
	HE1	2.998
	HE2	2.938
	HG1	1.535
	HG12	1.371
	HG2	1.395
	HN	8.089
Ile23	HA	4.629
	HB	1.708

	HD	0.673
	HD1	0.672
	HG1	1.525
	HG11	1.508
	HG12	1.379
	HG2	0.779
Pro24	HN	7.049
	HA	4.012
	HB1	2.266
	HB2	1.163
	HD	2.510
	HD1	3.206
	HD2	2.507
Pro25	HG	1.608
	HA	4.425
	HB1	2.353
	HB2	1.903
	HD1	3.712
	HD2	3.476
	HG1	2.056
Gly26	HG2	2.016
	HA1	4.198
	HA2	3.799
Trp27	HN	8.582
	HA	5.263
	HB	3.183
	HD1	6.970
	HE1	9.736
	HE3	7.399
	HH2	6.279
	HN	7.962
	HZ2	6.900
HZ3	6.823	
Asp28	HA	5.341
	HB1	2.670
	HB2	2.570
	HN	9.283
Ile29	HA	5.228
	HB	1.583
	HD	0.756
	HG1	1.549
	HG11	1.127
	HG12	1.057
	HG2	0.843
Ser30	HN	8.895
	HA	5.864
	HB1	3.958
	HB2	3.691

Phe31	HN	9.114
	HA	6.028
	HB	3.598
	HN	9.232
	Hm-Ar	6.774
	Ho-Ar	7.411
	Hp-Ar	6.723
Thr32	HA	5.566
	HB	3.914
	HG	1.053
Ser33	HN	9.302
	HA	4.286
	HB1	2.572
	HB2	2.137
Asn34	HN	8.409
	HA	4.350
	HB1	3.094
	HB2	2.686
Gly35	HN	9.301
	HA1	4.114
	HA2	3.544
Lys36	HN	8.647
	D2	1.288
	HA	4.450
	HB	1.509
	HD1	1.405
	HD2	1.292
	HE1	2.576
	HE2	2.482
	HG1	1.078
	HG2	0.815
Val37	HN	7.708
	HZ	7.261
	HA	4.059
	HB	1.941
	HG1	0.919
Trp38	HG2	0.864
	HN	8.483
	HA	4.718
	HB1	3.538
	HB2	3.059
	HD1	6.906
	HE1	9.825
	HE3	7.785
	HH2	7.263
Ser30	HN	8.897
	HZ2	7.441
	HZ3	7.230

Trp39	HA	4.750
	HB1	3.520
	HB2	3.051
	HD1	7.151
	HE1	9.831
	HE3	7.093
	HH2	6.847
	HN	7.868
	HZ2	7.297
HZ3	6.537	
Thr40	HA	5.110
	HB	3.718
	HG	0.924

Ala41	HN	7.432
	HA	3.922
	HB	1.127
	HN	8.685
Arg42	HA	4.564
	HB1	1.829
	HB2	1.693
	HD	3.198
	HE	7.265
	HG	1.608
NH2	HN1	8.513
	HN2	7.124
		7.828

Sequential (i, i+1)	Intensity
Glu1HA-Thr2HN	s
Thr2HA-Trp3HN	s
Thr2HB-Trp3HN	m
Thr2HB-Trp3HZ3	w
Thr2HG-Trp3HN	m/w
Thr2HN-Trp3HN	m
Trp3HA-Ile4HN	s
Trp3HB-Ile4HN	m
Trp3HE3-Ile4HN	w
Trp3HN-Ile4HN	m
Ile4HA-Glu5HN	s
Ile4GB-Glu5HN	m
Ile4HG2-Glu5HN	m
Ile4HD-Glu5HN	w
Ile4HN-Glu5HN	m
Glu5HA-Phe6HN	s
Glu5HN-Phe6HN	m
Phe6HA-Thr7HN	s
Phe6HB1-Thr7HN	m
Phe6HB2-Thr7HN	m/w
Phe6HN-Thr7HN	m
Thr7HA-Val8HN	m
Thr7HB-Val8HN	m
Thr7HG-Val8HN	m
Thr7HN-Val8HN	m
Val8HA-Thr9HN	s
Val8HB-Thr9HN	m
Val8HG1-Thr9HN	s
Val8HG2-Thr9HN	w
Val8HN-Thr9HG	m
Val8HN-Thr9HN	m
Thr9HA-Lys10HN	m
Thr9HB-Lys10HN	s

Thr9HG-Lys10HN	m
Thr9HN-Lys10HN	w
Lys10HA-Glu11HN	m/w
Lys10HB1-Glu11HN	m
Lys10HB2-Glu11HN	m
Lys10HD1-Glu11HN	w
Lys10HD2-Glu11HN	w
Lys10HG-Glu11HN	w
Lys10HN-Glu11HN	m
Glu11HA-Cp12HN	s
Glu11HB1-Cp12HN	m
Glu11HB2-Cp12HN	m
Glu11HG-Cp12HN	w
Glu11HN-Cp12HN	m
Cp12HA-Tyr13HN	s
Cp12HB-Tyr13HN	w
Cp12HN-Tyr13HB	w
Cp12HN-Tyr13HN	m
Tyr13HA-Asp14HN	m/w
Tyr13HB-Asp14HN	s
Tyr13HD-Asp14HN	w
Tyr13HN-Asp14HN	m
Asp14HA-Lys15HN	m/w
Asp14HB1-Lys15HN	m
Asp14HB2-Lys15HN	m
Asp14HN-Lys15HN	m
Lys15HA-Cp16HN	m
Lys15HN-Cp16HN	m
Cp16HA-Lys17HN	s
Cp16HB-Lys17HN	w
Cp16HN-Lys17HN	m
Lys17HA-Gln18HN	m
Lys17HB1-Gln18HN	s
Lys17HB2-Gln18HN	m

Lys17HG-Gln18HN	m
Lys17HN-Gln18HB2	w
Lys17HN-Gln18HN	m/s
Gln18HA-Cp19HN	m
Gln18HB1-Cp19HN	m
Gln18HB2-Cp19HN	m
Cp19CH22-Ala20HN	w
Cp19CH23-Ala20HN	m
Cp19HA-Ala20HN	s
Cp19HB-Ala20HN	w
Gln18HN-Cp19HN	w
Cp19HA-Ala20HB	m
Cp19HN-Ala20HN	m
Ala20HA-Arg21HN	m/w
Ala20HB-Arg21HN	m
Ala20HN-Arg21HG1	m/w
Ala20HN-Arg21HN	m
Ala20HN-Arg21HA	m
Arg21HA-Lys22HN	m
Arg21HB-Lys22HN	s
Arg21HN-Lys22HN	s
Arg21HN-Lys22HA	w
Lys22HN-Ile23HG2	m
Lys22HN-Ile23HN	s
Ile23HD-Pro24HD2	m/w
Ile23HD1-Pro24HD1	s
Ile23HG2-Pro24HD1	m
Ile23HG2-Pro24HD2	m
Ile23HN-Pro24HD1	m/w
Pro24HB1-Pro25HD1	m/w
Pro24HB1-Pro25HD2	m
Pro24HG-Pro25HD1	w
Pro24HG-Pro25HD2	m/w
Pro25HA-Gly26HN	s
Pro25HB1-Gly26HN	m
Pro25HB2-Gly26HN	m
Pro25HD1-Gly26HN	w
Pro25HG1-Gly26HN	w
Pro25HG2-Gly26HN	w
Gly26HA1-Trp27HD1	w
Gly26HA2-Trp27HD1	w
Gly26HA2-Trp27HN	m
Gly26HN-Trp27HB	m
Gly26HN-Trp27HN	s
Gly26HN-Trp27HD1	w
Trp27HA-Asp28HN	s
Trp27HB-Asp28HN	w
Trp27HD1-Asp28HB2	w

Trp27HD1-Asp28HN	w
Trp27HE3-Asp28HB2	m/w
Trp27HN-Asp28HB1	m/w
Trp27HN-Asp28HN	w
Asp28HA-Ile29HN	s
Asp28HB1-Ile29HN	m
Asp28HB1-Ile29HG2	m/w
Asp28HB1-Ile29HD	m
Asp28HB2-Ile29HN	m
Asp28HN-Ile29HN	w
Ile29HA-Ser30HA	w
Ile29HA-Ser30HN	s
Ile29HB-Ser30HN	m/w
Ile29HD-Ser30HN	w
Ile29HG12-Ser30HN	w
Ile29HG2-Ser30HN	m
Ile29HN-Ser30HN	w
Ile29HN-Ser30HA	w
Ser30HA-Phe31HN	s
Ser30HA-Phe31Ho-Ar	w
Ser30HA-Phe31HB	w
Ser30HB1-Phe31HN	m
Ser30HB2-Phe31HN	m
Ser30HN-Phe31HN	m
Ser30HN-Phe31Ho-Ar	w
Ser30HN-Phe31HA	m/w
Phe31HA-Thr32HN	s
Phe31HA-Thr32HG	w
Phe31HB-Thr32HN	m
Phe31HN-Thr32HN	m
Phe31HN-Thr32HG	w
Phe31Ho-Ar-Thr32HA	w
Thr32HA-Ser33HN	s
Thr32HB-Ser33HN	s
Thr32HG-Ser33HN	m
Thr32HN-Ser33HN	m
Ser33HA-Asn34HN	s
Ser33HB1-Asn34HN	m
Ser33HB2-Asn34HN	m/w
Asn34HA-Gly35HN	s
Asn34HB1-Gly35HN	m
Asn34HB2-Gly35HN	m
Asn34HN-Gly35HN	m
Gly35HA1-Lys36HN	m
Gly35HA2-Lys36HN	m
Gly35HN-Lys36HN	m
Lys36HA-Val37HN	s
Lys36HB-Val37HN	m

Lys36HG2-Val37HN	m/w
Lys36HN-Val37HN	w
Val37HA-Trp38HD1	m
Val37HA-Trp38HN	s
Val37HG1-Trp38HD1	w
Val37HN-Trp38HN	w
Trp38HA-Trp39HN	m
Trp38HB1-Trp39HN	m
Trp38HD1-Trp39HN	w
Trp38HN-Trp39HN	m
Trp39HA-Thr40HN	s
Trp39HB1-Thr40HN	m/w
Trp39HB2-Thr40HN	w
Trp39HE3-Thr40HN	w

Trp39HE3-Thr40HA	m
Trp39HZ3-Thr40HA	w
Thr40HA-Ala41HN	s
Thr40HB-Ala41HN	m/w
Thr40HG-Ala41HN	m
Thr40HN-Ala41HN	m
Ala41HA-Arg42HN	s
Ala41HB-Arg42HN	m
Ala41HN-Arg42HN	m/w
Arg42HA-NH2HN1	s
Arg42HA-NH2HN2	s
Arg42HB2-NH2HN2	m/w
Arg42HN-NH2HN1	w
Arg42HN-NH2HN2	w

Non sequential (i, i+2)	Intensity
Thr2HA-Ile4HN	w
Thr2HB-Ile4HN	m/w
Trp3HZ2-Glu5HN	w
Thr9HA-Glu11HN	m
Thr9HG-Glu11HN	w
Lys10HN-Cp12HN	w
Cp12HA-Asp14HN	m/w
Cp12HN-Asp14HN	w
Asp14HB1-Cp16HN	m/w
Lys17HN-Cp19HN	w
Gln18HN-Ala20HN	w
Cp19HB-Arg21HN	m/w
Ala20HA-Lys22HN	m/w
Ala20HN-Lys22HN	w
Arg21HA-Ile23HN	m/w
Arg21HG1-Ile23HN	m
Pro25HA-Trp27HN	m
Trp27HB-Ile29HN	w
Trp27HZ3-Ile29HN	w
Ile29HA-Phe31HA	m
Ile29HA-Phe31Ho-Ar	m
Ile29HD-Phe31Hm-Ar	m
Ile29HD-Phe31Ho-Ar	m
Ile29HD-Phe31Hp-Ar	m
Ile29HG2-Phe31HA	w
Ile29HG2-Phe31Hm-Ar	m
Ile29HG2-Phe31Ho-Ar	m
Ile29HG2-Phe31Hp-Ar	m

Ile29HN-Phe31HN	w
Ile29HN-Phe31HA	w
Ile29HN-Phe31Ho-Ar	w
Ser30HB2-Thr32HG	m/w
Ser30HN-Thr32HN	w
Ser33HA-Gly35HN	m/w
Ser33HN-Gly35HN	m/w
Asn34HA-Lys36HN	w
Asn34HB2-Lys36HN	w
Asn34HN-Lys36HN	w
Lys36HB-Trp38HD1	m/w
Lys36HB-Trp38HZ2	m
Lys36HE1-Trp38HZ2	w
Lys36HN-Trp38HD1	w
Lys36HN-Trp38HE1	s
Val37HA-Trp39HN	m
Val37HA-Trp39HZ2	m/w
Val37HG1-Trp39E3	m/w
Val37HG1-Trp39HN	m/w
Val37HG2-Trp39HN	m
Trp39HE3-Ala41HN	w
Trp39HH2-Ala41HN	w
Trp39HH2-Ala41HA	m/w
Trp39HH2-Ala41HB	m
Trp39HZ2-Ala41HA	m/w
Trp39HZ2-Ala41HB	m
Trp39HZ3-Ala41HN	w
Trp39HZ3-Ala41HB	m

Non sequential (i, i+3)	Intensity
Thr9HA-Cp12HN	m/w
Lys10HA-Tyr13HD	m

Lys10HA-Tyr13HE	w
Lys10HA-Tyr13HN	m/w
Lys10HA-Tyr13HB	m

Glu11HA-Asp14HN	s
Glu11HA-Asp14HB2	w
Glu11HA-Asp14HB1	m
Glu11HN-Asp14HN	w
Cp12HB-Lys15HN	m
Tyr13HA-Cp16HN	w
Tyr13HA-Cp16HA	m
Asp14HA-Lys17HN	m/w
Cp16HB-Cp19HN	m
Lys17HA-Ala20HN	m
Lys17HA-Ala20HB	m
Lys17HB1-Ala20HN	w
Lys17HB1-Ala20HB	m
Lys17HD-Ala20HN	w
Lys17HG-Ala20HN	m/w
Lys17HN-Ala20HB	w
Gln18HA-Arg21HN	m

Cp16HN-Cp19HB	m
Cp19HB-Lys22HN	m
Ala20HA-Ile23HN	m
Ala20HN-Ile23HD	m/w
Pro24HB1-Trp27HD1	m
Pro24HB2-Trp27HD1	m
Pro24HB2-Trp27HE1	w
Pro24HD1-Trp27HD1	m
Pro24HD2-Trp27HD1	m
Pro24HG-Trp27HD1	m
Thr32HB-Gly35HN	m/w
Thr32HG-Gly35HN	w
Ser33HN-Lys36HN	m
Ser33HN-Lys36HB	m/w
Lys36HB-Trp39HE1	m
Trp38HN-Ala41HN	m

Non sequential (i, i+4)	Intensity
Val8HG1-Cp12CH22	m
Val8HG1-Cp12CH23	m/w
Val8HG1-Cp12CH24	m
Val8HG1-Cp12HA	m/s
Val8HG2-Cp12CH22	m/w
Val8HG2-Cp12CH23	w
Val8HG2-Cp12CH24	m/w
Thr9HG-Tyr13HE	w
Lys10HA-Asp14HN	m/w
Lys10HN-Asp14HN	w
Cp12CH24-Cp16HN	m/w
Tyr13HA-Lys17HN	w
Tyr13HB-Lys17HN	w
Tyr13HN-Lys17HA	w
Tyr13HN-Lys17HN	w
Tyr13HN-Lys17HX	w

Tyr13HN-Lys17HE	m
Tyr13HN-Lys17HD	m
Tyr13HN-Lys17HB1	m
Asp14HA-Gln18HN	m/w
Lys17HA-Arg21HN	w
Ile23HA-Trp27HD1	w
Ile23HD-Trp27HD1	m/w
Ile23HD-Trp27HE3	m
Ile23HD-Trp27HN	m/w
Ile23HG2-Trp27HE3	w
Ile23HG2-Trp27HN	w
Thr32HA-Lys36HN	w
Thr32HG-Lys36HN	m
Ser33HN-Val37HG1	w/m
Ser33HN-Val37HG2	w
Ser33HN-Val37HA	m

Long range contacts	Intensity
Trp3HE3-Trp38HZ2	m
Trp3HN-Trp38HD1	m
Trp3HN-Trp38HE1	m
Trp3HN-Trp38HN	w
Thr7HG-Trp38HD1	w
Val8HB-Tyr13HE	w
Val8HG1-Tyr13HD	w
Val8HG1-Tyr13HE	w
Val8HG1-Tyr13HN	w
Val8HG1-Phe31Hm-Ar	m/w
Val8HG1-Phe31Ho-Ar	m/w
Val8HG1-Phe31Hp-Ar	w

Val8HG2-Tyr13HE	w
Val8HG2-Phe31Hm-Ar	w
Val8HG2-Phe31Ho-Ar	m
Val8HG2-Phe31Hp-Ar	w
Tyr13HA-Phe31Hm-Ar	m
Tyr13HA-Phe31Hp-Ar	m
Tyr13HA-Trp39HZ3	w
Tyr13HD-Phe31Hm-Ar	m/w
Tyr13HD-Phe31Ho-Ar	w
Tyr13HD-Phe31Hp-Ar	w
Tyr13HE-Phe31Hm-Ar	m
Tyr13HE-Phe31Ho-Ar	m/w
Tyr13HE-Phe31Hp-Ar	m

Tyr13HE-Trp38HD1	w
Tyr13HE-Trp39E3	w
Tyr13HE-Trp39HB2	m/w
Tyr13HE-Trp39HB1	m/w
Tyr13HE-Ile29HG11	w
Tyr13HE-Trp39HD1	m/w
Asp14HA-Lys17HN	m/w
Asp14HB1-Lys15HN	m
Cp16HA-Phe31Hm-Ar	m/w
Cp16HA-Phe31Hp-Ar	m/w
Lys17HA-Trp39HH2	m/w
Lys17HB1-Phe31Hm-Ar	m/w
Lys17HB1-Phe31Hp-Ar	w
Lys17HB1-Trp39HH2	m/w
Lys17HB1-Trp39HZ3	m/w
Lys17HB1-Ile29HG2	m
Lys17HB1-Ile29HD	m
Lys17HX-Ile29HD	m
Lys17HX-Ile29HG2	m
Lys17HX-Phe31Hm-Ar	m
Lys17HX-Phe31Hp-Ar	m
Lys17HN-Trp39HZ3-	w
Lys17HN-Ile29HG11	w
Lys17HN-Trp39HH2	m/w
Ala20HB-Trp27HD1	m
Ala20HB-Ile29HN	m/w
Ala20HB-Trp39HH2	m/w
Ala20HN-Ile29HD	m
Ala20HN-Ile29HG11	w
Ala20HN-Ile29HG2	w
Lys22HA-Ile23HN	m
Lys22HA-Trp27HD1	w
Ile23HD-Asp28HA	w
Ile23HD-Asp28HN	m/w
Ile23HD-Ser30HN	w
Ile23HG2-Asp28HN	w
Trp27HB-Trp39HH2	m
Trp27HB-Trp39HZ3	m/w
Trp27HE1-Arg42HG	w
Trp27HH2-Trp39HD1	w
Trp27HH2-Trp39HB2	m/w
Trp27HH2-Trp39HE1	w
Trp27HZ3-Trp39B2	m
Trp27HZ3-Trp39HB1	m
Trp27HE3-Thr40HA	w

Asp28HA-NH2HN1	w
Asp28HA-NH2HN2	m/w
Asp28HA-Arg42HA	m/w
Asp28HB1-Thr40HG	m
Asp28HB2-Thr40HG	m
Asp28HN-Thr40HG	w
Ile29HD-Trp39HZ3	w
Ile29HG12-Val37HN	m
Ile29HG2-Trp39E3	w/m
Ile29HG2-Trp39HZ3	m
Ile29HN-Val37HB	w
Ile29HN-Val37HG2	m
Ile29HN-Val37HG1	m
Ser30HA-Thr40HA	m
Ser30HA-Trp39HE3	m
Ser30HA-Val37HG2	m/w
Ser30HA-Val37HG1	m
Ser30HB1-Val37HG1	m/w
Ser30HB1-Val37HG2	m/w
Ser30HN-Val37HG1	w
Phe31HB-Trp39HE3	m
Phe31HB-Trp39HN	m/w
Phe31HN-Val37HG2	m/w
Phe31HN-Val37HG1	w
Phe31HN-Trp38HB2	w
Phe31HN-Trp39HE3	m/w
Phe31HN-Trp39HN	w
Phe31HN-Thr40HA	m
Phe31Ho-Ar-Trp39HZ3	w/m
Phe31Ho-Ar-Trp38HB2	m
Thr32HA-Trp38HD1	m
Thr32HA-Trp38HE1	w
Thr32HA-Trp38HN	m
Thr32HA-Val37HG1	w
Thr32HA-Val37HA	m
Thr32HB-Trp38HD1	m/w
Thr32HB-Trp38HN	w
Thr32HG-Val37HN	w
Thr32HG-Trp38HD1	m/w
Ser33HB1-Trp38HD1	m
Ser33HB1-Trp38HE1	m
Ser33HB2-Trp38HD1	m
Ser33HB2-Trp38HE1	m/w
Ser33HN-Trp38HB1	w
Ser33HN-Trp38HB2	w

5.11. Crystallography

For the crystallization, purified and lyophilised peptides were dissolved in milliQ water at concentrations ranging between 10-20 mg/mL. The peptide solutions in water were mixed with each of the screen conditions on volume-volume ratios of 1:1, 1:2 and 2:1. Crystallization conditions were screened either manually by hanging drop vapor diffusion or using a Phoenix crystallization robot, from the Art Robbins Instruments, by sitting drop vapor diffusion. All crystals were grown at 20 °C. For the structural elucidation we used synchrotron radiation. This method offers exceptional X-ray brilliance and a highly collimated beam, along with the advantage of tuning the wavelength to suit specific experiments, such as single wavelength anomalous dispersion (SAD) or multi-wavelength anomalous dispersion (MAD). Data collection took place at synchrotron light sources, namely EMBL Hamburg, Petra III, and BESSY Berlin. Crystals were either transported on crystallisation plates or frozen in a dry-shipper at temperatures below 100K for the experiments.

6. References

- (1) Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res* **2000**, 28 (1), 235–242.
- (2) Kendrew, J. C., Bodo, G., Dintis, H. M., Parrish, R. G., and Wyckoff, H. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* **1958**, 181 (4610), 662–666.
- (3) Pauling, L.; Corey, R. B.; Branson, H. R. The Structure of Proteins; Two Hydrogen-Bonded Helical Configurations of the Polypeptide Chain. *Proc Natl Acad Sci U S A* **1951**, 37 (4), 205–211.
- (4) Sung, S. S. Peptide Folding Driven by Van Der Waals Interactions. *Protein Science* **2015**, 24 (9), 1383–1388.
- (5) Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. Stereochemistry of Polypeptide Chain Configurations. *J Mol Biol* **1963**, 7 (1), 95–99.
- (6) Anfinsen, C. B. Principles That Govern Protein Folding. *Science (1979)* **1973**, 181 (4096), 223–230.
- (7) Dill, K. A. Theory for the Folding and Stability of Globular Proteins. *Biochemistry* **1985**, 24 (6), 1501–1509.

- (8) Haber, E.; Anfinsen, C. B. Side-Chain Interactions Governing the Pairing of Half-Cystine Residues in Ribonuclease. *J Biol Chem* **1962**, *237* (6), 1839–1844.
- (9) Kauzmann, W. Some Factors in the Interpretation of Protein Denaturation. *Adv Protein Chem* **1959**, *14* (C), 1–63.
- (10) Baldwin, R. L. Dynamic Hydration Shell Restores Kauzmann's 1959 Explanation of How the Hydrophobic Factor Drives Protein Folding. *Proc Natl Acad Sci U S A* **2014**, *111* (36), 13052–13056.
- (11) Matthews, B. W. Hydrophobic Interactions in Proteins. *eLS* **2001**.
- (12) Walther, K. A.; Gräter, F.; Dougan, L.; Badilla, C. L.; Berne, B. J.; Fernandez, J. M. Signatures of Hydrophobic Collapse in Extended Proteins Captured with Force Spectroscopy. *Proc Natl Acad Sci U S A* **2007**, *104* (19), 7916–7921.
- (13) Kamtekar, S.; Schiffer, J. M.; Xiong, H.; Babik, J. M.; Hecht, M. H. Protein Design by Binary Patterning of Polar and Nonpolar Amino Acids. *Science* (1979) **1993**, *262*, 1680–1685.
- (14) Chan, H. S. Amino Acid Side-chain Hydrophobicity. *eLS* **2002**, 1–7.
- (15) Zwanzig, R. *Two-State Models of Protein Folding Kinetics*; 1997; Vol. 94.
- (16) Julia Hockenmaier, Aravind K. Joshi, and K. A. D. Routes Are Trees: The Parsing Perspective on Protein Folding. *Proteins: Structure, Function and Bioinformatics* **2007**, *66*, 1–15.
- (17) Hubbard, R. E.; Kamran Haider, M. Hydrogen Bonds in Proteins: Role and Strength. *eLS* **2010**, No. February, 1–7.
- (18) Li, J.; Hou, C.; Ma, X.; Guo, S.; Zhang, H.; Shi, L.; Liao, C.; Zheng, B.; Ye, L.; Yang, L.; He, X. Entropy-Enthalpy Compensations Fold Proteins in Precise Ways. *Int J Mol Sci* **2021**, *22* (17), 1–20.
- (19) Sosnick, T. R.; Barrick, D. The Folding of Single Domain Proteins—Have We Reached a Consensus? *Curr Opin Struct Biol* **2011**, *21* (1), 12–24.
- (20) Dill, K. A.; Chan, H. S. From Levinthal to Pathways to Funnels. *Nat Struct Biol* **1997**, *4* (1), 10–19.
- (21) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, Pathways, and the Energy Landscape of Protein Folding: A Synthesis. *Proteins: Structure, Function, and Bioinformatics* **1995**, *21* (3), 167–195.
- (22) Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R. The Protein Folding Problem. *Annu Rev Biophys* **2008**, *37*, 289–316.
- (23) Hartl, F. U.; Hayer-Hartl, M. Converging Concepts of Protein Folding in Vitro and in Vivo. *Nat Struct Mol Biol* **2009**, *16* (6), 574–581.

- (24) Zhuravleva, A.; Korzhnev, D. M. Protein Folding by NMR. *Prog Nucl Magn Reson Spectrosc* **2017**, *100*, 52–77.
- (25) Borgia, A.; Williams, P. M.; Clarke, J. Single-Molecule Studies of Protein Folding. *Annu Rev Biochem* **2008**, *77*, 101–125.
- (26) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science (1979)* **2010**, *330* (6002), 341–346.
- (27) Šali, A.; Shakhnovich, E.; Karplus, M. How Does a Protein Fold? *Nature* **1994**, *369* (6477), 248–251.
- (28) Kryshtafovych, A.; Schwede, T.; Topf, M.; Fidelis, K.; Moult, J. Critical Assessment of Methods of Protein Structure Prediction (CASP)—Round XIV. *Proteins: Structure, Function and Bioinformatics* **2021**, *89* (12), 1607–1617.
- (29) Kryshtafovych, A.; Antczak, M.; Szachniuk, M.; Zok, T.; Kretsch, R. C.; Rangan, R.; Pham, P.; Das, R.; Robin, X.; Studer, G.; Durairaj, J.; Eberhardt, J.; Sweeney, A.; Topf, M.; Schwede, T.; Fidelis, K.; Moult, J. New Prediction Categories in CASP15. *Proteins: Structure, Function and Bioinformatics* **2023**, No. May, 1–8.
- (30) Pedersen, J. T.; Moult, J. Ab Initio Protein Folding Simulations with Genetic Algorithms: Simulations on the Complete Sequence of Small Proteins. *Proteins: Structure, Function and Genetics* **1997**, *29* (SUPPL. 1), 179–184.
- (31) Gö, N. Protein Folding as a Stochastic Process. *J Stat Phys* **1983**, *30* (2), 413–423.
- (32) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of State Calculations by Fast Computing Machines. *J Chem Phys* **1953**, *21* (6), 1087–1092.
- (33) Simons, K. T.; Kooperberg, C.; Huang, E.; Baker, D. Assembly of Protein Tertiary Structures from Fragments with Similar Local Sequences Using Simulated Annealing and Bayesian Scoring Functions. *J Mol Biol* **1997**, *268* (1), 209–225.
- (34) Bonneau, R.; Tsai, J.; Ruczinski, I.; Chivian, D.; Rohl, C.; Strauss, C. E. M.; Baker, D. Rosetta in CASP4: Progress in Ab Initio Protein Structure Prediction. *Proteins: Structure, Function and Genetics* **2001**, *45* (SUPPL. 5), 119–126.
- (35) Rohl, C. A.; Strauss, C. E. M.; Misura, K. M. S.; Baker, D. Protein Structure Prediction Using Rosetta. *Methods Enzymol* **2004**, *383* (2003), 66–93.
- (36) Rhiju Das, Bin Qian, Srivatsan Raman, Robert Vernon, J. T.; Philip Bradley, Sagar Khare, Michael D. Tyka, Divya Bhat, Dylan Chivian, David E. Kim, William H. Sheffler, Lars Malmstrom, Andrew M. Wollacott, Chu Wang, Ingemar Andre, and D. B. Structure Prediction for CASP7 Targets Using

- Extensive All-Atom Refinement with Rosetta@home. *Proteins: Structure, Function and Bioinformatics* **2007**, *69* (8), 118–128.
- (37) Bradley, P.; Misura, K. M. S.; Baker, D. Toward High-Resolution de Novo Structure Prediction for Small Proteins. *Science (1979)* **2005**, *309* (5742), 1868–1871.
- (38) Alzubaidi, L.; Zhang, J.; Humaidi, A. J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M. A.; Al-Amidie, M.; Farhan, L. *Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions*; Springer International Publishing, **2021**; Vol. 8.
- (39) Sherstinsky, A. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. *Physica D* **2020**, *404*, 132306.
- (40) Kryshtafovych, A.; Schwede, T.; Topf, M.; Fidelis, K.; Moulton, J. Critical Assessment of Methods of Protein Structure Prediction (CASP)—Round XIII. *Proteins: Structure, Function and Bioinformatics* **2019**, *87* (12), 1011–1020.
- (41) Senior, A. W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Židek, A.; Nelson, A. W. R.; Bridgland, A.; Penedones, H.; Petersen, S.; Simonyan, K.; Crossan, S.; Kohli, P.; Jones, D. T.; Silver, D.; Kavukcuoglu, K.; Hassabis, D. Improved Protein Structure Prediction Using Potentials from Deep Learning. *Nature* **2020**, *577* (7792), 706–710.
- (42) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589.
- (43) V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. S. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters Viktor. *Proteins: Structure, Function, and Bioinformatics* **2006**, *65*, 712–725.
- (44) Mariani, V.; Biasini, M.; Barbato, A.; Schwede, T. IDDT: A Local Superposition-Free Score for Comparing Protein Structures and Models Using Distance Difference Tests. *Bioinformatics* **2013**, *29* (21), 2722–2728.
- (45) McCammon, J. A., Gelin, B. R., and Karplus, M. Dynamics of Folded Proteins. *Nature* **1977**, *267*, 585–590.
- (46) Ponder, J. W.; Case, D. A. Force Fields for Protein Simulations. *Adv Protein Chem* **2003**, *66*, 27–85.

- (47) Byrne, A.; Williams, D. V.; Barua, B.; Hagen, S. J.; Kier, B. L.; Andersen, N. H. Folding Dynamics and Pathways of the Trp-Cage Mini-proteins. *Biochemistry* **2014**, *53* (38), 6011–6021.
- (48) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-Folding Proteins Fold. *Science (1979)* **2011**, *334* (6055), 517–520.
- (49) Geng, H.; Chen, F.; Ye, J.; Jiang, F. Applications of Molecular Dynamics Simulation in Structure Prediction of Peptides and Proteins. *Comput Struct Biotechnol J* **2019**, *17*, 1162–1170.
- (50) Iftimie, R.; Minary, P.; Tuckerman, M. E. Ab Initio Molecular Dynamics: Concepts, Recent Developments, and Future Trends. *Proc Natl Acad Sci U S A* **2005**, *102* (19), 6654–6659.
- (51) Meli, M.; Morra, G.; Colombo, G. Simple Model of Protein Energetics to Identify Ab Initio Folding Transitions from All-Atom MD Simulations of Proteins. *J Chem Theory Comput* **2020**, *16* (9), 5960–5971.
- (52) Ding, F.; Tsao, D.; Nie, H.; Dokholyan, N. V. Ab Initio Folding of Proteins with All-Atom Discrete Molecular Dynamics. *Structure* **2008**, *16* (7), 1010–1018.
- (53) Chowdhury, S.; Lee, M. C.; Xiong, G.; Duan, Y. Ab Initio Folding Simulation of the Trp-Cage Mini-Protein Approaches NMR Resolution. *J Mol Biol* **2003**, *327* (3), 711–717.
- (54) Jones, S.; Thornton, J. M. Principles of Protein-Protein Interactions. *Proc Natl Acad Sci U S A* **1996**, *93* (1), 13–20.
- (55) Duncan G. G. McMillan, Sophie J. Marritt, Mackenzie A. Firer-Sherwood, Liang Shi, David J. Richardson, Stephen D. Evans, Sean J. Elliott, Julea N. Butt, and L. J. C. J. Protein-Protein Interaction Regulates the Direction of Catalysis and Electron Transfer in a Redox Enzyme Complex. *J Am Chem Soc* **2013**, *135*, 10550–10556.
- (56) Westermarck, J.; Ivaska, J.; Corthals, G. L. Identification of Protein Interactions Involved in Cellular Signaling. *Molecular and Cellular Proteomics* **2013**, *12* (7), 1752–1763.
- (57) Nero, T. L.; Morton, C. J.; Holien, J. K.; Wielens, J.; Parker, M. W. Oncogenic Protein Interfaces: Small Molecules, Big Challenges. *Nat Rev Cancer* **2014**, *14* (4), 248–262.
- (58) Gavin C. K. W. Koh, Pablo Porras, Bruno Aranda, Henning Hermjakob, and S. E. O. Analyzing Protein-Protein Interaction Networks. *J Proteome Res* **2012**, *11*, 2014–2031.
- (59) Martino, E.; Chiarugi, S.; Margheriti, F.; Garau, G. Mapping, Structure and Modulation of PPI. *Front Chem* **2021**, *9* (October), 1–8.

- (60) He, X.; Xu, C. Immune Checkpoint Signaling and Cancer Immunotherapy. *Cell Res* **2020**, *30* (8), 660–669.
- (61) Ishida, Y.; Agata, Y.; Shibahara, K.; Honjo, T. Induced Expression of PD-1, a Novel Member of the Immunoglobulin Gene Superfamily, upon Programmed Cell Death. *EMBO Journal*. 1992, pp 3887–3895.
- (62) Sanmamed, M. F.; Chen, L. A Paradigm Shift in Cancer Immunotherapy: From Enhancement to Normalization. *Cell* **2018**, *175* (2), 313–326.
- (63) Iwai, Y.; Hamanishi, J.; Chamoto, K.; Honjo, T. Cancer Immunotherapies Targeting the PD-1 Signaling Pathway. *J Biomed Sci* **2017**, *24* (1), 1–11.
- (64) Pardoll, D. M. The Blockade of Immune Checkpoints in Cancer Immunotherapy. *Nat Rev Cancer* **2012**, *12* (4), 252–264.
- (65) Chen, L.; Han, X. Anti-PD-1/PD-L1 Therapy of Human Cancer: Past, Present, and Future. *Journal of Clinical Investigation* **2015**, *125* (9), 3384–3391.
- (66) Smyth, M. J.; Teng, M. W. L. 2018 Nobel Prize in Physiology or Medicine. *Clin Transl Immunology* **2018**, *7* (10), 2–5.
- (67) Tang, Q.; Chen, Y.; Li, X.; Long, S.; Shi, Y.; Yu, Y.; Wu, W.; Han, L.; Wang, S. The Role of PD-1/PD-L1 and Application of Immune-Checkpoint Inhibitors in Human Cancers. *Front Immunol* **2022**, *13* (September), 1–19.
- (68) Liu, C.; Yang, M.; Zhang, D.; Chen, M.; Zhu, D. Clinical Cancer Immunotherapy: Current Progress and Prospects. *Front Immunol* **2022**, *13* (October), 1–22.
- (69) Zou, W.; Wolchok, J. D.; Chen, L. PD-L1 (B7-H1) and PD-1 Pathway Blockade for Cancer Therapy: Mechanisms, Response Biomarkers, and Combinations. *Sci Transl Med* **2016**, *8* (328).
- (70) Mellman, I.; Coukos, G.; Dranoff, G. Cancer Immunotherapy Comes of Age. *Nature* **2011**, *480* (7378), 480–489.
- (71) Cao, L.; Goresnik, I.; Coventry, B.; Case, J. B.; Miller, L.; Kozodoy, L.; Chen, R. E.; Carter, L.; Walls, A. C.; Park, Y. J.; Strauch, E. M.; Stewart, L.; Diamond, M. S.; Veessler, D.; Baker, D. De Novo Design of Picomolar SARS-CoV-2 Miniprotein Inhibitors. *Science (1979)* **2020**, *370* (6515), 426–431.
- (72) Zhou, X.; Zuo, C.; Li, W.; Shi, W.; Zhou, X.; Wang, H.; Chen, S.; Du, J.; Chen, G.; Zhai, W.; Zhao, W.; Wu, Y.; Qi, Y.; Liu, L.; Gao, Y. A Novel D-Peptide Identified by Mirror-Image Phage Display Blocks TIGIT/PVR for Cancer Immunotherapy. *Angewandte Chemie - International Edition* **2020**, *59* (35), 15114–15118.
- (73) Wang, T.; Wu, X.; Guo, C.; Zhang, K.; Xu, J.; Li, Z.; Jiang, S. Development of Inhibitors of the Programmed Cell Death-1/Programmed Cell Death-Ligand 1 Signaling Pathway. *J Med Chem* **2019**, *62* (4), 1715–1730.

- (74) Konstantinidou, M.; Zarganes-Tzitzikas, T.; Magiera-Mularz, K.; Holak, T. A.; Dömling, A. Immune Checkpoint PD-1/PD-L1: Is There Life Beyond Antibodies? *Angewandte Chemie - International Edition* **2018**, *57* (18), 4840–4848.
- (75) Cabrele, C.; Martinek, T. A.; Reiser, O.; Berlicki, Ł. Peptides Containing β -Amino Acid Patterns: Challenges and Successes in Medicinal Chemistry. *J Med Chem* **2014**, *57* (23), 9718–9739.
- (76) Zhang, X.; Schwartz, J.-C. D.; Guo, X.; Bhatia, S.; Cao, E.; Lorenz, M.; Cammer, M.; Chen, L.; Zhang, Z.-Y.; Edidin, M. A.; Nathenson, S. G.; Almo, S. C. Structural and Functional Analysis of the Costimulatory Receptor Programmed Death-1. *Immunity* **2004**, *20* (5), 651.
- (77) Zak, K. M.; Grudnik, P.; Magiera, K.; Dömling, A.; Dubin, G.; Holak, T. A. Structural Biology of the Immune Checkpoint Receptor PD-1 and Its Ligands PD-L1/PD-L2. *Structure* **2017**, *25* (8), 1163–1174.
- (78) Lin, D. Y. W.; Tanaka, Y.; Iwasaki, M.; Gittis, A. G.; Su, H. P.; Mikami, B.; Okazaki, T.; Honjo, T.; Minato, N.; Garboczi, D. N. The PD-1/PD-L1 Complex Resembles the Antigen-Binding Fv Domains of Antibodies and T Cell Receptors. *Proc Natl Acad Sci U S A* **2008**, *105* (8), 3011–3016.
- (79) Zak, K. M.; Kitel, R.; Przetocka, S.; Golik, P.; Guzik, K.; Musielak, B.; Dömling, A.; Dubin, G.; Holak, T. A. Structure of the Complex of Human Programmed Death 1, PD-1, and Its Ligand PD-L1. *Structure* **2015**, *23* (12), 2341–2348.
- (80) Pelay-Gimeno, M.; Glas, A.; Koch, O.; Grossmann, T. N. Structure-Based Design of Inhibitors of Protein-Protein Interactions: Mimicking Peptide Binding Epitopes. *Angewandte Chemie - International Edition* **2015**, *54* (31), 8896–8927.
- (81) Dockal, M.; Hartmann, R.; Fries, M.; Thomassen, M. C. L. G. D.; Heinzmann, A.; Ehrlich, H.; Rosing, J.; Osterkamp, F.; Polakowski, T.; Reineke, U.; Griessner, A.; Brandstetter, H.; Scheiflinger, F. Small Peptides Blocking Inhibition of Factor Xa and Tissue Factor-Factor VIIa by Tissue Factor Pathway Inhibitor (TFPI). *Journal of Biological Chemistry* **2014**, *289* (3), 1732–1741.
- (82) Lee, M. S.; Gippert, G. P.; Soman, K. V.; Case, D. A.; Wright, P. E. Three-Dimensional Solution Structure of a Single Zinc Finger DNA-Binding Domain. *Science (1979)* **1989**, *245* (4918), 635–637.
- (83) Chiu, T. K.; Kubelka, J.; Herbst-Irmer, R.; Eaton, W. A.; Hofrichter, J.; Davies, D. R. High-Resolution x-Ray Crystal Structures of the Villin Headpiece Subdomain, an Ultrafast Folding Protein. *Proc Natl Acad Sci U S A* **2005**, *102* (21), 7517–7522.

- (84) Sberro, H.; Fremin, B. J.; Zlitni, S.; Edfors, F.; Greenfield, N.; Snyder, M. P.; Pavlopoulos, G. A.; Kyripides, N. C.; Bhatt, A. S. Large-Scale Analyses of Human Microbiomes Reveal Thousands of Small, Novel Genes. *Cell* **2019**, *178* (5), 1245-1259.e14.
- (85) Craik, D. J.; Fairlie, D. P.; Liras, S.; Price, D. The Future of Peptide-Based Drugs. *Chem Biol Drug Des* **2013**, *81* (1), 136–147.
- (86) Baker, E. G.; Bartlett, G. J.; Porter Goff, K. L.; Woolfson, D. N. Miniprotein Design: Past, Present, and Prospects. *Acc Chem Res* **2017**, *50* (9), 2085–2092.
- (87) Ciesiołkiewicz, A.; Lizandra Perez, J.; Berlicki, Ł. Miniproteins in Medicinal Chemistry. *Bioorg Med Chem Lett* **2022**, *71* (June), 1–10.
- (88) Crook, Z. R.; Nairn, N. W.; Olson, J. M. Miniproteins as a Powerful Modality in Drug Development. *Trends Biochem Sci* **2020**, *45* (4), 332–346.
- (89) Ożga, K.; Berlicki, Ł. Miniprotein-Based Artificial Retroaldolase. *ACS Catal* **2022**, *12* (24), 15424–15430.
- (90) Crook, Z. R.; Sevilla, G. P.; Friend, D.; Brusniak, M. Y.; Bandaranayake, A. D.; Clarke, M.; Gewe, M.; Mhyre, A. J.; Baker, D.; Strong, R. K.; Bradley, P.; Olson, J. M. Mammalian Display Screening of Diverse Cystine-Dense Peptides for Difficult to Drug Targets. *Nat Commun* **2017**, *8* (1).
- (91) Araghi, R. R.; Bird, G. H.; Ryan, J. A.; Jenson, J. M.; Godes, M.; Pritz, J. R.; Grant, R. A.; Letai, A.; Walensky, L. D.; Keating, A. E. Iterative Optimization Yields Mcl-1–Targeting Stapled Peptides with Selective Cytotoxicity to Mcl-1–Dependent Cancer Cells. *Proc Natl Acad Sci U S A* **2018**, *115* (5), 886–895.
- (92) Aghaabdollahian, S.; Ahangari Cohan, R.; Norouzian, D.; Davami, F.; Asadi Karam, M. R.; Torkashvand, F.; Vaseghi, G.; Moazzami, R.; Latif Dizaji, S. Enhancing Bioactivity, Physicochemical, and Pharmacokinetic Properties of a Nano-Sized, Anti-VEGFR2 Adnectin, through PASylation Technology. *Sci Rep* **2019**, *9* (1), 1–14.
- (93) Eliassen, R.; Daly, N. L.; Wulff, B. S.; Andresen, T. L.; Conde-Frieboes, K. W.; Craik, D. J. Design, Synthesis, Structural and Functional Characterization of Novel Melanocortin Agonists Based on the Cyclotide Kalata B1. *Journal of Biological Chemistry* **2012**, *287* (48), 40493–40501.
- (94) Marwari, S.; Poulsen, A.; Shih, N.; Lakshminarayanan, R.; Kini, R. M.; Johannes, C. W.; Dymock, B. W.; Dawe, G. S. Intranasal Administration of a Stapled Relaxin-3 Mimetic Has Anxiolytic- and Antidepressant-like Activity in Rats. *Br J Pharmacol* **2019**, *176* (20), 3899–3923.
- (95) Blundell, T. L.; Pitts, J. E.; Tickle, I. J. X-Ray Analysis (1.4-Å Resolution) of Avian Pancreatic Polypeptide: Small Globular Protein Hormone. *Proc Natl Acad Sci U S A* **1981**, *78* (7 I), 4175–4179.

- (96) Lavergne, V.; J. Taft, R.; F. Alewood, P. Cysteine-Rich Mini-Proteins in Human Biology. *Curr Top Med Chem* **2012**, *12* (14), 1514–1533.
- (97) Arolas, J. L.; Aviles, F. X.; Chang, J. Y.; Ventura, S. Folding of Small Disulfide-Rich Proteins: Clarifying the Puzzle. *Trends Biochem Sci* **2006**, *31* (5), 292–301.
- (98) Tamaoki, H.; Miura, R.; Kusunoki, M.; Kyogoku, Y.; Kobayashi, Y.; Moroder, L. Folding Motifs Induced and Stabilized by Distinct Cystine Frameworks. *Protein Eng* **1998**, *11* (8), 649–659.
- (99) Craik, D. J.; Daly, N. L.; Waine, C. The Cystine Knot Motif in Toxins and Implications for Drug Design. *Toxicon* **2001**, *39* (1), 43–60.
- (100) Le-Nguyen, D.; Heitz, A.; Chiche, L.; Hajji, M. E. L.; Castro, B. Characterization and 2D NMR Study of the Stable [9-21, 15-27] 2 Disulfide Intermediate in the Folding of the 3 Disulfide Trypsin Inhibitor EETI II. *Protein Science* **1993**, *2*, 165–174.
- (101) Heitz, A.; Le-Nguyen, D.; Chiche, L. Min-21 and Min-23, the Smallest Peptides That Fold like a Cystine-Stabilized β -Sheet Motif: Design, Solution Structure, and Thermal Stability. *Biochemistry* **1999**, *38* (32), 10615–10625.
- (102) Chiche, L.; Heitz, A.; Gelly, J.-C.; Gracy, J.; Chau, P.; Ha, P.; Hernandez, J.-F.; Le-Nguyen, D. Squash Inhibitors: From Structural Motifs to Macrocyclic Knottins. *Curr Protein Pept Sci* **2005**, *5* (5), 341–349.
- (103) Postic, G.; Gracy, J.; Périn, C.; Chiche, L.; Gelly, J. C. KNOTTIN: The Database of Inhibitor Cystine Knot Scaffold after 10 Years, toward a Systematic Structure Modeling. *Nucleic Acids Res* **2018**, *46* (D1), D454–D458.
- (104) Wang, C. K. L.; Kaas, Q.; Chiche, L.; Craik, D. J. CyBase: A Database of Cyclic Protein Sequences and Structures, with Applications in Protein Discovery and Engineering. *Nucleic Acids Res* **2008**, *36* (SUPPL. 1), 206–210.
- (105) Mulvenna, J. P.; Wang, C.; Craik, D. J. CyBase: A Database of Cyclic Protein Sequence and Structure. *Nucleic Acids Res* **2006**, *34* (Database issue), 192–194.
- (106) Camarero, J. A. Cyclotides, a Versatile Ultrastable Micro-Protein Scaffold for Biotechnological Applications. *Bioorg Med Chem Lett* **2017**, *27* (23), 5089–5099.
- (107) Saether, O.; Craik, D. J.; Campbell, I. D.; Sletten, K.; Juul, J.; Norman, D. G. Elucidation of the Primary and Three-Dimensional Structure of the Uterotonic Polypeptide Kalata B1. *Biochemistry* **1995**, *34* (13), 4147–4158.
- (108) Claeson, P.; Göransson, U.; Johansson, S.; Luijendijk, T.; Bohlin, L. Fractionation Protocol for the Isolation of Polypeptides from Plant Biomass. *J Nat Prod* **1998**, *61* (1), 77–81.

- (109) Krätzner, R.; Debreczeni, J. É.; Pape, T.; Schneider, T. R.; Wentzel, A.; Kolmar, H.; Sheldrick, G. M.; Uson, I. Structure of Ecballium Elaterium Trypsin Inhibitor II (EETI-II): A Rigid Molecular Scaffold. *Acta Crystallogr D Biol Crystallogr* **2005**, *61* (9), 1255–1262.
- (110) Kintzing, J. R.; Cochran, J. R. Engineered Knottin Peptides as Diagnostics, Therapeutics, and Drug Delivery Vehicles. *Curr Opin Chem Biol* **2016**, *34*, 143–150.
- (111) Poth, A. G.; Chan, L. Y.; Craik, D. J. Cyclotides as Grafting Frameworks for Protein Engineering and Drug Design Applications. *Biopolymers* **2013**, *100* (5), 480–491.
- (112) Camarero, J. A.; Campbell, M. J. The Potential of the Cyclotide Scaffold for Drug Development. *Biomedicines* **2019**, *7* (2), 1–19.
- (113) Heitz, A.; Chiche, L.; Le-Nguyen, D.; Castro, B. Folding of the Squash Trypsin Inhibitor EETI II: Evidence of Native and Non-Native Local Structural Preferences in a Linear Analogue. *Eur J Biochem* **1995**, *233* (3), 837–846.
- (114) Colgrave, M. L.; Craik, D. J. Thermal, Chemical, and Enzymatic Stability of the Cyclotide Kalata B1: The Importance of the Cyclic Cystine Knot. *Biochemistry* **2004**, *43* (20), 5965–5975.
- (115) Moore, S. J.; Cochran, J. R. *Engineering Knottins as Novel Binding Agents*, 1st ed.; Elsevier Inc., 2012; Vol. 503.
- (116) Lahti, J. L.; Silverman, A. P.; Cochran, J. R. Interrogating and Predicting Tolerated Sequence Diversity in Protein Folds: Application to E. Elaterium Trypsin Inhibitor-II Cystine-Knot Mini-protein. *PLoS Comput Biol* **2009**, *5* (9).
- (117) Santiveri, C. M.; Santoro, J.; Rico, M.; Jiménez, M. A. Factors Involved in the Stability of Isolated β -Sheets: Turn Sequence, β -Sheet Twisting, and Hydrophobic Surface Burial. *Protein Science* **2004**, *13* (4), 1134–1147.
- (118) Doig, A. J. A Three Stranded β -Sheet Peptide in Aqueous Solution Containing N-Methyl Amino Acids to Prevent Aggregation. *Chemical Communications* **1997**, No. 22, 2153–2154.
- (119) Cory D. DuPai, Bryan W. Davies, C. O. W. A Systematic Analysis of the Beta Hairpin Motif in the Protein Data Bank 2. *bioRxiv* **2020**, *21* (1), 1–9.
- (120) Sudol, M.; Hunter, T. New Wrinkles for an Old Domain. *Cell* **2000**, *103* (7), 1001–1004.
- (121) Macias, M. J.; Hyvönen, M.; Baraldi, E.; Schultz, J.; Sudol, M.; Saraste, M.; Oschkinat, H. Structure of the WW Domain of a Kinase-Associated Protein Complexed with a Proline-Rich Peptide. *Nature*. 1996, pp 646–649.
- (122) Macias, M. J.; Gervais, V. Structural Analysis of WW Domains and Design of a WW Prototype. *America (NY)* **2000**, 375–379.

- (123) Stewart, A. L.; Park, J. H.; Waters, M. L. Redesign of a WW Domain Peptide for Selective Recognition of Single-Stranded DNA. *Biochemistry* **2011**, *50* (13), 2575–2584.
- (124) Zhu, W.; Tanday, N.; Flatt, P. R.; Irwin, N. Pancreatic Polypeptide Revisited: Potential Therapeutic Effects in Obesity-Diabetes. *Peptides (N.Y.)* **2023**, *160* (November 2022), 170923.
- (125) Baker, E. G.; Williams, C.; Hudson, K. L.; Bartlett, G. J.; Heal, J. W.; Goff, K. L. P.; Sessions, R. B.; Crump, M. P.; Woolfson, D. N. Engineering Protein Stability with Atomic Precision in a Monomeric Miniprotein. *Nat Chem Biol* **2017**, *13* (7), 764–770.
- (126) Porter Goff, K. L.; Nicol, D.; Williams, C.; Crump, M. P.; Zieleniewski, F.; Samphire, J. L.; Baker, E. G.; Woolfson, D. N. Stabilizing and Understanding a Miniprotein by Rational Redesign. *Biochemistry* **2019**, *58* (28), 3060–3064.
- (127) Pinheiro, S.; Soteras, I.; Gelpí, J. L.; Dehez, F.; Chipot, C.; Luque, F. J.; Curutchet, C. Structural and Energetic Study of Cation- π -Cation Interactions in Proteins. *Physical Chemistry Chemical Physics* **2017**, *19* (15), 9849–9861.
- (128) Craven, T. W.; Cho, M. K.; Traaseth, N. J.; Bonneau, R.; Kirshenbaum, K. A Miniature Protein Stabilized by a Cation- π Interaction Network. *J Am Chem Soc* **2016**, *138* (5), 1543–1550.
- (129) Craven, T. W.; Bonneau, R.; Kirshenbaum, K. PPII Helical Peptidomimetics Templated by Cation- π Interactions. *ChemBioChem* **2016**, 1824–1828.
- (130) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. Designing a 20-Residue Protein. *Nat Struct Biol* **2002**, *9* (6), 425–430.
- (131) Barua, B.; Lin, J. C.; Williams, V. D.; Kummler, P.; Neidigh, J. W.; Andersen, N. H. The Trp-Cage: Optimizing the Stability of a Globular Miniprotein. *Protein Engineering, Design and Selection* **2008**, *21* (3), 171–185.
- (132) Snow, D. C., Zagrovic, B., and Pande, V. S. The Trp Cage: Folding Kinetics and Unfolded State Topology via Molecular Dynamics Simulations. *J Am Chem Soc* **2002**, *124*, 14548–14549.
- (133) Miller, J.; McLachlan, A. D.; Klug, A. Repetitive Zinc-Binding Domains in the Protein Transcription Factor IIIA from *Xenopus* Oocytes. *EMBO J* **1985**, *4* (6), 1609–1614.
- (134) Cassandri, M.; Smirnov, A.; Novelli, F.; Pitolli, C.; Agostini, M.; Malewicz, M.; Melino, G.; Raschellà, G. Zinc-Finger Proteins in Health and Disease. *Cell Death Discov* **2017**, *3* (1).
- (135) Struthers, M. D.; Cheng, R. P.; Imperiali, B. Design of a Monomeric 23-Residue Polypeptide with Defined Tertiary Structure. *Science (1979)* **1996**, *271* (5247), 342–345.

- (136) Žoldák, G.; Stigler, J.; Pelz, B.; Li, H.; Rief, M. Ultrafast Folding Kinetics and Cooperativity of Villin Headpiece in Single-Molecule Force Spectroscopy. *Proc Natl Acad Sci U S A* **2013**, *110* (45), 18156–18161.
- (137) Bi, Y.; Cho, J. H.; Kim, E. Y.; Shan, B.; Schindelin, H.; Raleigh, D. P. Rational Design, Structural and Thermodynamic Characterization of a Hyperstable Variant of the Villin Headpiece Helical Subdomain. *Biochemistry* **2007**, *46* (25), 7497–7505.
- (138) Mandell, D. J.; Kortemme, T. Backbone Flexibility in Computational Protein Design. *Curr Opin Biotechnol* **2009**, *20* (4), 420–428.
- (139) Fowler, D. M.; Araya, C. L.; Fleishman, S. J.; Kellogg, E. H.; Stephany, J. J.; Baker, D.; Fields, S. High-Resolution Mapping of Protein Sequence-Function Relationships. *Nat Methods* **2010**, *7* (9), 741–746.
- (140) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science (1979)* **2003**, *302* (5649), 1364–1368.
- (141) Huang, P. S.; Boyken, S. E.; Baker, D. The Coming of Age of de Novo Protein Design. *Nature* **2016**, *537* (7620), 320–327.
- (142) Dahiyat, B. I.; Mayo, S. L. De Novo Protein Design: Fully Automated Sequence Selection. *Science (1979)* **1997**, *278* (5335), 82–87.
- (143) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. DREIDING: A Generic Force Field for Molecular Simulations. *Journal of Physical Chemistry* **1990**, *94* (26), 8897–8909.
- (144) Goldstein, R. F. Efficient Rotamer Elimination Applied to Protein Side-Chains and Related Spin Glasses. *Biophys J* **1994**, *66* (5), 1335–1340.
- (145) Andrea G. Cochran, Nicholas J. Skelton, and M. A. S. Tryptophan Zippers: Stable, Monomeric Beta-Hairpins. **2001**, *98* (10), 5578–5583.
- (146) Cochran, A. G.; Tong, R. T.; Starovasnik, M. A.; Park, E. J.; McDowell, R. S.; Theaker, J. E.; Skelton, N. J. A Minimal Peptide Scaffold for β -Turn Display: Optimizing a Strand Position in Disulfide-Cyclized β -Hairpins. *J Am Chem Soc* **2001**, *123* (4), 625–632.
- (147) Bureau, H. R.; Quirk, S.; Hernandez, R. The Relative Stability of Trpzip1 and Its Mutants Determined by Computation and Experiment. *RSC Adv* **2020**, *10* (11), 6520–6535.
- (148) Ramírez-Alvarado, M.; Blanco, F. J.; Serrano, L. De Novo Design and Structural Analysis of a Model β -Hairpin Peptide System. *Nat Struct Biol* **1996**, *3* (7), 604–612.

- (149) Alba, E. De; Santoro, J.; Rico, M.; Jiménez, M. A. De Novo Design of a Monomeric Three-Stranded Antiparallel β -Sheet. *Protein Science* **2008**, *8* (4), 854–865.
- (150) Röhl, C. A.; Martin Scholtz, J.; Bald, R. L.; York, E. J.; Stewart, J. M. Kinetics of Amide Proton Exchange in Helical Peptides of Varying Chain Lengths. Interpretation by the Lifson-Roig Equation. *Biochemistry* **1992**, *31* (5), 1263–1269.
- (151) Stanger, H. E.; Syud, F. A.; Espinosa, J. F.; Giriati, I.; Muir, T.; Gellman, S. H. Length-Dependent Stability and Strand Length Limits in Antiparallel β -Sheet Secondary Structure. *Proc Natl Acad Sci U S A* **2001**, *98* (21), 12015–12020.
- (152) Chou, P. Y.; Fasman, G. D. Empirical Predictions of Protein Conformation. *Annu Rev Biochem* **1978**, *47*, 251–276.
- (153) Xiong, H.; Buckwalter, B. L.; Shieh, H. M.; Hecht, M. H. Periodicity of Polar and Nonpolar Amino Acids Is the Major Determinant of Secondary Structure in Self-Assembling Oligomeric Peptides. *Proc Natl Acad Sci U S A* **1995**, *92* (14), 6349–6353.
- (154) O’Neil, K. T.; DeGrado, W. F. A Thermodynamic Scale for the Helix-Forming Tendencies of the Commonly Occurring Amino Acids. *Science (1979)* **1990**, *250* (4981), 646–651.
- (155) Marqusee, S.; Baldwin, R. L. Helix Stabilization by Glu-...Lys+ Salt Bridges in Short Peptides of de Novo Design. *Proc Natl Acad Sci U S A* **1987**, *84* (24), 8898–8902.
- (156) Toniolo, C. Structural Role of Valine and Isoleucine Residues in Proteins. A Proposal. *Macromolecules* **1978**, *11* (2), 437–438.
- (157) Aurora, R.; Rose, G. D. Helix Capping. *Protein Science* **1998**, *7* (1), 21–38.
- (158) Koga, N.; Tatsumi-Koga, R.; Liu, G.; Xiao, R.; Acton, T. B.; Montelione, G. T.; Baker, D. Principles for Designing Ideal Protein Structures. *Nature* **2012**, *491* (7423), 222–227.
- (159) Richter, F.; Leaver-Fay, A.; Khare, S. D.; Bjelic, S.; Baker, D. De Novo Enzyme Design Using Rosetta3. *PLoS One* **2011**, *6* (5), 1–12.
- (160) Tyka, M. D.; Keedy, D. A.; André, I.; Dimaio, F.; Song, Y.; Richardson, D. C.; Richardson, J. S.; Baker, D. Alternate States of Proteins Revealed by Detailed Energy Landscape Mapping. *J Mol Biol* **2011**, *405* (2), 607–618.
- (161) Sheffler, W.; Baker, D. RosettaHoles: Rapid Assessment of Protein Core Packing for Structure Prediction, Refinement, Design, and Validation. *Protein Science* **2009**, *18* (1), 229–239.
- (162) Bhardwaj, G.; Mulligan, V. K.; Bahl, C. D.; Gilmore, J. M.; Harvey, P. J.; Cheneval, O.; Buchko, G. W.; Pulavarti, S. V. S. R. K.; Kaas, Q.; Eletsky, A.;

- Huang, P. S.; Johnsen, W. A.; Greisen, P. J.; Rocklin, G. J.; Song, Y.; Linsky, T. W.; Watkins, A.; Rettie, S. A.; Xu, X.; Carter, L. P.; Bonneau, R.; Olson, J. M.; Coutsiyas, E.; Correnti, C. E.; Szyperski, T.; Craik, D. J.; Baker, D. Accurate de Novo Design of Hyperstable Constrained Peptides. *Nature* **2016**, *538* (7625), 329–33
- (163) Huang, P. S.; Oberdorfer, G.; Xu, C.; Pei, X. Y.; Nannenga, B. L.; Rogers, J. M.; DiMaio, F.; Gonen, T.; Luisi, B.; Baker, D. High Thermodynamic Stability of Parametrically Designed Helical Bundles. *Science (1979)* **2014**, *346* (6208), 481–485.
- (164) Rocklin, G. J.; Chidyausiku, T. M.; Goresnik, I.; Ford, A.; Houliston, S.; Lemak, A.; Carter, L.; Ravichandran, R.; Mulligan, V. K.; Chevalier, A.; Arrowsmith, C. H.; Baker, D. Global Analysis of Protein Folding Using Massively Parallel Design, Synthesis, and Testing. *Science (1979)* **2017**, *357* (6347), 168–175.
- (165) Coutsiyas, E. A.; Seok, C.; Jacobson, M. P.; Dill, K. A. A Kinematic View of Loop Closure. *J Comput Chem* **2004**, *25* (4), 510–528.
- (166) Mandell, D. J.; Coutsiyas, E. A.; Kortemme, T. Sub-Angstrom Accuracy in Protein Loop Reconstruction by Robotics-Inspired Conformational Sampling. *Nat Methods* **2009**, *6* (8), 551–552.
- (167) Stawikowski, M.; Fields, G. B. Introduction to Peptide Synthesis. *Curr Protoc Protein Sci* **2012**, *1* (SUPPL.69), 1–13.
- (168) Gellman, S. H. Foldamers : A Manifesto. **1998**, *31* (4), 173–180.
- (169) Martinek, T. A.; Fülöp, F. Peptidic Foldamers: Ramping up Diversity. *Chem Soc Rev* **2012**, *41* (2), 687–702.
- (170) Fülöp, F.; Martinek, T. A.; Tóth, G. K. Application of Alicyclic β -Amino Acids in Peptide Chemistry. *Chem Soc Rev* **2006**, *35* (4), 323–334.
- (171) Appella, D. H.; Christianson, L. A.; Karle, I. L.; Powell, D. R.; Gellman, S. H. β -Peptide Foldamers: Robust Helix Formation in a New Family of β -Amino Acid Oligomers. *J Am Chem Soc* **1996**, *118* (51), 13071–13072.
- (172) Seebach, D.; Matthews, J. L. β -Peptides: A Surprise at Every Turn. *Chemical Communications* **1997**, *1* (21), 2015–2022.
- (173) Srinivasulu, G.; Kiran Kumar, S.; Sharma, G. V. M.; Kunwar, A. C. 11/9-Mixed Helices in the α/β -Peptides Derived from Alternating α - and β -Amino Acids with Proteinogenic Side Chains. *Journal of Organic Chemistry* **2006**, *71* (22), 8395–8400.
- (174) Szefczyk, M.; Węglarz-Tomczak, E.; Fortuna, P.; Krzysztoń, A.; Rudzińska-Szostak, E.; Berlicki, Ł. Controlling the Helix Handedness of $\text{A}\alpha\beta$ -Peptide Foldamers through Sequence Shifting. *Angewandte Chemie - International Edition* **2017**, 2087–2091.

- (175) Choi, S. H.; Guzei, I. A.; Spencer, L. C.; Gellman, S. H. Crystallographic Characterization of Helical Secondary Structures in 2:1 and 1:2 α/β -Peptides. *J Am Chem Soc* **2009**, *131* (8), 2917–2924.
- (176) Schmitt, M. A.; Choi, S. H.; Guzei, I. A.; Gellman, S. H. New Helical Foldamers: Heterogeneous Backbones with 1:2 and 2:1 α/β -Amino Acid Residue Patterns. *J Am Chem Soc* **2006**, *128* (14), 4538–4539.
- (177) Berlicki, Ł.; Pilsl, L.; Wéber, E.; Mándity, I. M.; Cabrele, C.; Martinek, T. A.; Fülöp, F.; Reiser, O. Unique α,β - and $\alpha,\alpha,\beta,\beta$ -Peptide Foldamers Based on Cis- β -Aminocyclopentanecarboxylic Acid. *Angewandte Chemie - International Edition* **2012**, *51* (9), 2208–2212.
- (178) Horne, W. S.; Gellman, S. H. Foldamers with Heterogeneous Backbones. *Acc Chem Res* **2008**, *41* (10), 1399–1408.
- (179) Price, J. L.; Hadley, E. B.; Steinkruger, J. D.; Gellman, S. H. Detection and Analysis of Chimeric Tertiary Structures by Backbone Thioester Exchange: Packing of an α Helix against an α/β -Peptide Helix. *Angewandte Chemie - International Edition* **2010**, *49* (2), 368–371.
- (180) Drewniak, M.; Węglarz-Tomczak, E.; Oźga, K.; Rudzińska-Szostak, E.; Macegoniuk, K.; Tomczak, J. M.; Bejger, M.; Rypniewski, W.; Berlicki, Ł. Helix-Loop-Helix Peptide Foldamers and Their Use in the Construction of Hydrolase Mimetics. *Bioorg Chem* **2018**, *81* (July), 356–361.
- (181) Oźga, K.; Drewniak-Świtalska, M.; Rudzińska-Szostak, E.; Berlicki, Ł. Towards Foldameric Miniproteins: A Helix-Turn-Helix Motif. *Chempluschem* **2021**, *86* (4), 646–649.
- (182) Collie, G. W.; Pulka-Ziach, K.; Lombardo, C. M.; Fremaux, J.; Rosu, F.; Decossas, M.; Mauran, L.; Lambert, O.; Gabelica, V.; Mackereth, C. D.; Guichard, G. Shaping Quaternary Assemblies of Water-Soluble Non-Peptide Helical Foldamers by Sequence Manipulation. *Nat Chem* **2015**, *7* (11), 871–878.
- (183) Horne, W. S.; Price, J. L.; Keck, J. L.; Gellman, S. H. Helix Bundle Quaternary Structure from α/β -Peptide Foldamers. *J Am Chem Soc* **2007**, *129* (14), 4178–4180.
- (184) Reinert, Z. E.; Horne, W. S. Protein Backbone Engineering as a Strategy to Advance Foldamers toward the Frontier of Protein-like Tertiary Structure. *Org Biomol Chem* **2014**, *12* (44), 8796–8802.
- (185) Dale F. Kreidler, David E. Mortenson, Katrina T. Forest, S. H. G. Effects of Single A-to- β Residue Replacements on Structure and Stability in a Small Protein: Insights from Quasiracemic Crystallization. *J Am Chem Soc* **2016**, *138* (20), 6498–6505.

- (186) Bejger, M.; Fortuna, P.; Drewniak-Świtalska, M.; Plewka, J.; Rypniewski, W.; Berlicki, Ł. A Computationally Designed β -Amino Acid-Containing Miniprotein. *Chemical Communications* **2021**, 57 (49), 6015–6018.
- (187) Lengyel, G. A.; Frank, R. C.; Horne, W. S. Hairpin Folding Behavior of Mixed α/β -Peptides in Aqueous Solution. *J Am Chem Soc* **2011**, 133 (12), 4246–4249.
- (188) David E Mortenson, Dale F Kreitler, Nicole C Thomas, Ilia A Guzei, Samuel H Gellman, K. T. F. Evaluation of β -Amino Acid Replacements in Protein Loops: Effects on Conformational Stability and Structure. *ChemBioChem* **2018**, 19, 604–612.
- (189) Drewniak-Świtalska, M.; Barycza, B.; Rudzińska-Szostak, E.; Morawiak, P.; Berlicki, Ł. Constrained Beta-Amino Acid-Containing Miniproteins. *Org Biomol Chem* **2021**, 19 (19), 4272–4278.
- (190) George, K. L.; Horne, W. S. Foldamer Tertiary Structure through Sequence-Guided Protein Backbone Alteration. *Acc Chem Res* **2018**, 51 (5), 1220–1228.
- (191) Reinert, Z. E.; Horne, W. S. Folding Thermodynamics of Protein-like Oligomers with Heterogeneous Backbones. *Chem Sci* **2014**, 5 (8), 3325–3330.
- (192) Reinert, Z. E.; Lengyel, G. A.; Horne, W. S. Protein-like Tertiary Folding Behavior from Heterogeneous Backbones. *J Am Chem Soc* **2013**, 135 (34), 12528–12531.
- (193) George, K. L.; Seth Horne, W. Heterogeneous-Backbone Foldamer Mimics of Zinc Finger Tertiary Structure. *J Am Chem Soc* **2017**, 139 (23), 7931–7938.
- (194) Lombardo, C. M.; Kumar, V.; Douat, C.; Rosu, F.; Mergny, J. L.; Salgado, G. F.; Guichard, G. Design and Structure Determination of a Composite Zinc Finger Containing a Nonpeptide Foldamer Helical Domain. *J Am Chem Soc* **2019**, 141 (6), 2516–2525.
- (195) Hellinger, R.; Muratspahić, E.; Devi, S.; Koebach, J.; Vasileva, M.; Harvey, P. J.; Craik, D. J.; Gründemann, C.; Gruber, C. W. Importance of the Cyclic Cystine Knot Structural Motif for Immunosuppressive Effects of Cyclotides. *ACS Chem Biol* **2021**, 16 (11), 2373–2386.
- (196) Gründemann, C.; Stenberg, K. G.; Gruber, C. W. T20K: An Immunomodulatory Cyclotide on Its Way to the Clinic. *Int J Pept Res Ther* **2019**, 25 (1), 9–13.
- (197) Romero, Haylie K., Christensen, S. B., Cesare Mannelli, L. D., Gajewiak, J., Ramachandra, R., Elmslie, K. S. Vetter, D. E., Ghelardini, C., Iadonato, S. P., Mercado, J. L., Olivera, B. M. and McIntosh, J. M. Inhibition of A9 α 10 Nicotinic Acetylcholine Receptors prevents Chemotherapy-Induced Neuropathic Pain. *Proc Natl Acad Sci U S A* **2017**, 114, 1825–1832.
- (198) Wang, C. K.; Gruber, C. W.; Cemazar, M.; Siatskas, C.; Tagore, P.; Payne, N.; Sun, G.; Wang, S.; Bernard, C. C.; Craik, D. J. Molecular Grafting onto a Stable

Framework Yields Novel Cyclic Peptides for the Treatment of Multiple Sclerosis. *ACS Chem Biol* **2014**, *9* (1), 156–163.

- (199) Sia, S. K.; Kim, P. S. Protein Grafting of an HIV-1-Inhibiting Epitope. *Proc Natl Acad Sci U S A* **2003**, *100* (17), 9756–9761.
- (200) Chevalier, A.; Silva, D. A.; Rocklin, G. J.; Hicks, D. R.; Vergara, R.; Murapa, P.; Bernard, S. M.; Zhang, L.; Lam, K. H.; Yao, G.; Bahl, C. D.; Miyashita, S. I.; Goresnik, I.; Fuller, J. T.; Koday, M. T.; Jenkins, C. M.; Colvin, T.; Carter, L.; Bohn, A.; Bryan, C. M.; Fernández-Velasco, D. A.; Stewart, L.; Dong, M.; Huang, X.; Jin, R.; Wilson, I. A.; Fuller, D. H.; Baker, D. Massively Parallel de Novo Protein Design for Targeted Therapeutics. *Nature* **2017**, *550* (7674), 74–79.
- (201) Levy, J. H.; O'Donnell, P. S. The Therapeutic Potential of a Kallikrein Inhibitor for Treating Hereditary Angioedema. *Expert Opin Investig Drugs* **2006**, *15* (9), 1077–1090.
- (202) Bryan, C. M.; Rocklin, G. J.; Bick, M. J.; Ford, A.; Majri-Morrison, S.; Kroll, A. V.; Miller, C. J.; Carter, L.; Goresnik, I.; Kang, A.; DiMaio, F.; Tarbell, K. V.; Baker, D. Computational Design of a Synthetic PD-1 Agonist. *Proc Natl Acad Sci U S A* **2021**, *118* (29), 1–9.
- (203) De Veer, S. J.; Kan, M. W.; Craik, D. J. Cyclotides: From Structure to Function. *Chem Rev* **2019**, *119* (24), 12375–12421.
- (204) Pennington, M. W.; Czerwinski, A.; Norton, R. S. Peptide Therapeutics from Venom: Current Status and Potential. *Bioorg Med Chem* **2018**, *26* (10), 2738–2758.
- (205) Gran, L. An Oxytocic Principle Found in *Oldenlandia Affinis* DC. An Indigenous, Congolese Drug “Kalata-Kalata” Used to Accelerate the Delivery. *rsk. Farm. Selskap.* **1970**, *32*, 173–180.
- (206) Gruber, C. W.; Elliott, A. G.; Ireland, D. C.; Delprete, P. G.; Dessen, S.; Göransson, U.; Trabi, M.; Wang, C. K.; Kinghorn, A. B.; Robbrecht, E.; Craika, D. J. Distribution and Evolution of Circular Mini-proteins in Flowering Plants. *Plant Cell* **2008**, *20* (9), 2471–2483.
- (207) Tam, J. P.; Lu, Y. A.; Yang, J. L.; Chiu, K. W. An Unusual Structural Motif of Antimicrobial Peptides Containing End-to-End Macrocyclic and Cystine-Knot Disulfides. *Proc Natl Acad Sci U S A* **1999**, *96* (16), 8913–8918.
- (208) Svängård, E.; Burman, R.; Gunasekera, S.; Lövborg, H.; Gullbo, J.; Göransson, U. Mechanism of Action of Cytotoxic Cyclotides: Cycloviolacin O2 Disrupts Lipid Membranes. *J Nat Prod* **2007**, *70* (4), 643–647.
- (209) Gustafson, K. R.; Sowder, R. C.; Henderson, L. E.; Parsons, I. C.; Kashman, Y.; Cardellina, J. H.; McMahon, J. B.; Buckheit, R. W.; Pannell, L. K.; Boyd, M. R. Circulin-a and Circulin-B - Novel Hiv-Inhibitory Macrocyclic Peptides

- from the Tropical Tree *Chassalia-Parvifolia*. *J Am Chem Soc* **1994**, *116* (20), 9337–9338.
- (210) Henriques, S. T.; Huang, Y. H.; Chaousis, S.; Sani, M. A.; Poth, A. G.; Separovic, F.; Craik, D. J. The Prototypic Cyclotide Kalata B1 Has a Unique Mechanism of Entering Cells. *Chem Biol* **2015**, *22* (8), 1087–1097.
- (211) Gründemann, C.; Koehbach, J.; Huber, R.; Gruber, C. W. Do Plant Cyclotides Have Potential as Immunosuppressant Peptides? *J Nat Prod* **2012**, *75* (2), 167–174.
- (212) Henriques, S. T.; Huang, Y. H.; Rosengren, K. J.; Franquelim, H. G.; Carvalho, F. A.; Johnson, A.; Sonza, S.; Tachedjian, G.; Castanho, M. A. R. B.; Daly, N. L.; Craik, D. J. Decoding the Membrane Activity of the Cyclotide Kalata B1: The Importance of Phosphatidylethanolamine Phospholipids and Lipid Organization on Hemolytic and Anti-HIV Activities. *Journal of Biological Chemistry* **2011**, *286* (27), 24231–24241.
- (213) Barry, D. G.; Daly, N. L.; Clark, R. J.; Sando, L.; Craik, D. J. Linearization of a Naturally Occurring Circular Protein Maintains Structure but Eliminates Hemolytic Activity. *Biochemistry* **2003**, *42* (22), 6688–6695.
- (214) Gründemann, C.; Thell, K.; Lengen, K.; Garcia-Käufer, M.; Huang, Y. H.; Huber, R.; Craik, D. J.; Schabbauer, G.; Gruber, C. W. Cyclotides Suppress Human T-Lymphocyte Proliferation by an Interleukin 2-Dependent Mechanism. *PLoS One* **2013**, *8* (6), 1–12.
- (215) Constantinescu, C. S.; Farooqi, N.; O'Brien, K.; Gran, B. Experimental Autoimmune Encephalomyelitis (EAE) as a Model for Multiple Sclerosis (MS). *Br J Pharmacol* **2011**, *164* (4), 1079–1106.
- (216) Thell, K.; Hellinger, R.; Sahin, E.; Michenthaler, P.; Gold-Binder, M.; Haider, T.; Kuttke, M.; Liutkevičiute, Z.; Göransson, U.; Gründemann, C.; Schabbauer, G.; Gruber, C. W. Oral Activity of a Nature-Derived Cyclic Peptide for the Treatment of Multiple Sclerosis. *Proc Natl Acad Sci U S A* **2016**, *113* (15), 3960–3965.
- (217) Troeira Henriques, S.; Craik, D. J. Cyclotide Structure and Function: The Role of Membrane Binding and Permeation. *Biochemistry* **2017**, *56* (5), 669–682.
- (218) Liao, X.; Rabideau, A. E.; Pentelute, B. L. Delivery of Antibody Mimics into Mammalian Cells via Anthrax Toxin Protective Antigen. *ChemBioChem* **2014**, *15* (16), 2458–2466.
- (219) Rabideau, A. E.; Pentelute, B. L. Delivery of Non-Native Cargo into Mammalian Cells Using Anthrax Lethal Toxin. *ACS Chem Biol* **2016**, *11* (6), 1490–1501.
- (220) Iadonato, S. P, Muñoz, E. J. Modifications and Uses of Conotoxin Peptides, 2016.

- (221) Mehta, L.; Dhankhar, R.; Gulati, P.; Kapoor, R. K.; Mohanty, A.; Kumar, S. Natural and Grafted Cyclotides in Cancer Therapy: An Insight. *Journal of Peptide Science* **2020**, *26* (4–5), 1–16.
- (222) Wang, C. K.; Craik, D. J. Linking Molecular Evolution to Molecular Grafting. *Journal of Biological Chemistry* **2021**, *296*, 100425.
- (223) Muratspahić, E.; Koehbach, J.; Gruber, C. W.; Craik, D. J. Harnessing Cyclotides to Design and Develop Novel Peptide GPCR Ligands. *RSC Chem Biol* **2020**, *1* (4), 177–191.
- (224) Henriques, S. T.; Craik, D. J. Importance of the Cell Membrane on the Mechanism of Action of Cyclotides. *ACS Chem Biol* **2012**, *7* (4), 626–636.
- (225) Ji, Y.; Majumder, S.; Millard, M.; Borra, R.; Bi, T.; Elnagar, A. Y.; Neamati, N.; Shekhtman, A.; Camarero, J. A. In Vivo Activation of the P53 Tumor Suppressor Pathway by an Engineered Cyclotide. *J Am Chem Soc* **2013**, *135* (31), 11623–11633.
- (226) Gunasekera, S.; Foley, F. M.; Clark, R. J.; Sando, L.; Fabri, L. J.; Craik, D. J.; Daly, N. L. Engineering Stabilized Vascular Endothelial Growth Factor-A Antagonists: Synthesis, Structural Characterization, and Bioactivity of Grafted Analogues of Cyclotides. *J Med Chem* **2008**, *51* (24), 7697–7704.
- (227) Thongyoo, P.; Roqué-Rosell, N.; Leatherbarrow, R. J.; Tate, E. W. Chemical and Biomimetic Total Syntheses of Natural and Engineered MCoTI Cyclotides. *Org Biomol Chem* **2008**, *6* (8), 1462–1470.
- (228) Yin, H.; Zhou, X.; Huang, Y. H.; King, G. J.; Collins, B. M.; Gao, Y.; Craik, D. J.; Wang, C. K. Rational Design of Potent Peptide Inhibitors of the PD-1:PD-L1 Interaction for Cancer Immunotherapy. *J Am Chem Soc* **2021**, *143* (44), 18536–18547.
- (229) Pascolutti, R.; Sun, X.; Kao, J.; Maute, R. L.; Ring, A. M.; Bowman, G. R.; Kruse, A. C. Structure and Dynamics of PD-L1 and an Ultra-High-Affinity PD-1 Receptor Mutant. *Structure* **2016**, *24* (10), 1719–1728.
- (230) Pan, X.; Kortemme, T. Recent Advances in de Novo Protein Design: Principles, Methods, and Applications. *Journal of Biological Chemistry* **2021**, *296*, 100558.
- (231) Horne, W. S.; Johnson, L. M.; Ketas, T. J.; Klasse, P. J.; Lu, M.; Moore, J. P.; Gellmana, S. H. Structural and Biological Mimicry of Protein Surface Recognition by α/β -Peptide Foldamers. *Proc Natl Acad Sci U S A* **2009**, *106* (35), 14751–14756.
- (232) Johnson, L. M.; Mortenson, D. E.; Yun, H. G.; Horne, W. S.; Ketas, T. J.; Lu, M.; Moore, J. P.; Gellman, S. H. Enhancement of α -Helix Mimicry by an α/β -Peptide Foldamer via Incorporation of a Dense Ionic Side-Chain Array. *J Am Chem Soc* **2012**, *134* (17), 7317–7320.

- (233) Kelly, S. M.; Jess, T. J.; Price, N. C. How to Study Proteins by Circular Dichroism. *Biochim Biophys Acta Proteins Proteom* **2005**, *1751* (2), 119–139.
- (234) Greenfield, N. J. Using Circular Dichroism Spectra to Estimate Protein Secondary Structure. *Nat Protoc* **2007**, *1* (6), 2876–2890.
- (235) Micsonai, A.; Wien, F.; Kernya, L.; Lee, Y. H.; Goto, Y.; Réfrégiers, M.; Kardos, J. Accurate Secondary Structure Prediction and Fold Recognition for Circular Dichroism Spectroscopy. *Proc Natl Acad Sci U S A* **2015**, *112* (24), 3095–3103.
- (236) Bliven, S.; Prlić, A. Circular Permutation in Proteins. *PLoS Comput Biol* **2012**, *8* (3), 1–5
- (237) Lo, W. C.; Lee, C. C.; Lee, C. Y.; Lyu, P. C. CPDB: A Database of Circular Permutation in Proteins. *Nucleic Acids Res* **2009**, *37* (SUPPL. 1), 328–332.
- (238) Viguera, A. R.; Serrano, L.; Wilmanns, M. Different Folding Transition States May Result in the Same Native Structure. *Nat Struct Biol* **1996**, *3* (10), 874–880.
- (239) Whitehead, T. A.; Bergeron, L. M.; Clark, D. S. Tying up the Loose Ends: Circular Permutation Decreases the Proteolytic Susceptibility of Recombinant Proteins. *Protein Engineering, Design and Selection* **2009**, *22* (10), 607–613.
- (240) Topell, S.; Hennecke, J.; Glockshuber, R. Circularly Permuted Variants of the Green Fluorescent Protein. *FEBS Lett* **1999**, *457* (2), 283–289.
- (241) Qian, Z.; Lutz, S. Improving the Catalytic Activity of Candida Antarctica Lipase B by Circular Permutation. *J Am Chem Soc* **2005**, *127* (39), 13466–13467.
- (242) Byrne, A.; Kier, B. L.; Williams, D. V.; Scian, M.; Andersen, N. H. Circular Permutation of the Trp-Cage: Fold Rescue upon Addition of a Hydrophobic Staple. *RSC Adv* **2013**, *3* (43), 19824–19829.
- (243) Fleishman, S. J.; Leaver-Fay, A.; Corn, J. E.; Strauch, E. M.; Khare, S. D.; Koga, N.; Ashworth, J.; Murphy, P.; Richter, F.; Lemmon, G.; Meiler, J.; Baker, D. Rosettascripts: A Scripting Language Interface to the Rosetta Macromolecular Modeling Suite. *PLoS One* **2011**, *6* (6), 1–10.
- (244) Bonneau, R.; Ruczinski, I.; Tsai, J.; Baker, D. Contact Order and Ab Initio Protein Structure Prediction. *Protein Science* **2002**, *11* (8), 1937–1944.
- (245) Rocklin, G. J.; Chidyausiku, T. M.; Goresnik, I.; Ford, A.; Houliston, S.; Lemak, A.; Carter, L.; Ravichandran, R.; Mulligan, V. K.; Chevalier, A.; Arrowsmith, C. H.; Baker, D. Global Analysis of Protein Folding Using Massively Parallel Design, Synthesis, and Testing. *Science (1979)* **2017**, *357* (6347), 168–175

- (246) Su, H.; Wang, W.; Du, Z.; Peng, Z.; Gao, S. H.; Cheng, M. M.; Yang, J. Improved Protein Structure Prediction Using a New Multi-Scale Network and Homologous Templates. *Advanced Science* **2021**, *8* (24), 1–11.
- (247) Du, Z.; Su, H.; Wang, W.; Ye, L.; Wei, H.; Peng, Z.; Anishchenko, I.; Baker, D.; Yang, J. The TrRosetta Server for Fast and Accurate Protein Structure Prediction. *Nat Protoc* **2021**, *16* (12), 5634–5651.
- (248) D.A. Case, V. Babin, J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R. E.; Duke, H. Gohlke, A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. K.; T.S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K.M. Merz, F. Paesani, D.R. Roe, A. Roitberg, C. S.; R. Salomon-Ferrer, G. Seabra, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X.; Kollman, W. and P. A. AMBER 14. *University of California, San Francisco* **2014**.
- (249) D.A. Case, H.M. Aktulga, K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, G. A.; Cisneros, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R.; Harris, S. Izadi, S.A. Izmailov, C. Jin, K. Kasavajhala, M.C. Kaymak, E. King, A. Kovalenko, T. K.; T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K. M.; Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K.A. O’Hearn, A. Onufriev, F. Pan, S. P.; R. Qi, A. Rahnamoun, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, N. R.; Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, Y. Xue, D.M. York, S.; Zhao, and P. A. K. Amber 2021. *University of California, San Francisco*. **2021**.
- (250) Bogetti, A. T.; Piston, H. E.; Leung, J. M. G.; Cabalteja, C. C.; Yang, D. T.; Degrave, A. J.; Debiec, K. T.; Cerutti, D. S.; Case, D. A.; Horne, W. S.; Chong, L. T. A Twist in the Road Less Traveled: The AMBER Ff15ipq-m Force Field for Protein Mimetics. *Journal of Chemical Physics* **2020**, *153* (6).
- (251) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindah, E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.
- (252) Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase. *J Comput Chem* **2003**, *24* (16), 1999.
- (253) Sorin, E. J.; Pande, V. S. Exploring the Helix-Coil Transition via All-Atom Equilibrium Ensemble Simulations. *Biophys J* **2005**, *88* (4), 2472–2493.
- (254) Buer, B. C.; Levin, B. J.; Marsh, E. N. G. Influence of Fluorination on the Thermodynamics of Protein Folding. *J Am Chem Soc* **2012**, *134* (31), 13027–13034.
- (255) Reinert, Z. E.; Horne, W. S. Folding Thermodynamics of Protein-like Oligomers with Heterogeneous Backbones. *Chem Sci* **2014**, *5* (8), 3325–3330.

- (256) P.V. Konarev, V.V. Volkov, A.V. Sokolova, M.H.J. Koch, D. I. S. PRIMUS: A Windows PC-Based System for Small-Angle Scattering Data Analysis. *J. Appl. Cryst.* **2003**, *36*, 1277–1282.
- (257) D. Franke, M. V. Petoukhov, P. V. Konarev, A. Panjkovich, A. Tuukkanen, H. D.; T. Mertens, A. G. Kikhney, N. R. Hajizadeh, J. M. Franklin, C. M. Jeffries, D. I. S. ATSAS 2.8: A Comprehensive Data Analysis Suite for Small-Angle Scattering from Macromolecular Solutions. *J. Appl. Cryst.* **2017**, *50*, 1212–1225.
- (258) Goddard, T.D.; Kneller, D. G. Sparky, 3 Rd Ed. *University of California; San Francisco, CA, USA* **2001**.