Author:
Porter Goff, Kate

Title:
Design, synthesis and characterisation of new protein folds

Design, synthesis and characterisation of new protein folds

Kathryn Louise Porter Goff

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of PhD in the Faculty of Science.

School of Chemistry, November 2018

53941 words

48433 words (excluding references)

Abstract

The primary sequence of amino acids in a polypeptide chain determines a proteins 3D shape, how it folds and its function. Understanding this process is known as the protein-folding problem. The loss of conformational entropy upon folding must be overcome by the formation of many weak, but cooperative, non-covalent interactions. Studying miniproteins reduces the complexity of the protein-folding problem and allows us to systematically probe contributions to protein folding and stability.

Recently in the Woolfson group, a 34-residue monomeric miniprotein, PPα, was made using a fragment-based design approach that borrowed sequences from two natural protein structures. This antiparallel construct comprises a polyproline-II helix, loop and α helix. PPα is stabilised by the interdigitation of proline from the polyproline-II helix into a stack of tyrosine residues presented by the α helix. There remains much to be understood about the sequence-to-structure and sequence-to-stability relationships in PPα. Further, such knowledge and understanding offers opportunities for using PPα as a reliable modular building block to design and engineer entirely new protein folds and functions.

In chapter 3 of this thesis, the complete rational redesign of PPα is described, moving away from a fragment-based design towards a *de novo* framework. An optimised PPα topology is presented that has been analysed extensively by a variety of biophysical techniques including NMR spectroscopy. Optimised PPα has significantly improved thermal stability compared to the parent design. The enhanced stability is attributed to a number of factors including electrostatic steering effects and general tightening of the structure resulting in improved non-covalent interactions, specifically CH–π interactions. General and intimate side chain-side chain interactions are probed in chapter 4 revealing sequence-to-structure/stability relationships for the miniprotein fold.

With the knowledge of how to design robust PPα modules, chapter 5 explores designing two new topologies based on PPα that, to our knowledge, have not been observed before in nature. Firstly, through iterative rounds of design, we achieve a stable and well-folded α helix:loop:polyproline-II helix topology (αPP). Subsequently, PPα and αPP are combined to give a PPαPP construct with the expanded topology polyproline-II helix:loop:α helix:loop:polyproline-II helix. Through mutagenesis studies it is shown that the PPα component of the PPαPP topology is the more dominant interface and when this is removed the stability of the topology is dramatically reduced.

Two further strategies to improve the stability of PPα are discussed in chapter 6. Firstly, the chain length of optimised PPα is increased. Similar to previous observations for α-helical coiled coils, the stability of PPα increases with increasing chain length in a non-linear, cooperative manner. Secondly, successful enzymatic cyclisation of PPα using an engineered peptide ligase yields cyclic PPα variants with enhanced thermal stabilities.

In summary, an optimised PPα has been designed and sequence-to-stability relationships for the miniprotein elucidated. The first steps towards expanded PPα topologies are presented. These should pave the way for the development of more-advanced PPα-based protein folds with applications in protein engineering and synthetic biology.

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ..... DATE:.....

Acknowledgments

Firstly, thank you to my supervisor, Dek, for your endless enthusiasm, support and guidance throughout my PhD. I have thoroughly enjoyed being a part of your group for the past three and a half years and have learnt so much.

Thank you to everyone in the Woolfson group, both past and present, for being such a fantastic lab and good fun to work with. I want to especially thank Emily for her support, advice and intellect throughout my PhD and for helping proofread this thesis. Thanks also to Chris Williams for running the NMR experiments and solving the structure presented herein.

I am grateful to the EPSRC and the University of Bristol for funding my PhD project and to the Bristol Chemical Synthesis CDT management team – Kevin, Emma, Mar and Laura – for their support.

Thank you to my friends for their constant encouragement, in particular: my Bristol gal pals, for all the laughs and hot meals, particularly while writing up, and my climbing friends, for the trips away climbing in beautiful surroundings. A special thank you to my family – Mum and Dad for your unwavering support and interest in my work and Ben for your good humour. Finally, thank you Will for all your support and keeping me smiling during my PhD.

List of Abbreviations

AUC	Analytical ultracentrifugation
BUDE	Bristol University docking engine
CD	Circular dichroism
Cl-HOBt	6-Chloro-1-hydroxybenzotriazole
COSY	Correlation spectroscopy
DIC	*N*,*N*-Diisopropylcarbodiimide
DIPEA	*N*,*N*-Diisopropylethylamine
DHB	2,5-Dihydroxybenzoic acid
DMF	*N*,*N*-Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DPH	1,6-Diphenyl-1,3,5-hexatriene
f/f_0	Frictional ratio
Fmoc	Fluorenylmethyloxycarbonyl
HATU	Hexafluorophosphate azabenzotriazole tetramethyl uronium
HPLC	High-performance liquid chromatography
HSQC	Heteronuclear single quantum coherence spectroscopy
ISAMBARD	Intelligent system for analysis, model building and rational design
K_d	Dissociation constant
KIH	Knobs-into-holes
MALDI-TOF	Matrix-assisted laser desorption/ionisation time-of-flight
MD	Molecular dynamics
MRE	Mean residue ellipticity
Mw	Molecular weight
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser effect spectroscopy
PBS	Phosphate buffered saline
PDB	Protein data bank
PPII	Polyproline type II
RMSD	Root-mean-square deviation

rt	Room temperature
SASA	Solvent accessible surface area
SE	Sedimentation equilibrium
SPPS	Solid-phase peptide synthesis
SV	Sedimentation velocity
TFA	Trifluoroacetic acid
TIPS	Triisopropylsilane
T_M	Thermal midpoint of unfolding
TOCSY	Total correlation spectroscopy
UV	Ultraviolet
$\bar{v}$	Partial specific volume
ZF	Zinc finger

Amino acids: Alanine (Ala, A), Arginine (Arg, R), Asparagine (Asn, N), Aspartic acid (Asp, D), Cysteine (Cys, C), Glutamic acid (Glu, E), Glutamine (Gln, Q), Glycine (Gly, G), Histidine (His, H), Isoleucine (Ile, I), Leucine (Leu, L), Lysine (Lys, K), Methionine (Met, M), Phenylalanine (Phe, F), Proline (Pro, P), Serine (Ser, S), Threonine (Thr, T), Tryptophan (Trp, W), Tyrosine (Tyr, Y), Valine (Val, V).

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 The protein-folding problem

Max Pertuz and John Kendrew were awarded the 1962 Nobel Prize in Chemistry for their pioneering work using X-ray crystallography to determine the molecular structure of the globular proteins myoglobin and haemoglobin; the former of which was the first three-dimensional structure of a protein (**Figure 1-1**).[1-2] They found that both molecules were built from Linus Pauling's α helices, but folded and packed together in intricate manners. Kendrew described the main features of myoglobin as its "complexity and lack of symmetry" adding "the arrangement seems to be almost totally lacking in the kind of regularities which one instinctively anticipates."[3] Their work raised the question of how protein structures are explained by physical principles and subsequently the notion of a folding problem first emerged.

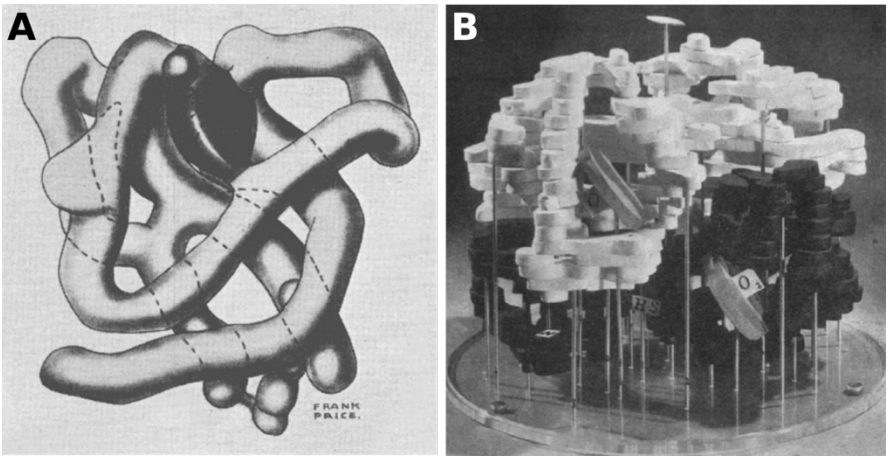


Figure 1-1: Structures of myoglobin and haemoglobin (A) Drawing of the tertiary structure of myoglobin as deduced from the 6 Å fourier synthesis. (B) Haemoglobin model, haem groups indicated as disks. Images taken from Kendrew and Pertuz papers.[1-2]

The protein-folding problem is concerned with how a protein's primary linear amino-acid sequence dictates its three-dimensional native structure. More specifically, protein-folding studies aim to understand the sequence-to-structure relationships that govern how proteins fold.[4-5] Since the problem was first posed over half a century ago, Dill *et al.* describe the protein-folding problem as having grown from a distinct research question into a self-perpetuating field of science.[6]

A landmark in protein science was the postulate that resulted from Christian Anfinsen's experiments on the renaturation of Ribonuclease A.[7] He hypothesised that the native structure of a protein is the thermodynamically stable state, and that this depends only on its amino-acid sequence and the solution conditions, not on the kinetic folding route.[8] Further, folding does not depend on whether the protein is synthesised *in vivo* or *in vitro*. This led Cyrus Levinthal to pose the question of *how*

do proteins fold to their native state? The Levinthal paradox involves the consideration that a polypeptide chain cannot converge on its native state in real-time by a random search because the conformational space available to an unfolded polypeptide chain is too large.[9-10] Levinthal reasoned that the protein must fold through a directed process, or protein folding pathways.

Energy landscapes and folding funnels are used as a conceptual framework for understanding protein folding kinetics (**Figure 1-2**).[11] A protein folds by taking steps that are generally incrementally downhill in energy towards a single stable low-energy native state. Folding can occur *via* multiple pathways in the conformational space landscape,[12] hence why denatured peptides starting from different open conformations can refold to the same native structure. However, this is not always the case, there are exceptions in which the biologically active form is kinetically trapped, for example serpins.[13]

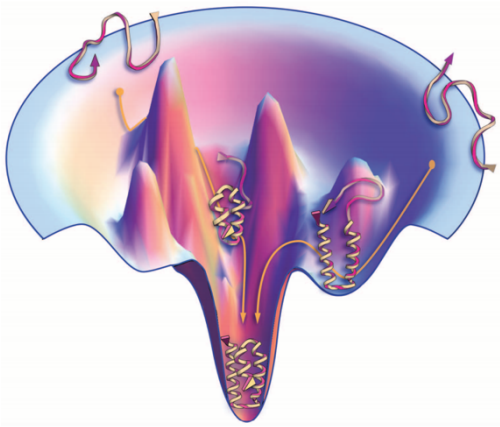


Figure 1-2: The funnel-shaped energy landscape of proteins has few low energy native like conformations and more unfolded structures. Image taken from an article by Dill and MacCullum.[6]

The folded state of a protein is determined by the cooperation of many weak non-covalent interactions including hydrogen bonds, van der Waals interactions, electrostatic interactions and hydrophobic interactions.[14] Hydrogen bonds are important because they form between backbone amide and carbonyl groups in nearly all secondary structure components including α helices and β sheets. Similarly, the atoms within a folded protein are tightly packed implying van der Waals interactions are also important.[15] While some amino-acid side chains attract or repel one another as a result of negative and positive charges these groups are usually limited to the protein surface and so electrostatic interactions are unlikely to be the dominant interaction for folding.[4, 16] Given that it is the side chains of amino acids that are responsible for the differences between proteins, and not the backbone hydrogen bonds, it is thought that hydrophobic interactions between non-polar side chains are the dominant driving force for folding. Supporting this view, globular proteins have hydrophobic cores where non-polar amino acids are buried from water. Also, studies have shown that there is a 1-2 kcal mol^{-1} benefit for sequestering a non-polar side chain from water into a hydrophobic oil-like media.[17]

Opposing the folding process is a large loss in conformational entropy as the protein collapses to its folded state. This penalty is only just outweighed by the enthalpy of forming the non-covalent interactions and as a result, native proteins are only marginally stable, approximately 5-10 kcal mol^{-1} more stable than their denatured states.[6, 14] This subtle balance between forces makes predicting the three-dimensional structure of proteins particularly challenging.[18-19] One way to address the protein-folding problem is to study small folding motifs termed miniproteins. In these systems the number of factors that influence the folded state is reduced making them good platforms for probing contributions to protein folding and stability both *in silico* and *in vitro*.[20-23] Larger proteins have generally evolved for function as opposed to for stability and so it is harder to deconvolute their sequences into the components that are responsible for stability over function.

1.2 Miniproteins

Miniproteins are defined as short peptides (approximately ≤ 40 amino acid residues) that usually have well-defined globular-like folds with sequestered hydrophobic cores.[21] They comprise two or more elements of secondary structure and have cooperative folding. While single secondary structure components exist and can be cooperatively folded, for example single α helices (SAHs), this is rare and they do not contain a hydrophobic core like globular proteins.[24] While the entropic cost of folding is lower for miniproteins, their hydrophobic cores are smaller and there are fewer non-covalent interactions to stabilise the folded state. This means that many miniproteins rely on other methods of stabilisation including metal binding,[25] covalent-crosslinking,[26] un-natural side-chain staples,[27] cyclisation[28] and oligomerisation.[29] These miniproteins will be discussed briefly later. However, the focus of this thesis is on miniproteins stabilised by non-covalent interactions.

As well as providing systems to probe various aspects of protein-folding specifically sequence-to-structure and -stability relationships, miniproteins also have potentially useful applications. For instance, they provide scaffolds onto which functional domains can be grafted specifically motifs for catalysis, binding and recognition that interfere with protein-protein interactions. Also, miniproteins can be used as building blocks to design and engineer larger, more complex folds and systems.[30-32] Furthermore, they offer a starting point from which to access the *dark matter of protein space*, that is, protein folds that have not been observed before in nature.[33-34]

1.2.1 Miniproteins stabilised by metal binding and covalent crosslinking

1.2.1.1 Zinc fingers

Zinc fingers (ZnFs) are one of the most widespread protein-folding motifs involved in DNA binding and recognition. The most widely studied and engineered ZnF is the Cys2His2 (C2H2)-type.[35] The C2H2-ZnF is a small independently folded domain with a compact globular structure in which two histidine and two cysteine residues coordinate in a tetrahedral manner to a Zn ion. The first NMR structure corresponding to a single ZnF from the *Xenopus* protein Xfin revealed a 25-residues peptide composed of a β hairpin, a loop and an α helix (**Figure 1-3A**). Just above where the Zn

binds, Leu and Phe residues form a small hydrophobic pocket; the rest of the amino acids are solvent-exposed.[25] The four amino-acid residues located at the *N*-terminus of the α helix participate in DNA recognition by interacting with hydrogen donors and acceptors exposed in the major groove of DNA.[36] The modular assembly of metal-binding ZnF domains has allowed the engineering of artificial proteins and enzymes that recognise defined regions of DNA that can activate, repress or alter user-specified genes. These have been used in the field of targeted genome editing.[37] While natural ZnFs require metal binding for folding, their ββα architectures have been engineered to fold in the absence of metal ions (see section 1.2.2.2).

1.2.1.2 EF-hands

The EF-hand calcium-binding motif plays an essential role in cell signalling.[38] The motif, which is approximately 35-residues long, comprises two short α helices separated by a metal-binding loop.[39] The Ca^{2+} ion binds to the loop in a preferred coordination geometry of seven ligands arranged in a pentagonal bipyramid. EF-hands usually occur in pairs to form discrete units in larger domains, most commonly containing two, four or six EF-hands. The EF-hand pair is stabilised through hydrophobic interactions between helices as well as π-π stacking interactions between single aromatic residues in the loop of each EF-hand. While stand-alone single-EF hands do exist, they are generally stabilised through dimerization (**Figure 1-3B**).[38]

1.2.1.3 Cysteine knots

Cysteine-knot miniproteins, often termed knottins, consist of approximately 30 – 40 amino acid residues.[40] They have an unusual architecture where two disulfide bonds form a ring through which a third disulfide bond is threaded. The resulting structure has exceptional stability and resistance to proteolysis.[41] A subset of these proteins are termed cyclotides, which, as well as having a knot framework, are characterised by their head-to-tail cyclised peptide backbone. Kalata B1 was the first cyclotide miniprotein to be structurally characterised (**Figure 1-3C**).[42] It has three antiparallel β strands stabilised by mainchain and side-chain hydrogen bonds and hydrophobic clusters. It also has a number of well-ordered tight turns stabilised by an extensive hydrogen bond network. A number of Gly residues are conserved in the turn regions of the cyclotide framework where adaptability of the backbone geometry is important. A Glu residue is also conserved that is involved in key hydrogen-bonding interactions.

Cysteine knots play roles in various biological processes acting as enzyme inhibitors, hormones, growth factors, toxins, and host-defence molecules.[43] Their exceptional stabilities and well-defined structures make them promising scaffolds for molecular engineering and pharmaceutical applications.[26] While knottin cores are conserved, the loops between Cys residues are tolerant to mutation and have shown a broad range of sequence diversity thus making them suitable candidates for grafting bioactive epitopes.[42]For example, the Cochran lab have used directed evolution to engineer knottin peptides to bind with nanomolar affinity to integrin receptors which are overexpressed in a variety of cancers.[44-45]

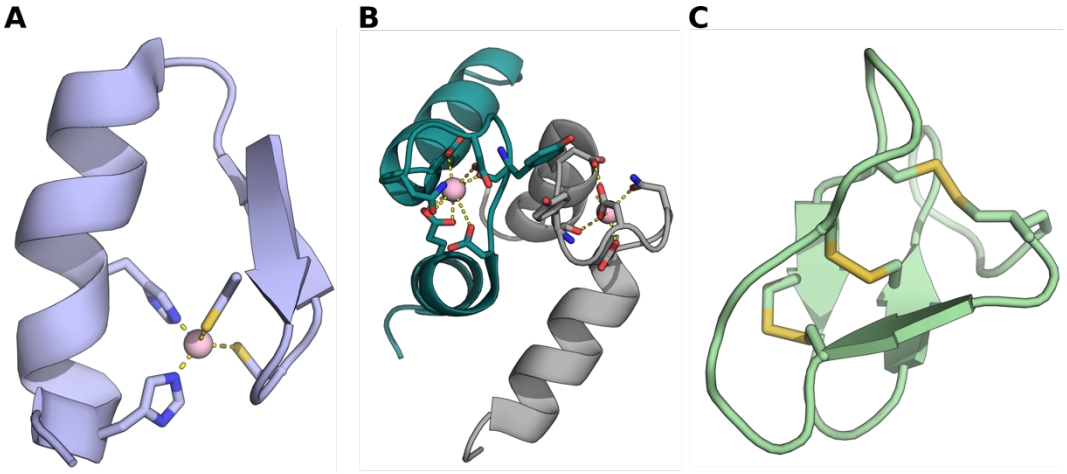


Figure 1-3: NMR structures of miniproteins stabilised by metal binding and covalent crosslinking. (A) Single ZnF (PDB: 1ZnF). (B) Homodimeric EF-hand (PDB: 1CTA). (C) Kalata B1 cyclotide (PDB: 1NB1). Metal ions shown as pink spheres. Hydrogen bonds shown as dashed lines. Images generated using PyMol.

Overall, considerable progress has been made in the field of miniproteins stabilised by covalent-crosslinking and metal binding. Naturally occurring pharmacologically active miniproteins have shapes that are highly complementary to binding pockets on their targets. These peptides combine the stability and cell permeability of small molecules with the binding affinity and target specificity of larger protein-therapeutics and therefore are attractive drug molecules.[46-49] Consequently, effort has focused on reengineering naturally occurring miniproteins to generate new bioactive molecules.[44] However, these approaches are hampered by the limited variety of naturally occurring, constrained miniproteins and the lack of global shape complementarity with the target. In response to this problem, the Baker group have recently developed computational methods for the accurate *de novo* design of miniproteins stabilised by disulfide crosslinks.[50] 12 experimentally determined X-ray and NMR structures of miniproteins that are 18 to 47 residues long containing both canonical and non-canonical amino acids have been solved revealing very close matches to the computationally-designed models (**Figure 1-4**). These miniproteins span a broad range of topologies incorporating multiple secondary structure components. The miniproteins were designed using the Rosetta software suite.[51] Large numbers of peptide backbones were randomly generated, combinatorial sequence design calculations were performed to identity sequences that stabilised each backbone conformation, and finally designed sequence/structures were assessed to determine which ones were in deep energy minima, that is, which ones had a large energy gap between the designed structure and alternative conformations.[50]

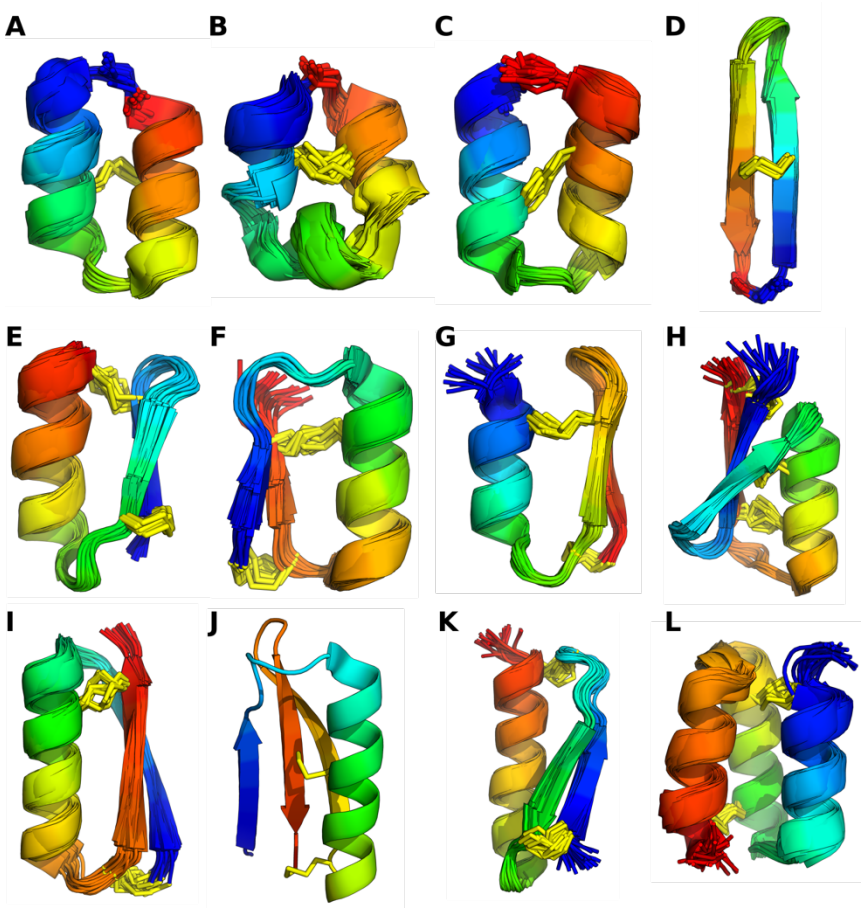


Figure 1-4: NMR and crystal structures of the hyper-stable constrained miniproteins designed by the Baker lab. (A-D) Backbone cyclised miniproteins. (E-G) Miniproteins incorporating both L- and D-amino acids. (H-L) Genetically encodable miniproteins. Note: (J) is an X-ray crystal structure (A-I,K,L) are NMR structures showing overlays of all states in each ensemble. Coloured blue to red from *N* to *C* terminus. Disulfide bond crosslinks shown in yellow. PDB codes: 5JG9, 2ND2, 2ND3, 5JHI, 5JI4, 5KVN, 5KWO, 5KWP, 5KX2, 5KWZ, 5KX1, 5KX0. Images generated using PyMol.

1.2.2 Miniproteins stabilised by non-covalent interactions

The focus of this thesis is on non-covalently stabilised miniproteins. While the majority of miniproteins observed and studied are stabilised by covalent crosslinks and metal ions, it is non-covalently stabilised miniproteins that will allow in depth examination of fundamental sequence-to-stability relationships without the complications of stability aids or larger protein contexts. Over the past half-century the majority of non-covalently stabilised miniproteins described have been fragments of larger natural proteins that have undergone iterative rounds of redesign and optimisation to enhance stability or introduce functional motifs. This has resulted in general design rules for miniproteins and a greater appreciation for and understanding of the non-covalent interactions present in such structures. More recently, high-throughput methods have also been used to design miniproteins.[52-53]

1.2.2.1 Pancreatic polypeptides

The polyproline-II:loop:α helix topology was first observed in the X-ray crystal structure of the avian pancreatic peptide (aPP) dimer (**Figure 1-5A**).[29] Pancreatic peptides are usually 36 residues in length and adopt a well-defined hairpin structure in water. The compact fold is stabilised by the interdigitation of Pro residues from the polyproline-II helix and aromatic residues stacked on the α helix forming a hydrophobic core. aPP is a symmetric dimer stabilised by the intermolecular π-stacking interaction between Tyr residues and is maintained in solution above micromolar concentrations.[54-55] Indeed, most natural pancreatic peptides dimerise in solution. This has driven engineering and design of monomeric variants with a focus on developing miniprotein-based ligands as therapeutics.

The Schepartz lab have used the aPP fold extensively as a scaffold to present α-helical and polyproline-II-helical recognition epitopes. For example, a synthetic monomeric aPP analogue named pGolemi was engineered that binds strongly (K_d 700 nM) to the Mena EVH1(enabled/VASP homology 1) domain of the Actin assembly-inducing protein (Acta) from *Listeria monocytogenes*.[56-57] EVH1 domains regulate actin filament dynamics and, similar to SH3 and WW domains, they recognise proline-rich sequences that are folded into polyproline-II helices. pGolemi was designed by grafting a polyproline-binding domain onto the *N*-terminal polyproline-II helix of aPP. As well as having high affinity for EVH1, 10-fold higher than the best previously known EVH1 ligand Acta11,[58] pGolemi also discriminates between paralogs and reduces bacterial motility.

A similar sequence-grafting strategy has been used for the aPP variant p007, which presents the α-helical recognition epitope found on the bZIP protein GCN4 and binds DNA with nanomolar affinity and high specificity.[59]Also substitution of Arg residues for residues located on the α helix of aPP help facilitate cell permeability without significant loss of helicity or thermal stability.[60] Furthermore, the bovine pancreatic peptide (bPP) has been designed as an artificial esterase to catalyse the hydrolysis of *p*-nitrophenyl esters. To do this His residues have been grafted onto the solvent exposed face of the α helix.[61]

1.2.2.2 ββα folds and metal free zinc fingers

Small, independent ββα units are best exemplified by ZnFs as mentioned in section 1.2.1.1. While folding of natural ZnFs is driven by metal binding as opposed to the hydrophobic core, as noted above, several research groups have targeted this fold for metal free designs. In 1996, Struthers *et al.* obtained a 23-residue monomeric metal free ββα motif through an iterative design process in which amino acids were selected that would enhance the inherent secondary structure of the polypeptide.[62] The final design incorporated a suitable turn that more effectively promoted β-hairpin formation, a D-Pro and a non-natural amino acid. The NMR structure revealed that the α helix and β hairpin are held tightly together by a defined hydrophobic core. Interestingly, the solution structure reveals the motif is more open than in the natural ZnFs. Struthers states this groove may be useful for catalysis or the design of enzyme mimics. Truncating the loop between β hairpin and α helix

prevents the hairpin folding back on the α helix exposing a hydrophobic surface and promoting homo-oligomerisation.[63]

A year later, a computational design using only proteogenic amino acids was reported by Dahiyat *et al*.[64] A combinatorial design algorithm was used to screen a combinatorial library of 1.9×10^{27} amino acid sequences for compatibility with the target ββα topology.[65-66] The design was validated experimentally by NMR spectroscopy (**Figure 1-5B**). The backbone RMSD between the average NMR state and the design target was 1.98 Å. The peptide had a T_M of 39 °C, with weakly cooperative and completely reversible unfolding curve. The broad melt transition observed is consistent with a low enthalpy of folding, which is to be expected for a peptide with a small hydrophobic core.

1.2.2.3 Villin headpiece

Another approach to miniprotein design is to pare down larger natural proteins. Villin is an F-actin-bundling protein involved in microvilli formation in absorptive epithelia.[67] In 1997, NMR spectroscopy of the *C*-terminal 35 residues of the chicken villin (HP-35) in aqueous solution revealed a well folded peptide consisting of three α-helical segments with each helix contributing residues to the central hydrophobic core (**Figure 1-5C**).[68-69] HP-35 is surprisingly thermostable with a T_M of 70 °C.

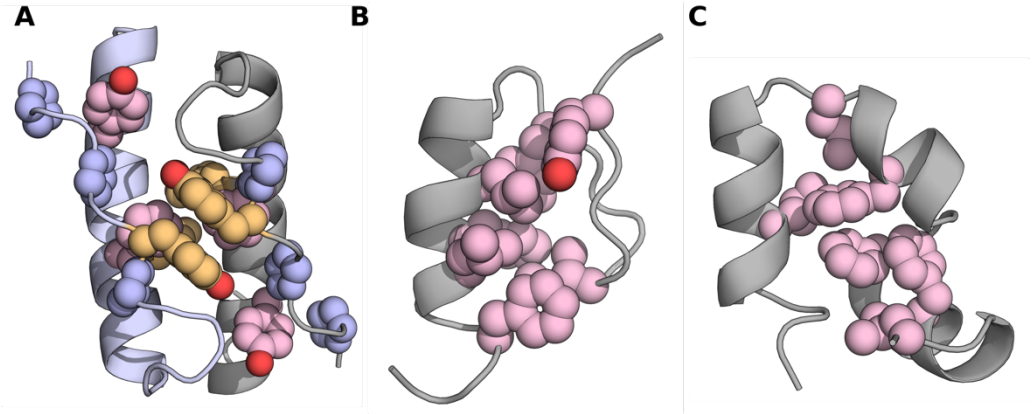


Figure 1-5: Structures of miniproteins. (A) X-ray structure of avian pancreatic peptide dimer (PDB: 1PPT). (B) NMR structure of metal free ZnF (PDB: 1PSV). (C) NMR structure of Villin Headpiece (PDB:1VII). Colour key: aromatic and hydrophobic residues (pink), π-π stacking between Tyr residues in aPP dimer (yellow), Pro (blue). Images generated using PyMol.

1.2.2.4 β Hairpins and Trp-zippers

Many research groups have explored β hairpin formation in naturally derived and designed systems. Early design efforts resulted in free-standing β structures based on natural fragments, which were only moderately folded in aqueous media.[70-71] However since then more-folded structures have been attained through largely iterative processes as well as some combinatorial screening.[72] Trp-zippers (12-16 residues in length) are well-studied examples of short β-hairpins with interlocked Trp residues showing edge-to-face packing against one another (**Figure 1-6A**).[73-74] The β sheets are highly twisted and the peptide is also stabilised by capping residues at the termini to permit cross-

strand hydrogen bonding. These miniproteins are highly folded monomers exhibiting exceptional thermal stabilities (T_M 50 – 79 °C) and reversible cooperative unfolding curves. Chignolin is another example of a minimal β hairpin (10 residues) and is stabilised by the edge-to-face packing of Tyr and Trp residues.[75] The peptide shows cooperative folding albeit with a broad thermal melt curve (T_M 39 °C). The chignolin peptide has been used in phage display to identify a nanomolar binder of the Fc region of antibody immunoglobulin G.[76] Trp-pocket β hairpins are stabilised by cation−π interactions in which Lys packs against a diTrp cleft on the opposite strand.[77] Some Trp-pockets are near fully folded and can only be unfolded with high concentrations of denaturant or elevated temperatures.

Overall, there has been substantial work in the field and as a result a number of general guidelines can be considered to ensure the successful design and stabilisation of monomeric β hairpins[78]: a hydrophobic cluster on one surface of the hairpin close to the loop region;[79-80] inter-strand side-chain interaction with particular focus on Trp-Trp interactions;[73, 81-82] high β-sheet and turn propensities;[83] and charged, aromatic residues or β-capping motifs at the termini.[84]

1.2.2.5 Designed three-stranded β sheets

Initial designs of three-stranded β-sheets have been achieved by appending a third strand onto previously characterised β hairpins.[85-88] Most of these constructs however are largely unfolded in purely aqueous media and require methanol to fold. One successful example, however, incorporates a Pro-Gly turn and *N*-methylated amino acids in the first and third strands to prevent amyloid-like aggregation.[88] The 23-residue peptide is a soluble monomer and folds upon heating. No high-resolution, structural data has been reported, the peptide is characterised by circular dichroism (CD) spectroscopy, size-exclusion chromatography and 1D NMR line shape analysis.

A 20-residue peptide named Betanova has been designed that forms a monomeric, three-stranded, antiparallel β-sheet.[85] The NMR structure reveals that the peptide is stabilised by an aromatic-rich hydrophobic cluster on one face of the β sheet. However, subsequent more in-depth studies show that Betanova is only partially folded.[89] This has led to the redesign of the fold through computationally informed point mutations that resulted in a more-stable structure.[90]

The Gellman lab has designed a parallel triple-stranded β-sheet by incorporation of non-peptide linkers that connect pairs of strands in a parallel orientation.[91] The solution structure reveals the desired peptide fold and CD spectroscopy shows a cooperative thermal unfolding transition. Enhanced folding is observed for the three-stranded-sheet relative to each two-stranded parallel hairpin component.

WW domains are natural antiparallel three-stranded β-sheets and are one of the smallest naturally occurring folds. Their name reflects the conserved Trp residues in the first and third strands. WW domains mediate protein-protein interactions with short proline-rich motifs.[92] The first structure of a WW domain was for human Yap65 in complex with its ligand.[93] Following this, shorter WW

domains have been isolated of 34 and 37 residues in length in absence of bound ligands.[94] Based on a systematic analysis of conserved residues in the WW family, a 36-residue folded prototype WW sequence has been designed and an NMR structure determined (**Figure 1-6B**).[94] It must be noted, however, that the prototype is less thermally stable than natural β-sheets (T_M 44 °C). WW domains have been re-engineered to incorporate different functions for example a DNA binding pocket that shows 10-fold selectivity for single-stranded DNA over duplex DNA.[95]

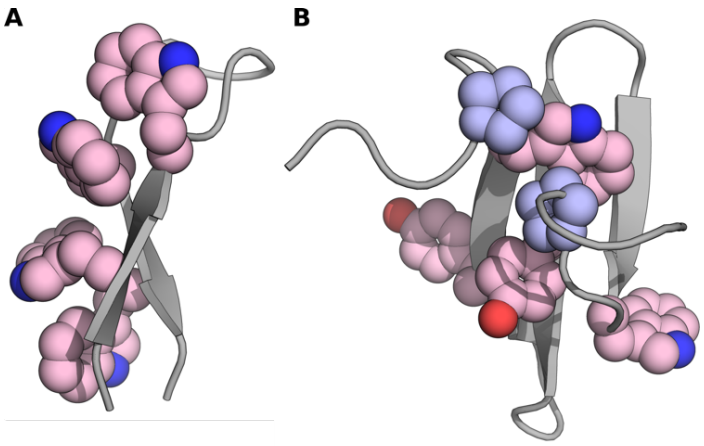


Figure 1-6: Structures of miniproteins. (A) NMR structure of Trp-Zipper (PDB:1LE3). (B) NMR structure of the prototype WW domain (PDB: 1E0M). Colour key: aromatic residues (pink), Pro (blue). Images generated using PyMol.

1.2.2.6 The Trp-cage

The Trp-cage is a highly and cooperatively folded 20-residue miniprotein truncated from the 39-residue extendin-4 (EX4) peptide found in Gila monster saliva. Original truncation of the *C*-terminal portion of EX4 resulted in a peptide that was only folded in the fold-promoting cosolvent 2,2,2-trifluroethanol (TFE).[96] However, a series of incremental sequence modifications has led to a well-ordered fold comprising an *N*-terminal α helix followed by a structured loop with a hydrophobic core centred on a single Trp residue surrounded by Pro side chains (**Figure 1-7A**).[96] It is noted that the high Pro content may contribute to the stability of the peptide by limiting the conformational entropy loss associated with folding. In contrast to this, the peptide also has high Gly content – the most flexible and least rigid of all the natural amino acids. Gly flexibility may however allow the backbone to contort and further finetune the tertiary fold. Together, Gly, Pro and Trp form an organised fold with tightly buried side-chain interactions.[97] Further mutations to the Trp-cage has seen increases in thermal stability with T_Ms up to 64 °C.[98] Cyclisation of one variant gave a T_M of 95 °C.[99] The Trp-cage has lent itself as a prime target for computational study as a result of its small, well folded nature and the large amount of structural data.[20, 100]

1.2.2.7 βαβ Designs

Extended βαβ repeats are widely observed in natural proteins, for example: TIM barrels consist of eight α helices and β sheets that alternate along the chain forming a closed barrel like structure;[101] the Rossmann fold is a structural motif found in proteins that bind nucleotides and forms an open

sheet structure;[102] and Leu-rich repeats form a structural motif that resembles an α/β horseshoe.[103] The alternating secondary structures lead to parallel β sheets and parallel α helices.

The first *de novo* design of a stand-alone water-soluble βαβ motif was reported in 2009 (**Figure 1-7B**).[104] The design comprises a 12-residue α helix, selected after a statistical analysis of the helix length in natural α/β proteins, paired with two five-residue β strands which were of corresponding length. Sequences were selected based on secondary-structure preferences. An amphipathic helix was designed with Leu and Ala on the hydrophobic face and electrostatic interactions between Glu and Lys were arranged on the opposite face. Ile and Val were placed on the parallel β sheet forming a hydrophobic pocket in the core of the miniprotein. Initial designs were molten globule. To obtain the stable folded state, a Trp-Trp pair in the β strands, similar to Trp-zippers, was introduced.[73] An NMR structure of the final 36-residue design reveals face-to-face packing of the pair. The βαβ motif is highly thermally stable up to 90 °C. This is exceptional for a peptide containing only natural proteinogenic amino acids and without disulfide bonds or metal binding.

1.2.2.8 TrpPlexus

TrpPlexus is a 19-residue miniprotein with a β strand:loop:polyproline-II helix topology designed using a fragment-based approach (**Figure 1-7C**).[105] The *N*-terminal β strand is rich in Arg residues, while the *C*-terminal polyproline-II strand is Trp rich with a WSXWX motif. Despite the name, the polyproline-II helix does not contain any Pro residues. The sequences for the β strand and polyproline-II helix are borrowed from a fibronectin type III (f3) binding domain and are connected with a D-Pro,Gly loop. The NMR structure of TrpPlexus shows the interdigitation of Trp and Arg residues, which leads to cation-π interactions. This network of solvent exposed cation-π interactions offers an alternative to hydrophobic-core packing. TrpPlexus is monomeric with a broad thermal unfolding transition. However, the unfolding transition sharpened when unfolding was measured as a function of guanidinium chloride concentration showing a two-state sigmoidal transition.

Further work by the Kirshenbaum group synthesised a disulfide cyclised variant and showed its tolerance for *N*-substituted Gly and Pro resides in the polyproline-II helix.[28] The ability to incorporate a broad variety of monomer types in the polyproline-II helix provides opportunities for developing peptide and peptidomimetic inhibitors that can target PXXP motif-binding proteins and modulate protein-protein interactions in cellular signalling processes.

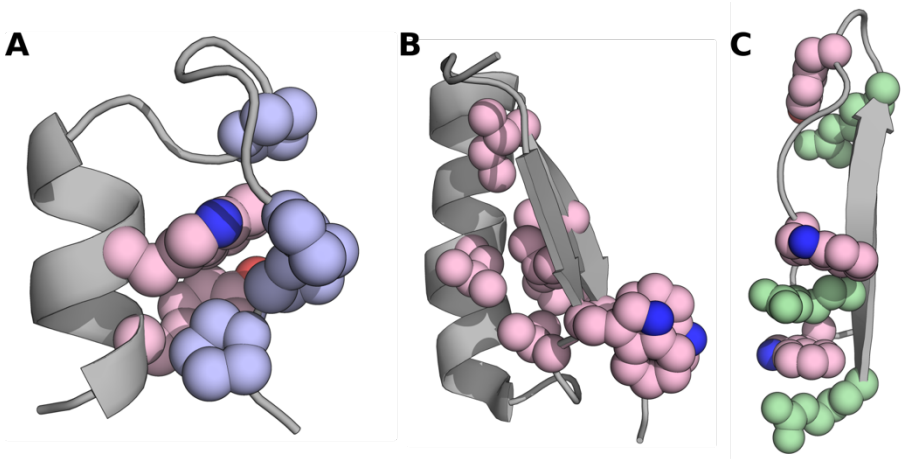


Figure 1-7: Structures of miniproteins. (A) NMR structure of the Trp-Cage (PDB:1L2Y). (B) NMR structure of a βαβ motif (PDB: 2KI0). (C) NMR structure of TrpPlexus (PDB: n/a, see ref [105]. Colour key: aromatic residues (pink), Pro (blue), Arg (green). Images generated using PyMol.

1.2.2.9 High-throughput methods to miniprotein design

In recent years, high-throughput methods have been used in miniprotein design. The Baker group have used parallel protein design on a large scale to build thousands of new miniprotein variants and determine what sequence motifs stabilise these small structures.[52, 106] In their approach miniprotein structures were designed computationally using a fragment-based approach in Rosetta. Four target topologies are described - ααα, βαββ, αββα, ββαββ - of which only the ααα topology is observed in nature within their target size range of 40 to 43 amino acids. Libraries of amino-acid sequences were generated to fit best each of these structures. These proteins were obtained *via* high-throughput DNA synthesis and cloning. The resulting miniproteins were displayed on the surface of yeast so that every cell had many copies, with each protein genetically fused to an expression tag that was fluorescently labelled with an antibody and subsequently exposed to protease. Stable variants were distinguished from unstable variants based on their resistance to protease treatment. Protease cleavage of the miniprotein led to loss of the tag and loss of fluorescence. The cells that retained fluorescence after proteolysis were then isolated by fluorescence-activated cell sorting and identified by deep sequencing. The isolated miniproteins were characterised and their sequences analysed for sequence-to-stability relationships which could be fed back into the design cycle. Iteration between design and experiment improved design success rate from 6% to 47%.

From sequence analysis of these new stably folded miniproteins Rocklin *et al.* highlight key sequence and structural features: firstly, the amount of buried non-polar surface area from hydrophobic amino acids. While this is well established, they quantify its importance showing that stable variants need at least 30 $Å^2$ per buried residue. The study also revealed the importance of having charged side-chain residues at the termini of α helices that balance the terminal partial charges of the helices. Overall, four high-resolution miniprotein structures of each target topology were obtained by NMR spectroscopy (**Figure 1-8**).

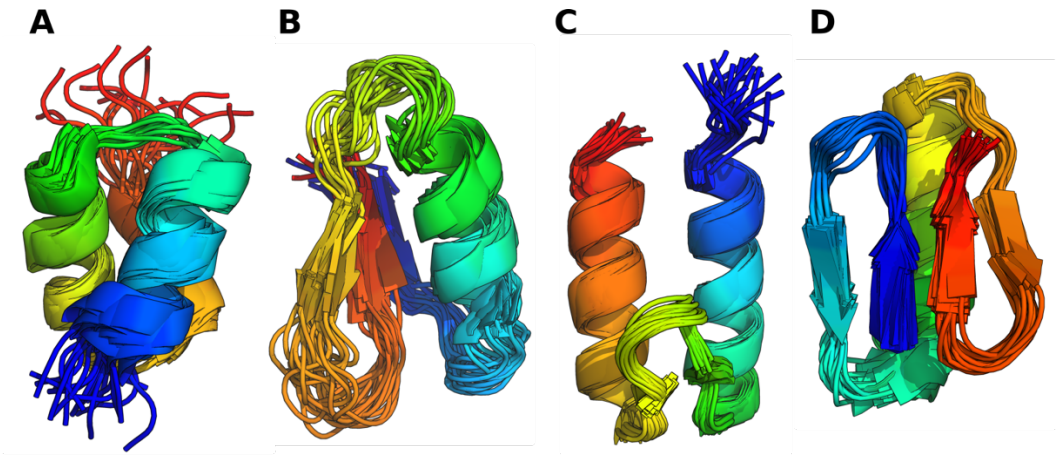


Figure 1-8: Designed peptide from the Baker lab using high-throughput methods. (A) ααα topology (PDB: 5UOI). (B) βαββ topology (PDB: 5UP5). (C) αββα topology (PDB: 5UYO). (D) ββαββ topology (PDB: 5UP1). Coloured blue to red from *N* to *C* terminus. Images generated using PyMol.

1.2.3 Common features of miniproteins

Whether the entropic cost of folding is less for small proteins or not, enthalpically favoured interactions still need to be made to form a stable, folded 3D structure. Inherent to their small size, miniproteins make fewer non-covalent interactions. Experimentally, this is apparent from the broader melt curves in CD spectroscopy, which are indicative of low enthalpies of folding.[64] This is consistent with miniproteins having small hydrophobic cores. Therefore, to form stable miniproteins the non-covalent interactions that are made need to be optimal.

The stable miniproteins described above have many common features in terms of both sequence and structure. These can be analysed to understand and refine the most important non-covalent interactions needed for creating stable miniprotein folds. A bioinformatics analysis of miniproteins in the PDB by the Woolfson group revealed that some classes of amino acids are more prevalent in miniproteins compared to larger proteins,[21] including: electron-rich aromatics, in particular Tyr and Trp residues; long and charged amino acids for example Arg, Glu and Lys; and Pro residues. In contrast, it was found that small non-polar amino acids are more prevalent in larger proteins. This suggests that these aromatics and the longer charged residues may form more optimal non-covalent interactions such as CH-π interactions, cation-π interactions, π-π interactions and salt bridges. Furthermore, it is noted that CH-π interactions are six times more dense in miniproteins compared to larger proteins. The prevalence of Pro may be due to reducing conformational entropy.

In the examples studied, the observed hydrophobic cores often have aromatic residues in contact with residues such as Pro and Arg leading to networks of CH-π and cation-π interactions, respectively. High proportions of Pro residues also likely reflects their ability to reduce the entropic cost of folding through their fixed torsion angle. Related to this, the backbones of miniproteins are more contorted and make better use of structured loops compared to larger counterparts. The

contorted backbones likely helps sequester hydrophobic residues and pack them more tightly in their core.

The α helix secondary structure is a common building block in both miniproteins as well as larger proteins and are observed in similar proportions for each class. While examples of free-standing α helices have been observed in nature and have also been designed they are more commonly observed in tertiary and quaternary structures, the most common of which is the α-helical coiled coil.[24, 107]

1.3 Coiled coils: sequence, structure and function

α-Helical coiled coils are ubiquitous folding motifs consisting of a two or more helices wrapped around each other to form supercoils.[108] Constituting approximately 3 % of all protein-encoding regions of the known genome,[109-110] coiled coils have a broad range of functions including structural roles, DNA binding and mediating protein-protein interactions.[111-113] While predicting the structure of proteins from their primary amino-acid sequence is difficult for the majority of proteins, there are well-established rules and sequence-to-structure and -stability relationships for coiled coils. Arguably, it is the best understood of all protein structures. Furthermore, we are able to predict their structures and also design new sequences that fold into well-defined assemblies.[114-117] The sequence and structural features of coiled-coil peptides are discussed below.

The primary sequence of the majority of coiled-coil regions consists of a seven-residue repeat commonly known as a heptad repeat in which residues are labelled ***abcdefg***. These repeats have hydrophobic (H) and polar (P) residues arranged in the pattern HPPHPPP; *i.e.* hydrophobics occupy the ***a*** and ***d*** positions of the ***a*** through ***g*** repeat. While there are variations on this pattern, generally consisting of different combinations of HPP and HPPP sequences, we will focus on the most common heptad pattern.[118-119] The identity of the hydrophobic and polar amino acid residues dictates the final structure of the assembly, including helix orientation (parallel or antiparallel), oligomerisation state, and homotypic or heterotypic assemblies.[120-124]

The primary sequence of a coiled coil folds into α-helical secondary structure elements which are stabilised by backbone hydrogen bonds between donor H^N atoms at residue i and carboxyl oxygen atoms at residue i+4 (**Figure 1-9A**). These hydrogen bonds lie parallel to the helical axis.[125] The canonical right-handed α helix can be described by a number of parameters, including: the rise per residue, equal to 1.5 Å; the residues per turn, equal to 3.6; the helix radius, equal to 2.3 Å; and phi (Φ) and psi (Ψ) torsion angles, in the region of –63° and –43° respectively (**Figure 1-9B &C**).[126-127]

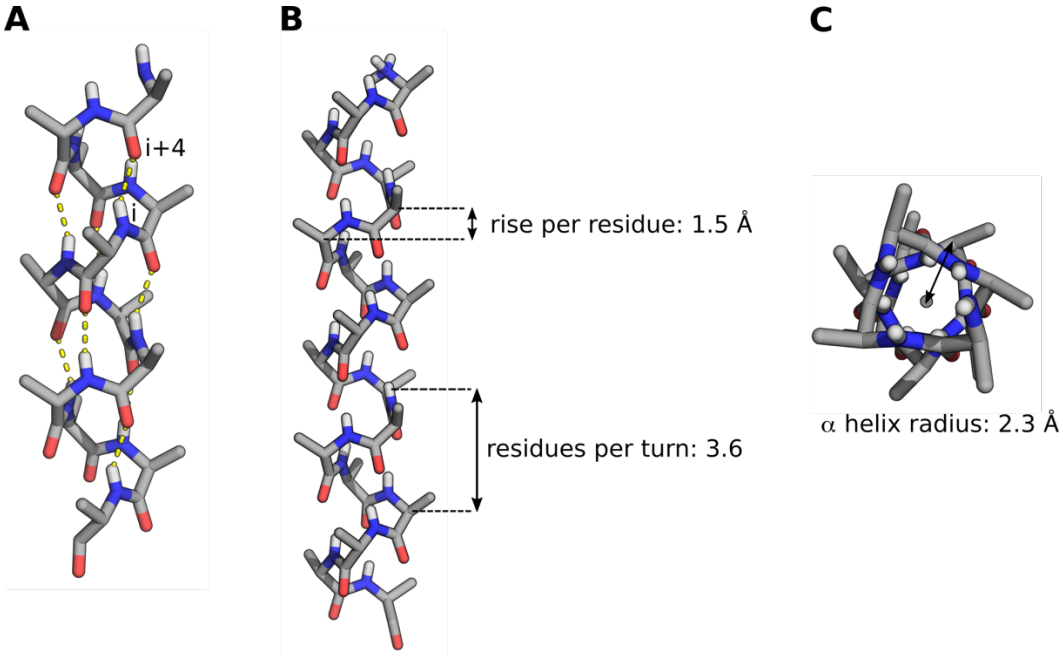


Figure 1-9: The parameters that describe α helix secondary structure. (A) α helix showing hydrogen bonding between backbone NH and C=O groups of residues i to i+4 apart. (B) α helix showing the rise per residue and residue per turn parameters. (C) View down α helix from the *N*-terminus highlighting helical axis and radius of the α helix.

Other secondary structures found in folded proteins include: polyproline-II helices (discussed in detail in Section 1.4.1), β sheets, 3_{10} helices, and π helices. Their corresponding torsion angles are represented visually in the Ramachandran plot in **Figure 1-10** and their parameters detailed in **Table 1-1**.

Table 1-1: Examples of secondary structures and their parameters.

Secondary structure	Residues per turn	Rise per residue / Å	Radius of helix / Å
α helix	+3.6	1.5	2.3
3_{10} helix	+3	2	1.9
π helix	+4.3	1.1	2.8
Polyproline-II	–3.0	3.1	1.6
β strand	–2.3	3.3	1.0

3_{10} and π helices are stabilised by hydrogen bonds between peptide amide and carbonyl groups of residues (i, i+3) and (i, i+5) respectively. Both structures are rare, the 3_{10} helix is usually found at the ends of α helices, while the π helix has only been observed in a handful of structures persisting for, at most, only a couple of turns.[128-129] The majority of designed 3_{10} helices contain non-natural amino acids, in particular α-aminoisobutyric acid (AIB) residues which contain a second methyl group attached to the C_α atom.[130] Alongside α helices, β strands are the other major structural element found in globular proteins.[131] β strands are arranged adjacent to other strands and associate by hydrogen bonding to form a β sheet-like structure which, in the majority of cases, is twisted. β sheets are made up of extended strands and fall within a wide shallow energy minimum of

Ramachandran space. They can interact in parallel or antiparallel orientations, each with a distinctive hydrogen-bonding pattern and side chains alternating above and below the sheet.

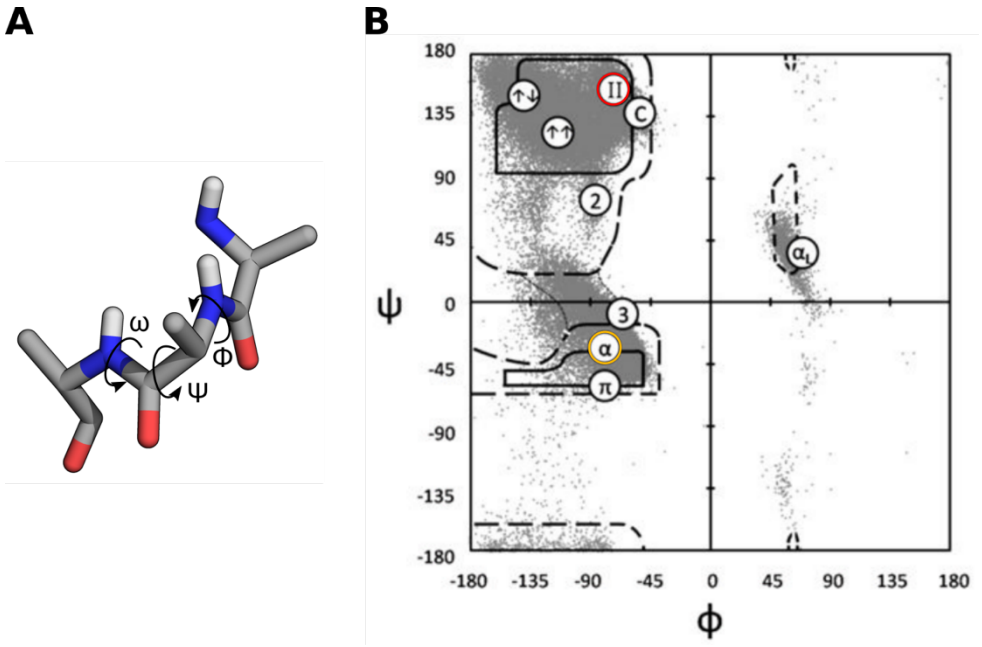


Figure 1-10: Torsion angles and Ramachandran space. (A) section of an α helix showing the torsion angles Φ, Ψ, and ω. The planarity of the amide bond restricts ω to be 180°. (B) Ramachandran plot highlighting α helix region in orange and polyproline-II helix region in red. Modified from reference [126].

Given that the α helix repeats itself every 3.6 residues and hydrophobic side chains are spaced at combinations of three and four residues in the HPPHPPP sequence repeat, an amphipathic helix is formed when the heptad is projected onto a helical wheel (**Figure 1-11A&B**).[132-133] The hydrophobic effect is a key driving force in most biological self-assembling processes. This phenomenon results in biological molecules in aqueous solution folding or self-assembling to minimize the hydrophobic surface area in contact with the bulk solvent. Therefore, in the context of coiled coils, amphipathic helices come together to bury their hydrophobic ***a*** and ***d*** residues (**Figure 1-11C**).

When α helices wrap around each other they form left-handed super-helical coiled coils. The supercoiling results from a mismatch in periodicity between the heptad repeat (H residues on average every 3.5 residues) and the α helix (3.6 residues per turn) and thus an ***a***/***d*** hydrophobic seam winds slowly around the right-handed helix in a left-handed manner. Given the regular nature of coiled coils, their structures can be described by three geometric parameters, as first described by Crick in 1953[134-135]: radius, interface angle and pitch (**Figure 1-11D**). A range of software applications have built upon these parameters allowing for the parametric modelling of coiled coils. Hardbury *et al.* used the Crick equations to help design a right-handed coiled-coil tetramer,[136-137] Woolfson *et al.* modified the Crick equations to allow non-canonical repeats to be modelled[138] and De Grado *et al.* has developed a method for fitting Crick parameters to known coiled-coil structures.[139] More recently, CC-Builder and ISAMBARD have been developed for building and optimising models of coiled coils of a broad range of oligomeric states.[140-142] Further, the

ISAMBARD framework can be used to model any protein fold that is parameterizable; that is, folds that have a regular structure and can be described by mathematical parameters.

When the α helices of coiled coils wind around each other the side chains at the interface do not simply contact each other. Instead, the α helices take part in intimate side-chain-packing interactions. A side chain referred to as a "knob" on one helix interdigitates with a diamond-shaped "hole" made up of four residues on a partnering helix. This type of intimate packing was first postulated by Crick and is termed "knobs-into-holes" (KIH) packing (**Figure 1-11E&F**).[143] For example, in parallel coiled coils, an ***a*** knob docks into a ***dgad*** hole and a ***d*** knob docks into an ***adea*** hole.

1.3.1 Oligomer-state specification

The connection between sequence and KIH packing was first discussed in seminal work by Harbury *et al*.[120, 144] In their experiments they found that mutating the core ***a*** and ***d*** positions of leucine-zipper peptide GCN4-p1 to combinations of Leu, Ile and Val residues resulted in changes to the oligomeric state of the assembly: when Ile and Leu were placed at ***a*** and ***d***, respectively, a dimer was formed; Ile at both positions resulted in a trimer; and Leu at ***a*** and Ile at ***d*** gave a tetramer. This result was rationalised in terms of how the ***a*** and ***d*** side chains project towards their partner helix. In dimers, the ***a*** position projects out of the interface while the ***d*** position projects towards the interface. These projections change with changes in oligomerisation state. In dimers, ***a*** and ***d*** participate in parallel and perpendicular packing, respectively. Parallel packing is where the C_α–C_β bond vector of the knob residue points parallel to the vector between the C_α positions of the hole residues on the corresponding helix while in perpendicular packing the knob C_α–C_β bond vector points perpendicular. In tetramers this packing is swapped over, while in trimers packing is somewhere between the two and is referred to as acute packing. Leu is most tolerated at perpendicular sites while beta-branched Ile (and Val) is preferred at parallel or acute sites. Perpendicular and parallel packing are discussed further in Section 4.1.1.

Harbury's experiments paved the way for future work in the field of coiled-coil design. Extensive work has since been carried out into the design and specification of both lower (< 5 helices in assembly) and higher-order coiled coils (>4 helices in assembly).[32, 145-147] The oligomeric state of such coiled coils is determined by the interhelical interfaces present in their assemblies. The addition of helices from dimer through to tetramer leads to increased inclusion of peripheral residues (as well as ***a*** and ***d***) within the hydrophobic core of the assembly. In particular, residues at ***g*** become knobs as well as just being part of the hole residues. These are referred to as Type I interfaces compared to the classical Type N interfaces, which are present in coiled-coil dimers (**Figure 1-12**). Increasing the hydrophobic core to include ***e*** knob residues as well leads to Type II interfaces and specifies pentamers through to heptamers. Type III assemblies have two distinct hydrophobic seams with a single intervening residue between them and specify even larger assemblies such as the 12-helix barrel of TolC.[148]

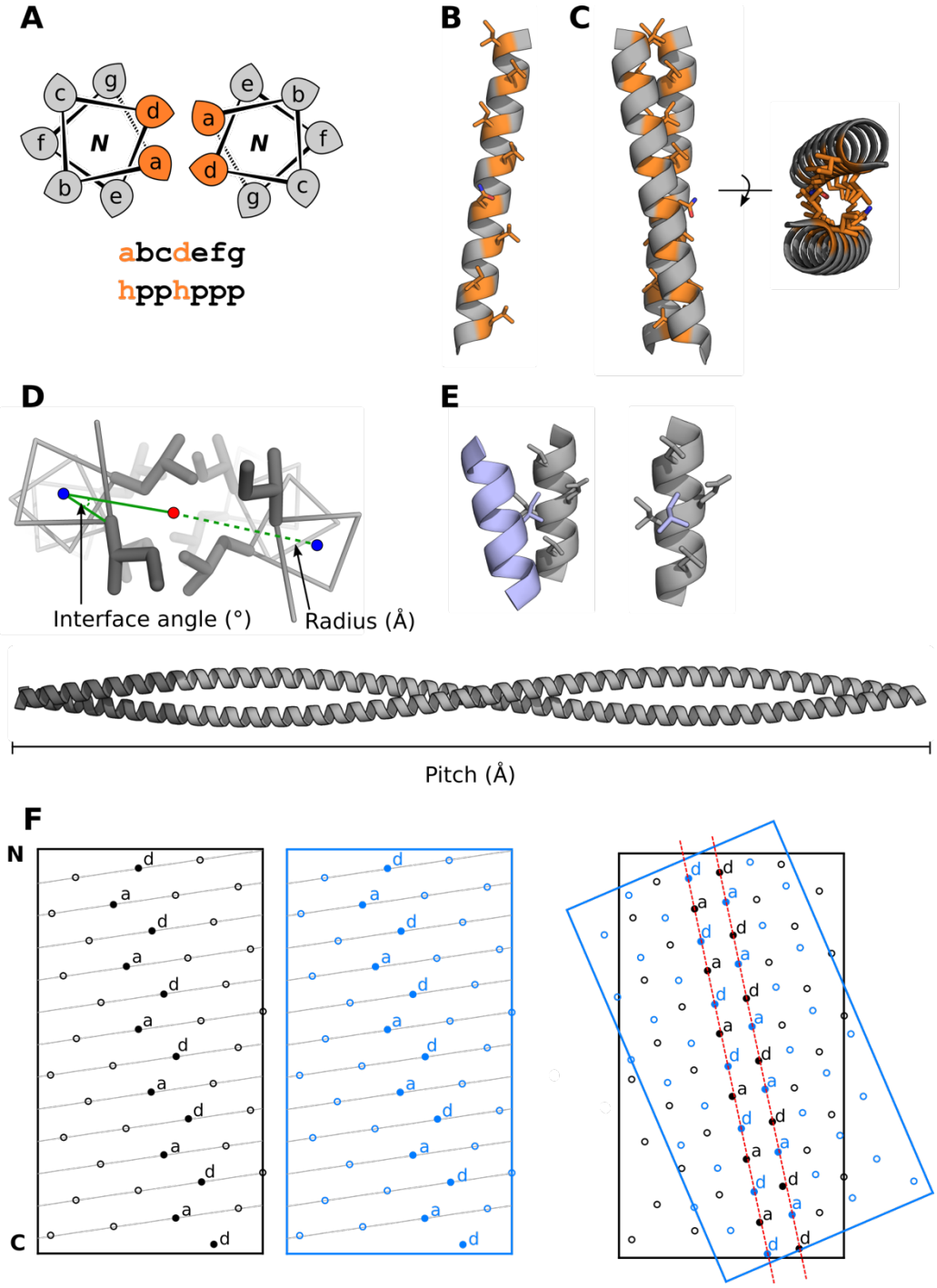


Figure 1-11: Structure of α-helical coiled coils. (A) Projecting the heptad HPPHPPP repeat onto a helical wheel results in a seam of hydrophobic residues. Orientation of the leaf shape shows the direction of the C_α-C_β bond vector. N refers to the helix terminal nearest the viewer. (B) Amphipathic helix. (C) Coiled-coil dimer showing hydrophobic seam in orange. View down the coiled-coil axis shown on right. (D) Geometric parameters describing coiled coils: Interface angle (°), radius (Å), and pitch (Å). (E) Knobs-into-holes packing in coiled coils. (F) Helical nets for two identical α helices (left, N termini at the top, positions of C_α shown as circles) and superimposed helices (right) to show tight KIH packing in a coiled-coil interface. The core ***a*** and ***d*** residues are highlighted as filled circles.

Using these sequence-to-structure relationships as rules for protein design, a basis set of *de novo* designed peptides has been achieved for oligomeric states from dimer through to heptamer.[32, 145] X-ray crystal structures of these assemblies are shown in **Figure 1-13**. The higher-order coiled coils (>4 helices) have central open-ended channels and, therefore, are α-helical barrels. For these barrels the ***a*** and ***d*** sites are predominantly occupied by Leu and Ile residues. The residues at ***e*** and ***g*** essentially fine-tune oligomeric state selection: for CC-pent (five helices in the assembly) the flanking ***e*** and ***g*** residues are large Ile and Glu residues, respectively; in CC-Hex the ***e*** position is Ala and the ***g*** position is Glu or Ser; while in CC-Hept both ***e*** and ***g*** positions are Ala. Thus, in general, the smaller the residues at ***e*** and ***g*** the larger the helix-helix-helix interface angle and therefore the larger the barrel.

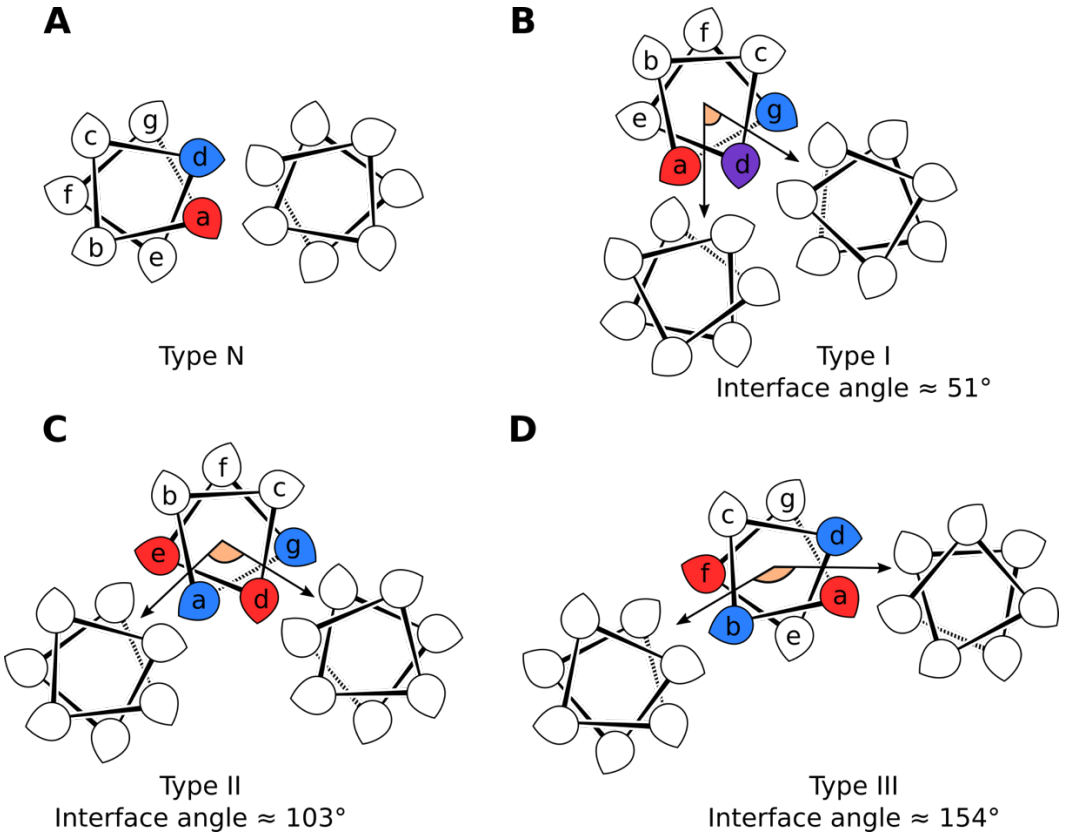


Figure 1-12: Coiled-coil interfaces. (A) Type N interface, found predominantly in dimers. (B) Type 1 interface found in trimers and tetramers. (C) Type II interface found in tetramers to heptamers. (D) Type III interface found in octamers and above. Coloured residues show "knobs" in KIH packing. Colour key: parallel packing (red), perpendicular packing (blue), acute packing (purple). The angle between two hydrophobic seams is highlighted in orange is termed the interface angle. Adapted from reference [113].

These defined and well-characterised coiled coils have been used as building blocks for the reliable construction of more complex assemblies. Brodsky *et al.* used a homotrimeric coiled coil to promote folding of a recombinant bacterial collagen.[149] Many collagen helices appear to require adjacent globular domains to assist folding. Brodsky *et al.* successfully replaced a natural trimerization domain with a *de novo* designed peptide. The coiled-coil "toolkit" has been engineered to make peptide fibres with detailed structural characterisation by cryoTEM.[150] Further, using a peptide

origami approach, Jerela *et al.* have combined six coiled-coil pairs in a single polypeptide to direct the folding of a tetrahedron nanostructure.[151] Functional residues can also be introduced into the lumen of the higher order coiled coils. A Cys-His-Glu catalytic triad was successfully introduced into the pore of CC-Hept and showed hydrolytic activity.[152] This is the first example of a functional catalytic triad engineered into a *de novo* protein scaffold.

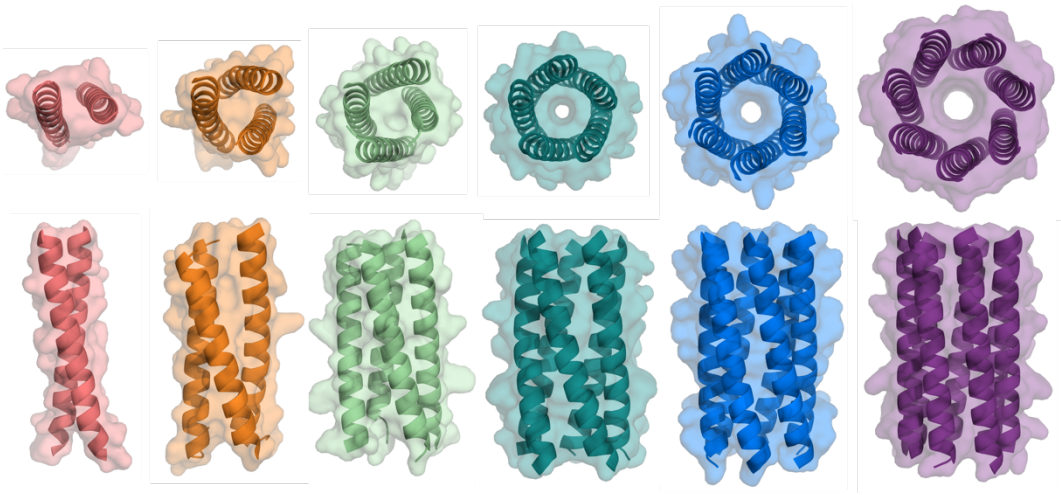


Figure 1-13: X-ray crystal structures of the *de novo* designed basis set ranging in oligomeric state from two to seven helices. CC-Di (red, PDB 4DZM), CC-Tri (orange, PDB: 4DZL), CC-Tet (green, PDB: 3R4A), CC-Pent (teal, PDB: 4PN8), CC-Hex (blue, PDB: 3R3K), CC-Hept (purple, PDB: 4PNA). Images generated using PyMol.

1.4 Hybrid coiled-coil miniprotein: PPα

Recently, the Woolfson group used a fragment based design approach to create a stable, monomeric coiled-coil-hybrid miniprotein named PPα.[153] While coiled coils are stabilised by burying the hydrophobic residues in each amphipathic helix by oligomerisation, PPα is stabilised by the buttressing of an α helix with a polyproline-II helix. Overall, PPα comprises an *N*-terminal polyproline-II helix, a loop region and a *C*-terminal α helix.

1.4.1 Polyproline-II helix secondary structure

Polyproline-II helices have long been recognised as the dominant conformation in the collagen triple helix (**Figure 1-14A**).[154] The characteristic structure is composed of a three-residue repeat Xaa-Yaa-Gly sequence where Xaa and Yaa are often Pro and Hyp (hyroxyproline), respectively. The triple helix contains a one residue stagger, which allows Gly to pack in the core of the structure. More recently, the polyproline-II helix has emerged as a structural class not only of fibrous proteins but also of folded and natively unfolded proteins more generally. Bioinformatics analyses of the PDB show that 2 % of all residues in the PDB adopt polyproline-II conformation.[155-158] While this is significantly lower than that of the α helix and β structure, it is comparable to that of 3_{10} helices. Despite its name, the presence of Pro residues in polyproline-II helix is not a prerequisite (See

TrpPlexus, Section 1.2.2.8). Indeed, up to 46 % of polyproline-II helices in folded peptides and proteins do not contain Pro residues.[159]

The polyproline-II helix is an extended left-handed helix defined by torsion angles in the region of $\Phi = -75°$ and $\Psi = 145°$. It has a rise per residue of 3.1 Å compared to 1.5 Å for α helices and in the ideal form has three residues per turn; this number can vary slightly in natural structures (**Figure 1-14B**).[160] As a result of its extended character and the fact that the backbone amino and carboxyl groups point away from the helical axis, the polyproline-II helix does not support regular patterns of intrachain hydrogen bonds. Instead the helix is stabilised by main-chain hydrogen bonds with water molecules as well as side chain-main chain hydrogen bonds, specifically, the Gln side chain can participate in a hydrogen bonds with the backbone carboxyl oxygen of the preceding residue.[156] Also stabilisation is achieved through protein-protein interactions.[159] Polyproline-II helices are more flexible in comparison to α helices, although this flexibility is partially reduced when Pro is present in the sequence.

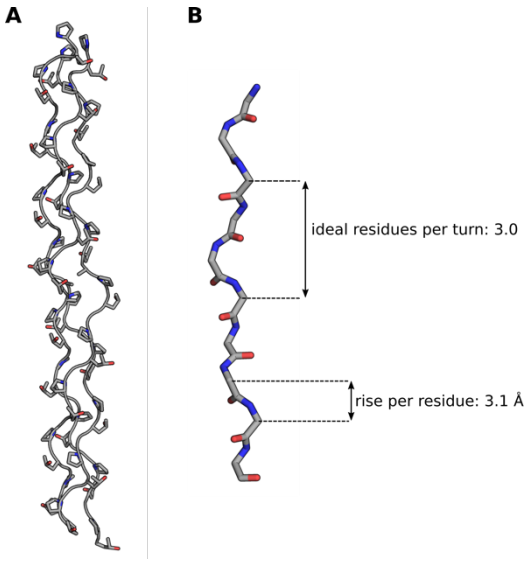


Figure 1-14: Structure of collagen and polyproline-II helices. (A) Crystal structure of a collagen-like peptide (PDB: 1CAG). (B) Segment of a polyproline-II helix highlighting ideal residue per turn of 3 and a rise per residue of 3.1 Å. Images generated in PyMol.

Polyproline-II helices play an important role in both protein-protein as well as protein-nucleic acid interactions and mediate a wide range of molecular functions.[161] They are considered the most widely spread binding motif in proteins. Their important role in binding is a result of their flexible structure and preferred location on the surface of proteins. The absence of intrachain hydrogen bonds means the helix is more flexible and can adjust its conformation to form hydrogen bonds to its target. Examples of recognition domains that bind ligands in the polyproline-II conformation include SH3, WW, EVH1, GYF, UEV and profilin domains.[160] Specific examples are discussed in more detail in Section 1.2.1 and Chapter 6.

Compared to α helix and β sheets, polyproline-II helices have not gained widespread interest. While this is mainly due to their low frequency of occurrence it is also because there are only a few methods

for polyproline-II assignment. The most widely used secondary structure assignment method, DSSP, which is used by the PDB does not assign polyproline-II helices. DSSP assigns secondary structure on the basis of particular hydrogen bonding patterns. Several secondary structure assignment methods that do include polyproline-II helices, for example XTLSSTR,[162] PROSS,[163] and SEGNO[164] give divergent results. A new consensus assignment method (DSSP-PPII) however, based on the 'coil' class of DSSP, is in good agreement with earlier methods.[165] DSSP-PPII assigns polyproline-II helices based on at least two consecutive residues being within the torsion angle range $\Phi = -75 \pm 30°$ and $\Psi = 145 \pm 30°$.

1.4.2 PPα design and characterisation

The PPα design borrowed components from two natural structures: the bacterial surface adhesin and antigen (AgI/II) in *Streptococcus mutans*, and the bovine pancreatic peptide hormone (**Figure 1-15**).[55, 166] In both structures a polyproline-II helix and α helix combine to form a tertiary structure in which Pro residues on the former interdigitate with Tyr residues presented by the latter. *Streptococcus mutans* is the causative agent of human tooth decay and AgI/II has become a target for protective immunity. AgI/II is a cell surface localised protein adhesin that mediates adhesion to the tooth surface and influences biofilm formation. While a mechanistic understanding of the functional properties of the peptide is unknown, removal of AgI/II results in decreased virulence. A high-resolution (1.8 Å) crystal structure of the A_3VP_1 fragment of AgI/II revels an extended α helix that is intertwined with a polyproline-II helix with a left-handed supercoil. The α helix extends to approximately 155 Å in length. The bovine pancreatic peptide is a shorter peptide that is part of the pancreatic peptide family and forms dimers.

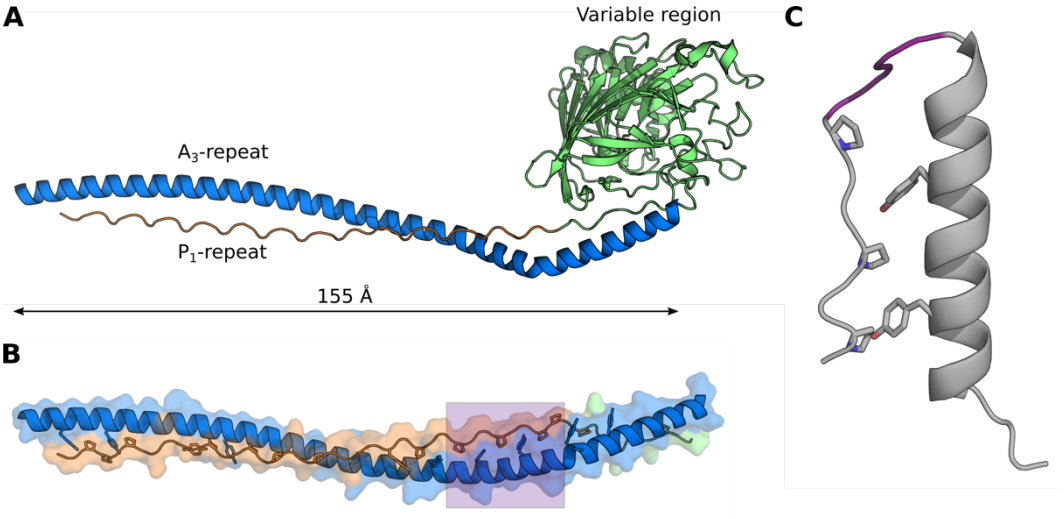


Figure 1-15: Natural protein components used in the design of PPα. (A) Bacterial surface adhesin (AgI/II) from *Streptococcus mutans* (PDB: 3IOX) highlighting the variable region, the extended α helix region (A_3-repeat) and the polyproline-II helix (P_1-repeat). (B) Surface structure of the α and polyproline-II helices highlighting interdigitation of Pro and Tyr residues. The purple box shows the fragment used in the design of PPα. (C) Bovine pancreatic peptide (PDB: 1BBA). Region in purple is the loop region borrowed in the design of PPα. Images generated in Pymol.

Two short segments of polyproline-II helix and α helix were chosen from the bacterial surface adhesin (highlighted in purple in **Figure 1-15B**) for the PPα design. Segments were selected that best matched the different periodicities of the two secondary structure. Given that engineering loops in protein design is challenging, the two helices were connected with the loop from the bovine pancreatic peptide. A model generated for the topology was found to be stable over 100 ns of molecular dynamics (MD) simulation in water. A 34-residue peptide named PPα-Tyr, was synthesised and characterised in solution. CD spectroscopy showed that the peptide was folded and stable with approximately 50 % helicity at 5 °C and a T_M of 39 °C. Sedimentation analytical ultracentrifugation (SE AUC) showed PPα-Tyr was monomeric in solution. NMR spectroscopy was used to determine the solution structure of PPα-Tyr. The NMR ensemble was consistent with the design model. KIH-type packing was observed between the two helices where Pro residues on the polyproline-II helix dock into a diamond shaped hole formed on the α helix top and tailed by Tyr residues (**Figure 1-16**).

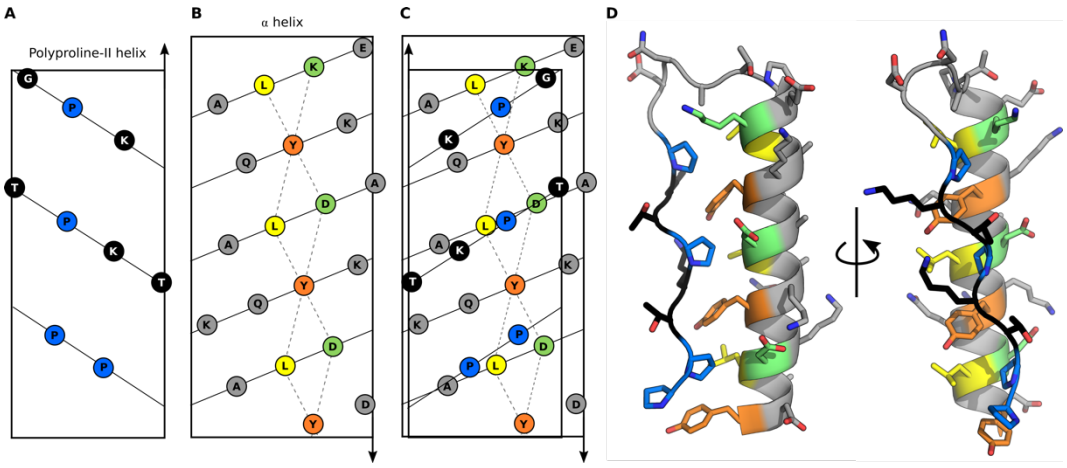


Figure 1-16: Design of PPα. (A) Helical net of the polyproline-II helix component of PPα. (B) Helical net of α helix component of PPα. (C) Overlay of the helical nets of the polyproline-II and α helices showing KIH pacing of Pro residues into the diamond holes top and tailed with Tyr. (D) *In silico* model of PPα-Tyr after 100 ns of MD simulations in water. Adapted from Baker *et al.*[153]

1.4.3 CH-π interactions in PPα

Close contacts were observed between the Tyr residues and the surrounding side chains prompting a search for potential CH–π interactions in the NMR ensemble. 45 CH–π interactions (2.25 per ensemble structure) were detected based on parameters from previous studies.[167-168]These were attributed as follows: 24 CH–π interactions between Pro and Tyr; 4 between Leu and Tyr; 15 between Lys and Tyr; and 2 between Gly and Tyr. This suggested that beyond the hydrophobic effect and van der Waals' forces from side-chain packing, PPα-Tyr is stabilised by CH–π interactions. To probe this hypothesis further stability studies were carried out in which a series of eight para-substituted phenylalanine mutants of PPα were synthesised, from electron rich *p*-methoxyphenylalanine through to electron-poor *p*-nitrophenylalanine. Proteinogenic Trp and Phe

and His mutants were also synthesised. NMR structures of the three most stable mutants, PPα-OMe (p-methoxyphenylalanine substituents) and PPα-Me (p-methylphenylalanine substituents), showed folded polyproline-II:loop: α helix topologies with intimate contacts and CH–π interactions between Pro and the modified aromatic rings (**Figure 1-17**).

The T_Ms of the *para*-substituted phenylalanine variants were plotted against the corresponding Hammett constant (σ_p). The Hammett constant was used as a proxy for the electron density in the aromatic ring. It was observed that peptides with electron-rich aromatic π systems were more thermally stable than those with electron withdrawing substituents. This suggests there is an electronic contribution to the Pro-aromatic interactions. Furthermore, it provides evidence for CH–π interactions; electron density from the ring is redistributed into the CH bond and therefore these interactions are favoured by the electron-donating substituents in the series. The Hammett plot plateaued for PPα-φCN and PPα-φNO_2 mutants consistent with cyano and nitro substituted benzene rings having reduced interaction energies with CH groups. Interestingly, PPα-φNH_2 has a lower than expected T_M for which there is no clear explanation.

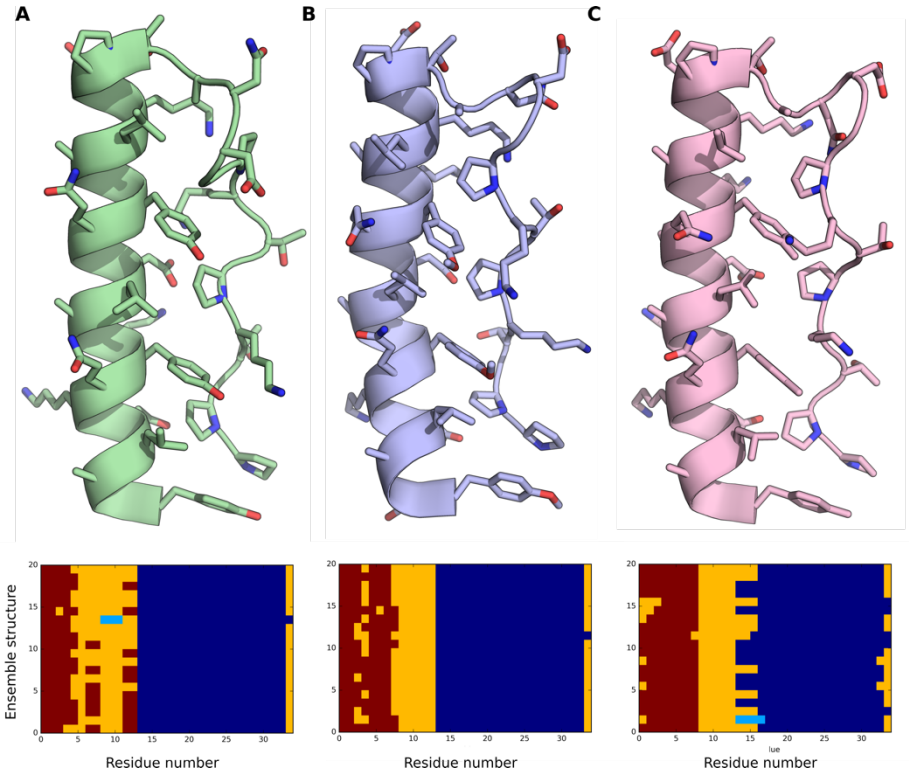


Figure 1-17: NMR ensemble structures and corresponding secondary structure assignments. (A) PPα-Tyr (B) PPα-φOCH_3 (All Tyr residues substituted for *p*-methoxyphenylalanine) (C) PPα-φCH_3 (All Tyr residues substituted for *p*-methylphenylalanine). Key: Polyproline-II helix (red); bend or hydrogen-boned turn (yellow); 3_{10} helix (minimum length 3 residues, light blue); α helix (minimum length 4 residues, dark blue). Residue 0 = *N*-terminal acetyl cap.

Thermodynamic parameters were calculated through a van't Hoff analysis. A range of free energies of unfolding (ΔG_{unf}) were determined, which varied linearly with σ_p. Interestingly, the ΔG_{unf} values

were spread over a 0.9 kcal mol^{-1} range which is similar to the literature estimate for CH–π interactions (1.5-2.8 kcal mol^{-1}). Given that each ensemble structure from the NMR data had between two and three CH–π interactions, small differences in energy shift the equilibrium constants by nearly an order of magnitude. This implies that non-covalent interactions affect the energetics of folding and association significantly.

A bioinformatics analysis of the PDB supported these conclusions: Pro-Tyr and Pro-Trp interactions are observed more frequently than expected by chance. In contrast, Pro-Phe interactions are underrepresented. Further, Pro-Tyr and Pro-Trp make many more CH–π interactions than any other side-chain interactions. Woolfson *et al.* state that these findings indicate that CH–π interactions which are traditionally considered as weak non-covalent interactions can considerably impact protein folding and stability and could be particularly important in the design and optimisation of miniproteins.

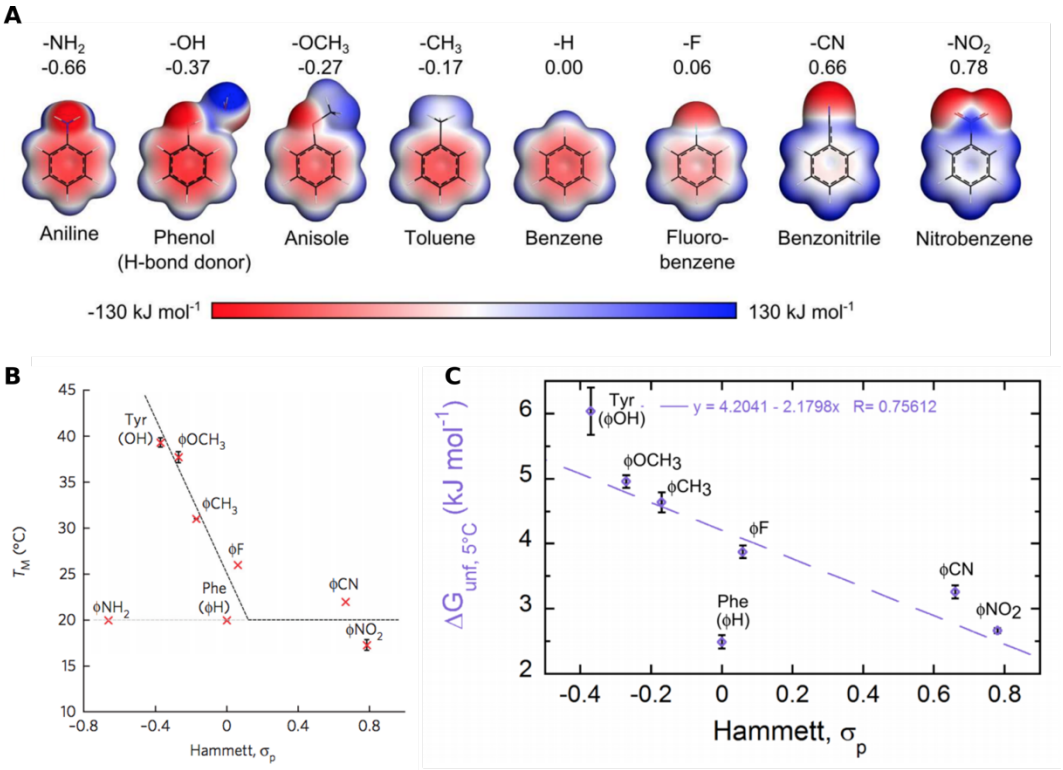


Figure 1-18: Folding and stability of PPα and mutants. (A) Electrostatic surface potential for para-substituted benzene side chains. Scale: ≤ -130 kJ mol-1 (electropositive, blue) to ≥ 130 kJ mol-1 (electronegative, red). (B) Plot of T_M against the Hammett σ_p parameter for the *para*-substituted aromatic substituents. (C) Free energy of unfolding (ΔG_{unf}) at 5°C vs. Hammett σ_p parameter for *para*-substituents of the PPα mutant series. Figure modified from reference [153].

1.5 Scope of thesis

Miniproteins reduce the complexity of the protein folding problem and allow us to probe contributions to protein folding and stability. Further, they provide stable, well-folded structures that can be used as modular building blocks in the design and engineering of larger and more-complex assemblies with the potential of accessing the *dark matter of protein space*. Recently, a fragment-based approach has been used to design the PPα miniprotein with the topology polyproline-II helix:loop: α helix. There is much to be understood about the sequence-to-structure and sequence-to-stability relationships in PPα. In this thesis, PPα has been completely redesigned, moving away from a structure based on natural sequence towards a *de novo* framework with enhanced thermal stability. General and intimate side chain-side chain interactions are examined revealing sequence-to-structure relationships for the fold. Following iterative rounds of design, two new topologies are explored based on the PPα fold. Finally, further strategies to improve PPα are described, specifically chain elongation and cyclisation.

Chapter 2 Materials and methods

2.1 General considerations

Fmoc protected amino acids, *N*,*N*-dimethylformamide (DMF) and 6-Chloro-1-hydroxybenzotriazole (HOBt-Cl) were purchased from AGTC Bioproducts, (Hessle, UK). H-Rink Amide-Chemmatrix® resin was purchased from PACS BioMatrix Inc. (Sain-Jean-sur Richelieu, Canada). Hexafluorophosphate azabenzotriazole tetramethyl uronium (HATU) was purchased from Carbosynth Ltd. (Compton, UK) and morpholine from Merck Millipore (Burlington, USA). All other chemicals were purchased from Sigma Aldrich (Gillingham, UK) or Fisher Scientific (Loughborough, UK), and used without further purification. Solution phase biophysical characterization was undertaken in phosphate-buffered saline (PBS, 8.2 mM sodium phosphate, 1.8 mM potassium phosphate, 2.7 mM potassium chloride, 137 mM sodium chloride, pH 7.4) in ultrapure water from a Synergy® UV water purification system (Merck Millipore).

2.2 Peptide synthesis and purification

2.2.1 Automated solid-phase peptide synthesis

Peptides were synthesized *via* standard Fmoc solid-phase peptide synthesis (SSPS) methods on H-Ring Amide-Chemmatrix® resin on a 0.1 mmol scale using a CEM (Buckingham, UK) Liberty Blue microwave-assisted automated peptide synthesiser with inline UV monitoring. Fmoc-protected amin acids were used at 0.2 M in DMF.

Fmoc deprotection was performed with 20% (v/v) morpholine in peptide-grade DMF, with the exception of sequences prone to aspartimide formation where 5% (v/v) formic acid was added to the deprotection solution. Peptides were synthesised in two stages. Stage one: single couplings under microwave conditions using *N*,*N*-Diisopropylcarbodiimide (DIC)/HOBt-Cl activation (0.5 M, DMF/ 1 M, DMF) for the α-helix component (i.e. for oPPα from residue 34 (Tyr) to residue 14 (Pro), peptide synthesized from *C* to *N* terminus). Stage two: double couplings under non-microwave conditions using HATU/ *N*,*N*-Diisopropylethylamine (DIPEA) activation (0.45 M, DMF/ 0.6 M, DMF) for the loop and polyproline II helix (i.e. for oPPα from residue 13 (Thr) to residue 1 (Pro)). Microwave conditions for this region led to peptide degradation during synthesis.

Following automated synthesis, the peptides were washed with DMF (4 x 5 mL), and *N*-terminally acetylated using an excess of acetic anhydride (0.25 mL), and pyridine (0.30 mL) in DMF (5 mL), with slow inversion for 20 min at room temperature (rt). The acetylated peptides were washed with DCM (4 x 5 mL) then cleaved from the resin and side chain deprotected using trifluoroacetic acid (TFA) / triisopropylsilane (TIPS) / H_2O (95 : 2.5 : 2.5 , 15 mL) for 2 h. Following cleavage, the TFA solution was reduced in volume to *ca.* 5 mL under a positive flow of nitrogen. Cold diethyl ether (40 mL) was then added to precipitate the peptide, which was isolated by refrigerated centrifugation

at 3 krpm, re-dissolved in water / acetonitrile (50 : 50, 10 mL) and lyophilised to yield the crude peptide as a white solid (**Figure 2-1**).

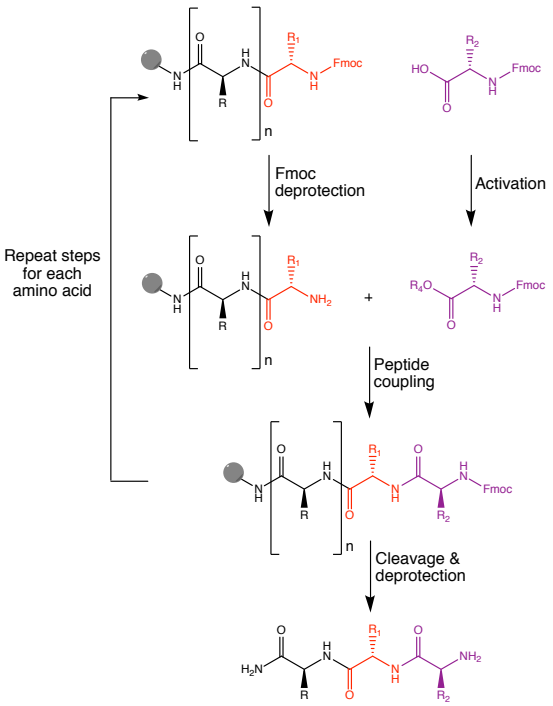


Figure 2-1: Overview of SPPS showing Fmoc deprotection, activation, peptide coupling, and cleavage and deprotection.

2.2.2 Peptide purification

Crude peptides were purified by reverse-phase high performance liquid chromatography (HPLC) on a Phenomenex Luna C18 column (5 µm particle size; 100 Å pore size; 150 × 10 mm). A gradient of water (0.1 % TFA, buffer A) and acetonitrile (0.1 % TFA, buffer B) between 20 and 80 % or 20 and 60% buffer B over 40 min at a flow rate of 3 mL min^{-1} with absorbance recorded at 220 and 280 nm was typically used. Pure fractions were identified by MALDI-TOF (matrix-assisted laser desorption/ionisation time-of-flight) mass spectrometry and analytical HPLC (*vide infra*), and were combined and lyophilised.

2.2.3 MALDI-TOF mass spectrometry

MALDI-TOF mass spectrometry was performed on a Bruker (Coventry, UK) UltrafleXtreme II MALDI-TOF mass spectrometer, operating in positive-ion reflector mode. Peptides were co-crystallised on a ground-steel target plate using 2,5-dihydroxybenzoic acid (DHB) or α-cyano-4-hydroxycinnamic acid (α-CHA) as the matrix. Representative mass spectra for each peptide are shown in Section 8.1. Theoretical masses are quoted as the average masses and were calculated using Peptide Synthetics' online peptide mass calculator (http://www.peptidesynthetics.co.uk/tools/, Peptide Protein Research Ltd.).

2.2.4 Analytical HPLC

Peptide purity was confirmed by reverse-phase analytical HPLC on a JASCO 2000 series HPLC system using a Phenomenex® Kinetex C18 (5 μM particle size; 100 x 4.5 mm) column, monitoring at 220 and 280 nm wavelengths Typically, a gradient of water (0.1 % TFA, buffer A) and acetonitrile (0.1 % TFA, buffer B) between 20 and 80 % buffer B over 20 min was used. Representative analytical HPLC traces for each peptide is shown in Section 8.1.

2.2.5 Peptide concentration determination

Pure peptides were dissolved in ultrapure deionised water and concentrations were determined on a Thermo Scientific (Waltham, USA) Nanodrop 2000 UV / visible spectrophotometer. Concentrations were determined by absorbance of Tyr at 280 nm using the Beer-Lambert Law (**Equation 2-1**). ε_{280}(Tyr) = 1280 mol^{-1} dm^3cm^{-1}. The extinction coefficient of oPPα-Phe and PPII was determined using a literature protocol based on the absorbance at 214 nm.[169] Pure peptides and stocks were stored at -20 °C and thawed before use.

$$A = \varepsilon . c . l$$

Equation 2-1: The Beer-Lambert Law. A is absorbance, ε is the extinction coefficient (mol^{-1} dm^3cm^{-1}), c is the concentration (mol dm^{-3}) and l is the pathlength (cm).

2.3 Solution phase biophysical characterisation

2.3.1 Circular dichroism spectroscopy

2.3.1.1 Introduction to circular dichroism

Circular Dichroism (CD) measures the difference in absorbance of left and right circularly polarised light by a chiral molecule (*e.g.* peptides).[170] Circularly polarised light is the result of two linearly polarised light waveforms that are perpendicular to one another and phase shifted by a quarter wavelength. CD spectroscopy can be used to determine the secondary structure present in peptides and proteins due to the distinctive spectra produced in the far UV (260 - 190 nm). Peptides produce two key CD signals as a result of the amide bond at *ca.* 220 and 190 nm. These signals arise from electronic transitions from the n to π* (nπ*) and π to π* (ππ*) orbitals respectively. These spectral features are characteristic of random coiled peptides. The periodic alignment of amide bonds in the peptide backbone within regions of secondary structure modify these basic electronic transitions, giving rise to information about the conformation of the peptide.[171]

α-Helical secondary structure give a characteristic "double dip" spectrum with minima at 208 and 222 nm and a maximum at 190 nm. β-Sheets have a minimum at *ca.* 215 nm (nπ*) and a maximum at *ca.* 195 nm (ππ*) and polyproline II helices have a minimum at *ca.* 197 nm and a maximum at 220 nm (**Figure 2-2**).[172]

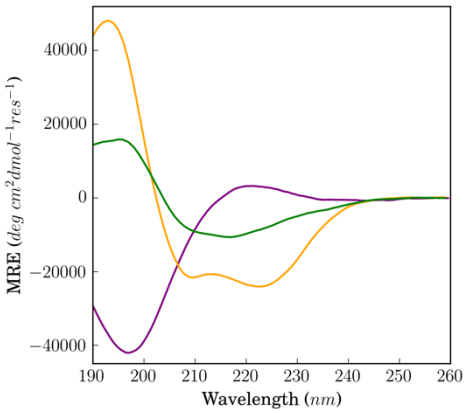


Figure 2-2: CD spectra of protein secondary structure. α-Helical secondary structure (yellow, myoglobin), β-sheet secondary structure (green, β-Lactoglobulin) and polyproline-II helix (purple, bovine collagen type II). Spectra collected from Protein Circular Dichroism Data Bank.[173]

Examination of CD spectra in the near UV range of 250 to 340 nm can provide useful information on the environment and conformation of aromatic side chains and consequently the structure of the peptide or protein. This is because the ππ* transition on the aromatic rings of phenylalanine, tyrosine and tryptophan are responsive to their local surroundings.[174]

2.3.1.2 Circular dichroism experimental details

CD spectra were collected using a JASCO 815 or JASCO 810 spectropolarimeter fitted with a Peltier temperature controller. Peptide samples were prepared in PBS at pH 7.4. CD measurements were carried out in 1 mm (50 – 200 μM total peptide concentration, 250 μL total sample volume), 5 mm (10 μM total peptide concentration, 1250 μL total sample volume) or 10 mm (near UV measurements, 100 μM total peptide concentration, 2500 μL total sample volume) pathlength cuvettes (Starna Scientific; Ilford, UK). For peptide concentrations greater than 500 μM a sandwich cell cuvette with pathlength 0.1 mm was used (100 μL total sample volume). CD spectra were baseline corrected and recorded as the average of five scans from 260 – 190 nm at 5 °C using a scanning speed of 100 nm min^{-1}, a bandwidth of 1 nm, a 1 nm step size and a 1 s response time. Thermal denaturation curves were obtained from 5 to 95 °C and back to 5 °C (temperature slope = 40 °C hr^{-1}) by monitoring the absorbance at 222 nm (1 nm bandwidth) at 1 °C intervals with 16 s delay and 16 s response times. For near UV measurements 5 °C scans were recorded from 250 – 340 nm and denaturation curves were obtained by monitoring at 276 nm. The midpoint of the denaturation curve (T_M) was determined by taking the maximum value from the first derivative of the thermal transition.

Data was buffer subtracted and then spectra were converted from ellipticities (mdeg) to mean residue ellipticities (MRE deg cm^2 $dmol^{-1}$ res^{-1}) by normalising for peptide concentration, number of peptide bonds and the cuvette pathlength as demonstrated in **Equation 2-2.** Representative spectra and thermal denaturation measurements for all peptides are shown in Section 8.2.

$$[MRE] = \frac{(\theta . 100)}{c . l . b}$$

Equation 2-2: Conversion of mdeg (θ) to mean residue ellipticity. *c* is the concentration of the sample (mol dm^{-3}), *l* is the sample pathlength (cm) and *b* is the number of amide bonds in the sample.

The fraction of α helix (%) for each peptide was calculated using **Equation 2-3**:

$$Fraction\ helix\ (\%) = 100\left(\frac{[MRE\ ]_{222} - [MRE]_{coil}}{-42\ 000\left(1 - \frac{3}{n}\right) - [MRE\ ]_{coil}}\right)$$

Equation 2-3: Percentage fraction helix at 222 nm. Where $[MRE]_{coil} = 640 - 45T = 415$ deg cm^2 $dmol^{-1}$ res^{-1} at 5°C. T is the temperature (°C); n is the number of bonds including the *C*-terminal amide.[175-176]

2.3.2 Analytical ultracentrifugation

2.3.2.1 Introduction to analytical ultracentrifugation

Analytical ultracentrifugation experiments follow the sedimentation of macromolecules over time when a centrifugal force is applied. Two types of experiments are conventionally carried out: sedimentation velocity (SV) and sedimentation equilibrium (SE). Each experiment gives information on the size of the macromolecule, while each also providing additional information.[177]

2.3.2.2 Sedimentation velocity

In SV experiments samples are spun at high rotor speeds (typically >40 000 rpm) leading to the depletion of peptide away from the centre of the rotor creating a pellet at the bottom of the cell. Data is collected across the radial distance of the cell following the sedimentation of the peptide over time. The data is fitted to the Lamm equations in a continuous c(s) distribution model returning a sedimentation coefficient profile.[178] This allows for determination of the sedimentation coefficient (Svedberg units, $S = 10^{-13}$ s) and the frictional coefficient (f/f_0) of the peptide, both of which relate to the mass of the peptide being studied. f/f_0 gives information about the shape of the peptide, a theoretical value of 1 would imply a totally spherical peptide. An advantage of the SV experiment is information on the homogeneity of the sample is obtained; multiple peaks are observed in the sedimentation coefficient profile if oligomers or aggregates of the peptide are present (**Figure 2-3**).

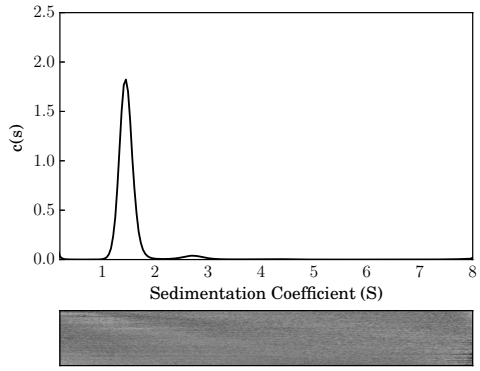


Figure 2-3: Representative SV data. Top: SV sedimentation coefficient distribution after fitting to continuous c(s) distribution. Sharp peak demonstrates homogenous sample. Bottom: Residuals for the above fit shown as greyscale, showing the difference between the raw and fitted data in bitmap format.

2.3.2.3 Sedimentation equilibrium

In SE experiments samples are spun at lower speeds and consequently the sedimentation of sample is balanced by its desire to diffuse back up the cell due to the creation of a concentration gradient. An equilibrium is established when no change in the concentration distribution of any component is detectable and when the opposing forces are balanced.[179] When this is achieved, scans are taken over the radial distance of the sample at a range of different speeds. Unlike for SV, no distinct boundaries are observed; instead a smooth gradient is seen. The curves are fitted to an exponential equation, and the molecular weight of the peptide is returned. If, at the concentration of the peptide, multiple species are sampled then more complex models for associating system can be fitted (e.g. monomer-dimer equilibrium) and a K_D can be determined (**Figure 2-4**).

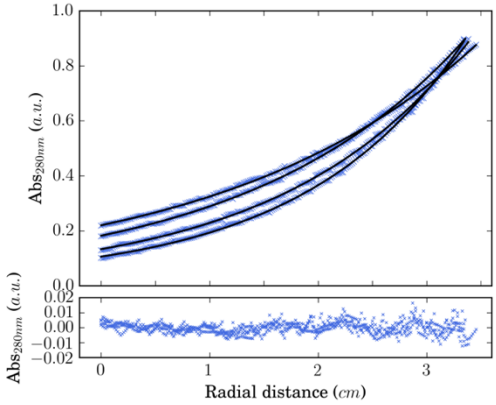


Figure 2-4: Representative SE data and fit. Top: data measured at four centrifugal speeds (crosses) and fitted to single ideal species model curves (lines). Bottom: Residuals for the above fit.

2.3.2.4 Sedimentation Velocity Method

SV AUC experiments were conducted at 20 °C in either a Beckman XL-A or XL-I ultracentrifuge equipped with an An-60 or An-50 Ti rotor. 410 µL of solution at 100 µM peptide concentration in PBS were loaded in the sample channel, and the reference channel was loaded with 420 µL of PBS buffer. Samples were centrifuged at 50 krpm, with absorbance scans taken across the cell at a radial

range of 5.8 – 7.3 cm at 5 min intervals to a total of 120 scans. Data was fitted to a continuous c(s) distribution model using SEDFIT (http://www.analyticalultracentrifugation.com) at a 95 % confidence level.[180] The baseline, meniscus, frictional coefficient (f/f_0), and systematic time-invariant and radial-invariant noise were fitted. Residuals for the c(s) distribution fit are shown as a bitmap in which the greyscale shade indicates the difference between fit and raw data.

2.3.2.5 Sedimentation equilibrium method

SE AUC experiments were conducted at 20 °C in either a Beckman XL-A or XL-I ultracentrifuge equipped with an An-60 Ti rotor and cells equipped with a 2-channel aluminium centrepiece and sapphire windows. Peptide solutions were made up in PBS at 80 – 130 µM (110 µl) peptide concentration. The reference channel contained 120 µl of PBS buffer. Samples were centrifuged from 44 – 60 krpm in increments of 4 krpm. Absorbance was measured at 280 nm across the cell at a radial distance of 5.8 – 7.3 cm at each speed after 8 h. To check the sample had reached equilibrium absorbance was measured again after a further hour to before moving onto the next speed. Data was fitted to a single ideal species model with Ultrascan II (http://ultrascan2.uthscsa.edu/).[181] The partial specific volume ($\bar{v}$) for peptides and the solvent density (1.0054 g mL^{-1}) were calculated using Sednterp (http://www.jphilo.mailway. com/download.html). 95% confidence limits were calculated using Monte Carlo analysis of the obtained fits. All SE data, fitted curves and residuals are shown in Section 8.3.

2.4 Nuclear magnetic resonance spectroscopy

Peptide was prepared at 1 mM concentration in PBS. The pH was adjusted to pH 7.4 with 10 mM NaOH and the sample freeze-dried before being reconstituted in the appropriate volume of D_2O (10%) in H_2O (90%). The pH and concentration were confirmed.

NMR data were acquired at 278 K on a Bruker Ascend spectrometer operating at 700 MHz equipped with a 1.7 mm micro-cryoprobe (BrisSynBio, University of Bristol). The peptides were assigned using standard 2D homonuclear spectra: TOCSY (60 ms mixing time) and NOESY (100 and 250 ms mixing times). Both were acquired with spectral widths of 9,375 Hz, 4,096 complex points in f2 and 1,024 complex points in f1. To help with backbone and side-chain assignment, natural abundance ^{15}N (96 (t1) × 1792 (t2) complex points) and ^{13}C (124 (t1) × 2,048 (t2) complex points) HSQC spectra were also acquired.

NMR data were processed by Dr Chris Williams (Crump Group, Bristol University) with NMRPipe and qMDD[38, 68, 182]. Peak picking and assignment were carried out in CCPNMR Analysis 2.4.1.[183] NOE assignment and structure calculation were carried out with ARIA 2.3.1[184] and CNS v 1.2[185]. The final structures were water refined using the standard ARIA protocol. Dihedral restraints for the α-helix were generated using DANGLE[186] and validated with the NOE spectra before inclusion into the structure calculation. Three bond J couplings were extracted from a high resolution COSY and supplemented into the structure calculation during the final stages of refinement. The final refined ensemble was composed of 20 structures with the lowest energy and no violations (>5°) and

validated using PSVS v1.5.[187] All peptide structure images were rendered with Pymol (http://www.pymol.org).

2.5 X-ray crystallography: peptide crystallisation trial methods

Freeze dried peptides were resuspended in deionised water to a concentration of 15 mg mL^{-1} for sitting-drop vapour-diffusion crystallisation trials using standard commercial screens (JCSG, Structure Screen 1+2 Morpheus™, ProPlex and PACT Premier™; Molecular Dimensions (Newmarket, UK)) at 20 °C and 4 °C with 0.3 μL of the screening solution.

2.6 Computational design and analysis

ISAMBARD (Intelligent System for Analysis, Model Building And Rational Design) is an open-source Python package with a suite of tools for biomolecular structure analysis, protein design, model building and evaluation.[142]

2.6.1 Loop finder

LoopFinder, a protein design tool within ISAMBARD, was used to extract loops from the Protein Data Bank (PDB) that fit between a pair of entering and exiting residues in the αPP topology. Loops were filtered based on the following criteria:

- Loops from crystal structures with ≤ 2 Å resolution
- ≤ 3 Å fit RMSD between loop and αPP model backbone coordinates
- 3 – 6 residue loop lengths
- Loops that link α and β secondary structure
- Loops containing cysteine were omitted

2.7 Bioinformatics

2.7.1 CH–π interaction identification

CH–π interactions were identified within the ISAMBARD framework.[142] CH–π interactions were identified using parameters based on those used to find CH–π interactions involving carbohydrates in protein crystal structures[168] and adapted to account for any CH protons interacting with an amino acid aromatic ring. This adaptation provides an update to a previous analysis of CH–π interactions.[167] CH–π interactions were determined between all amino acid CH bonds and the aromatic ring of Tyr. Interactions were classed as CH–π positive if the following conditions were met: CH–π distance (between the CH proton to centre of the aromatic ring) ≤3.5 Å; CH–π angle (between the vector of the CH bond and the normal to the plane of the aromatic ring) ≤ 55°; C projection distance (between the projection of the CH carbon to the plane of the aromatic ring and the centre of the ring) ≤ 2.0 Å for 6-membered Tyr (**Figure 2-5**).

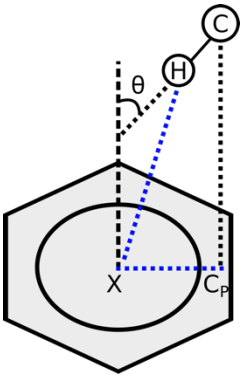


Figure 2-5: Parameters defining a CH–π interaction. CH–π distance (between the CH proton to centre of the aromatic ring, i.e. H–X, blue) ≤ 3.5 Å. CH–π angle (between the vector of the C–H bond and the normal to the plane of the aromatic ring) θ ≤ 55°. C-projection distance (between the projection of the CH carbon to the plane of the aromatic ring and the centre of the ring, i.e. X – C_p, blue) ≤ 2.0 Å for 6-membered ring (Tyr).

2.7.2 Solvent accessibility analysis

Solvent accessible surface area (SASA) of both oPPα and PPα were calculated within the ISAMBARD framework using the programme NACCESS which uses the Lee and Richards method described in Section 1.4.6.[188-189] SASA was calculated for the oPPα and PPα complex, loop region was ignored. SASA was then calculated for the individual α helix and polyproline-II helix secondary structure components. SASA of the complex was then taken as a proportion of the total SASA of the individual components. This was done for all 20 structures in the NMR ensemble and then averaged.

2.8 Molecular dynamics simulation

2.8.1 Introduction to molecular dynamics

Molecular dynamics (MD) is one of the principle tools for modelling proteins and probing their stability, folding and molecular recognition of other biologically important molecules. A molecule is described as a series of charged points (atoms) linked by springs (bonds). The potential energy function allows us to calculate the force experienced by any atom given the position of the other atoms in the system. How these forces affect the trajectories of atoms and molecules are determined by solving Newton's equations of motion.

2.8.2 Molecular dynamics simulation setup

Models were constructed using Pymol and set up for MD simulation using the Gromacs 4.6.7 suite of tools.[190-191] Hydrogen atoms were added consistent with pH 7.4 using pdb3gmx, and the TIP3P water model and Amber99sb-ILDN force field were chosen.[192] A cubic periodic boundary box was set up with dimensions 2 nm greater than the longest dimension of the model with editconf. This box was filled with water molecules using genbox and 137 mM NaCl using genion. The system was energy minimized and position-restrained MD run for 200 ps as an NPT (normal pressure and temperature) ensemble at 278 K, 1 Bar using the Verlet cut-off scheme and under PME (Particle

Mesh Ewald) boundary conditions as an initial relaxation and equilibration step. The restraints were then removed and a further 10 or 100 ns of MD was performed using 1 GPU and 6 cores of an X86 workstation. Structures were saved every 10 ps. Post-simulation trajectories were processed using trjconv to remove periodic boundary condition effects and solvent and inspected using VMD.[193] All atom and backbone RMSDs and RMSFs were calculated using Gromacs.

Chapter 3 Stabilising the PPα miniprotein by rational design

The work described in this chapter was designed by the author of this thesis (KPG), Dr Emily Baker and Prof. Dek Woolfson. Bristol Chemical Synthesis CDT rotation student Frank Zieleniewski synthesised and preliminary characterised oPPα-E↔K; full characterisation was followed up by KPG. Data collection and assignment of the oPPα NMR structure was done by Dr Chris Williams (Prof. Matthew Crump's Group) and critical analysis of the structure was done by KPG.

3.1 Chapter introduction

Previously, the fragment-based design and characterization of a 34-residue miniprotein, PPα, has been described.[153] PPα comprises a polyproline-II helix, loop and α helix that are adapted from two natural proteins: an antigenic bacterial surface adhesin (AgI/II) from *Streptococcus mutans*[166] and the bovine pancreatic polypeptide hormone.[29] PPα is a monomer in aqueous solution and reversibly unfolds with a midpoint unfolding temperature (T_M) of 39 ˚C. A Nuclear Magnetic Resonance (NMR) spectroscopy structure reveals that PPα is stabilized by the interdigitation of Pro and Tyr residues between the two helices and numerous CH–π interactions.

We envisage using PPα as a building block to build larger and more-complex assemblies with applications in protein engineering and synthetic biology.[194-195] PPα could provide a new component to the toolkit of peptides already developed to help make the engineering of biological systems easier and more accessible.[151] Such building blocks should be predictable with respect to sequence, structure, and stability in order to be used reliably in a variety of systems while maintaining structural and functional integrity.[32] As such, *de novo* designed building blocks of reduced complexity where the role of every residue in the sequence is understood provide advantages over more-complex sequences borrowed from nature.

Arguably one of the best understood protein folding units in terms of sequence-to-structure relationships is the α-helical coiled coil where two or more amphipathic α-helices supercoil around a central hydrophobic core.[132, 134, 196] α-Helical coiled coils with a range of oligomeric states have been *de novo* designed and fully characterised so that they can be used easily and reliably in different contexts.[32, 145] In this chapter, we apply a similar strategy to PPα to garner a better understanding of the PPα topology. A rational redesign of PPα is described, that moves away from a fragment-based design towards a *de novo* framework. The extent to which the stability of PPα can be enhanced through this process is explored. An optimised and fully characterised PPα framework with significantly enhanced thermal stability compared to the first-generation design is delivered.

3.2 Rational redesign of PPα

One of the key defining features of the PPα topology is the interdigitation of Pro residues on the polyproline-II helix with Tyr residues on the α helix. Furthermore, these residues are conserved in many natural examples of polyproline-II:α helix interfaces, such as the AgI/II bacterial surface adhesin and the pancreatic peptide family.[29, 55, 166] Baker *et al.* previously mutated the residue Tyr to a series of other proteinogenic and non-proteinogenic amino acids and found that Tyr gave the most thermally stable PPα fold.[153] Therefore, the Pro-Tyr interface was maintained in our PPα redesign with Pro placed at position ***1*** and Tyr placed at the ***d*** position of the polyproline-II and α-helical repeats respectively (**Figure 3-1**). Leu at the ***a*** position is also conserved in many related natural peptides and forms part of the diamond shaped hole into which Pro projects. Therefore, Leu at ***a*** was initially preserved in the new PPα designs. However, the mutability of this position is discussed in more detail in Chapter 4.

The redesign of PPα was inspired by a number of α-helical coiled-coil design principles. At the heart of all coiled coils, the burial of hydrophobic surfaces in intimate side-chain packing termed knobs-into-holes (KIH) interactions[197] provides the driving force for coiled-coil assembly and specification.[113, 132, 196,198] Alongside this, additional electrostatic interactions can provide further stabilisation. Charged residues at the ***e*** and ***g*** positions that flank the hydrophobic core can form interhelical Coulombic interactions in natural systems.[199-200] These have been used extensively to guide and stabilise *de novo* coiled-coil designs.[132-133, 201-202] Lys-Glu pairs have been used to control either homomeric or heteromeric assembly in peptides with otherwise identical ***a*** and ***d*** core interfaces (**Figure 3-1A**).[203]

By analogy, PPα might be stabilized by optimising similar interactions between the ***2*** and ***3*** positions of the polyproline-II helix and the ***e*** and ***g*** positions of the α helix (**Figure 3-1B**). It is noted that the direction of the C_α-C_β bond vectors of the side chains of the ***2*** and ***3*** positions in the polyproline-II helix point away from the α helix. This may mean that formal salt bridges are not made between the ***2***/***e***, and ***3***/***g*** positions due to the distance between the opposite charge groups being outside of the range for salt bridging. Nonetheless, charged residues may still help stabilise and promote folding of PPα through electrostatic steering where long-range electrostatic forces specifically increase the rate of helix association without affecting the rate of dissociation.[204-205]

There are two extreme possibilities for pairing charged residues in PPα: a Lys-based polyproline-II helix plus a Glu-based α helix or *vice versa* (**Figure 3-1C&D**). To determine the optimal combination, both mutants (oPPα and oPPα-E↔K) were considered (**Table 3-1**). While there are other possibilities for pairing charged residues with a combination of both Lys and Glu on the same helix we only investigated the two extreme and simplest cases.[206-208] More elaborate charge pattern designs might be useful for creating orthogonal sets of PPα peptides that can be used together without any cross-interaction.

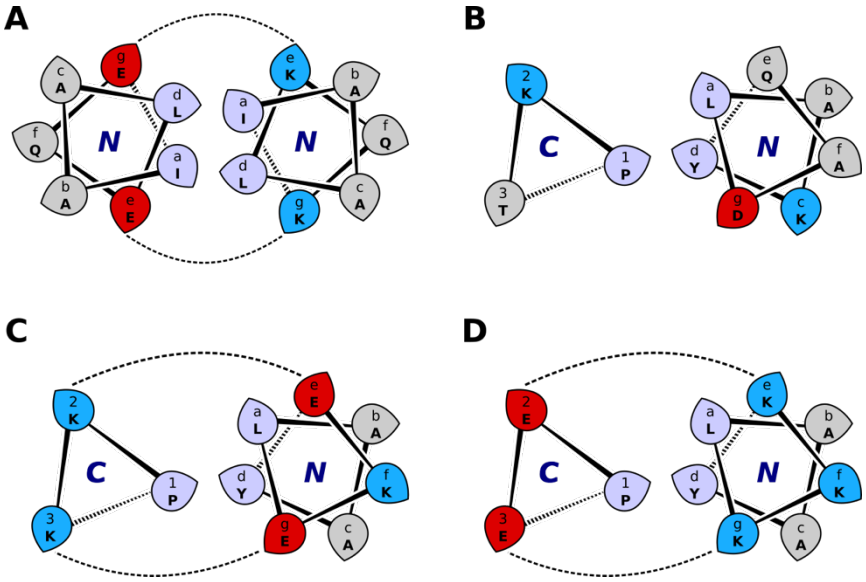


Figure 3-1: Helical wheel representations of α-helical coiled coils and the PPα fold. Leaf tips point along the direction of the C_α-C_β bond vectors. Lower case letters on the α helix and numbers on the polyproline-II helix (left) indicate the helical register. *C* and *N* refer to the termini of the helices nearest the viewer. Interhelix Coulombic interactions (dashed lines) (**A**) Sequence of a heteromeric dimer highlighting interhelix Coulombic interactions between **g** and **e** positions.[203] (**B**) Sequence of PPα with no interhelix Coulombic interactions. (**C**) Sequence of oPPα highlighting potential Coulombic or long-range electrostatic interactions between the **2** and **3** positions of the PPII helix and the **e** and **g** positions of the α helix. (**D**) Sequence of oPPα-E↔K highlighting alternative extreme arrangement of charged interactions.

With the initial α helix foreground ***e***, ***a***, ***d*** and ***g*** positions of the starting sequences selected, focus turned to the remaining background positions (***b***, ***c*** and ***f*** positions). These residues were designed to contribute to the overall stability of the peptide by promoting α helicity, without promoting aggregation. Therefore, helix propensity was optimised for the solvent-exposed face of the α helix.

Helix propensity is a measure of how the side chains of a particular amino acid affects the helix-to-coil equilibrium; some residues have more of a preference for the α-helical conformation compared to others.[209-211] Many different helix propensity scales have been developed based on different strategies and sources including statistical surveys,[212-213] host-guest analysis[214] and molecular dynamics calculations.[215-216] One of the first helix propensity scales has been presented by Chou and Fasman and is based on the frequency of occurrence of amino acids in α helices in proteins.[217-218] Later, DeGrado *et al* have designed an α-helical dimer in equilibrium with a random coil monomer. All 20 natural amino acids were substituted for a solvent exposed guest site on the helix and a list of free energy difference values (ΔΔG) for each amino acid were calculated.[214] Perhaps the most comprehensive and widely used scale has been derived by Scholtz *et al.* based on experimental measurements of helix propensity in 11 systems including both proteins and peptides.[209] From this analysis, the helix propensities of the amino acids are as follows (kcal mol^{-1}): Ala 0, Leu 0.21, Arg

0.21, Met 0.24, Lys 0.26, Gln 0.39, Glu 0.40, Ile 0.41, Trp 0.49, Ser 0.50, Tyr 0.53, Phe 0.54, Val 0.61, His 0.61, Asn 0.65, Thr 0.66, Cys 0.68, Asp 0.69, and Gly 1.

The α helix is stabilised mainly by favourable enthalpic contributions[219] from the formation of backbone hydrogen bonds, whereas the random coil is favoured by conformational entropy.[220] The 1 kcal mol $^{-1}$ difference in helix propensity between Gly and Ala is the result of the reduced phi-psi space available to residues when the H in Gly is replaced by a methyl group in Ala.[221-222] The methyl group reduces the entropy of the backbone in the random coil however it does not lose side chain conformational entropy on folding to form an α helix.[215] All other residues experience an unfavourable change in side chain conformational entropy upon folding to form an α helix.[209, 215, 220]

Based on this helix propensity scale, mutations to Ala at ***b*** and ***c*** and Lys at ***f*** for both oPPα and oPPα-E↔K were made. While Leu has relatively high helical propensity it was avoided in the oPPα background positions to ensure the monomeric nature of PPα was preserved. Met was also avoided as it is susceptible to oxidation and so too was Arg as it can have undesirable side reactions during synthesis. Lys was placed at position ***f***, that is, the position spatially between ***b*** and ***c***, to prevent a two-residue hydrophobic Ala interface forming, which may cause oligomerisation of the topology.

The length of the new PPα designs were maintained at three α-helical heptads so that a direct comparison in stability could be made between the new designs and parent PPα. Also, to allow comparison of the new PPα design with parent PPα, the same loop region from PPα was used. Furthermore, loop design is notoriously difficult, unlike for α helices and β-strands, there are not well determined amino-acid preferences for forming well-ordered loops.[223]

Table 3-1: Sequences for PPα variants.

Peptide	Sequence and helical register
	efgabcd efgabcd efgabcd
parent PPα[153]	Ac- PPTKPTKP GDNAT PEKLAKY QADLAKY QKDLADY -NH_2
o-α	Ac- PEKLAAY EKELAAY EKELAAY -NH_2
oPPα	Ac- PPKKPKKP GDNAT PEKLAAY EKELAAY EKELAAY -NH_2
oPPα-E↔K	Ac- PPEEPEEP GDNAT PEKLAAY KKKLAAY KKKLAAY -NH_2
oPPα-Phe	Ac- PPKKPKKP GDNAT PEKLAAF EKELAAF EKELAAF -NH_2

The residues at the initial ***e-f-g*** positions of the first heptad of oPPα and oPPα-E↔K were different to those in the second and third heptad, P-E-K compared to E-K-E. This is because Pro and Glu were conserved as the *N* terminal residues in the α helix of related natural sequences and have been shown to be good helix capping residues.[224-225]

3.3 Biophysical characterisation

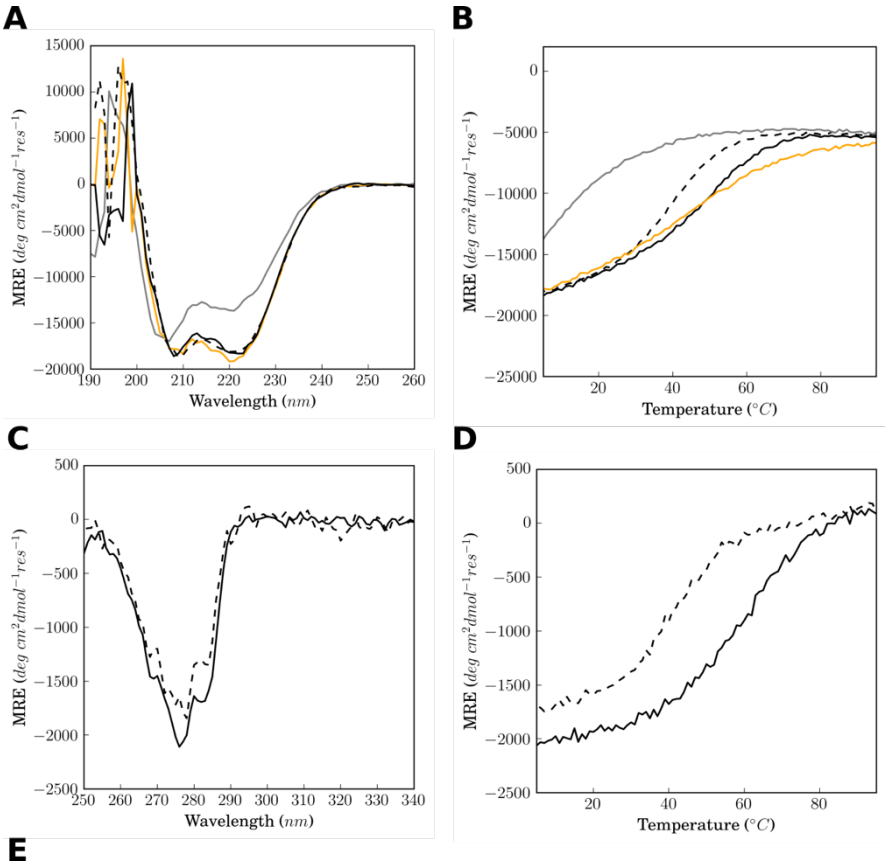
3.3.1 Circular dichroism spectroscopy

oPPα and oPPα-E↔K were synthesized by solid-phase peptide synthesis, purified by reverse-phase HPLC, and confirmed by MALDI-TOF mass spectrometry (see Section 8.1). The peptides were characterised by circular dichroism (CD) spectroscopy (**Figure 3-2**). Both peptides were folded with approximately 50% α-helical structure at 5 °C, similar to parent PPα. Thermal denaturation experiments of the far-UV CD signal at 222 nm, which reports on the secondary structure present, revealed that oPPα had a cooperative and fully reversible thermal unfolding transition with a T_M of 51 °C. This is a substantial increase of 12 ˚C over the T_M for PPα.

However, oPPα-E↔K was 10 °C less stable than oPPα with a T_M of 40 °C, which is still more stable than parent PPα. It is noted that oPPα-E↔K had a broader thermal unfolding transition compared to oPPα. The reduction in thermal stability can potentially be attributed to the large charge associated with the α helix; +6 compared to -2 for oPPα.

Consistent with PPα,[153] oPPα was considerably more stable than the constituent α helix (o-α). The α helix was 38 % folded and a T_M could not be extracted from the thermal melt curve. This is striking given that almost half the oPPα sequence is loop region or polyproline-II helix both of which do not contribute to the CD signal but are included in the spectra normalisation. This demonstrates the significant stabilising effect of buttressing the α helix against a polyproline-II helix. The folding and stability data is summarised in **Figure 3-2**.

Next oPPα was characterised by near-UV CD spectroscopy, which reports on the tertiary structure of the peptide through the environment of the aromatic side chains, in this case Tyr (**Figure 3-2C&D**). The thermal transition of oPPα monitored by near-UV CD spectroscopy gave unfolding and refolding curves coincident with the far-UV CD traces. While the degree of folding of both oPPα and PPα were similar at 5 °C in the far-UV region, the magnitude of the CD signal at 5 °C in the near-UV region was slightly greater for oPPα. This could indicate that the Tyr side chains of oPPα are slightly more rigid and less mobile than in PPα and suggests tighter packing between the two helices.[226]



Peptide name	$MRE_{222\ nm}$ / deg cm^2 $dmol^{-1}$ res^{-1}	% Helicity[175-176]	T_M / °C
PPα	-18319	48	39
o-α	-13746	38	n/a
oPPα	-18319	48	51 ± 0.8
oPPα-E↔K	-17810	47	40 ± 2.1

Figure 3-2: Folding and stability of oPPα and variants. (A) CD spectra recorded at 5 °C in the far-UV range for parent PPα, oPPα, oPPα-E↔K and o-α. MRE is mean residue ellipticity. (B) Temperature dependence of the far-UV CD signal monitored at 222 nm. (C) Aromatic CD spectra recorded at 5 °C in the near-UV range. (D) Temperature dependence of the near-UV CD signal monitored at 276 nm. Conditions: 100 μM peptide, PBS, pH 7.4. Colour key: PPα (black, dash), oPPα (black), oPPα- E↔K (orange), and o-α (grey). (E) Table summarising CD data on the folding and stability for oPPα and variants.

3.3.2 Analytical ultracentrifugation

Next, the constructs were investigated by analytical ultracentrifugation (AUC) to determine their molecular weights in solution. There are two standard AUC experiments usually used to determine molecular weight: sedimentation velocity (SV) and sedimentation equilibrium (SE). In SV experiments samples are spun at very high centrifugal forces and sedimentation is measured over time. SV experiments reveal information about the size and shape of the peptide as well as the homogeneity of the system. SV experiments could not be performed on PPα variants, however,

because rotational speeds > 60 krpm would be required for full peptide sedimentation and this exceeds the maximum speed limit of the analytical ultracentrifuge. Therefore, only SE experiments were performed where the sample is spun at lower speeds over longer times to allow an equilibrium to be set up between sedimentation and back diffusion. At equilibrium, information on the weight of the peptide can be extracted from the radial distribution of peptide concentration. For associating peptides, SE experiments can also provide association constants provided the dissociated and associated states are both accessed in the experimental data.

SE AUC experiments were carried out on both variants. Experimental data readily fitted to single-ideal species models returning molecular weights consistent with monomers (**Figure 3-3**). To ensure high confidence in the masses returned, residuals were plotted which indicate the difference between fit and raw data. 95% confidence limits were calculated using Monte Carlo analysis of the obtained fits.

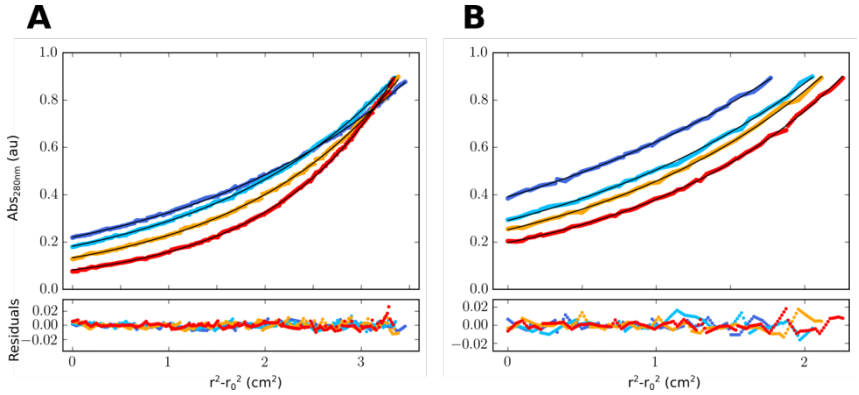


Figure 3-3: Sedimentation-equilibrium AUC data (top, dots) and fitted (lines) single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), and 60 (red) krpm for oPPα and oPPα-E↔K. Residuals (bottom) for the above fit. (**A**) oPPα ($\bar{v}$ = 0.748 cm^3 g^{-1}). The fit returned the following mass: oPPα, 3854 Da, (1.0 x monomer mass), 95% confidence limits = 3839 – 3869 Da. (**B**) oPPα-E↔K ($\bar{v}$ = 0.748 cm^3 g^{-1}). The fit returned the following mass: oPPα-E↔K, 3545 Da, (0.9 monomer mass), 95% confidence limits = 3511 – 3581 Da.

To probe and understand the structure of oPPα further, attempts were made to crystallise oPPα for structure determination by X-ray crystallography. Crystal trays were laid down as described in Section 1.5 using standard commercial screens at both 20 and 4 °C however no crystals have been observed to date.

3.4 Characterisation using NMR spectroscopy

As an alternative to X-ray crystallography, high-resolution NMR spectroscopy was used to determine the solution-phase structure of oPPα. Data for oPPα were acquired at both 600 and 700 MHz. As a result of the low-complexity of the oPPα sequence there was significant overlap in the signals in the 1D ^{1}H NMR spectrum (**Figure 3-4**). The peptide was assigned using standard 2D spectra: TOCSY, NOESY and HSQC. Previously assigned NMR spectra for PPα aided the

assignment of those for oPPα. Data were collected and the NMR structure solved by Dr Chris Williams, BrisSynBio, University of Bristol.

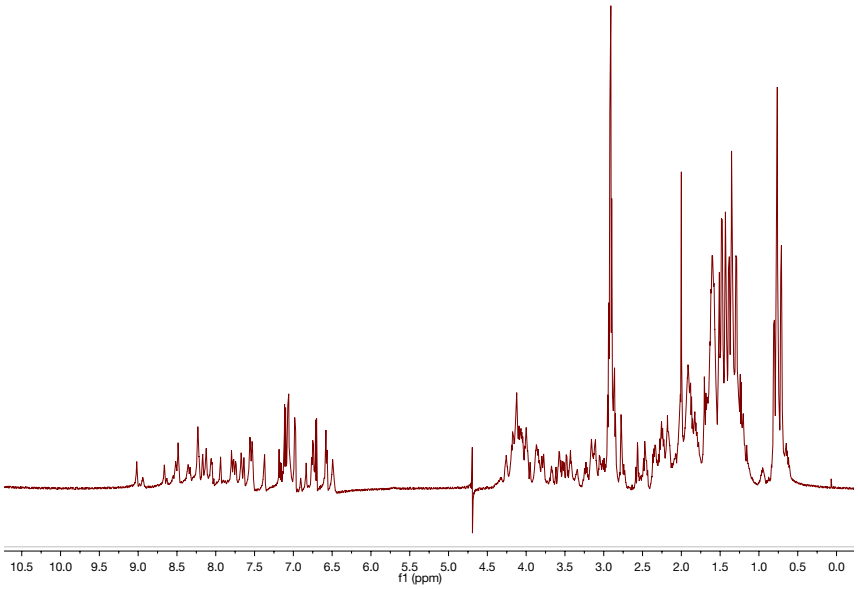


Figure 3-4: 1D ^{1}H NMR spectrum of oPPα acquired at 700 MHz showing the complexity of the system and overlap of peaks.

3.4.1 Sequential assignment approach

The sequential assignment approach consists of two stages.[227-229] Firstly, spin systems associated with particular amino acids are identified. This is achieved by running TOCSY (or COSY) experiments. The second stage is the assignment of each spin system to a particular residue in the amino-acid sequence. TOCSY experiments cannot be used to do this since there is no coupling across the amide bond. Instead, assignments are made from through-space nuclear Overhauser effects (NOEs) from the NOESY spectrum (**Figure 3-5**). Regular secondary structures (e.g. α helices and β strands) have regular and distinct NOE patterns.

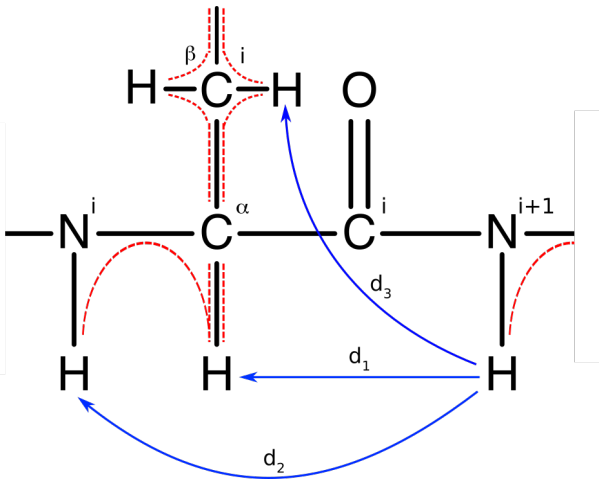


Figure 3-5: Polypeptide backbone showing the though-space distances (d1, d2, and d3) used for the sequential resonance assignment (blue arrows). Through bond spin-spin coupling connectivities within amino acid residues are shown as red lines.

3.4.2 [^{1}H-^{1}H]-TOCSY: Total correlation spectroscopy

Similar to a COSY experiment a TOCSY spectrum gives ^{1}H-^{1}H correlations due to scalar (through-bond) couplings. Unlike a COSY spectrum where coupling is limited to protons up to three bonds apart (^{4}J coupling constants are close to zero), a TOCSY experiment potentially correlates all H nuclei of the same spin system. There is no scalar coupling through the amide bond and so protons from different amino acids belong to different spins systems. Therefore, from the TOCSY spectrum, characteristic peak patterns for a spin systems can be assigned to a particular type of amino acid. The TOCSY spectrum for oPPα is shown in **Figure 3-6** highlighting the aliphatic side chain protons. Resonance assignments have been omitted for clarity.

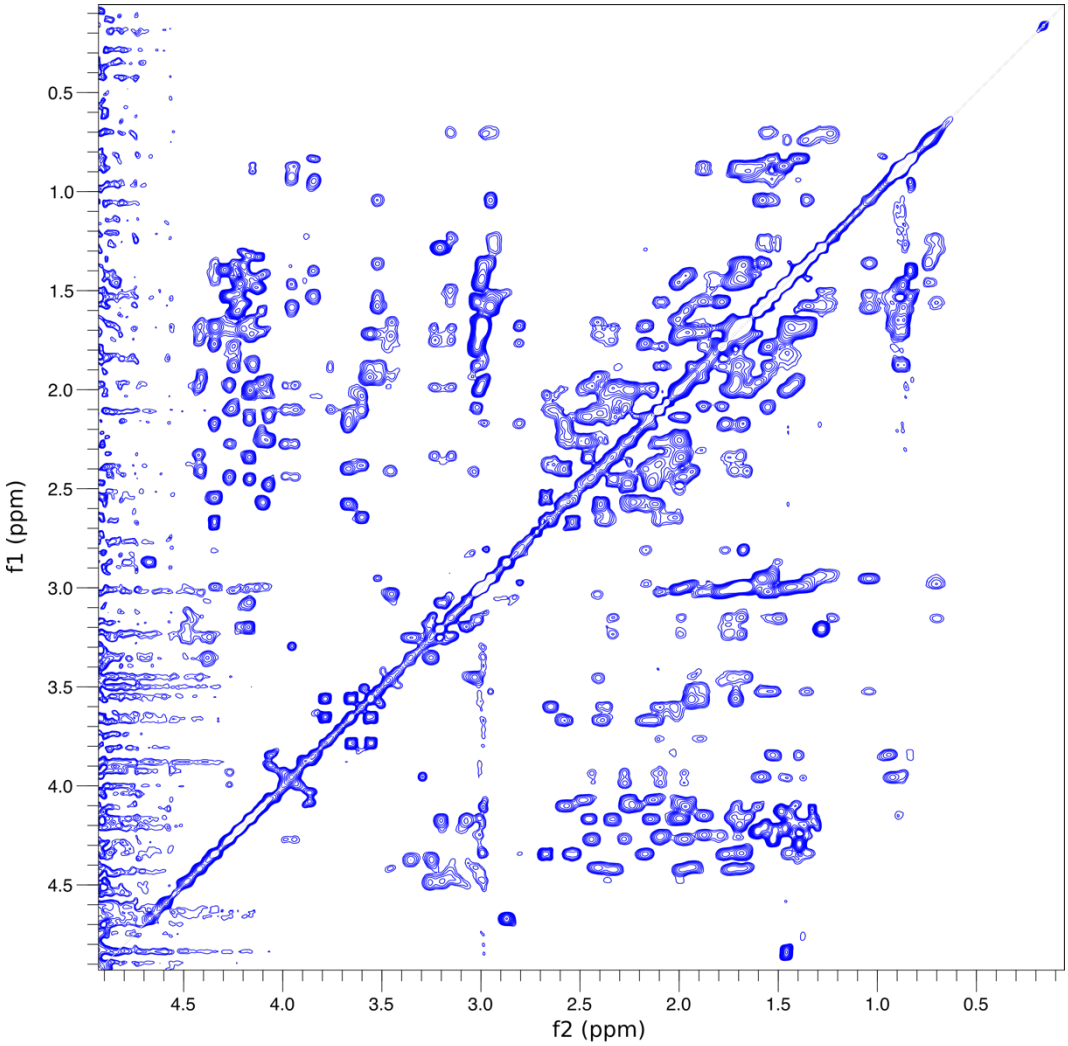


Figure 3-6: TOCSY spectrum of oPPα acquired at 600 MHz showing aliphatic side chain protons. Residual water band signal visible at 5 ppm on the f2 axis.

3.4.3 [^{1}H-^{1}H]-NOESY: Nuclear Overhauser effect spectroscopy

Cross peaks in a NOESY spectrum are the result of dipolar couplings resulting from interactions of spins *via* space and not through bonds as is the case for TOCSY experiments. The magnitude of these NOEs depends on the distance between interacting spins and therefore the NOESY spectrum contains all information about the spatial proximities of protons and consequently the 3D structure of the peptide.

The amide region of the NOESY spectrum of oPPα is shown in **Figure 3-7**. Sequential H^N_i-H^N_{i+1} NOEs observed in the spectrum indicated a well-defined α helix. To identity which amino acids these correlations belonged comparisons were made with the TOCSY spectrum.

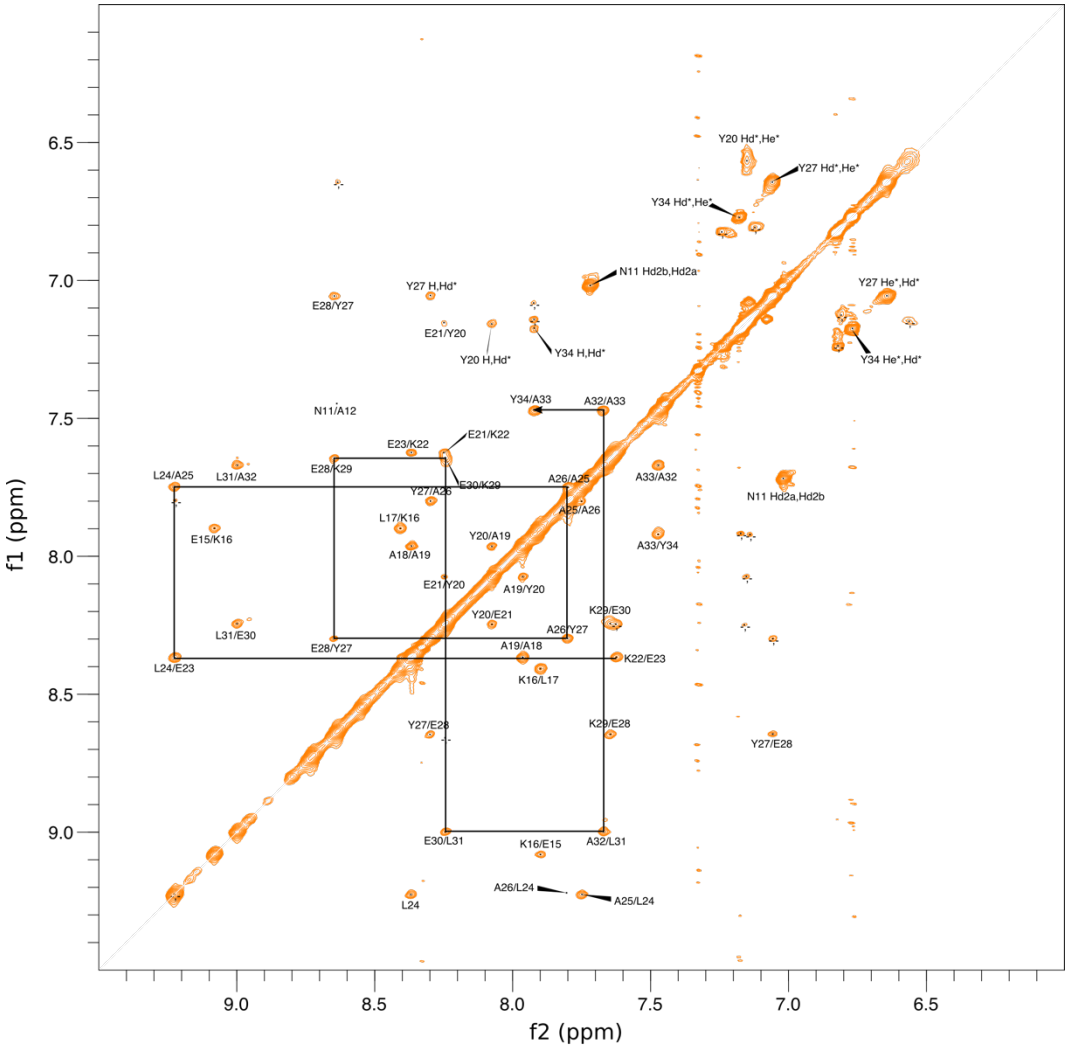


Figure 3-7: NOESY spectrum of oPPα acquired at 700 MHz (250 ms mixing time) in the amide/aromatic region. Strong sequential $H^N_i - H^N_{i+1}$connectivities in the helical segment highlighted between K22 to Y34 (black lines).

3.4.4 HSQC: Heteronuclear single quantum coherence spectroscopy

To help with backbone and side chain assignment, natural abundance ^{15}N and ^{13}C HSQC spectra were also acquired. The HSQC experiment shows one bond connectives, for example the ^{1}H-^{15}N HSQC shows peaks for all backbone amides and some side chain nitrogen atoms of Asn, Arg, Gln, His, Lys and Trp. Pro does not have an amide proton and so does not give cross peaks in the spectrum.

An overlay of the ^{1}H-^{15}N HSQC spectra for oPPα and PPα is shown in **Figure 3-8**. Like PPα, oPPα appeared to be well folded as determined by the good chemical shift dispersion. There were similarities between the two spectra. For example, peaks corresponding to the loop region (which is the same for both peptides), either overlaid or appeared in very similar regions of the spectrum (residues G9, D10, N11, A12, and T13). Cross peaks from E15 which is present in both peptides and corresponds to the start of the *N* terminus of the α helix also overlay well.

Unlike for PPα, clustering of peaks was observed for oPPα. This is a result of the repetitive and degenerate peptide sequence; there were a lot of the same residues in similar environments. This is particularly true for residues in the α-helical region, for example overlap is observed for residues A25 and A26, Y27 and E30, and L17, A18 and E23.

In oPPα, residues K4 and K7 were shifted downfield in f1 and upfield in f2 relative to residues K3 and K6. This is the result of the proximity of the K4 and K7 residues to the aromatic rings of Tyr27 and Tyr20, respectively.

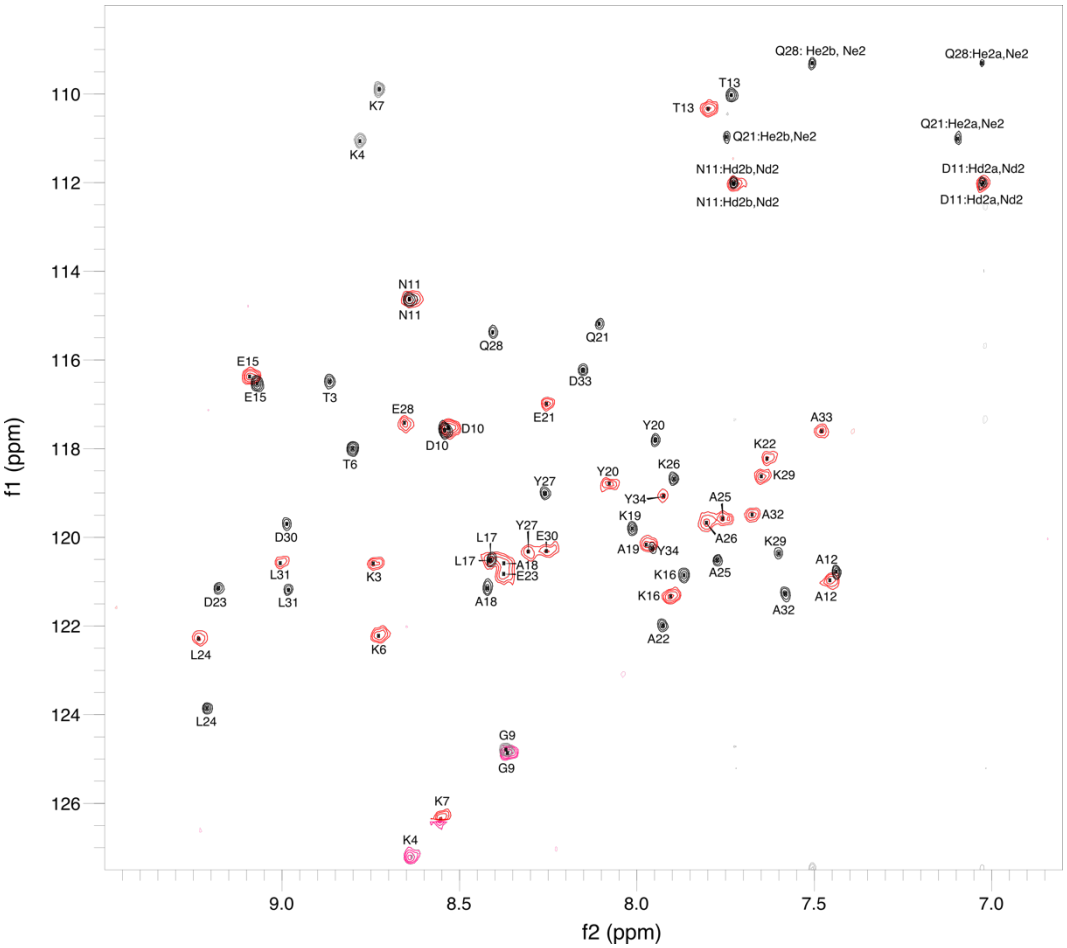


Figure 3-8: Assigned natural abundance ^{1}H-^{15}N HSQC of oPPα (red) and parent PPα (black) acquired on a 700 MHz spectrometer. Pink and grey peaks are folded into the spectrum for oPPα and parent PPα respectively. Residues are labelled using their one-letter codes. Coupling in carboxamide groups of Asp and Gln are specifically labelled.

3.4.5 Analysis of the NMR structure of oPPα

Despite the repetitive sequence, 95.8% of the ^{1}H NMR resonances for oPPα were assigned. The core of the structure, in particular the interdigitating Pro and Tyr residues, was well defined, while some of the solvent-exposed residues were more dynamic. The most dynamic region of the structure was the loop. The RMSDs across the ensemble were 0.52 ± 0.13 Å (backbone) and 1.05 ± 0.15 Å (all atom). These backbone and all atom RMSDs are comparable to those for the 900 MHz NMR structure of PPα; 0.51± 0.12 and 0.83 ± 0.12 Å, respectively (**Figure 3-9A**). The NMR structure of oPPα was closely similar to the NMR structure of PPα (**Figure 3-9B&C**). A representative structure from the oPPα ensemble (PDB 6GWX, model 8) matched PPα (PDB 5LO2, model 14) with RMSDs of 0.5 Å and 0.9 Å, respectively, measured over the backbone and all atoms.

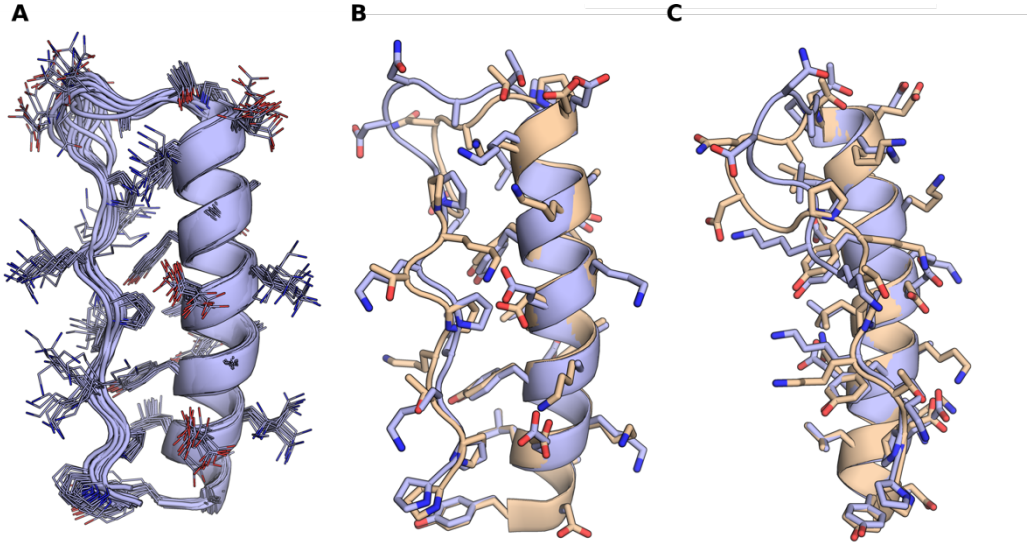


Figure 3-9: (**A**) Overlay of all 20 states in the NMR ensemble of oPPα. (**B**) Overlay of PPα (PDB 5LO2, model 14) and oPPα (PDB 6GWX, model 8). (**C**) Overlay of PPα (PDB 5LO2, model 14) and oPPα (PDB 6GWX, model 8) rotated through 90 ° along the axis of the α helix compared to (**B**). Colour key: oPPα (purple), PPα (wheat). Note PDB 6GWX is deposited in PDB but has not been released yet.

To explore the origins of the improved thermal stability of oPPα, the ensembles of structures for PPα and oPPα were compared in more detail. First, distances between the Lys and Glu residues introduced into oPPα (***2:e*** and ***3:g*** distances) were measured to determine whether interhelical salt bridges were present. An accepted definition of a salt bridge, defined by Barlow and Thornton, is based on a cut-off distance of 4 Å between N-O atom pairs.[230] Overall across all 20 ensemble states the average distances between the N_{ζ} of Lys and the O_{ε} of Glu were 8.7 ± 1.8 Å and 11.5 ± 2.4 Å for ***2:e*** and ***3:g*** respectively and therefore outside of any accepted definition for formal salt-bridge interactions;[231] thus, none of the potential salt bridges were made. However, while it is clear formal salt bridges do not stabilise oPPα, general electrostatic steering during folding likely contribute to stability.[24, 204, 232-236] In the literature, Schreiber and Fersht show that long-range electrostatic interactions increased the association rate constant between two proteins, barnase (Bacterial Ribonuclease) and its intracellular inhibitor (barstar), while also helping orientate the two proteins

for docking.[204] Electrostatic steering has been shown to enhance the association rate between proteins in other studies too both experimentally and computationally.[233, 237]

A key difference between the NMR structures was in the loop region and the orientation of the Pro8 residue closest to the loop. In PPα, the loop kinks out causing Pro8 to be more solvent exposed compared to the equivalent Pro8 residue in oPPα (**Figure 3-10A**). In oPPα, Pro8 docked into the diamond-shaped hole on the α helix more intimately than in PPα. The twisting out of Pro8 is likely result of the preceding residue, Lys7. In oPPα, Lys7 formed a long-range electrostatic interaction with Glu21 (**Figure 3-10B**). In PPα, there is no electrostatic steering effect between the two helices, instead a single interhelical salt bridge is made between Lys7 and Asp23 at 4.4 ± 1.3 Å with the knock on effect of Pro8 docking less tightly into the diamond shaped hole on the α helix (**Figure 3-10C**). We contend that electrostatic steering between the two helices has resulted in tightening of the oPPα NMR structure compared to PPα, which has contributed to the enhanced thermal stability.

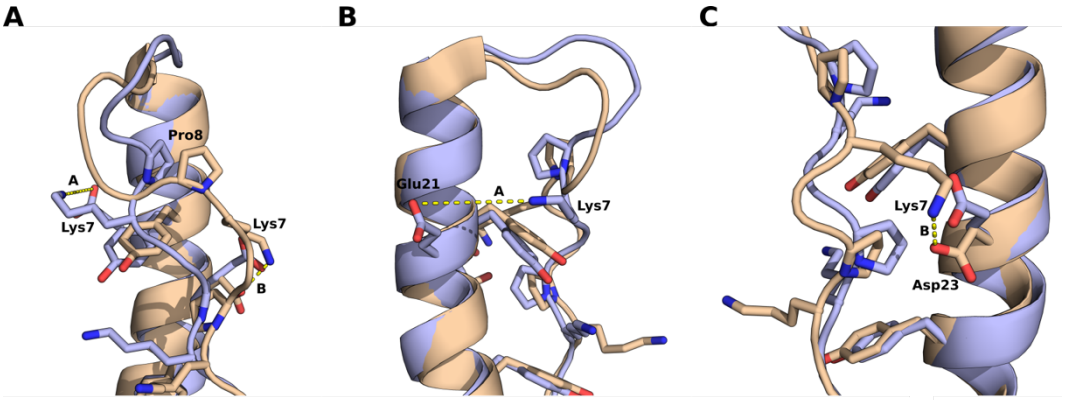


Figure 3-10: Overlay of oPPα (PDB 6GWX, model 14) and PPα (PDB 5LO2, model 14) (A) Different orientations of residue Pro8 observed in PPα and oPPα. (B) Long-range electrostatic interaction (6.8 Å) between Lys7 and Glu21 in oPPα. (C) Salt bridge (2.7 Å**)** between Lys7 and Asp23 in PPα. Colour key: oPPα (purple), PPα (wheat).

3.4.6 Analysis of solvent accessibility and side-chain burial in oPPα

Next, we explored other factors that may have contributed to the enhanced stability of oPPα over PPα. The driving force for the assembly and stabilisation of the PPα fold is the interdigitation of Pro and Tyr residues and the burying of the hydrophobic interface. While polar hydrophilic residues are stabilised by interactions with water, hydrophobic side chains prefer to be buried and out of contact with water.[238-239] The solvent accessible surface area (SASA) refers to the surface area of a biomolecule that is accessible to solvent, usually water. This is measured by rolling a probe of radius 1.4 Å over the van der Waal's surface of a macromolecule of interest, and the path traced out by the probe's centre is the accessible surface.[188] SASAs were calculated for both the oPPα and PPα NMR structures within the ISAMBARD framework using the programme Naccess.[189] First, SASAs were calculated for intact oPPα and PPα complex (**Figure 3-11A**) with the loop region ignored in the calculations as it is the same for both the oPPα and PPα structures. SASAs were then calculated for the individual α and PPII helix components for each structure (**Figure 3-11B&C**).

In order to be able to make reliable comparisons between the oPPα and PPα structures, the SASA of the complex was taken as a proportion of the total SASA of the individual components. This is summarised in **Equation 3-1**. This was done for every state in the NMR ensemble and the mean average calculated.

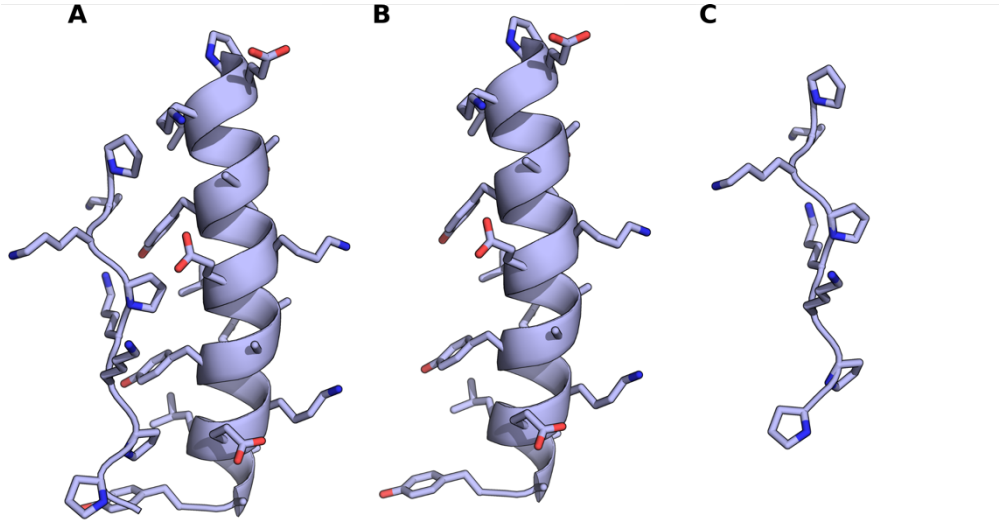


Figure 3-11: (**A**) oPPα complex with the loop removed. (**B**) α helix component (**C**) PPII helix component.

$$Relative\ SASA = \frac{SASA_{PP\alpha\ complex}}{(SASA_{\alpha} + SASA_{PPII})}$$

Equation 3-1: SASA of the PPα or oPPα complex as a proportion of the SASA of the individual α and PPII helix components.

A low value for the relative SASA is indicative of the interface between the two helices being less solvent accessible and thus more buried compared to a high value for normalised SASA. In other words, a normalised SASA value of 0 indicates fully buried and a value of 1 indicates fully solvent exposed.

Overall, the normalised SASA of oPPα was equal to that of PPα (oPPα 0.78 ± 0.006, PPα 0.78 ± 0.010). While the overall SASA is the same for both structures, on a per residue basis there are some differences between the structures (**Figure 3-12**). There is a shift in SASA for some interface amino acids, for example, in oPPα Tyr20, Leu17 and Lys4 are considerably more buried than in PPα. However, overall the analysis suggests there is no significant difference in SASA between the two NMR structures and therefore the enhanced thermal stability of oPPα is not a result of higher side-chain burial.

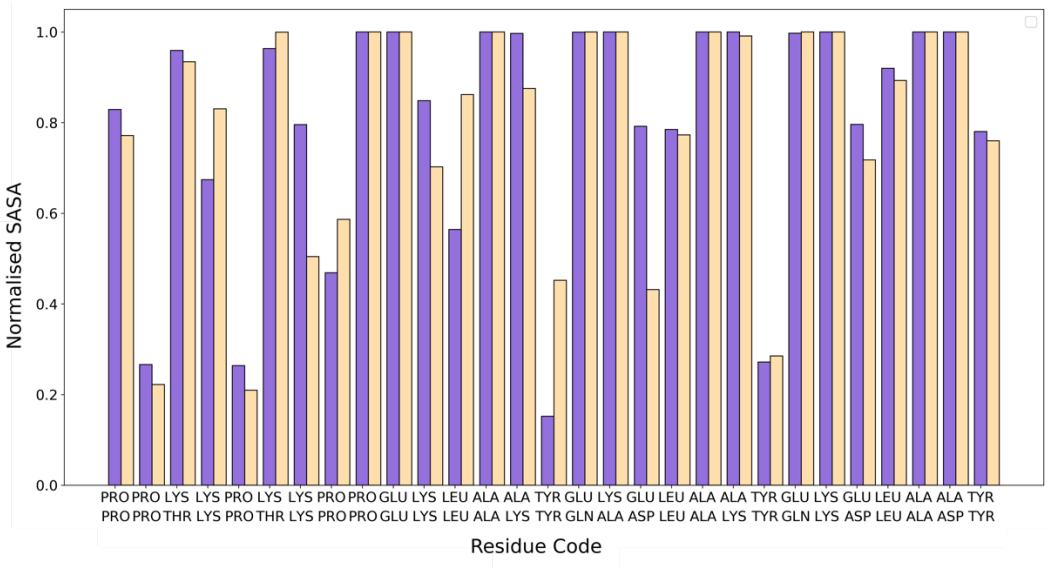


Figure 3-12: Normalised SASA per residue in oPPα vs PPα. Colour key: oPPα (purple), PPα (wheat). Top line on x axis: oPPα sequence. Bottom line on x axis: PPα sequence.

3.4.7 Analysis of CH–π interactions in oPPα

Previous work shows that PPα is stabilized by CH–π interactions,[153] therefore similar interactions were searched for in the oPPα ensemble. CH–π interactions were identified within the ISAMBARD framework.[168-142] Interactions were classed as CH-positive if the following criteria were met: CH–π distance (between the CH proton to centre of the aromatic ring) ≤3.5 Å; CH–π angle (between the vector of the CH bond and the normal to the plane of the aromatic ring) ≤55°; H projection distance (between the projection of the CH bond to the plane of the aromatic ring and the centre of the ring) ≤2.0 Å for Tyr (**Figure 3-13**).

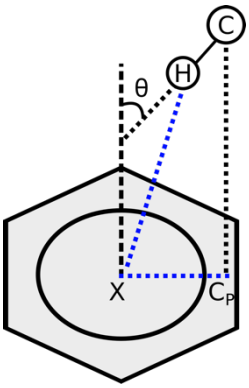


Figure 3-13: Parameters defining a CH-π interaction. CH–π distance (between the CH proton to centre of the aromatic ring, i.e. H–X, blue) ≤ 3.5 Å. CH–π angle (between the vector of the C–H bond and the normal to the plane of the aromatic ring) θ ≤ 55°. C-projection distance (between the projection of the CH carbon to the plane of the aromatic ring and the centre of the ring, i.e. X – C_p, blue) ≤ 2.0 Å for 6-membered ring (Tyr).

The analysis revealed 87 CH–π interactions across all 20 ensemble states in oPPα compared to 68 interactions in PPα, equating to an increase of 28 %. This works out at an average of 4.35 CH–π interactions per structure for oPPα compared to 3.40 per structure for PPα (**Table 3-2**, **Figure 3-14**).

The greater number of CH–π interactions in the oPPα NMR structure, approximately one extra CH–π interaction per ensemble state, likely contributes to the enhanced thermal stability observed for oPPα over PPα.

Table 3-2: Table of CH–π interactions in oPPα compared to PPα.

	oPPα	**PPα**
Pro donor: total CH–π interactions	5	22
Pro donor: CH–π interactions per structure	0.25	1.1
Leu donor: total CH–π interactions	25	4
Leu donor: CH–π interactions per structure	1.25	0.2
Lys donor: total CH–π interactions	57	15
Lys donor: CH–π interactions per structure	2.85	0.75
Gly donor: total CH–π interactions	0	20
Gly donor: CH–π interactions per structure	0	1
Acetyl group donor: total CH–π interactions	0	7
Acetyl group donor: CH–π interactions per structure	0	0.35
Total CH–π interactions	**87**	**68**
Average CH–π interactions per structure	**4.35**	**3.40**

Interestingly, the dominant CH donors differed between the two ensembles: in oPPα fewer CH donors emanated from Pro side chains (oPPα: 5 interactions, PPα:, 22 interactions), but many more came from Leu (oPPα: 25 interactions, PPα: 4 interactions) and from the Lys residues (oPPα: 57 interactions, PPα: 15 interactions) (**Table 3-2**,**Figure 3-14**).

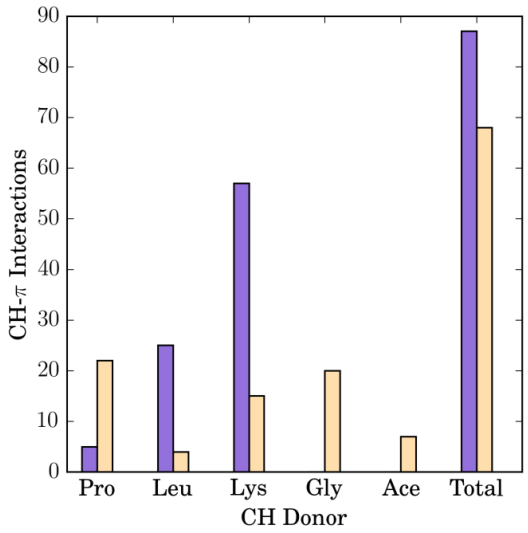


Figure 3-14: Bar chart summarising CH–π interactions across oPPα NMR ensemble compared to PPα. Colour key: oPPα (purple), PPα (wheat).

On closer inspection, in oPPα the side chains of Lys4 and Lys7 lie across the top faces of Tyr27 and Tyr20, respectively, leading to CH–π interactions between the C_α and C_β protons of the former and the aromatic rings of the latter. The side chains of Leu24 and Leu31 sit on the underside of Tyr20 and Tyr27 respectively, leading to CH–π interactions between the C_β protons of the former and the aromatic rings of the latter. Tyr residues appear to be anchored in place by CH–π interactions from Lys and Leu above and below the ring respectively (**Figure 3-15**).

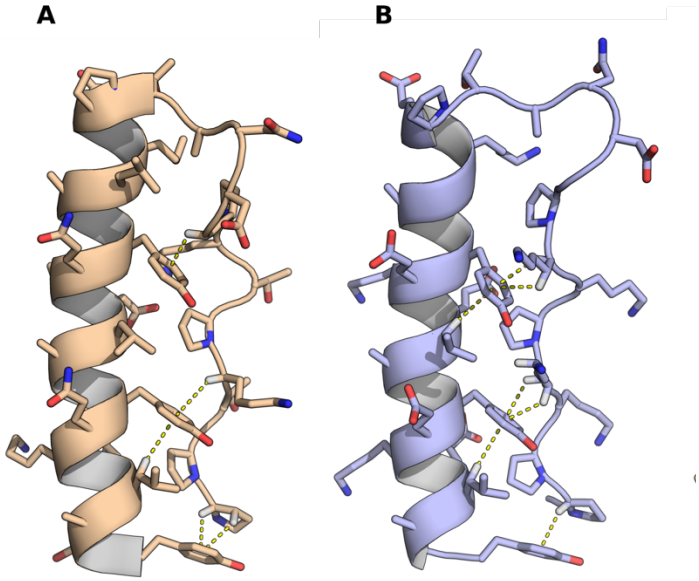


Figure 3-15: Representative models from the NMR structure ensembles of PPα and oPPα highlighting CH–π interactions. (**A**) PPα (PDB ID 5LO2, model 14). (**B**) oPPα (PDB ID 6GWX, model 8). Key: oPPα (purple), PPα (wheat), CH–π interactions(dashed lines).

The contrasting conformation of Lys side chains between structures is best emphasised on overlaying all NMR states for both PPα and oPPα (**Figure 3-16**). There is a change in conformation of the Lys4 residue between PPα and oPPα. In oPPα, Lys4 points more towards the interface of the two helices and lies across the face of Tyr27 to a greater extent than in PPα. In PPα Lys4 is more solvent exposed, consistent with the calculated SASA measurements for Lys4 (**Figure 3-12**).

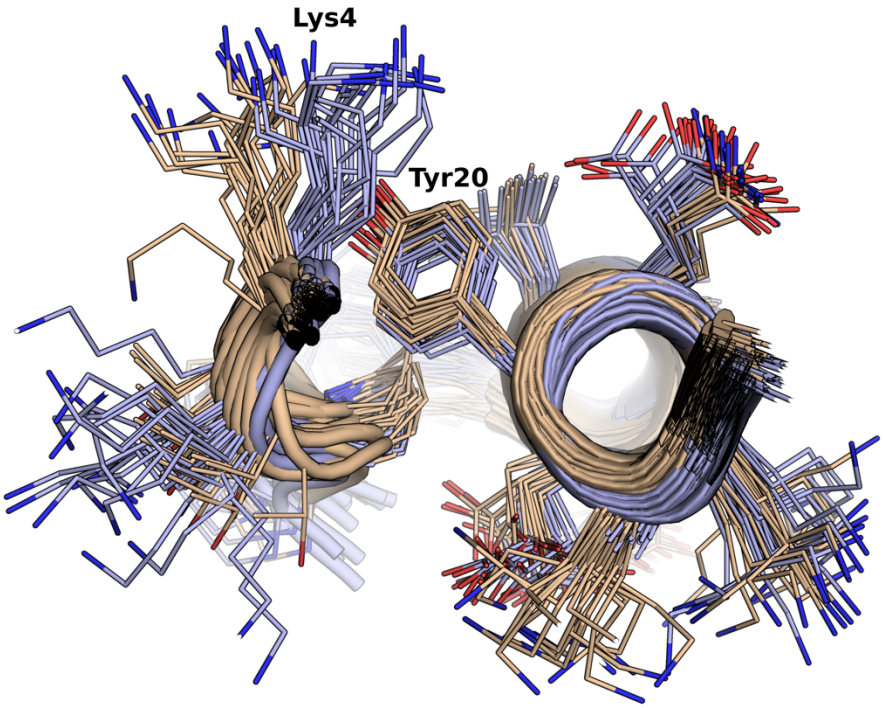


Figure 3-16: Overlay of oPPα and PPα NMR ensemble structures highlighting conformational change in Lys4 across the face of the aromatic ring in Tyr27 between structures. Key: oPPα (purple) and PPα (wheat).

3.5 Further substitutions in oPPα

Baker *et al.* show that the stability of PPα is influenced by the electron donating ability of the aromatic residue at the ***d*** position: peptides containing electron rich residues Tyr and Trp are more thermally stable compared to peptides containing electron poor residues Phe and His.[153] To explore the importance of the CH-π interactions further, all of the electron-rich Tyr residues in oPPα were replaced by more-electron-poor phenylalanine (oPPα-Phe) (**Figure 3-17**). The electrostatic surface potentials of Phe and Tyr side-chains are shown in **Figure 3-17D**. As judged by CD spectroscopy, oPPα and oPPα-Phe had similar helicities at 5 °C. However, the thermal stability of oPPα-Phe dropped by approximately 20 °C compared to oPPα giving a T_M of 33 °C. This drop is consistent with previous studies with PPα and thus indicates that while the identity of the CH–π donors is shifted the interactions are still at play. oPPα-Phe was monomeric by sedimentation equilibrium AUC.

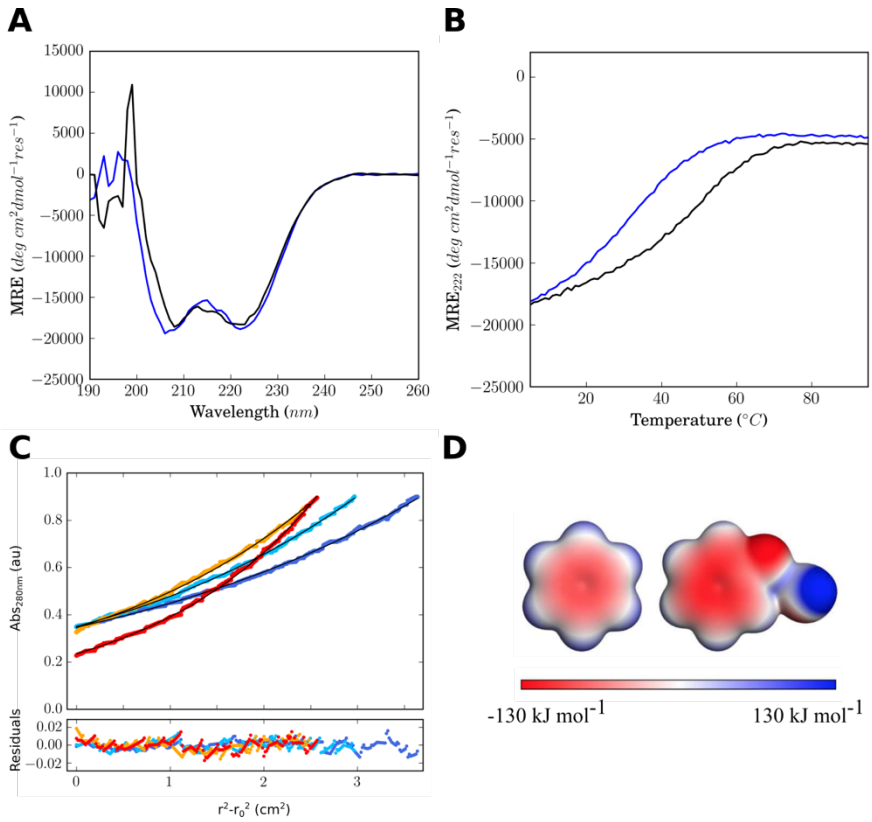


Figure 3-17: Folding and stability of oPPα-Phe. (A) CD spectra recorded at 5 °C for oPPα (black), oPPα-Phe (blue) (B) Temperature dependence of the far-UV CD signal monitored at 222 nm. (C) AUC Sedimentation-equilibrium for of oPPα-Phe. Data (top, dots) and fitted single-ideal species model curves (lines) at 44 (blue), 52 (aqua), 56 (orange), and 60 (red) krpm for oPPα peptides. Bottom: Residuals for the above fits using the same colour scheme. oPPα-Phe ($\bar{v}$ = 0.756 cm^3 g^{-1}). The fit returned a mass of 3771 Da, (1.0 monomer mass), 95% confidence limits = 3742 – 3801 Da. (D) Electrostatic surface potentials of the side chains of Phe (left) and Tyr (right). Scale: ≤ -130 kJ mol^{-1} (electropositive, blue) through ≥ 130 kJ mol^{-1} (electronegative, red). Modified from reference [153]. Conditions for CD: 100 µM peptide, PBS, pH 7.4. Conditions for AUC: 130 µM peptide, PBS, pH 7.4.

3.6 Chapter conclusion

In conclusion, the PPα fold has been optimised by rational protein redesign to give oPPα, which has a completely *de novo* framework and significantly enhanced thermal stability. A solution-phase NMR structure supported by a range of biophysical techniques including CD spectroscopy and AUC confirms that, similar to the parent peptide, oPPα forms a stably folded monomeric and compact structure comprising a polyproline-II helix linked *via* loop to an α helix. The two helices interact *via* KIH-like packing with the interdigitation of Pro and Tyr residues. CD spectroscopy reveals that oPPα unfolds reversibly with a midpoint unfolding temperature of 51 °C, a substantial increase of approximately 12 °C compared with PPα. Overall, 10 mutations are made to the PPα fold to yield the optimised sequence.

The increase in stability observed in oPPα compared to PPα is the result of a combination of factors. It was postulated that introducing Lys-Glu pairs between the ***2*** and ***3*** position of the PPII helix and the ***e*** and ***g*** position of the α helix would result in the formation of interhelical Coulombic interactions that would help stabilise the miniprotein fold. This is a strategy used frequently in the *de novo* design of coiled coils to guide and stabilise assemblies.[132-133] However, the distance between the Lys and Glu residues in oPPα indicated that salt bridges are not present in the NMR ensemble; distances are greater than the accepted cut-off for salt bridge interactions of 4 Å. While formal salt bridges may not stabilise oPPα, general electrostatic steering during folding likely contributes to the stability.[204, 232]

Overall tightening of the oPPα NMR structure compared to PPα is observed, which will also contribute to the enhanced stability. The oPPα NMR structure reveals all Pro residues on the polyproline-II helix face and dock into diamond shaped holes top and tailed by Tyr residues forming KIH-like interactions. However, there is an anomaly in PPα where Pro8, the Pro residue closest to the loop region, kinks out and is more solvent exposed; it does not dock so tightly into the diamond hole formed on the α helix. This is quantified by the higher SASA value for Pro8 in PPα compared to oPPα.

This tightening of the structure improves the non-covalent interactions, specifically the frequency of CH–π interactions. Analysis of the oPPα NMR structure reveals a significant increase in the number of CH–π interactions in the ensemble compared to PPα; 87 CH–π interactions (4.35 per structure) in oPPα compared to 68 (3.40 per structure) in PPα *i.e.* one extra CH–π interaction per ensemble state. Interestingly, the dominant CH donor shifts from Pro in PPα to Lys and Leu in oPPα. Closer inspection of the NMR ensemble shows Tyr residues are stabilised by CH-π interactions from Lys above the aromatic ring and Leu below the ring. Lys appears to lie across the face of Tyr.

Overall, the enhanced stability of oPPα can be attributed to electrostatic steering effects and overall tightening of the structure resulting in improved non-covalent binding, specifically an increase in the number of CH–π interactions. Further sequence-to-stability relationships need to be investigated to fully understand this miniprotein. In Chapter 4 such relationships will be discussed, in particular the effect of mutating the interface diamond-shaped-hole residues.

Chapter 4 Probing sequence-to-stability relationships in oPPα

The work described in this chapter was designed by the author of this thesis (KPG), Dr Emily Baker and Prof. Dek Woolfson. Bristol Chemical Synthesis rotation CDT students Frank Zieleniewski and Jenny Samphire synthesised and preliminary characterised peptides oPPα-A@a, oPPα-I@a, oPPα-V@a, oPPα-E@a, oPPα-K@a. Full characterisation of these peptides was followed up by KPG. Molecular dynamics simulations were set up and analysed by KPG.

4.1 Chapter introduction

Chapter 3 described the redesign of PPα into a *de novo* framework. This resulted in an optimised PPα fold with significantly enhanced thermal stability. The increase in the thermal stability of oPPα can be rationalised by electrostatic steering from interhelix Lys-Glu pairs and improved non-covalent interactions such as CH–π interactions as a result of the overall tightening of the structure. This chapter focuses on the interface residues involved in the KIH-like packing to further understand the PPα fold.

We have a good understanding of sequence-to-structure relationships in α-helical coiled coils.[132-133, 240] This has led to success in the rational design of lower order coiled coils[32, 203] (both homo- and hetero- meric dimers trimers, tetramers) as well as more complex α-helical barrels with central channels.[145] Successful rationally designed coiled coils are now being used as scaffolds onto which functions are being grafted[152] and as building blocks for more complicated assemblies.[241-242] In a similar way, deeper understanding of the PPα fold should help pave the way for the successful design of more complicated assemblies based around the polyproline-II helix:loop:α helix topology. This chapter explores sequence-to-structure and -stability relationships in oPPα by probing the diamond shaped hole residues that accommodate the Pro knob residue.

4.1.1 Knobs-into-holes packing

As discussed in previous chapters, intimate side-chain interdigitation, named knobs-into-holes (KIH) packing, directs helix-helix interactions in coiled coils. Specifically, hydrophobic side chains at ***a*** and ***d*** on one helix act as knobs and dock into diamond-shaped holes of four residues on the adjacent helix. For parallel helices, two types of KIH interactions are possible depending on the identity of the knob residue: an ***a*** knob can dock into a hole formed by ***dgad*** residues while a ***d*** knob interacts with ***adea*** residues (**Figure 4-1A**).[243]

As a result of the helical geometry, all the C_α-C_β bonds of side-chains project out of the helix backbone along a defined vector.[134] In parallel dimers ***a*** and ***d*** knob residues point towards the diamond shaped hole residues on the adjacent helix differently. The C_α-C_β bond vector of the ***d*** knob residue points directly into the hole towards the opposite helix and is referred to as perpendicular

packing.[132] Whereas the C_α-C_β bond vectors of ***a*** knobs point away from the helix-helix interface, which is termed parallel packing (**Figure 4-1C**).[132, 143, 244] As a result, ***d*** residues favour non-β-branched less bulky residues like Leu while ***a*** residues are more accommodating of different shaped side-chains and thus prefer β-branched Ile and Val.[132, 245]

In the PPα fold, Tyr at ***d*** is critical for folding and stability[153] and forms two sides of the diamond hole defined by side chains at successive ***dgad*** sites on the α helix. Only position ***1*** (Pro) on the polyproline-II helix acts as a knob residue pointing at the diamond hole on the α helix as a result of the three residue per turn geometry of the polyproline-II helix (**Figure 4-1B&D**).

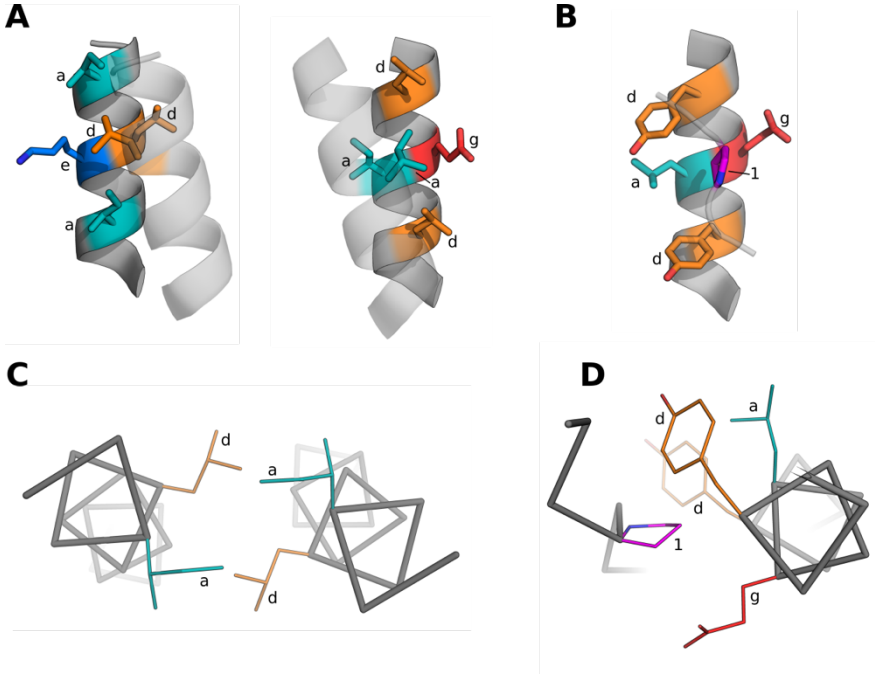


Figure 4-1: (A) Representation of knobs-into-holes packing for parallel coiled-coil dimers: (left) 'd' hole (right) 'a' hole. (B) Knobs-into-hole like packing in PPα. (C) Coiled-coil dimer (CC-Di, PDB: 4DZM) highlighting parallel and perpendicular packing. (D) oPPα (NMR structure, state 1) highlighting the ***a*** (cyan) and ***d*** (orange) sites. Key: cyan: ***a*** position, orange: ***d*** position**,** blue: ***g*** position, pink: ***1*** position.

4.1.2 Related natural PPα peptides

As described in the Introduction, the fragment based design of PPα borrowed from two natural proteins: pancreatic peptides[29] and the antigenic bacterial surface adhesin[166] (AgI/II). The pancreatic polypeptide family, consists of pancreatic peptides,[29, 246-247] peptide YY[248] and neuropeptide Y[249] and all have sequence similarities and some conserved residues (**Table 4-1**). The conservation of residues at the ***dgad*** positions in natural sequences was explored.

Table 4-1: Sequences of the pancreatic polypeptide family.

Peptide name	Sequence and helical register	PDB Code
	321321 efgabcd efgabcd efgabcd	
oPPα	Ac- PPKKPKKP GDNAT PEKLAAY EKELAAY EKELAAY -NH_2	6GWQ*
Parent PPα	Ac- PPTKPTKP GDNAT PEKLAKY QADLAKY QKDLADY -NH_2	5LO2
aPP	GPSQPTYP GDDAP VEDLIRF YNDLQQY LNVVTRH RY -NH_2	1PPT†
bPP	APLEPEYP GDNAT PEQMAQY AAELRRY INMLTRP RY -NH_2	1BBA
hPP	APLEPVYP GDNAT PEQMAQY AADLRRY INMLTRP RY -NH_2	-
pPP	APLEPVYP GDDAT PEQMAQY AAELRRY INMLTRP RY -NH_2	-
nPY	YPSKPDNP GEDAP AEDLARY YSALRHY INLITRQ RY -NH_2	1RON
pYY	YPIKPEAP GEDAS PEELNRY YASLRHY LNLVTRQ RY -NH_2	2DEZ
AgI/II	PPVKPTAP // EAKLAKY QADLAKY QKDLADY PV	3IOX

aPP (avian pancreatic peptide), bPP (bovine pancreatic peptide), hPP (human pancreatic peptide), pPP (porcine pancreatic peptide), nPY (neuropeptide Y), pYY (peptide tyrosine tyrosine), AgI/II (antigenic bacterial surface adhesin) residues 450-472 and 838-845. *NMR structure deposited in the PDB but will not be released to the public until publication. †All structures are NMR structures apart from aPP which is an X-ray crystal structure.

Table 4-1, showing the sequences of the pancreatic polypeptide family, can also be interpreted visually as a sequence logo: a graphical representation of an amino acid sequence alignment (Error! Reference source not found.).[250-251] The logo consists of a stack of letters, one stack for each position in the sequence. The height of the letter within the stack indicates the relative frequency of each amino acid at that position. Letters at the top of a stack are the most frequent across all sequences.

Both **Table 4-1**and Error! Reference source not found. show that Tyr at ***d*** and Pro at position ***1*** are highly conserved in related folds of natural proteins. Leu at position ***a*** is also conserved. The ***a*** position (residue24) is entirely Leu across all sequences for the middle α-helical heptad (heptad 2) while the frequency of other residues at the ***a*** position increases in heptads ***1*** and ***3***. For example for heptad 3, while Leu (residue30) is still the most frequent, other hydrophobic residues Ile and Val are present in some sequences. The ***g*** site is less conserved in related natural peptides and in the majority of cases it is a polar residue.

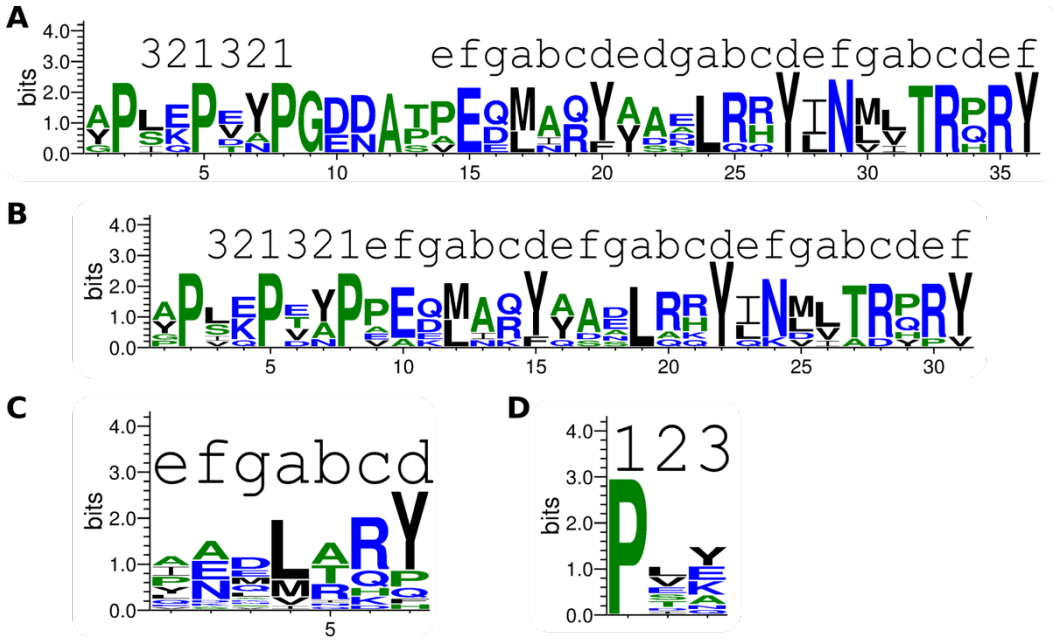


Figure 4-2: Sequence logo of the related natural PPα like folds. (A) Sequence logo for aPP, bPP, hPP, pPP, nPY and pYY. (B) Sequence logo for aPP, bPP, hPP, pPP, nPY, pYY and AgI/II. Loop region omitted. (C) Sequence logo for an α-helical heptad repeat for aPP, bPP, hPP, pPP, nPY, pYY and AgI/II. (D) Sequence logo for the three residue repeating unit in the polyproline-II helices of aPP, bPP, hPP, pPP, nPY, pYY. Sequence logo for Sequence logos generated with Weblogo.[250-251] Colour key: polar residues (blue), hydrophobic residues (black), Ala, Pro and Gly residues (green).

As well as having sequence similarities, the pancreatic polypeptide family are also structurally very similar. **Figure 4-3A** shows overlays of the middle heptad of oPPα, parent PPα, avian pancreatic peptide (aPP), bovine pancreatic peptide (bPP), neuropeptide Y (nPY) and antigenic bacterial surface adhesin (AgI/II). All structures overlay well. The majority have similar side chain conformations for the ***dgad*** residues, in particular for the Pro and Tyr residues. There is some variation in the puckering of the ring in the Pro residues.

nPY is slightly different. While it has sequence similarity to the other peptides it has different structural features. nPY has distinctly different side-chain conformations for the Tyr residues (**Figure 4-3A**) and in the NMR structure the α helix does not interact with the polyproline-II helix; NOEs were not found between the α and polyproline-II helices (**Figure 4-3B**).[249] This can be rationalised by studies that have shown nPY is a dimer in solution with the two monomers interacting *via* their α helices in an antiparallel fashion.[252] From a biological perspective, nPY is a neurotransmitter acting on both the central and peripheral nervous systems and the self-associated forms of nPY is thought to be its storage form while the monomer is the active form.[249]

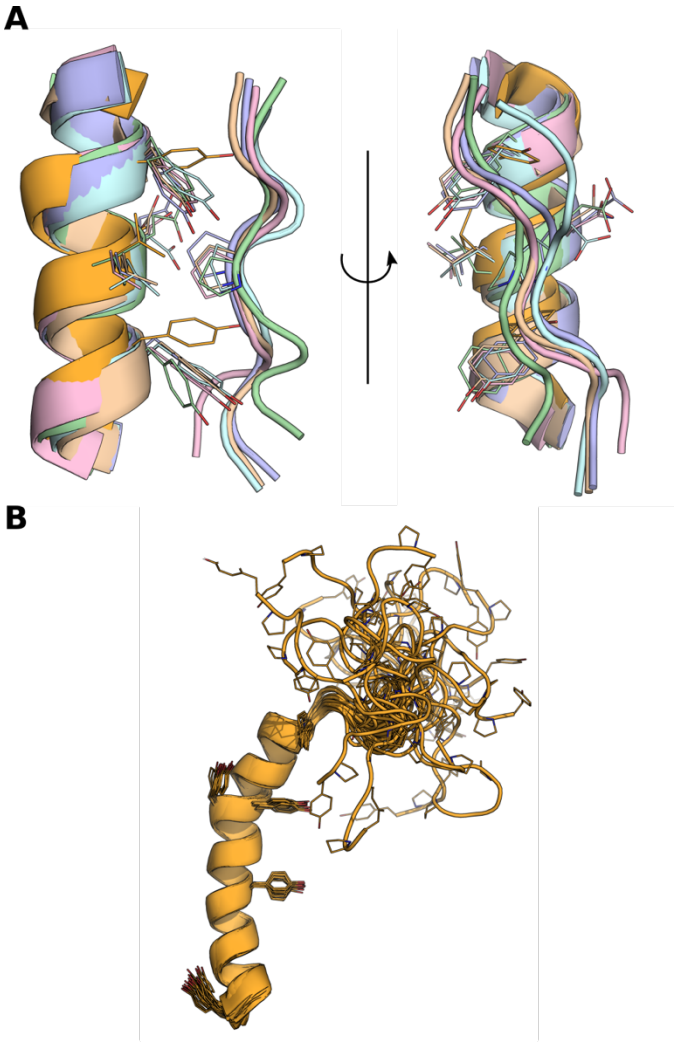


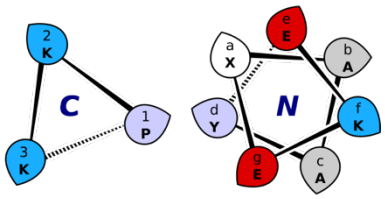
Figure 4-3: (A) Overlay of the middle heptad for aPP (pink), bPP (green), nPY (orange, only α helix shown), AgI/II (wheat), oPPα (purple), parent PPα (pink) highlighting the similar conformations of the side chains in the diamond hole residue positions ***dgad***. Structures aligned in Pymol. Note nPY PPII helix not present since peptide does not adopt PPα topology. (B) Overlay of all 26 states in the NMR ensemble of nPY (PDB: 1RON).

4.2 Probing the diamond shaped hole residues in oPPα

4.2.1 Mutations to the *a* position of oPPα

While Pro and Tyr residues are essential for the formation and stability of the PPα fold the amenability of the ***a*** and ***g*** positions to mutation were investigated. First, to explore what other residues, if any, might substitute for Leu at ***a***, a series of mutants (*X@**a*** mutants) in which all three sites were replaced by Ala, β-branched Ile or Val or charged Glu or Lys residues were synthesised and characterised in solution (**Figure 4-4**). Non-bulky Ala was initially selected to determine the contribution of Leu to the stability of the peptide in a similar manner to alanine-scanning. β-branched Ile and Val were selected because the sequence logos above show they are present at the ***a*** position in the heptad repeats of natural related sequences while Glu and Lys were selected to determine the effect of charged residue at the ***a*** position.

A



X = L, I, V, E, K, A

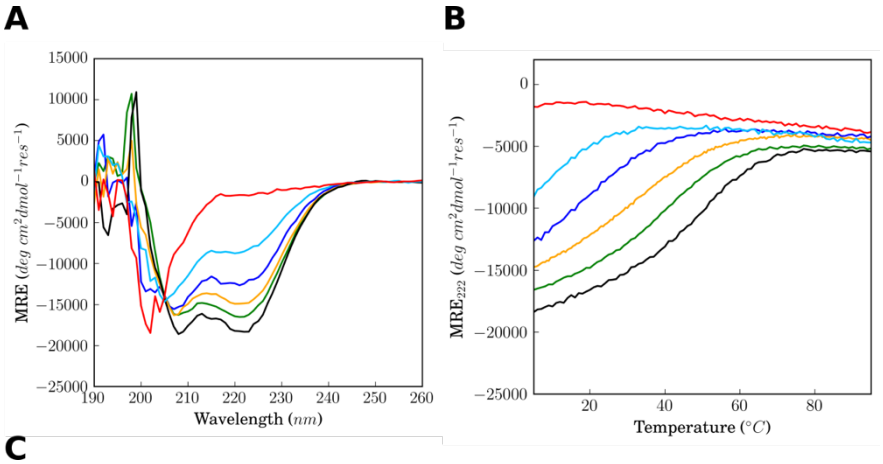
B

Peptide name	Sequence and helical register			
	321321	efgabcd	efgabcd	efgabcd
oPPα	Ac- PPKKPKKP	GDNAT	PEK**L**AAY	EKE**L**AAY EKE**L**AAY -NH_2
oPPα-I@***a***	Ac- PPKKPKKP	GDNAT	PEK**I**AAY	EKE**I**AAY EKE**I**AAY -NH_2
oPPα-V@***a***	Ac- PPKKPKKP	GDNAT	PEK**V**AAY	EKE**V**AAY EKE**V**AAY -NH_2
oPPα-E@***a***	Ac- PPKKPKKP	GDNAT	PEK**E**AAY	EKE**E**AAY EKE**E**AAY -NH_2
oPPα-K@***a***	Ac- PPKKPKKP	GDNAT	PEK**K**AAY	EKE**K**AAY EKE**K**AAY -NH_2
oPPα-A@***a***	Ac- PPKKPKKP	GDNAT	PEK**A**AAY	EKE**A**AAY EKE**A**AAY -NH_2

Figure 4-4: (A) Helical wheel representation of the series of mutations at the ***a*** position (X@***a*** mutations). (B) Sequences of the series of peptides with mutations to the ***a*** position.

4.2.1.1 Characterisation of the X@***a*** series by CD spectroscopy

These peptides were first analysed by CD spectroscopy to determine the effect of ***a*** position mutations on the overall structure and stability of the peptides (**Figure 4-5**). CD spectra and thermal unfolding curves revealed a broad range of stabilities for the X@***a*** variants ranging in helicity from 5 to 48 % at 5 °C and ranging in T_M from < 5 °C to 51 °C. Of all the X@***a*** mutants, Leu at ***a*** was the most stable, which is consistent with it being the naturally preferred residue at the ***a*** position.



C

Peptide name	$MRE_{222\ nm}$ / deg cm^2 $dmol^{-1}$ res^{-1}	% Helicity[175-176]	T_M / °C
oPPα	-18319	48	51 ± 0.8
oPPα-I@***a***	-16493	43	40 ± 0.1
oPPα-V@***a***	-14854	37	34 ± 1.4
oPPα-K@***a***	-12448	33	19 ± 2.3
oPPα-A@***a***	-8522	23	-
oPPα-E@***a***	-1642	5	-

Figure 4-5: Folding and stability of oPPα and X@***a*** variants. (A) CD spectra recorded at 5 °C for oPPα variants: oPPα-E@***a*** (red), oPPα-A@***a*** (sky blue), oPPα-K@***a*** (blue), oPPα-I@***a*** (green), oPPα-V@***a*** (orange), and oPPα (black) for comparison. (B) Temperature dependence of the far-UV CD signal monitored at 222 nm for the X@***a*** mutants, same colour scheme as in (A). (C) Table summarising the MRE at 222 nm at 5 °C, the % helicity and also the T_M for all the X@***a*** peptide variants. Conditions: 100 μM peptide, PBS, pH 7.4.

The stability of the oPPα-A@***a*** variant was drastically compromised with respect to oPPα. The CD spectra at 5 °C showed 23% α-helical character, less than half that of oPPα, despite the favourable increase in helical propensity of Ala compared to Leu. Thermal stability was also substantially compromised. We attribute this to the non-bulky nature of the methyl functional group eliminating any preferable side chain interactions. Further, this implies that Leu at ***a*** plays an important role in the folding and stability of the peptide and contributes to the interface.

oPPα variants with β–branched residues at ***a***, oPPα-I@***a*** and oPPα-V@***a***, were well folded, albeit with reduced thermal stabilities compared to oPPα with T_Ms values of 40 °C and 34 °C vs 51 °C respectively. This is perhaps unsurprising given the similar size and isomeric nature of Ile and Leu. Furthermore, both Ile and Val are present in the third heptad of some natural sequences (**Table 4-1**) hinting at their amenability at the ***a*** position. The reduction in T_M values suggests that the extra steric bulk of the β carbon may hinder access of the Pro knob to the hole. However, it must also be noted that this cannot be disconnected from the fact that the α-helix propensities of Ile and Val are appreciably lower than that of Leu.[209]

The least folded variant, oPPα-E@***a***, was unfolded showing no α-helical characteristics at 208 and 222 nm. We postulate this is due in part to the proximal glutamates at the ***e*** position if the α helix were folded, but also the bulky nature of the Glu side-chain may potentially prevent Pro from docking so tightly into the diamond hole. This is despite a potentially positive interaction with the ***2*** position. Also the helical propensity of Glu is slightly lower compared to Leu which will contribute to the reduced stability. This result suggests large negatively charged side chains are not compatible at the ***a*** position. In comparison, oPPα-K@***a*** gives a partially folded peptide with a T_M of 33 °C. Lys likely forms a favourable i→i+4 interaction with Glu at the ***e*** position contributing to stability.[24] However the close proximity to Lys at position ***2*** will reduce electrostatic steering effects between the two helices thus hampering stability. It is encouraging nevertheless that larger positively charged residues can be accommodated.

To summarise, Leu at ***a*** gave the most folded and thermally stable variant. However, other residues such as hydrophobic Ile, Val and positively charged residue Lys can also be tolerated. This is encouraging for future applications of oPPα where the sequence will need to be amenable to mutation to allow for function to be introduced.

4.2.1.2 Characterisation of the X@a series by AUC

To determine the oligomeric states of the X@***a*** peptide series in solution SE AUC experiments were carried out. For all peptides in the series, apart from oPPα-E@***a***, experimental data readily fitted to single-ideal species models returning molecular weights consistent with monomers (**Figure 4-6**). oPPα-E@***a*** precipitated at approximately 50 krpm and therefore data could not be fitted.

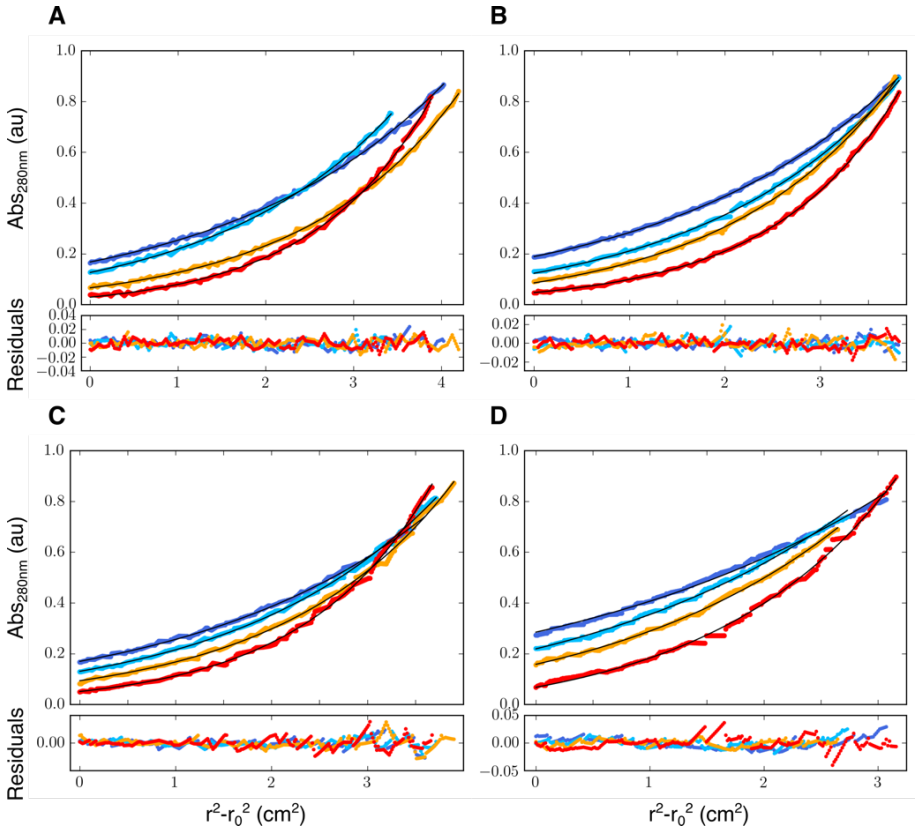


Figure 4-6: AUC Sedimentation-equilibrium data (top, dots) and fitted single-ideal species model curves (lines) at 44 (blue), 52 (aqua), 56 (orange), and 60 (red) krpm. (A) oPPα-A@***a*** ($\bar{v}$ = 0.733 cm^3 g^{-1}). (B) oPPα-K@***a*** ($\bar{v}$ = 0.742 cm^3 g^{-1}). (C) oPPα-V@***a*** ($\bar{v}$ = 0.743 cm^3 g^{-1}). (D) oPPα-I@***a*** ($\bar{v}$ = 0.748 cm^3 g^{-1}). The fit returned the following masses: oPPα-A@***a***, 3510 Da (1.0 × monomer mass), 95% confidence limits = 3495 – 3526 Da; oPPα-K@***a***, 3692 Da, (1.0 × monomer mass), 95% confidence limits = 3680 – 3703 Da; oPPα-V@***a***, 3702 Da, (1.0 × monomer mass), 95% confidence limits = 3683 – 3722 Da; oPPα-I@***a***, 3074 Da, (1.0 × monomer mass), 95% confidence limits = 3037 – 3111 Da. Conditions: 130 μM peptide, PBS, pH 7.4.

4.2.2 Probing a fully hydrophobic interface

To further assess the optimal interface, the ***g*** position of the α helix was also investigated. In oPPα, polar residue Glu at ***g*** contributes to electrostatic steering between the α and polyproline-II helices. Therefore, any mutations at ***g*** could be detrimental in this respect so Glu was installed at the ***c*** position to maintain charge complementarity between the two helices. This allowed the identity of the ***g*** position to be explored. Note also that Glu at the ***c*** position results in a potential Glu → Lys i→i+4 side chain salt bridge with the ***f*** position (**Figure 4-7**).

The α helix also needed to be considered since in oPPα the first heptad is different to the following two heptads and starts with the residues P-E-K at ***e-f-g***. As mentioned in Chapter 3, this is because Pro and Glu are conserved as the *N*-terminal residues in the α helix of related natural sequences (**Table 4-1**, Error! Reference source not found.). This is likely because Pro and Glu have been shown to be good helix capping residues with high frequency of occurrence at the N1 and N2 positions

respectively in α helix motifs.[224-225, 253] Pro lacks the flexibility of other residues because its backbone dihedral angle ϕ is constrained at ≈ –70° similar to an ideal helix.[224] Lys was selected to balance the charge associated with Glu. Therefore, the first turn of the α helix following the loop region was maintained across mutants of oPPα. The ***g*** position was therefore not mutated in the first heptad but only mutated for heptads two and three.

4.2.2.1 Leu at the ***g*** position

The ***g*** site was initially mutated to Leu with Leu maintained at the ***a*** position since our analysis above shows Leu offers the greatest stability. This resulted in a fully hydrophobic diamond hole and a wider hydrophobic interface (oPPα-L@***a***-L@***g***) (**Figure 4-7**).

Peptide name	Sequence and helical register
	321321 efgabcd efgabcd efgabcd
oPPα	Ac- PPKKPKKP GDNAT PEKLA**A**Y EK**E**LA**A**Y EK**E**LA**A**Y -NH_2
oPPα-L@***a***-L@***g***	Ac- PPKKPKKP GDNAT PEKLA**E**Y EK**L**LA**E**Y EK**L**LA**E**Y-NH_2

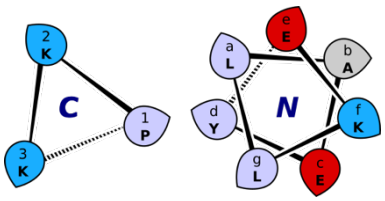


Figure 4-7: (Top) Sequence of peptide with a fully hydrophobic diamond hole (oPPα-L@a-L@g). (B) Helical wheel representation of oPPα-L@a-L@g.

The CD spectra at 5 °C showed that oPPα-L@***a***-L@***g*** had approximately the same helicity as oPPα (**Figure 4-8A**). However, a much broader transition in the thermal denaturation curve was observed with a T_M of 76 °C (**Figure 4-8B**). This is a substantial increase over the T_M of oPPα (51 °C). The broad nature of the thermal melt is unlike the melt curves of other oPPα mutants. The different unfolding behaviour suggests that the interactions in oPPα-L@***a***-L@***g*** are potentially different to what has previously been observed.

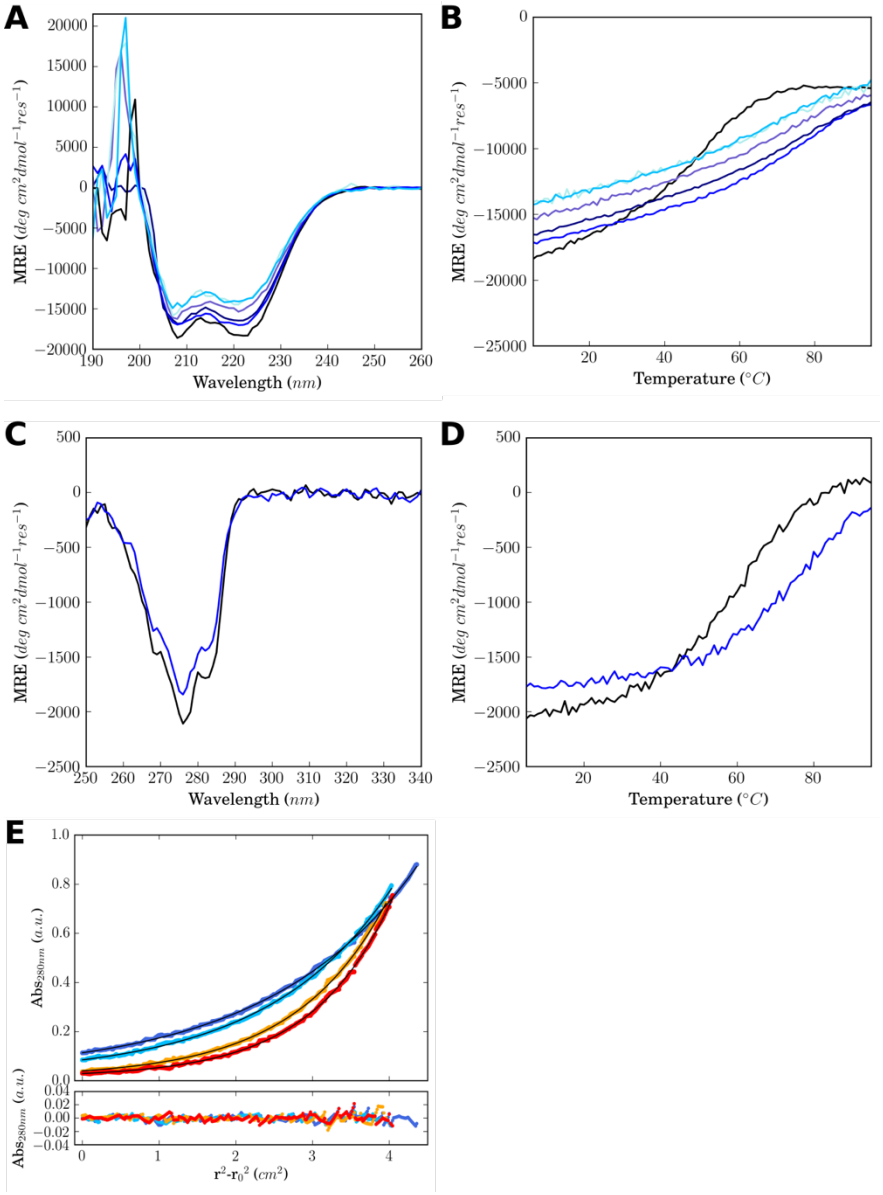


Figure 4-8: Folding and stability of oPPα-L@***a***-L@***g***. (A) Far-UV CD spectra recorded at 5 °C for oPPα, oPPα-L@***a***-L@***g***. (B) Temperature dependence of the far-UV CD signal monitored at 222 nm. For both (A) and (B) spectra overlaid for concentrations 5, 10, 50, 100 and 200 μM. The lines for each concentration range in colour from light blue (low concentration) to dark blue (high concentration). (C) Near-UV CD spectra recorded at 5 °C for oPPα, oPPα-L@***a***-L@***g***. (D) Temperature dependence of the near-UV CD signal monitored at 276 nm for oPPα and oPPα-L@***a***-L@***g***. (E) AUC Sedimentation-equilibrium data (top, dots) and fitted single-ideal species model curves (lines) at 44 (blue), 52 (aqua), 56 (orange), and 60 (red) krpm for oPPα-L@***g***-L@***a***. Bottom: Residuals for the above fits. oPPα-L@***a***-L@***g*** $\bar{v}$ = 0.754 cm^3 g^{-1} The fit returned a mass of 4912 Da (1.2 × monomer mass), 95% confidence limits = 4899 – 4926. Colour key: oPPα (black), oPPα-L@***a***-L@***g*** (blue). Conditions: 100 μM (CD Spectroscopy) and 130μM (SE AUC) peptide, PBS, pH 7.4.

To investigate this further, near-UV CD spectroscopy experiments were also performed (**Figure 4-8C&D**). The magnitude of the CD signal in the near-UV region is very similar compared to oPPα indicating similar stability at 5 °C. Monitoring the thermal transition showed unfolding and folding curves that were coincident with those recorded in the far-UV with a T_M of 76 °C.

The broad melt curve combined with the fully hydrophobic diamond hole interface perhaps suggests that there is non-specific association of the two helices. The two hydrophobic faces of the polyproline-II and α helices come together driven by the hydrophobic effect but perhaps do not interact so precisely as in PPα and oPPα. In addition, some possible self-association may be occurring.

To test this oligomerisation hypothesis, CD spectra were run at a range of different concentrations to determine whether there was any concentration dependence to the CD signal. Concentration dependence of the CD signal is indicative that the peptide could be oligomerising into a higher order oligomeric state.[226, 254-256] As well as running CD spectroscopy experiments at 100 μM, experiments were also run at 5, 10, 50 and 200 μM (**Figure 4-8A&B**). A range of peptide helicities and T_Ms were observed (**Table 4-2**). CD spectra at 5 and 10 μM were consistent with one another however were less folded compared to oPPα-L@***a***-L@***g*** at 100 μM exhibiting lower helicity at 5 °C and a lower T_M of 71 °C. The CD signal at 50 μM was between 10 and 100 μM with a T_M of 72 °C while at 200 μM the T_M value was 77 °C, slightly higher than at 100 μM. This concentration dependence of the CD signal is indicative that some peptide oligomerisation is taking place.

Table 4-2: Concentration dependence of CD signal for oPPα-L@**a**-L@**g**

Concentration / μM	**$MRE_{222\ nm}$ / deg cm^2 dmol^{-1} res^{-1}**	**% Helicity[175-176]**	**T_M / °C**
5	-14065	37	71
10	-14259	37	71
50	-15265	40	72
100	-17333	46	76 ± 1.2
200	-16525	43	77

To further examine the oligomeric state of oPPα-L@***a***-L@***g*** SE AUC experiments were performed (**Figure 4-8E**). SE AUC determined that the peptide had a mass of 1.2 × the mass of the monomer. This non-discrete value combined with the concentration dependence of the CD signal suggests that oPPα-L@***a***-L@***g*** is not fully monomeric and is in equilibrium with a higher order species, potentially a dimer.

4.2.3 Probing which position is more dominant: ***a*** or ***g***

Leu is conserved at the ***a*** position in related natural sequences. Leu at both ***a*** and ***g*** results in a higher midpoint of unfolding however the discrete monomeric nature of the miniprotein is lost. The broad transition in response to temperature may also be the result of non-specific association.

While it is clear Leu at both ***a*** and ***g*** is not preferable, the preference for one position over the other has not been investigated. To make a direct comparison between the preference for Leu at either the ***a*** or ***g*** positions the following two peptides were synthesised (**Figure 4-9**). Again, Glu was installed at the ***c*** position to maintain complementary electrostatic interactions between the two helices. PEK was also retained which limited mutations to the second and third heptads of the α helix. In oPPα-L@***a***-A@***g*** the residues are spaced $L_i \rightarrow Y_{i+4}$ whereas in oPPα-A@***a***-L@***g*** the$_y$ are spaced $Y_i \rightarrow L_{i+4}$.

Peptide name	Sequence and helical register
	321321 efgabcd efgabcd efgabcd
oPPα-L@***a***-A@***g***	Ac- PPKKPKKP GDNAT PEKLAEY EK**AL**AEY EK**AL**AEY-NH_2
oPPα-A@***a***-L@***g***	Ac- PPKKPKKP GDNAT PEKLAEY EK**LA**AEY EK**LA**AEY-NH_2

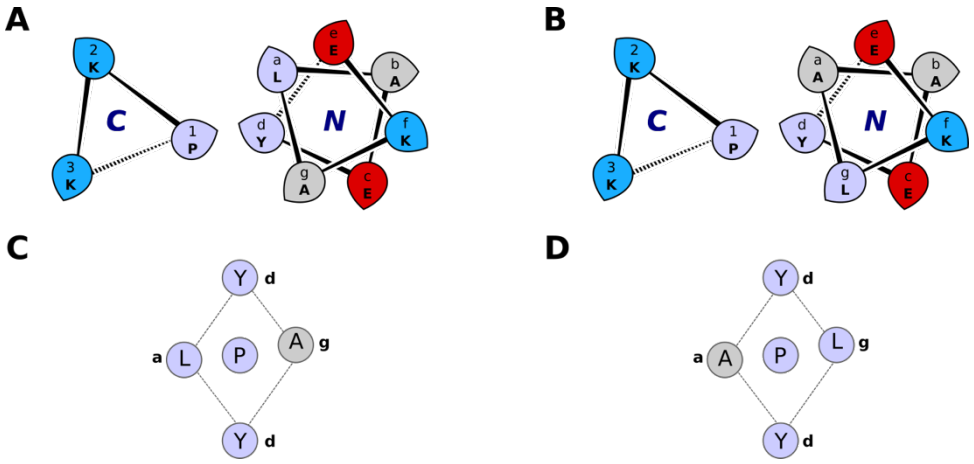


Figure 4-9: (Top) Peptides sequences for oPPα-L@***a***-A@***g*** and oPPα-A@***a***-L@***g***. (Bottom) Helical wheel representations of oPPα-L@a-A@g (A) and oPPα-L@***g***-A@***a*** (B). Diamond holes formed in the second and third heptads of oPPα-L@***a***-A@***g*** (C) and oPPα-A@***a***-L@***g*** (D).

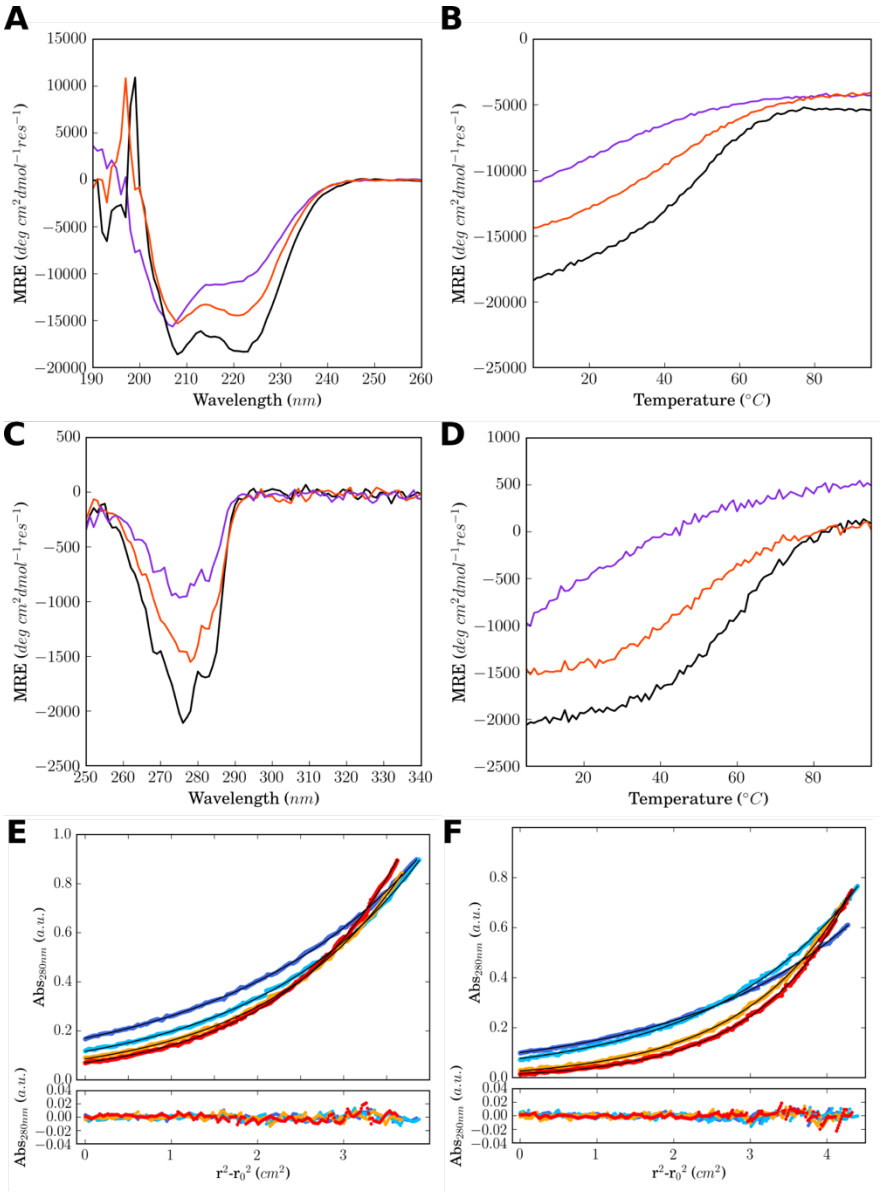


Figure 4-10: (A) Folding and stability of oPPα-L@***a***-A@***g*** and oPPα-A@***a***-L@***g***. (A) Far-UV CD spectra recorded at 5 °C for oPPα-L@***a***-A@***g*** and oPPα-A@***a***-L@***g***. (B) Temperature dependence of the far-UV CD signal monitored at 222 nm. (C) Near-UV CD spectra recorded at 5 °C for oPPα-L@***a***-A@***g*** and oPPα-A@***a***-L@***g***. (D) Temperature dependence of the near-UV CD signal monitored at 276 nm. (E&F) AUC Sedimentation-equilibrium data (top, dots) and fitted single-ideal species model curves (lines) at 44 (blue), 52 (aqua), 56 (orange), and 60 (red) krpm for oPPα-L@***g***-L@***a***. Bottom: Residuals for the above fits. (E) oPPα-L@***a***-A@***g*** $\bar{v} = 0.745$ cm^3 g^{-1}. The fit returned a mass of 4019 Da (1.0 × monomer mass), 95% confidence limits = 4006.9 – 4030.6. (F) oPPα-A@***a***-L@***g*** $\bar{v} = 0.745$ cm^3 g^{-1} The fit returned a mass of 4101 Da (1.1 × monomer mass), 95% confidence limits = 4089.0 – 4113.3. Conditions: 100 µM (CD Spectroscopy) and 130µM (SE AUC) peptide, PBS, pH 7.4. Colour key: oPPα (black), oPPα-L@***a***-A@***g*** (purple), oPPα-A@***a***-L@***g*** (red).

oPPα-L@***a***-A@***g*** was significantly less folded than oPPα at 5 °C with 29% helicity and a T_M of 21 °C (**Figure 4-10A&B**). Surprisingly, oPPα-A@***a***-L@***g*** was more folded compared to oPPα-L@***a***-A@***g*** (38 % helicity at 5 °C), albeit less so than oPPα . The thermal melt curve of oPPα-A@***a***-L@***g*** was slightly broader compared to oPPα and a lower T_M of 43 °C was calculated.

Near-UV CD spectra at 5 °C for oPPα-L@***a***-A@***g*** showed a less intense signal compared to oPPα-A@***a***-L@***g*** indicative of less rigid and more mobile Tyr side chains.[226] The thermal unfolding curves in the near-UV range for both were consistent with traces in the far-UV range and T_Ms were coincident (**Figure 4-10C&D**).

oPPα-L@***a***-A@***g*** and oPPα-A@***a***-L@***g*** were both monomeric in solution by SE AUC (**Figure 4-10E&F**) yielding masses of 1.0 and 1.1 × monomer mass.

These results suggest that, with Glu at ***c***, Leu is best placed at the ***g*** position compared to the ***a*** position. This is surprising given that Leu is very conserved at the ***a*** position in related natural folds. It is also surprising given that on mutating Leu at ***a*** to Ala it would be expected that this would result in the loss of CH-π interactions between Leu and Tyr that are observed in both PPα and oPPα and stabilise the structure. However other interactions must compensate for this.

4.2.4 Swapping the ***a*** and ***g*** positions

Based on this result, the effect of Leu at ***g*** in the context of oPPα was investigated by synthesising oPPα-***a*↔*g*** where the ***a*** and ***g*** positions were swapped (**Figure 4-11**). The ***e*** and ***c*** as well as the ***b*** and ***f*** positions were also swapped so as to keep both peptides equivalent and comparable. Introducing complementary or unfavourable charged interactions in the α helix of oPPα-***a*↔*g*** that were not observed in oPPα was avoided.

Peptide name	Sequence and helical register
	321321 efgabcd efgabcd efgabcd
oPPα	Ac- PPKKPKKP GDNAT PEKLAAY EKELAAY EKELAAY-NH_2
oPPα-***a*↔*g***	Ac- PPKKPKKP GDNAT PEKLAAY AALEKEY AALEKEY -NH_2

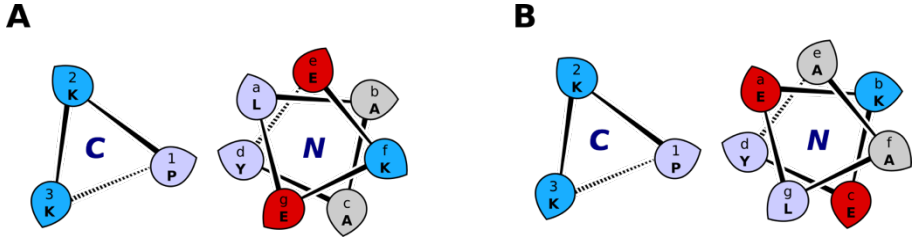


Figure 4-11: (Top) Peptide sequence of oPPα-***a*↔*g***. (Bottom) Helical wheel of oPPα (A) and oPPα-***a*↔*g*** (B).

oPPα-***a*↔*g*** was less folded at 5 °C compared to oPPα and yielded a T_M value of 19 °C, a substantial decrease compared to the T_M of oPPα. The near UV CD spectra was consistent with what is observed in the far-UV range. SE AUC experiments found oPPα-***a*↔*g*** to be monomeric in solution.

This result is inconsistent with what is observed when probing Leu at both ***a*** and ***g*** in isolation (oPPα-L@***a***-A@***g*** and oPPα-A@***a***-L@***g***), which showed that Leu at ***g*** appeared to provide greater stability compared to Leu at ***a***. This highlights the difficulties in probing non-covalent interactions even in the context of miniproteins. The diamond hole residues work in combination to provide stability to the peptide and looking at individual contributions of Leu at ***a*** or ***g*** does not accurately reflect the effect this has on stability in the context of oPPα. Comparing oPPα with oPPα-***a*↔*g*** confirmed that placing Leu at the ***a*** position and Glu at the ***g*** position is the favourable way round.

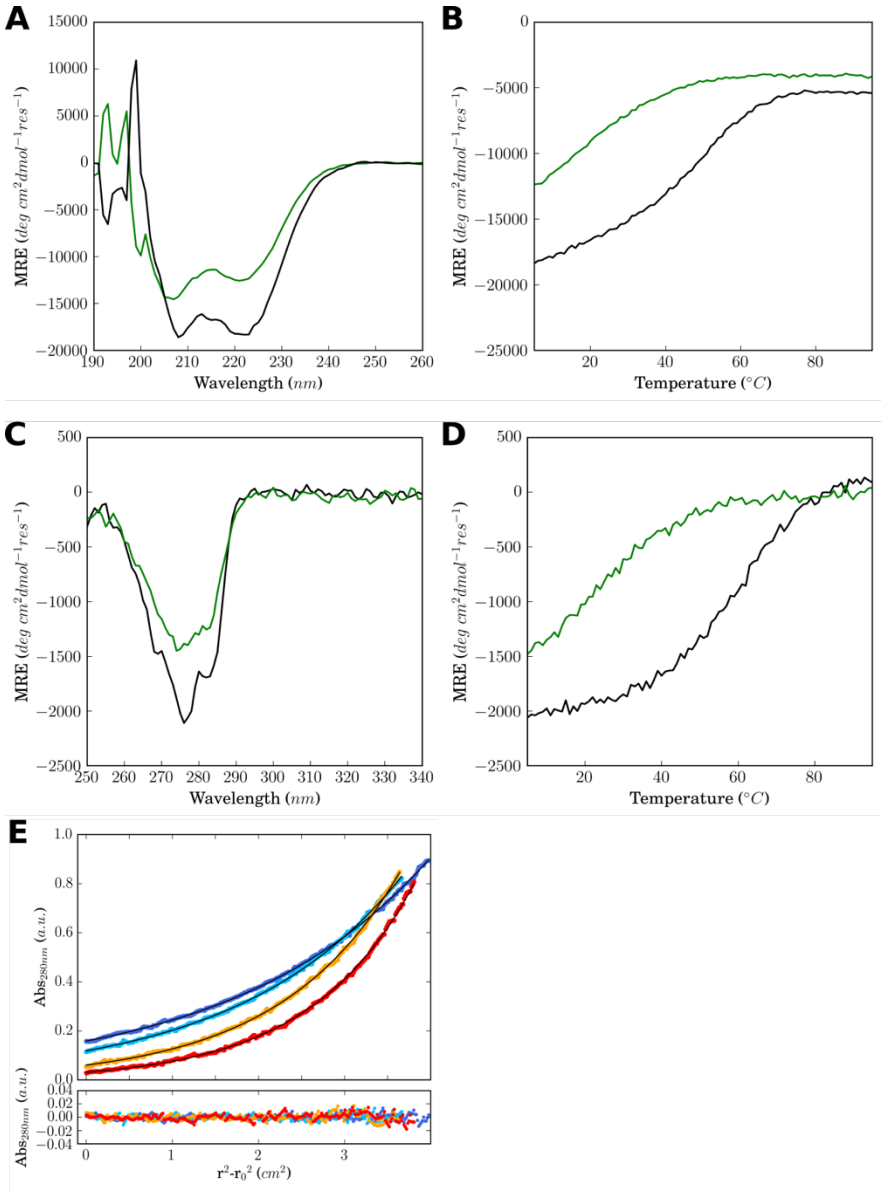


Figure 4-12: (A) Folding and stability of oPPα-***a↔g*** (A) Far-UV CD spectra recorded at 5 °C for oPPα-***a↔g*** and oPPα. (B) Temperature dependence of the far-UV CD signal monitored at 222 nm. (C) Near-UV CD spectra recorded at 5 °C for oPPα-***a↔g*** and oPPα. (D) Temperature dependence of the near-UV CD signal monitored at 276 nm. (E&F) AUC Sedimentation-equilibrium data (top, dots) and fitted single-ideal species model curves (lines) at 44 (blue), 52 (aqua), 56 (orange), and 60 (red) krpm for oPPα-***a↔g***. Bottom: Residuals for the above fits. (E) oPPα-***a↔g*** $\bar{v} = 0.748$ cm^3 g^{-1} The fit returned a mass of 4079 Da (1.1 × monomer mass), 95% confidence limits = 4065 – 4093. Conditions: 100 µM (CD Spectroscopy) and 130µM (SE AUC) peptide, PBS, pH 7.4. Colour key: oPPα (black), oPPα-***a↔g*** (green.

4.3 Molecular dynamics simulations of oPPα and variants

Obtaining experimental structures was not possible for all of the oPPα variants because of their reduced stabilities. Therefore, molecular dynamics (MD) simulations in water were used to explore the root causes of the difference in stabilities in the oPPα variants. MD simulations are one of the principle tools in the theoretical study of biological molecules to assess their time-dependent behaviours. MD simulations can provide information on protein stability, conformational changes, protein folding, molecular recognition and ion transport in biological systems.[257-258]

4.3.1 MD simulations of the X@a series of oPPα peptide

Point mutations were made to conformer eight of the NMR structure of oPPα in Pymol to generate models for the X@a mutants. A cubic periodic boundary box was set up and the box was filled with water molecules and 137 mM NaCl to mimic experimental conditions. Simulations were run using the Gromacs suite of tools and the Amber99sb-ILDN force field. After an initial relaxation and equilibration period each model peptide underwent a further 100 ns of MD. Simulations were performed in triplicate for each X@***a*** model from the same starting structure to ensure as much conformational space was sampled. However without infinite simulation time the quality of sampling in simulations will always be limited. Theoretically, with an infinitely long MD run time, the overall global thermodynamic energy minimum can be reached. Simulations were also run for the oPPα NMR structure (state 8).

All-atom RMSDs were calculated across the MD simulation; each structure from the trajectory was compared to the reference structure at the 0 ns timepoint. RMSDs were used to give an indication of how stable the model peptide was. RMSDs that fluctuate most across the time course were deemed less stable relative to RMSDs that did not fluctuate so much and were more consistent across the time course.

Firstly, the oPPα NMR structure was simulated and was found to be stable over three 100 ns simulations showing no significant fluctuations in the RMSD over the trajectory. For the oPPα X@***a*** variants, the RMSDs of the model structures through the MD trajectories correlated crudely with the observed experimental thermal stabilities. oPPα and oPPα-I@***a,*** which have the highest T_Ms, were stable with little fluctuation throughout all simulations as judged by RMSD. Whereas, all of the other variants deviated progressively from the initial structures as the simulations ensued as indicated by the increasing RMSD values (**Figure 4-13**).

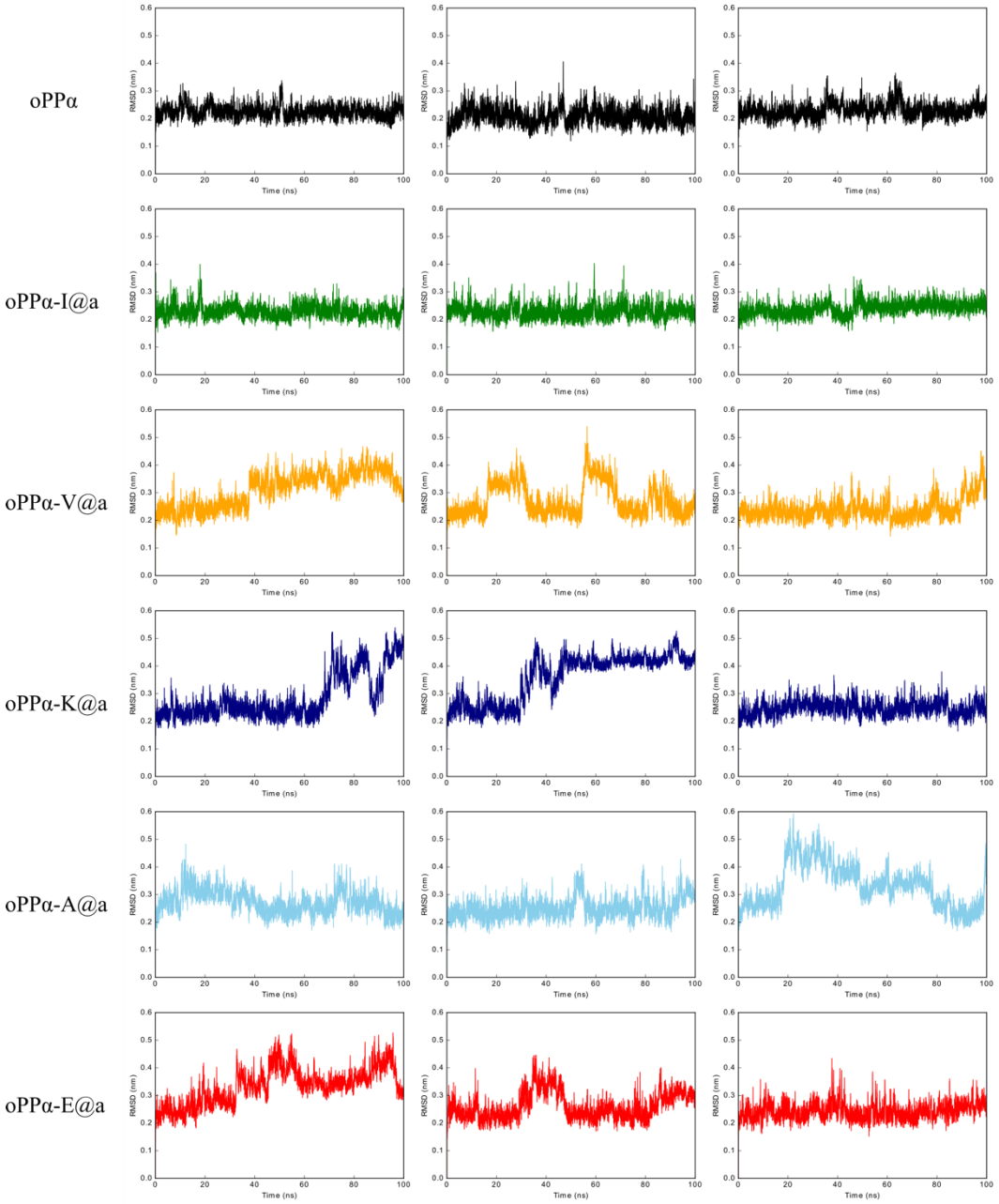


Figure 4-13: RMSD analyses of molecular dynamics trajectories. All atom RMSD vs. simulated time for oPPα-X@***a*** variants run in triplicate. RMSDs calculated relative to 0 ns time point. Colour key: oPPα (black), oPPα-I@***a*** (green), oPPα-V@***a*** (orange), oPPα-K@***a*** (blue), oPPα-A@***a*** (sky blue), oPPα-E@***a*** (red).

To quantitatively compare the MD structures, the RMSD between the 0 and 100 ns structures were calculated for each repeat, averaged and the standard deviation determined (**Figure 4-3**). Again, there was a rough trend in which the all atom and backbone RMSDs increase as the stability of the peptide decreases (i.e. as the T_M of the peptide drops). Also, as well as an increase in RMSD for the less stable peptides, there was an increase in the standard deviation too. This suggests that the less stable peptides are more dynamic.

Table 4-3: RMSD between 0 and 100 ns timepoints for oPPα-X@a Series.

	Backbone RMSD / Å	**All atom RMSD / Å**	**T_M / °C**
oPPα	1.28 ± 0.21	2.49 ± 0.26	51
oPPα-I@***a***	1.43 ± 0.10	2.62 ± 0.11	40
oPPα-V@***a***	2.44 ± 0.19	3.31 ±0.25	34
oPPα-K@***a***	2.80 ± 1.22	3.82 ± 1.17	19
oPPα-A@***a***	2.43 ±0.26	3.45 ± 0.38	-
oPPα-E@***a***	2.78 ±1.02	3.78 ± 0.97	-

As well as calculating RMSDs between the 0 and 100 ns structures, root mean square fluctuations (RMSFs) were also determined. The RMSF calculates, for each residue, the fluctuation about the average position. While the average structure is not that informative the fluctuations gives insight into the flexibility of different parts of the peptide and correspond to crystallographic b-factors. The RMSF and average structure for each mutant in the X@a series are represented in **Figure 4-14**. The colours of the structures are distributed over the b-factors values where blue represents the most stable and red represents the most fluctuating regions. The bulges in the cartoons emphasise regions of high fluctuation in b factors in the structure.

For each oPPα mutants the interface ***a***, ***d*** and ***g*** positions were the most stable, in particular the residues Tyr20, Tyr27, X@a24, and Pro5 showed the least fluctuation around the average structure and were the most rigid. For oPPα and oPPα-I@***a***, flexibility was limited mainly to the loop region and all Tyr, Leu and Pro residues were rigid. oPPα-V@***a*** showed some fraying towards the *N*-terminus of the α helix. Significant fraying of both the *N* and *C* terminus was observed for oPPα-K@***a*** and oPPα-E@***a***. Notably, only slight fraying of the *C*-terminus of oPPα-A@***a*** was observed however large fluctuations in the loop region and the Pro8 residue were noted.

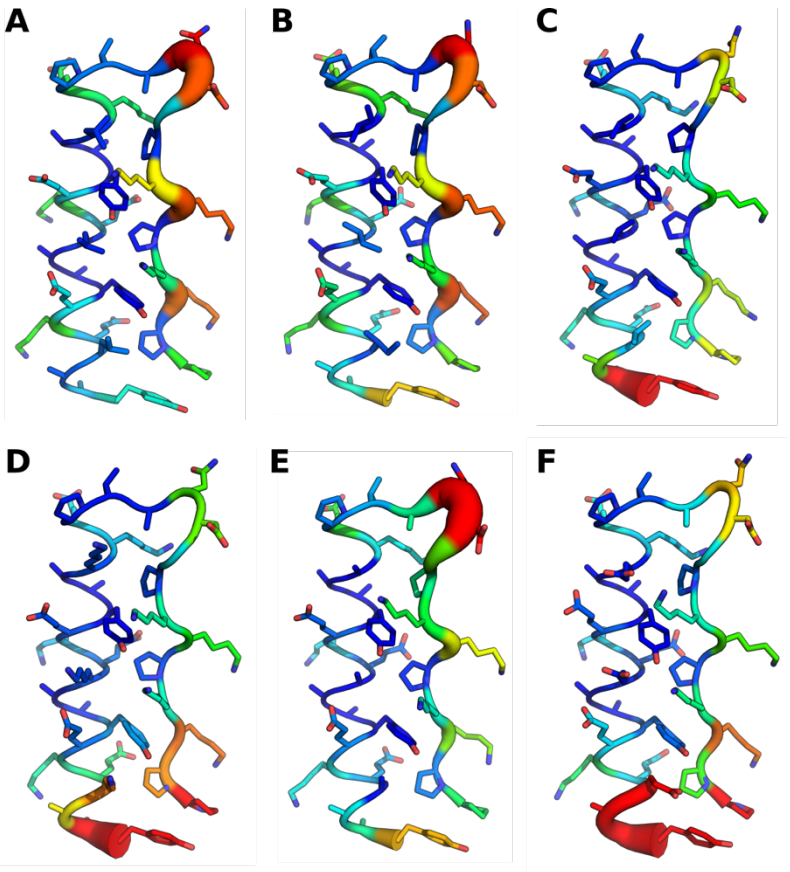


Figure 4-14: Average structures of each X@a mutant highlighting RMSFs. (A) oPPα (B) oPPα-I@a (C) oPPα-V@a (D) oPPα-K@a (E) oPPα-A@a (F) oPPα-E@a. Colour gradient: blue (most stable) through to red (most fluctuating).

Snapshots of each mutant were also overlaid at 0 and 100 ns timepoints to compare the gross change in structure from the start to the end of the simulation (**Figure 4-15**). For oPPα and oPPα-I@***a*** the snapshots were very similar and the 100 ns structures overlaid well with the initial structures. oPPα-V@***a*** overlaid well however slight fraying of the *C*-terminus of the α helix was observed, consistent with the RMSF. Significant fraying of both the *N*-terminus of the PPII helix and *C*-terminus of the α helix was observed for oPPα-K@***a*** and oPPα-E@***a***. No KIH interaction between Pro and the ***dagd*** residues is observed in the final heptad of the oPPα-E@***a*** structure. Also bulging of the α helix in oPPα-E@***a*** was observed across the trajectories. While, only slight fraying of the *C*-terminus of oPPα-A@***a*** is observed unfolding of the *N*-terminal α-helix towards the loop region is seen.

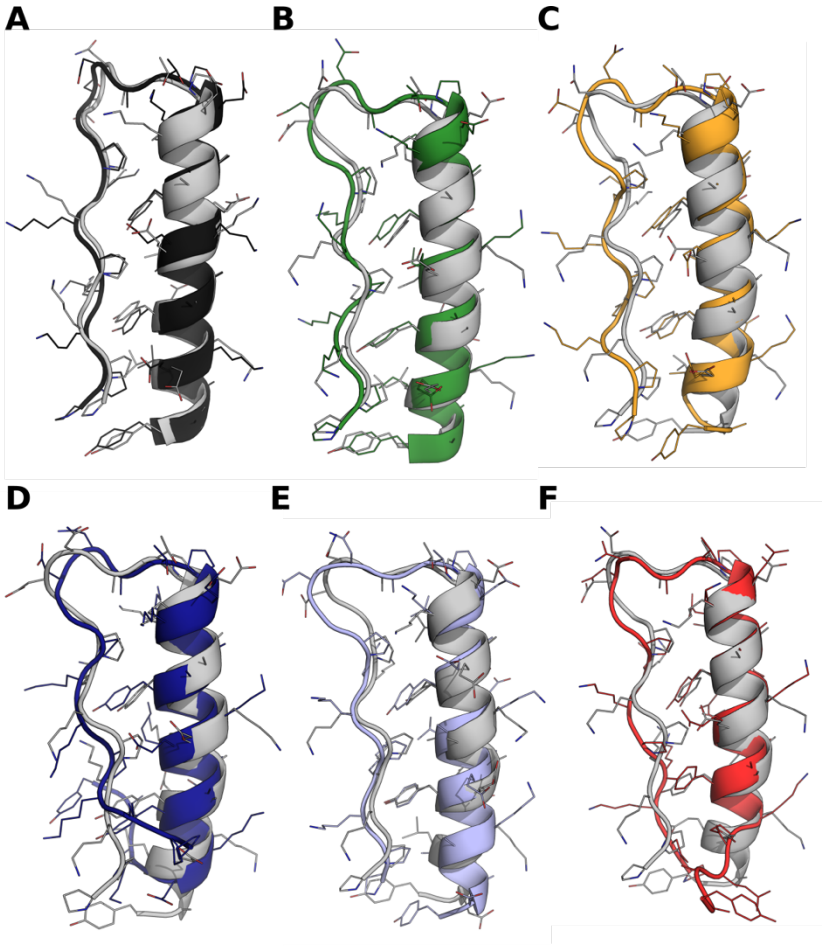


Figure 4-15: Overlays of the 0 and 100 ns snapshots from the MD simulations of the oPPα-X@***a*** mutants. (A) oPPα (B) oPPα-I@***a*** (C) oPPα-V@***a*** (D) oPPα-K@***a*** (E) oPPα-A@***a*** (F) oPPα-E@***a***. Structures shown in grey are of 0 ns timepoints.

Figure 4-16 shows the diamond shaped holes of the second (middle) heptad of the oPPα-X@***a*** variants at 0, 20, 40, 60, 80, and 100 ns timepoints. The overlays highlight the difference in stabilities of the diamond holes for each mutant model. oPPα and oPPα-I@***a*** show the least variation and residues at the ***d***, ***a***, ***g*** and ***l*** positions overlay relatively well with little change in conformation. Interestingly, while there is some more variation in the conformation of the Tyr side chains for oPPα-K@***a*** and oPPα-E@***a*** compared to the other mutants, there is not a significant difference in the diamond holes for all mutants.

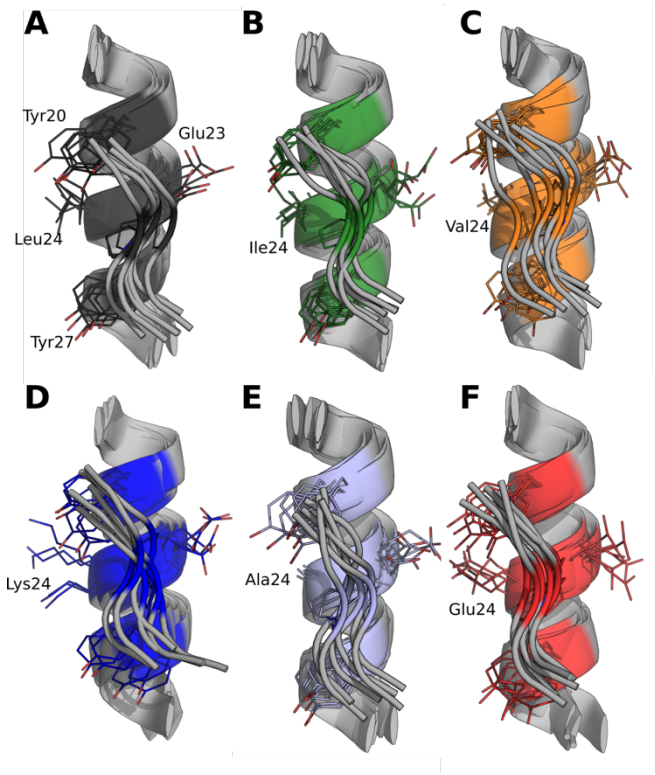


Figure 4-16: *In silico* models for designed oPPα variants after 100 ns of molecular dynamics simulation in water. Structures highlight ***a*** position mutations and corresponding 'knobs-into-holes packing' for the middle heptad within the α-helix sequence. (A) oPPα, (B) oPPα-I@***a***, (C) oPPα-V@***a***, (D) oPPα-K@***a*** (red). (E) oPPα-A@***a*** (F) oPPα-E@***a***.

4.3.1.1 Correlation between the BUDE score and T_M in oPPα-X@a mutants

BUDE (Bristol University Docking Engine) predicts the free energy of binding (kJ mol^{-1}) between two molecules using an empirical free-energy force field and can perform protein-protein docking in real space, virtual screening and ligand binding site identification on protein surfaces.[140, 259-260] It was hypothesised that there may be a correlation between the BUDE[259-260] score of the peptide after MD and the stability (T_M) of the peptide. To assess the fit of the polyproline-II helix with the α helix the former was treated as the ligand and the latter as the receptor. It was hypothesised that peptides which had a higher T_M and were deemed more thermally stable would also exhibit a better (i.e. more negative) BUDE score. BUDE scores for each oPPα-X@***a*** model was implemented within the ISAMBARD framework.[142]

The BUDE scores for the oPPα-X@***a*** series are shown in **Figure 4-17**. Unfortunately, there appears to be no correlation between BUDE score and peptide stability; all models have very similar BUDE scores. It appears that the subtle differences in the ***a*** position are not significant enough for BUDE to accurately determine relative stabilities. BUDE overweights charge-charge interactions and so ***a*** mutations may not be enough to disentangle these non- covalent interactions.

	T_M / °C	BUDE score before MD	Average BUDE score after MD
oPPα	51	-391.9	-369.8 ± 4.5
oPPα-I@***a***	40	-377.4	-366.4 ± 13.5
oPPα-V@***a***	34	-375.7	-341.7 ± 16.0
oPPα-K@***a***	19	-385.8	-374.7 ± 15.8
oPPα-A@***a***	-	-359.0	-367.8 ± 18.1
oPPα-E@***a***	-	-374.2	-373.0 ± 43.7

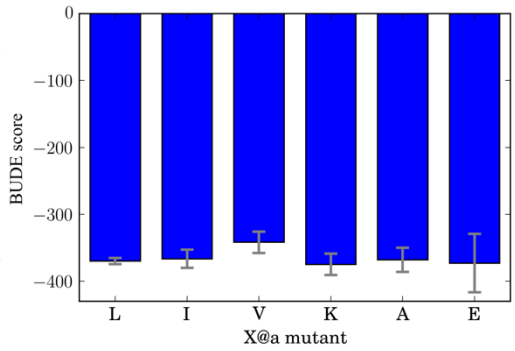


Figure 4-17: (Top) Table of BUDE scores of the oPPα-X@***a*** models before and after 100 ns of MD. (Bottom) Bar chart showing BUDE score after MD for each oPPα-X@***a*** model.

4.3.2 MD simulations examining Leu at ***a*** and ***g*** positions

Point mutations were made to the NMR structure of oPPα in Pymol to also generate models for oPPα-L@***a***-L@***g***, oPPα-L@***a***-A@***g***, oPPα-A@***a***-L@***g*** and oPPα-***a*↔*g***. Simulations were performed in triplicate for each model and all atom RMSDs were calculated across the MD simulations (**Figure 4-18**).

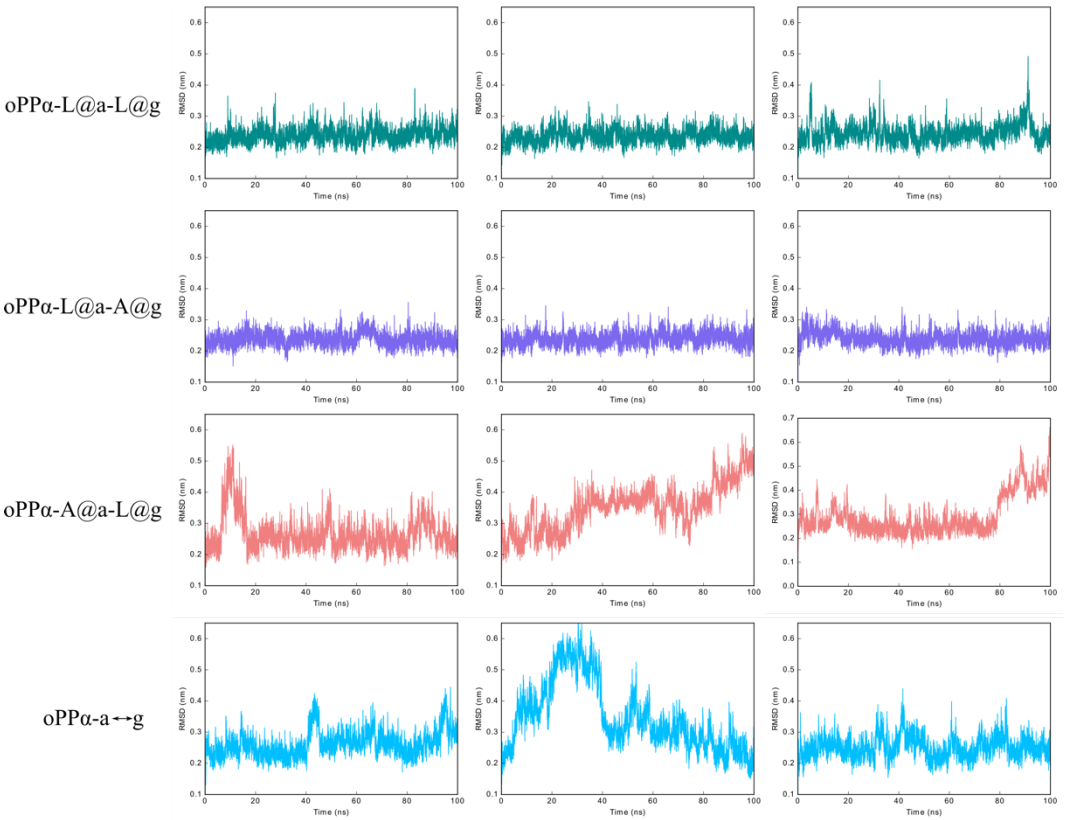


Figure 4-18: RMSD analysis of molecular dynamics trajectories. RMSDs calculated relative to 0 ns time point. Colour key: oPPα-L@***a***-L@***g*** (teal), oPPα-L@***a***-A@***g*** (purple), oPPα-A@***a***-L@***g*** (pink) and oPPα-***a*↔*g*** (sky blue).

There were no significant fluctuations in the RMSD over the time course for both oPPα-L@***a***-L@***g*** and oPPα-L@***a***-A@***g*** and this is reflected in the low all atom RMSDs between the 0 and 100 ns time points; 2.92 ± 0.11 and 2.58 ± 0.40 respectively (**Table 4-4**). Whereas oPPα-A@***a***-L@***g*** and oPPα-***a*↔*g*** showed much greater variation across the ensembles with all atom RMSDs of 4.51 ± 1.61 and 2.91 ± 0.70 respectively. In particular the high standard deviation in the RMSD for oPPα-A@***a***-L@***g*** reflects the less stable nature of the structure.

Table 4-4: Comparison of average RMSDs between 0 and 100 ns timepoints.

	Backbone RMSD / Å	**All atom RMSD / Å**
oPPα-L@***a***-L@***g***	1.75 ± 0.10	2.92 ± 0.11
oPPα-L@***a***-A@***g***	1.57 ± 0.50	2.58 ± 0.40
oPPα-A@***a***-L@***g***	3.36 ± 1.83	4.51 ± 1.61
oPPα-***a*↔*g***	1.51 ± 0.45	2.91 ± 0.70

The RMSF and average structure were again determined and are shown in **Figure 4-19**. For oPPα-L@***a***-L@***g*** and oPPα-L@***a***-A@***g*** fluctuations were limited mainly to the loop region. The interface ***a***, ***d*** and ***g*** positions were the most stable. oPPα-A@***a***-L@***g*** and oPPα-***a*↔*g*** showed significant fraying at both the *N* and *C* termini.

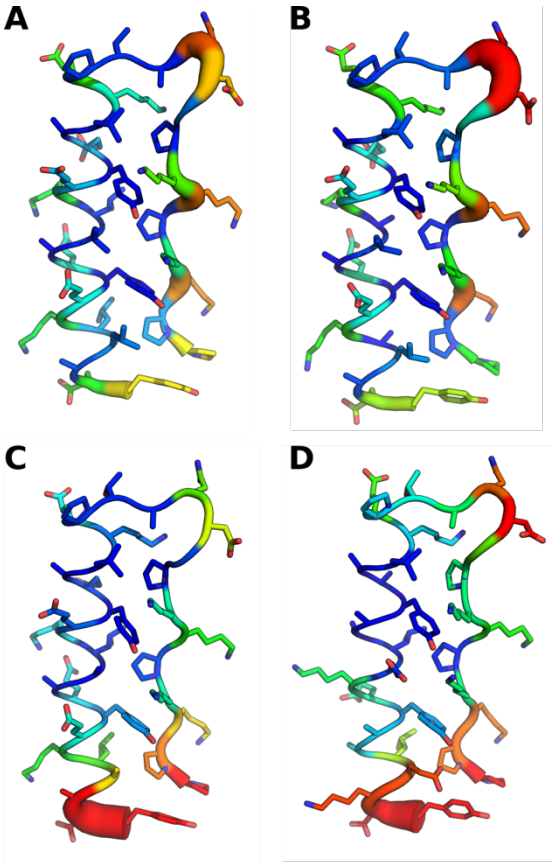


Figure 4-19: Average structures of each mutant highlighting RMSFs. (A) oPPα-L@***a***-L@***g*** (B) oPPα-L@***a***-A@***g*** (C) oPPα-A@***a***-L@***g*** (D) oPPα-***a*↔*g*** Colour gradient: blue (most stable) through to red (most fluctuating).

Snapshots of each mutant were overlaid at 0 and 100 ns timepoints to compare how the structures changed overtime. For oPPα-L@***a***-L@***g*** and oPPα-L@***a***-A@***g*** the 0 and 100 ns structures overlaid well whereas there was much more deviation as a result of fraying termini for oPPα-A@***a***-L@***g*** and oPPα-***a*↔*g***, consistent with the RMSFs.

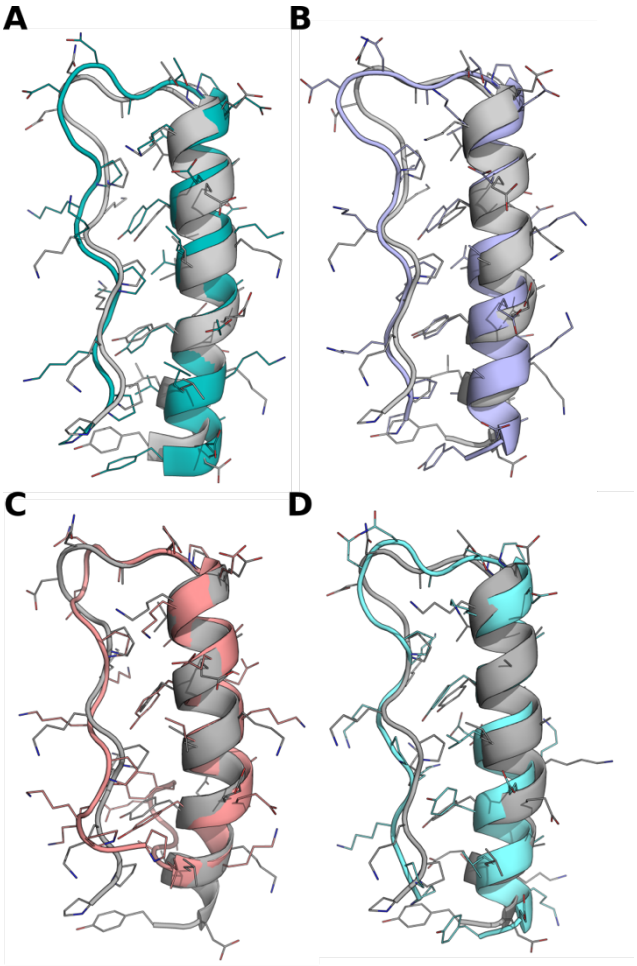


Figure 4-20: Overlays of the 0 and 100 ns snapshots from MD simulations. (A) oPPα-*L*@***a***-*L*@***g*** (B) oPPα-L@***a***-*A*@***g*** (C) oPPα-*L*@***a***-*L*@***g***. (D) oPPα-***a*↔*g***. Structures shown in grey are of 0 ns timepoints.

Figure 4-21 shows the diamond shaped holes of the second (middle) heptad of the oPPα variants at 0, 20, 40, 60 , 80 and 100 ns timepoints. The overlays highlight the difference in stabilities of each mutant model. oPPα-L@***a***-L@***g*** shows the least variation and residues at the ***d***, ***a***, ***g*** and ***1*** positions overlay relatively well with little change in conformation. Overall the time points for oPPα-L@***a***-A@***g*** overlay well, there is more movement in the polyproline-II helix across the α helix compared to oPPα-L@***a***-L@***g***. oPPα-A@***a***-L@***g*** and oPPα-***a*↔*g*** show the most variation between timepoints. There is substantial variation in the conformation of the Tyr residues for oPPα-A@***a***-L@***g*** and both show significant movement of the polyproline-II helix.

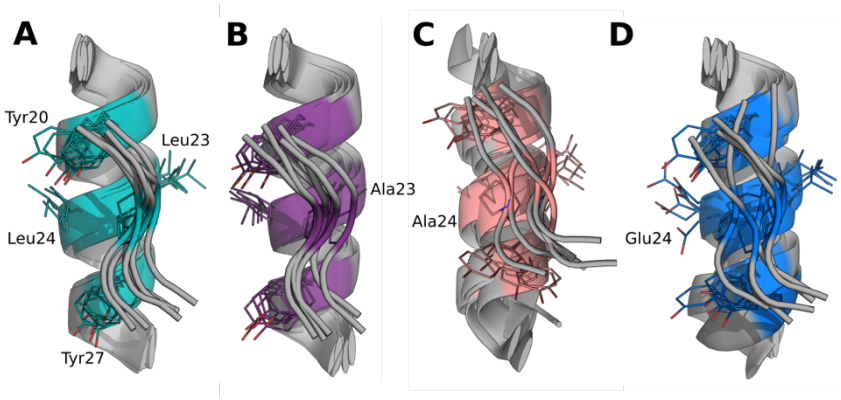


Figure 4-21: Overlay of MD structures at 0, 20, 40, 60, 80 and 100 ns time points. Structures highlight 'knobs-into-holes packing' for the middle heptad within the α-helix sequence. (A) oPPα-*L*@***a***-*L*@***g***, (B) oPPα-*L*@***a***-*A*@***g*** (C) oPPα-*A*@***a***-*L*@***g*** (D) oPPα-***a*↔*g***.

Table 4-5: Combined solution-phase biophysical characterisation of oPPα and mutants.

Entry	Name	Sequence and register 321321 efgabcd efgabcd efgabcd	MRE_{222} (deg cm^2 $dmol^{-1}$ res^{-1}) [a]	% Helicity	T_M / °C [b]	Oligomeric State [c]
1	Parent PPα	Ac- PPTKPTKP GDNAT PEKLAKY QADLAKY QKDLADY $-NH_2$	-18319	48	39	0.9
2	oPPα	Ac- PPKKPKKP GDNAT PEKLAAY EKELAAY EKELAAY $-NH_2$	-18319	48	51 ± 0.8	1.0
3	oPPα-E↔K	Ac- PPEEPEEP GDNAT PEKLAAY KKKLAAY KKKLAAY $-NH_2$	-17810	47	40 ± 2.1	0.9
4	oPPα-Phe	Ac- PPKKPKKP GDNAT PEKLAAF EKELAAF EKELAAF $-NH_2$	-18105	47	33 ± 0.6	1.0
5	oPPα-I@***a***	Ac- PPKKPKKP GDNAT PEKIAAY EKEIAAY EKEIAAY $-NH_2$	-16493	43	40 ± 0.1	1.0
6	oPPα-V@***a***	Ac- PPKKPKKP GDNAT PEKVAAY EKEVAAY EKEVAAY $-NH_2$	-14854	37	34 ± 1.4	1.0
7	oPPα-K@***a***	Ac- PPKKPKKP GDNAT PEKKAAY EKEKAAY EKEEAAY $-NH_2$	-12448	33	19 ± 2.3	1.0
8	oPPα-A@***a***	Ac- PPKKPKKP GDNAT PEKAAAY EKEAAAY EKEKAAY $-NH_2$	-8522	23	-	1.0
9	oPPα-E@***a***	Ac- PPKKPKKP GDNAT PEKEAAY EKEEAAY EKEEAAY $-NH_2$	-1642	5	-	-
10	oPPα-L@***a***-L@***g***	Ac- PPKKPKKP GDNAT PEKLAEY EKLLAEY EKLLAEY $-NH_2$	-17333	45	76 ± 1.2	1.2
11	oPPα-L@***a***-A@***g***	Ac- PPKKPKKP GDNAT PEKLAEY EKALAEY EKALAEY $-NH_2$	-10849	29	21	1.0
12	oPPα-A@***a***-L@***g***	Ac- PPKKPKKP GDNAT PEKLAEY EKLAAEY EKLAAEY $-NH_2$	-14365	38	43	1.1
13	oPPα-***a***↔***g***	Ac- PPKKPKKP GDNAT PEKLAAY AALEKEY AALEKEY $-NH_2$	-12477	33	19 ± 1.3	1.1

[a] MRE_{222} from CD spectroscopy at 5 °C and 100 μM peptide in PBS. [b] The point of inflection in sigmoidal thermal denaturation curves, calculated from the maxima of the first derivative of the thermal transition. [c] Oligomeric state calculated from SE AUC experiments.

4.4 Chapter conclusion

oPPα is stabilised by KIH-like packing between the α and polyproline-II helices. Pro residues on the polyproline-II helix dock into diamond shaped holes formed by successive ***dagd*** residues on the α helix. Mutations were made to the diamond hole to determine sequence to stability relationships for oPPα.

Firstly, the ***a*** position was mutated to a series of hydrophobic, polar and charged residues. Ile and Val are tolerated at the ***a*** position yielding well folded peptides with T_M values of 43 and 37 °C respectively. Lys at ***a*** is tolerated to some extent while Ala and Glu are not tolerated. Therefore, as observed in related natural peptides Leu at ***a*** appears to be the most favoured.

A fully hydrophobic diamond hole was also probed by placing Leu at both ***a*** and ***g***. The behaviour of this variant in response to temperature was altered and gave a broad CD melt curve with a T_M of 76 °C. Concentration dependence of the CD signal was indicative of peptide oligomerisation which was confirmed by SE AUC returning a molecular mass equivalent to 1.2 × the monomer mass of the peptide. Collectively this indicated that non-specific association of the helices is occurring, consistent with a broader hydrophobic seam.

A direct comparison between the preference for Leu at either the ***a*** or ***g*** positions was made by comparing oPPα-L@***a***-A@***g*** with oPPα-A@***a***-L@***g***. Despite nature selecting Leu at the ***a*** position oPPα-A@***a***-L@***g*** was more stable with a higher T_M value. This led to investigating the effect of Leu at ***g*** in the context of oPPα with Glu at the ***a*** position; the ***a*** and ***g*** positions were switched. oPPα-a↔g was substantially less folded than oPPα with a T_M of 19 °C. This result suggests that the diamond hole residues work in combination to provide stability to the peptide and observing individual contributions of Leu at ***a*** or ***g*** does not accurately reflect the effect this has on stability in the context of oPPα.

MD studies of the X@***a*** peptide series show a crude correlation between T_M and RMSD across the trajectory. Analysis of the MD simulations revealed that instability of the model peptide mutant mainly manifested itself as fraying of the *N* and *C* termini. While MD is a powerful computational technique for studying biological systems there are some limitations to it. Firstly, the output of a simulation is the result of the atomic force field and so results will only be realistic if the force field mimics the forces experienced by 'real' atoms. Designing a good forcefield is very challenging. A balance must be struck between the accurate representation of real atoms within a forcefield, a simple enough forcefield to speed up evaluation, and a forcefield that is applicable to many systems. While the AMBER forcefield is accepted as one of the best forcefields for studying proteins and nucleic acids, there is still room for improvement. Secondly, time limitation is a major problem with MD simulations. It is currently not feasible to run simulations longer than the nanosecond timescale using standard MD simulations. This may not be long enough to observe the process of protein folding.

Overall, Leu is the most preferred residue at the ***a*** position, however Ile, Val and Lys are also tolerated. Preference for Leu at the ***a*** position over the ***g*** position in oPPα is also observed. A fully hydrophobic diamond hole results in increased thermal stability of the peptide at the cost of a discrete monomeric structure.

The optimised and thermally stable PPα structure and the sequence-to-stability relationships determined herein can now be used as a foundation on which to build larger and more complex assemblies based on the PPα fold.

Chapter 5 Towards larger oPPα topologies

5.1 Chapter introduction

The previous chapters have described the successful redesign of PPα and subsequent sequence-to-stability studies for the optimised topology. This chapter describes attempts to use this optimised PPα as a building block in the construction of larger and more-complex *de novo* assemblies. Towards this goal, expanding the PPα assembly to incorporate additional secondary structure elements and to give new tertiary folds with PPαPP or αPPα topologies were explored (**Figure 5-1**).

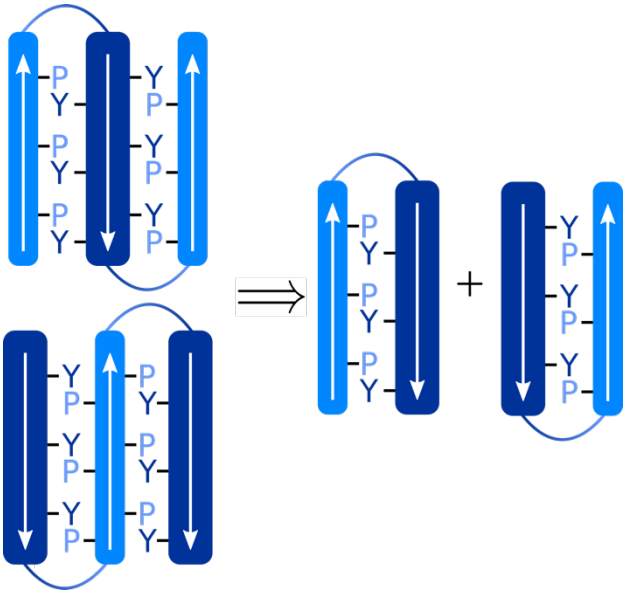


Figure 5-1: Cartoon representations of the PPαPP and αPPα folds broken down into their constituent components: PPα and αPP. Colour key: Polyproline-II helix (light blue), α helix (dark blue). Arrows represent orientation of helices.

We hypothesised that optimal PPα and αPP components might be combined to give these constructs. With an optimised PPα component already in hand, efforts turned towards designing an αPP component. The αPP topology differs from the oPPα topology since the α helix is *N* terminal in the former. We anticipated that finding an appropriate loop to join the α and polyproline-II helices together would be important. To test this hypothesis the two secondary structure components of oPPα were synthesised individually and their interactions investigated (**Figure 5-2A**). CD spectroscopy was performed on a 1:1 mix of the peptides corresponding to the α and polyproline-II helices and the resultant CD spectra were compared to the theoretical average spectrum of the non-interacting helices (**Figure 5-2B&C**).

If the α and polyproline-II helices were not interacting then the resultant CD spectrum at 5 °C would overlay exactly with the theoretical average spectrum of the two individual helices. The polyproline-II helix CD spectrum resembled that of a polyproline-II helix with a minimum at approximately 197 nm and a maximum at approximately 220 nm while the α helix was 38 % α-helical. The resultant CD spectrum of the 1:1 mixture gave an α helicity of 34 % while the theoretical average of the non-interacting helices was 26 % α-helical (oPPα was 48 % folded at the same concentration). The thermal melt curves for the 1:1 mix of the α and polyproline-II helical components and the theoretical average of the two were also very similar. Together, these results indicated that in the absence of the loop, the peptides corresponding to the α and polyproline-II helices showed very limited interaction and the α:polyproline-II complex was much less folded and stable than the single chain oPPα. Furthermore, this demonstrated that selecting an appropriate loop to join the secondary structure components together was going to be important.

A

Peptide name	Sequence and helical register
	efgabcd efgabcd efgabcd
α component	Ac-PEKLAAY EKELAAY EKELAAY-NH_2
Polyproline-II component	Ac-PPKKPKKPG-NH_2

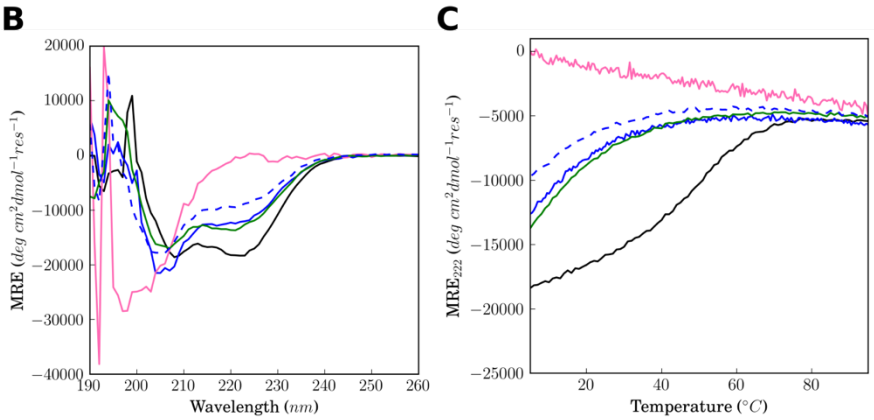
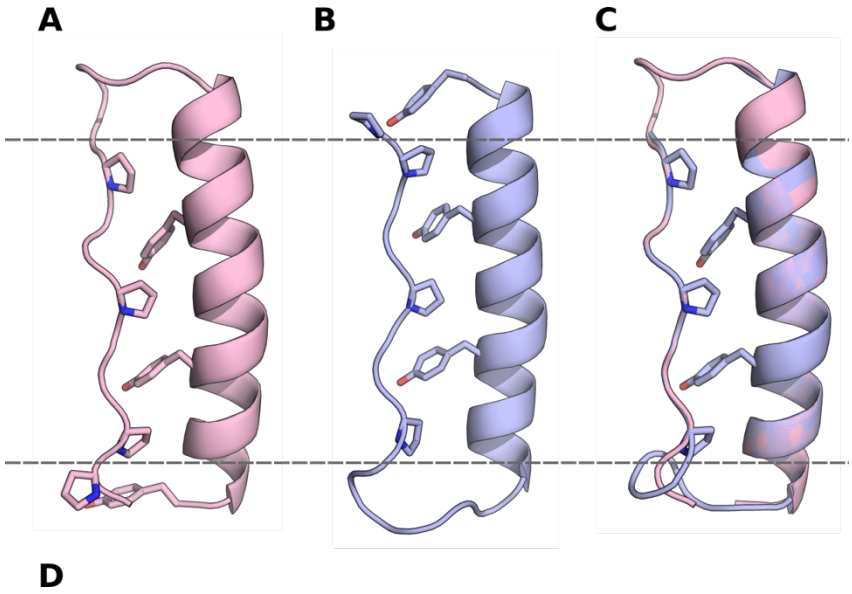


Figure 5-2: CD spectra of α and polyproline-II helices. (A) Sequences of α and polyproline-II helical components. (B) CD spectra at 5 °C for α helix, polyproline-II helix, 1:1 mix of α and polyproline helices and theoretical average of non-interacting helices. (C) Temperature dependence of the CD signal monitored at 222 nm. Conditions: 100 μM each peptide, PBS, pH 7.4. Colour key: α helix component (green), PPII component (pink), 1:1 mix of the peptides corresponding to the α and PPII helices (blue), theoretical mix of helices (blue dash), oPPα (black).

5.2 Designing an αPP topology

The αPP topology was designed to maintain the interface between the two helices as in oPPα (**Figure 5-3A-C**). The NMR structure of oPPα revealed close contacts between the *N*-terminal PP-cap and *C*-terminal Tyr, therefore a PP-cap was installed at the now *C*-terminal polyproline-II helix

and Tyr at the *N*-terminal α helix. This required the α-helical register to be shifted by one residue to maintain Tyr at **d** and Leu at **a**. The resulting heptad runs **d** through **c**. Loop design is notoriously difficult and therefore, as a starting point, the natural loop used in PPα was initially also used for αPP to give αPP-GDNAT (**Figure 5-3D**).[223]



Peptide name	Sequence and helical register
	321321 efgabcd efgabcd efgabcd
oPPα	Ac- PPKKPKKP GDNAT PEKLAAY EKELAAY EKELAAY-NH_2
	defgabc defgabc defgabc defgabc 123123123
αPP-GDNAT	Ac-YEKELAA YEKELAA YEKELAA GDNAT PKKPKKPPG-NH_2
αPP-GDNAT-4	Ac-YEKELAA YEKELAA YEKELAA YEKELAA GDNAT PKKPKKPKKPPG-NH_2

Figure 5-3: (A) NMR structure state 1 of PPα (B) Model of αPP generated within the ISAMBARD framework. (C) Overlay of PPα and αPP. The region between the two black dashed lines represent the interface region that is the same in both structures. The region above the top grey line and below the bottom grey line are different between the oPPα and αPP topologies. (D) Sequences of initial αPP designs compared to oPPα.

Loop design is challenging for a number of reasons: firstly, the backbone hydrogen-bonding potential of a loop is not automatically satisfied as it is for α helices and β sheets.[223] For a loop to adopt a low-energy discrete conformation the backbone/side-chain polar groups need to be hydrogen bonded to other groups either in the loop, the rest of the protein, or be solvent exposed and able to hydrogen bond with water. However, if too much of the loop is exposed to water then the loop may not adopt a unique conformation. Secondly, unlike for α helices and β sheets where there are general rules and sequence-to-structure relationships for forming well-folded secondary structures,[83, 261-262] the relationships for forming an ordered loops are less clear. Most ordered loops have diverse sequences that form specific and hard-to-predict interactions within the loop and with the rest of the protein.[223, 263]

A Gly residues was included at the *C*-terminus of αPP so that a Pro residue was not adjacent to the Rink Amide resin during peptide synthesis. When Pro was positioned at the *C*-terminus a single Pro residue was deleted during synthesis.

The αPP-GDNAT peptide was first characterised by CD spectroscopy. αPP-GDNAT was significantly less folded than oPPα with an α-helicity of 32 % compared to 48 % for oPPα (**Figure 5-4A)**. Interestingly, the minimum at 208 nm is substantially lower than at 222 nm for αPP-GDNAT. As discussed in Section 2.3.1 a distinctive α-helical peptide is characterised by minima at 208 and 222 nm of approximately equal intensities. While the ratio of these two minima is sometimes used as an additional gauge of α-helicity the origin and effect of peptide sequence on this ratio remains ill-defined.[264-266] It is likely the difference in MRE values is linked to the less folded nature of the peptide and the increased random coil. αPP-GDNAT was also significantly less thermally stable than oPPα and a T_M could not be extracted from the data (**Figure 5-4B**). SE AUC data confirmed that the peptide was monomeric in solution (**Figure 5-4C**).

While the α helix:polyproline-II helix interface in αPP should be the same as in oPPα, the peptide is significantly less folded and stable. This suggests that although the α helix:polyproline-II helix interface provides the key driving force for peptide folding the role of the loop region is also critical for folding and cannot be overlooked. While the natural loop adopts conformations optimal for joining the secondary structure components in oPPα the loop is not optimised for αPP.

αPP-GDNAT was also extended and synthesised as a four heptad variant to determine the effect of increased length on the stability of αPP. αPP-GDNAT-4 was significantly more folded and thermally stable than αPP-GDNAT with an α-helicity of 50 % and a T_M of 49 °C. The stability of four heptad variant αPP-GDNAT-4 is comparable to oPPα (three heptads). Further, from work discussed in Section 6.2, αPP-GDNAT (three heptads) is comparable in stability to shortened oPPα variant oPPα-2 (two heptads). This further suggests that the GDNAT loop is not optimised for αPP. The first heptad after the loop potentially compensates for the suboptimal loop resulting in fraying or unfolding at the *C* terminus of the α helix component. Subsequently, this leads to loss in helicity and stability of the peptide topology. The effect of chain length on oPPα will be discussed in more depth in Section 6.2.

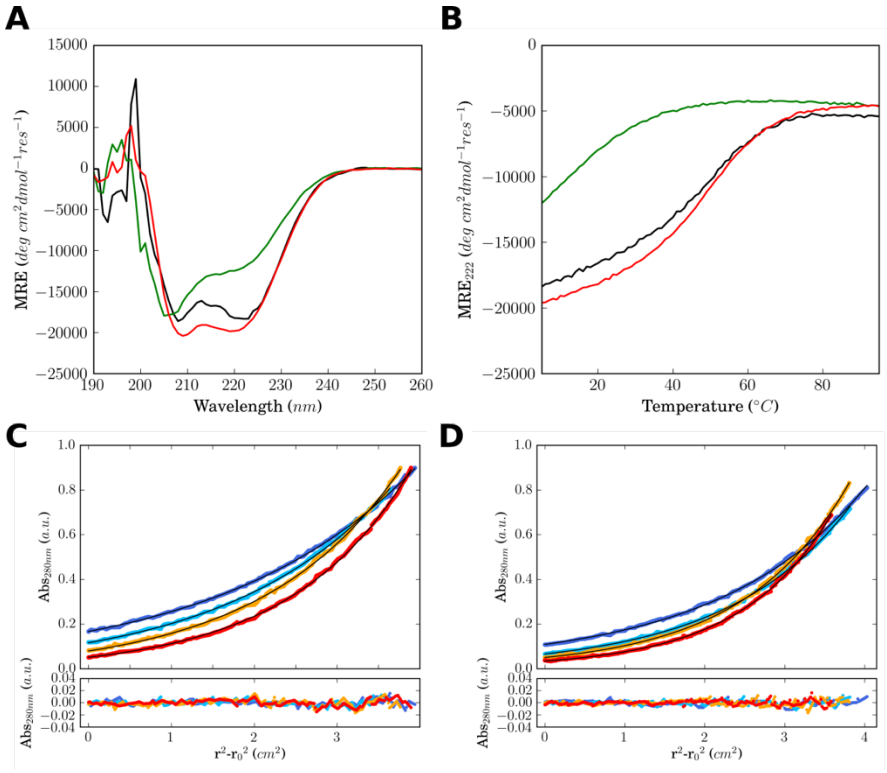


Figure 5-4: Folding and stability of αPP-GDNAT and αPP-GDNAT-4 compared to oPPα. (A) CD spectra recorded at 5 °C for αPP-GDNAT and αPP-GDNAT-4. (B) Temperature dependence of the CD signal monitored at 222 nm. (C-D) AUC Sedimentation-equilibrium data (top, dots) and curves fitted single-ideal species model (lines) at 44 (blue), 52 (aqua), 56 (orange), and 60 (red) krpm for αPP-GDNAT. Bottom: Residuals for the above fits. (C) αPP-GDNAT: $\bar{v} = 0.743$ cm^3 g^{-1}. The fit returned a mass of 3860 Da (1.0 × monomer mass), 95% confidence limits = 3849.7 – 3870.4 Da. (D) αPP-GDNAT-4: $\bar{v} = 0.747$ cm^3 g^{-1}. The fit returned a mass of 4979 Da (1.0 × monomer mass), 95% confidence limits = 4955.8 – 5001.8 Da. Conditions: 100 µM peptide, PBS, pH 7.4. Colour key: αPP-GDNAT (green), αPP-GDNAT-4 (red), oPPα (black).

As a control, a simple and more flexible Gly-Ser loop was selected of the same length to join the α and polyproline-II helices together (**Figure 5-5A**).[151, 267] αPP-GSGSG was slightly more folded than αPP-GDNAT with an α-helicity of 36 % however this was still not comparable to the folding observed for oPPα (**Figure 5-5B**). Further, an imbalance in the minima at 208 and 222 nm was observed. A T_M of 20 °C was extracted from the thermal unfolding curve of αPP-GSGSG, compared to 51 °C for oPPα (Figure 5-5C). SE AUC experiments revealed the αPP variant was monomeric in solution (**Figure 5-5D**).

A

Peptide name	Sequence and helical register
	defgabc defgabc defgabc 123123
αPP-GSGSG	Ac-YEKELAA YEKELAA YEKELAA GSGSG PKKPKKPPG-NH_2

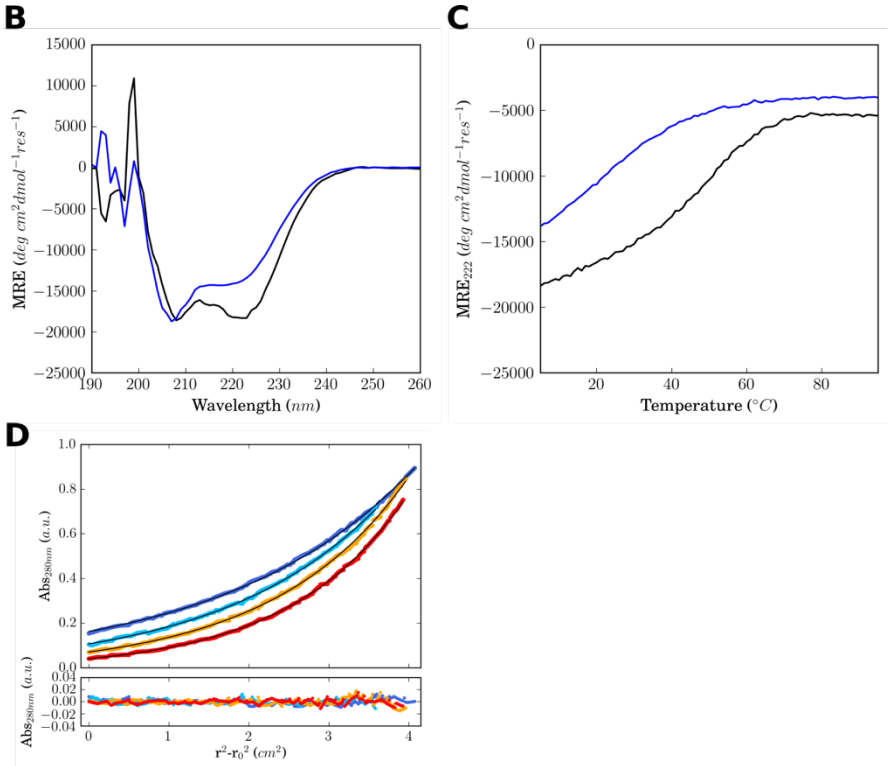


Figure 5-5: Sequence and biophysical characterisation of αPP-GSGSG compared to oPPα. (A) Sequence of oPPα-GSGSG. (B) CD spectra recorded at 5 °C for αPP-GSGSG. (C) Temperature dependence of the CD signal monitored at 222 nm. (D) AUC Sedimentation-equilibrium data (top, dots) and curves fitted single-ideal species model (lines) at 44 (blue), 52 (aqua), 56 (orange), and 60 (red) krpm for αPP-GSGSG. Bottom: Residuals for the above fits. $\bar{v}$ = 0.744 cm^3 g^{-1}. The fit returned a mass of 3864 Da (1.0 × monomer mass), 95% confidence limits = 3851.1 – 3877.5 Da. Conditions: 100 μM peptide, PBS, pH 7.4. Colour key: αPP-GSGSG (blue), oPPα (black).

While the flexible linker allowed the loop to adopt a broad range of conformations, including ones favourable for the antiparallel αPP topology, the lack of rigidity potentially meant the loop did not adopt a single unique conformation. The reduced folding of both αPP-GDNAT and αPP-GSGSG implied a more structured loop optimised for the αPP topology was required and thus a more sophisticated approach to loop design was needed.

5.2.1 Capping Motifs

There are other factors as well as an optimal loop that may help improve the stability of the αPP topology including capping motifs. We note that loops and capping motifs were investigated simultaneously but for the purpose of this chapter they are discussed separately.

Capping residues at the *N* termini of α helices are important for stabilising helix formation. A bioinformatics study of an early subset of α helices in the PDB shows preferences of amino acids for particular positions at the ends of α helices (**Figure 5-6**).[225] Specifically, there is a strong preference for Asn at the *N*-cap position. This is because Asn can form a hydrogen bond from its Oδ position to the backbone NH group of residue N3, which is the third residue after the *N*-cap and is exposed in the first turn of the α helix. There is also a strong preference for Pro at position N1. One reason for this is because the backbone dihedral angle (ϕ) of Pro is fixed at approximately 70° and therefore, with only one rotatable angle, the loss of entropy on forming an α helix structure is lower than for other amino acids.[224] The study notes side-chain hydrogen bonds to the backbone NHs of N2 and N3 but almost never to N1. It has been hypothesised that another important contribution to why Pro is preferred at N1 is simply to occupy a position which would otherwise have an exposed NH group.[268] Pro is often thought of as a "helix breaker" because it is not often found in the interior positions of α helices however, given its preference at position N1, a better description of Pro may be as a helix initiator.[268] A preference for charged residues at N2 is also noted, in particular Asp and Glu.[24] Overall, an important theme to help stabilise the terminal ends of α helices is the need to satisfy potential hydrogen bonds.

N-cap	**N1**	**N2**	**N3**	**N4**	**N5**	**mid**	**C-cap**
Asn	Pro	Asp/Glu	Glu/Asp	Leu	Lys	...	Gly

Figure 5-6: Summary of some of the position specific amino acid preferences in α helices. Top entry is the position in the α helix and the bottom entry is the amino acid preference at that position.

In αPP the *N*-cap of the α helix is Tyr, the N1 position is Glu and the N2 position is Lys. While Glu is the second-most preferred residue at the N1 position, Tyr and Lys are significantly less favourable.[225] Therefore, we propose that the addition of favourable residues at the *N*-terminus of the α helix of αPP may help stabilise the overall topology. A series of peptides were synthesised with various *N*-terminal capping residues added (**Table 5-1**).

Table 5-1: Sequences of αPP variants with various capping residues.

Peptide name	**Sequence and helical register**					
		defgabc	defgabc	defgabc		123123
αPP-NP	Ac-NP	YEKELAA	YEKELAA	YEKELAA	GDNAT	PKKPKKPPG-NH_2
αPP-NPE	Ac-NPE	YEKELAA	YEKELAA	YEKELAA	GDNAT	PKKPKKPPG-NH_2
αPP-NPELAA	Ac-NPELAA	YEKELAA	YEKELAA	YEKELAA	GDNAT	PKKPKKPPG-NH_2

αPP-NP was first synthesised with Asn at the *N*-cap position and Pro at the N1 position. This peptide was significantly more folded and thermally stable than the corresponding peptide with the same loop region (αPP-GDNAT) but without the Asn-Pro cap. αPP-NP had an α helicity of 39 % compared to 32 % for αPP-GDNAT and a measurable T_M of 15 °C. The next peptide synthesised in

the series, αPP-NPE, had a Asn-Pro-Glu cap. Glu was chosen at the N2 position over Asp since Glu is much more prevalent in the α-helix sequence of oPPα and has higher α-helical propensity generally.[209] This is despite the fact that Asp is the most preferred residue at this position with Glu the second most preferred.[225] This additional Glu residue enhanced the folding and stability further to an α-helicity of 42 % and a T_M of 22 °C.

In an attempt to make a seamless transition between the Asn-Pro-Glu capping residues and the first heptad in the αPP sequence the residues Leu-Ala-Ala were added after Asn-Pro-Glu. In effect, a truncated heptad (NPELAA) with enhanced helix capping potential was added onto the *N*-terminus of αPP. While this peptide does not represent a realistic αPP variant, it was synthesised to determine how much helicity and stability could be garnered from a helix cap. αPP-NPELAA was slightly more folded than αPP-NPE with an α-helicity of 44 % however the T_M was approximately 20 °C which is slightly lower than for αPP-NPE. All peptide variants were monomeric in solution as determined by SE AUC. The biophysical characterisation data of these peptides are shown in **Figure 5-7** and their folding and stability statistics summarised in **Table 5-2**.

Table 5-2: Table summarising the MRE_{222} at 5 °C, the % helicity and the T_M for the series of peptides with helix caps.

Peptide name	$MRE_{222\ nm}$ / deg cm^2 dmol^{-1} res^{-1}	% Helicity[175-176]	T_M / °C
oPPα	-18319	48	51
αPP-GDNAT	-12046	32	*n/a*
αPP-NP	-14910	39	15
αPP-NPE	-16020	42	22
αPP-NPELAA	-17057	44	20

Interestingly, as the folding of the peptides in this series increased the ratio between the 208 and 222nm minima become closer to equal intensities. This series of peptides demonstrates the importance of residues close to the *N*-terminus which will impart stability to the peptide. The optimal cap in terms of enhancing thermal stability appears to be Asn-Pro-Glu. This cap was therefore used for subsequent αPP design variants.

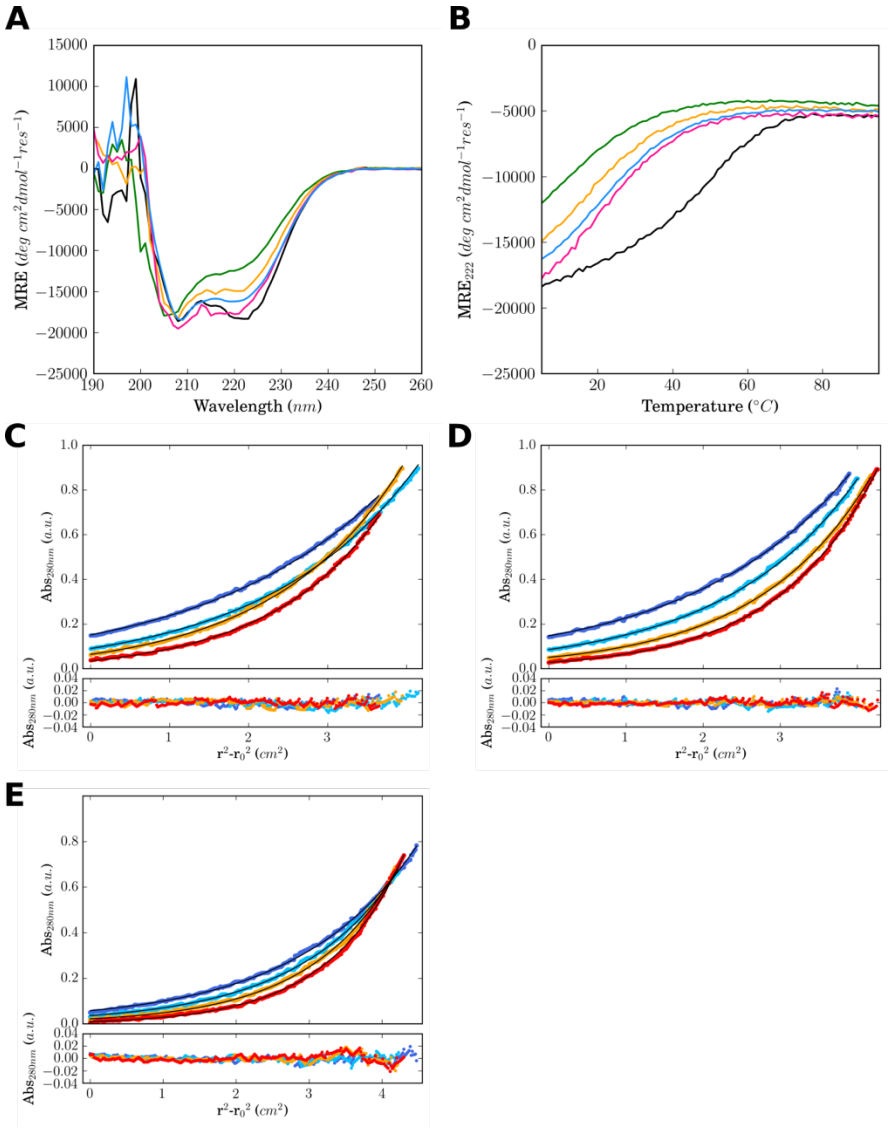


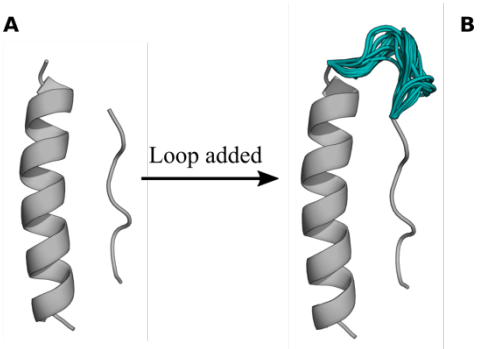
Figure 5-7: Biophysical characterisation of αPP variants with various capping motifs. (A) CD spectra recorded at 5 °C for αPP-NP, αPP-NPE, αPP-NPELAA. (B) Temperature dependence of the CD signal monitored at 222 nm. (C-E) AUC Sedimentation-equilibrium data (top, dots) and curves fitted to a single-ideal species model (lines) at 44 (blue), 52 (aqua), 56 (orange), and 60 (red) krpm. Bottom: Residuals for the above fits. (C) αPP-NP: $\bar{v}$ = 0.740 cm^3 g^{-1}. The fit returned a mass of 3936 Da (1.0 × monomer mass), 95% confidence limits = 3920 – 3951 Da. (D) αPP-NPE: $\bar{v}$ = 0.738 cm^3 g^{-1}. The fit returned a mass of 4296 Da (1.0 × monomer mass), 95% confidence limits = 4284 – 4308 Da. (E) αPP-NPELAA: $\bar{v}$ = 0.742 cm^3 g^{-1}. The fit returned a mass of 4646 Da (1.0 × monomer mass), 95% confidence limits = 4628 – 4664 Da. Conditions: 100 μM peptide, PBS, pH 7.4. Colour key: αPP-GDNAT (green), oPPα (black), αPP-NP (yellow), αPP-NPE (blue), αPP-NPELAA (pink).

5.2.2 Optimising loop region in αPP

With the new helix cap improving αPP folding and stability, efforts turned to the more challenging task of finding an optimal loop to connect the α helix and polyproline-II helix components. For reasons discussed earlier in this chapter, *de novo* loop design is significantly more challenging than designing regular elements of secondary structures like α helices and β strands. Therefore, loops were taken from proteins in the PDB. LoopFinder, a protein design tool within ISAMBARD, was used to extract loops from the PDB that fitted between the residues that 'entered' and 'exited' the loop. The entering residue was defined as the last residue of secondary structure entering the loop and by the same logic the exiting residue was defined as the first residue of secondary structure exiting the loop.

In oPPα a five-residue loop was optimal to bridge the gap between the polyproline-II and α helices. Based on this, it was hypothesised that a loop of similar length would be adequate for αPP. Therefore, only loops of between three and six residues were extracted from the PDB. Loops were filtered based on the following criteria: Firstly, loops were only selected from NMR or crystal structures with a resolution ≤ 3 Å. A resolution greater than this was not deemed reliable enough to ensure an accurate fit between the residues entering and exiting the loop and the loop residues selected. Also, this constraint helped refine and reduce the number of loops returned. Secondly, only loops with ≤ 2 Å RMSD fit between loop and entering and exiting residue backbone coordinates were selected. This ensured the transition in backbone coordinates from αPP to the loop were continuous. To ensure the loops selected fitted well with the secondary structure elements either side of the loop, the DSSP defined secondary structure of the entering and exiting residues were restricted to α and β, respectively. The β secondary structure element was chosen since polyproline-II helix is not formally defined by DSSP. Loops containing cysteine residues were also omitted for ease of handling; the thiol group can readily undergo oxidation if not handled with care.

Based on these criteria, 2212 loops were extracted from the PDB by LoopFinder. Of these, over half were three-residue loops, a quarter were four and five residues, and only 28 loops were 6 residues long (**Figure 5-8**).



Loop Length	Number of Loops
3	1362
4	433
5	389
6	28
Total	**2212**

Figure 5-8: (A) Overlay of the top 100 five-residue loops i.e. those that have the lowest RMSD. (B) Frequency of loops of ranging lengths selected from the PDB using LoopFinder.

Subsequently, loops were refined and selected for ease of synthesis and low RMSD with αPP. Loops with amino-acid sequences that were particularly hydrophobic were avoided because the loop in αPP is solvent exposed. Also, only loops with the α and β helices in an antiparallel orientation, similar to αPP, were selected. The loop sequences chosen are shown in **Table 5-3**.

Table 5-3: Sequences of αPP peptides with ranging loop lengths.

Peptide name	Sequence and helical register				
	defgabc	defgabc	defgabc		123123
αPP-NPE-3	Ac-NPE YEKELAA	YEKELAA	YEKELAA	**GKP**	PKKPKKPPG-NH_2
αPP-NPE-4	Ac-NPE YEKELAA	YEKELAA	YEKELAA	**NPQG**	PKKPKKPPG-NH_2
αPP-NPE-5	Ac-NPE YEKELAA	YEKELAA	YEKELAA	**NGGQG**	PKKPKKPPG-NH_2
αPP-NPE-6	Ac-NPE YEKELAA	YEKELAA	YEKELAA	**QAKDQQ**	PKKPKKPPG-NH_2

The loops selected for joining the constituent components of αPP are shown in their native protein environments in **Figure 5-9**. All loops are proximal to the surface of the protein as opposed to being buried within the core and thus are relatively solvent exposed. While the conformation that a loop adopts depends on the surrounding protein environment, the fact that these loops are solvent exposed bodes well for their use in αPP where they will also be in a solvent exposed environment.

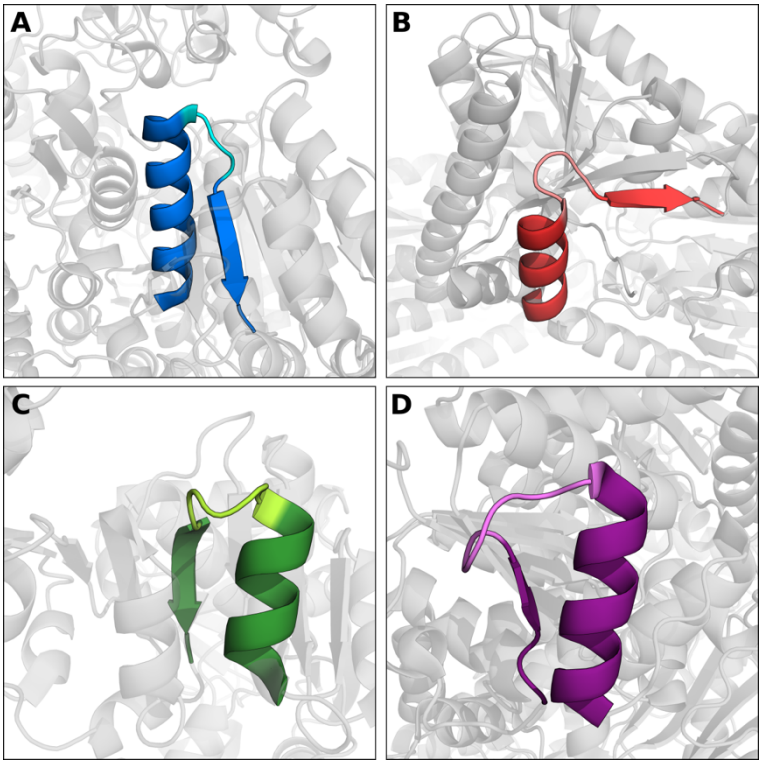


Figure 5-9: Loops selected for αPP in their natural protein environment. (A) 3 residue loop, GKP, from PDB: 1T9A. (B) 4 residue loop, NPQG, from PDB: 3T7D. (C) 5 residue loop, NGGQG, from PDB: 3QWA. (D) 6 residue loop, QAKDQQ, from PDB: 4RHE. Colour key: Loop (light), entering and exiting secondary structure (dark), surrounding protein (grey).

5.2.3 MD Studies

To give an indication of the stability of the series of αPP peptides with varying loop lengths, 100 ns MD simulations were performed for each variant (**Figure 5-10**). Note that for this MD study models of the peptides were generated without the Asn-Pro-Glu capping residues however this did not interfere with determining the effectiveness of the loop region.

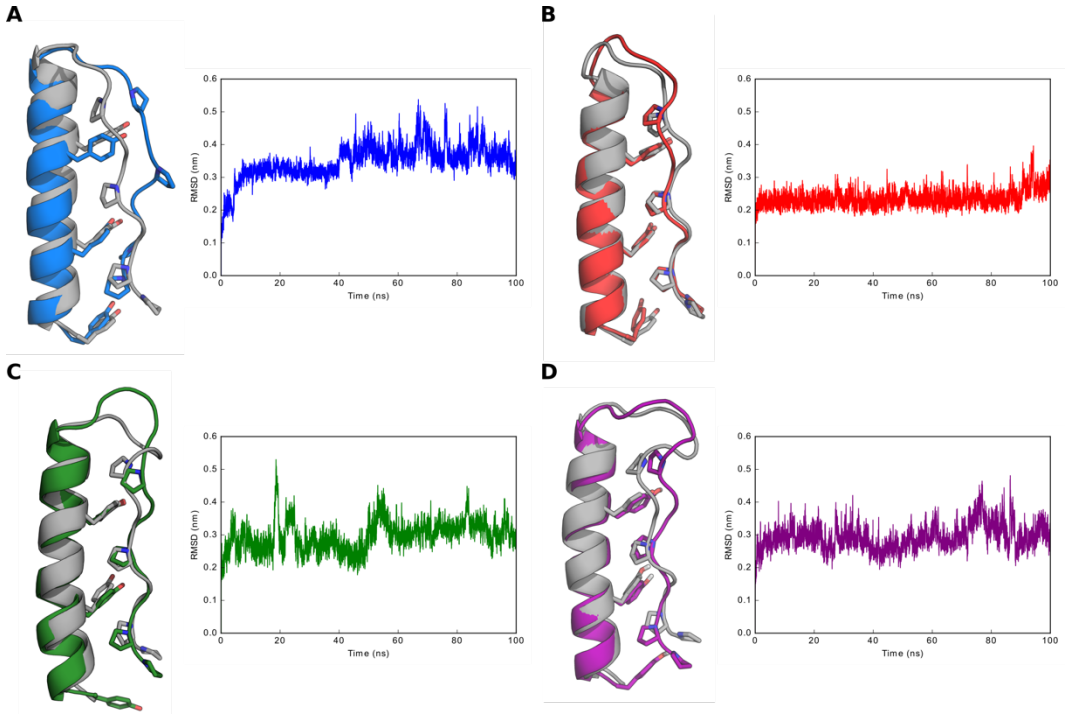


Figure 5-10: Snapshots of MD simulations at 0 and 100 ns, and fluctuations in all-atom RMSD across the trajectory. (A) αPP-3 (loop: GKP) (B) αPP-4 (loop: NPQG) (C) αPP-5 (loop: NGGQG) (D) αPP-6 (loop: QAKDQQ). Colour key: initial 0 ns snapshot (grey), 100 ns snapshot (blue, red, green and purple).

αPP-NPE-3 with the three-residue GKP loop was least stable across the MD trajectory resulting in the proline residues on the polyproline-II helix flipping to the solvent-exposed surface and no longer interdigitated with the tyrosine residues on the α helix. It appears the loop was too short to keep the polyproline-II helix in the correct orientation against the α helix and the interface between the α and polyproline-II helix was lost. Consistent with this, the RMSD across the trajectory showed the greatest deviation. In contrast, αPP-NPE-4 (with the four-residue NPQG loop) was stable across the trajectory and the initial and 100 ns snapshots overlay well, in particular the Pro and Tyr interface residues. This suggests the four-residue loop was a good fit and allowed the interface between the two helices to be maintained. αPP-NPE-5 and αPP-NPE-6 were also relatively stable across their trajectories. While the Tyr and Pro residues from the 0 and 100 ns timepoints do not overlay as well as compared to αPP-NPE-4, the interface between the two helices is maintained and Pro and Tyr are still interdigitated. As a result of this MD study αPP-NPE-3 was not taken forward for synthesis.

Subsequently, only the αPP variants with four, five and six residue loops were synthesised. These peptides were then characterised by CD spectroscopy to determine their folding and stability in

solution. All peptides in the series exhibited very similar α-helicities at 5 °C (approximately 40 %), however they showed a broader range of thermal stabilities. The variant with the shortest length loop, αPP-NPE-4, was the least thermally stable in the series with a T_M of 18 °C. Interestingly, αPP-NPE-4 was less thermally stable than αPP-NPE, which has the natural GDNAT loop. The variant with the five-residue loop, αPP-NPE-5 was the most thermally stable in the series with a T_M of 28 °C. The variant with the longest loop, αPP-NPE-6 fell between the two with a T_M of 20 °C. All the peptides in this series were monomeric in solution as determined by SE AUC. The CD and AUC biophysical data are shown in **Figure 5-11** and the folding and thermal stability of these peptides are summarised in **Table 5-4** .

Peptide name	**$MRE_{222\ nm}$ / deg cm^2 dmol^{-1} res^{-1}**	**% Helicity$^{175-176}$**	**T_M / °C**
oPPα	-18319	48	51
αPP-NPE	-16020	42	22
αPP-NPE-4	-15395	40	18
αPP-NPE-5	-15789	41	28
αPP-NPE-6	-15190	39	20

Table 5-4: Summary of the α-helicity at 5 °C and thermal stability of the series of peptides with varying loop length.

While none of the peptides were as stable as the PPα counterpart the αPP-NPE-5 variant with the NGGQG loop had the greatest thermal stability. Therefore, this loop was taken forward to create a PPαPP design and combine PP:α and α:PP interfaces. It was hoped that the presence and well folded behaviour of the PPα component may help template the folding of the *C*-terminal second polyproline-II helix.

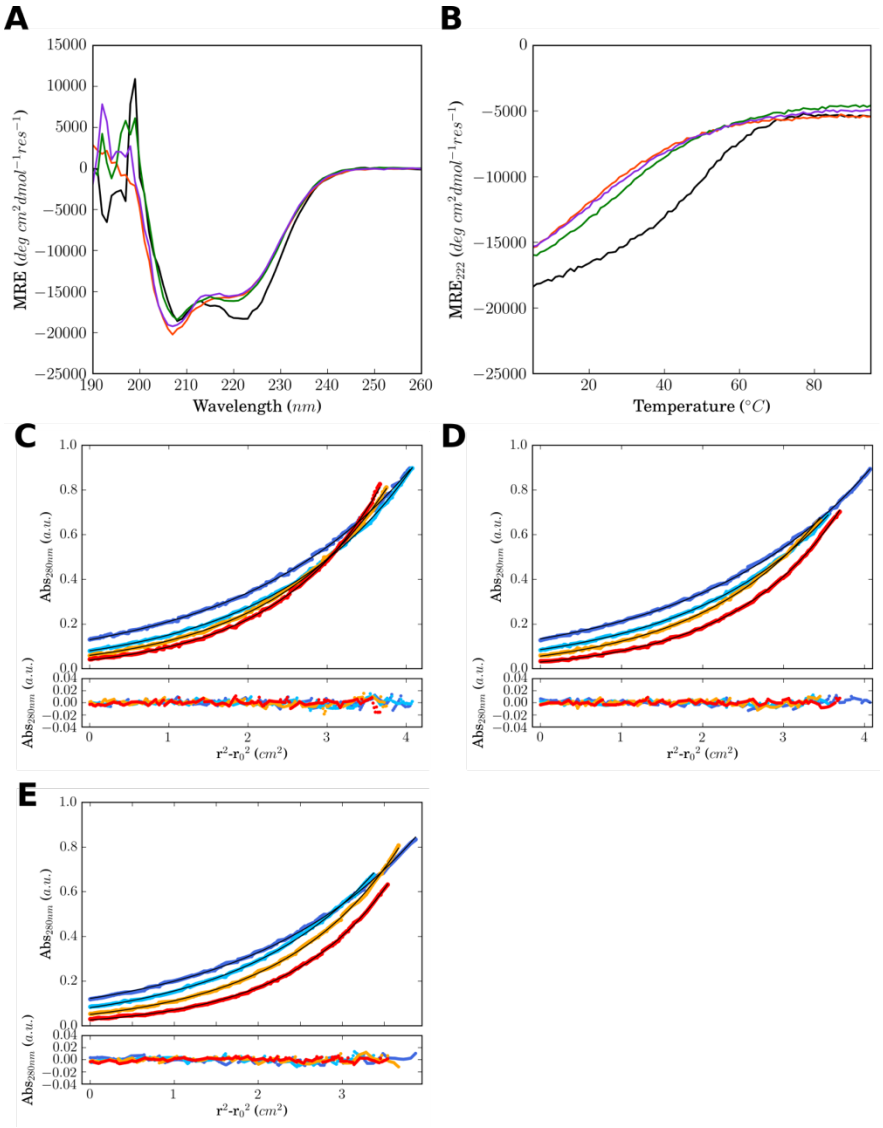


Figure 5-11: Biophysical characterisation of αPP variants with varying loop lengths and composition selected using LoopFinder. (A) CD spectra recorded at 5 °C for αPP-NPE-4, αPP-NPE-5, αPP-NPE-6. (B) Temperature dependence of the CD signal monitored at 222 nm. (C-E) AUC Sedimentation-equilibrium data (top, dots) and curves fitted to a single-ideal species model (lines) at 44 (blue), 52 (aqua), 56 (orange), and 60 (red) krpm. Bottom: Residuals for the above fits. (C) αPP-NPE-4: $\bar{v}$ = 0.741 cm^3 g^{-1}. The fit returned a mass of 4172 Da (1.0 × monomer mass), 95% confidence limits = 4160 –4184 Da. (D) αPP-NPE-5: $\bar{v}$ = 0.737 cm^3 g^{-1}. The fit returned a mass of 4180 Da (1.0 × monomer mass), 95% confidence limits = 4166 – 4195 Da. (E) αPP-NPE-6: $\bar{v}$ = 0.739 cm^3 g^{-1}. The fit returned a mass of 4562 Da (1.0 × monomer mass), 95% confidence limits = 4548 – 4576 Da. Conditions: 100 μM peptide, PBS, pH 7.4. Colour key: oPPα (black), αPP-NPE-4 (red), αPP-NPE-5 (green), αPP-NPE-6 (purple).

5.3 Expanding the PPα topology: PPαPP

With both an oPPα and αPP component in hand efforts turned towards combing the two topologies in an attempt to create a PPαPP topology (**Figure 5-12A**). In α-helical barrels, each helix interacts with the adjacent helix *via* a hydrophobic seam. There are three main types of interfaces as discussed in Section 1.3.1: the two hydrophobic seams share one residue (type I interface), the two seams are adjacent (type II interface) or the seams are separated by one intervening residue (type III interface).[113, 145] Type III-like interfaces were used in the design of PPαPP (**Figure 5-12B**).

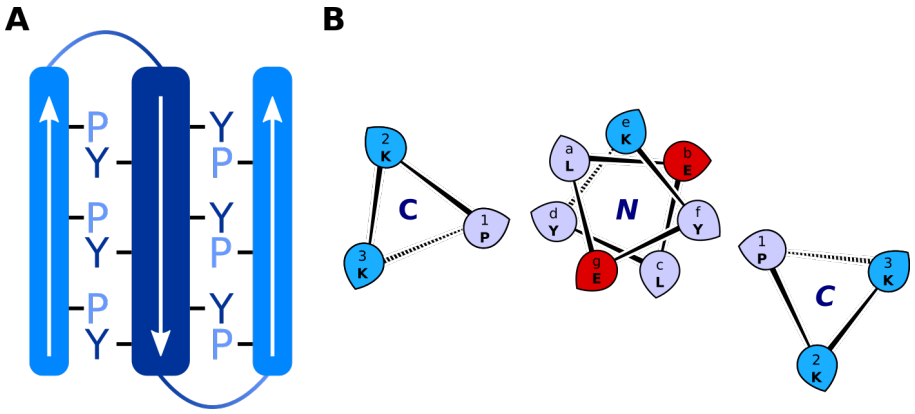


Figure 5-12: (A) Cartoon representation of the PPαPP topology with the polyproline-II helices shown in light blue and the α helix shown in dark blue. Arrows demonstrate the orientation of the helices. (B) Helical wheel representation of PPαPP with overlaid peptide sequence highlighting the interhelical interfaces. Colour key: hydrophobic seams (purple), Glu (red), Lys (blue). *C* and *N* represent the termini of the helices nearest the viewer.

The ***a***, ***d*** and ***g*** positions (Leu-Tyr-Glu) that make up the diamond shaped hole on the α helix of oPPα were maintained in the design of PPαPP. An extra Leu-Tyr-Glu interface was placed on the previously solvent-exposed face of oPPα α helix at the ***c***, ***f*** and ***b*** positions. Lys was placed at the ***e*** position to balance charge on the α helix. Also, a potential i→i+3 salt bridge could form between Lys at ***e*** and Glu at ***b*** which would help stabilise the α helix. The natural GDNAT loop was used to connect the PPα components of the topology while the NGGQG loop from αPP-NPE-5 was used to connect the αPP component. Overall the sequence for the design of PPαPP is shown in **Table 5-5**. PPαPP was first characterised by CD spectroscopy (**Figure 5-13**).

Table 5-5: Sequence of the PPαPP Design.

Peptide name	**Sequence and helical register**
	321321 efgabcd efgabcd efgabcd 123123
oPPα	Ac-PPKKPKKP GDNAT PEKLAAY EKELAAY EKELAAY-NH_2
PPαPP	Ac-PPKKPKKP GDNAT PYKLELY KYELELY KYELELY NGGQG PKKPKKPPG -NH_2

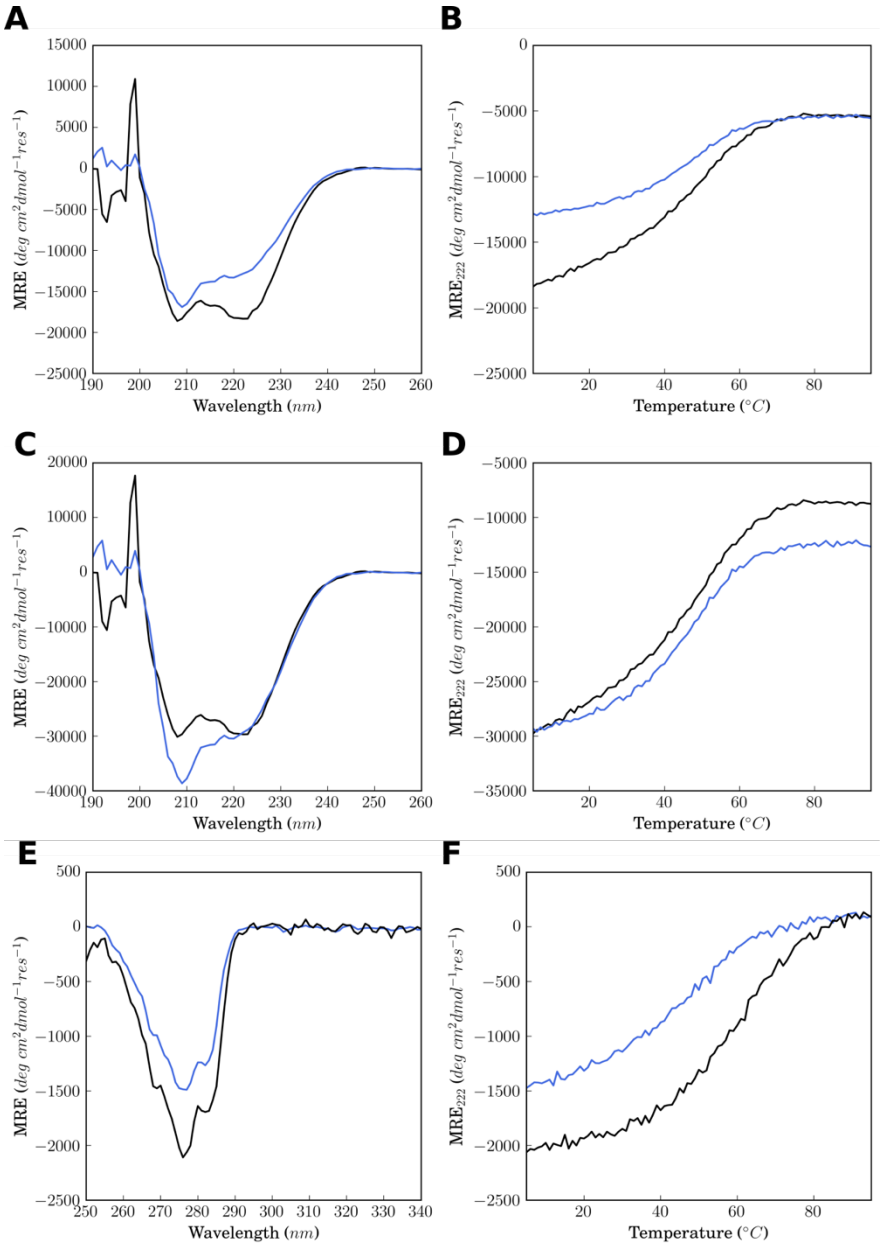


Figure 5-13: Folding and stability of PPαPP. (A) CD spectra recorded at 5 °C for oPPα and PPαPP. (B) Temperature dependence of the far-UV CD signal monitored at 222 nm. (C) CD spectra recorded at 5 °C normalised for expected number of α-helical residues. (D) Temperature dependence of the far-UV CD signal monitored at 222 nm normalised for expected number of α-helical residues. (E) Aromatic CD spectra recorded at 5 °C. (F) Temperature dependence of the near-UV CD signal monitored at 276 nm. Conditions: 100 μM peptide, PBS, pH 7.4. Colour key: oPPα (black), PPαPP (blue).

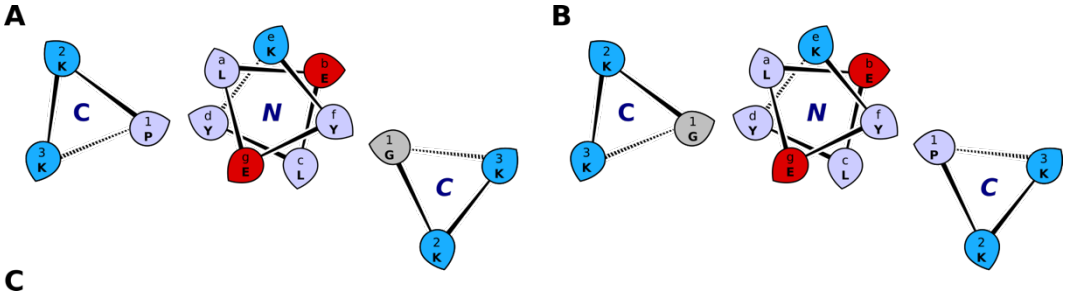
PPαPP was significantly less folded at 5 °C in the far-UV region compared to oPPα with an α-helicity of 33 %. However, the thermal stability of PPαPP was only slightly lower than oPPα with a T_M of 48 °C. The reduced folding of PPαPP is the result of the presence of the extra polyproline-II helix which will have limited contribution to the CD signal at 222 nm. Therefore, to give a better indication of the effect of the extra polyproline-II helix on the stability of PPαPP, the CD data were

renormalised for α-helical residues only in both topologies (**Figure 5-13C&D**). Upon renormalisation the α-helicity of PPαPP at 222 nm matched that of oPPα.

There is also a significant difference in the ratio of the minima at 208 and 222 nm. This is again likely a result of the additional polyproline-II helix. Polyproline-II helices have a deep minimum at approximately 197 nm which will contribute to the signal at 208 nm for PPαPP. (See Section 2.3.1).

PPαPP was also monitored by near-UV CD spectroscopy to monitor the aromatic environment. The thermal transition of PPαPP by near-UV CD spectroscopy showed unfolding and refolding curves that were consistent with the far-UV CD traces. The T_M when measured through the aromatic CD signal at 276 nm was 48 °C which is the same as the T_M when measured through the far-UV CD trace.

To probe further the contributions of each helix interface to the overall stability of the PPαPP topology each interface was knocked out in turn. An interface was removed by mutation of all the Pro residues on the polyproline-II helix to Gly residues. The sidechain of Gly is a single hydrogen atom and therefore should not contribute to helix packing. Each interface was knocked out individually leading to two peptides: PPαPG and PGαPP (**Figure 5-14**).



Peptide name	Sequence and helical register
	321321 efgabcd efgabcd efgabcd 123123
PPαPP	Ac-PPKKPKKP GDNAT PYKLELY KYELELY KYELELY NGGQG PKKPKKPPG -NH_2
PPαPG	Ac-PPKKPKKP GDNAT PYKLELY KYELELY KYELELY NGGQG GKKGKKGGG -NH_2
PGαPP	Ac-GGKKGKKG GDNAT PYKLELY KYELELY KYELELY NGGQG PKKPKKPPG -NH_2

Figure 5-14: Sequences and helical wheel representation of PPαPG and PGαPP. (A) PPαPG (B) PGαPP. (C) Sequences of the PPαPG and PGαPP peptides.

Both peptides were analysed by CD spectroscopy (**Figure 5-15**). The CD spectra of PPαPG informed about the folding and stability of the PPα interface while the PGαPP related to the αPP interface.

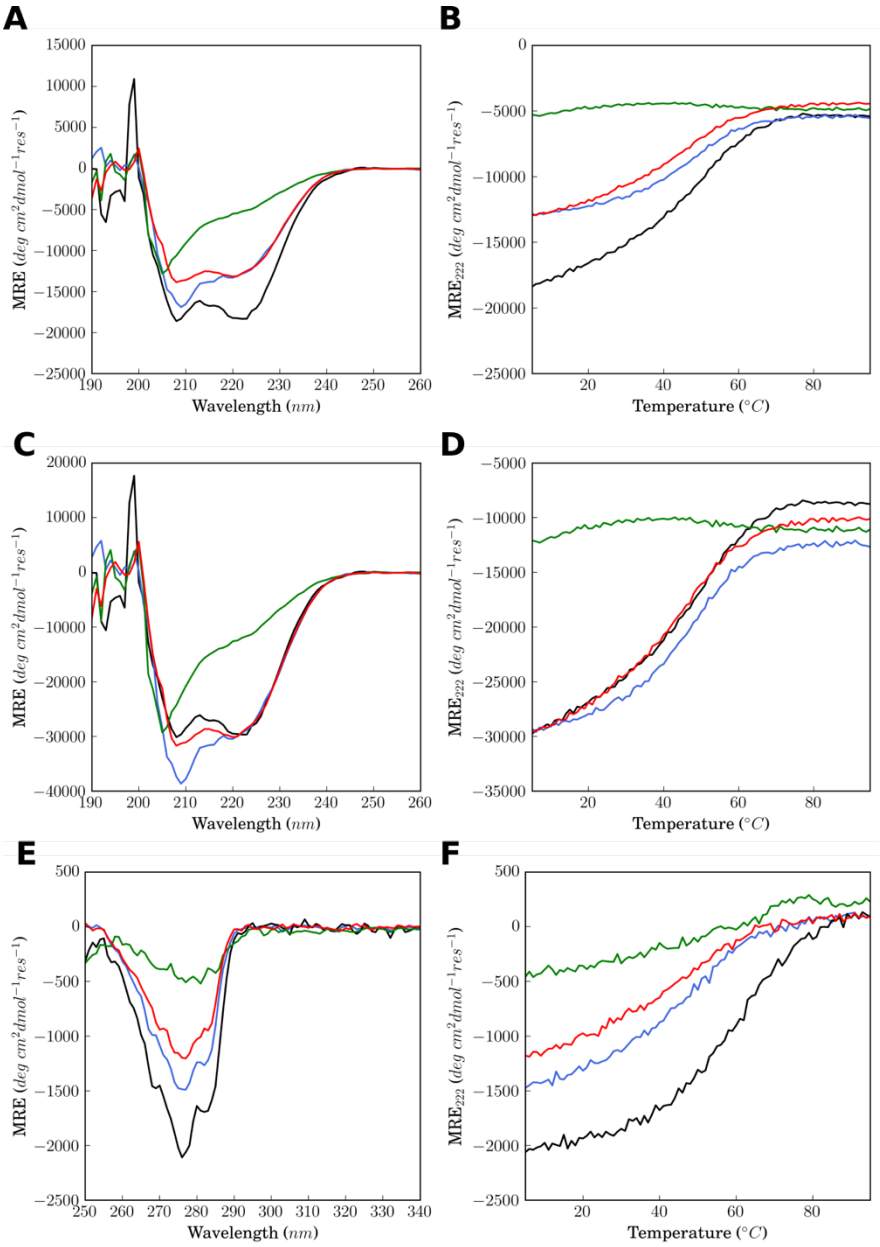


Figure 5-15: Folding and stability of PPαPG and PGαPP. (A) CD spectra recorded at 5 °C for oPPα, PPαPP, PPαPG and PGαPP. (B) Temperature dependence of the far-UV CD signal monitored at 222 nm. (C) CD spectra recorded at 5 °C normalised for α-helical residues. (D) Temperature dependence of the far-UV CD signal monitored at 222 nm normalised for α-helical residues. (E) Aromatic CD spectra recorded at 5 °C. (F) Temperature dependence of the near-UV CD signal monitored at 276 nm. Conditions: 100 μM peptide, PBS, pH 7.4. Colour key: oPPα (black), PPαPP (blue), PPαPG (red), and PGαPP (green).

PPαPG was well folded with the same helicity at 5 °C as observed for PPαPP. The thermal stability of PPαPG was slightly reduced with a T_M of 46 °C compared to 48 °C for PPαPP. Conversely, the folding and thermal stability of PGαPP was dramatically reduced. The peptide showed only 14 %

α-helicity at 5 °C and a T_M could not be extracted. A summary of the folding and stability of these peptides is shown in **Table 5-6**.

Table 5-6: Summary of the folding and thermal stabilities of PPαPP and variants.

Peptide name	$MRE_{222\ nm}$ / deg cm^2 dmol^{-1} res^{-1}	% Helicity[175-176]	T_M / °C
oPPα	-18319	48	51
PPαPP	-12818	33	48
PPαPG	-12939	33	46
PGαPP	-5283	14	n/a

The slight reduction in thermal stability of PPαPG compared to PPαPP suggests the αPP interface plays a lesser role in the overall stability of the topology while the PPα interface is the most dominant. Further, the significantly reduced folding and stability of PGαPP suggests that without the PPα interface the PPαPP topology will not form. Taken together this implies it is the PPα component of PPαPP that reduces the energy for second polyproline-II helix by providing a template.

It was anticipated that the PPα interface within the PPαPP topology would be the most dominant given we have previously shown oPPα to be more folded and stable compared to αPP. However, the fact that the PGαPP peptide was essentially unfolded was unexpected.

Interestingly, the minima at 208 and 222 nm are of equal intensities for PPαPG, this is in contrast to what is observed for PPαPP. This may be explained by the fact that without the proline residues, a polyproline-II helix conformation is not adopted; instead just random coil. While the CD spectra of random coil and polyproline-II helices are similar there are some slight differences. One such difference is that the minima at 197 nm in the CD spectra of polyproline-II helices is reduced in magnitude and redshifted.[172] The effect of this in the context of PPαPG would result in the signal at 208 nm to also be reduced in magnitude and therefore result in an intensity that matches the signal at 222 nm.

The oligomeric states of PPαPP, PPαPG and PPαPPP were determined by SE AUC (**Figure 5-16**). Both PPαPP and PPαPG were monomeric in solution and fitted well to single ideal species models. PGαPP precipitated at approximately 50 kprm and therefore the oligomeric state of the peptide could not be determined. The presence of precipitation emphasised the unfolded and less predictable nature of the peptide.

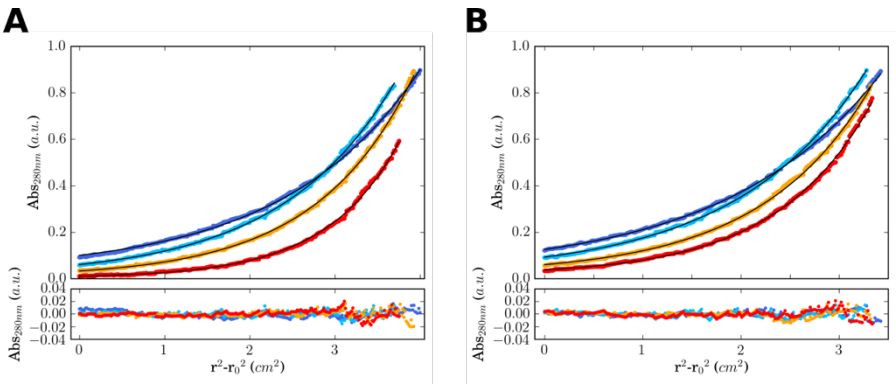


Figure 5-16: Sedimentation-equilibrium AUC data (top, dots) and curves fitted to single-ideal species model (lines) at 44 (blue), 52 (aqua), 56 (orange), and 60 (red) krpm. Bottom: Residuals for the above fits. (A) PPαPP: $\bar{v}$ = 0.753 cm^3 g^{-1}. The fit returned a mass of 5951 Da (1.1 × monomer mass), 95% confidence limits = 5932 – 5970 Da. (B) PPαPG: $\bar{v}$ = 0.726 cm^3 g^{-1}. The fit returned a mass of 5919 Da (1.1 × monomer mass), 95% confidence limits = 5897 – 5941 Da. Conditions: 100 μM peptide, PBS, pH 7.4.

5.4 Chapter conclusion

In conclusion, an iterative design process has been used to design a stable and folded αPP topology. Through the iterative design process it became apparent that simply maintaining the interface between the two helices of the foregoing oPPα design and using either a flexible (αPP-GSGSG) or suboptimal loop region (αPP-GDNAT) to connect the α and polyproline-II helices is not enough to form a stable αPP topology. Subsequently, two important factors were considered for generating a well folded αPP topology: introducing *N*-terminal capping motifs and an optimising the loop region.

N-terminal capping motifs help initiate helix formation and stabilise the α-helix component. Based on work by Richardson *et al.*, which gives amino acid preferences for specific locations at the ends of helices, an Asn-Pro-Glu motif was added to the *N* terminus of the α helix of αPP.[225] This resulted in an improvement in the folding and thermal stability of the αPP topology compared to the initial design with an increase in T_M from < 20 to 22 °C.

MD simulations suggest that a three-residue loop is too short to connect the α and polyproline-II helices effectively but a loop of length four to six residues should be sufficient. Therefore, LoopFinder was used to extract loops ranging from four to six residues from the PDB that would fit between a pair of entering and exiting residues of the αPP topology. The best loop selected from LoopFinder is five residues in length and resulted in an αPP variant with a T_M of 28 °C. This is a significant improvement on the initial αPP design (T_M < 20 °C). However, this is still not comparable to the stability of the oPPα counterpart suggesting αPP could be optimised further.

Therefore, a more comprehensive study of loops selected by LoopFinder will be required to find an optimal αPP loop. A quantitative approach would help refine and select the best loops for synthesis

from the large number (>1000) of loops outputted by LoopFinder. Parameters to be considered include RMSD with entering and exiting residues, hydrophobicity of loop specifically the hydrophobic moment, orientation of secondary structure either side of loop, loop rigidity, and ease of loop synthesis with regard to sequence.

While LoopFinder offers a good starting point for selecting potential loops to connect secondary structure components together there are some drawbacks to this approach. LoopFinder assumes the loop will adopt the same conformation in the native protein environment as in its new miniprotein environment, which may not be the case.

Overall, while an αPP topology that is comparable in folding and stability to that of oPPα (T_M 51 °C) has not been delivered a reasonably well folded peptide with a T_M of 28 °C (αPP-NPE-5) has been designed and characterised and has been taken forward to expand the PPα topology. Also, significant progress forward has been made compared to the initial αPP design which had a T_M of < 20 °C.

The NGGQG loop from αPP-NPE-5 is used in the design of PPαPP. PPαPP has similar levels of folding and thermal stability to oPPα. To determine the contribution of each interface to the overall stability of the topology two variants (PPαPG and PGαPP) are described with each helix interface knocked out in turn. The interface was removed by mutation of the Pro residues in the polyproline-II helix to Gly. CD spectroscopy of these peptides revealed that the PPα interface in the PPαPP topology was significantly more dominant; PGαPP was essentially unfolded.

The next stage of characterisation will seek to verify the structure through NMR spectroscopy. High resolution structural data will be required to determine whether the PPαPP peptide has formed the topology desired. NOEs between helices will help in determining whether the second polyproline-II helix is interacting with the α helix or not.

Overall, two new topologies have been designed (αPP and PPαPP) that, to the best of our knowledge, have not been observed before in nature and thus help move us into the dark matter of protein space.[33] Solution phase biophysical characterisation for the peptides discussed in this chapter are summarised in **Table 5-7**.

Table 5-7: Combined solution-phase biophysical characterisation for the peptides discussed in this chapter.

Entry	Name	Sequence and register	MRE_{222} (deg cm^2 $dmol^{-1}$ res^{-1}) [a]	% Helicity	T_M / °C [b]	Oligomeric State [c]
		321321 efgabcd efgabcd efgabcd				
1	oPPα	Ac- PPKKPKKP GDNAT PEKLAAY EKELAAY EKELAAY-NH_2	-18319	48	51	1.0
αPP		defgabc defgabc defgabc defgabc 123123123				
2	αPP-GDNAT	Ac-YEKELAA YEKELAA YEKELAA GDNAT PKKPKKPPG-NH_2	-12046	32	< 20	1.0
3	αPP-GDNAT-4h	Ac-YEKELAA YEKELAA YEKELAA YEKELAA GDNAT PKKPKKPKKPPG-NH_2	-19601	50	49	1.0
4	αPP-GSGSG	Ac-YEKELAA YEKELAA YEKELAA GSGSG PKKPKKPPG-NH_2	-13826.8	36	20	1.0
Capping Motifs		defgabc defgabc defgabc 123123				
5	αPP-NP	Ac-NP YEKELAA YEKELAA YEKELAA GNDAT PKKPKKPPG-NH_2	-14910	39	15	1.0
6	αPP-NPE	Ac-NPE YEKELAA YEKELAA YEKELAA GDNAT PKKPKKPPG-NH_2	-16020	42	22	1.0
7	αPP-NPELAA	Ac-NPELAA YEKELAA YEKELAA YEKELAA GDNAT PKKPKKPPG-NH_2	-17057	44	20	1.0
Loops		defgabc defgabc defgabc 123123				
8	αPP-NPE-4	Ac-NPE YEKELAA YEKELAA YEKELAA NPQG PKKPKKPPG-NH_2	-15395	40	18	1.0
9	αPP-NPE-5	Ac-NPE YEKELAA YEKELAA YEKELAA NGGQG PKKPKKPPG-NH_2	-15789	41	28	1.0
10	αPP-NPE-6	Ac-NPE YEKELAA YEKELAA YEKELAA QAKDQQ PKKPKKPPG-NH_2	-15190	39	20	1.0
PPαPP		321321 defgabc defgabc defgabc 123123				
11	PPαPP	Ac-PPKKPKKP GDNAT PYKLELY KYELELY KYELELY NGGQG PKKPKKPPG-NH_2	-12818	33	48	1.1
12	PPαPG	Ac-PPKKPKKP GDNAT PYKLELY KYELELY KYELELY NGGQG GKKGKKGGG-NH_2	-12939	33	46	1.1
13	PGαPP	Ac-GGKKGKKG GDNAT PYKLELY KYELELY KYELELY NGGQG PKKPKKPPG-NH_2	-5283	14	n/a	n/a

[a] MRE_{222} from CD spectroscopy at 5 °C and 100 μM peptide in PBS. [b] The point of inflection in sigmoidal thermal denaturation curves, calculated from the maxima of the first derivative of the thermal transition. [c] Oligomeric state calculated from SE AUC experiments.

Chapter 6 Modulating stability in oPPα through elongation and cyclisation

The cyclisation work described in this chapter was done in collaboration with Marcel Schmidt (MS) at EnzyPep, Netherlands. The work was conceived and designed by the author of this thesis (KPG), Prof. Dek Woolfson, and MS. MS performed the enzymatic cyclisation and KPG carried out the synthesis and biophysical characterisation. Dr Prasun Kumar performed the bioinformatics analysis.

6.1 Chapter introduction

6.1.1 Protein-protein interactions

Miniproteins provide a compact and stable scaffold onto which functional motifs can be grafted including binding and recognition motifs that have the potential to disrupt protein-protein interactions.[21, 59, 269] Both transient and long-lived, protein-protein interactions are important for controlling a diverse range of cellular functions.[270-273] As a result, protein interfaces have become targets for perturbing cell function.[274-275] Further, the use of miniproteins to disrupt protein-protein interactions in important cellular recognition domains offers exciting and challenging opportunities to develop miniprotein-based drug therapeutics.[276-278] More-challenging motifs that have the potential to introduce catalytic function into the simplified miniprotein system are also being explored.[61]

SH3 (Src homololgy region 3) and other domains such as EVH1 and WW domains have similar interfaces to that of PPα. SH3 domains are one of the most abundant molecular recognition domains in the proteome and are found in a wide variety of intracellular and membrane-associated proteins.[279] They play important roles in protein-protein interactions essential to cell function.[280] The SH3 domain has a characteristic fold consisting of five or six β strands arranged as two antiparallel β sheets. The surface of the domain has a hydrophobic ligand binding pocket with highly conserved aromatic residues in which Pro-rich sequences bind. The Pro-rich ligands contain a conserved PXXP binding motif and adopt a polyproline-II helix conformation (**Figure 6-1A**).[281-286] In isolation, the Pro-rich ligands bind with modest affinity (5 -100 μM) and with little selectivity within families of SH3 domains; many SH3 domains bind multiple PXXP containing peptides.[287-288] However, SH3 domains are often found in association with other interaction domains (e.g. SH2 domains) and work in a cooperative manner to give highly specific protein-protein interactions.[289]

The avian pancreatic peptide (aPP) has been used as a scaffold whereby some residues on the polyproline-II helix have been replaced by a Pro-rich sequence motif (RP1) that interacts with the SH3 domain of abelson tyrosine kinase (Ab1-SH3) (**Figure 6-1B**).[281] The binding of this new aPP mutant (APP-RP1) to Ab1-SH3 is comparable to the binding of RP1with Ab1-SH3.

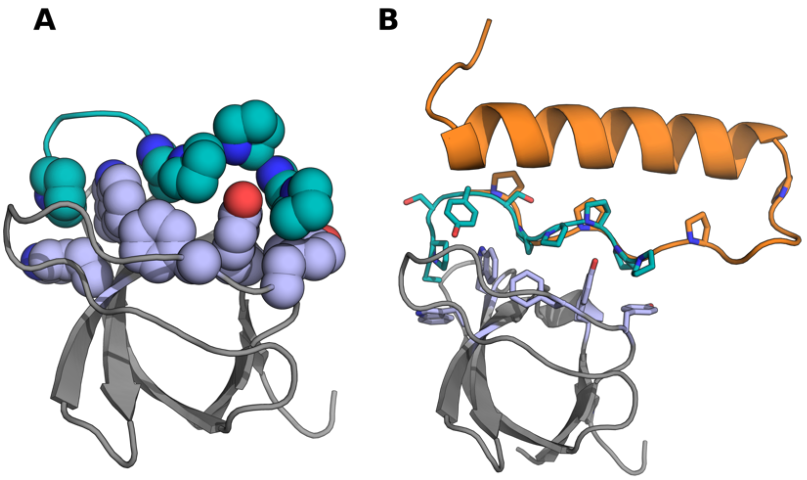


Figure 6-1: (A) Interaction of the SH3 domain of abelson kinase (Ab1-SH3) with a proline rich sequence motif called RP1 (PDB: 1BBZ). (B) Structural alignment of aPP (PDB: 1PPT) with RP1 motif which is complexed to Ab1-SH3. Colour key: RP1(teal), Ab1-SH3 (grey, aromatic residues highlighted in purple) and aPP (orange).

The aPP scaffold has been used in a similar way to bind to the polyproline-recognition domain EVH1. EVH1 domains are found in a large number of signalling proteins and are often involved in modulating the actin cytoskeleton or in signal transduction in postsynaptic compartments of chemical synapses.[290] Optimisation of the polyproline-II helix of aPP produced a variant with high affinity for the ActA target protein in *Listeria monocytogenes*, EVH1 mena.[56] The α helix of aPP has also been optimised for sequence specific DNA recognition. A mutant of aPP named p007 presents the α-helical recognition epitope found on the bZIP protein GCN4. The recognition epitope was grafted onto the solvent exposed face of the α helix and subsequently binds DNA with nanomolar affinity.[59]

6.1.2 Controlling the stability of protein-protein interactions

Similar to aPP, the compact, well-folded nature of the oPPα miniprotein lends itself as a good scaffold for targeting and disrupting protein-protein interactions. The sequence-to-stability relationships established for the fold in Chapter 4 will help determine which residues are amenable to mutation. In complicated association-dissociation processes involving complex dynamics it will be important to be able to control and tune the stability and affinity of protein-protein interactions.

One way to tune stability is to vary the length of the peptide components. The stability of α-helical coiled coils increases dramatically with peptide length in a direct but non-linear cooperative manner; the increase in stability on going from three to four heptads is greater than going from four to five heptads.[291-292] As chain length increases, the importance of hydrophobic interactions in relation to stability decreases.[291] The correlation between peptide chain length and stability has successfully been used to engineer and quantify a range of stabilities into heterodimeric coiled coils with micromolar to sub-nanomolar dissociation constants.[203] Such peptides will be useful for applications in protein engineering and synthetic biology where protein-protein interactions with specific

stabilities are required. The effect of chain length on the stability of oPPα will therefore be examined in this chapter.

As well as varying chain length to modulate peptide stability, miniproteins can also be stabilised by cyclisation. Cyclic miniproteins are particularly attractive for *in vivo* applications where the peptides will need to be resistant to proteolysis and degradation in the cellular environment.[293] Linear peptide based drugs have limited resistance to proteolysis[294] compared to cyclic variants which have much better metabolic stability due to their reduced sensitivity to proteolytic cleavage.[295] A cyclic miniprotein also offers potentially enhanced binding properties as a result of its limited conformational flexibility which in turn reduces the entropic penalty upon binding. Given the importance of proteolytic stability for applications of miniproteins *in vivo*, and more ambitiously as drug therapeutics, cyclisation of oPPα has been explored and will be discussed herein.

6.1.3 Peptide cyclisation strategies

A peptide can be cyclised in four different ways: head-to-tail (*C*-terminus to *N*-terminus), head-to-side chain, side chain-to-tail, or side chain-to-side chain.[296] For each of these strategies various different synthetic methodologies have been explored. One strategy involves an azide-alkyne cycloaddition to introduce a 1,2,3-triazole into the peptide backbone, which are both thermally and physiologically stable.[297-298] Further, the triazole is readily obtained by 1,3-dipolar cycloadditions between an azide and alkyne group. However, while the synthesis is synthetically facile a non-native peptide bond is generated at the ligation site. Another cyclisation strategy is the synthesis of larger more flexible peptide macrocycles followed by an intramolecular ring contraction to give the target structure.[299-300] This has been demonstrated as an effective way to reduce the entropic penalty of macrocyclization, particularly for very small ring sizes. Ring closing metathesis has also been used to cyclise side chain protected peptides in anhydrous organic solvents.[301-302] However, for many of these synthetic methodologies heavy dilution is required to prevent polymerisation; there are risks of epimerisation; and sidechain protected peptides can have poor solubility.[303]

Native chemical ligation (NCL) is an alternative strategy often used to cyclise unprotected peptides in aqueous solution. This reaction utilises an *N*-terminal Cys and a *C*-terminal thioester that react to form a native amide bond through a trans-thioesterification followed by S-to-N acyl migration.[304] The strategy was initially developed to ligate peptide segments together to facilitate the synthesis of large peptides. However, by incorporating the *N*-terminal Cys and *C*-terminal thioester into the same peptide, cyclic peptides can readily be synthesised.[305] However, this strategy usually requires the presence of a Cys in the peptide sequence although desulfurization methods have been developed.[306-307]

6.1.3.1 Enzymatic cyclisation

Enzymes offer an alternative to traditional synthetic methodologies for cyclisation. Examples of such enzymes include sortases,[308] trypsin,[309] asparaginyl endoproteases (AEP),[310-311] and subtilisin variants such as peptiligase.[312] While these enzymes offer excellent stereo- and chemo- selectivity

many suffer from incomplete ligation, low catalytic efficiency, and leave a "footprint" region. A footprint is an unavoidable enzyme recognition sequence at the coupling site. However, a peptiligase based enzyme variant named omniligase-1 has recently been reported to efficiently catalyse cyclisation and is footprint free.[313-314]

Peptiligase is protein engineered from a stabilised cation independent variant of substilisin BPN' (Sbt149).[313] The calcium binding domain has been removed and 18 stabilising mutations made including the introduction of a disulfide bridge. Subsequently, the introduction of two further mutations (Ser212 to Cys and Pro216 to Ala) yields peptiligase. Omniligase-1 is an improved variant of peptiligase that has recently become commercially available.[314] Schmidt *et al* (EnzyPep, Netherlands) demonstrate the use of omniligase-1 as a versatile enzyme for head-to-tail peptide cyclisation of linear peptides bearing a *C*-terminal ester (**Figure 6-2**). Omniligase-1 has been shown to have broad peptide substrate scope.[314]

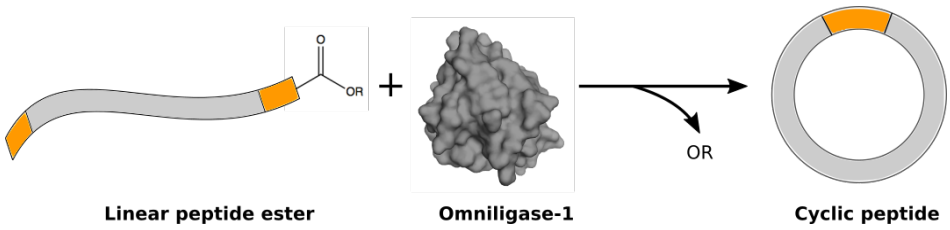


Figure 6-2: Schematic showing cyclisation of linear peptide esters using omniligase-1.

Omniligase-1 has six specific substrate recognition pockets: four recognise the *C*-terminal part of the peptide (S1-S4), and two are involved in binding the *N*-terminal acyl acceptor part of the peptide (S1' and S2').[314] To ensure efficient ligation, hydrophobic amino acids are preferred at position 4 (P4). Positions 1, 1' and 2' (P1, P1' and P2') should avoid Pro. (**Figure 6-3**).

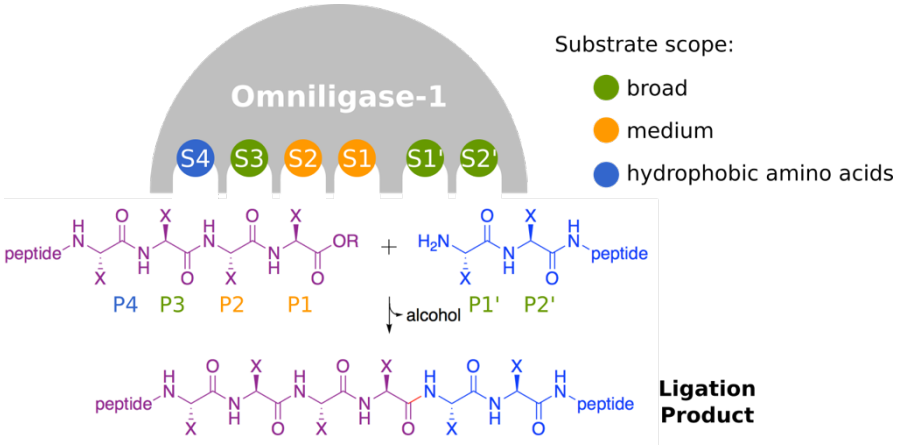


Figure 6-3: Schematic representation of the enzyme pocket of Omniligase-1 highlighting the substrate scope at each position.

6.2 Effects of chain length on the formation and stability of the oPPα topology

With the optimized PPα sequence in hand, the effect of chain length on the formation and stability of the oPPα topology was investigated. First, and simply, oPPα variants with 1-, 3- and 4- unit polyproline-II-helical repeats and correspondingly 2-, 4- and 5- heptad α-helical repeats were synthesised (**Table 6-1**). oPPα completes the series with two polyproline-II units and three α-helical repeats. For all sequences the final two Pro residues at the *N* terminus were maintained.

Table 6-1: Sequences of the series of peptides with chain length ranging from 2 to 5 heptads.

Peptide	Sequence and helical register
	321321321321 efgabcd efgabcd efgabcd efgabcd efgabcd
oPPα-2	Ac- PPKKP GDNAT PEKLAAY EKELAAY-NH_2
oPPα	Ac- PPKKPKKP GDNAT PEKLAAY EKELAAY EKELAAY-NH_2
oPPα-4	Ac- PPKKPKKPKKP GDNAT PEKLAAY EKELAAY EKELAAY EKELAAY-NH_2
oPPα-5	Ac-PPKKPKKPKKPKKP GDNAT PEKLAAY EKELAAY EKELAAY EKELAAY EKELAAY-NH_2

The oPPα-X series of peptides were first analysed by CD spectroscopy (**Figure 6-5**). Unsurprisingly, the shortest peptide, oPPα-2, was the least folded in the series with an α helicity of 35 % at 5 °C, and a correspondingly low T_M of 19 °C. The larger construct, oPPα-4, was significantly more folded and thermally stable than oPPα with a T_M of 66 °C and the longest peptide, oPPα-5, was the most thermally stable in the series with a T_M of 72 °C.

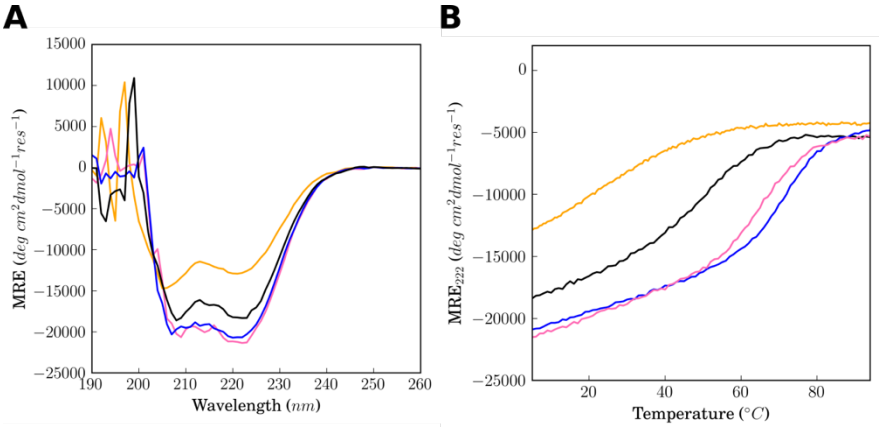


Figure 6-4: CD data showing the effect of chain length on the folding and stability in oPPα. (A) CD spectra recorded at 5 °C for the oPPα-X series: oPPα-2 (orange), oPPα (black), oPPα-4 (pink) and oPPα-5 (blue). Temperature dependence of the CD signal monitored at 222 nm for oPPα-X series, same colour scheme as for (A). Conditions: 100 μM peptide, PBS, pH 7.4.

Overall, stability increases with chain length in the oPPα topology. However, similar to what is observed for α-helical coiled coils, this occurs in a non-linear cooperative manner (**Figure 6-5**). The most significant increase in thermal stability is observed between oPPα-2 and oPPα-3 with an increase of 32 °C. A smaller increase in stability of 15 °C is seen between oPPα-3 and oPPα-4 while there is only a 6 °C difference in the stability between oPPα-4 and oPPα-5.

Peptide name	**$MRE_{222\ nm}$ / deg cm^2 $dmol^{-1}$ res^{-1}**	**% Helicity[175-176]**	**T_M / °C**
oPPα-2	-12822	35	19
oPPα	-18319	48	51
oPPα-4	-21356	54	66
oPPα-5	-20670	52	72

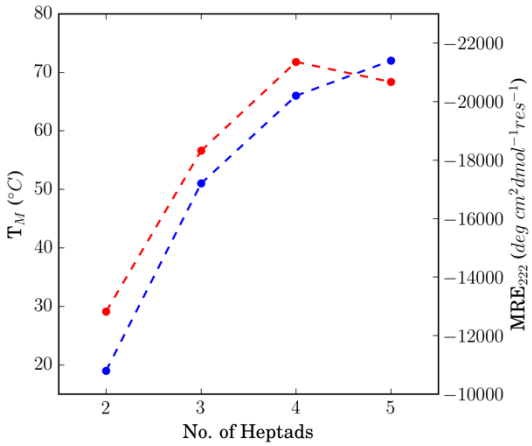


Figure 6-5: Effect of chain length on the folding and stability of oPPα. (Top) Table summarising the MRE_{222} at 5 °C, the % helicity and the T_M for all the oPPα-X peptides. (Bottom) Plot of heptad length vs T_M (blue) and MRE_{222} at 5 °C (red) for oPPα-X peptides. Dashed lines to help guide the eye.

To determine the oligomeric state of the new longer oPPα constructs SE AUC experiments were performed. Experimental data for each peptide readily fitted to a single ideal species models and confirmed all peptides existed as monomers in solution (**Figure 6-6**).

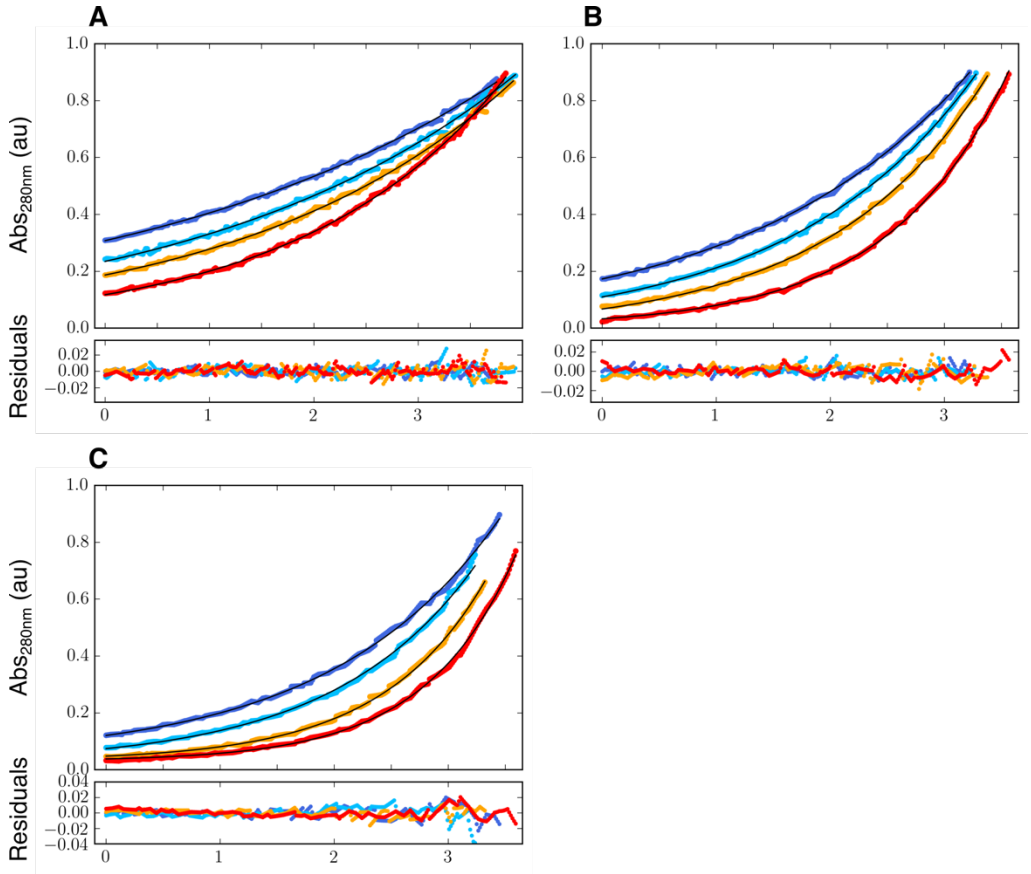
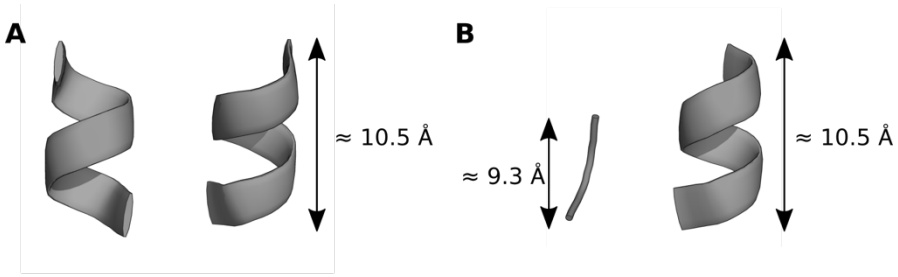


Figure 6-6: Sedimentation-equilibrium AUC data (top, dots) and fitted (lines) single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), and 60 (red) krpm for oPPα-2, oPPα-4 and oPPα-5. Residuals (bottom) for the above fit using the same colour scheme. (A) oPPα-2 ($\bar{v}$ = 0.742 cm^3 g^{-1}), (B) oPPα-4 ($\bar{v}$ = 0.730 cm^3 g^{-1}), (C) oPPα-5 ($\bar{v}$ = 0.753 cm^3 g^{-1}). The fits returned the following masses: oPPα-2, 2503 Da, (0.9 × monomer mass), 95% confidence limits = 2491 – 2514 Da; oPPα-4, 4846 Da, (1.0 × monomer mass), 95% confidence limits = 4826 – 4867 Da; oPPα-5, 6402 Da, (1.0 × monomer mass), 95% confidence limits = 6375 – 6429 Da. Conditions for AUC: 130 μM, PBS, pH 7.4.

6.2.1 Periodicity mismatch in the oPPα topology

In Crick's model of α-helical coiled coils, the 7-residue repeating unit (heptad) of each helix spans the same distance and therefore alignment between the two helices is maintained along the coiled coil.[135, 197] However, in oPPα the 3- and 7-residue repeats of the polyproline-II and α helices do not quite match structurally. Polyproline-II helices have a rise per residue of approximately 3.1 Å and therefore span ≈ 9.3 Å per helical repeat; while α helices have a rise per residue of approximately 1.5 Å and span ≈ 10.5 Å per heptad repeat (**Figure 6-7A&B**). This could result in mismatches over the lengths of the helical rods of up to 6 Å for oPPα-5 which in turn could destabilize the helix–helix interactions that we seek to optimize (**Figure 6-7C**).



C

Peptide name	PPII helix theoretical length / Å	α Helix theoretical length / Å	Mismatch / Å
oPPα-2	18.6	21	2.4
oPPα	27.9	31.5	3.6
oPPα-4	37.2	42	4.8
oPPα-5	46.5	52.5	6

Figure 6-7: Sequence repeats in α-helical coiled-coils and oPPα. (A) Heptad repeat in α-helical coiled coil highlighting distance spanned (1.5 x 7 = 10.5 Å). (B) 3- and 7- residue repeats of the polyproline-II and α helices highlighting distance spanned (polyproline-II helix 3.1 x 3 = 9.3 Å). (C) Table highlighting the mismatch in periodicities between the polyproline-II and α helices at varying chain lengths.

Methods to reduce the mismatch in periodicities and realign the two helices were investigated. Skips, stutters and stammers are relatively prevalent in α-helical coiled coils.[315-316] A skip is defined as the insertion of one reside into the heptad pattern, a stutter is the insertion of four residues and a stammer is the insertion of three residues. By the same logic, a skip can also be thought of as the insertion of two stutters. Stutters straighten the helices by decreasing the degree of supercoiling in a left handed coiled coil while stammers increase supercoiling.[315] When stutters and stammers are repeated regularly along the helix, as opposed to just being local discontinuities, they lead to coiled coils that deviate globally from the Crick model and can lead to a range of periodicities. For example inclusion of stutters leads to hendecad repeats (7+4=11) and pentadecad repeats (7+4+4=15).[118] The supercoil of these coiled coils is determined by the deviation of their periodicity from the 3.63 residues per turn of an undistorted coiled coil. Crick coiled coils with periodicities of 3.5 (7/2, heptad) are left-handed, hendacads with 3.67 (11/3) are virtually straight, and pentadecads with 3.75 (15/4) are right-handed. Stammers are less common in natural coiled coils since the super coil needs to compensate for their 3.3 periodicity (10/3). This imposes a lot of strain on the helix and is at the limit of allowed α-helical backbone conformation.[118]

Translating this logic to oPPα, helix realignment could be achieved in two ways; either the introduction of a stutter into the α helix or a skip into the polyproline-II helix. Indeed the bacterial surface adhesin Ag I/II from *S. mutans* appears to adopt the latter in the form of a Pro skip (PXXPPXX) on the polyproline-II helix that intersperses the PXX repeat (**Figure 6-8**).[166] The α-PP interface in AgI/II is three full polyproline-II helix repeat units before a skip is introduced therefore

a Pro skip could improve the stability of the longest oPPα-5 peptide which has four full polyproline-II helix repeat units.

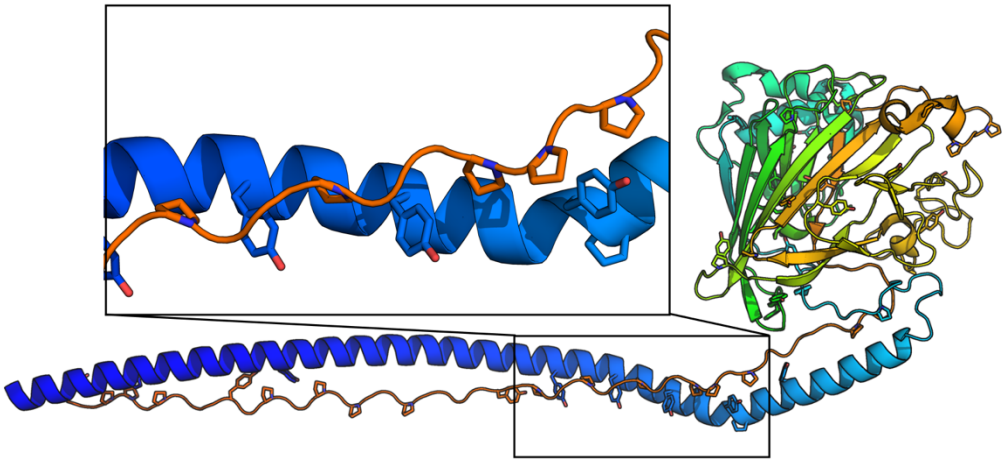


Figure 6-8: The bacterial surface adhesin (PDB: 3IOX) from *S. mutans.* highlighting the interdigitation of the Pro and Tyr residues between the extended α helix and polyproline-II helix and the Pro skip on the polyproline-II helix breaking the PXX repeat. AgI/II is coloured blue to orange from *N* to *C* terminus.

A skip was chosen to be incorporated onto the polyproline-II helix over a stutter on the α helix since naturally related proteins use Pro skips on the polyproline-II helix. Also introduction of a stutter on the α helix was anticipated to be more disruptive. As such, a Pro skip was introduced to the middle of the polyproline-II repeat of the oPPα-5 sequence (**Table 6-2**).

Table 6-2: Sequence of oPPα-5 with the incorporation of a skip on the polyproline-II helix.

Peptide	**Sequence and helical register**
	321321321321 efgabcd efgabcd efgabcd efgabcd efgabcd
oPPα-5	Ac-PPKKPKKPKKPKKP GDNAT PEKLAAY EKELAAY EKELAAY EKELAAY EKELAAY-NH_2
oPPα-5-skip	Ac-PPKKPKK**PP**KKPKKP GDNAT PEKLAAY EKELAAY EKELAAY EKELAAY EKELAAY-NH_2

oPPα-5-skip showed similar helicity to oPPα-5 at 5 °C but was slightly less thermally stable with a T_M of 68 °C. This suggests that the oPPα fold is plastic in nature and can tolerate minor mismatches in periodicities without the need for helix realignment. However, this may not be the case for longer, fibrous-like assemblies such as AgI/II. The slightly reduced T_M of oPPα-5 may also suggest that the skip position is not optimally placed. Future investigation is required to vary the position of the skip along the polyproline-II helix. oPPα-5-skip was also found to be monomeric in solution by SE AUC.

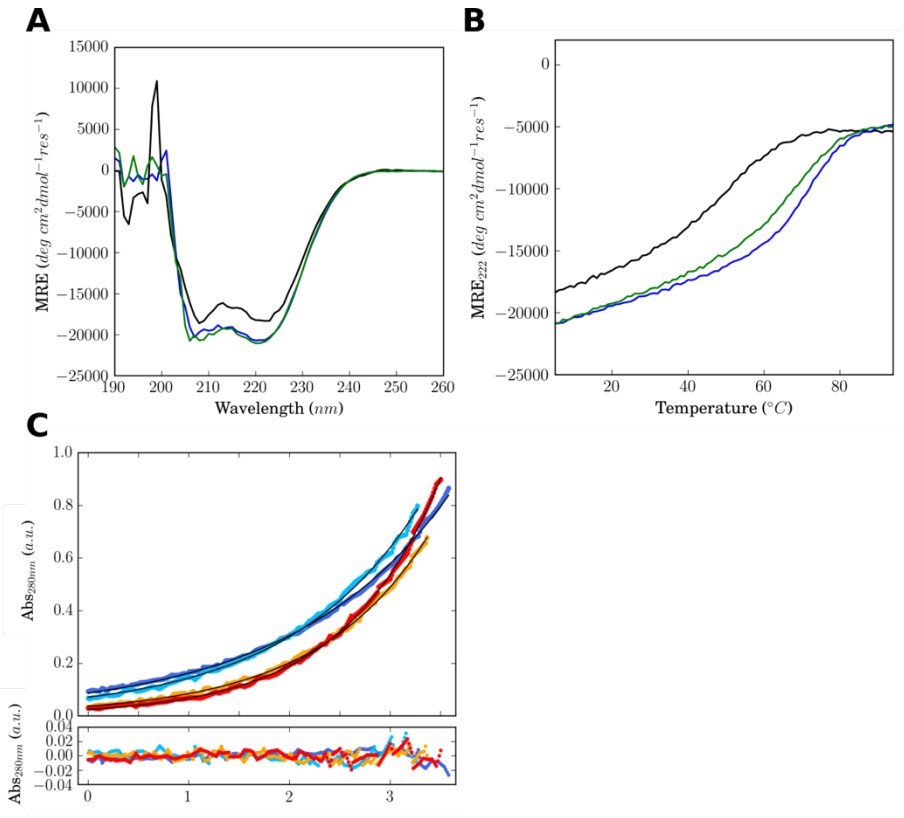


Figure 6-9: Folding and stability of oPPα-5-skip. (A) CD spectra recorded at 5 °C for oPPα (black), oPPα-5: (blue), and oPPα-5-skip (green). (B) Temperature dependence of the CD signal monitored at 222 nm. Same colour scheme as for (A). Conditions for CD experiments: 100 μM peptide, PBS, pH 7.4.(C) AUC Sedimentation-equilibrium data (top, dots) and fitted single-ideal species model curves (lines) at 44 (blue), 52 (aqua), 56 (orange), and 60 (red) krpm for oPPα-5-skip. Bottom: Residuals for the above fits using the same colour scheme. oPPα-5-skip $\bar{v}$ = 0.753 cm^3 g^{-1}. The fit returned a mass of 6085 Da (1.0 × monomer mass), 95% confidence limits = 6126 – 8257 Da. Conditions for AUC: 130 μM, PBS, pH 7.4.

6.2.2 Helix capping in the oPPα topology

As mentioned earlier oPPα possesses an *N* terminal PP. This double Pro was maintained across the various chain lengths. This was because the NMR structures of oPPα and PPα show the PP interacting with the *C* terminus of the α helix suggesting PP is a good helix cap. To determine the effect on stability of this PP cap oPPα-4 was also synthesised without the final Pro residue (**Table 6-3**).

Table 6-3: Sequence of oPPα-4 with the removal of the double Pro cap.

Peptide	Sequence and helical register
	321321321 efgabcd efgabcd efgabcd efgabcd
oPPα-4	Ac- **PP**KKPKKPKKP GDNAT PEKLAAY EKELAAY EKELAAY EKELAAY-NH_2
oPPα-4-no-cap	Ac- **P**KKPKKPKKP GDNAT PEKLAAY EKELAAY EKELAAY EKELAAY-NH_2

oPPα-4 and oPPα-4-nocap were similar in the CD spectra. oPPα-4-no-cap was marginally less folded at 5 °C compared to oPPα-4 and also slightly less thermally stable with a T_M of 62 °C. This suggests that the double Pro cap helps anchor the *N*-terminus of the polyproline-II helix against the α helix. SE AUC experiments showed oPPα-4-no-cap was monomeric in solution.

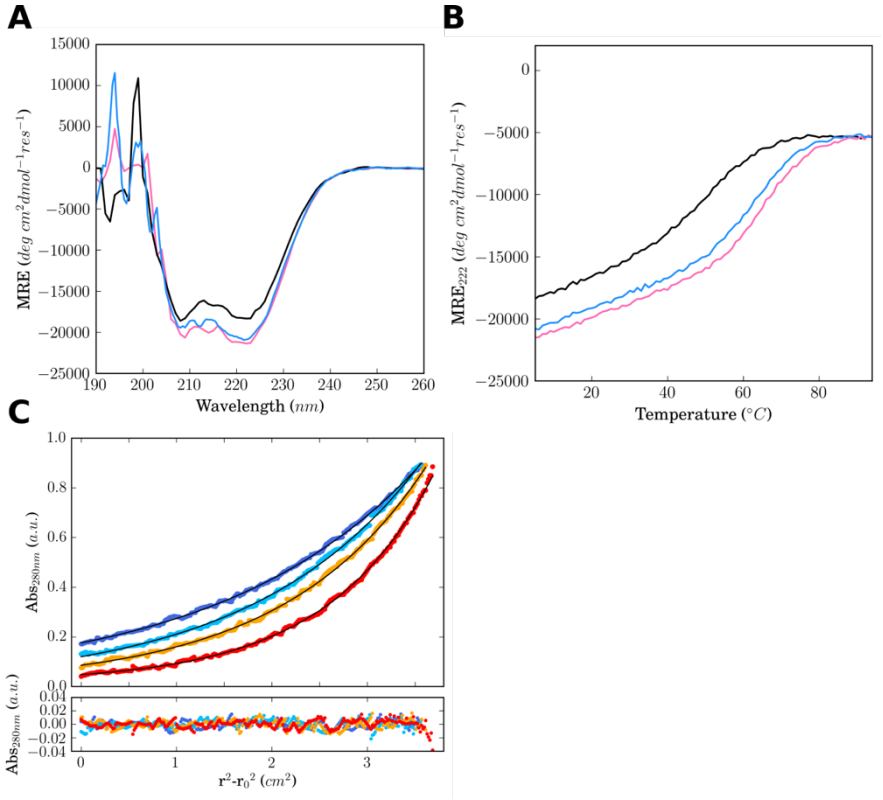


Figure 6-10: Folding and stability of oPPα-4-no-cap. (A) CD spectra recorded at 5 °C for oPPα (black), oPPα-4: (pink), and oPPα-4-no-cap (sky blue). (B) Temperature dependence of the CD signal monitored at 222 nm. Same colour scheme as for (A). Conditions for CD experiments: 100 μM peptide, PBS, pH 7.4.(C) AUC Sedimentation-equilibrium data (top, dots) and curves fitted to a single-ideal species model (lines) at 44 (blue), 52 (aqua), 56 (orange), and 60 (red) krpm for oPPα-4-no-cap. Bottom: Residuals for the above fits using the same colour scheme. oPPα-4-no-cap $\bar{v}$ = 0.751 cm^3 g^{-1}. The fit returned a mass of 4488 Da (0.9 × monomer mass), 95% confidence limits = 4466 – 4510 Da. Conditions for AUC: 130 μM, PBS, pH 7.4.

6.2.3 Models and MD simulations of the oPPα-X series

Models of the oPPα-X series of peptides were constructed in PyMol based on the bacterial surface adhesin AgI/II structure (PDB: 3IOX). Models were then energy minimised over 10 ns of MD simulations in water (**Figure 6-11**).

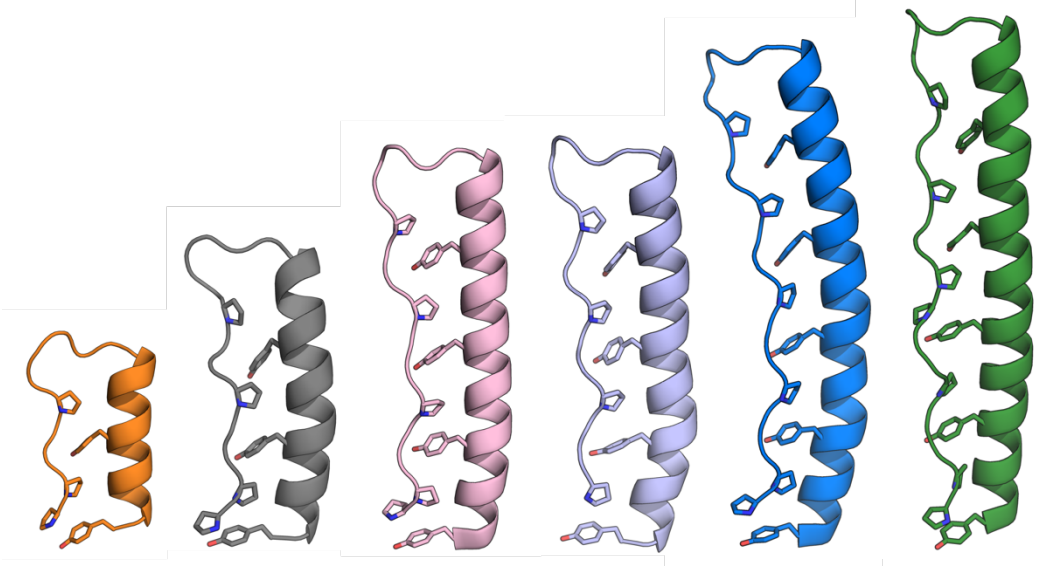


Figure 6-11: Models of the oPPα-X Series after 10 ns of MD simulations in water. oPPα-2 (orange), oPPα (NMR structure, state 8, black), oPPα-4 (pink), oPPα-4-no-cap (lilac), oPPα-5 (blue), oPPα-5-skip (green).

While 10 ns is not long enough for the peptides to reach their most stable conformation or unfold, the short simulation gives an idea of the stability of the peptides. Also given the repetitive and degenerate nature of these sequences solving an NMR structure for them would be very challenging and therefore these MD simulations provide an initial model to help visualise these structures. All models were found to be well folded over the 10 ns trajectory and had relatively low RMSD values between the initial and 10 ns timepoints (**Table 6-4**).

Table 6-4: RMSD between 0 and 10 ns timepoints for the oPPα-X series during MD simulations.

Peptide	All atom RMSD / Å	Backbone RMSD / Å
oPPα-2	2.012	1.103
oPPα-4	3.018	2.081
oPPα-4-no-cap	2.056	1.063
oPPα-5	2.835	1.690
oPPα-5-skip	2.287	1.363

In brief summary, a set of oPPα variants have been presented ranging in length from two to five α-helical heptads. The different lengths allow access to peptides with a range of stabilities. As observed for α-helical coiled coils, stability increases with chain length in a non-linear manner. For the lengths tested a skip region is not necessary in order to maintain a periodicity match between the two helices. The double PP at the *N*-terminus of the polyproline-II helix appears to act as a helix cap adding to the constructs stability.

6.3 Enzymatic cyclisation of oPPα with omniligase-1

6.3.1 Cyclic oPPα design

Next, the cyclisation of oPPα was explored as an alternative strategy to increase stability. The cyclisation strategy discussed in **Section 6.1.3.1** was applied to oPPα. In order to cyclise oPPα a linker between the *N*-terminal end of the polyproline-II helix and the *C*-terminal end of the α helix had to be introduced. Inspection of the oPPα NMR structure revealed the distance between the *N*-terminal nitrogen of Pro1 and the *C*-terminal carboxyl carbon of Tyr34 is 8.4 ± 0.4 Å. This is similar to the distance between the *C*-terminal carboxyl carbon of Pro8 and the *N*-terminal nitrogen of Pro14 (8.8 ± 0.4 Å) for which a five-residue loop connects the two secondary structure components (**Figure 6-12A**). Therefore, a similar length loop could be satisfactory in joining the *N*-terminal end of the polyproline-II helix and the *C*-terminal end of the α helix while still maintaining the interface between the two helices. A loop of four to six residues in length was therefore deemed appropriate.

The position of cyclisation was considered next. The cyclisation site could be part of the loop region or within either the polyproline-II or α helix components. Given Pro should be avoided in the recognition sequence for omniligase-1 the polyproproline-II helix was discounted as a cyclisation site.

Therefore, both cyclisation in the α helix and loop region were investigated (**Figure 6-12B**). There were advantages and disadvantages to both cyclisation strategies. Loop cyclisation would allow a sequence to be selected that is most optimal for the enzyme and therefore give the cyclisation the best chance of success. α-Helix cyclisation requires a six residue sequence frame on the α helix to be chosen that is compatible with omniligase-1. This could result in a sequence frame that is not fully optimal for omniligase-1 or require mutations that disrupt the oPPα assembly. However, the α helix cyclisation strategy allows a sequence for Loop2 (**Figure 6-12B**) to be selected that is optimal for bridging the gap between the two secondary structure components without worrying about enzyme compatibility. Also, we anticipate that the peptide undergoing α-helix cyclisation would be partially unfolded before enzyme addition, as a result of the break in α-helical secondary structure, therefore facilitating formation of the enzyme-substrate intermediate and successful cyclisation. The peptide undergoing loop cyclisation would likely be folded pre-enzyme addition and therefore may hinder cyclisation.

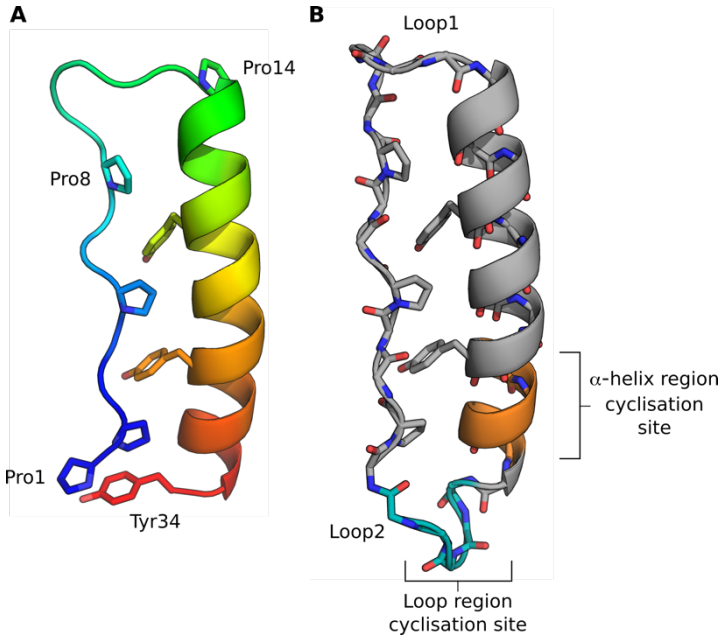


Figure 6-12: oPPα and cyclic variant. (A) State 8 of NMR strucutre of oPPα. (B) Model of cyclic oPPα generated in ISAMBARD highlighting the two different cyclisation sites examined. Colour key: *N*-terminus (blue) and *C*-terminus (red).

Based on unpublished work by Schmidt *et al.*, a six residue loop sequence was selected based on its compatibility with omniligase-1. The $P_4P_3P_2P_1P_1'P_2'$sequence chosen was LSTKDL. The enzyme recognition sequence avoided Pro and ensured a hydrophobic residue at P_4 (leucine). For cyclisation in the α helix region, a six residue recognition sequence frame was selected as AYEKEL between heptad two and three ensuring P_4 was a hydrophobic residue (Ala) (**Table 6-5**).

Table 6-5: Sequence of the cyclised oPPα peptides.

Cyclisation Location	Peptide	Sequence
Loop	oPPα-LSTKDL	DL PPKKPKKP GDNAT PEKLAAY EKELAAY EKELAAY LSTK -OCam-L-OH
	cPPα-LSTKDL	DL PPKKPKKP GDNAT PEKLAAY EKELAAY EKELAAY LSTK
α helix	oPPα-GPQY	ELAAY GPQY PPKKPKKP GDNAT PEKLAAY EKELAAY EK-OCam-L-OH
	cPPα-GPQY	ELAAY GPQY PPKKPKKP GDNAT PEKLAAY EKELAAY EK

Loop2, connecting the *N*-terminal end of the polyproline-II helix with the *C*-terminal end of the α helix, was chosen based on a bioinformatics analysis of loop residues in the PDB. A Top8000 dataset from the Richardson lab[317] was used. This dataset comprises 7957 high resolution structures. Only structures of the highest quality are selected to this dataset and criteria for selection include a crystal structure resolution of < 2 Å and a MolProbity[318] score of < 2. MolProbity is a structure validation web service that assesses the quality of a model at both the global and local level, for example ensuring Ramachandran outliers are limited.

The Richardson dataset was further refined by culling PDB entries that had greater than 25% sequence identity using the Pisces Server.[319] This yielded 3252 PDB files. From this dataset, structures containing an *N*-terminal α helix connected to a *C*-terminal polyproline-II helix *via* a loop region were selected. The program Assignment of Secondary Structure in Proteins (ASSP)[320] was used to assign secondary structure. The most common methods of secondary structure assignment such as Dictionary of Protein Secondary Structure (DSSP)[321] and STRIDE[322] were not used since they omit definitions for polyproline-II helices.[323]

The ASSP method assigns secondary structure by first calculating local geometric parameters: twist, rise per residue, virtual torsion angle, and radius. The full length of the protein chain is scanned and parameters for blocks of four C_α atoms are calculated. The continuity in the protein structure is checked based on these parameters and continuous stretches are further divided into secondary structure elements. Assignments include right- or left- handed α helix, right- or left- handed 3_{10} helix, right- or left- handed π helix, left-handed polyproline-II helix, extended conformation (β-strands). If a stretch of protein does not fall into any of these categories it is labelled as undefined. The minimum possible length for α and polyproline-II helices are four and three residues respectively.

In our analysis a loop region was specified by having 'undefined' secondary structure. 734 structures were obtained that had an *N*-terminal α helix connected to a *C*-terminal polyproline-II helix by such a region. A range of loop lengths from zero to four residues were extracted (**Table 6-6**). Interestingly, no loops longer than four residues were found. This is likely a result of longer loop regions containing other regions of secondary structure and not just strictly 'undefined' peptide. Note that PPα is not present in the dataset since it does not have an *N*-terminal α helix component; the α helix is *C*-terminal.

Table 6-6: Frequency of loops of ranging lengths connecting *N*-terminal α helix to *C*-terminal polyproline-II helix.

Loop Length	Number of Loops
0	127
1	250
2	150
3	136
4	116

Since four-to-six residue loops were most desirable and there were no five or six residue loops identified, we focused on the four-residue loops in the dataset. Of the 116 structures 60 % were in a antiparallel conformation. The propensity for residues at each position in the four residue loop was then calculated (**Figure 6-13**). The most preferred residue at each position were (Gly/Asn, Gly/Pro, Pro/Gln, Gly/Tyr) equating to (Position 1, Position 2, Position 3, Position 4). Based on this analysis the four residue loop GPQY was selected for the cyclic structure in which the ligation site was located in the α helix.

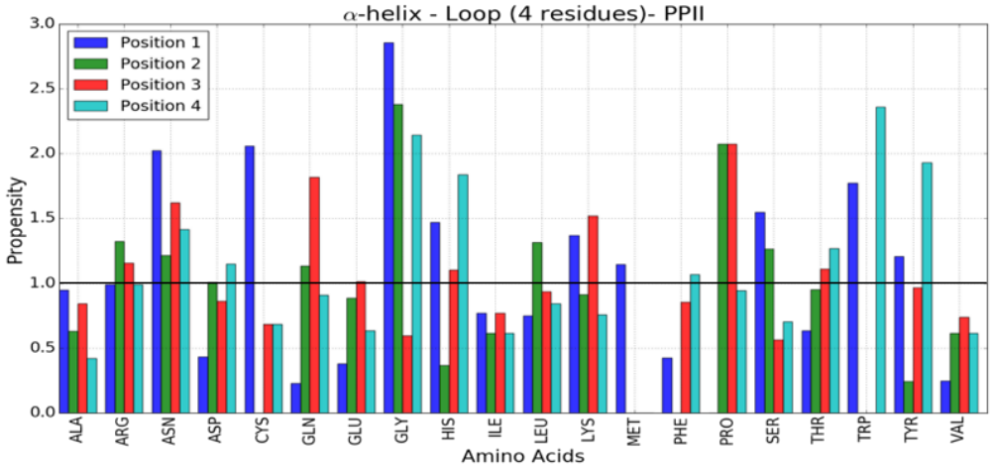


Figure 6-13: Barchart highlighting the propensity for the amino acid residues in the four residue loop connecting an α helix to a polyproline-II helix.

6.3.2 Synthesis and biophysical characterisation of cyclic oPPα variants

The linear peptides bearing *C*-terminal activated carboxyamidomethyl (Cam) ester were synthesised by SPPS and then enzymatically cyclised with omniligase-1 by Marcel Schmidt. Both cyclic oPPα variants were obtained, despite evidence of unwanted ester hydrolysis side reactions, with cPPα-GPQY being the most facile. As anticipated, the pre-folded nature of the peptides, in particular cPPα-LSTKDL, may have inhibited the formation of the enzyme-substrate intermediate.

By CD spectroscopy, both peptides appeared to be less folded at 5 °C compared to oPPα (**Figure 6-14**). However, thermal unfolding experiments of the far-UV CD revealed cPPα-GPQY and cPPα-LSTKDL both underwent cooperative and fully reversible transitions with T_M of 70 °C and 73 °C, respectively. This is a substantial increase over that of oPPα (51 °C).

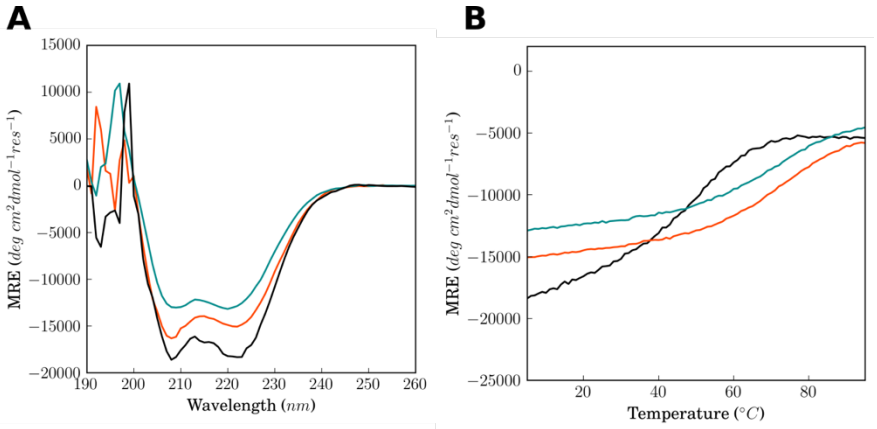


Figure 6-14: Folding and stability of cPPα variants. (A) CD spectra recorded at 5 °C for oPPα (black), cPPα-LSTKDL (red) and cPPα-GPQY (teal). (B) Temperature dependence of the CD signal monitored at 222 nm. Same colour scheme as for (A). Conditions: 100 μM peptide, PBS.

We posit that some of the reduction in helicity of both cyclic peptides at 5 °C can be attributed to the introduction of either a four or six residue loop region which should not contribute to the α-

helical CD signal however is included in the normalisation calculation. Therefore the data for the cyclic peptides were renormalised for the residues in oPPα only (i.e. 34 residues not 40 or 38 residues for oPPα-LSTKDL and oPPα-GPQY, respectively). The renormalized CD spectra are shown in **Figure 6-15**. The degree of folding of cPPα-LSTKDL at 5 °C appears comparable to that of oPPα, while cPPα-GPQY remains less folded than oPPα. This suggests that a four residue loop is not long enough to bridge the gap between the α helix and polyproline-II helix secondary structure components and may have caused fraying of the structure around the loop region.

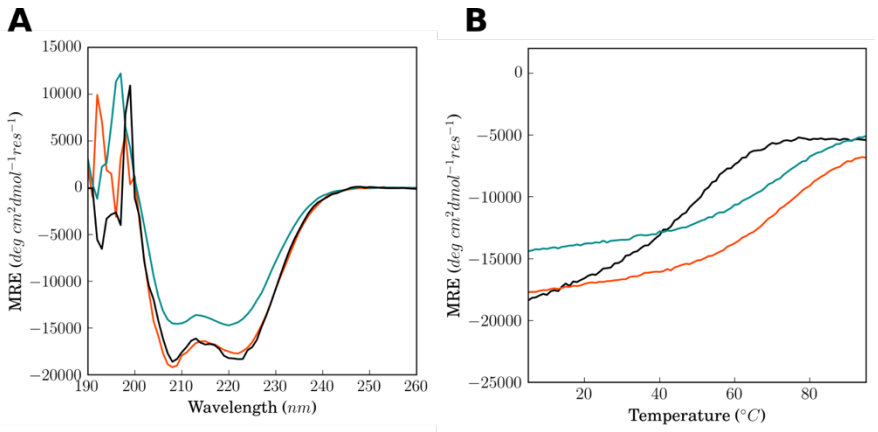


Figure 6-15: Folding and stability of cPPα variants normalised for 34 oPPα residues only. (A) CD spectra recorded at 5 °C for oPPα (black), cPPα-LSTKDL (red) and cPPα-GPQY (teal). (B) Temperature dependence of the far-UV CD signal monitored at 222 nm. Same colour scheme as for (A). Conditions: 100 μM peptide, PBS.

SE AUC experiments confirmed that both cyclic peptides were monomeric in solution. Experimental data readily fitted models to single-ideal species (**Figure 6-16)**.

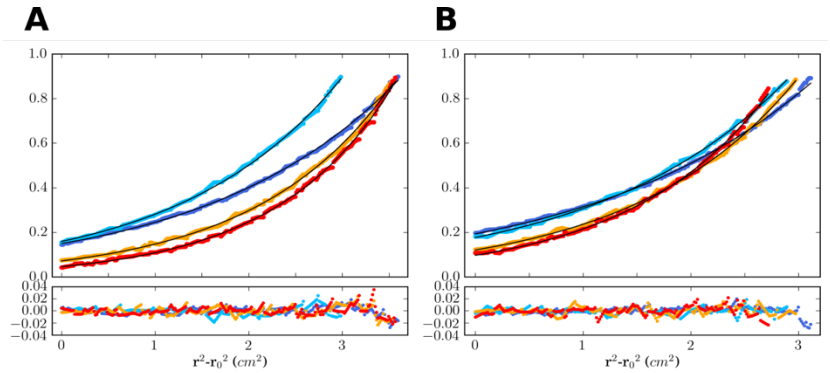


Figure 6-16: Sedimentation-equilibrium AUC data (top, dots) and fitted (lines) curves to single-ideal species model at 44 (blue), 52 (aqua), 56 (orange), and 60 (red) krpm for for cPPα- LSTKDL and cPPα-GPQY. Residuals (bottom) for the above fit. (**A**) oPPα-LSTKDL ($\overline{v}$ = 0.751 $cm^3\ g^{-1}$). The fit returned the following mass: oPPα-LSTKDL, 4100 Da, (0.9 x monomer mass), 95% confidence limits = 4079 – 4121 Da. (**B**) oPPα-GPQY ($\overline{v}$ = 0.743 $cm^3\ g^{-1}$). The fit returned the following mass: oPPα-GPQY, 4315 Da, (1.0 monomer mass), 95% confidence limits = 4268 – 4365 Da. Conditions for AUC: 130 μM, PBS, pH 7.4.

6.4 Chapter conclusion

Two strategies were explored for improving the stability of oPPα. Firstly, the chain length of oPPα was varied. A set of oPPα variants of ranging length from two to five α-helical heptads were designed with T_{MS}s ranging from 19 to 72 °C. Similar to what is observed in α-helical coiled coils the stability of these peptides increases with increasing chain length in a non-linear but cooperative manner.[291-292] The greatest increase in stability is observed between oPPα-2 (T_M 19 °C) and oPPα (T_M 51 °C) with minimal increase in stability between the longest variants oPPα-4 and oPPα-5 (66 to 72 °C).

It was hypothesised that increasing the lengths of the helices may have led to a mismatch in the periodicities resulting from the different secondary structures. Therefore, a Pro skip was introduced to realign Pro and Tyr side chains however this had little effect on the folding or thermal stability of oPPα-5. This suggests that for the lengths investigated the oPPα fold is plastic to some extent and can accommodate small mismatches in periodicities but this may not be the case for longer, fibrous like assemblies.

This set of oPPα variants will be useful for applications in protein engineering and synthetic biology where protein-protein interactions of varying stabilities are required. They will also be useful in the emerging field of peptide origami.[324] While the base complementarity of DNA has been exploited to rationally design artificial nanostructures that can arrange into 2D and 3D structures (DNA origami),[325] the design of such peptide counterparts is significantly more challenging as a result of many cooperative and long range interactions. The first example is the self-assembly of a tetrahedral peptide fold designed from 12 orthogonal coiled-coil peptides joined by short flexible peptide linkers.[151] Graph theory revealed that both parallel and antiparallel coiled-coil components are needed to form a single chain tetrahedron. While both orientations are present in nature significantly more parallel dimers have been isolated and designed. Therefore, these antiparallel oPPα peptides may be useful additions to the toolkit of coiled-coil peptide building blocks and will aid the design of future polyhedra and the construction of complex modular protein assemblies more generally.

Secondly, successful enzymatic cyclisation of oPPα yielding two cyclic peptide variants with significantly enhanced thermal stabilities compared to the linear oPPα peptide. Both peptides had cooperative and fully reversible unfolding transitions with T_{MS} of 70 and 73 °C. Future work will focus on obtaining a high resolution structure of the cPPα mutants. Given the enhanced thermal stability of the cyclic oPPα variants initial attempts were made to crystallise the peptides. Unfortunately crystal trays (laid using standard commercial screens at 20 °C) resulted in no crystals to date however further crystal trays will be laid at higher concentrations and lower temperatures to attempt to reach optimal crystallisation conditions.

Further investigations are also required to determine the proteolytic stability of these cyclic peptides in a biological fluid. If proteolytic stability is high then these oPPα variants will be useful scaffolds for grafting recognition motifs to modulate protein-protein interactions in a cellular environment. More ambitiously they could have the potential to be good architectures for developing miniprotein based drug therapeutics.

Chapter 7 Conclusions and future work

7.1 Overall conclusions

Through the work described in this thesis, an optimised PPα miniprotein has been designed and fully characterised in solution. Sequence-to-stability relationships have been probed through mutagenesis studies. The newly designed oPPα has also been used as a foundation for building two new topologies, which, to the best of our knowledge, have not been observed before in nature: αPP and PPαPP.

Firstly, the PPα minprotein, which is based on sequences from natural peptides, has been rationally redesigned into a completely *de novo* PPα framework. The redesign of PPα was inspired by design principles for α-helical coiled coils including knobs-into-holes packing, electrostatic interactions between helices, and α-helical propensity. The optimised PPα miniprotein has been extensively characterised in solution by a range of biophysical techniques including high-resolution NMR spectroscopy. The parent and optimised design have similar helicities at 5 °C. However, oPPα has significantly enhanced thermal stability. We attribute the enhanced stability of the peptide to a number of factors. While formal salt bridges are not observed between charged side chains on the α and polyproline-II helices, electrostatic steering effects likely contribute to the elevated stability and general tightening of the structure. Furthermore, the number of CH–π interactions detected in the NMR structure of oPPα is increased, with an average of one extra CH–π interaction per ensemble state.

To enhance our understanding of the oPPα fold, sequence-to-stability relationships have been investigated through a series of amino acid mutations to the ***a*** and ***g*** positions on the α helix which are involved in KIH-type packing between the two helices. Studies found that the ***a*** position is tolerant of mutation to small hydrophobic residues Val and Ile as well as charged residue Lys. However, this position is not tolerant of Ala and Glu, which yield unfolded peptides. Overall, as observed in related natural peptides, Leu is the most favoured residue at the ***a*** position. A fully hydrophobic hole for accepting the Pro side chain has been investigated by placing Leu at both the ***a*** and ***g*** positions. While expanding the hydrophobic seam gives a peptide that is thermally very stable, concentration-dependent circular dichroism spectroscopy experiments combined with sedimentation equilibrium analytical ultracentrifugation experiments indicate this is at the cost of a discrete monomeric structure. Finally, Leu at ***a*** and Glu at ***g*** is preferred over Glu at ***a*** and Leu at ***g*** in the oPPα framework.

Optimised PPα and the sequence-to-stability relationships determined in this thesis can be used as a foundation on which to build larger and more complex assemblies based on the PPα fold for

applications in synthetic biology and protein engineering. Towards this goal, an iterative design process has been used to design a stable and folded αPP topology. Initial designs show that simply maintaining the interface between the two helices and joining them using a flexible or suboptimal loop region is not enough to form a stable αPP topology. Both *N*-terminal capping motifs and loop optimisation are important factors to be considered to give a folded and stable αPP topology. *N*-terminal capping motifs were chosen based on the work by Richardson *et al.* and loops were selected using LoopFinder within the ISAMBARD framework.[142, 225] While an αPP topology equivalent in folding and stability to that of oPPα has not been delivered, a reasonably well folded αPP variant has been designed and characterised. Further, significant progress has been made compared to the initial αPP design. oPPα and the new αPP topology were subsequently used in the design of a new extended PPαPP topology. PPαPP shows similar folding and stability compared to oPPα. Mutagenesis experiments reveal that the PPα interface of the topology is the more dominant compared to the αPP interface and when it is removed the stability of the topology is reduced dramatically. An NMR or X-ray crystal structure is needed to verify the structure and determine whether the PPαPP peptide has formed the desired topology.

Finally, two strategies have been implemented to enhance the stability of oPPα: increasing chain length and peptide cyclisation. Consistent with observations for α-helical coiled-coils, the stability of oPPα increases with chain length in a non-linear cooperative manner. The greatest increase in stability is observed between variants with two and three α-helical heptad repeats, while a minimal increase in stability is observed between the longest variants of lengths four and five α-helical heptad repeats. It is hypothesised that increasing the lengths of the helices may lead to a mismatch in the periodicities between the different secondary structures. To test this, a Pro skip was introduced to realign Pro and Tyr side chains. However, this had little effect on the folding or thermal stability of the variant with five α-helical heptad repeats. This suggests the PPα fold is plastic enough to accommodate small mismatches in periodicities within the lengths tested. However, this may not be the case for longer, fibrous like assemblies. This set of oPPα variants will be useful for applications where protein-protein interactions of varying stabilities are required.

Successful enzymatic cyclisation of oPPα has produced two cyclic peptide variants with significantly enhanced thermal stabilities compared to the linear oPPα peptide. Both have cooperative and fully reversible unfolding transitions with T_{Ms} exceeding that of oPPα. Such miniproteins are potentially useful scaffolds for grafting functional motifs for applications *in vivo*.

7.2 Future work

Two new topologies have been designed (αPP and PPαPP) that help protein design move into the dark matter of protein space.[33] Future targets for expanding the PPα topology include αPPα and PPαPPα, given that the PPα component of PPαPP is the more dominant. Further, more ambitious

targets include expanding oPPα to form circular 3D objects. For example, one can envisage forming a single chain hybrid polyproline-II-α-helical barrel structure with alternating polyproline-II and α helices (**Figure 7-1A**). Alternatively an 'inverted' TIM-barrel type structure could be designed with

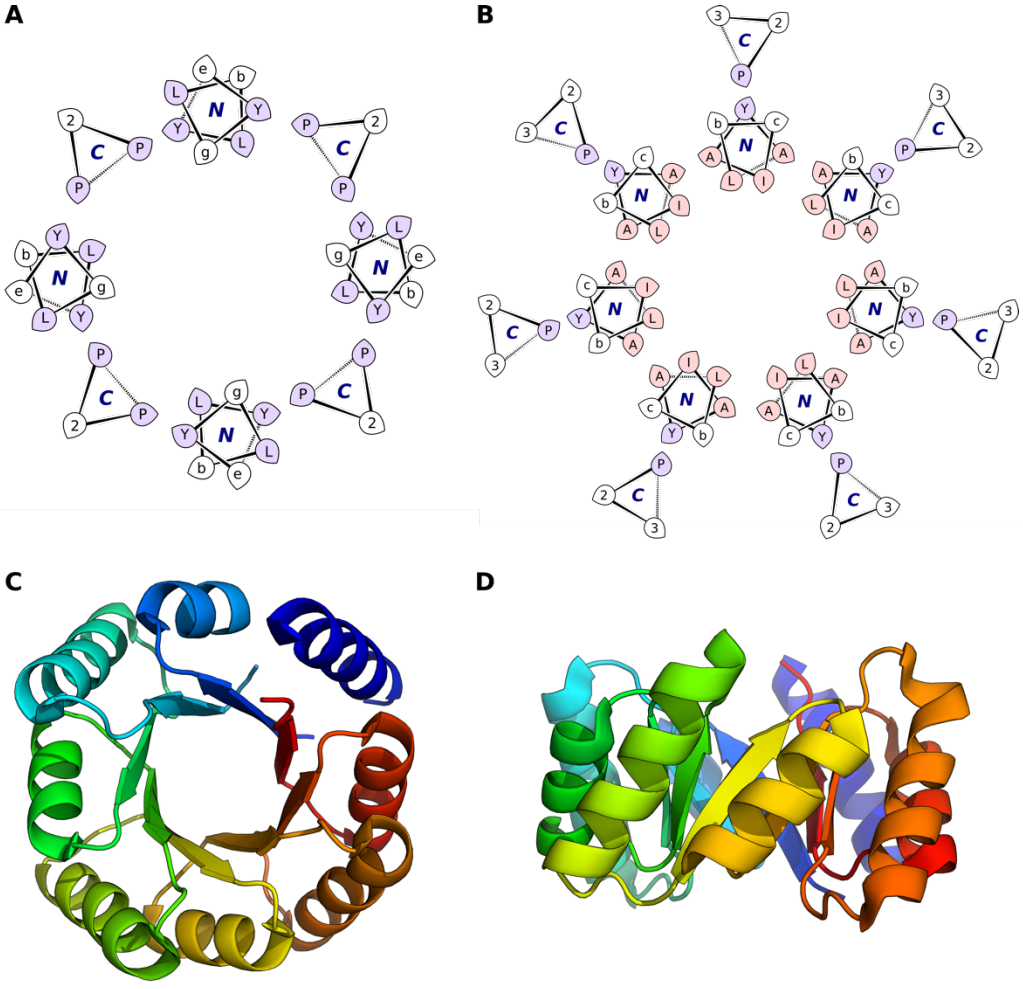


Figure 7-1: Design targets incorporating PPα. (A) Single chain hybrid polyproline-II-α-helical barrel structure with alternating polyproline-II and α helices. (B**)** An 'inverted' TIM-barrel type structure with α helices forming a central α-helical barrel with each helix connected by an external polyproline-II helix. Loops not shown for (A) and (B). (C) Top-down view of *de novo* designed TIM-barrel, PDB 5BVL. (D) Side view of TIM-barrel.

α helices forming a central α-helical barrel with each helix connected by an external polyproline-II helix (**Figure 7-1B-D**).[326]

Analogues to α-helical barrels (specifically CC-Hept), such single-chain barrels could find applications in catalysis and small-molecule binding.[145, 152] Currently CC-Hept is a rudimentary hydrolase. However this is limited by the seven-fold symmetry; one mutation results in seven copies

across the assembly. A single-chain variant would allow point mutations to be made in the assembly and more sophisticated active sites within the lumen of the barrel to be designed.

The compact and stable nature of the oPPα miniprotein makes it a useful scaffold for targeting and disrupting protein-protein interactions. Further the sequence-to-stability relationships established for the fold will help determine which residues are amenable to mutation. The set of oPPα variants of ranging lengths will be useful for controlling and tuning the stability and affinity of the protein-protein interactions. This will be particularly important in systems with complex association-dissociation dynamics.

The cyclic variants will be good starting points for *in vivo* applications where the peptides will need to be resistant to proteolysis and degradation. Future work should focus on determining their proteolytic stability in a biological fluid. If proteolytic stability is high then these variants will be useful scaffolds for grafting recognition motifs for modulating protein-protein interactions in a cellular environment. Motifs could be grafted onto either solvent exposed face of the polyproline-II helix or the α helix. The related natural pancreatic peptides have been used as scaffolds to present α helical and polyproline-II recognition epitopes.[56, 59, 61] Further, and more ambitiously, the cyclic oPPα miniproteins could be good architectures for developing miniprotein-based therapeutics.

Chapter 8 Appendix

8.1 HPLC traces and MALDI-TOF MS for the designed peptide sequences

8.1.1 Chapter 3: Stabilising the PPα miniprotein by rational design

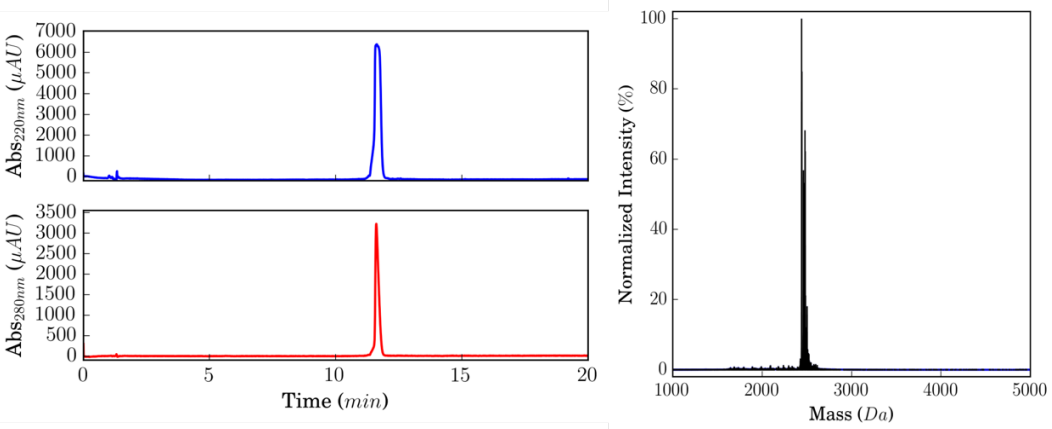


Figure 8-1: o-α - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 2440.242 Da, observed mass = 2440.904 Da.

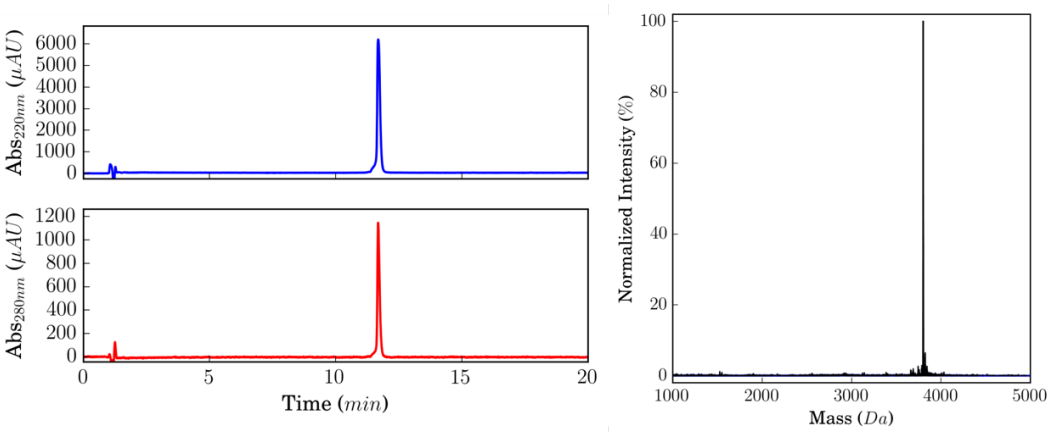


Figure 8-2: oPPα - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 3799.009 Da, observed mass = 3800.901 Da.

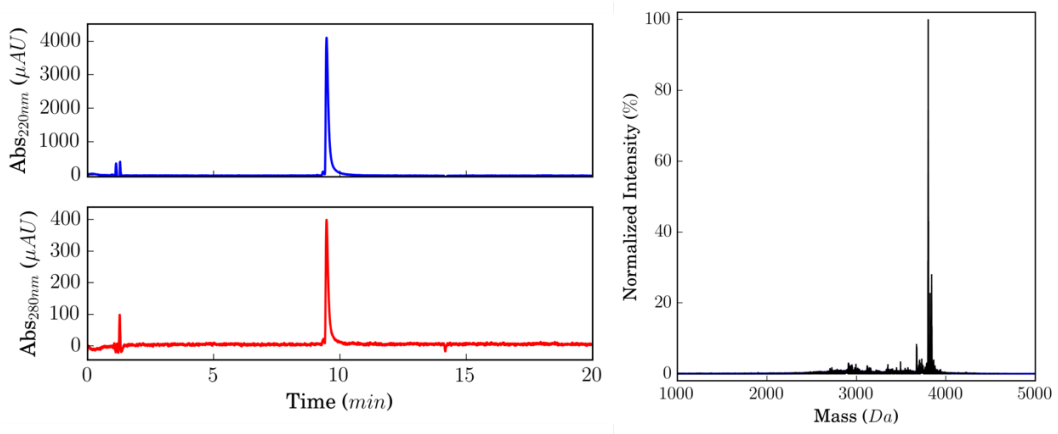


Figure 8-3: oPPα-E↔K - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 3802.219 Da, observed mass = 3802.219 Da.

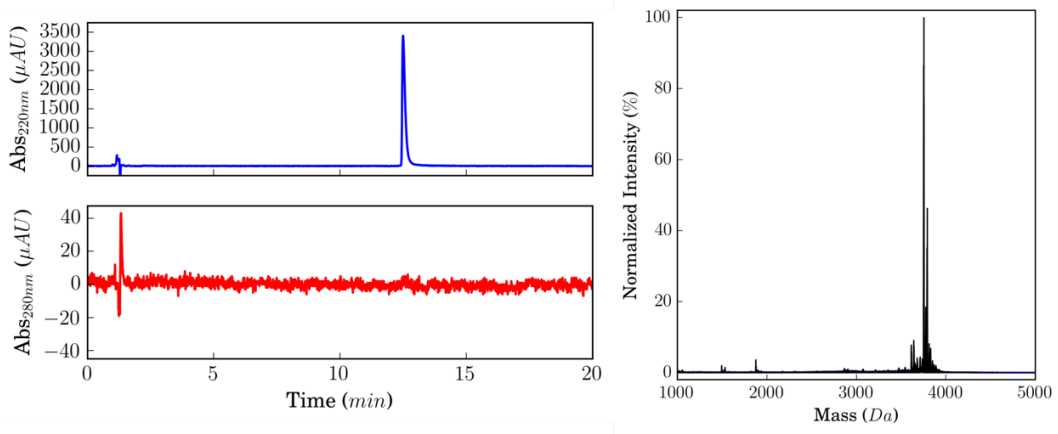


Figure 8-4: oPPα-Phe - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 3752.024 Da, observed mass = 3752.826 Da.

8.1.2 Chapter 4: Probing sequence-to-stability relationships in oPPα

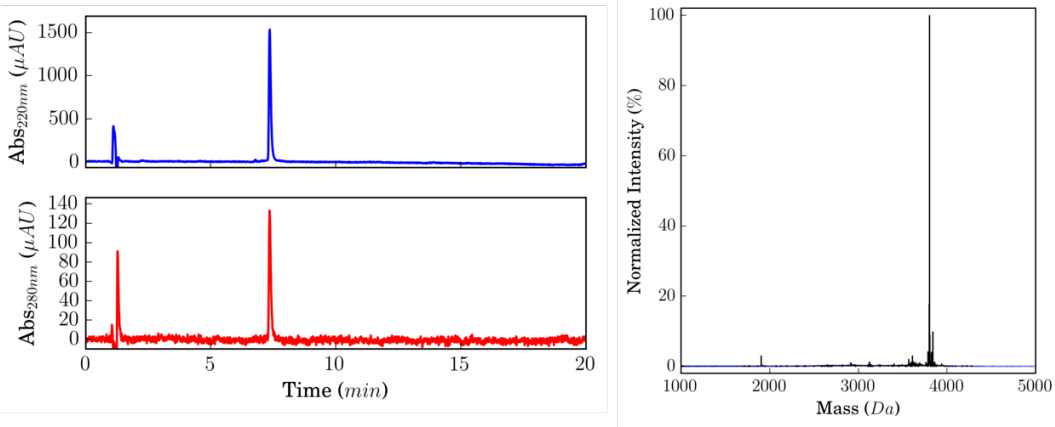


Figure 8-5: oPPα-I@a - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 3752.024 Da, observed mass = 3752.826 Da.

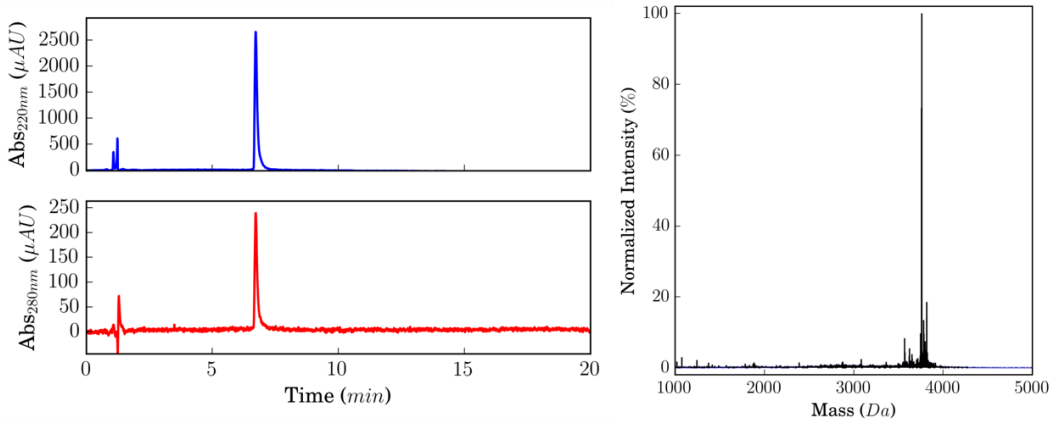


Figure 8-6: oPPα-V@a - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 3752.024 Da, observed mass = 3752.826 Da.

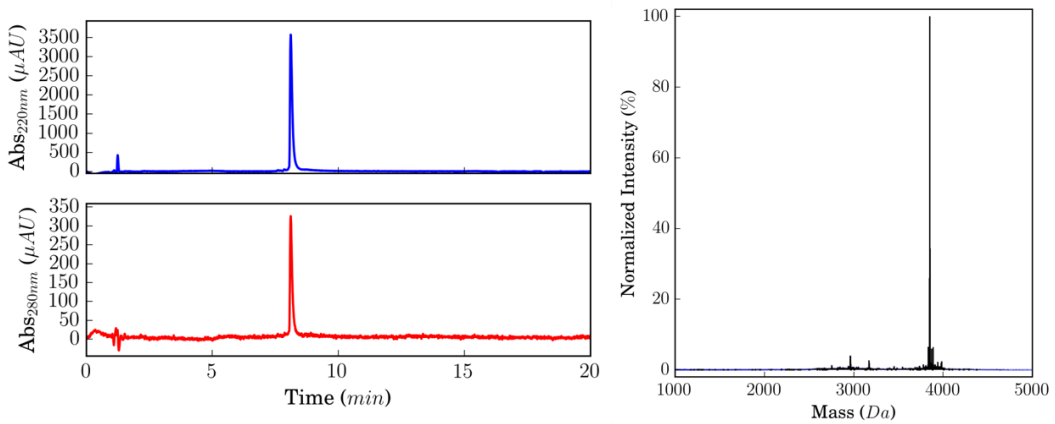


Figure 8-7: oPPα-E@a - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 3752.024 Da, observed mass = 3752.826 Da.

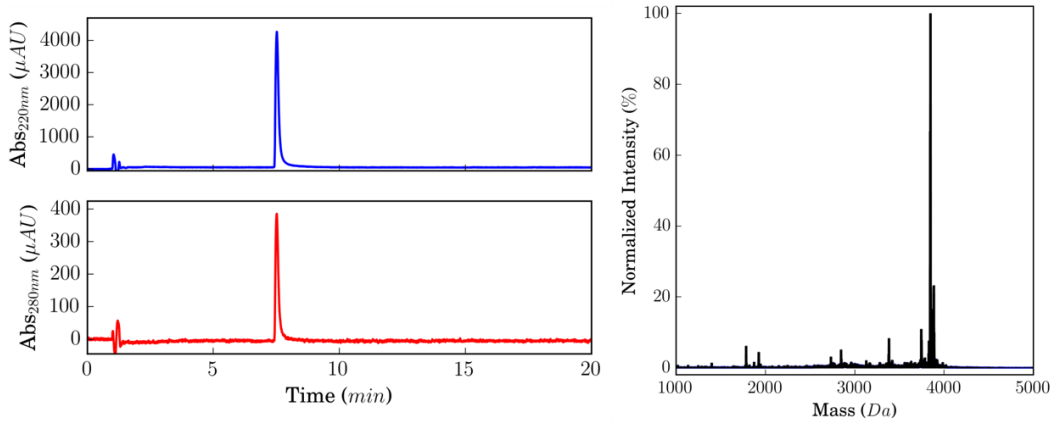


Figure 8-8: oPPα-K@a - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 3846.374 Da, observed mass = 3846.239 Da.

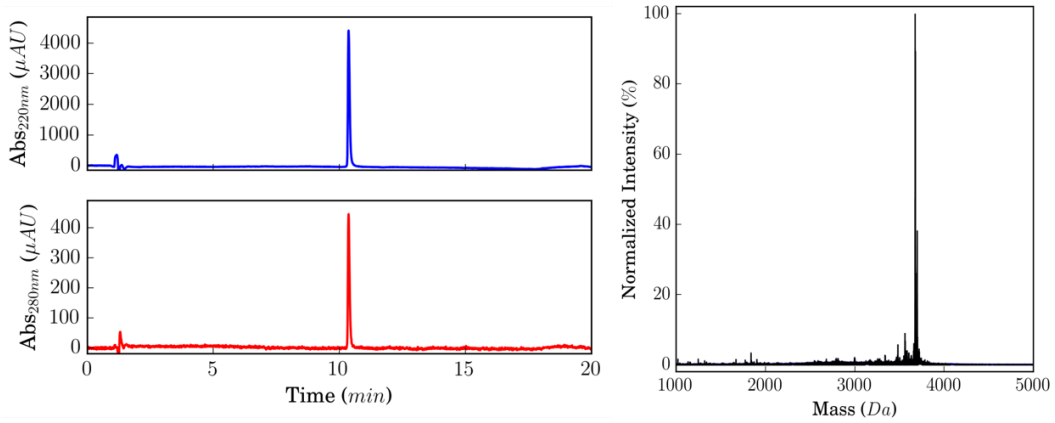


Figure 8-9: oPPα-A@a - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 3675.393 Da, observed mass = 3675.088 Da.

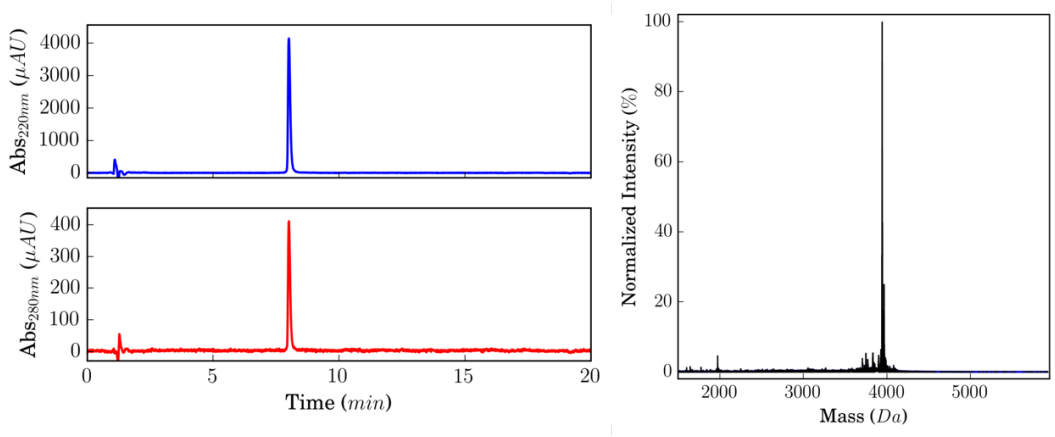


Figure 8-10: oPPα-L@a-L@g - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 3941.109 Da, observed mass = 3946.465 Da.

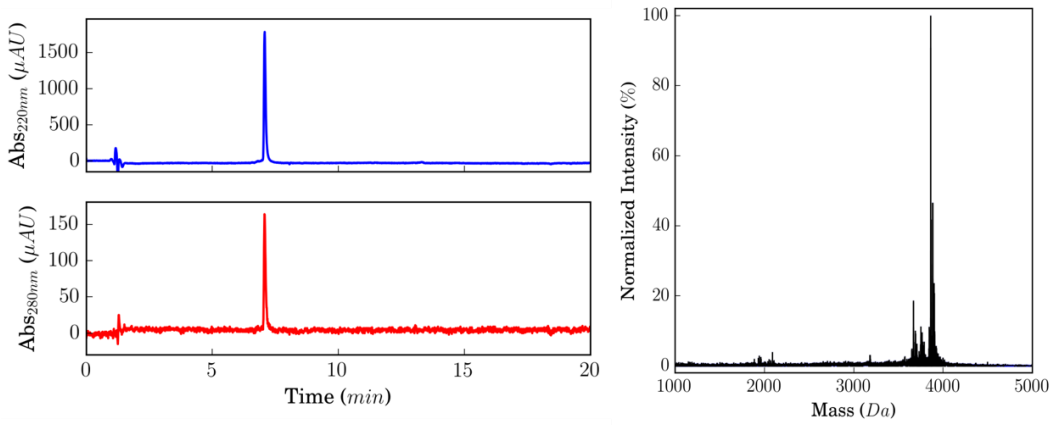


Figure 8-11: oPPα-L@a-A@g - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 3857.015 Da, observed mass = 3863.115 Da.

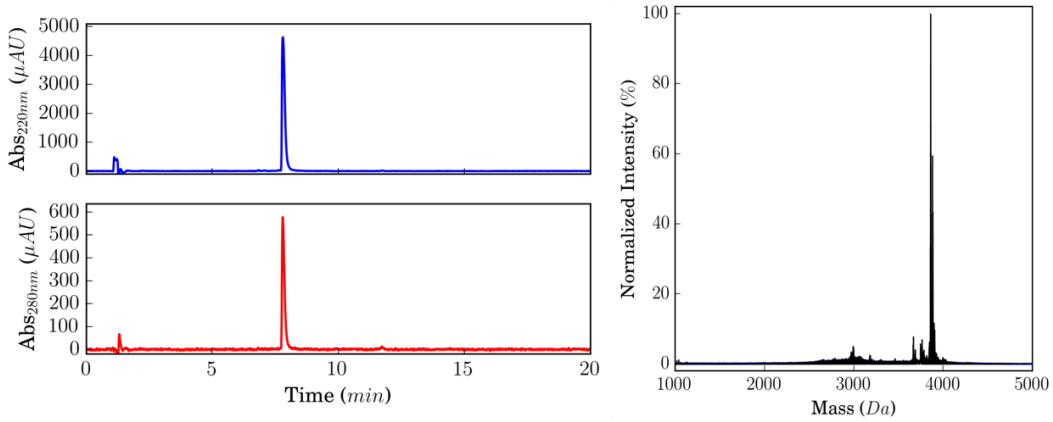


Figure 8-12: oPPα-A@a-L@g - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 3857.015 Da, observed mass = 3860.997 Da.

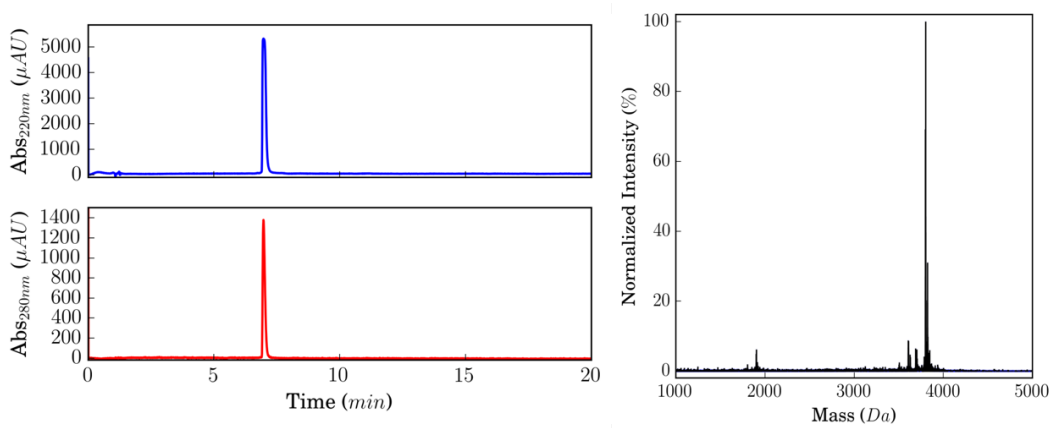


Figure 8-13: oPPα-a↔g - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 3799.099 Da, observed mass = 3800.896 Da.

8.1.3 Chapter 5: Towards larger oPPα topologies

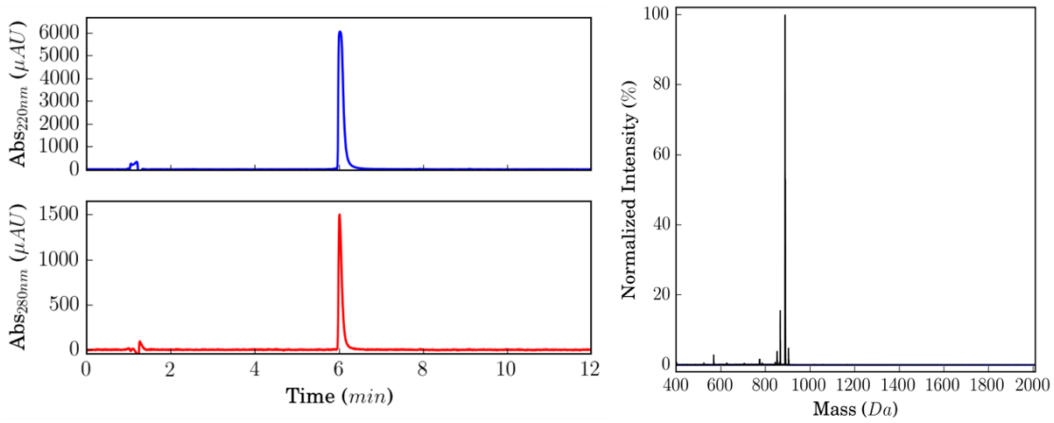


Figure 8-14: PPII - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 864.460 Da, observed mass = 865.969 Da.

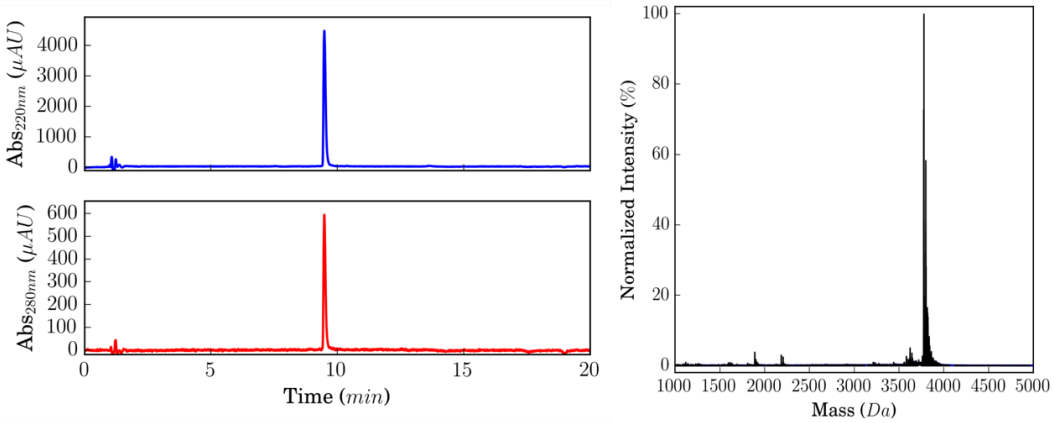


Figure 8-15: oPPα-GSGSG - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 3777.265 Da, observed mass = 3777.630 Da.

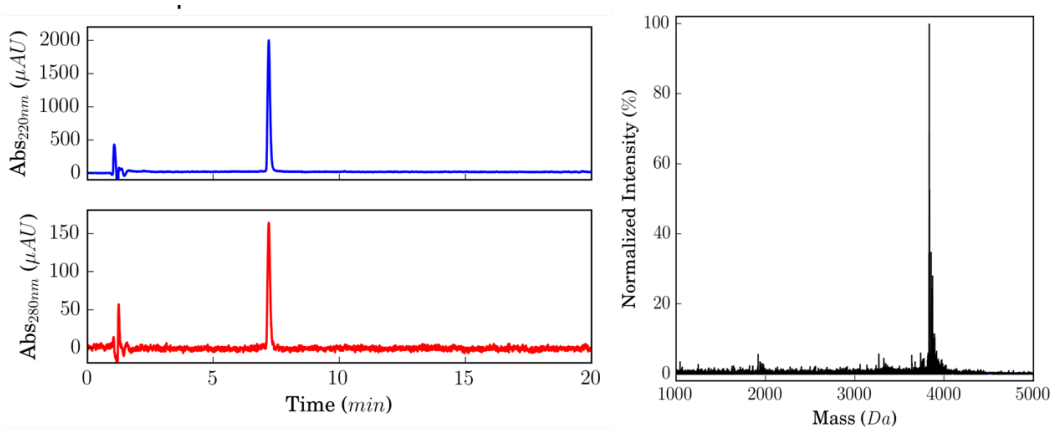


Figure 8-16: oPPα-GDNAT - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 3888.020 Da, observed mass = 3889.910 Da.

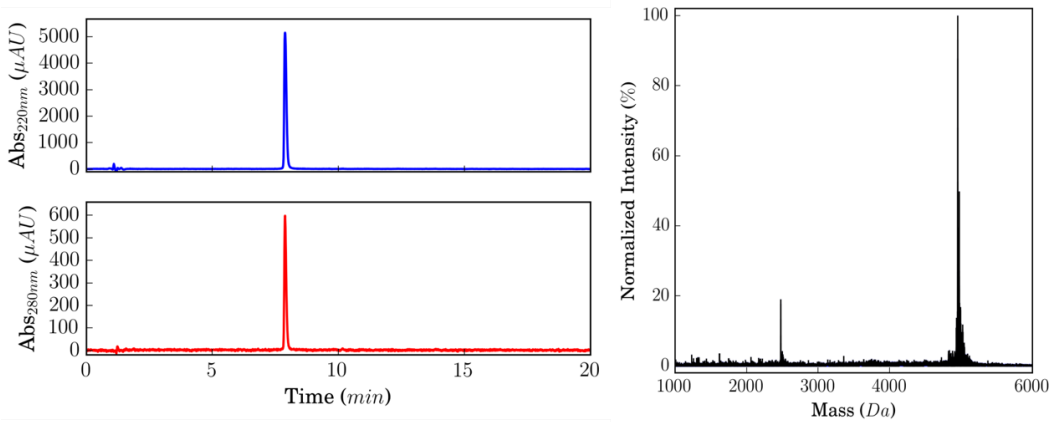


Figure 8-17: oPPα-GDNAT-4 - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 4951.624 Da, observed mass = 4954.896 Da. $[M+H]^{2+}$ also visible.

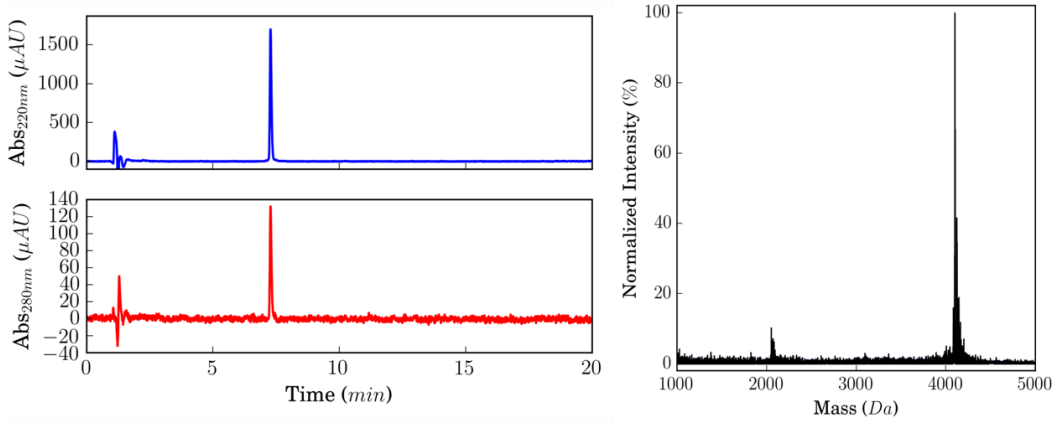


Figure 8-18: oPPα-NP - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 4101.601 Da, observed mass = 4103.111 Da. $[M+H]^{2+}$ also visible.

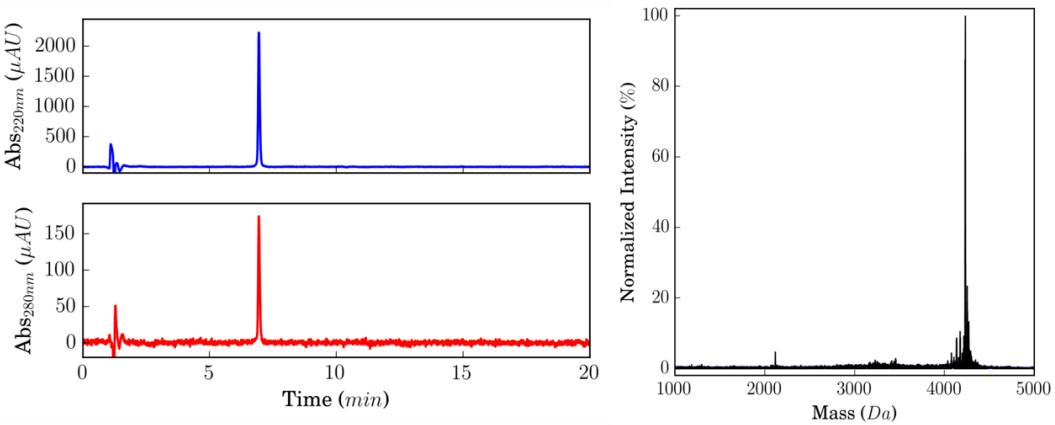


Figure 8-19: oPPα-NPE - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 4230.716 Da, observed mass = 4231.905 Da.

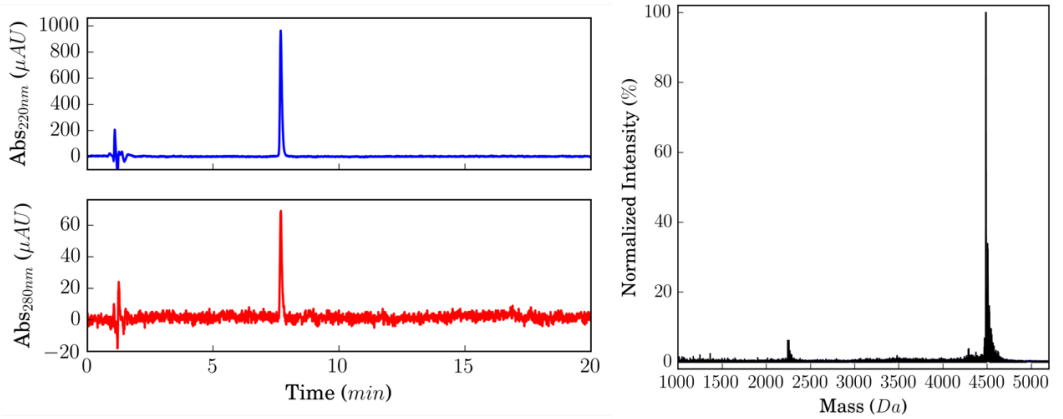


Figure 8-20: oPPα-NPELAA - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 4486.033 Da, observed mass = 4486.418 Da.

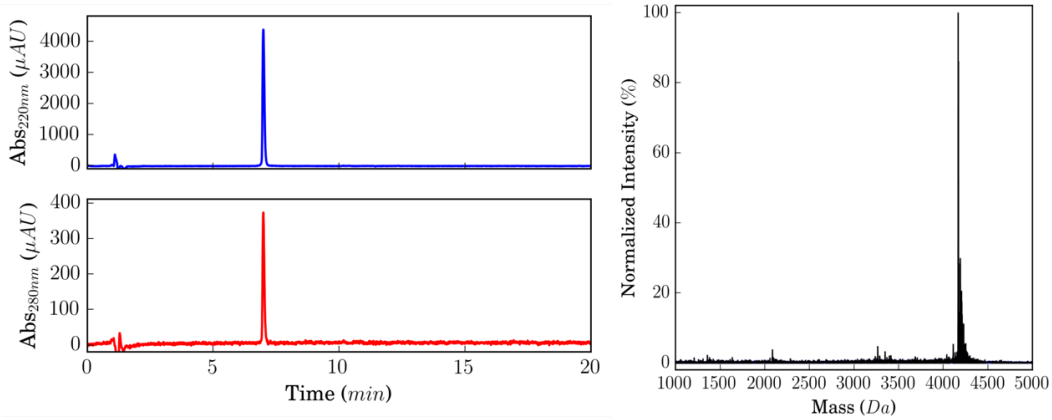


Figure 8-21: oPPα-NPE-4 - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 4168.691 Da, observed mass = 4168.664 Da.

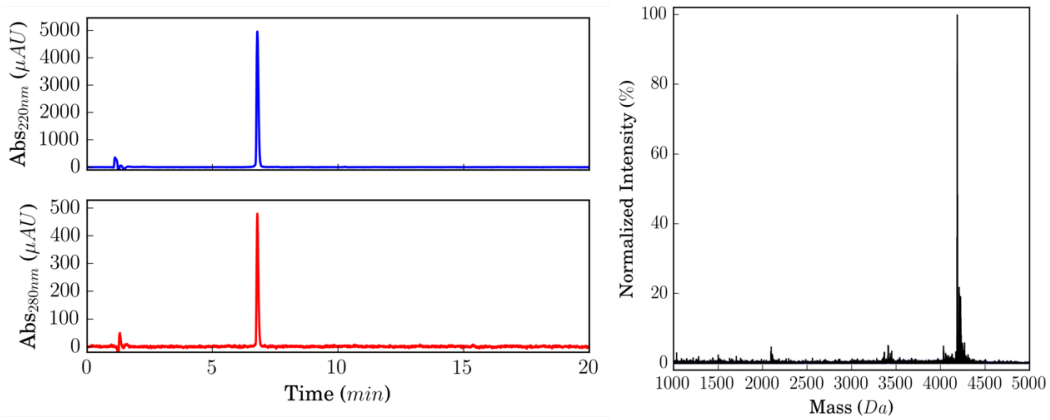


Figure 8-22: oPPα-NPE-5 - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 4185.678 Da, observed mass = 4187.968 Da.

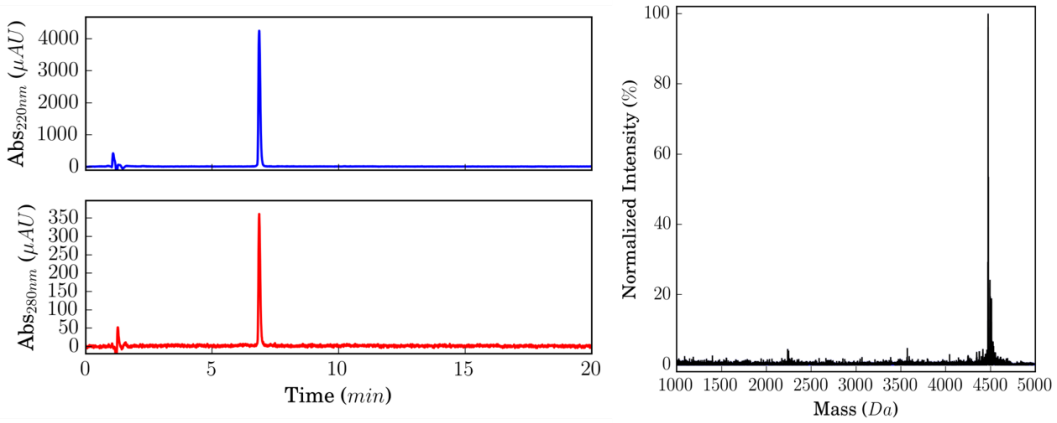


Figure 8-23: oPPα-NPE-6 - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 4471.021 Da, observed mass = 4473.628 Da.

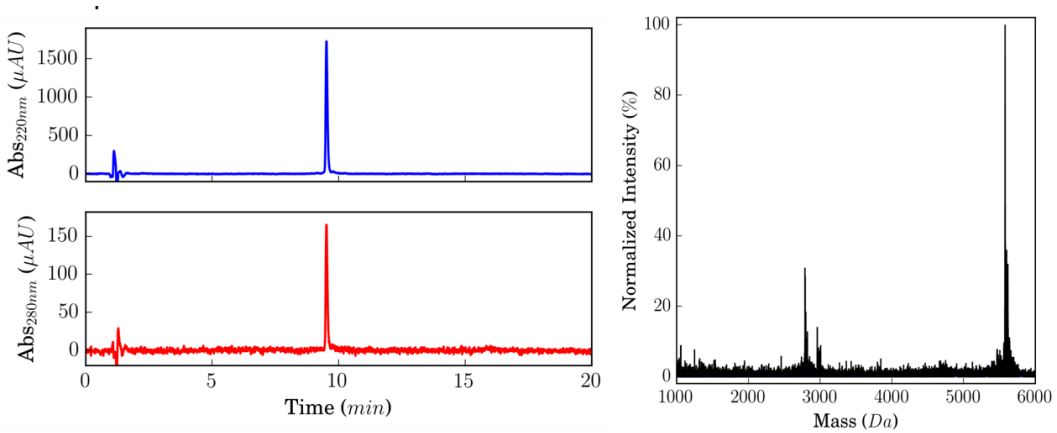


Figure 8-24: PPαPP - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 5575.464 Da, observed mass = 5579.034 Da. $[M+H]^{2+}$ also visible.

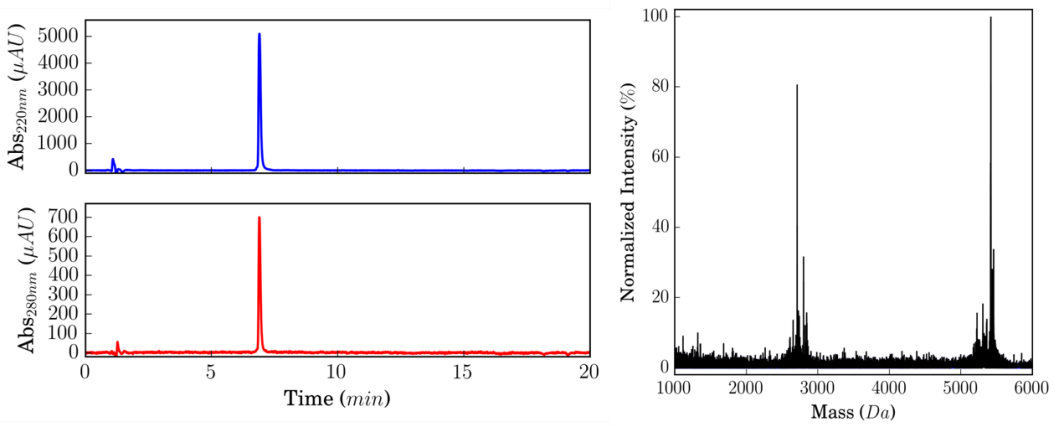


Figure 8-25: PPαPG - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 5415.205 Da, observed mass = 5418.360 Da. $[M+H]^{2+}$ also visible.

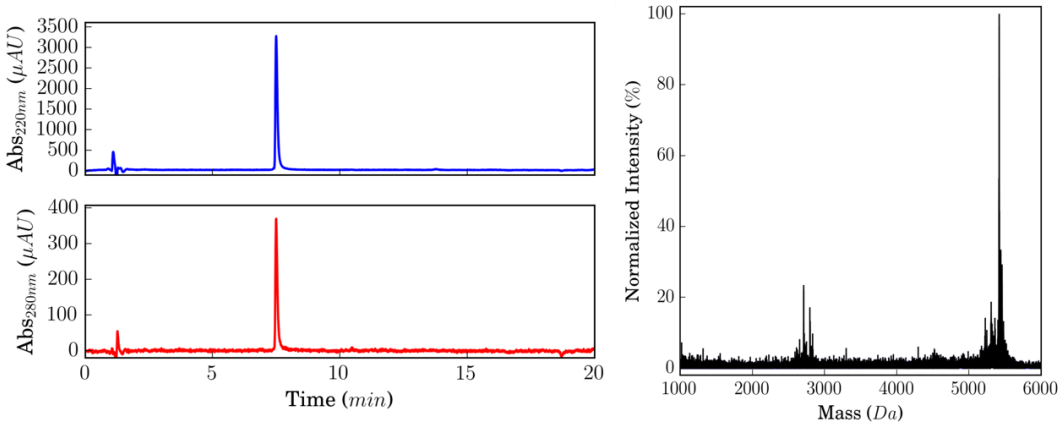


Figure 8-26: PGαPP - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 5415.205 Da, observed mass = 5418.926 Da. $[M+H]^{2+}$ also visible.

8.1.4 Chapter 6: Modulating stability in oPPα through elongation and cyclisation

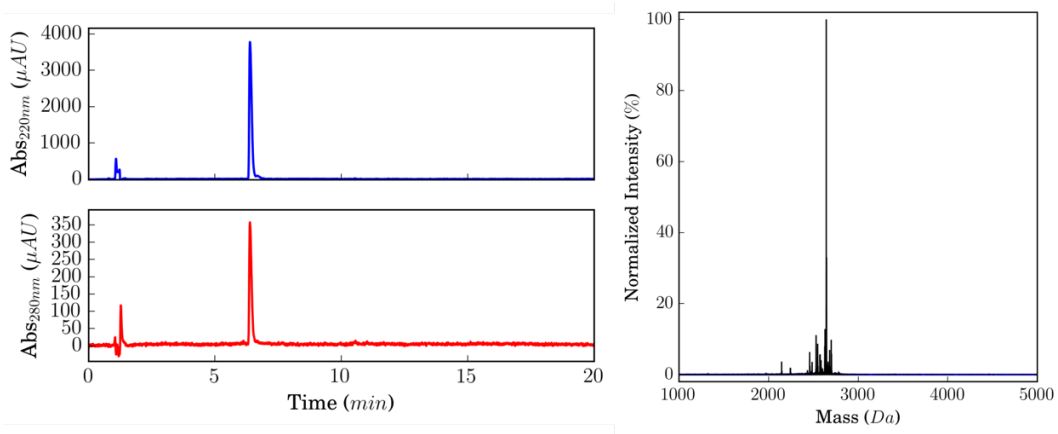


Figure 8-27: oPPα-2 - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 4863.574 Da, observed mass = 4863.025 Da.

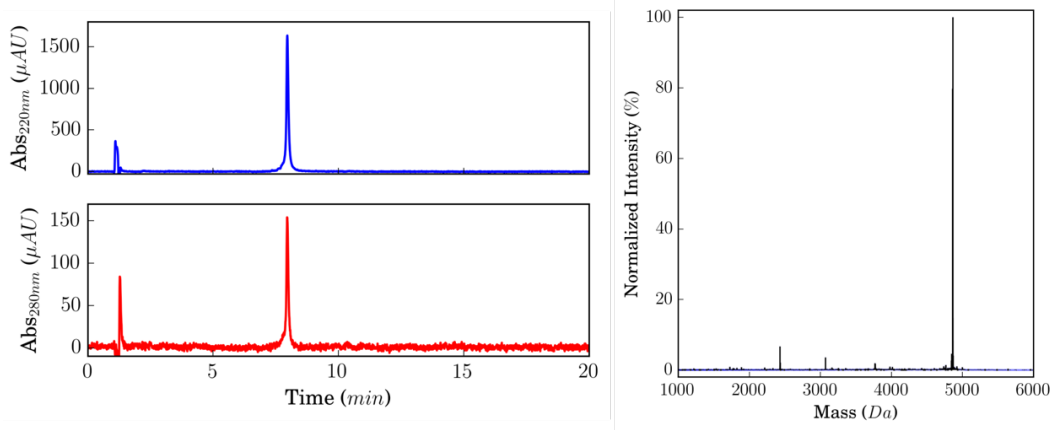


Figure 8-28: oPPα-4-nocap - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 4863.574 Da, observed mass = 4863.025 Da.

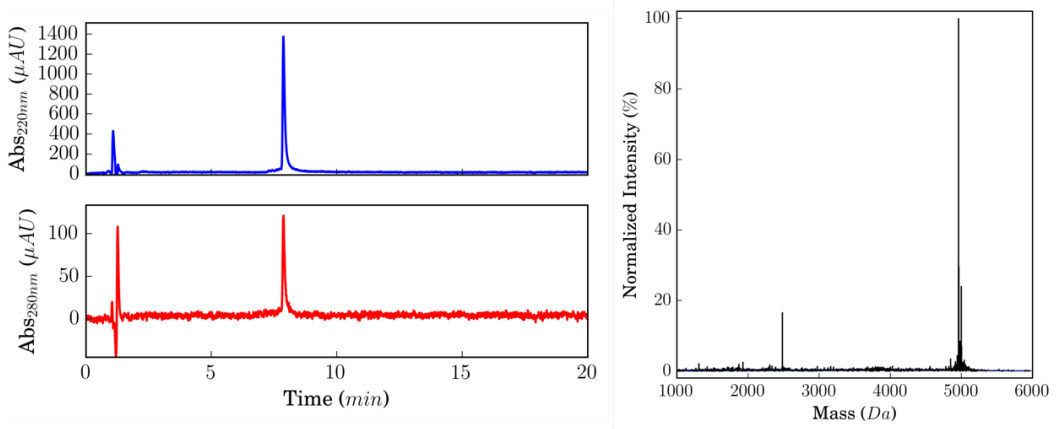


Figure 8-29: oPPα-4 - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 4960.690 Da, observed mass = 4961.102 Da. $[M+H]^{2+}$ also visible.

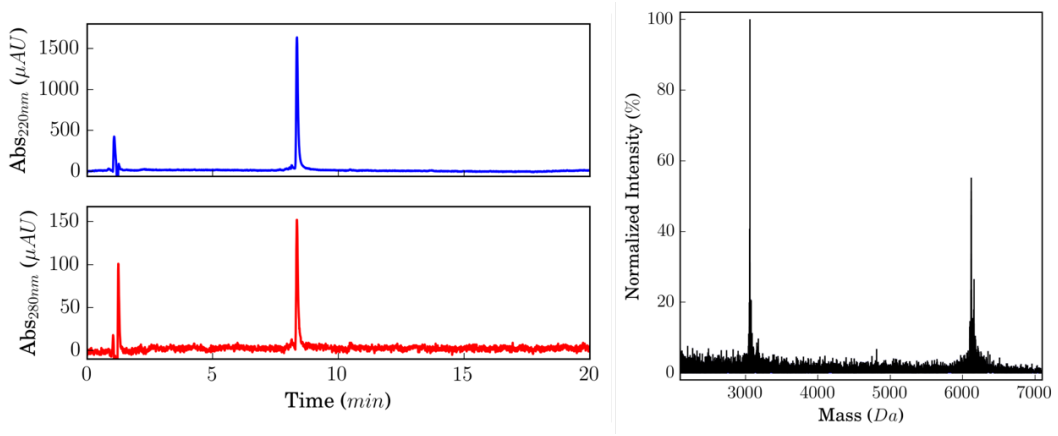


Figure 8-30: oPPα-5 - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 6119.050 Da, observed mass = 6119.708 Da. Observed mass $[M+H]^{2+}$ = 3059.692.

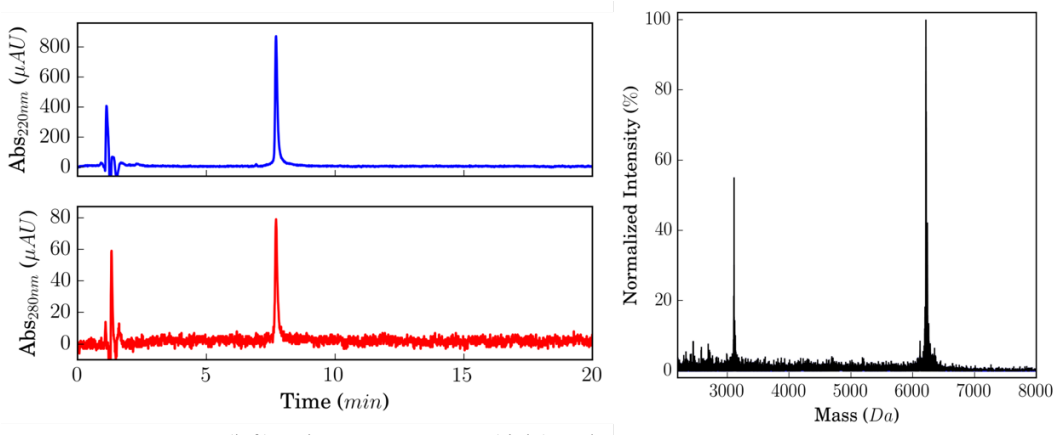


Figure 8-31: oPPα-5-skip - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 6211.45 Da, observed mass = 6215.346 Da. $[M+H]^{2+}$ also visible.

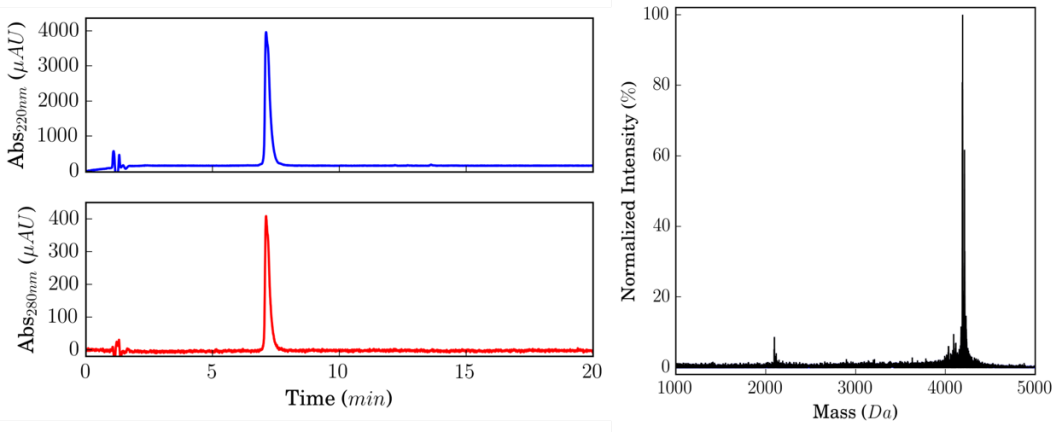


Figure 8-32: cPPα-GPQY - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 4185.18 Da, observed mass = 4190.124 Da. $[M+H]^{2+}$ also visible.

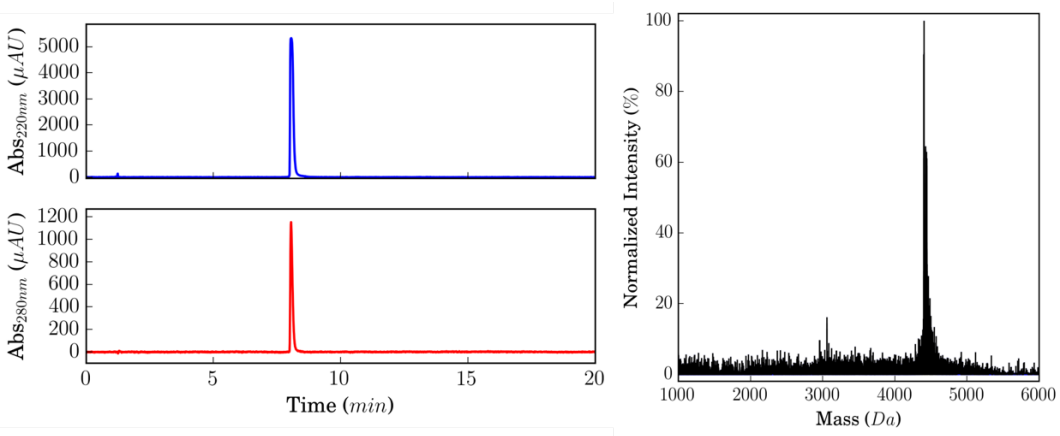


Figure 8-33: cPPα-LSTKDL - HPLC traces (left) and MALDI-TOF MS (right). Calculated mass $[M+H]^+$ = 4397.35 Da, observed mass = 4404.301 Da.

8.2 Circular dichroism data for designed peptide sequences

8.2.1 Stabilising the PPα miniprotein by rational design

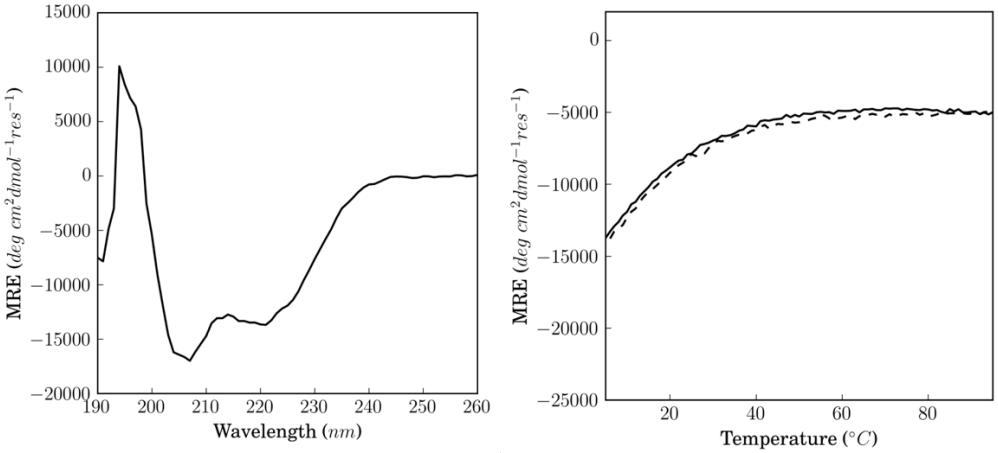


Figure 8-34: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for o-α. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

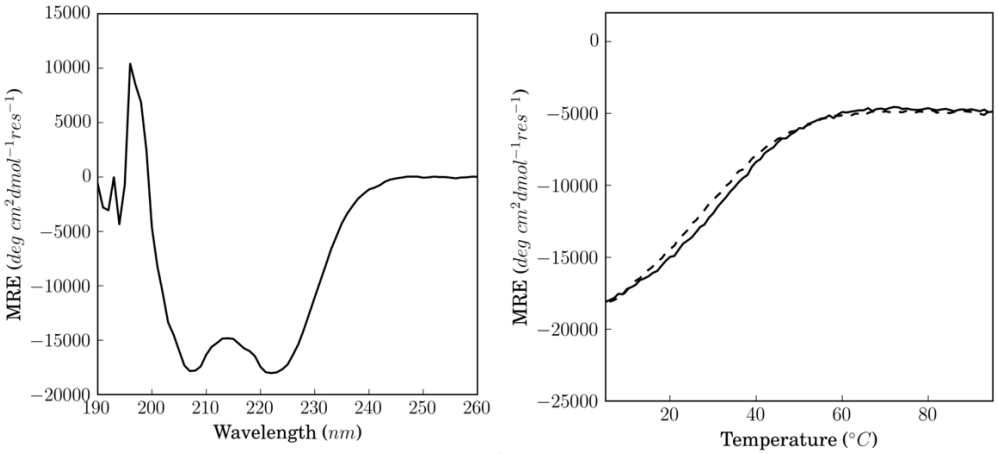


Figure 8-35: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for oPPα. Conditions: 100 μM peptide, PBS , pH 7.4. Key: melt (black), cool (dashed line).

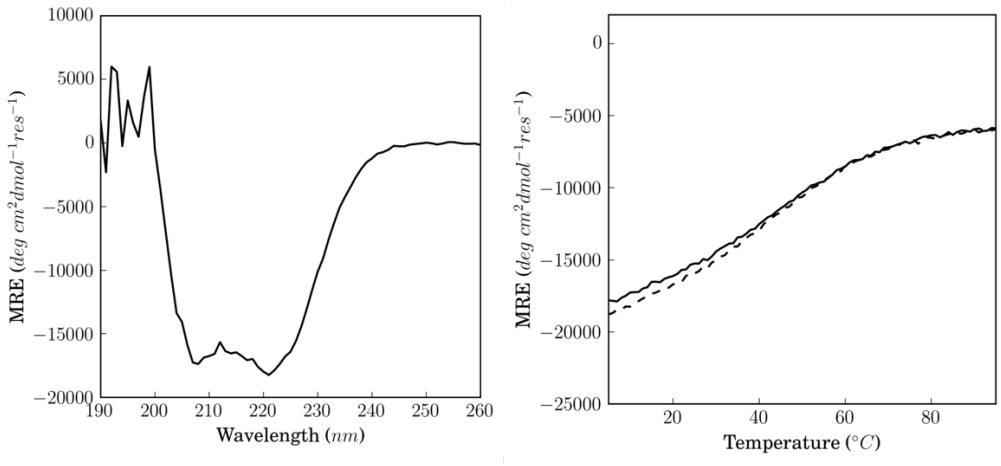


Figure 8-36: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for oPPα-E↔K. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

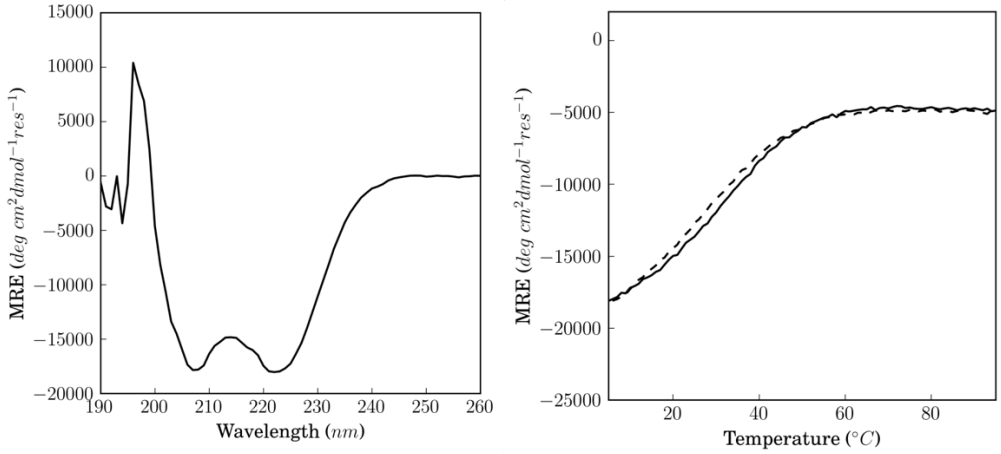


Figure 8-37: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for oPPα-Phe. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

8.2.2 Probing Sequence-to-Stability Relationships in oPPα

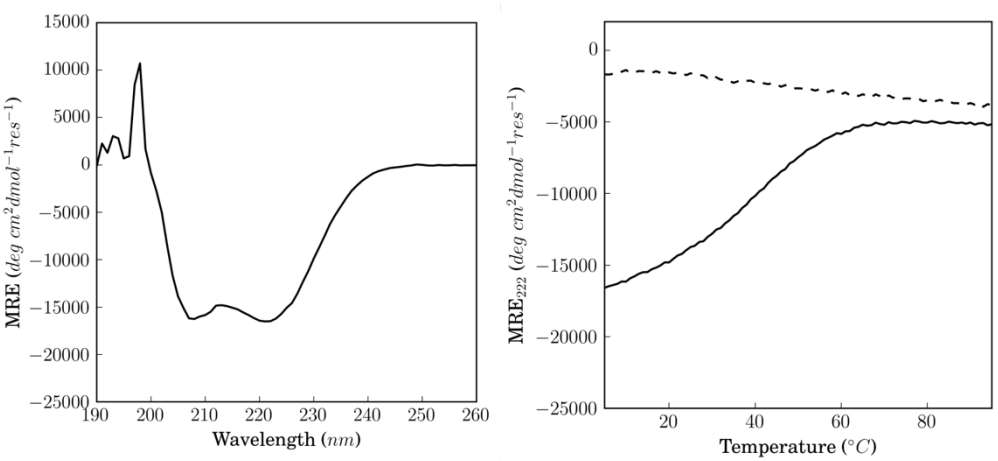


Figure 8-38: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for oPPα-I@a. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

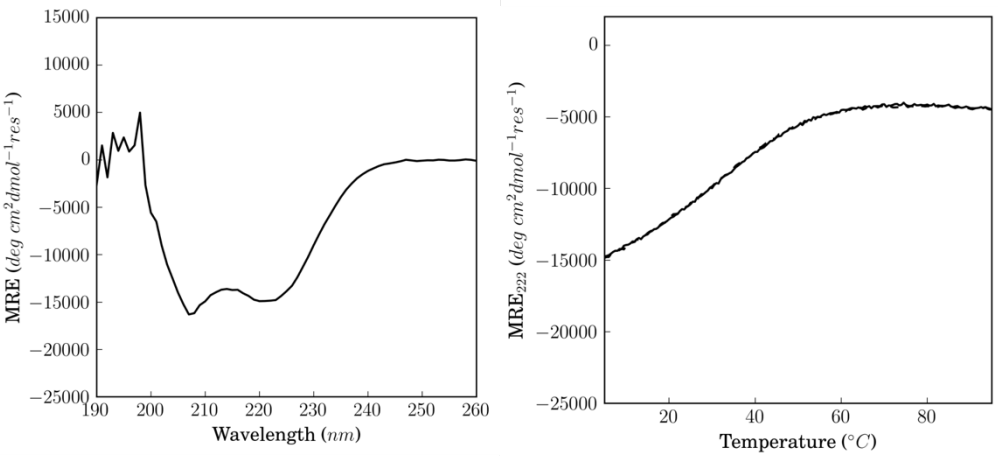


Figure 8-39: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for oPPα-V@a. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

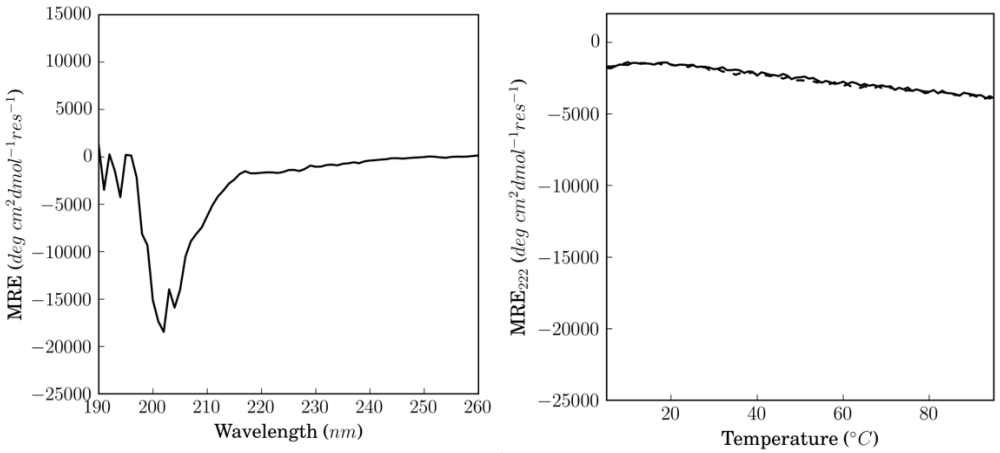


Figure 8-40: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for oPPα-E@a. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

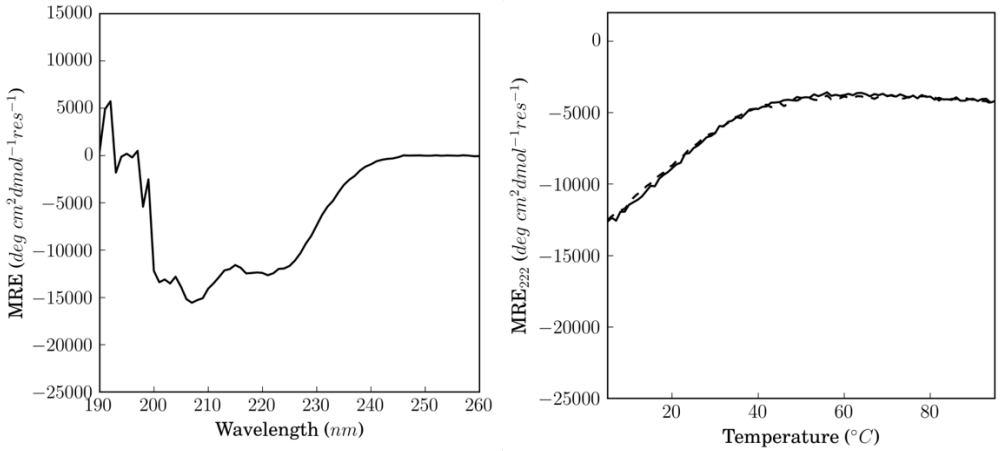


Figure 8-41: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for oPPα-K@a. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

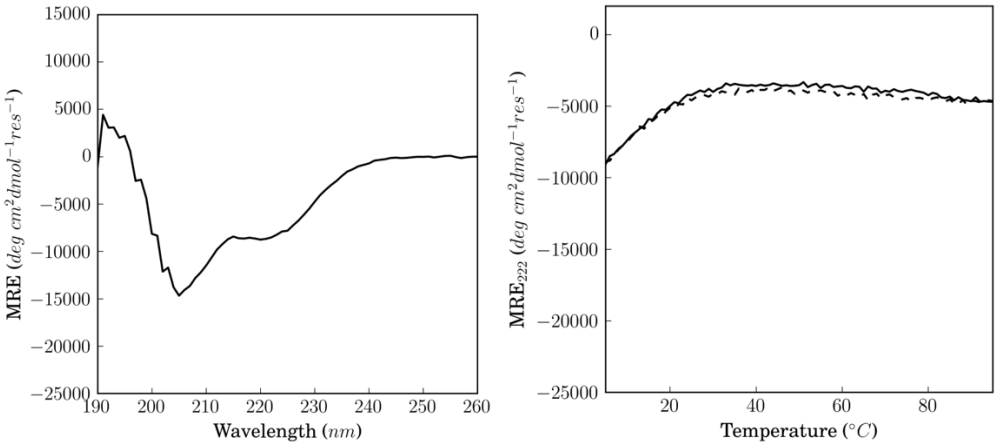


Figure 8-42: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for oPPα-A@a. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

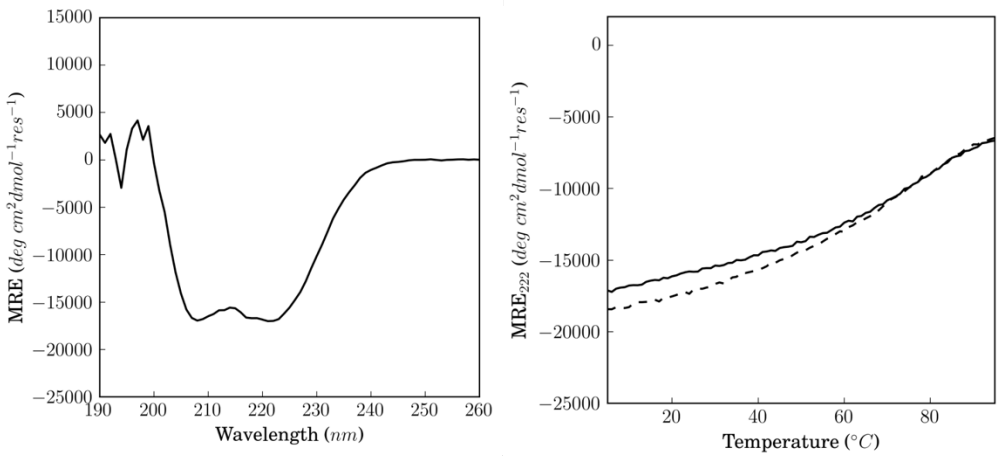


Figure 8-43: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for oPPα-L@a-L@g. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

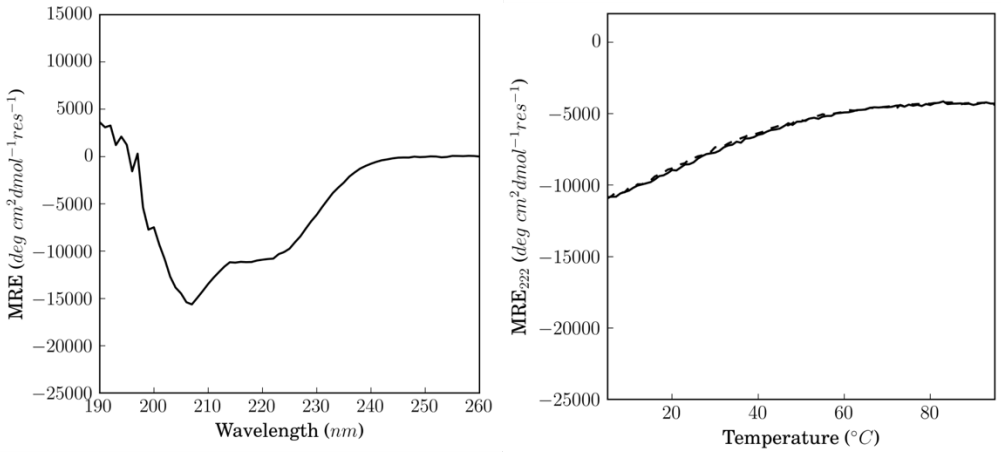


Figure 8-44: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for oPPα-L@a-A@g. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

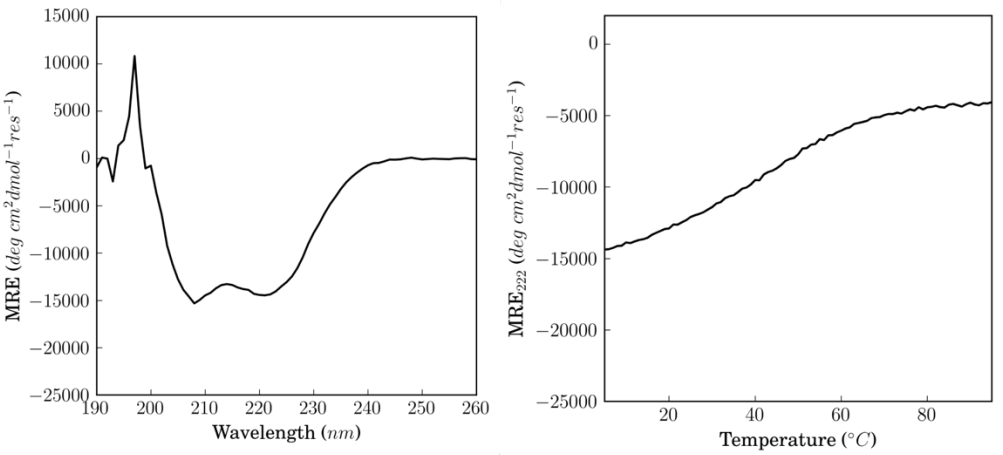


Figure 8-45: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for oPPα-A@a-L@g. Conditions: 100 μM peptide, PBS, pH 7.4.

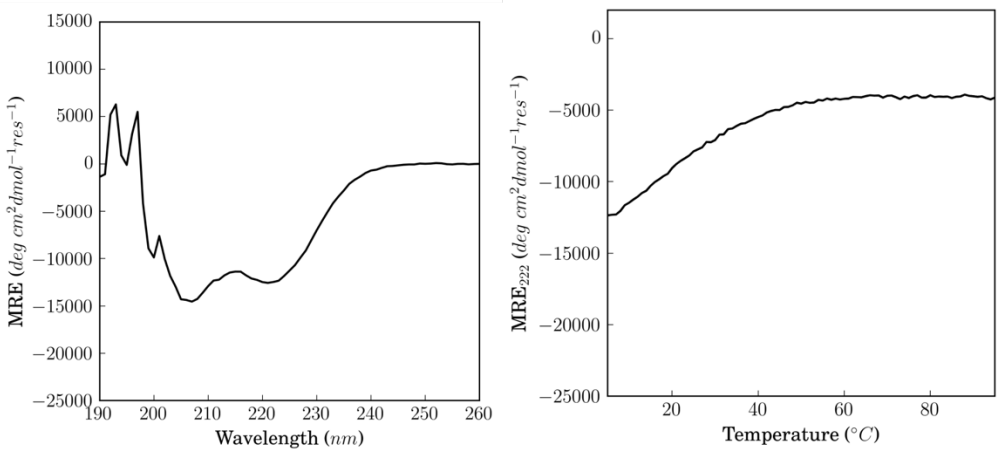


Figure 8-46: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for oPPα-a↔g. Conditions: 100 µM peptide, PBS, pH 7.4.

8.2.3 Towards larger oPPα topologies

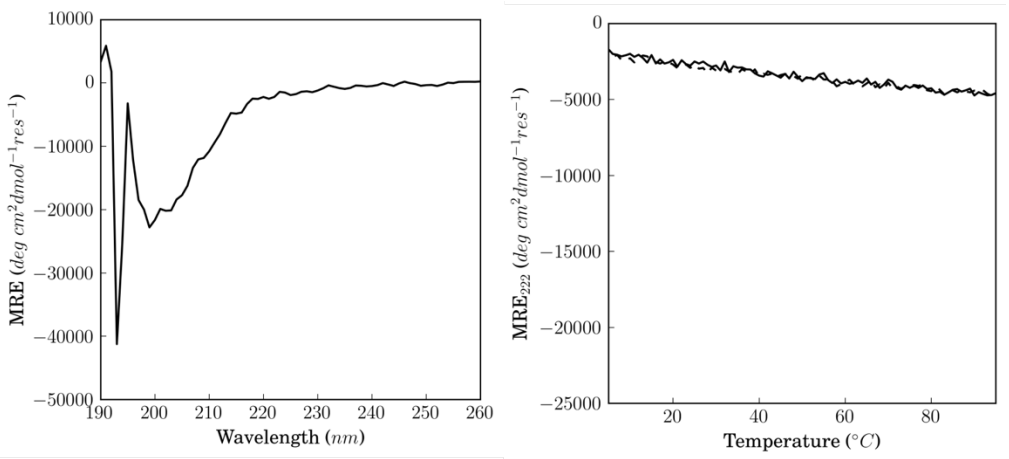


Figure 8-47: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for PPII. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

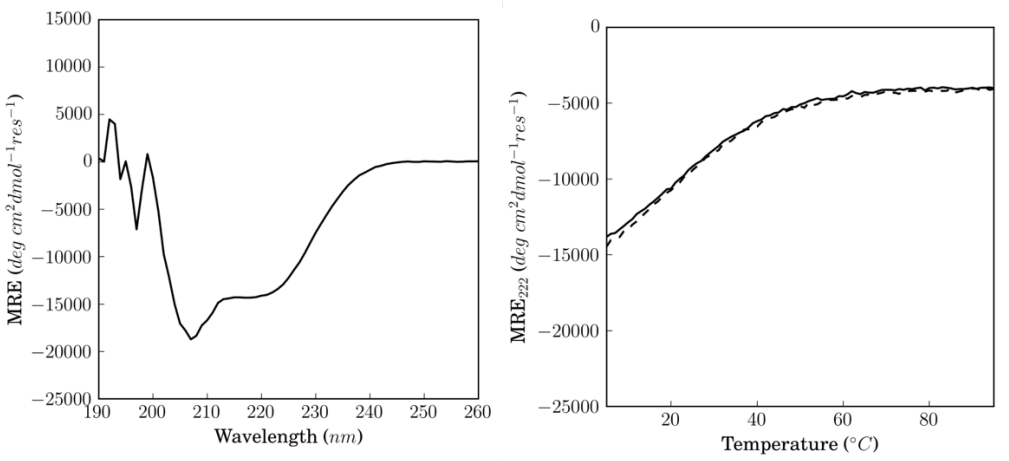


Figure 8-48: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for αPP-GSGSG. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

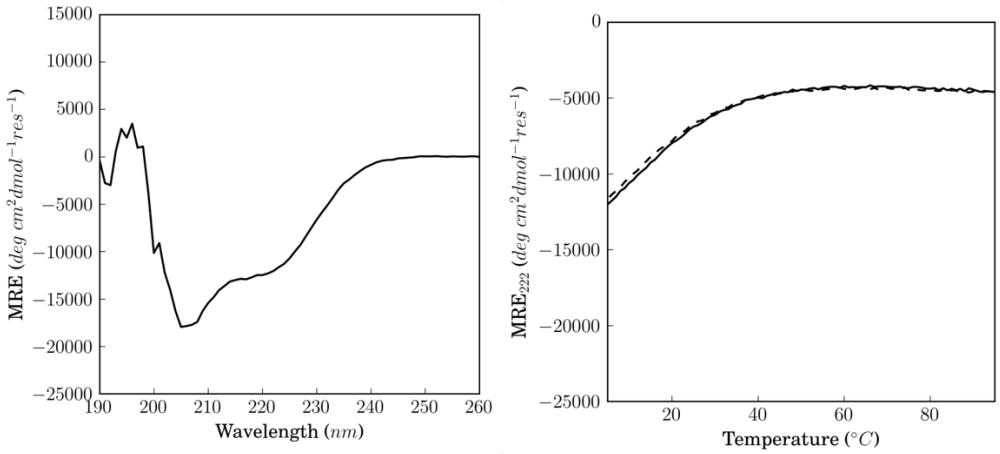


Figure 8-49: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for αPP-GDNAT. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

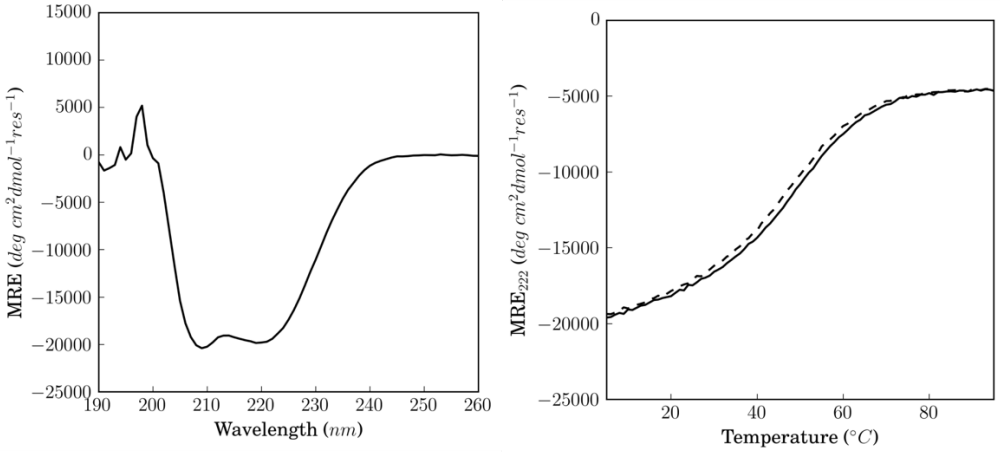


Figure 8-50: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for αPP-GDNAT-4. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

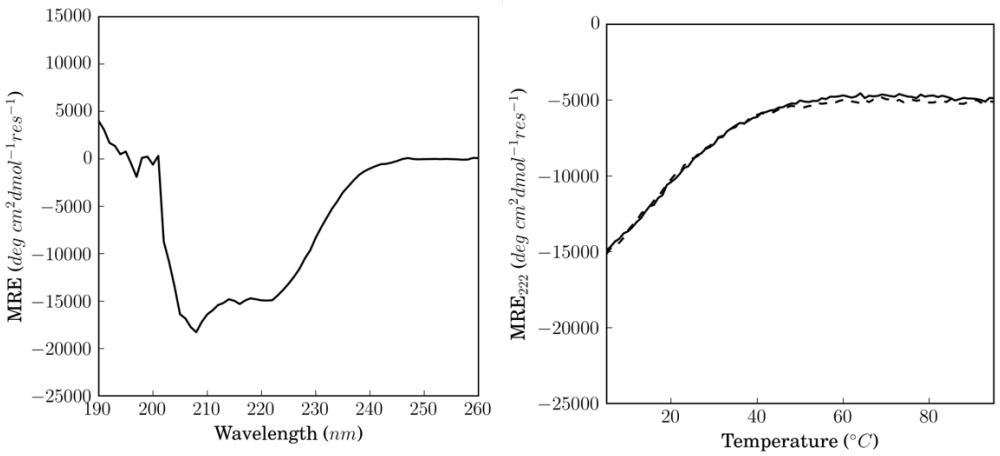


Figure 8-51: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for αPP-GDNAT-NP. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

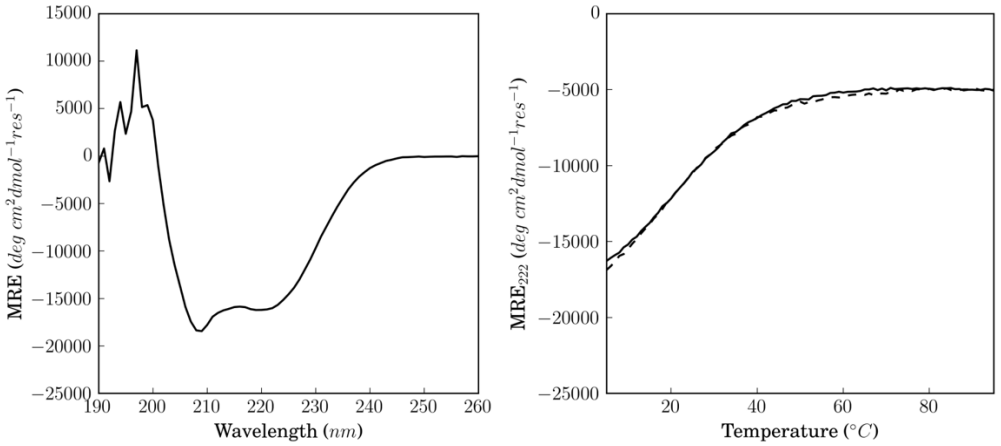


Figure 8-52: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for αPP-GDNAT-NPE. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

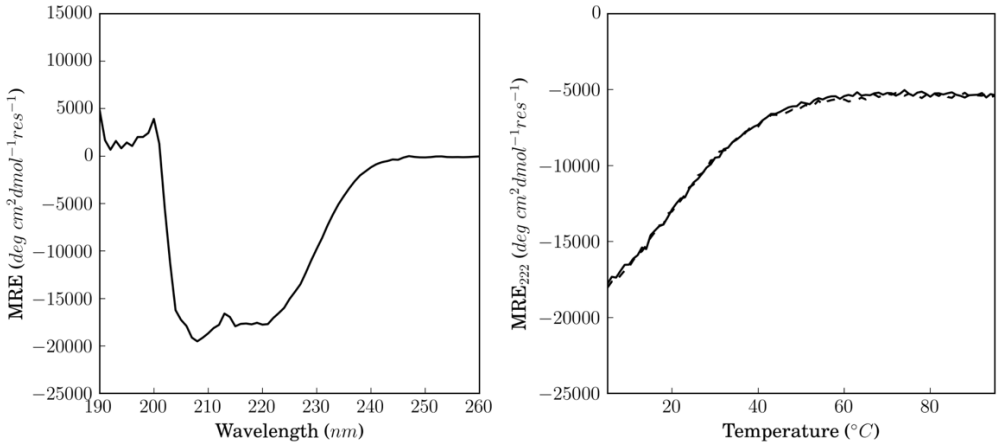


Figure 8-53: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for αPP-GDNAT-NPELAA. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

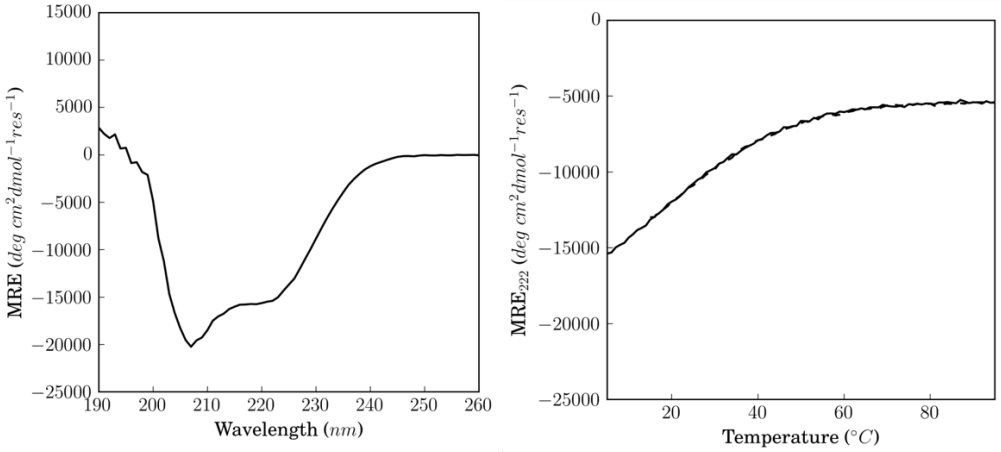


Figure 8-54: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for αPP-NPE-4. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

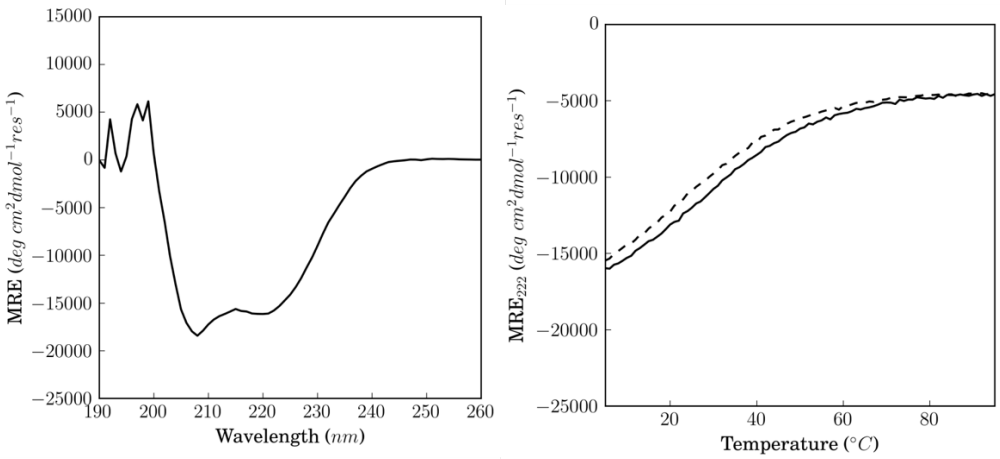


Figure 8-55: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for αPP-NPE-5. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

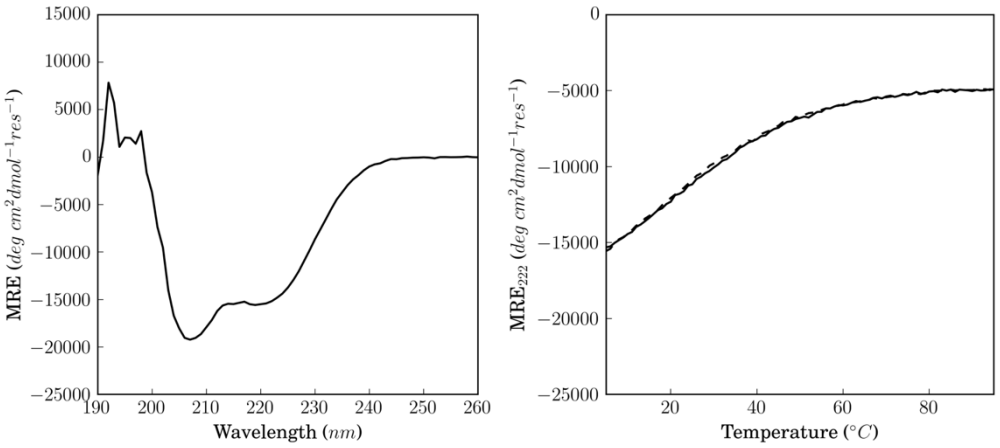


Figure 8-56: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for αPP-NPE-6. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

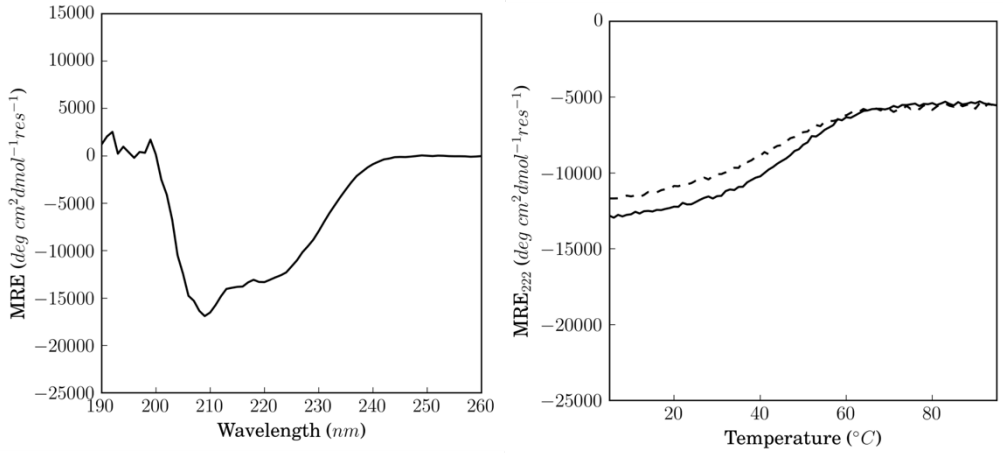


Figure 8-57: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for PPαPP. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

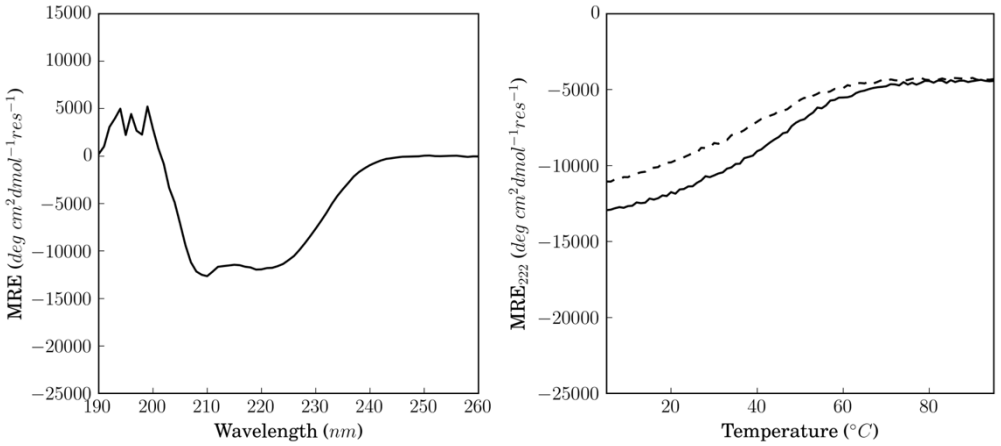


Figure 8-58: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for PPαPG. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

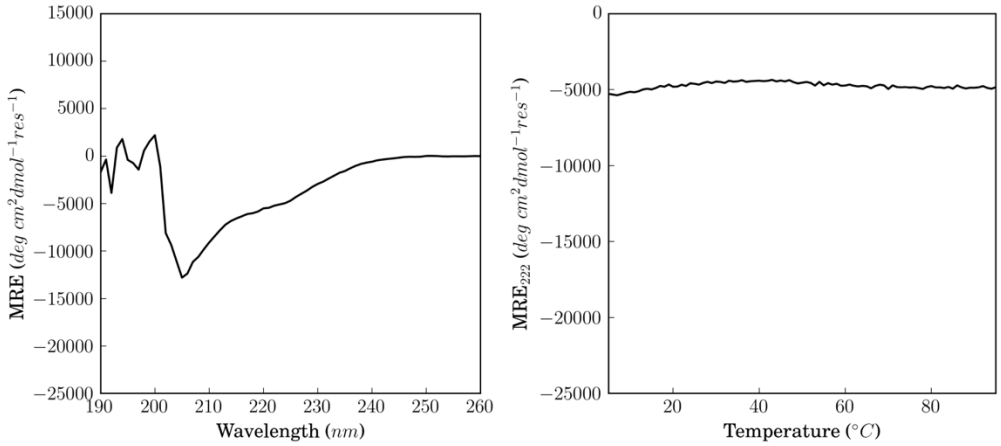


Figure 8-59: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for PGαPP. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

8.2.3 Modulating stability in oPPα through elongation and cyclisation

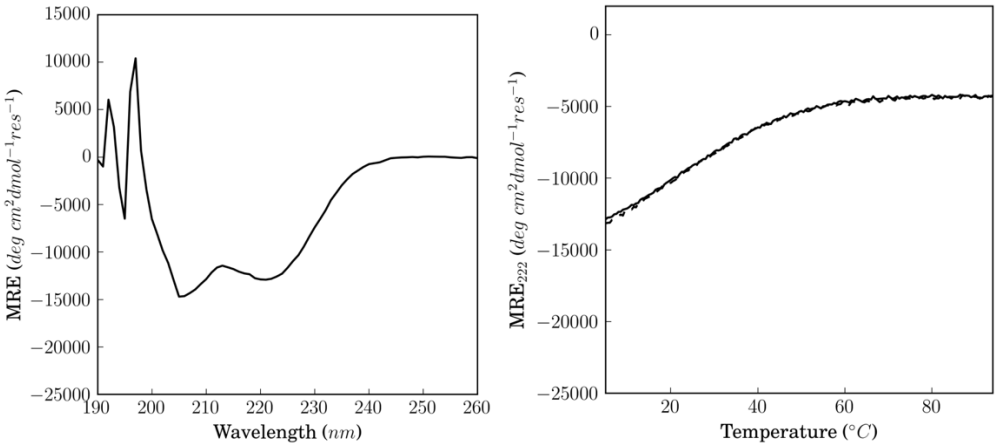


Figure 8-60: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for oPPα-2. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

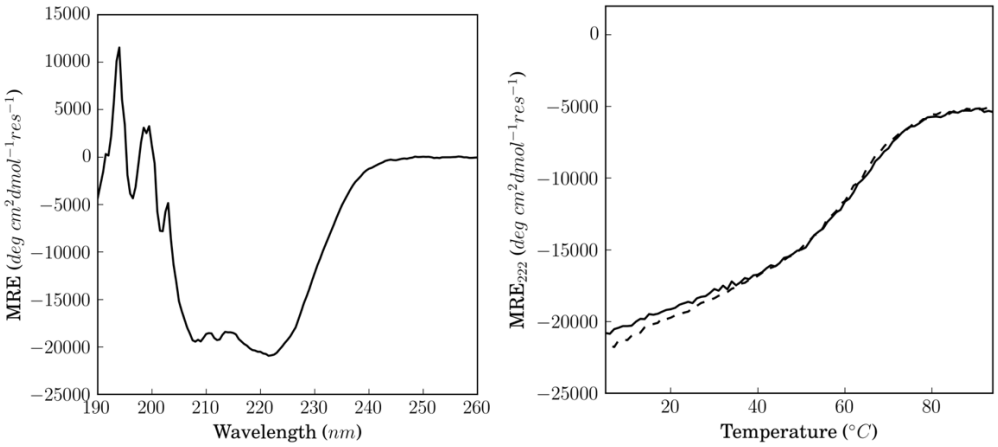


Figure 8-61: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for oPPα-4-nocap. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

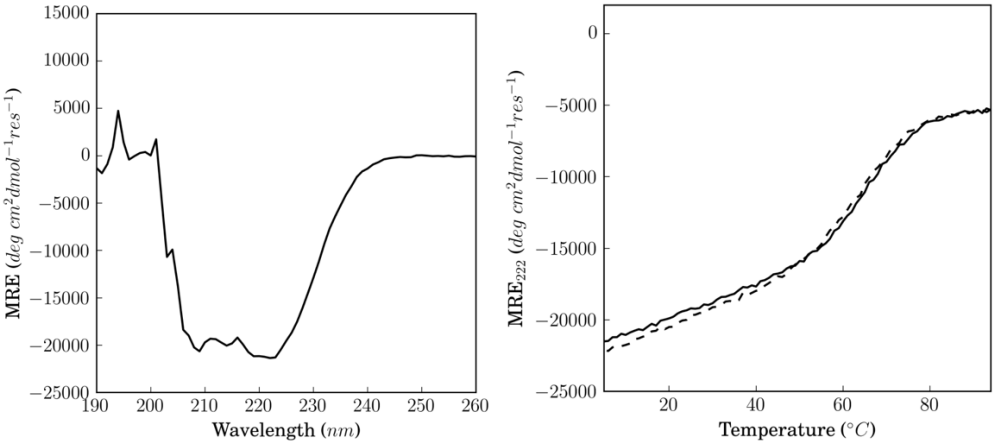


Figure 8-62: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for oPPα-4. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

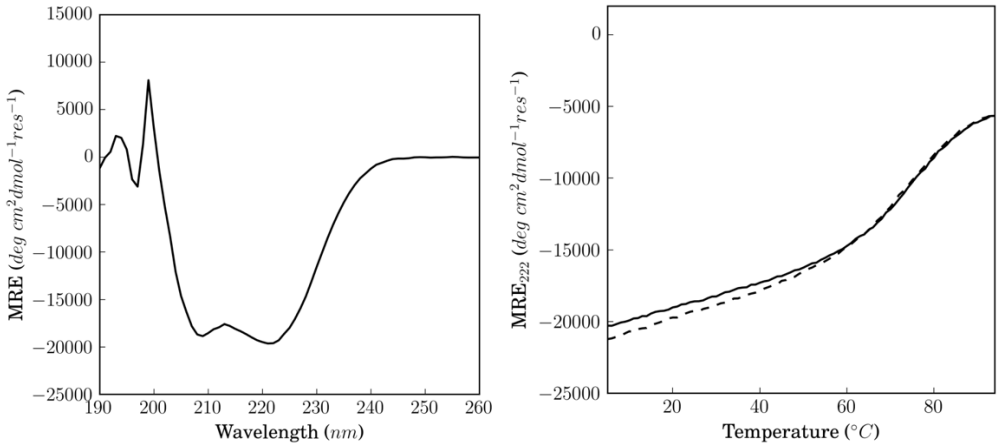


Figure 8-63: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for oPPα-5. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

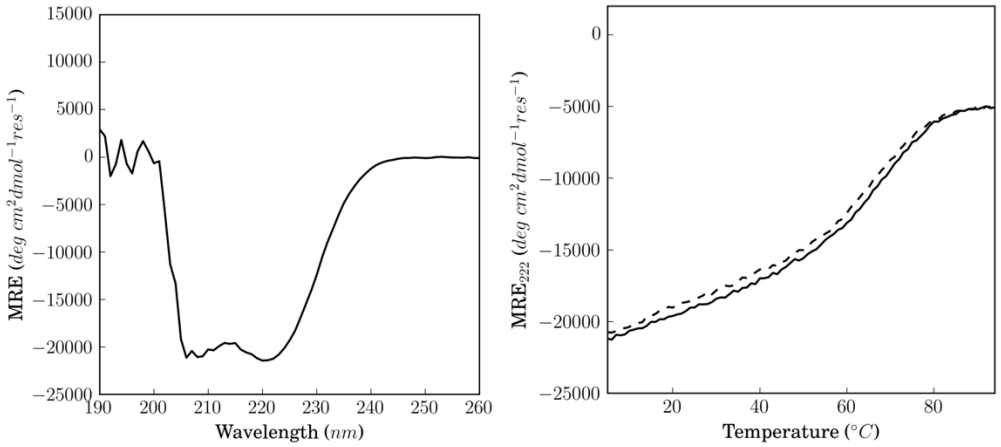


Figure 8-64: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for oPPα-5-skip. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

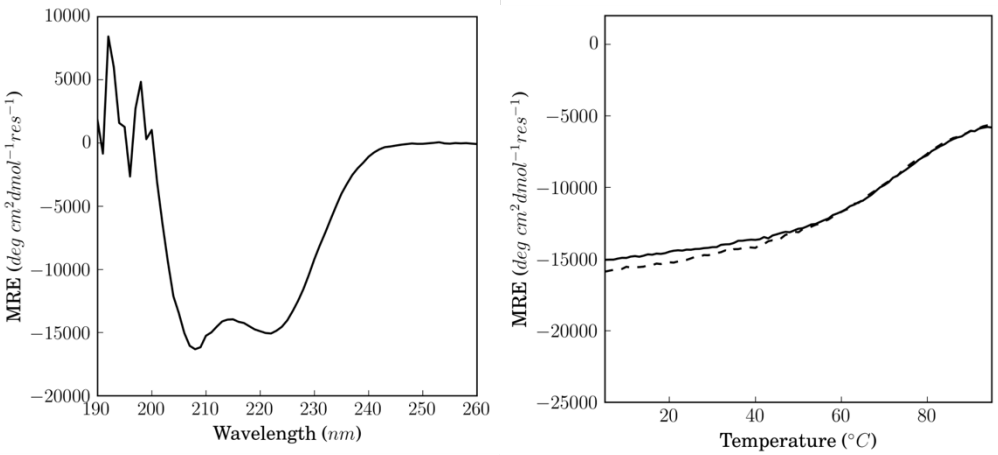


Figure 8-65: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for cPPα-LSTKDL. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

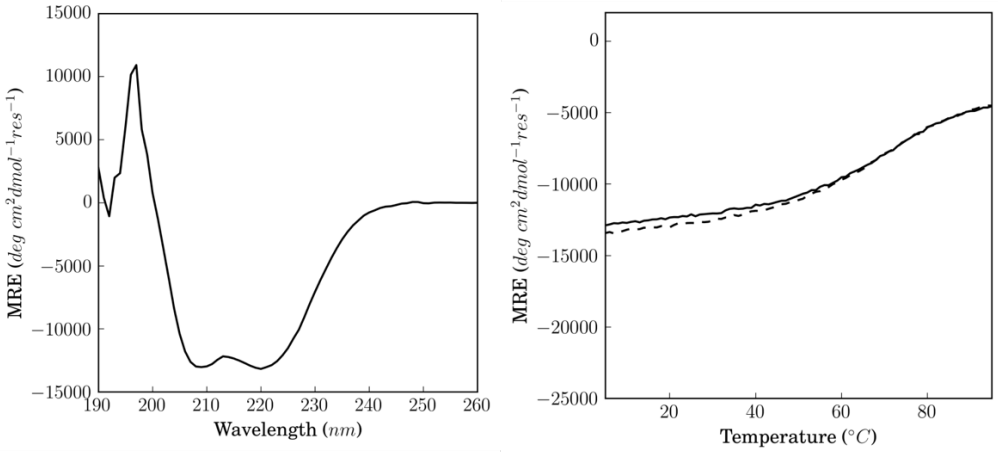


Figure 8-66: CD spectrum at 5°C (left) and temperature dependent CD signal, monitored at 222 nm (right) for cPPα-GPQY. Conditions: 100 μM peptide, PBS, pH 7.4. Key: melt (black), cool (dashed line).

8.3 Analytical ultracentrifugation data for designed peptide sequences

8.3.1 Stabilising the PPα miniprotein by rational design

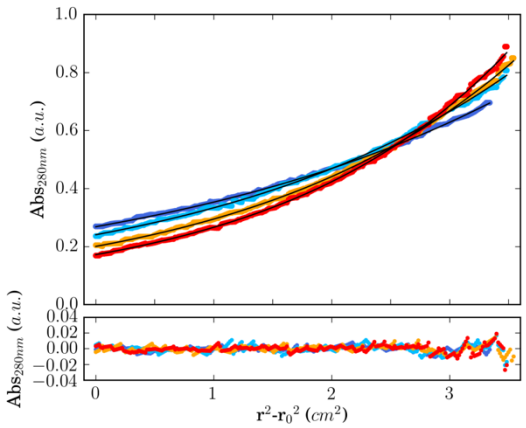


Figure 8-67: AUC data for o-α ($\bar{v}$ = 0.747 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 2928 Da (1.2 × monomer mass, 95 % confidence limits 2880 – 2917 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

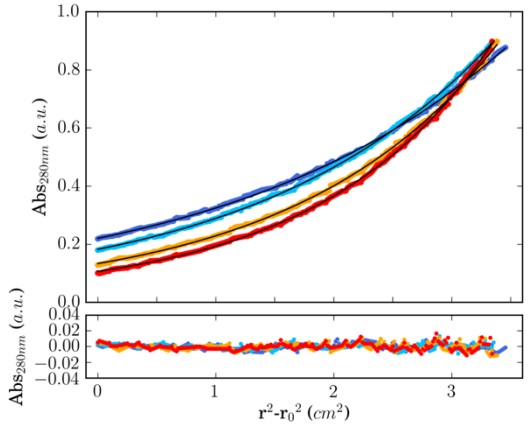


Figure 8-68: AUC data for oPPα ($\bar{v}$ = 0.748 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 3854 Da (1.0 × monomer mass, 95 % confidence limits 3839 – 3869 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

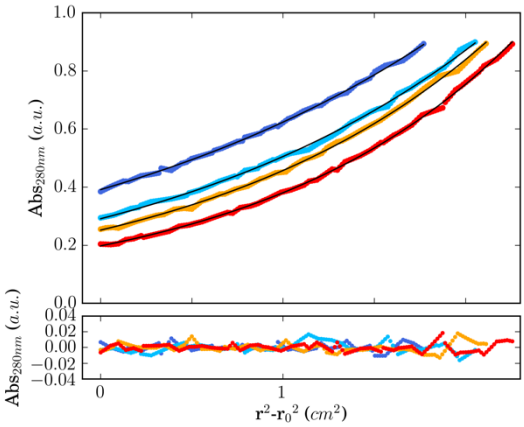


Figure 8-69: AUC data for oPPα-E↔K ($\bar{v}$ = 0.748 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 35454 Da (0.9 × monomer mass, 95 % confidence limits 3511 – 3581Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

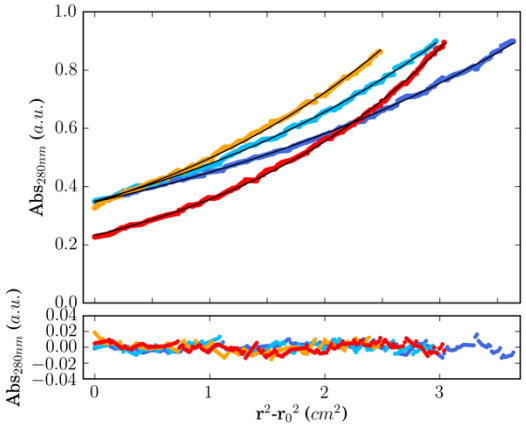


Figure 8-70: AUC data for oPPα-Phe ($\bar{v}$ = 0.756 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 3771 Da (1.0 × monomer mass, 95 % confidence limits 3742 – 3801 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

8.3.2 Probing Sequence-to-Stability Relationships in oPPα

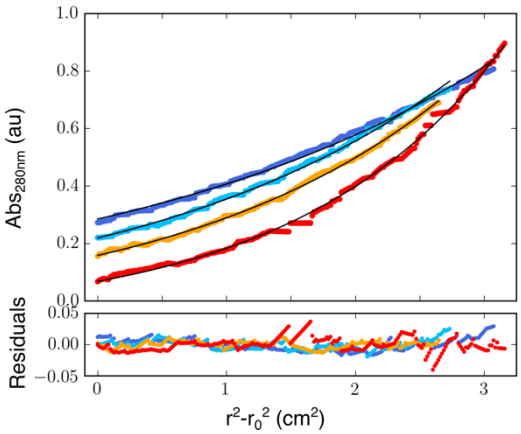


Figure 8-71: AUC data for oPPα-I@a ($\bar{v}$ = 0.748 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 3074 Da (1.0 × monomer mass, 95 % confidence limits 3037 – 3111 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

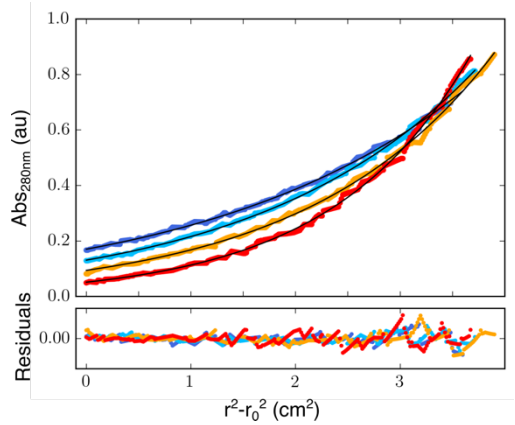


Figure 8-72: AUC data for oPPα-V@a ($\bar{v}$ = 0.743 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 3702 Da (1.0 × monomer mass, 95 % confidence limits 3683 – 3722 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

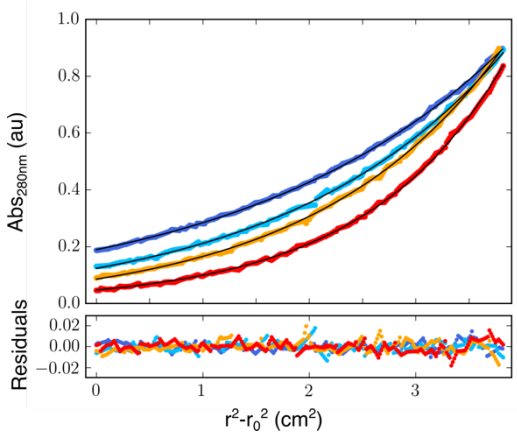


Figure 8-73: AUC data for oPPα-K@a ($\bar{v}$ = 0.742 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 3692 Da (1.0 × monomer mass, 95 % confidence limits 3680 – 3703 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

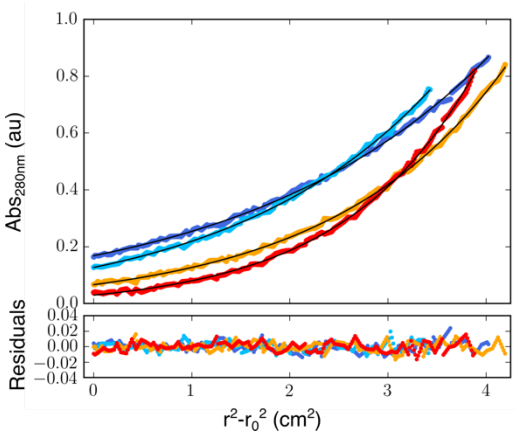


Figure 8-74: AUC data for oPPα-A@a ($\bar{v}$ = 0.733 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 3510 Da (1.0 × monomer mass, 95 % confidence limits 3495 – 3526 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

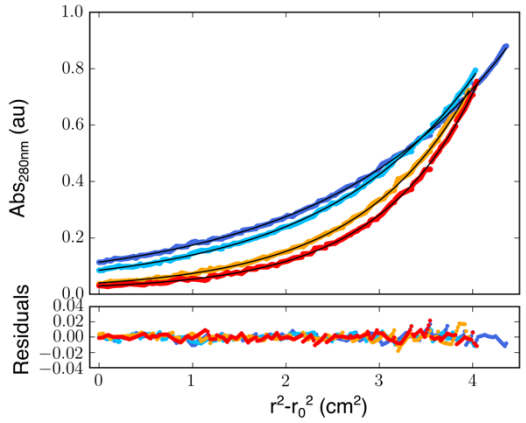


Figure 8-75: AUC data for oPPα-L@a-L@g ($\bar{v}$ = 0.754 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 4912 Da (1.2 × monomer mass, 95 % confidence limits 4899 – 4926 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

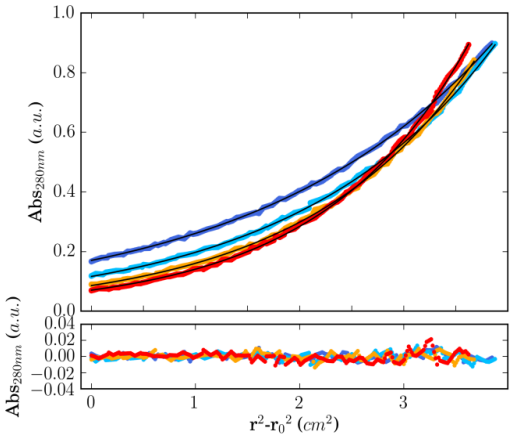


Figure 8-76: AUC data for oPPα-L@a-A@g ($\bar{v}$ = 0.745 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 4019 Da (1.0 × monomer mass, 95 % confidence limits 4006 – 4030 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

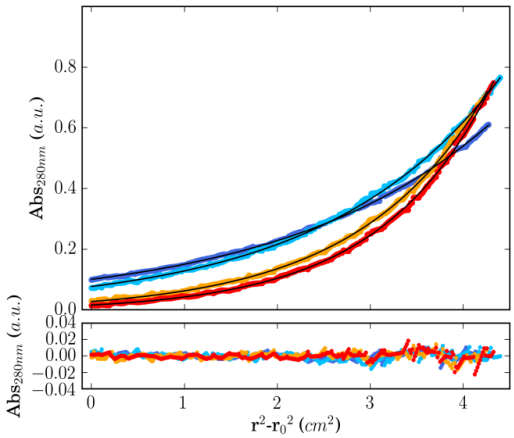


Figure 8-77: AUC data for oPPα-A@a-L@g ($\bar{v}$ = 0.745 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 4101 Da (1.1 × monomer mass, 95 % confidence limits 4089 – 4113 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

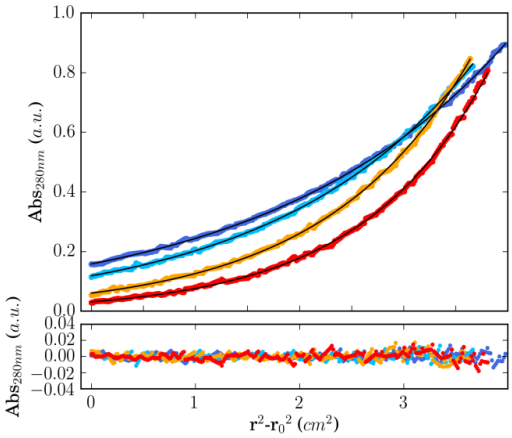


Figure 8-78: AUC data for oPPα-***a↔g*** ($\bar{v}$ = 0.748 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 4079 Da (1.1 × monomer mass, 95 % confidence limits 4065 – 4093 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

8.3.3 Towards larger oPPα assemblies

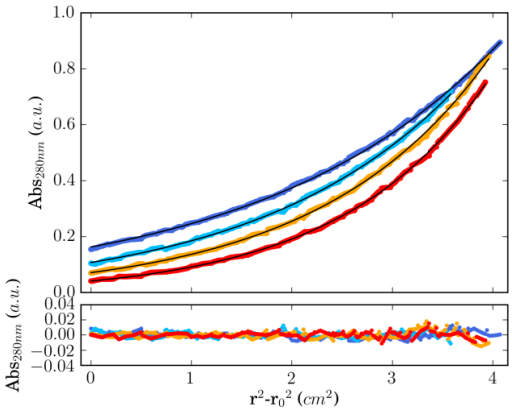


Figure 8-79: AUC data for αPP-GSGSG ($\bar{v}$ = 0.744 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 3864 Da (1.0 × monomer mass, 95 % confidence limits 3851 – 3878 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

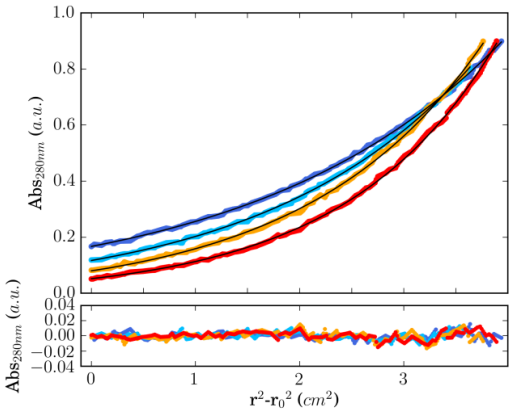


Figure 8-80: AUC data for αPP-GDNAT ($\bar{v}$ = 0.743 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 3860 Da (1.0 × monomer mass, 95 % confidence limits 3850 – 3870 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

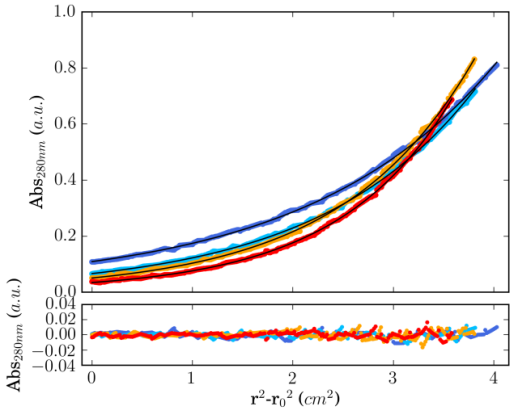


Figure 8-81: AUC data for αPP-GDNAT-4 ($\bar{v}$ = 0.747 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 4979 Da (1.0 × monomer mass, 95 % confidence limits 4956 – 5002 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

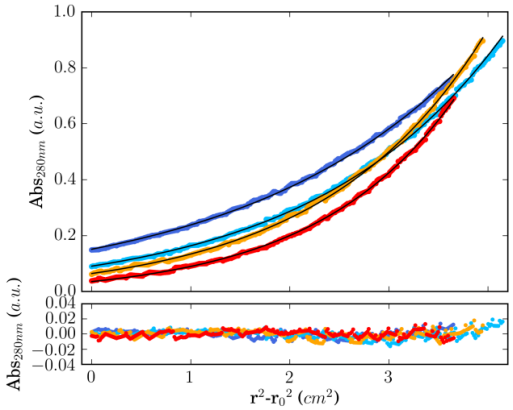


Figure 8-82: AUC data for αPP-NP ($\bar{v}$ = 0.740 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 3936 Da (1.0 × monomer mass, 95 % confidence limits 3920 – 3951 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

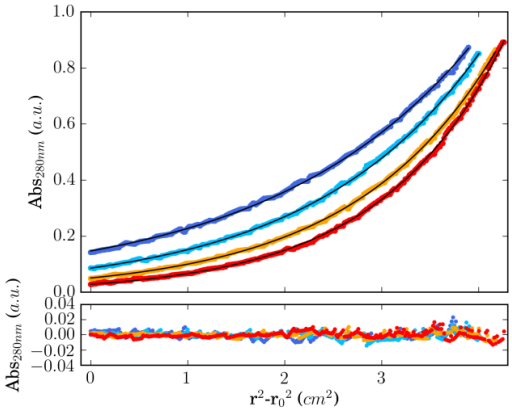


Figure 8-83: AUC data for αPP-NPE ($\bar{v}$ = 0.738 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 4296 Da (1.0 × monomer mass, 95 % confidence limits 4284 – 4308 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

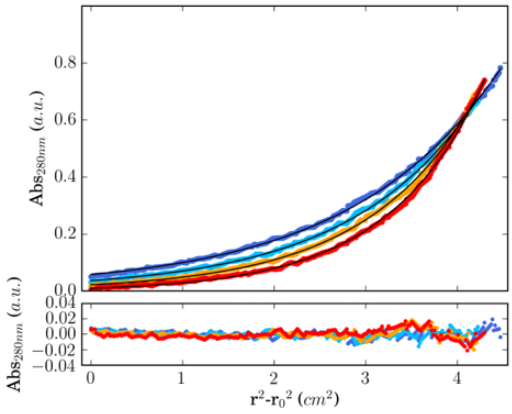


Figure 8-84: AUC data for αPP-NPELAA ($\bar{v}$ = 0.742 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 4646 Da (1.0 × monomer mass, 95 % confidence limits 4628 – 4664 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

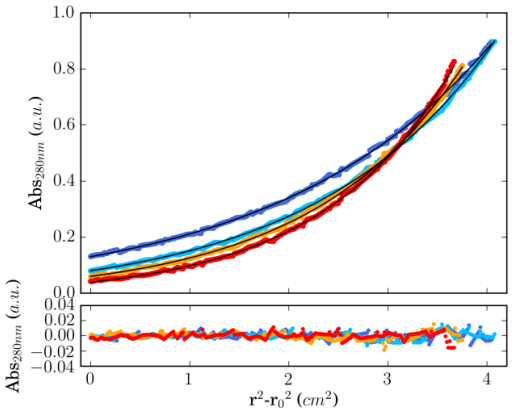


Figure 8-85: AUC data for αPP-NPE-4 ($\bar{v}$ = 0.741 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 4172 Da (1.0 × monomer mass, 95 % confidence limits 4160 – 4184 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

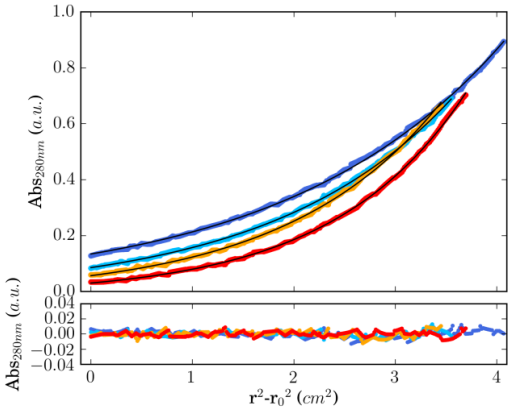


Figure 8-86: AUC data for αPP-NPE-5 ($\bar{v}$ = 0.737 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 4180 Da (1.0 × monomer mass, 95 % confidence limits 4166 – 4195 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

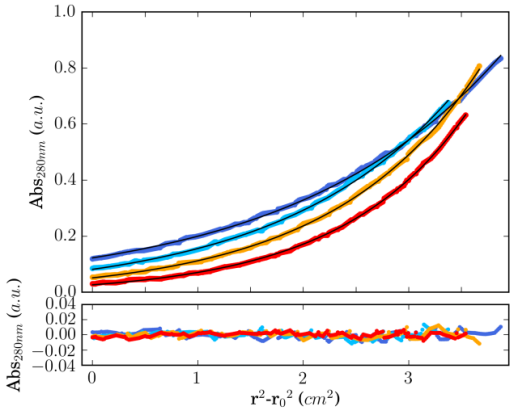


Figure 8-87: AUC data for αPP-NPE-6 ($\bar{v}$ = 0.739 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 4562 Da (1.0 × monomer mass, 95 % confidence limits 4548 – 4576 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

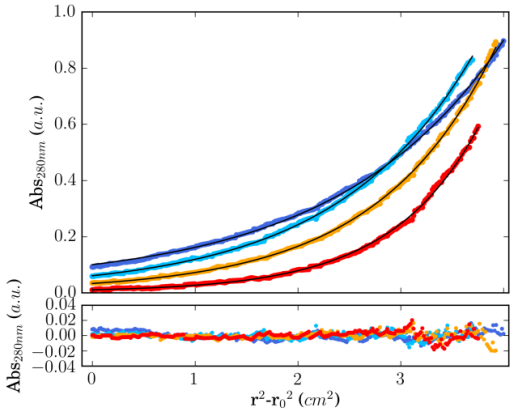


Figure 8-88: AUC data for PPαPP ($\bar{v}$ = 0.753 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 5951 Da (1.0 × monomer mass, 95 % confidence limits 5932 – 5970 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

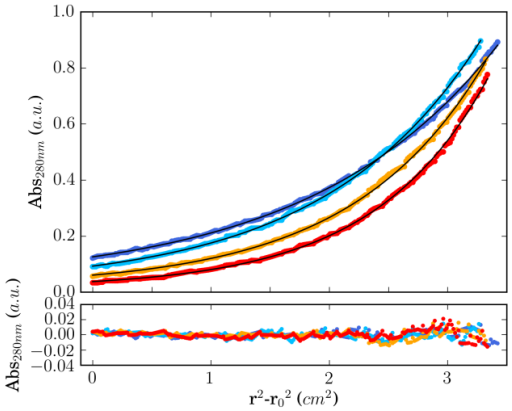


Figure 8-89: AUC data for PPαPG ($\bar{v}$ = 0.726 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 5919 Da (1.0 × monomer mass, 95 % confidence limits 5897 – 5941 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

8.3.4 Modulating stability in oPPα through elongation and cyclisation

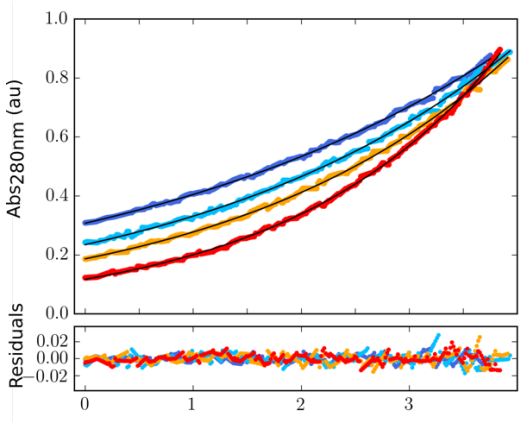


Figure 8-90: AUC data for oPPα-2 ($\bar{v}$ = 0.742 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 2503 Da (1.0 × monomer mass, 95 % confidence limits 2491 – 2514 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

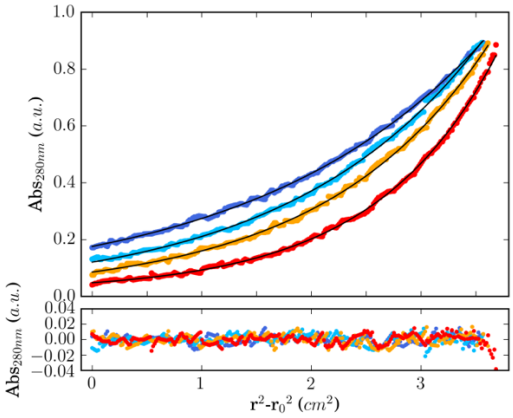


Figure 8-91: AUC data for oPPα-4-nocap ($\bar{v}$ = 0.751 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 4488 Da (1.0 × monomer mass, 95 % confidence limits 4466 – 4510 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

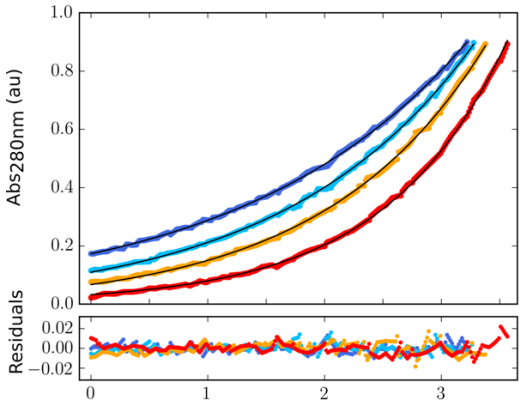


Figure 8-92: AUC data for oPPα-4 ($\bar{v}$ = 0.730 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 4846 Da (1.0 × monomer mass, 95 % confidence limits 4826 – 4867 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

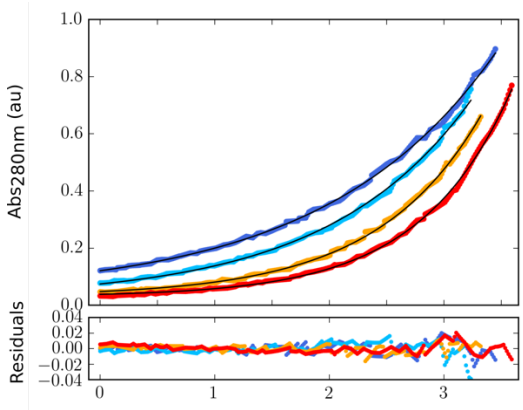


Figure 8-93: AUC data for oPPα-5 ($\bar{v}$ = 0.753 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 6402 Da (1.0 × monomer mass, 95 % confidence limits 6375 – 6429 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

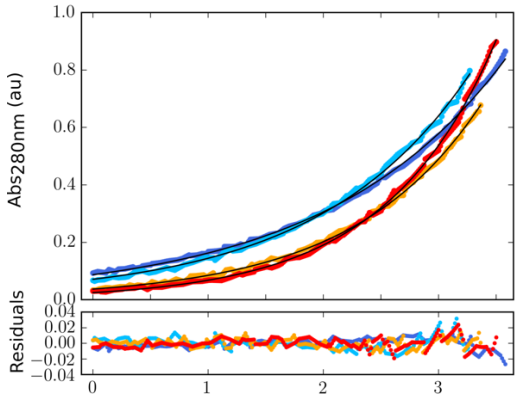


Figure 8-94: AUC data for oPPα-5-skip ($\bar{v}$ = 0.753 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 6085 Da (1.0 × monomer mass, 95 % confidence limits 6126 – 8257 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

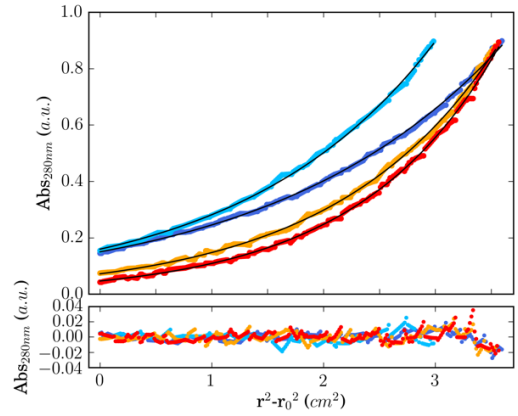


Figure 8-95: AUC data for cPPα-LSTKDL ($\bar{v}$ = 0.751 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 4100 Da (0.9 × monomer mass, 95 % confidence limits 4079 – 4121 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

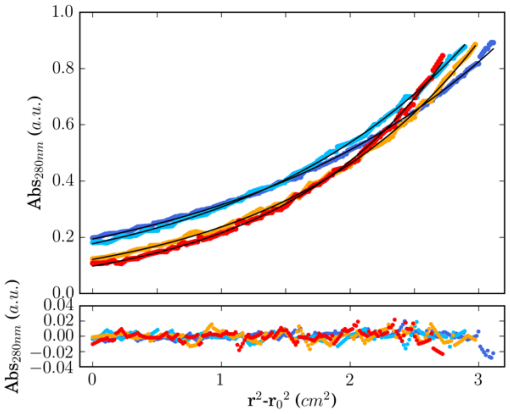


Figure 8-96: AUC data for cPPα-GPQY ($\bar{v}$ = 0.743 cm^3 g^{-1}). Sedimentation equilibrium data (top, dots) and fitted single-ideal species model curves at 44 (blue), 52 (aqua), 56 (orange), 60 (red) krpm. The fit returns a molecular weight of 4315 Da (1.0 × monomer mass, 95 % confidence limits 4268 – 4365 Da). Bottom: Residuals for the above fit (same colour scheme as above). Conditions: 130 µM peptide, PBS, pH 7.4, 20 °C.

8.4 NMR assignments, structure calculations and validation statistics for structural ensembles.

Table 8-1: Comparison of NMR data between PPα and oPPα.

	oPPα	**PPα 900MHz**	**PPα 700MHz**
Degree of Assignment			
H (%)	95.8	87	87
Heavy atoms (N and C) (%)	29.3	32.4	32.4
Number of Restraints			
Total Distance restraints	624	583	459
Intra-residue (|i-j|=0)	260	245	228
Sequential (|i-j|=1)	118	103	121
Medium Range (2 ≤ |i-j| < 5)	77	67	11
Long Range (|i-j| ≥ 5)	39	35	20
Ambiguous	130	133	79
Dihedral angle restraints	48	55	43
Restraints statistics[a]			
Rms of distance violations (Å)	0.029 ± 0.011	0.062±0.019	0.092±0.028
Rms of dihedral violations (Å)	0.030 ± 0.060	0.03 ±0.036	0.09±0.097
Violations > 0.5 Å	0	0	0
Violations > 0.3 Å	1.75 ± 1.4	5.65 ± 1.52	2.2±1.6
Violations > 0.1 Å	7.1 ± 1.9	12.45±1.63	4.6±2.1
Rms from idealised covalent geometry[a]			
Bonds (Å)	0.003± 0.0001	0.0027 ± 0.0001	0.0027± 0.0001
Angles (°)	0.42±0.01	0.429 ± 0.010	0.42±0.01
Impropers (°)	1.01±0.07	0.932 ± 0.084	0.98±0.10
Structural quality[c]			
Ramachandran Plot[d]			
Most favoured regions (%)	99.4	96.8	97.0
Allowed regions (%)	0.6	3.2	3.0
Generously allowed(%)	0	0	0
Disallowed regions (%)	0	0	0
Verify3D Z-score	-1.61	-2.89	-2.73

ProsaII Z- score	2.07	1.08	1.28
Procheck Z- score (Φ/Ψ)	2.05	1.97	1.61
Procheck Z-score (all)	2.25	1.54	0.470
MolProbility Z- score	0.86	0.13	0.30
No of close contacts	0	1	0
Coordinate precision rmsd (Å)			
All backbone atoms (Å)	0.52±0.13	0.514 ± 0.121	0.67±0.15
All heavy atoms (Å)	1.05±0.15	0.825 ± 0.122	1.18±0.14

Table 8-2: CH–π interactions in oPPα. 87 interactions (4.35 ± 1.13 per ensemble structure).

Model	CH donor	CH acceptor
1	Lys4 CA-HA Lys4 CB-HB2 Lys7 CA-HA Leu24 CB-HB2 Leu31 CB-HB2	Tyr27 Tyr27 Tyr20 Tyr20 Tyr27
2	Lys4 CA-HA Lys4 CB-HB2 Lys7 CA-HA Leu24 CB-HB2	Tyr27 Tyr27 Tyr20 Tyr20
3	Lys4 CA-HA Lys4 CB-HB2 Lys7 CA-HA Leu24 CB-HB2	Tyr27 Tyr27 Tyr20 Tyr20
4	Lys4 CA-HA Lys4 CB-HB2 Lys7 CA-HA Lys7 CB-HB2 Leu24 CB-HB2 Leu31 CB-HB2	Tyr27 Tyr27 Tyr20 Tyr20 Tyr20 Tyr27
5	Lys4 CA-HA Lys4 CB-HB2 Lys7 CA-HA Leu24 CB-HB2	Tyr27 Tyr27 Tyr20 Tyr20
6	Pro1 CA-HA Lys4 CB-HB2 Lys7 CA-HA Leu24 CB-HB2	Tyr34 Tyr27 Tyr20 Tyr20
7	Pro1 CA-HA Lys4 CA-HA LYS4 CB-HB2 Lys7 CA-HA	Tyr34 Tyr27 Tyr27 Tyr20
8	Pro1 CA-HA Lys4 CA-HA Lys4 CB-HB2 Lys7 CA-HA Lys7 CB-HB2 Leu24 CB-HB2 Leu31 CB-HB2	Tyr34 Tyr27 Tyr27 Tyr20 Tyr20 Tyr20 Tyr27
9	Lys4 CA-HA Lys7 CA-HA Leu24 CB-HB2 Leu31 CB-HB2	Tyr27 Tyr20 Tyr20 Tyr27
10	Lys4 CA-HA Lys4 CB-HB2 Lys7 CA-HA Leu24 CB-HB2	Tyr27 Tyr27 Tyr20 Tyr20
11	Lys4 CA-HA Lys7 CA-HA Leu24 CB-HB2	Tyr27 Tyr20 Tyr20
12	Lys4 CA-HA Lys4 CB-HB2	Tyr27 Tyr27

	Lys7 CA-HA Leu24 CB-HB2 Leu31 CB-HB2	Tyr20 Tyr20 Tyr27
13	Lys4 CA-HA Lys4 CB-HB2 Lys7 CA-HA Leu24 CB-HB2	Tyr27 Tyr27 Tyr20 Tyr20
14	Lys4 CA-HA Lys4 CB-HB2 Lys7 CA-HA Lys7 CB-HB2 Leu31 CB-HB2	Tyr27 Tyr27 Tyr20 Tyr20 Tyr27
15	Lys4 CA-HA Leu24 CB-HB2	Tyr27 Tyr20
16	Lys4 CA-HA Lys7 CA-HA Leu24 CB-HB2	Tyr27 Tyr20 Tyr20
17	Lys4 CA-HA Lys7 CA-HA Lys7 CB-HB2 Leu31 CB-HB2	Tyr27 Tyr20 Tyr20 Tyr27
18	Lys4 CA-HA Lys7 CA-HA Lys7 CB-HB2 Leu31 CB-HB2	Tyr27 Tyr20 Tyr20 Tyr27
19	Pro2 CD-HD2 Lys4 CA-HA Lys4 CB-HB2 Lys7 CA-HA Leu24 CB-HB2 Leu31 CB-HB2	Tyr34 Tyr27 Tyr27 Tyr20 Tyr20 Tyr27
20	Pro1 CA-HA Lys4 CA-HA Lys7 Ca-HA Lys7 CB-HB2 Leu24 CB-HB2	Tyr34 Tyr27 Tyr20 Tyr20 Tyr20

Pro

donor total = 5 i.e. 0.25 CH–π interactions per structure.

Leu donor total = 25 i.e. 1.25 CH–π interactions per structure.

Lys donor total = 57 i.e. 2.85 CH–π interactions per structure.

References

1. Kendrew, J.; Dickerson, R.; Strandberg, B.; Hart, R.; Davies, D.; Phillips, D.; Shore, V. *Nature,* **1960**, *185*, 427.

2. Perutz, M. F.; Rossmann, M. G.; Cullis, A. F.; Muirhead, H.; Will, G.; North, A. *Nature,* **1960**, *185*, 416.

3. Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R.; Wyckoff, H.; Phillips, D. C. *Nature,* **1958**, *181*, 662.

4. Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R. *Annu. Rev. Biophys.,* **2008**, *37*, 289−316.

5. Travaglini-Allocatelli, C.; Ivarsson, Y.; Jemth, P.; Gianni, S. *Curr. Opin. Struct. Biol.,* **2009**, *19*, 3.

6. Dill, K. A.; MacCallum, J. L. *Science,* **2012**, *338*, 1042.

7. Anfinsen, C. B.; Haber, E.; Sela, M.; White, F. *Proc. Natl. Acad. Sci. U.S.A.,* **1961**, *47*, 1309.

8. Anfinsen, C. B. *Science,* **1973**, *181*, 223.

9. Levinthal, C. *Mossbauer spectroscopy in biological systems,* **1969**, *67*, 22.

10. Honig, B. *J. Mol. Biol.,* **1999**, *293*, 283.

11. Dill, K. A.; Chan, H. S. *Nat. Struct. Mol. Biol.,* **1997**, *4*, 10.

12. Radford, S. E.; Dobson, C. M.; Evans, P. A. *Nature,* **1992**, *358*, 302.

13. Wang, Z.; Mottonen, J.; Goldsmith, E. J. *Biochemistry,* **1996**, *35*, 16443.

14. Dill, K. A. *Biochemistry,* **1990**, *29*, 7133.

15. Chen, J.; Stites, W. E. *Biochemistry,* **2001**, *40*, 15280.

16. Nakamura, H. *Q. Rev. Biophys.,* **1996**, *29*, 1.

17. Wolfenden, R. *J. Gen. Physiol.,* **2007**, *129*, 357.

18. Pace, C. N.; Scholtz, J. M.; Grimsley, G. R. *FEBS Lett.,* **2014**, *588*, 2177.

19. Robertson, A. D.; Murphy, K. P. *Chem. Rev.,* **1997**, *97*, 1251.

20. Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. *Science,* **2011**, *334*, 517.

21. Baker, E. G.; Bartlett, G. J.; Porter Goff, K. L.; Woolfson, D. N. *Acc. Chem. Res.,* **2017**, *50*, 2085−2092.

22. Nguyen, H.; Maier, J.; Huang, H.; Perrone, V.; Simmerling, C. *J. Am. Chem. Soc.,* **2014**, *136*, 13959.

23. Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. *Proc. Natl. Acad. Sci. U.S.A.,* **2012**, *109*, 17845.

24. Baker, E. G.; Bartlett, G. J.; Crump, M. P.; Sessions, R. B.; Linden, N.; Faul, C. F.; Woolfson, D. N. *Nat. Chem. Biol.,* **2015**, *11*, 221−8.

25. Lee, M. S.; Gippert, G. P.; Soman, K. V.; Case, D. A.; Wright, P. E. *Science,* **1989**, *245*, 635.

26. Craik, D. J.; Daly, N. L.; Waine, C. *Toxicon,* **2001**, *39*, 43.

27. Kale, S. S.; Villequey, C.; Kong, X.-D.; Zorzi, A.; Deyle, K.; Heinis, C. *Nat. Chem.,* **2018**, *10*, 715.

28. Craven, T. W.; Bonneau, R.; Kirshenbaum, K. *ChemBioChem,* **2016**, *17*, 1824.

29. Blundell, T. L.; Pitts, J. E.; Tickle, I. J.; Wood, S. P.; Wu, C. W. *Proc. Natl. Acad. Sci. U.S.A.,* **1981**, *78*, 4175−4179.

30. Bharat, T. A.; Eisenbeis, S.; Zeth, K.; Höcker, B. *Proc. Natl. Acad. Sci. U.S.A.,* **2008**, *105*, 9942.

31. Zhu, H.; Sepulveda, E.; Hartmann, M. D.; Kogenaru, M.; Ursinus, A.; Sulz, E.; Albrecht, R.; Coles, M.; Martin, J.; Lupas, A. N. *Elife,* **2016**, *5*, e16761.

32. Fletcher, J. M.; Boyle, A. L.; Bruning, M.; Bartlett, G. J.; Vincent, T. L.; Zaccai, N. R.; Armstrong, C. T.; Bromley, E. H.; Booth, P. J.; Brady, R. L. *ACS Synth. Biol.,* **2012**, *1*, 240−250.

33. Taylor, W. R.; Chelliah, V.; Hollup, S. M.; MacDonald, J. T.; Jonassen, I. *Structure,* **2009**, *17*, 1244.

34. Woolfson, D. N.; Bartlett, G. J.; Burton, A. J.; Heal, J. W.; Niitsu, A.; Thomson, A. R.; Wood, C. W. *Curr. Opin. Struct. Biol.,* **2015**, *33*, 16−26.

35. Razin, S.; Borunova, V.; Maksimenko, O.; Kantidze, O. *Biochemistry (Mosc.),* **2012**, *77*, 217.

36. Wolfe, S. A.; Nekludova, L.; Pabo, C. O. *Annu. Rev. Biophys. Biomol. Struct.,* **2000**, *29*, 183.

37. Gersbach, C. A.; Gaj, T.; Barbas III, C. F. *Acc. Chem. Res.,* **2014**, *47*, 2309.

38. Gifford, J. L.; Walsh, M. P.; Vogel, H. J. *Biochem. J.,* **2007**, *405*, 199.

39. Shaw, G. S.; Hodges, R. S.; Sykes, B. D. *Biochemistry,* **1992**, *31*, 9572.

40. Avrutina, O., Synthetic Cystine-Knot Miniproteins–Valuable Scaffolds for Polypeptide Engineering. In *Protein targeting compounds*, Springer: 2016; pp 121.

41. Lavergne, V.; J Taft, R.; F Alewood, P. *Curr. Top. Med. Chem.,* **2012**, *12*, 1514.

42. Rosengren, K. J.; Daly, N. L.; Plan, M. R.; Waine, C.; Craik, D. J. *J. Biol. Chem.,* **2003**, *278*, 8606.

43. Daly, N. L.; Craik, D. J. *Curr. Opin. Chem. Biol.,* **2011**, *15*, 362.

44. Kimura, R. H.; Levin, A. M.; Cochran, F. V.; Cochran, J. R. *Proteins: Struct., Funct., Bioinf.,* **2009**, *77*, 359.

45. Silverman, A. P.; Levin, A. M.; Lahti, J. L.; Cochran, J. R. *J. Mol. Biol.,* **2009**, *385*, 1064.

46. Craik, D. J.; Fairlie, D. P.; Liras, S.; Price, D. *Chem. Biol. Drug Des.,* **2013**, *81*, 136.

47. Kintzing, J. R.; Cochran, J. R. *Curr. Opin. Chem. Biol.,* **2016**, *34*, 143.

48. Gebauer, M.; Skerra, A. *Curr. Opin. Chem. Biol.,* **2009**, *13*, 245.

49. Zahnd, C.; Kawe, M.; Stumpp, M. T.; de Pasquale, C.; Tamaskovic, R.; Nagy-Davidescu, G.; Dreier, B.; Schibli, R.; Binz, H. K.; Waibel, R. *Cancer Res.,* **2010**, 1595.

50. Bhardwaj, G.; Mulligan, V. K.; Bahl, C. D.; Gilmore, J. M.; Harvey, P. J.; Cheneval, O.; Buchko, G. W.; Pulavarti, S. V.; Kaas, Q.; Eletsky, A. *Nature,* **2016**, *538*, 329.

51. Leaver-Fay, A.; Tyka, M.; Lewis, S. M.; Lange, O. F.; Thompson, J.; Jacak, R.; Kaufman, K. W.; Renfrew, P. D.; Smith, C. A.; Sheffler, W., ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. In *Methods Enzymol.*, Elsevier: 2011; Vol. 487, pp 545.

52. Rocklin, G. J.; Chidyausiku, T. M.; Goreshnik, I.; Ford, A.; Houliston, S.; Lemak, A.; Carter, L.; Ravichandran, R.; Mulligan, V. K.; Chevalier, A. *Science,* **2017**, *357*, 168.

53. 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. *Nature,* **2017**, *550*, 74.

54. Hodges, A. M.; Schepartz, A. *J. Am. Chem. Soc.,* **2007**, *129*, 11024.

55. Li, X.; Sutcliffe, M. J.; Schwartz, T. W.; Dobson, C. M. *Biochemistry,* **1992**, *31*, 1245−1253.

56. Golemi-Kotra, D.; Mahaffy, R.; Footer, M. J.; Holtzman, J. H.; Pollard, T. D.; Theriot, J. A.; Schepartz, A. *J. Am. Chem. Soc.,* **2004**, *126*, 4.

57. Link, N. M.; Hunke, C.; Mueller, J. W.; Eichler, J.; Bayer, P. *Biol. Chem.,* **2009**, *390*, 417−26.

58. Prehoda, K. E.; Lee, D. J.; Lim, W. A. *Cell,* **1999**, *97*, 471.

59. Yang, L.; Schepartz, A. *Biochemistry,* **2005**, *44*, 7469.

60. Smith, B. A.; Daniels, D. S.; Coplin, A. E.; Jordan, G. E.; McGregor, L. M.; Schepartz, A. *J. Am. Chem. Soc.,* **2008**, *130*, 2948.

61. Nicoll, A. J.; Allemann, R. K. *Org. Biomol. Chem.,* **2004**, *2*, 2175−80.

62. Struthers, M. D.; Cheng, R. P.; Imperiali, B. *Science,* **1996**, *271*, 342.

63. Mezo, A. R.; Cheng, R. P.; Imperiali, B. *J. Am. Chem. Soc.,* **2001**, *123*, 3885.

64. Dahiyat, B. I.; Mayo, S. L. *Science,* **1997**, *278*, 82.

65. Dahiyat, B. I.; Mayo, S. L. *Protein Sci.,* **1996**, *5*, 895.

66. Dahiyat, B. I.; Sarisky, C. A.; Mayo, S. L. *Science,* **1997**, *278*, 82.

67. Otto, J. J. *Curr. Opin. Cell Biol.,* **1994**, *6*, 105.

68. McKnight, C. J.; Matsudaira, P. T.; Kim, P. S. *Nat. Struct. Biol.,* **1997**, *4*, 180.

69. Chiu, T. K.; Kubelka, J.; Herbst-Irmer, R.; Eaton, W. A.; Hofrichter, J.; Davies, D. R. *Proc. Natl. Acad. Sci. U.S.A.,* **2005**, *102*, 7517.

70. Blanco, F. J.; Jimenez, M. A.; Herranz, J.; Rico, M.; Santoro, J.; Nieto, J. L. *J. Am. Chem. Soc.,* **1993**, *115*, 5887.

71. Cox, J. P.; Evans, P. A.; Packman, L. C.; Williams, D. H.; Woolfson, D. N. *J. Mol. Biol.,* **1993**, *234*, 483.

72. Pastor, M. T.; de la Paz, M. L.; Lacroix, E.; Serrano, L.; Pérez-Payá, E. *Proc. Natl. Acad. Sci. U.S.A.,* **2002**, *99*, 614.

73. Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. *Proc. Natl. Acad. Sci. U.S.A.,* **2001**, *98*, 5578.

74. Russell, S. J.; Cochran, A. G. *J. Am. Chem. Soc.,* **2000**, *122*, 12600.

75. Honda, S.; Yamasaki, K.; Sawada, Y.; Morii, H. *Structure,* **2004**, *12*, 1507.

76. Watanabe, H.; Yamasaki, K.; Honda, S. *J. Biol. Chem.,* **2014**, *289*, 3394.

77. Riemen, A. J.; Waters, M. L. *Biochemistry,* **2009**, *48*, 1525.

78. Mahalakshmi, R. *Arch. Biochem. Biophys.,* **2018**, *661*, 39.

79. Espinosa, J. F.; Gellman, S. H. *Angew. Chem. Int. Ed.,* **2000**, *39*, 2330.

80. Fesinmeyer, R. M.; Hudson, F. M.; Andersen, N. H. *J. Am. Chem. Soc.,* **2004**, *126*, 7238.

81. Santiveri, C. M.; Jiménez, M. A. *J. Pept. Sci.,* **2010**, *94*, 779.

82. Ramírez-Alvarado, M.; Blanco, F. J.; Serrano, L. *Nat. Struct. Mol. Biol.,* **1996**, *3*, 604.

83. Minor Jr, D. L.; Kim, P. S. *Nature,* **1994**, *367*, 660.

84. Kiehna, S. E.; Waters, M. L. *Protein Sci.,* **2003**, *12*, 2657.

85. Kortemme, T.; Ramírez-Alvarado, M.; Serrano, L. *Science,* **1998**, *281*, 253.

86. Sharman, G. J.; Searle, M. S. *J. Am. Chem. Soc.,* **1998**, *120*, 5291.

87. Schenck, H. L.; Gellman, S. H. *J. Am. Chem. Soc.,* **1998**, *120*, 4869.

88. Doig, A. *Chem. Commun.,* **1997**, *0*, 2153.

89. Kuznetsov, S. V.; Hilario, J.; Keiderling, T. A.; Ansari, A. *Biochemistry,* **2003**, *42*, 4321.

90. López, M.; Lacroix, E.; Ramírez-Alvarado, M.; Serrano, L. *J. Mol. Biol.,* **2001**, *312*, 229.

91. Kung, V. M.; Cornilescu, G.; Gellman, S. H. *Angew. Chem. Int. Ed.,* **2015**, *127*, 14544.

92. Staub, O.; Rotin, D. *Structure,* **1996**, *4*, 495.

93. Macias, M. J.; Hyvönen, M.; Baraldi, E.; Schultz, J.; Sudol, M.; Saraste, M.; Oschkinat, H. *Nature,* **1996**, *382*, 646.

94. Macias, M. J.; Gervais, V.; Civera, C.; Oschkinat, H. *Nat. Struct. Mol. Biol.,* **2000**, *7*, 375.

95. Stewart, A. L.; Park, J. H.; Waters, M. L. *Biochemistry,* **2011**, *50*, 2575.

96. Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nat. Struct. Mol. Biol.,* **2002**, *9*, 425.

97. Gellman, S. H.; Woolfson, D. N. *Nat. Struct. Mol. Biol.,* **2002**, *9*, 408.

98. Barua, B.; Lin, J. C.; Williams, V. D.; Kummler, P.; Neidigh, J. W.; Andersen, N. H. *Protein Eng., Des. Sel.,* **2008**, *21*, 171.

99. Scian, M.; Lin, J. C.; Le Trong, I.; Makhatadze, G. I.; Stenkamp, R. E.; Andersen, N. H. *Proc. Natl. Acad. Sci. U.S.A.,* **2012**, *109*, 12521.

100. Paschek, D.; Hempel, S.; García, A. E. *Proc. Natl. Acad. Sci. U.S.A.,* **2008**, *105*, 17754.

101. Huang, P.-S.; Feldmeier, K.; Parmeggiani, F.; Velasco, D. A. F.; Höcker, B.; Baker, D. *Nat. Chem. Biol.,* **2016**, *12*, 29.

102. Rao, S. T.; Rossmann, M. G. *J. Mol. Biol.,* **1973**, *76*, 241.

103. Kobe, B.; Kajava, A. V. *Curr. Opin. Struct. Biol.,* **2001**, *11*, 725.

104. Liang, H.; Chen, H.; Fan, K.; Wei, P.; Guo, X.; Jin, C.; Zeng, C.; Tang, C.; Lai, L. *Angew. Chem. Int. Ed.,* **2009**, *48*, 3301.

105. Craven, T. W.; Cho, M. K.; Traaseth, N. J.; Bonneau, R.; Kirshenbaum, K. *J. Am. Chem. Soc.,* **2016**, *138*, 1543.

106. Woolfson, D. N.; Baker, E. G.; Bartlett, G. J. *Science,* **2017**, *357*, 133.

107. Wolny, M.; Batchelor, M.; Bartlett, G. J.; Baker, E. G.; Kurzawa, M.; Knight, P. J.; Dougan, L.; Woolfson, D. N.; Paci, E.; Peckham, M. *Sci. Rep.,* **2017**, *7*, 44341.

108. Liu, J.; Rost, B. *Protein Sci.,* **2001**, *10*, 1970.

109. Rose, A.; Manikantan, S.; Schraegle, S. J.; Maloy, M. A.; Stahlberg, E. A.; Meier, I. *Plant Physiol.,* **2004**, *134*, 927.

110. Rackham, O. J.; Madera, M.; Armstrong, C. T.; Vincent, T. L.; Woolfson, D. N.; Gough, J. *J. Mol. Biol.,* **2010**, *403*, 480.

111. Strelkov, S. V.; Burkhard, P. *J. Struct. Biol.,* **2002**, *137*, 54.

112. Mason, J. M.; Arndt, K. M. *ChemBioChem,* **2004**, *5*, 170.

113. Woolfson, D. N.; Bartlett, G. J.; Bruning, M.; Thomson, A. R. *Curr. Opin. Struct. Biol.,* **2012**, *22*, 432.

114. Gruber, M.; Söding, J.; Lupas, A. N. *J. Struct. Biol.,* **2006**, *155*, 140.

115. Vincent, T. L.; Green, P. J.; Woolfson, D. N. *Bioinformatics,* **2012**, *29*, 69.

116. McDonnell, A. V.; Jiang, T.; Keating, A. E.; Berger, B. *Bioinformatics,* **2005**, *22*, 356.

117. Trigg, J.; Gutwin, K.; Keating, A. E.; Berger, B. *PLoS One,* **2011**, *6*, e23519.

118. Lupas, A. N.; Bassler, J. *Trends Biochem. Sci,* **2017**, *42*, 130.

119. Lupas, A. N.; Gruber, M., The structure of α-helical coiled coils. In *Adv. Protein Chem.*, Elsevier: 2005; Vol. 70, pp 37.

120. Harbury, P. B.; Zhang, T.; Kim, P. S.; Alber, T. *Science,* **1993**, *262*, 1401.

121. Oakley, M. G.; Hollenbeck, J. J. *Curr. Opin. Struct. Biol.,* **2001**, *11*, 450.

122. Monera, O. D.; Zhou, N. E.; Lavigne, P.; Kay, C. M.; Hodges, R. S. *J. Biol. Chem.,* **1996**, *271*, 3995.

123. O'Shea, E. K.; Lumb, K. J.; Kim, P. S. *Curr. Biol.,* **1993**, *3*, 658.

124. Negron, C.; Keating, A. E. *J. Am. Chem. Soc.,* **2014**, *136*, 16544.

125. Pauling, L.; Corey, R. B.; Branson, H. R. *Proc. Natl. Acad. Sci. U.S.A.,* **1951**, *37*, 205.

126. Hollingsworth, S. A.; Karplus, P. A. *Biomol. Concepts,* **2010**, *1*, 271.

127. 攀芬 *J. Mol. Biol.,* **1963**, *7*, 95.

128. Fodje, M.; Al-Karadaghi, S. *Protein Eng., Des. Sel.,* **2002**, *15*, 353.

129. Armen, R.; Alonso, D. O. V.; Daggett, V. *Protein Sci.,* **2003**, *12*, 1145.

130. Gessmann, R.; Brückner, H.; Petratos, K. *J. Pept. Sci.,* **2003**, *9*, 753.

131. Nesloney, C. L.; Kelly, J. W. *Biorg. Med. Chem.,* **1996**, *4*, 739.

132. Woolfson, D. N. *Subcell. Biochem.,* **2017**, *82*, 35−61.

133. Woolfson, D. N. *Adv. Protein Chem.,* **2005**, *70*, 79.

134. Crick, F. H. *Acta. Crystallogr.,* **1953**, *6*, 689−697.

135. Crick, F. H. *Acta. Crystallogr.,* **1953**, *6*, 685.

136. Harbury, P. B.; Tidor, B.; Kim, P. S. *Proc. Natl. Acad. Sci. U.S.A.,* **1995**, *92*, 8408.

137. Harbury, P. B.; Plecs, J. J.; Tidor, B.; Alber, T.; Kim, P. S. *Science,* **1998**, *282*, 1462.

138. Offer, G.; Hicks, M. R.; Woolfson, D. N. *J. Struct. Biol.,* **2002**, *137*, 41.

139. Grigoryan, G.; DeGrado, W. F. *J. Mol. Biol.,* **2011**, *405*, 1079.

140. Wood, C. W.; Bruning, M.; Ibarra, A. A.; Bartlett, G. J.; Thomson, A. R.; Sessions, R. B.; Brady, R. L.; Woolfson, D. N. *Bioinformatics,* **2014**, *30*, 3029.

141. Wood, C. W.; Woolfson, D. N. *Protein Sci.,* **2018**, *27*, 103.

142. Wood, C. W.; Heal, J. W.; Thomson, A. R.; Bartlett, G. J.; Ibarra, A. A.; Brady, R. L.; Sessions, R. B.; Woolfson, D. N. *Bioinformatics,* **2017**, *33*, 3043.

143. Walshaw, J.; Woolfson, D. N. *J. Struct. Biol.,* **2003**, *144*, 349.

144. Harbury, P. B.; Kim, P. S.; Alber, T. *Nature,* **1994**, *371*, 80.

145. Thomson, A. R.; Wood, C. W.; Burton, A. J.; Bartlett, G. J.; Sessions, R. B.; Brady, L. J.; Woolfson, D. N. *Science,* **2014**, *346*, 485.

146. Ghirlanda, G.; Lear, J. D.; Ogihara, N. L.; Eisenberg, D.; DeGrado, W. F. *J. Mol. Biol.,* **2002**, *319*, 243.

147. Lovejoy, B.; Choe, S.; Cascio, D.; McRorie, D. K.; DeGrado, W. F.; Eisenberg, D. *Science,* **1993**, *259*, 1288.

148. Koronakis, V.; Sharff, A.; Koronakis, E.; Luisi, B.; Hughes, C. *Nature,* **2000**, *405*, 914.

149. Yoshizumi, A.; Fletcher, J. M.; Yu, Z.; Persikov, A. V.; Bartlett, G. J.; Boyle, A. L.; Vincent, T. L.; Woolfson, D. N.; Brodsky, B. *J. Biol. Chem.,* **2011**, *286*, 17512.

150. Burgess, N. C.; Sharp, T. H.; Thomas, F.; Wood, C. W.; Thomson, A. R.; Zaccai, N. R.; Brady, R. L.; Serpell, L. C.; Woolfson, D. N. *J. Am. Chem. Soc.,* **2015**, *137*, 10554.

151. Gradišar, H.; Božič, S.; Doles, T.; Vengust, D.; Hafner-Bratkovič, I.; Mertelj, A.; Webb, B.; Šali, A.; Klavžar, S.; Jerala, R. *Nat. Chem. Biol.,* **2013**, *9*, 362.

152. Burton, A. J.; Thomson, A. R.; Dawson, W. M.; Brady, R. L.; Woolfson, D. N. *Nat. Chem.,* **2016**, *8*, 837.

153. Baker, E. G.; Williams, C.; Hudson, K. L.; Bartlett, G. J.; Heal, J. W.; Porter Goff, K. L.; Sessions, R. B.; Crump, M. P.; Woolfson, D. N. *Nat. Chem. Biol.,* **2017**, *13*, 764−770.

154. Shoulders, M. D.; Raines, R. T. *Annu. Rev. Biochem.,* **2009**, *78*, 929.

155. Berisio, R.; Loguercio, S.; De Simone, A.; Zagari, A.; Vitagliano, L. *Protein Pept. Lett.,* **2006**, *13*, 847.

156. Stapley, B. J.; Creamer, T. P. *Protein Sci.,* **1999**, *8*, 587.

157. Adzhubei, A. A.; Sternberg, M. J. *J. Mol. Biol.,* **1993**, *229*, 472.

158. Jha, A. K.; Colubri, A.; Zaman, M. H.; Koide, S.; Sosnick, T. R.; Freed, K. F. *Biochemistry,* **2005**, *44*, 9691.

159. Cubellis, M.; Caillez, F.; Blundell, T.; Lovell, S. *Proteins: Struct., Funct., Bioinf.,* **2005**, *58*, 880.

160. Adzhubei, A. A.; Sternberg, M. J.; Makarov, A. A. *J. Mol. Biol.,* **2013**, *425*, 2100.

161. Rath, A.; Davidson, A. R.; Deber, C. M. *J. Pept. Sci.,* **2005**, *80*, 179.

162. King, S. M.; Johnson, W. C. *Proteins: Struct., Funct., Bioinf.,* **1999**, *35*, 313.

163. Srinivasan, R.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.,* **1999**, *96*, 14258.

164. Cubellis, M. V.; Cailliez, F.; Lovell, S. C. *BMC Bioinformatics,* **2005**, *6*, 8.

165. Mansiaux, Y.; Joseph, A. P.; Gelly, J.-C.; de Brevern, A. G. *PLoS One,* **2011**, *6*, e18401.

166. Larson, M. R.; Rajashankar, K. R.; Patel, M. H.; Robinette, R. A.; Crowley, P. J.; Michalek, S.; Brady, L. J.; Deivanayagam, C. *Proc. Natl. Acad. Sci. U.S.A.,* **2010**, *107*, 5983−8.

167. Brandl, M.; Weiss, M. S.; Jabs, A.; Sühnel, J.; Hilgenfeld, R. *J. Mol. Biol.,* **2001**, *307*, 357−377.

168. Hudson, K. L.; Bartlett, G. J.; Diehl, R. C.; Agirre, J.; Gallagher, T.; Kiessling, L. L.; Woolfson, D. N. *J. Am. Chem. Soc.,* **2015**, *137*, 15152−60.

169. Kuipers, B. J. H.; Gruppen, H. *J. Argic. Food Chem.,* **2007**, *55*, 5445.

170. Sreerama, N.; Woody, R. W. *Methods Enzymol.,* **2004**, *383*, 318.

171. Wallace, B. A.; Janes, R. W., *Modern Techniques for Circular Dichroism and Synchrotron Radiation Circular Dichroism Spectroscopy*. IOS Press: 2009; Vol. 1.

172. Lopes, J. L. S.; Miles, A. J.; Whitmore, L.; Wallace, B. A. *Protein Sci.,* **2014**, *23*, 1765.

173. Whitmore, L.; Woollett, B.; Miles, A. J.; Klose, D. P.; Janes, R. W.; Wallace, B. A. *Nucleic Acids Res.,* **2011**, *39*, 480.

174. Li, Z.; Hirst, J. D. *Chem. Sci.,* **2017**, *8*, 4318.

175. Myers, J. M.; Pace, C. N.; Scholtz, J. M. *Proc. Nat. Acad. Sci. U.S.A.,* **1997**, *94*, 2833

176. Scholtz, J. M.; Qian, H.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Biopolymers,* **1991**, *31*, 1463.

177. Cole, J. L.; Lary, J. W.; P. Moody, T.; Laue, T. M., Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium. In *Biophysical Tools for Biologists, Volume One: In Vitro Techniques*, 2008; pp 143.

178. Laue, T. M.; Stafford, W. F. *Annu. Rev. Biophys. Biomol. Struct.,* **1999**, *28*, 75.

179. Scott, D. J.; Harding, S. E.; Rowe, A. J., A Brief Introduction to the Analytical Ultracentrifugation of Proteins for Beginners. In *Analytical Ultracentrifugation*, 2007; pp 1.

180. Schuck, P. *Biophys. J.,* **2000**, *78*, 1606.

181. Schuster, T. M.; Laue, T. M., *Modern Analytical Ultracentrifugation*. Springer: 1994.

182. Pabo, C. O.; Peisach, E.; Grant, R. A. *Annu. Rev. Biochem.,* **2001**, *70*, 313.

183. Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. *Proteins,* **2005**, *59*, 687.

184. Rieping, W.; Habeck, M.; Bardiaux, B.; Bernard, A.; Malliavin, T. E.; Nilges, M. *Bioinformatics,* **2007**, *23*, 381.

185. Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta. Crystallogr.,* **1998**, *D54*, 905.

186. Cheung, M. S.; Maguire, M. L.; Stevens, T. J.; Broadhurst, R. W. *J. Magn. Reson.,* **2010**, *202*, 223.

187. Bhattacharya, A.; Tejero, R.; Montelione, G. T. *Proteins,* **2007**, *66*, 778.

188. Lee, B.; Richards, F. M. *J. Mol. Biol.,* **1971**, *55*, 379.

189. Hubbard, S.; Thornton, J. *Department of Biochemistry and Molecular Biology, University College London,* **1993**.

190. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comp. Phys. Comm.,* **1995**, *91*, 43.

191. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.,* **2008**, *4*, 435.

192. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins,* **2010**, *78*, 1950.

193. Humphrey, W.; Dalke, A.; Sculten, K. *J. Mol. Graph.,* **1996**, *14*, 27.

194. Armstrong, C. T.; Boyle, A. L.; Bromley, E. H.; Mahmoud, Z. N.; Smith, L.; Thomson, A. R.; Woolfson, D. N. *Faraday Discuss.,* **2009**, *143*, 305.

195. Thompson, K. E.; Bashor, C. J.; Lim, W. A.; Keating, A. E. *ACS Synth. Biol.,* **2012**, *1*, 118.

196. Lupas, A. N.; Bassler, J. *Trends Biochem. Sci,* **2017**, *42*, 130.

197. Crick, F. H. *Acta. Crystallogr.,* **1953**, *6*, 689.

198. Zhu, B. Y.; Zhou, M. E.; Kay, C. M.; Hodges, R. S. *Protein Sci.,* **1993**, *2*, 383.

199. O'Shea, E. K.; Klemm, J. D.; Kim, P. S.; Alber, T. *Science,* **1991**, *254*, 539.

200. Lee, D. L.; Ivaninskii, S.; Burkhard, P.; Hodges, R. S. *Protein Sci.,* **2003**, *12*, 1395.

201. Fletcher, J. M.; Bartlett, G. J.; Boyle, A. L.; Danon, J. J.; Rush, L. E.; Lupas, A. N.; Woolfson, D. N. *ACS Chem. Biol.,* **2017**, *12*, 528.

202. Hodges, R. S.; Saund, A. K.; Chong, P.; St-Pierre, S.; Reid, R. E. *J. Biol. Chem.,* **1981**, *256*, 1214.

203. Thomas, F.; Boyle, A. L.; Burton, A. J.; Woolfson, D. N. *J. Am. Chem. Soc.,* **2013**, *135*, 5161−6.

204. Schreiber, G.; Fersht, A. R. *Nat. Struct. Biol.,* **1996**, *3*, 427−431.

205. Selzer, T.; Albeck, S.; Schreiber, G. *Nat. Struct. Mol. Biol.,* **2000**, *7*, 537.

206. Bromley, E. H.; Sessions, R. B.; Thomson, A. R.; Woolfson, D. N. *J. Am. Chem. Soc.,* **2008**, *131*, 928.

207. Nautiyal, S.; Woolfson, D. N.; King, D. S.; Alber, T. *Biochemistry,* **1995**, *34*, 11645.

208. Reinke, A. W.; Grant, R. A.; Keating, A. E. *J. Am. Chem. Soc.,* **2010**, *132*, 6025.

209. Scholtz, J. M.; Pace, C. N. *Biochem. J.,* **1998**, *75*, 422.

210. Blaber, M.; Zhang, X.; Matthews, B. W. *Science,* **1993**, *260*, 1637.

211. Myers, J. M.; Pace, C. N.; Scholtz, J. M. *Proc. Natl. Acad. Sci. U.S.A.,* **1997**, *94*, 2833.

212. Chou, P. Y.; Fasman, G. D. *Annu. Rev. Biochem.,* **1978**, *47*, 251.

213. Wojick, J.; Altmann, K. H.; Scheraga, H. A. *Biopolymers,* **1990**, *30*, 121.

214. O'Neil, K. T.; DeGrado, W. F. *Science,* **1990**, *250*, 646.

215. Hermans, J.; Anderson, A. G.; Yun, R. H. *Biochemistry,* **1992**, *31*, 5646.

216. Creamer, T. P.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.,* **1992**, *89*, 5937.

217. Chou, P. Y.; Fasman, G. D. *Biochemistry,* **1974**, *13*, 211.

218. Chou, P. Y.; Fasman, G. D. *Annu. Rev. Biochem.,* **1978**, *47*, 251.

219. Scholtz, J. M.; Marqusee, S.; Baldwin, R. L.; York, E. J.; Stewart, J. M.; Santoro, M.; Bolen, D. W. *Proc. Natl. Acad. Sci. U.S.A.,* **1991**, *88*, 2854.

220. Aurora, R.; Creamer, T. P.; Srinivasan, R.; Rose, G. D. *J. Biol. Chem.,* **1997**, *272*, 1413.

221. Nemethy, G. *J. Phys. Chem.,* **1966**, *70*, 998.

222. Lupque, I.; Mayorga, O. L.; Freire, E. *Biochemistry,* **1996**, *35*, 13681.

223. Hu, X.; Wang, H.; Ke, H.; Kuhlman, B. *Proc. Natl. Acad. Sci. U.S.A.,* **2007**, *104*, 17668.

224. Aurora, R.; Rosee, G. D. *Protein Sci.,* **1998**, *7*, 21.

225. Richardson, J. S.; Richardson, D. C. *Science,* **1988**, *240*, 1648.

226. Kelly, S. M.; Price, N. C. *Curr. Protein Peptide Sci.,* **2000**, *1*, 349.

227. Wagner, G.; Wüthrich, K. *J. Mol. Biol.,* **1982**, *155*, 347.

228. Wider, G.; Lee, K. H.; Wüthrich, K. *J. Mol. Biol.,* **1982**, *155*, 367.

229. Wüthrich, K. *Biopolymers,* **1983**, *22*, 131.

230. Barlow, D. J.; Thornton, J. M. *J. Mol. Biol.,* **1983**, *168*, 867−885.

231. Donald, J. E.; Kulp, D. W.; DeGrado, W. F. *Proteins,* **2011**, *79*, 898.

232. Schreiber, G.; Haran, G.; Zhou, H.-X. *Chem. Rev.,* **2009**, *109*, 839−860.

233. Hemsath, L.; Dvorsky, R.; Fiegen, D.; Carlier, M. F.; Ahmadian, M. R. *Mol .Cell,* **2005**, *20*, 313−24.

234. Blochliger, N.; Xu, M.; Caflisch, A. *Biophys. J.,* **2015**, *108*, 2362−2370.

235. Sheinerman, F. B.; Norel, R.; Honig, B. *Curr. Opin. Struct. Biol.,* **2000**, *10*, 153.

236. Schreiber, G. *Curr. Opin. Struct. Biol.,* **2002**, *12*, 41.

237. Kiel, C.; Selzer, T.; Shaul, Y.; Schreiber, G.; Herrmann, C. *Proc. Natl. Acad. Sci. U.S.A.,* **2004**, *101*, 9223.

238. Connolly, M. L. *Science,* **1983**, *221*, 709.

239. Farber, P. J.; Mittermaier, A. *Protein Sci.,* **2008**, *17*, 644.

240. Parry, D. A.; Fraser, R. B.; Squire, J. M. *J. Struct. Biol.,* **2008**, *163*, 258.

241. Fletcher, J. M.; Harniman, R. L.; Barnes, F. R.; Boyle, A. L.; Collins, A.; Mantell, J.; Sharp, T. H.; Antognozzi, M.; Booth, P. J.; Linden, N. *Science,* **2013**, 595−599.

242. Xu, C.; Liu, R.; Mehta, A. K.; Guerrero-Ferreira, R. C.; Wright, E. R.; Dunin-Horkawicz, S.; Morris, K.; Serpell, L. C.; Zuo, X.; Wall, J. S.; Conticello, V. P. *J. Am. Chem. Soc.,* **2013**, *135*, 15565.

243. Steinkruger, J. D.; Bartlett, G. J.; Hadley, E. B.; Fay, L.; Woolfson, D. N.; Gellman, S. H. *J. Am. Chem. Soc.,* **2012**, *134*, 2626.

244. Hadley, E. B.; Testa, O. D.; Woolfson, D. N.; Gellman, S. H. *Proc. Natl. Acad. Sci. U.S.A.,* **2008**, *105*, 530.

245. Testa, O. D.; Moutevelis, E.; Woolfson, D. N. *Nucleic Acids Res.,* **2008**, *37*, D315.

246. Li, X.; Sutcliffe, M. J.; Schwartz, T. W.; Dobson, C. M. *Biochemistry,* **1992**, *31*, 1245.

247. Kimmel, J.; Hayden, L. J.; Pollock, H. G. *J. Biol. Chem.,* **1975**, *250*, 9369.

248. Nygaard, R.; Nielbo, S.; Schwartz, T. W.; Poulsen, F. M. *Biochemistry,* **2006**, *45*, 8350.

249. Monks, S. A.; Karagianis, G.; Howlett, G. J.; Norton, R. S. *J. Biomol. NMR,* **1996**, *8*, 379.

250. Crooks, G. E.; Hon, G.; Chandonia, J.-M.; Brenner, S. E. *Genome Res.,* **2004**, *14*, 1188.

251. Schneider, T. D.; Stephens, R. M. *Nucleic Acids Res.,* **1990**, *18*, 6097.

252. Minakata, H.; Taylor, J. W.; Walker, M.; Miller, R.; Kaiser, E. T. *J. Biol. Chem.,* **1989**, *264*, 7907.

253. Kim, M. K.; Kang, Y. K. *Protein Sci.,* **1999**, *8*, 1492.

254. Weiss, M. A.; Ellenberger, T.; Wobbe, C. R.; Lee, J. P.; Harrison, S. C.; Struhl, K. *Nature,* **1990**, *347*, 575.

255. Fezoui, Y.; Weaver, D. L.; Osterhout, J. J. *Proc. Natl. Acad. Sci. U.S.A.,* **1994**, *91*, 3675.

256. Marky, L. A.; Breslauer, K. J. *Biopolymers,* **1987**, *26*, 1601.

257. Daggett, V. *Chem. Rev.,* **2006**, *106*, 1898.

258. Hospital, A.; Goñi, J. R.; Orozco, M.; Gelpí, J. L. *Adv. Appl. Bioinform.,* **2015**, *8*, 37.

259. McIntosh-Smith, S.; Wilson, T.; Ibarra, A. Í.; Crisp, J.; Sessions, R. B. *Comput. J.,* **2012**, *55*, 192.

260. McIntosh-Smith, S.; Price, J.; Sessions, R. B.; Ibarra, A. A. *Int. J. High Perform. Comput. Appl.,* **2015**, *29*, 119.

261. Regan, L. *Curr. Biol.,* **1994**, *4*, 656.

262. Hill, R. B.; Raleigh, D. P.; Lombardi, A.; DeGrado, W. F. *Acc. Chem. Res.,* **2000**, *33*, 745.

263. Nagi, A. D.; Regan, L. *Fold Des.,* **1997**, *2*, 67.

264. Bruch, M. D.; Dhingra, M. M.; Gierasch, L. M. *Proteins: Struct., Funct., Bioinf.,* **1991**, *10*, 130.

265. Wang, D.; Chen, K.; Kulp, J. L.; Arora, P. S. *J. Am. Chem. Soc.,* **2006**, *128*, 9248.

266. Kallenbach, N. R.; Lyu, P.; Zhou, H., CD spectroscopy and the helix-coil transition in peptides and polypeptides. In *Circular dichroism and the conformational analysis of biomolecules*, Springer: 1996; pp 201.

267. Boyle, A. L.; Bromley, E. H.; Bartlett, G. J.; Sessions, R. B.; Sharp, T. H.; Williams, C. L.; Curmi, P. M.; Forde, N. R.; Linke, H.; Woolfson, D. N. *J. Am. Chem. Soc.,* **2012**, *134*, 15457.

268. Presta, L. G.; Rose, G. D. *Science,* **1988**, *240*, 1632.

269. Fletcher, J. M.; Horner, K. A.; Bartlett, G. J.; Rhys, G. R.; Wilson, A.; Woolfson, D. N. *Chem. Sci.,* **2018**.

270. Nooren, I. M.; Thornton, J. M. *EMBO J.,* **2003**, *22*, 3486.

271. Perkins, J. R.; Diboun, I.; Dessailly, B. H.; Lees, J. G.; Orengo, C. *Structure,* **2010**, *18*, 1233.

272. Keskin, O.; Gursoy, A.; Ma, B.; Nussinov, R. *Chem. Rev.,* **2008**, *108*, 1225.

273. Nussinov, R.; Tsai, C.-J. *Cell,* **2013**, *153*, 293.

274. Shogren-Knaak, M. A.; Alaimo, P. J.; Shokat, K. M. *Annu. Rev. Cell Dev. Biol.,* **2001**, *17*, 405.

275. Huang, Z. *Pharmacol. Ther.,* **2000**, *86*, 201.

276. Milroy, L.-G.; Grossmann, T. N.; Hennig, S.; Brunsveld, L.; Ottmann, C. *Chem. Rev.,* **2014**, *114*, 4695.

277. Surade, S.; Blundell, T. L. *Chem. Biol.,* **2012**, *19*, 42.

278. Phan, T.; Nguyen, H. D.; Göksel, H.; Möcklinghoff, S.; Brunsveld, L. *Chem. Commun.,* **2010**, *46*, 8207.

279. Musacchio, A.; Gibson, T.; Lehto, V.-P.; Saraste, M. *FEBS Lett.,* **1992**, *307*, 55.

280. Mayer, B. J. *J. Cell Sci.,* **2001**, *114*, 1253.

281. Cobos, E. S.; Pisabarro, M. T.; Vega, M. C.; Lacroix, E.; Serrano, L.; Ruiz-Sanz, J.; Martinez, J. C. *J. Mol. Biol.,* **2004**, *342*, 355.

282. Cicchetti, P.; Mayer, B. J.; Thiel, G.; Baltimore, D. *Science,* **1992**, *257*, 803.

283. Feng, S.; Chen, J. K.; Yu, H.; Simon, J. A.; Schreiber, S. L. *Science,* **1994**, *266*, 1241.

284. Lim, W. A.; Richards, F. M.; Fox, R. O. *Nature,* **1994**, *372*, 375.

285. Musacchio, A.; Saraste, M.; Wilmanns, M. *Nat. Struct. Mol. Biol.,* **1994**, *1*, 546.

286. Ren, R.; Mayer, B. J.; Cicchetti, P.; Baltimore, D. *Science,* **1993**, *259*, 1157.

287. Dalgarno, D.; Botfield, M.; Rickles, R. *J. Pept. Sci.,* **1997**, *43*, 383.

288. Bartelt, R. R.; Light, J.; Vacaflores, A.; Butcher, A.; Pandian, M.; Nash, P.; Houtman, J. C. *Biochim. Biophys. Acta,* **2015**, *1853*, 2560.

289. Nash, P., SH3 Domains. In *Reference Module in Life Sciences*, Elsevier: 2017.

290. Ball, L. J.; Jarchau, T.; Oschkinat, H.; Walter, U. *FEBS Lett.,* **2002**, *513*, 45.

291. Su, J. Y.; Hodges, R. S.; Kay, C. M. *Biochemistry,* **1994**, *33*, 15501−15510.

292. Litowski, J.; Hodges, R. *J. Pep. Res.,* **2001**, *58*, 477−492.

293. Zorzi, A.; Deyle, K.; Heinis, C. *Curr. Opin. Chem. Biol.,* **2017**, *38*, 24.

294. Zhou, X.; Po, A. L. W. *Int. J. Pharm.,* **1991**, *75*, 97.

295. Baeriswyl, V.; Heinis, C. *Protein Eng., Des. Sel.,* **2012**, *26*, 81.

296. White, C. J.; Yudin, A. K. *Nat. Chem.,* **2011**, *3*, 509.

297. Bock, V. D.; Perciaccante, R.; Jansen, T. P.; Hiemstra, H.; van Maarseveen, J. H. *Org. Lett.,* **2006**, *8*, 919.

298. Turner, R. A.; Oliver, A. G.; Lokey, R. S. *Org. Lett.,* **2007**, *9*, 5011.

299. Lécaillon, J.; Gilles, P.; Subra, G.; Martinez, J.; Amblard, M. *Tetrahedron Lett.,* **2008**, *49*, 4674.

300. Meutermans, W. D.; Golding, S. W.; Bourne, G. T.; Miranda, L. P.; Dooley, M. J.; Alewood, P. F.; Smythe, M. L. *J. Am. Chem. Soc.,* **1999**, *121*, 9790.

301. Miller, S. J.; Blackwell, H. E.; Grubbs, R. H. *J. Am. Chem. Soc.,* **1996**, *118*, 9606.

302. Chapman, R. N.; Dimartino, G.; Arora, P. S. *J. Am. Chem. Soc.,* **2004**, *126*, 12252.

303. Rohrbacher, F.; Deniau, G.; Luther, A.; Bode, J. W. *Chem. Sci.,* **2015**, *6*, 4889.

304. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. *Science,* **1994**, *266*, 776.

305. Clark, R. J.; Craik, D. J. *J. Pept. Sci.,* **2010**, *94*, 414.

306. Yan, L. Z.; Dawson, P. E. *J. Am. Chem. Soc.,* **2001**, *123*, 526.

307. Camarero, J. A.; Cairó, J. J.; Giralt, E.; Andreu, D. *J. Pept. Sci.,* **1995**, *1*, 241.

308. Antos, J. M.; Truttmann, M. C.; Ploegh, H. L. *Curr. Opin. Struct. Biol.,* **2016**, *38*, 111.

309. Thongyoo, P.; Roqué-Rosell, N.; Leatherbarrow, R. J.; Tate, E. W. *Org. Biomol. Chem.,* **2008**, *6*, 1462.

310. Nguyen, G. K.; Wang, S.; Qiu, Y.; Hemu, X.; Lian, Y.; Tam, J. P. *Nat. Chem. Biol.,* **2014**, *10*, 732.

311. Yang, R.; Wong, Y. H.; Nguyen, G. K.; Tam, J. P.; Lescar, J.; Wu, B. *J. Am. Chem. Soc.,* **2017**, *139*, 5351.

312. Nuijens, T.; Toplak, A.; Quaedflieg, P. J. L. M.; Drenth, J.; Wu, B.; Janssen, D. B. *Adv. Synth. Catal.,* **2016**, *358*, 4041.

313. Toplak, A.; Nuijens, T.; Quaedflieg, P. J. L. M.; Wu, B.; Janssen, D. B. *Adv. Synth. Catal.,* **2016**, *358*, 2140.

314. Schmidt, M.; Toplak, A.; Quaedflieg, P. J. L. M.; Ippel, H.; Richelle, G. J. J.; Hackeng, T. M.; van Maarseveen, J. H.; Nuijens, T. *Adv. Synth. Catal.,* **2017**, *359*, 2050.

315. Gruber, M.; Lupas, A. N. *Trends Biochem. Sci,* **2003**, *28*, 679.

316. Brown, J. H.; Cohen, C.; Parry, D. A. *Proteins: Struct., Funct., Bioinf.,* **1996**, *26*, 134.

317. Read, R. J.; Adams, P. D.; Arendall, W. B., 3rd; Brunger, A. T.; Emsley, P.; Joosten, R. P.; Kleywegt, G. J.; Krissinel, E. B.; Lutteke, T.; Otwinowski, Z.; Perrakis, A.; Richardson, J. S.; Sheffler, W. H.; Smith, J. L.; Tickle, I. J.; Vriend, G.; Zwart, P. H. *Structure,* **2011**, *19*, 1395.

318. Chen, V. B.; Arendall, W. B., 3rd; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *Acta Crystallogr. Sect. D. Biol. Crystallogr.,* **2010**, *66*, 12.

319. Wang, G.; Dunbrack, R. L. *Bioinformatics,* **2003**, *19*, 1589.

320. Kumar, P.; Bansal, M. *Acta Crystallogr. Sect. D. Biol. Crystallogr.,* **2015**, *71*, 1077.

321. Kabsch, W.; Sander, C. *Biopolymers,* **1983**, *22*, 2577.

322. Frishman, D.; Argos, P. *Proteins,* **1995**, *23*, 566.

323. Mansiaux, Y.; Joseph, A. P.; Gelly, J. C.; de Brevern, A. G. *PLoS One,* **2011**, *6*, e18401.

324. Lapenta, F.; Aupič, J.; Strmšek, Ž.; Jerala, R. *Chem. Soc. Rev.,* **2018**, *47*, 3530.

325. Hong, F.; Zhang, F.; Liu, Y.; Yan, H. *Chem. Rev.,* **2017**, *117*, 12584.

326. Huang, P.-S.; Feldmeier, K.; Parmeggiani, F.; Velasco, D. A. F.; Höcker, B.; Baker, D. *Nat. Chem. Biol.,* **2016**, *12*, 29.