Author:
Edgell, Caitlin L

Title:
De novo designed protein-protein interaction domains for synthetic biology applications in cells

De novo designed protein-protein interaction domains for synthetic biology applications in cells

Caitlin L. Edgell

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

School of Biochemistry, October 2018

Word count: 81355
Excluding references: 69414

Abstract

Synthetic biology aims to make the realisation of novel biological phenomena easier, faster and more precise through the design and engineering of biological components. This can be attempted using “top-down” approaches, by engineering the genomes of extant organisms to make them behave in new ways. Alternatively, it can be performed in a “bottom-up” fashion, by designing the characteristics of individual molecules and combining them in increasingly complex systems. The latter allows for the design of entirely new behaviours through the construction and assembly of components that are not observed in nature. Important design targets for this bottom-up approach are components that temporally and spatially control the expression of genes, that is, transcription factors. These are the targets of the work presented in this thesis.

Transcription factors are ubiquitous and highly diverse. They are also the first point of control for the production of all other cellular components, both natural and designed. In prokaryotes, transcriptional control can be achieved by actively recruiting RNA polymerase to a gene of interest *via* protein-protein interactions or preventing RNA polymerase from binding to DNA by assembling oligomers at a promoter. Therefore, in their simplest forms, both transcriptional activators and repressors require (1) the ability to bind to DNA and (2) the ability to bind to other proteins. By combining protein domains that perform these functions, entirely artificial transcription factors can be generated. In the past, this has been achieved by combining naturally occurring protein domains. However, advances in rational protein design mean we can now design these domains *de novo*.

Herein, a set of *de novo* tetrameric coiled coil-based protein-protein interaction domains is rationally designed and characterised *in vitro*. These coiled coils are then used to direct protein-protein interactions in *E. coli*, demonstrated through the design of semi-artificial transcription factors, which can themselves be controlled at the transcriptional level. In the future, these constitutively interacting domains could be made even more useful through the incorporation of dynamic behaviours, such as ligand-induced conformational changes.

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: .....

Acknowledgements

Firstly, thank you to my supervisors, Dek and Nigel. I have loved pursuing my PhD and it has been a pleasure to work with both of you. You make a great team; long may it continue.

I have also been very lucky to belong to two excellent labs during my PhD: W520 in Chemistry and D40 in Biochemistry. Thank you to all past and present members of both labs for your advice, encouragement and conviviality. I want to especially thank Drew for helping me get started with my project and Abby, Caroline, Emily and Guto for their help proofreading this thesis.

Thank you to all the staff involved with the SynBio CDT, especially Claire, and to the BBSRC and EPSRC for their generous funding. It truly is something special you've created. Thank you also to the 2014 cohort. Our year in Oxford was an important formative time for me as a scientist (and a human being) and it was an honour to be part of such a talented, inspiring and friendly group of people.

Thank you Emily for always taking an interest in my work. I know we agree on the importance of a good work ethic. Thank you for inspiring me to strive. Thanks also to Judd for letting me stay in your house while I was writing, and for cooking me delicious vegetarian food. Finally, thank you to my family: thank you Thom for always reminding me to maintain a good sense of humour; thank you Dad for reminding me to work hard; and thank you Mum for reminding me not to work *too* hard.

List of Abbreviations

α-CHCA	α-Cyano-4-hydroxycinnamic acid matrix
ANS	8-Anilino-1-naphthalenesulfonic acid
ATF	Artificial transcription factor
AUC	Analytical ultracentrifugation
BUDE	Bristol University Docking Engine
CD	Circular dichroism
DBD	DNA-binding domain
DMF	N,N-Dimethylformamide
DMSO	Dimethyl sulfoxide
DPH	1,6-Diphenyl-1,3,5-hexatriene
f/f_0	Frictional ratio
Fmoc	Fluorenylmethyloxycarbonyl
GFP	Green fluorescent protein
HLH	Helix-loop-helix
HPLC	High-performance liquid chromatography
HTH	Helix-turn-helix
MALDI-TOF	Matrix-assisted laser desorption/ionisation time-of-flight
MRE	Mean residue ellipticity
Mw	Molecular weight
NMR	Nuclear magnetic resonance
OD	Optical density
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank
PID	Protein-protein interaction domain
RNAP	RNA polymerase
RT	Room temperature
s	Sedimentation coefficient
$s_{(20,w)}$	s normalised to standard conditions (water, 20 °C)
s.d.	Standard deviation
SE	Sedimentation equilibrium
SGC	Synthetic genetic circuit

SUMO	Small ubiquitin-related modifier
SV	Sedimentation velocity
TALE	Transcription activator-like effector
TFA	Trifluoroacetic acid
T_M	Thermal midpoint of unfolding
UV	Ultraviolet
$\bar{v}$	Partial specific volume
WT	Wild type
ZF	Zinc finger

Amino acids: Alanine, Ala, A; Arginine, Arg, R; Asparagine, Asn, N; Aspartate, 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 Synthetic biology

Synthetic biology aims to make the design of biological components and behaviours readily achievable and routine. The remit of synthetic biology is broad, encompassing the design of previously unseen biological molecules, the engineering of a cell's metabolome and the construction of orchestrated communities of organisms [1-3]. What these seemingly disparate ventures have in common is the desire (1) to generate useful products with real world applications, and (2) to use these products as tools to expand our understanding of both natural and synthetic systems.

Synthetic biology products can be considered as falling roughly into two categories (Figure 1-1) [4,5]. The first includes all products that have been derived or modified from naturally occurring components, including proteins or whole cells. The second covers products that have been designed *de novo* (from scratch). While the latter may aim to recreate natural structures or functions, the use of components with a non-natural origin allows the innate complexity that is present in most naturally occurring components to be minimised. Approaches taken in this category include *de novo* protein design, xenobiology and protocell design [3,6,7].

Naturally, there is overlap between the two categories. For example, natural proteins can be fused to *de novo* proteins to make semi-synthetic products, and proteins with non-natural structures or functions can be made using only the 20 canonical amino acids. Additionally, *de novo* components can be introduced into a naturally occurring organism to augment biology by bestowing novel behaviours on that organism.

1.1.1 Engineering approaches

Approaches that involve the modification of naturally occurring biological parts, from proteins to whole cells, fall under the remit of "top-down" synthetic biology [5]. Genetic engineering can be used to alter the properties and functions of well-

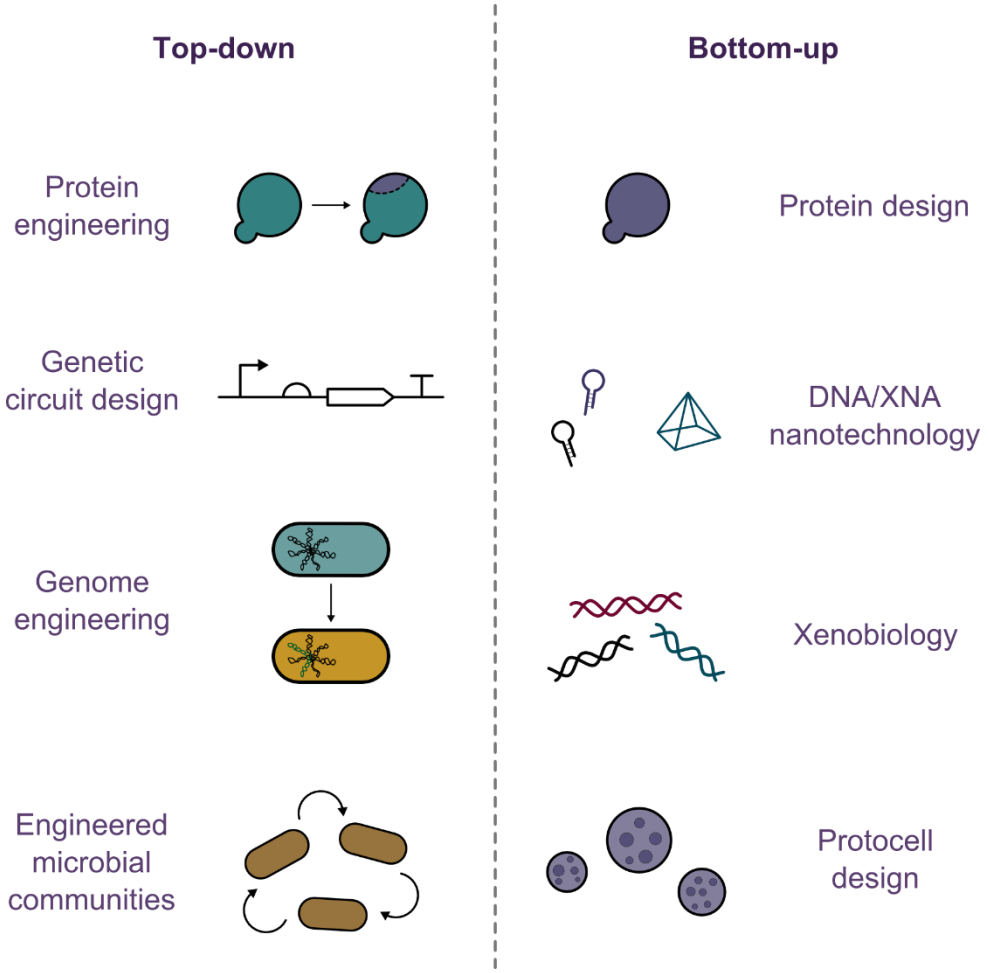


Figure 1-1 Summary of synthetic biology approaches. Left, top-down engineering approaches that involve modifying naturally occurring parts, including protein engineering, genetic circuit design, genome engineering and the engineering of microbial communities. Right, bottom-up *de novo* design approaches that involve using natural and non-natural building blocks to make new components and systems, including protein design, DNA nanotechnology, xenobiology and protocell biology.

characterised natural components. DNA sequencing and synthesis and genome editing technologies like CRISPR/Cas9 have greatly facilitated these endeavours.

Key examples of whole-cell engineering include ambitious projects such as: the chemical synthesis and transplantation of an entire bacterial genome to make JCVI-syn1.0, a version of *Mycoplasma mycoides* [8], as well as a minimal version of this genome [9]; recoding the genome of *Escherichia coli* to generate strains that use a reduced set of codons, freeing up the remaining codons for the introduction of non-standard amino acids [10,11]; and the Yeast 2.0 project, which aims to synthesise a version the genome of *Saccharomyces cerevisiae* [12-14]. Such projects have accelerated the development of tools for DNA synthesis, assembly and editing.

Other examples of biological engineering involve smaller genetic changes that still have profound effects on the behaviour of organisms. This includes metabolic engineering [2] and the design of synthetic genetic circuits [15]. In metabolic engineering, existing enzymatic pathways, and the mechanisms that regulate them, are modified to increase their productivity or alter their activity. This can involve: altering the substrate specificity of metabolic pathways, such that organisms can utilise alternate feedstocks [16-19]; redirecting enzymatic pathways to synthesise different products [20]; increasing the productivity of metabolic pathways [21]; and introducing heterologous pathways into industrially relevant hosts [22,23]. Conversely, in synthetic genetic circuit (SGC) design, biological parts are combined in new pathways to make circuits that impart various computer-like behaviours on living things, including multi-stable switching [24], logic-gating [25], enumeration [26], temporal oscillation [27] and light detection [28].

Finally, naturally occurring proteins can be modified to alter their function. For example, inducible transcription factors can be modified to respond to new ligands [29] and enzymes can be engineered to have altered substrate specificity [30]. Proteins can also be modified to have altered properties such as increased thermal stability [31,32] or to be simpler by using a reduced amino acid palette [33]. These engineered parts provide insight into protein sequence-to-structure relationships and also generate new components that can be used in metabolic engineering and SGC design.

1.1.2 De novo *approaches: biology from scratch*

The *de novo* design of biological components is a “bottom-up” approach to synthetic biology. Individual non-natural components, once realised, could be combined in increasingly complex systems until, in theory, something resembling a living cell is produced [4].

These *de novo* components can also be implemented inside existing organisms. This approach offers benefits over simply using naturally occurring biomolecules for the following reasons: (1) designed components, such as proteins, can be made much simpler than natural ones; (2) *de novo* components can be designed for orthogonality (*i.e*. with limited cross-interaction with natural hosts or systems); (3) designed components should be well understood so should respond predictably to modifications to the component or its environment; and (4) *de novo* components could be designed to perform activities that are not observed in nature.

Nucleic acids have been the focus of extensive study in synthetic biology and nanotechnology due to the simple Watson-Crick base pairing that dictates DNA and RNA interactions. As well as being vital to their biological functions these simple rules that dictate interactions between nucleic acids have been exploited in DNA origami and DNA nanotechnology to introduce new structures and functions into DNA [34-37]. Furthermore, xeno nucleic acids (XNAs) have been able to reproduce both natural and artificial functions of nucleic acids, and have been used to encode heritable genetic information [38], as catalysts [39] and in nanotechnology [40].

Further key advances in biomolecular design have occurred in the area of protein design. New protein structures and functions that have not been observed in nature can be designed using rational and computational approaches [41-44]. Using just the 20 proteinogenic amino acids means designed proteins can be readily introduced into organisms. However, the use of non-standard amino acids expands the chemical functionality of proteins greatly. For example, the site-specific incorporation of fluorescent amino acids has been an especially useful tool for the study of protein structure and function [45,46]. The development of codon suppression technologies, where mRNA codons, tRNAs and aminoacyl-tRNA synthetases are co-opted to incorporate non-standard amino acids during *in vivo* ribosomal synthesis [47-49], has facilitated the design of proteins that contain multiple non-standard amino acids [50]. At the extreme of this are foldamers, non-biological polymers that adopt specific structures but are made primarily of non-proteinogenic monomers, such as the α-branched amino acid 2-aminoisobutyric acid [51]. However, such polymers can generally only be made using synthetic methods.

While a truly *de novo* living cell is certainly a long way from being delivered, rationally designed components still represent useful tools for modifying the behaviour of existing organisms. SGC design is an area that could particularly benefit from the introduction of new components. Currently, most of the parts used in SGC design are taken from nature and, while nature contains a wealth of the types of component used in SGC design (*e.g.* transcription factors, promoters, reporter genes), only a small number of these are adequately characterised to be routinely used. There have been efforts to broaden the array of available parts by mining components, such as transcription factors, from natural genomes [52]. However, such parts must be characterised before they can be reliably implemented in SGCs. Therefore, it would be desirable to be able to routinely design components for specific tasks in SGCs as they arise. Transcription factors

represent a particularly useful design target given that they are powerful regulators of other cellular components.

1.2 Transcriptional regulation

Transcription is the process whereby specific sections of DNA are copied into separate RNA molecules by an RNA polymerase (RNAP). The RNA molecules can then go on to perform various cellular functions. For example, messenger RNAs (mRNAs) are translated into proteins by the ribosome. The correct spatial and temporal regulation of transcription is vital for the survival of an organism. Additional components, transcription factors, allow an organism to up- or down-regulate expression of specific genes in response to internal or external cues, altering the organism's behaviour accordingly.

Transcription in both eukaryotes and prokaryotes involves three steps – initiation, elongation and termination – all of which can be targets for regulating the expression level of a gene. However, the focus here will be mechanisms for the control of transcription initiation in bacteria.

1.2.1 Bacterial transcription

1.2.1.1 Bacterial promoters and RNA polymerase

In prokaryotes, all transcription is performed by a single type of RNAP. Each RNAP holoenzyme is made up of a number of different subunits; *E. coli* has a subunit composition of $\alpha_2\beta\beta'\omega\sigma$ (Figure 1-2) [53,54]. The RNAP core enzyme ($\alpha_2\beta\beta'\omega$ in *E. coli*) is a processive DNA-dependent RNA polymerase that synthesises RNA in a 5′ to 3′ direction.

The RNAP holoenzyme binds bacterial promoters at various regions, including the critical -10 and -35 boxes [55]. The -10 boxes are generally AT-rich to aid DNA melting during initiation. The exact sequences of the -10 and -35 regions influence the rate of transcription initiation by affecting the rate of strand unwinding and the strength of RNAP binding. Bacterial promoters may contain additional Upstream (UP) elements, AT-rich regions 40-60 bp upstream of the transcription start site that are also contacted by components of RNAP [55]. Finally, bacterial promoters may also contain other sequence elements including binding sites for transcription factors.

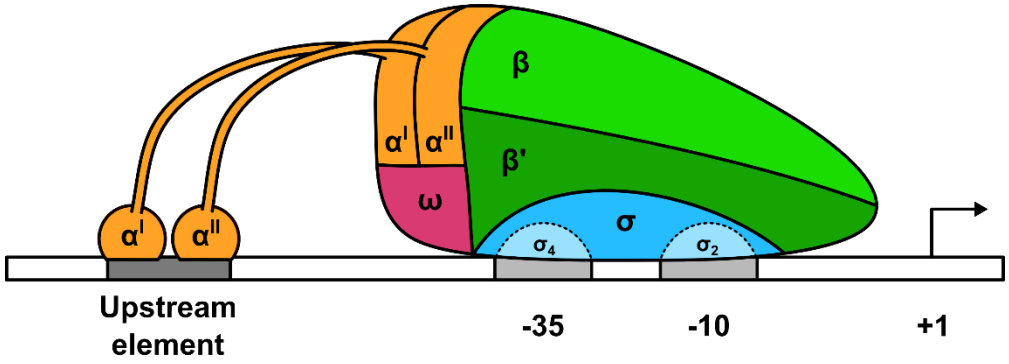


Figure 1-2 Schematic of *E. coli* RNA polymerase holoenzyme. The β and β′ subunits make up the active site; the α dimer facilitates RNAP assembly and contacts Upstream elements to activate transcription initiation; the ω subunit stabilises the β′ subunit; the σ subunit is responsible for promoter selection and binds the -10 and -35 boxes of the promoter (a σ^{70} family σ factor is shown). +1 is the transcription start site.

The active site of RNAP is made up of the β and β′ subunits, which form a "crab claw" shape [56]. A number of channels exist between the core of the claw and the protein surface: the primary channel, which accommodates downstream double-stranded (ds) DNA and the nascent RNA-DNA hybrid; the secondary channel, which allows nucleotide triphosphate entry; and the RNA exit channel. Catalysis occurs in the primary channel, at the base of the claw, and requires two Mg^{2+} ions coordinated *via* evolutionarily conserved residues [57].

A dimer of α subunits is located on the outside base of the claw [58]. Through interactions with the β and β′ subunits *via* their N-terminal domains (α-NTDs), the dimer has a role in initiating RNAP assembly [59,60]. The α subunits can also interact with promoter UP elements *via* their C-terminal domains (α-CTDs). This has a role in activating transcription initiation [61-64]. The α-CTDs also interact with transcription factors such as the catabolite activator protein (CAP) [65]. The α-NTDs and α-CTDs are joined by flexible linkers [66,67].

The ω subunit interacts predominantly with the β′ subunit [68] and has various roles including maintaining the structure and function of this subunit and in the assembly of the RNAP core enzyme [69-71].

Finally, the σ subunit, or σ factor, is responsible for promoter selection and is required for transcription initiation [72]. Unlike the other RNAP subunits, there are multiple different σ factors, which are active under different conditions; in *E. coli* there are seven σ factors [73]. The various σ factors have different promoter specificities and the RNAP holoenzyme is directed to specific genes depending on

the identity of its σ factor. The σ factors can be classified into two groups depending on their structure: the σ^{70} family, named for the *E. coli* "housekeeping" σ factor [74]; and the σ^{54} family [75]. The σ^{70} family members bind promoter DNA at the -10 and -35 hexamers *via* the σ_2 and σ_4 domains, respectively [74,76,77]. Conversely, the σ^{54} family members bind to a different type of promoter with conserved -12 and -24 regions [75].

1.2.1.2 The stages of transcription: initiation, elongation and termination

In all domains of life, transcription proceeds along three main steps as outlined briefly below [77,78]. The first step, initiation, is further split into a number of sub-steps (Figure 1-3) [79]. Initially, the RNAP holoenzyme binds to DNA non-specifically and undergoes promoter search [80,81]. On encountering a suitable promoter, RNAP binds tightly to the DNA *via* the σ subunit to form the closed promoter complex, RP_c. Additional contacts may be made to the DNA UP elements by the α subunits [61-63]. Regions of σ also bind to the active site and to the RNA exit channel, occluding them both [82]. The complex then undergoes isomerisation to form the open promoter complex, RP_o. This involves stabilisation of the unwound state of the AT-rich -10 box by the σ_2 domain. This leads to the formation of an open transcription bubble of 12-15 nucleotides (nt), which extends past the transcription start site (+1) [83]. Further conformational changes also remove σ from the active site, which is then occupied by single-stranded (ss) DNA. The RP_o complex then performs multiple rounds of abortive initiation where short RNA molecules are repeatedly synthesised [84]. Here, RNAP is bound to upstream promoter DNA while also pulling downstream DNA into the enzyme active site, resulting in DNA scrunching. The strain from the scrunched DNA and the continued occlusion of the RNA exit channel result in the dissociation of the abortive transcripts. Eventually, a long enough RNA molecule is made (11-15 nt) to expel σ from the RNA exit channel [82]. Further conformational changes occur in σ that weaken its affinity for both DNA and the RNAP core enzyme and it ultimately dissociates, forming a highly processive elongation complex (EC) that escapes the promoter and continues transcription elongation [77]. The RNAP core enzyme then progresses along the DNA. Elongation does not proceed at a uniform rate and the EC may pause. This can be a point of transcriptional control [85,86].

Transcription is completed by termination. In bacteria, two methods exist for termination at the end of genes [86]. The first, Rho-dependent termination, involves an ATP-dependent helicase, Rho [87,88]. This binds to *rut* sites in the nascent RNA

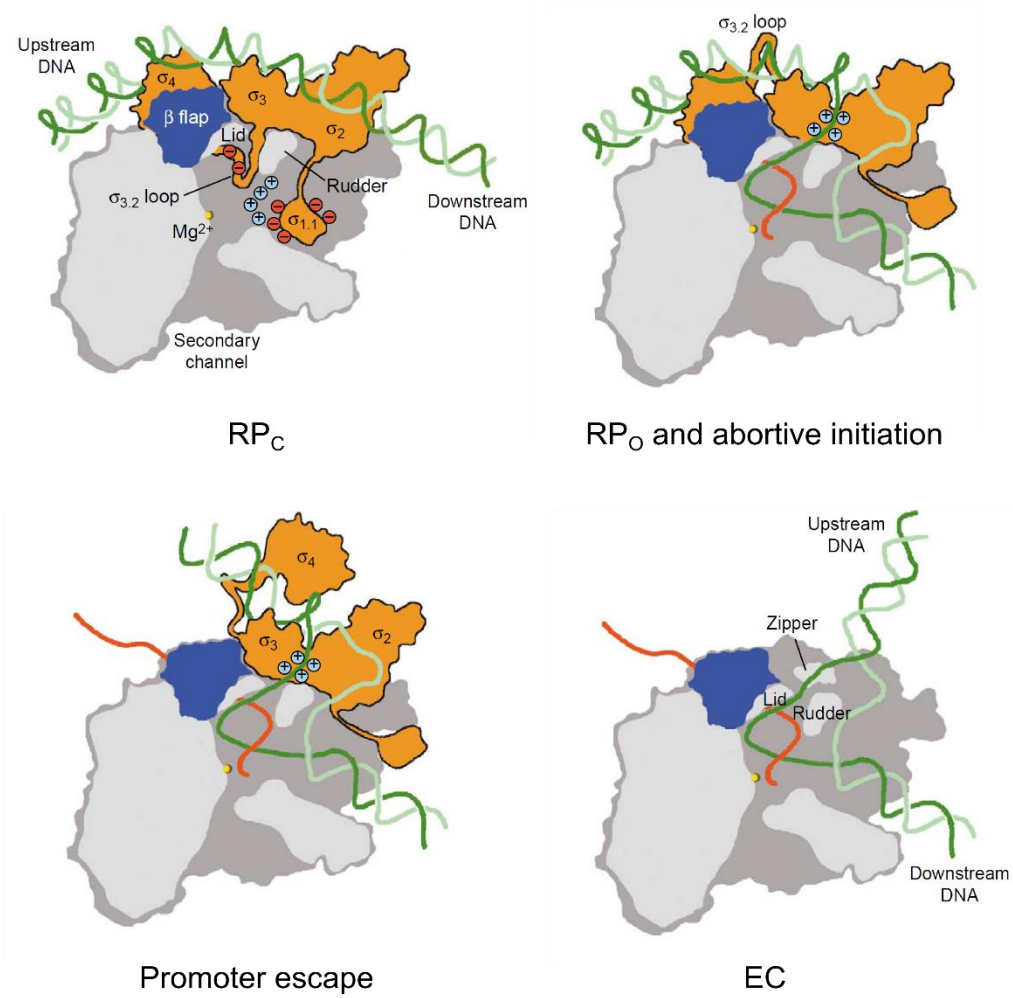


Figure 1-3 Stages of transcription initiation with the bacterial RNAP holoenzyme. Figure adapted from reference [77].

and then translocates along the RNA towards RNAP until it reaches a release site, whereupon RNAP is dissociated from the RNA. The second type of termination, Rho-independent or intrinsic termination, involves inverted repeats in the transcript that form an RNA hairpin loop, followed by a poly-U stretch [89,90]. The loop causes RNAP to pause and the weakly associated A:U hybrid falls apart, releasing RNAP.

1.2.1.3 Control mechanisms: alternative σ factors and transcription factors

Organisms can alter their transcriptional profiles to adapt to changing environmental conditions. Bacteria can achieve this in many ways, including through the use of alternative σ factors and transcription factors.

As introduced above, σ factors are the part of the RNAP holoenzyme responsible for promoter selection. In *E. coli* the housekeeping σ factor, σ^{70}, is the predominant σ factor active during logarithmic growth [73]. However, under different

circumstances, such as stress conditions, six alternative σ factors may become dominant [73]. The different σ factors are responsible for regulating the expression of multiple genes, which are expressed from related promoters [91]. This allows for widespread transcriptional changes using a relatively simple shared mechanism. For example, σ^{32} is the heat shock σ factor, which becomes active at elevated temperatures and helps the cell recover from the potential heat damage [92-94]. In the σ^{32} mRNA, the start codon is occluded within a stem and loop structure [95]. Elevated temperatures melt this RNA secondary structure allowing σ^{32} to be translated, leading to an increase in the level of this protein, which can then compete with the other σ factors for RNAP-binding [96,97]. The σ^{32} RNAP holoenzyme then initiates transcription at relevant promoters, leading to an increase in the expression of chaperones, proteases and other proteins that help the cell recover on returning to its optimal growth temperature [98].

The engineering of σ factors has been proposed as a means of generating new, orthogonal transcriptional regulators for synthetic biology [99]. For example, σ factors from *Bacillus subtilis* can be used for orthogonal gene expression in *E. coli* [100]. Additionally, mining of bacterial genomes for simple, two-domain σ sigma factors yielded a set of 20 σ factors with associated promoters that did not show any cross talk with each other [101]. The DNA binding domains of these σ factors could also be mixed and matched to generate new σ factors with new DNA-binding specificities [101].

Beyond alternative σ factors, bacteria possess a number of transcription factors that act to increase or decrease the expression level of a given gene above or below the basal level set by the σ factor and its binding to the -10 and -35 boxes. These transcription factors are classed as activators and repressors, respectively and they largely act at the various sub-steps of initiation. *E. coli* is predicted to have 300-350 different transcriptional regulators [102]. Many of these, such as CAP, regulate multiple genes while a small number act at a single promoter [103,104]. The latter group includes the model bacterial transcription factor, the Lac repressor. Transcription factors are themselves usually controllable, for example *via* small-molecule binding [105]. This allows an organism to sense features of its environment, such as metabolite availability, and change its transcriptional profile accordingly.

CAP is a transcriptional activator that acts at two different classes of promoter, class I and class II, in which the CAP-binding sites are located at different positions [106]. CAP uses a different mode of action at each class of promoter. In class I

promoters, CAP binds the DNA upstream of the -10 and -35 boxes and contacts the α-CTD of RNAP [65]. The additional contacts stabilise RNAP at the promoter in the closed conformation. In class II promoters, CAP binds the DNA at a site overlapping the -35 box and makes more extensive interactions with RNAP, binding the α-CTD, the α-NTD and the σ factor [107,108]. As well as initially stabilising RNAP at the promoter in the closed promoter complex, this also promotes isomerisation of RNAP to form the open promoter complex [106]. CAP is involved in the regulation of almost 200 genes and it is itself regulated by cyclic AMP (cAMP), only binding its specific DNA sites when complexed with cAMP [103,106]. Thus, *via* CAP, the availability of nutrients has a large effect on the metabolic profile of bacteria.

This recruitment/stabilisation method for transcription activation has been relatively simple to mimic with artificial systems. For example, introducing a protein-protein interaction between the α or the ω subunit of RNAP and a heterologous DNA-binding domain leads to an increase in expression of genes whose promoters contain the corresponding DNA sequence as the mutant RNAP is recruited to and stabilised at those promoters [109,110]. The recruitment/stabilisation method has also been used where the DNA-binding domain and RNAP are linked *via* a designed protein-protein interaction domain [111]. In these systems, the strength of activation corresponds to the strength of the protein-protein interaction.

Other methods for activating transcription in prokaryotes include distorting non-optimal promoters such that RNAP can bind to them more efficiently. For example the MerR family of activators bind at promoters with non-optimal -10 to -35 spacers, reconfiguring them so that RNAP can bind [112]. However, such modes of transcription activation are likely to be much more difficult to recreate.

The Lac repressor, LacI, is a transcriptional repressor that regulates transcription of the genes in the *lac* operon only [113]. Repression is caused by LacI binding an operator that overlaps the RNAP binding site [114]. This prevents RNAP from accessing the DNA, thus preventing the formation of the closed promoter complex and subsequent transcription initiation. However, LacI in fact binds two operators simultaneously [114]. This leads to DNA looping, which may occlude RNAP binding further and also increases the local concentration of LacI at the promoter [115,116]. Overall, this increases the achievable level of repression. LacI also responds to small molecules: binding of allolactose prevents LacI from binding strongly to operator DNA [117,118]. The LexA repressor also acts by blocking the formation of the

closed promoter complex and has a binding site that overlaps the -35 box [119]. Other methods for decreasing the level of transcription involve inhibiting open promoter complex formation [120] or inhibiting promoter escape after open promoter complex formation [121].

The mechanism of promoter occlusion has also been relatively simple to mimic. For example, a transcription activator-like effector designed to bind to the *lac* operon O_1 sequence has been found to repress transcription more strongly than wild type LacI when only the O_1 operator is present [122].

1.2.2 Artificial transcription factors

As introduced above, nature contains a wealth of transcription factors that operate *via* multiple different mechanisms. However, it is desirable to expand the toolkit of transcriptional regulators beyond those that exist in nature by making new, artificial transcription factors. This is because, as outlined in Section 1.1.2, such components can be simpler, more predictable, more orthogonal to natural systems and may be designed to perform actions not observed in nature.

Artificial transcription factors (ATFs) have diverse applications in synthetic biology, genomics and medicine. In synthetic biology, ATFs are useful in the construction of synthetic genetic circuits to introduce unnatural functions or behaviours into cells and organisms. Those designed to respond to ligands or physical stimuli are useful as biosensors for the detection of small molecules or other environmental changes [123].

Furthermore, ATFs have provided an alternative approach to genomic studies: ATFs designed to be strong repressors can achieve considerable gene knock-down, especially when combined with gene silencing technologies such as RNAi [124]. This use of ATFs to alter transcriptional profiles has also been used in “phenotypic engineering" to create variants of organisms with desired characteristics in a temporary and reversible manner, in contrast to more permanent genetic manipulations [125,126]. Finally, ATFs have been proposed as therapeutics for diseases involving aberrant or undesirable transcription such as cancer [127,128] and viral infections [129].

The simplest step in transcription to target with ATFs is the formation of the closed promoter complex. This is because the mechanisms to promote or prevent this step involve simple protein-protein and protein-DNA interactions: in activation, a

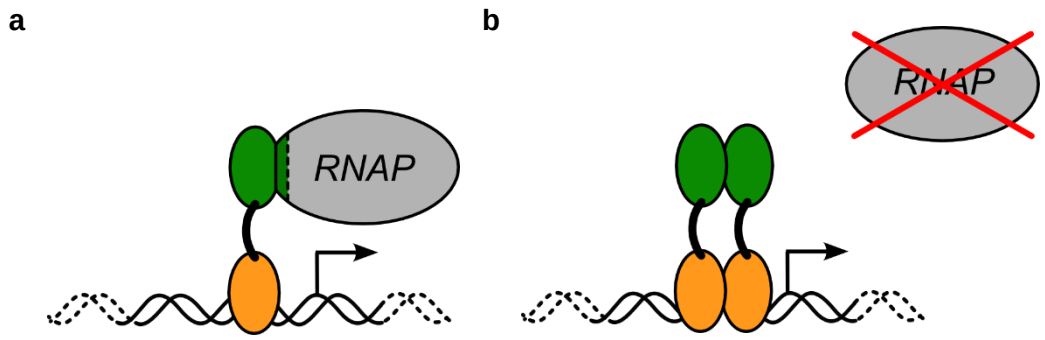


Figure 1-4 Schematics for transcription activation and repression by modular ATFs. (a) Transcription can be activated by recruiting a (modified) RNAP to a promoter. (b) Transcription can be repressed by oligomeric DNA-binding proteins that occlude the RNAP-binding site. Green ovals, a protein-protein interaction domain; orange ovals, a DNA-binding domain.

DNA-binding protein can recruit RNAP to a specific location in the DNA by interacting with one of its subunits (Figure 1-4a); in repression, oligomeric DNA-binding proteins can form stable complexes on the DNA, blocking RNAP binding (Figure 1-4b). Therefore, by combining protein-protein interaction domains and DNA-binding domains, it is possible to design new, modular transcription factors. Previously, this has been achieved using naturally occurring domains. However, thanks to advances in protein engineering and rational protein design, it is now possible to design these domains from scratch to make fully synthetic ATFs. Examples of ATFs containing natural, engineered and designed domains are discussed below.

1.2.2.1 DNA-binding domains (DBDs)

Many naturally occurring DBDs have been repurposed in ATFs. For example, cI from bacteriophage λ has been used to make artificial transcriptional activators [109,110]. DBDs from bacterial proteins have also been used: the DBD of LexA has been used to make a transcriptional repressor and the CadC DBD has been used to make a transcriptional activator [130].

The use of naturally occurring DBDs limits the use of ATFs to promoters that contain specific DNA sequences. It also increases the risk of cross talk with other promoters in natural hosts. While it is possible to alter the DNA-binding specificity of some DBDs [131,132], it is preferable to be able to design these proteins, and therefore the DNA sequences they bind to, from scratch. There are proteins, such as zinc-finger arrays and transcription activator-like effectors, that can be engineered in this way to bind to any DNA sequence.

Zinc fingers (ZFs) are widespread protein-folding motifs involved in DNA recognition (Figure 1-5a). The most widely studied and engineered type of ZFs are the Cys_2His_2 ZFs, in which a zinc ion is bound by two cysteines and two histidines [133,134]. ZF motifs can be arranged as tandem arrays in a single protein chain. Originally, ZF arrays with novel DNA-binding specificities were selected for using phage display [135-137]. This led to the discovery that (1) these proteins bind DNA in a modular fashion where each ZF motif recognises three bases of DNA, and (2) different ZF modules recognise different DNA triplets [138]. The DNA-binding code was deciphered and new ZF arrays were assembled by combining multiple triplet-recognising modules [139-141]. While it has since been found that these DNA-binding proteins are not as modular and designable as first believed [142], they have still found use in ATFs. For example, ZF arrays have been used to make transcriptional activators that recruit T7 polymerase to specific promoters [111]. They have also been fused to CAP to make transcriptional activators [126]. Additionally, ZF arrays have been used widely in designing mammalian ATFs [129,143].

Transcription activator-like effectors (TALEs) were first identified in plant pathogenic bacteria from the genus *Xanthomonas* where they act as transcriptional activators that facilitate bacterial infection [144]. TALEs consist of a variable number of 34 residue long tandem repeats. The repeats differ primarily at two key residues, 12 and 13, which are collectively called the repeat variable diresidue (RVD) [145]. Each repeat adopts a helix-loop-helix structure where the RVD is located within the loop and the helix-loop-helix repeats associate to form a superhelix that wraps around the DNA major groove (Figure 1-5b) [146,147]. The RVDs determine the DNA-binding specificity of the TALE but only residue 13 directly contacts the DNA. Each RVD is responsible for recognising one DNA base and the rules that dictate which amino acids bind to which nucleotides are known. This allows for the assembly of tandem repeats containing different RVDs to create TALEs that can bind to a desired DNA sequence [148,149]. TALEs are particularly promising as DBDs in ATFs as they can be designed to bind to target DNA of any length. Furthermore, the strength of DNA binding can be modulated by altering the length of the TALE [150].

Like ZF arrays, TALEs have been used extensively in the design of mammalian ATFs by fusing them to naturally occurring light- or ligand-inducible effector domains [151-153]. There are also some examples of bacterial ATFs that use TALEs as the DBD. For example, a TALE designed to bind to the *lac* O_1 operator can

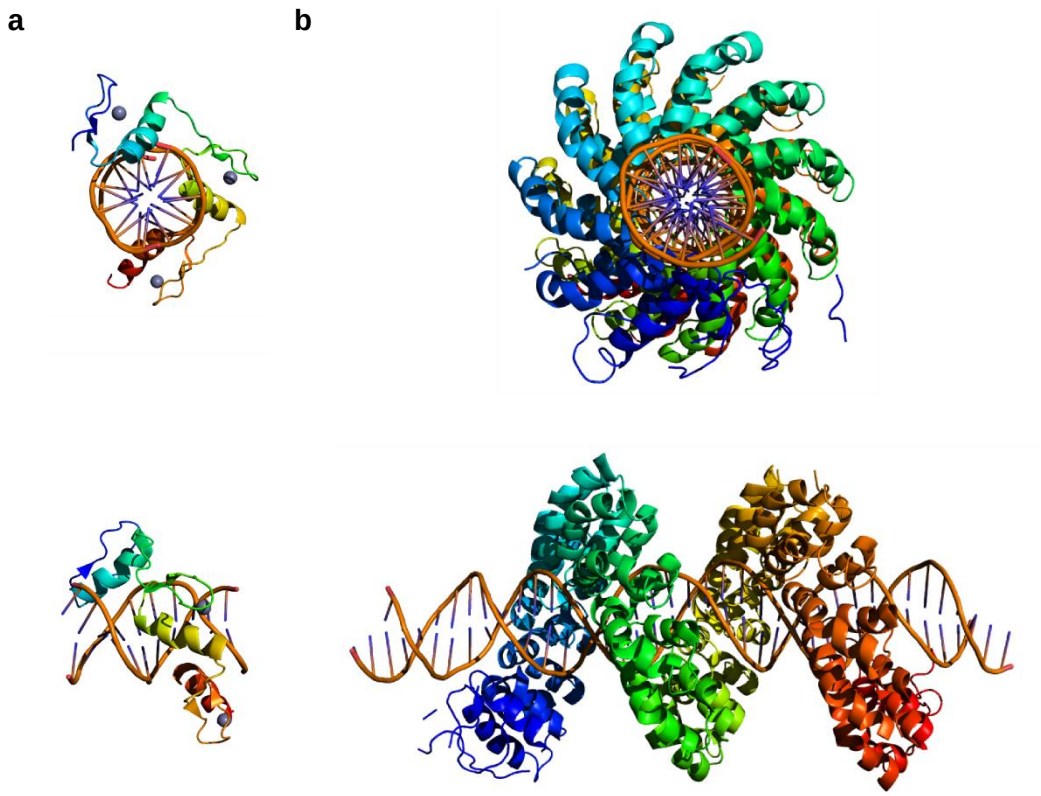


Figure 1-5 Engineerable DNA-binding domains for use in ATF design. (a) Zinc finger arrays (PBD ID: 1AAY [158]). Zinc ions are shown in lilac. (b) TALE (PDB ID: 3UGM [146]). Protein chains are coloured from N (blue) to C (red) terminus.

repress gene expression [122]. A modified version of this TALE with introduced protease sites provides a means of controlling the repression activity [154].

Finally, CRISPR/Cas9 ribonucleoprotein complexes can be used as DBDs [155]. In these complexes, a guide RNA (gRNA) targets a CRISPR-associated (Cas) nuclease to a specific DNA location [156]. A nuclease-deficient version of the Cas9 protein, called dCas9, can also be targeted to DNA without cleaving the DNA [156]. dCas9 has been used as the DBD in artificial transcriptional activators and repressors in *E. coli* [157].

1.2.2.2 Protein-protein interaction domains (PIDs)

Protein-protein interactions can be either constitutive or inducible. In constitutive interactions, provided the components of the interaction are present at sufficiently high concentrations, they will interact. In inducible interactions, the interaction occurs in the presence of some additional component, such as a small molecule, that may act allosterically or may itself form part of the interaction interface.

While inducible interactions are more controllable, this type of interaction is generally more difficult to design. This is reflected by the fact that the inducible protein-protein interactions used in ATFs are usually natural proteins or engineered versions of natural proteins. For example, a maltose-binding ATF was made by fusing a ZF DBD to a split maltose-binding protein (MBP) [159]. Maltose binding leads to a large conformational change in MBP, which also alters the conformation of the ZF such that it can no-longer bind DNA.

Another inducible system that has been widely exploited is the monomeric FK506-binding protein (FKBP), which binds to the small molecule drug rapamycin and similar synthetic ligands [160]. A version of FKBP with a single amino acid substitution can bind a synthetic ligand that does not bind the wild type protein [161]. Bivalent versions of this ligand, and others, can induce dimerisation of the protein [162]. Furthermore, another point mutation can turn FKBP into a dimer that can be reversibly dissociated by rapamycin binding [163]. Both mutants have been used to control the dimerisation of TALE DBDs in order to bring about DNA-looping [164].

Single-domain antibodies have also been proposed as inducible PIDs. For example, a V_HH fragment (a single-domain fragment of a camelid heavy-chain antibody) that undergoes dimerisation on binding its antigen, caffeine, has been used to mediate the interaction between monomers of dimeric DNA-binding proteins in order to facilitate DNA-binding [130,165].

Naturally occurring proteins have also been used as constitutive PIDs in bacterial ATFs. For example, the Gal4 dimerisation domain-Gal11^P interaction has been used to mediate the interaction between RNAP ω subunit and a DBD to activate transcription [110]. Gal4 is a yeast transcriptional activator and Gal11^P is a version of a component of yeast RNAP II [166]. These components have also been used to mediate the interaction between RNAP α subunit and novel ZF DBDs [167] and between ZF domains to make dimeric ZF DNA-binding proteins [168]. Novel peptide-peptide interactions have also been identified using phage display. Although only 15 amino acids long each, these could also mediate the dimerisation of ZF domains [169].

Higher-order oligomerisation domains have also been used. For example, the λ cI oligomerisation domain has been used to make a highly cooperative transcriptional repressors when fused to various DBDs [170].

Coiled coils and leucine zippers, which are present in many natural transcription factors, have also been used as constitutive interaction domains in ATFs. For example, the leucine-zipper region of the yeast bZIP transcription factor, GCN4, has been used to make dimeric ZF DNA-binding proteins [171,172]. More recently, *de novo* designed coiled coils have also been used as PIDs in transcriptional activators [111,173]. Heterodimeric coiled coils of different lengths, which have been shown to have different dissociation constants *in vitro* [174], gave different levels of transcriptional activation when the two components of the heterodimer were fused to a ZF DBD and T7 polymerase, respectively [111]. Finally, a collection of heterodimeric coiled coils has been used to reconstitute a split-transcription factor to restore transcription repression in yeast [175].

Beyond ATF design, natural and designed coiled coils have also been used successfully as PIDs in many other contexts, including in the assembly of multivalent antibodies [176,177] and to direct the assembly of photosynthetic reaction centres [178]. There are many examples of *de novo* designed coiled coils in the literature, and there is scope yet to design many more. They therefore represent a valuable resource as protein-protein interactions in ATF design.

1.3 Coiled coils: sequence, structure and function

The coiled coil is a ubiquitous protein-folding motif consisting of two or more α helices wrapped around each other to form a supercoil [179,180]. These protein domains perform a wide range of functions, including mediating protein-protein interactions. The sequences and structures of coiled coils have a number of characteristic features, which are described here.

1.3.1 Primary structure

The primary structure of a protein is the order of amino acids in its linear sequence when read from the N to the C terminus. Naturally occurring proteins usually consist of the 20 canonical amino acids. Every aspect of protein behaviour including structure, function, stability and folding is encoded within the primary sequence. Understanding how a protein's sequence ultimately dictates these properties (the protein-folding problem) has been a field of intense study [181]. While predicting the structure of a protein from its sequence is still difficult for the majority of proteins, there are some protein folds where the sequence-to-structure relationships are better understood. The coiled coil is one of these. As such, it is

not only possible to predict coiled coils and their structures from the amino acid sequence [182,183], it is also possible to design new sequences that will fold into defined structures with controllable characteristics.

The sequences of the majority of coiled coils consist of repeating seven residue motifs, called heptads, in which the residue positions are labelled *a*-*g*. There are other, less common repeat motifs but these will not be discussed here [184-186]. The *a* and *d* positions of the heptad are generally occupied by hydrophobic residues while the remaining positions are usually occupied by polar residues, giving a characteristic HP-pattern of **HPPHPPP** (**H**, hydrophobic; **P**, polar, Figure 1-6). The exact identity of these residues controls structural features such as oligomeric state, partner selection, helix orientation and supercoil geometry [187-190].

A huge variety of 3D structures can be derived from this seemingly simple heptad repeat. Additionally, coiled-coil properties beyond their tertiary structure can also be designed. For example, it is possible to modulate the strength of the interaction between the helices of a coiled coil to alter its stability [174,191,192].

1.3.2 Secondary structure

Proteins are made up of secondary structure elements that include α helices, 3_{10} helices and β strands. These structural elements are stabilised by regular backbone hydrogen bonds. For example, α helices, the secondary structure element that makes up α-helical coiled coils, contain hydrogen bonds between N–H donors in the backbone amide bond of residue i and carbonyl acceptors in the backbone amide bond of residue i+4 (Figure 1-7a) [193].

Secondary structure elements can also be described by a number of parameters including residues per turn, rise per residue, helix radius and the torsion angles, ϕ and ψ (Figure 1-7b-e). α-Helices have 3.6 residues per turn, a rise per residue of 1.5 Å and a backbone radius of 2.3 Å. The ϕ and ψ angles for α helices are centred around -63 ° and -43 °, respectively [194-196].

When a linear **HPPHPPP** heptad sequence is projected onto a helical wheel, the H residues at *a* and *d* align on the same side of the helix (Figure 1-6). This creates an amphipathic helix with both a hydrophobic and a polar face. Association of the hydrophobic face with other hydrophobic surfaces, resulting in burial of the hydrophobic residues, is a major driving force for protein folding.

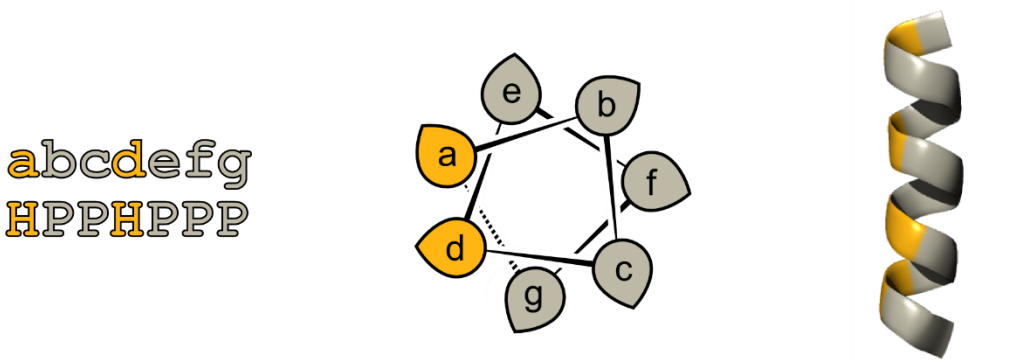


Figure 1-6 Patterning of hydrophobic and polar residues results in amphipathic helices. Projecting a HP-pattern (left) onto a helical wheel diagram (centre) demonstrates that hydrophobic residues are located on the same helical face. The orientation of the teardrops indicates the approximate direction of the C_α-C_β bond vector the residue at each position. The result is an amphipathic helix (right). Polar residues are highlighted in grey; hydrophobic residues are highlighted in yellow.

1.3.3 Tertiary structure

Variable numbers of right-handed amphipathic α helices wrap around each other to form left-handed superhelical coiled coils. The supercoiling comes about due to the mismatch in the number of residues per turn in an α helix (3.6) and the periodicity of the hydrophobic residues in the heptad (3.5). The result is a hydrophobic seam that spirals across the face of the helix. Helices can then wrap around each other *via* these faces. Like the individual helices, the coiled-coil superhelix can also be described by a number of parameters (Figure 1-8a) [197-199].

Furthermore, α helices in coiled coils take part in intimate interhelical interactions called knobs-into-holes (KIH) interactions where a “knob” residue on one helix docks into a diamond-shaped “hole” of four residues on the adjacent helix (Figure 1-8b) [200]. KIH interactions form along the length of a coiled coil such that a residue that forms part of a hole on one helix also acts as a knob residue, interacting with another hole on the adjacent helix. Thus, the helices become tightly interdigitated. Where residues act as both knobs and holes, these are termed complementary KIH interactions [201]. Isolated KIH interactions can also occur.

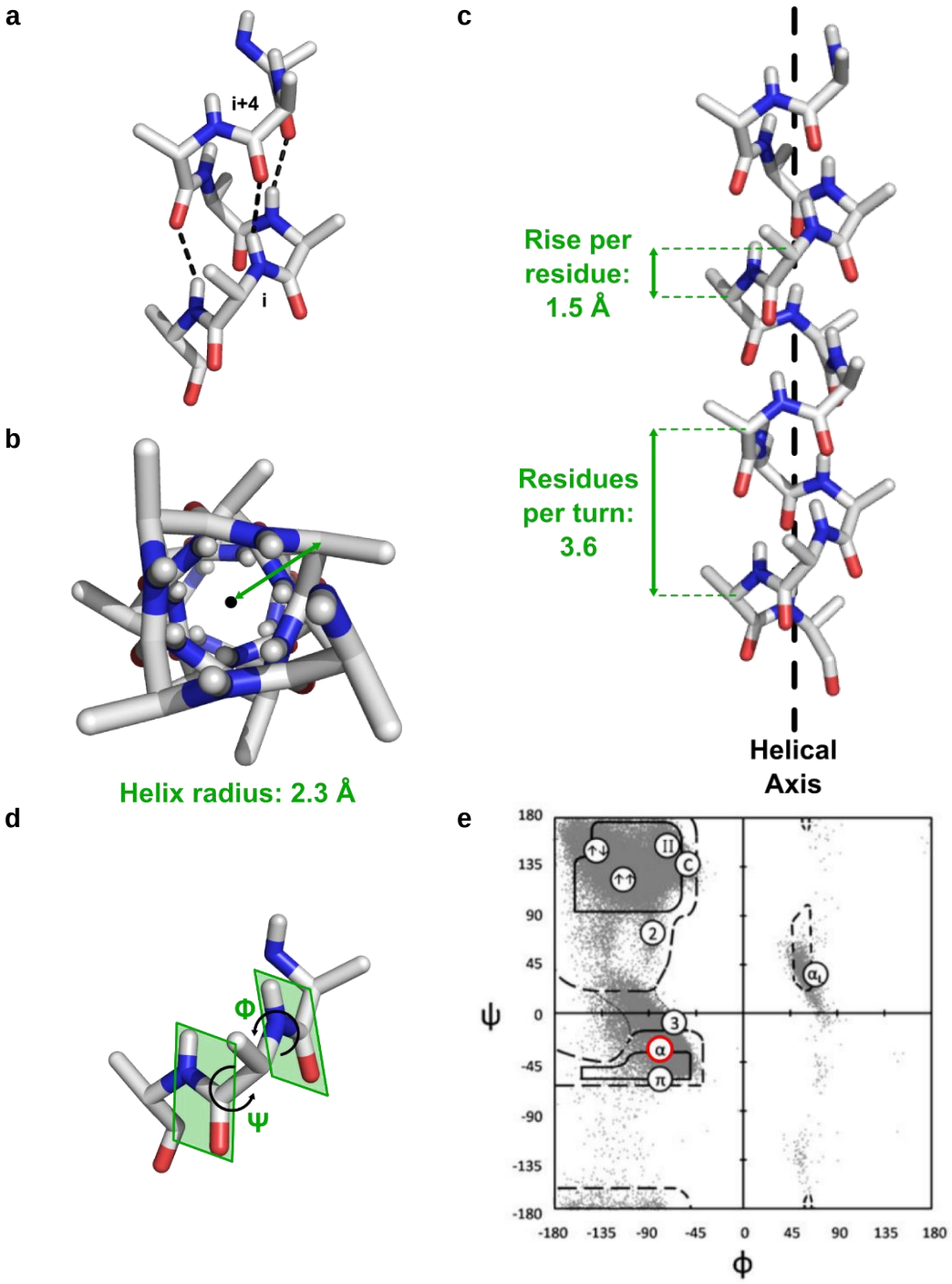


Figure 1-7 Structure and parameters of α helices. (a) Section of α helix demonstrating i to i+4 hydrogen bonds between backbone N–H and C=O groups. (b) An α helix viewed from the N terminus. The black dot represents the helical axis and the green arrow indicates the helical radius. (c) A section of α helix demonstrating the rise per residue and residues per turn parameters. The dashed black line is the helical axis. (d) A section of α helix demonstrating the torsion angles ϕ and ψ. Green boxes highlight planar amide bonds. (e) Ramachandran plot with α-helical region highlighted in red. Modified from reference [195]. In (b) and (c), idealised values for each parameter are shown in green.

One way to visualise the complementary KIH interactions that occur in coiled coils is using helical nets (Figure 1-8c) [200]. These diagrams are a visualisation of an "unrolled" helix with the positions of each C_α marked. When two of these unrolled helices are superimposed, the locations of the complementary KIH can be observed along the lengths of the helices (Figure 1-8c).

The interhelical interfaces in coiled coils fall into four categories: Type-N, Type-I, Type-II and Type-III (Figure 1-9). In Type-N interfaces, residues at the *a* and *d* positions primarily contribute to KIH packing and oligomeric state definition and form a single hydrophobic seam through which the amphipathic helices interact. Along with the *e* and *g* positions, which contribute to the hole residues, these residues make up a single motif called a *gade* motif and the heptads generally follow the classical **HPPHPPP** pattern. This type of interface is most compatible with dimer formation.

However, higher-order coiled coils can also contain an additional hydrophobic seam such that each helix interacts with its two neighbours *via* separate hydrophobic seams [202]. These "multifaceted" coiled coils are designated Type-I, -II or -III depending on the extent of the overlap between the two seams (Figure 1-9).

Expanding the hydrophobic seams beyond the *a* and *d* positions means that the more-peripheral *b*, *c*, *e* and *g* residues become part of the coiled-coil interface. This results in *gade*-like motifs where other positions in the heptad become equivalent to the *g*, *a*, *d* and *e* residues in the traditional Type-N interface. For example, in Type-I interfaces, the two *gade*-like interfaces are made up of the *gade* and *cdga* positions, which overlap by three residues; in Type-II interfaces they are made up of the *deab* and *cdga* positions, which overlap by two residues; in Type-III interfaces they are made up of the *gade* and *bcgf* positions, which overlap by a single residue. These *gade*-like motifs are essentially two individual Type-N interfaces that are superimposed in a larger Type-I, -II or -III interface. The Type-N interfaces can be thought of as forming individual dimeric interactions within a larger coiled-coil assembly, but with the KIH interactions occurring between different groups of residues, which are characteristic for the interface type.

The different types of multifaceted interface have different angles between the two hydrophobic seams [43,203,204]. These inter-seam angles dictate the positions of the adjacent helices relative to each other; as the inter-seam angle increases, so too will the angle between two flanking helices interacting with the same central helix.

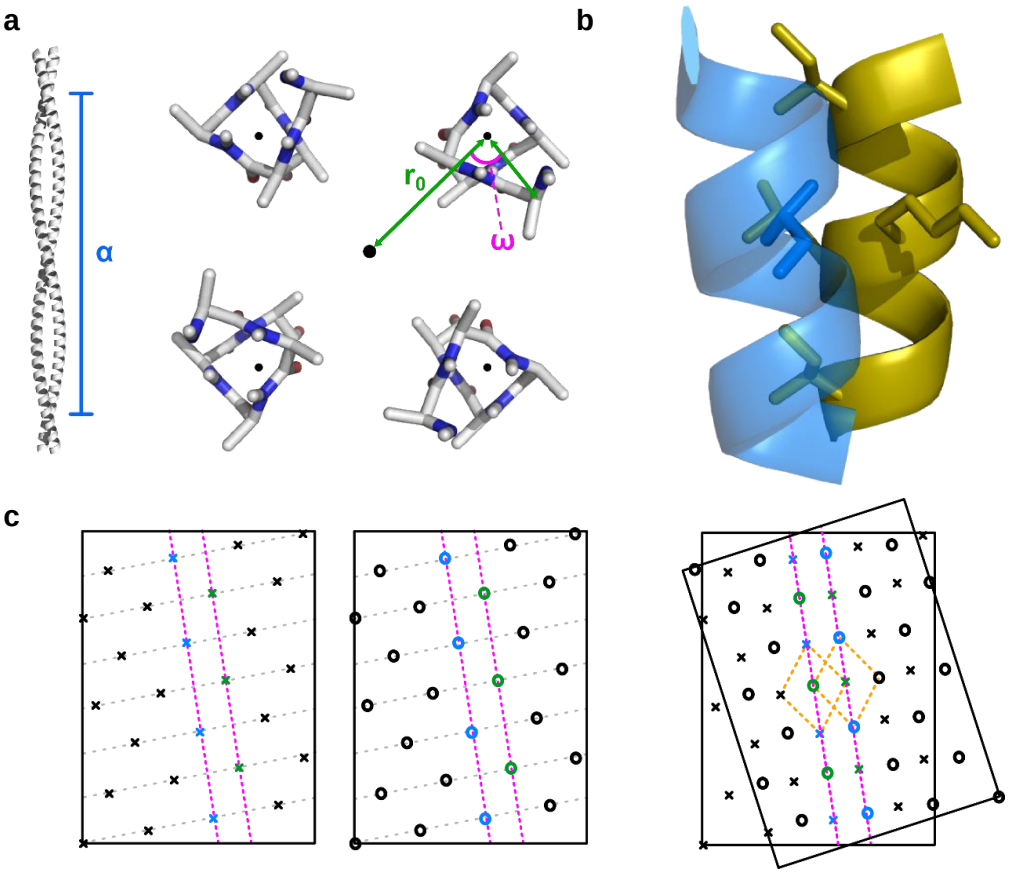


Figure 1-8 Structure and parameters of coiled coils. (a) Coiled coil parameters: α, pitch (Å); r_0, superhelix radius (Å); ω, helix phase (°) [198,199]. (b) Knobs-into-holes packing. A Leu knob (blue side chain) packs into an Ile-Lys-Ile-Leu hole (yellow side chains) in the adjacent helix. Figure based on a model of a tetrameric coiled coil. (c) Helical nets for two separate helices (left, crosses and circles indicate C_α positions, grey lines indicate backbones) and superimposed to show KIH packing in a dimeric coiled coil interface (right, complementary KIH interactions highlighted by yellow diamonds). Heptad *d* positions are blue, *a* positions are green. Magenta lines indicate trajectories of hydrophobic seams defined by *a* and *d* residues.

This helps determine oligomeric state. For example, the idealised angle between the seams in Type-I interfaces should be conducive with trimer formation; in Type-II with pentamer formation; and in Type-III with tetradecamer formation [204]. As the inter-seam angle increases, the oligomeric state must also increase in order to complete the ring of helices and "close" the coiled coil. However, while these are the idealised oligomeric states for these coiled-coil interfaces, they can also be compatible with other oligomeric states. For example, coiled coils with Type-II interfaces have been shown to form oligomeric states up to heptamer and are theoretically also compatible with tetramer formation [43].

Tetramers have also been observed with Type-N interfaces, with hydrophobic residues at only the *a* and *d* positions [187], and Type-I interfaces, with hydrophobic

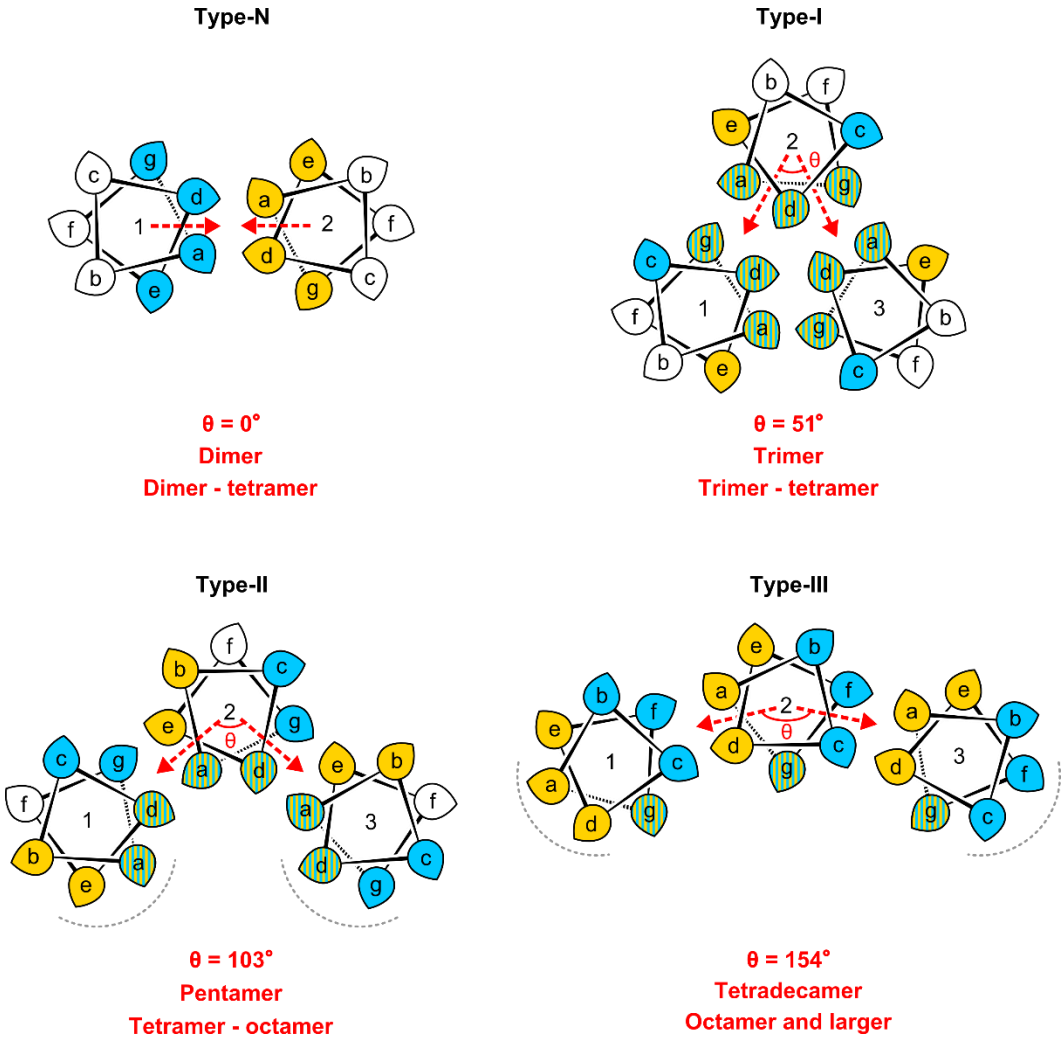


Figure 1-9 Types of coiled-coil interfaces. Top left: Type-N. Top right: Type-I. Bottom left: Type-II. Bottom right: Type-III. Red dashed arrows indicate the approximate orientation of the hydrophobic seams. *gade*-like motifs are shown in yellow and blue, shared residues are striped. Idealised inter-seam angles, optimal oligomeric states and range of potential oligomeric states available to each type of interface are shown in red below each helical wheel diagram [204]. Dashed grey arcs indicate that these helices make further interactions (not shown).

residues at *ade* or *adg* positions [205,206]. Thus, a variety of different interfaces appear to be compatible with tetrameric coiled coil formation. However, though the sequences of these coiled coils are similar to those of Type-N and Type-I coiled coils, they may still be forming Type-II-like structures, where the *a*, *d*, *e*, and *g* residues can all act as knobs in KIH interactions.

1.3.4 Sequence to structure: oligomeric-state specification

Seminal experiments by Harbury *et al* found that mutating the core *a*/*d* positions of the model leucine zipper GCN4-p1 to various combinations of Ile and Leu led to

changes in oligomeric state: *a*=Ile/*d*=Leu gave dimers; *a*=*d*=Ile or Leu, trimers; and *a*=Leu/*d*=Ile, tetramers [187].

This was rationalised because Ile prefers to take part in parallel KIH packing in which the knob C_α-C_β bond vector is parallel to the vector between the C_α atoms of the hole residues. Conversely, only Leu is compatible with perpendicular KIH packing in which the knob vector is oriented perpendicular to the hole vector [187,207]. This preference is largely driven by the unfavourable packing and steric clashes that would occur if each amino acid took part in the other type of packing. In tetramers, *a* forms a perpendicular knob and *d* forms a parallel knob, hence *a*=Leu/*d*=Ile cores. The identification of these simple sequence-to-structure relationships galvanized the field of *de novo* coiled-coil design and the so-called "Harbury rules" are still used to guide the design of lower-order coiled coils [208].

1.3.5 De novo *coiled-coil design*

Coiled-coil sequences designed entirely from scratch allow us to disentangle the relationships between a protein's sequence and its observed properties. They also provide a "blank canvas" that can be augmented with additional functions and improved properties either by further rational design, computational design or *via* library screening or directed evolution [209-212]. In short, in order to introduce novel, complex behaviours into an existing protein, a simplified *de novo* sequence often provides a better starting point than a natural sequence that is already itself highly complex.

Extensive investigation has been carried out into oligomeric state specification in coiled coils and, as a result, a "basis set" of *de novo* homooligomers, with oligomeric states ranging from dimer to heptamer has been designed (Figure 1-10) [43,207,208,213,214]. There are also examples of *de novo* designed heteromers, particularly heterodimers [174,188,215], but fewer examples for higher-order oligomers such as tetramers.

Furthermore, while many *de novo* coiled coils have been characterised extensively *in vitro*, relatively little is known about their biological properties, that is, how they behave inside living organisms. Ultimately, it is not yet known whether coiled coils designed to have optimal properties *in vitro*, with high thermal stabilities and simple, repetitive sequences, will also be optimal for use *in vivo*. This area, protein design in the cell, remains largely unexplored.

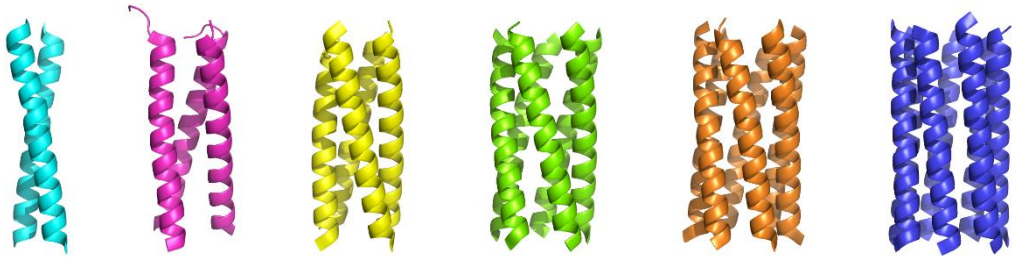


Figure 1-10 X-ray crystal structures of the basis set *de novo* designed coiled coils. From left to right: CC-Di, CC-Tri, CC-Tet, CC-Pent, CC-Hex and CC-Hept. PDB IDs: 4DZM [208]; 4DZL [208]; 3R4A [213]; 3PN8 [43]; 3R3K [213]; 4PNA [43].

1.3.6 Designed tetrameric coiled coils and 4-helix bundles

As designed constitutive PIDs, homo- and heterotetrameric coiled coils offer many advantages over lower-order coiled coils like dimers and trimers. Firstly, as the oligomeric state of a cooperatively folded protein is increased the sigmoidal concentration-dependent folding response becomes sharper (Figure 1-11). That is, tetramers are more sensitive to smaller changes in protein/peptide concentration than proteins with lower oligomeric states [170,216]. These small changes in concentration lead to larger changes in folding and, therefore, structure-dependent functions. Such systems show a large response to a small range of inputs (here, the input is protein concentration and the response is folding/association). This gives their response (folding) curves increasingly binary like characteristics compared to dimers and trimers and, as such, they make useful biological switches with more distinct on/off states [217,218]. Furthermore, such systems are generally more resistant to biological noise, such as fluctuations in gene expression [219-221].

A further reason why tetramers are more useful as PIDs is that a tetrameric assembly can contain up to four separate components and could also be designed to form hetero systems where all four components are unique. This makes them useful as PIDs because they can co-localise multiple proteins of interest. For example, they could be used to mediate the interaction between two dimeric DBDs to bring about looping. Also, they could be used to recruit multiple other protein domains, such as enzymes, to a specific DNA location (Figure 1-12).

Finally, tetrameric coiled coils have larger hydrophobic cores. They are therefore generally more stable and may be more tolerant to mutations and modifications

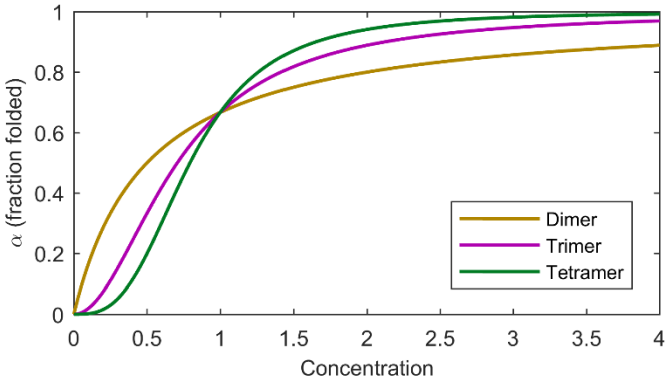


Figure 1-11 Theoretical concentration-dependent folding of different oligomeric states. As oligomeric state is increased from dimer to trimer to tetramer the sigmoidal concentration-dependent folding transitions become steeper. Concentration is in arbitrary concentration units. Fraction folded (α) = [folded]/([folded]+[unfolded]). $\alpha_n = C^n/(K_D.C + C^n)$, where n is the oligomeric state (n = 2, 3, 4).

intended to introduce more-sophisticated properties into *de novo* designed coiled coils, such as environment-responsive behaviours. For example, PIDs could be designed to undergo structural rearrangements in response to small molecule binding or chemical modifications without adversely affecting their stability. Therefore, tetrameric coiled coils are in fact a good candidate for both constitutive and inducible PIDs in ATFs (Figure 1-12).

Since the discovery that placing leucine at *a* positions and isoleucine at *d* positions in the GCN4-p1 heptad repeat leads to the formation of a parallel tetramer [187], this model coiled coil has been modified further in a number of ways to produce additional tetrameric coiled coils. For example, expanding the hydrophobic seam beyond the *a* and *d* residues has often been found to have a profound effect on the structure of the coiled coil. When valines are placed at all *e* positions in GCN4-p1, the peptide forms a parallel tetrameric coiled coil with offset helices [206]. This is in contrast to GCN4-pLI, which is blunt-ended [187]. When valine is placed at all *g* positions in GCN4-p1, the peptide forms an antiparallel tetramer, again in contrast to GCN4-pLI, which is a parallel coiled coil [205]. When alanine is placed at either all *e* or all *g* positions, the peptides also form antiparallel tetrameric coiled coils [205,222]. The peptides also form antiparallel coiled coils when alanine is placed at the *e* positions and leucine and valine are placed at the *a* and *d* positions, respectively [223]. Finally, GCN4-pLI itself has also been further modified. For example, glutamate or lysine residues have been placed at all *c* and *e* positions to produce two peptides, ecE and ecK. These peptides interact to form a heterotetramer [224].

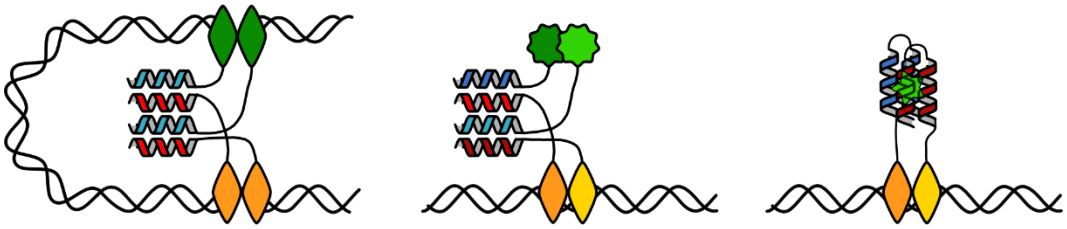


Figure 1-12 Schematics of proposed ATFs containing tetrameric coiled coils. Left, an A_2B_2 tetrameric coiled coil (light red and light blue helices) mediating an interaction between two dimeric DBDs (green and orange). Centre, an ABCD tetrameric coiled coil (light/dark red and light/dark blue helices) recruiting functional protein domains (light/dark green) to a dimeric DBD (yellow/orange). Right, a tetrameric coiled coil-based induced protein-protein interaction (dark red and dark blue helices) where a ligand (green) facilitates the interaction, resulting in formation of a dimeric DBD (yellow/orange).

Similar mutational studies have been performed with another natural coiled coil, the tetramerisation domain from the Lac repressor. This forms an antiparallel homotetramer. Modifications include increasing the length of the peptide to increase its stability [225] and placing oppositely charged residues at *b* and *c* positions to make two peptides that form a heterotetramer [192,226].

Further examples of novel tetrameric coiled coils include: a pair of disulphide-linked peptides with core alanine residues that form an antiparallel heteromer [227]; a computationally designed right-handed coiled coil (coiled coils are usually left-handed) [228]; a second computationally designed right-handed coiled coil in an antiparallel orientation where the helices are linked by loops to form a single-chain assembly [42]; a variant of a *de novo* heterodimer with a core substitution that forms a heterotetramer [229]; and an empirically designed parallel homotetramer [208]. The latter, CC-Tet, was designed according to the Harbury rules as part of a set of novel oligomerisation domains for use in synthetic biology [187,208].

Extensive design and characterisation of novel 4-helix bundles has also been carried out. These consist of four amphipathic interacting helices, often linked into helical hairpin or single-chain arrangements. As with coiled coils, their sequences follow a regular pattern of hydrophobic and polar residues, but they do not necessarily form the extensive knobs-into-holes interactions that are characteristic of coiled coils [230,231].

Early approaches to designing 4-helix bundles involved patterning hydrophobic and polar residues to form amphipathic α helices. These peptides usually contained a reduced palette of amino acids. For example, an early example

contained just leucine, glutamate, lysine and glycine [232,233]. The peptide formed a tetramer. When the helices were linked up to form helical hairpins *via* simple loops these hairpins also interacted to form dimeric 4-helix bundles [234]. The helices were also linked in a single chain to make a monomeric 4-helix bundle [235]. The structures were later made more native-like through mutagenesis [236] and were further modified to impart additional behaviours on them such as metal [237] or heme binding [238].

Further examples of 4-helix bundles have been identified from libraries of binary-patterned protein sequences. In these libraries, a HP-pattern (*e.g.* **PHPPHHP**) is defined, then the **H** and **P** positions are populated by a subset of hydrophobic (*e.g.* FILMV) and polar (*e.g.* DEHKNQ) residues at random [239-241]. Loop regions are also defined to produce single-chain proteins predicted to form monomeric helical bundles. The resulting libraries of proteins can then be probed for desirable properties. Folded 4-helix bundles with native-like structures can be readily extracted from such libraries [242-244]. Indeed, helical proteins have also been extracted from libraries where no binary pattern of **H**/**P** residues was defined [245,246]. Furthermore, the sequence diversity within the library encodes function, not just structure: 4-helix bundles capable of small molecule and co-factor binding and catalysis have been identified [211,212,247-249].

1.3.7 Ligand-binding coiled coils and helical bundles

Coiled coils and helical bundles that are able to bind to small molecules are a particularly attractive design target because, just like protein-protein interactions, protein-small molecule interactions are involved in a wide range of biological activities. These include: binding of substrates in enzymatic catalysis [250]; binding of cofactors to expand the chemical functionality of a protein [251]; sequestration and solubilisation of minerals [252,253]; transport across membranes [254]; and small molecule binding-induced conformational changes that allosterically alter the activity of a protein [255].

The latter is often observed in inducible transcription factors, such as the Lac repressor, which undergoes a conformational change on binding allolactose or isopropyl β-D-1-thiogalactopyranoside (IPTG) [117,255], and there are some examples of designed, ligand-inducible transcription factors (see Section 1.2.2.2) [29,256]. However, these attempts to design novel ligand-inducible transcription factors are yet to exploit the ability of some coiled coils and helical bundles to bind small

molecules (see below for details). For example, there are several examples of α-barrels that bind hydrophobic molecules in their axial pores. Furthermore, though the cores of lower order coiled coils (dimers to tetramers) are generally tightly packed, without pores or cavities, there are examples of natural and designed 3- and 4-helix bundles that not only contain channels or cavities but can also bind small molecules within these cavities.

Binding of small molecules has been associated with stabilisation of helical bundles and conformational switching, and so ligand binding may have an effect on the biophysical properties of a designed coiled coil. Therefore, while this thesis will predominantly focus on using coiled coils as constitutive PIDs, ligand-binding coiled coils and helical bundles may be of use as inducible protein-protein interactions in the future.

1.3.7.1 Small molecule binding to α-helical barrels

Coiled coils with oligomeric states of pentamer and above, the α-barrels, have axial, solvent accessible pores where the diameter is proportional to the oligomeric state [43,202,203]. Some naturally occurring α-barrels bind small molecules in these pores, most notably the coiled-coil region of the cartilage oligomeric matrix protein, COMPcc. This is a soluble homopentameric α-barrel that contains a pore divided into two chambers by a ring of glutamine residues [257,258]. These chambers, which are lined by aliphatic residues, can bind a wide range of molecules from benzene to vitamins to fatty acids [259,260].

Binding of dyes and other hydrophobic molecules has also been demonstrated for various *de novo* α-barrels with oligomeric states of pentamer to heptamer [261-264]. Moreover, it has been shown through dye displacement assays that different *de novo* α-barrels bind different small molecules with varying affinities, indicating that these coiled coils can discriminate between a range of molecules [264]. The selectivity of the barrels could also be modulated by introducing polar residues into the pore lumens that were proposed to make specific electrostatic interactions with ligands.

1.3.7.2 Small molecule binding to naturally occurring 4-helix bundles

The C-terminal coiled-coil region of the tetrabrachion protein from the thermophilic *Staphylothermus marinus* forms a right-handed tetrameric coiled coil. The crystal structure of the coiled coil (named RHCC for right-handed coiled coil) revealed it

contains four central cavities linked by a continuous, solvent accessible axial pore [265]. In the crystal structure these cavities are occupied by ordered waters. The presence of the cavities has led researchers to investigate RHCC as a drug delivery system: the potent chemotherapy agent cisplatin can be loaded into RHCC in a 1:1 cisplatin:tetramer ratio to form a stable complex [266]. While RHCC alone is neither toxic nor immunogenic, the RHCC-cisplatin complex could enter tumour cells and had comparable efficacy against various tumour cell lines to cisplatin alone. A *de novo*-designed right-handed coiled coil also contains solvent accessible cavities, although these are not linked by a central pore [228]. Such *de novo* systems may also be of use in the delivery of cytotoxic drugs.

The homotetrameric M2 protein from influenza A contains a transmembrane 4-helix bundle [267,268]. This transmembrane segment forms a proton channel and facilitates various functions of the protein, including endosome acidification and viral release [269]. Antiviral drugs, including amantadine, rimantadine and other adamantane-derived drugs are believed to bind in the M2 channel lumen, blocking proton translocation [270-273].

1.3.7.3 Small molecule binding to designed 3- and 4-helix bundles

Various attempts have also been made to introduce ligand-binding behaviour into the model leucine zipper GCN4-p1 and its tetrameric derivative, GCN4-pLI [187,274]. The first such attempt involved mutating the core asparagine in GCN4-p1 to alanine. The resulting variant was predominantly dimeric but was stabilised in a trimeric state by the addition of small hydrophobic molecules such as benzene [275]. The variant was co-crystallised with benzene occupying a cavity with a volume of 165 $Å^3$. Addition of the ligand also led to an increase in the thermal stability of the coiled coil. This ligand-induced trimerization mechanism was dubbed an allosteric switch.

The larger core volumes of tetramers compared to dimers and trimers makes them more amenable than smaller coiled coils to the introduction of cavities and channels. Indeed, the crystal structures of some parallel tetrameric coiled coils, such as GCN4-pLI, already contain very small internal cavities [187]. Attempts have been made to expand these cavities such that they may accommodate small molecules. Yadav *et al* replaced core leucine residues at various *a* positions in GCN4-pLI with various small residues [276]. When Leu9 was replaced by glycine or alanine, the resulting tetramers co-crystallised with iodobenzene in a solvent

accessible tunnel and a 220 $Å^3$ cavity, respectively. Subsequently, Mizuno *et al* demonstrated that a single-chain antiparallel GCN4-pLI variant with seven core alanine substitutions and a single core tryptophan substitution per helix could bind larger hydrophobic molecules such as adamantane [277]. The introduced cavity showed preferential binding to hydrophobic molecules and bound approximately 350-fold more strongly to adamantane than to 1-adamantanol.

Other attempts at introducing ligand binding into small helical proteins have focussed on helical bundles. One such attempt involved making cavity-creating leucine to alanine substitutions in a covalently templated 3-helix bundle [278]. Peptides with one or two alanine substitutions could be mixed-and-matched to make heteromeric assemblies through covalent attachment to a template peptide. The resulting α-helical bundles bound the hydrophobic dye 8-Anilino-1-naphthalenesulfonic acid (ANS). It is important to note that ANS is traditionally used to determine whether a protein adopts a unique, rather than molten globule, structure as it binds non-specifically to exposed hydrophobic patches in the latter. Therefore, the observed ANS binding could simply be an indication of a poorly structured core. However, based on their cooperative unfolding transitions, the covalently templated proteins appeared to have native-like properties. Therefore, here, ANS binding is an indication that the proteins contain accessible hydrophobic cavities that could be further modified to accommodate more-specific ligands.

De novo 4-helix bundles have also been designed to bind to halothane, a volatile molecule used as a general anaesthetic [279]. These bundles were initially designed as a model to probe halothane-protein interactions because the specific binding site of the molecule *in vivo* was unknown [280].

Extensive work has also been done to introduce porphyrin binding activity into 4-helix bundles through rational and computational design. These ligands, which include hemes and their non-natural analogues [238,281-284], act as cofactors to expand the chemical functionality of the proteins. Activities such as oxygen binding, oxidoreductase activity and light harvesting have been introduced into such *de novo* porphyrin-bound proteins [251,285,286].

Another approach that has been taken to introduce porphyrin binding to novel 4-helix bundles is through selection for heme binding from a library of binary patterned 4-helix bundles [247,248]. The specific structures, and therefore the functions, of such proteins are not explicitly designed. However, these simple

proteins have displayed a range of activities, which also includes small molecule binding [211,212]. In one study, core alanine mutations were made in a well-folded binary patterned 4-helix bundle [211]. Small hydrophobic molecules including benzene could bind to the resulting cavities in the proteins core. A more elaborate study involved screening three 4-helix bundles against a library of small molecules [212]. Strikingly, the proteins, which were not designed to be small molecule-binding proteins, each displayed binding to several ligands with reasonable affinity and selectivity.

1.3.7.4 Metal binding to helical bundles

The incorporation of metal binding into coiled coils is well explored [287] and will not be discussed here in detail. However, there are some examples of this phenomenon that are pertinent and interesting in the context of inducible protein-protein interactions and transcriptional regulation. Firstly, metal binding is able to induce coiled-coil folding. A peptide designed to contain core histidine residues is able to bind various metals, including nickel [288]. On metal coordination, this peptide forms a trimeric coiled coil. Further examples a metal binding-induced conformational changes include a peptide that switches from a trimeric coiled coil to a monomeric helix-loop-helix structure on coordinating zinc *via* two histidine residues [289] and a peptide that switches from a trimeric coiled coil to a Cys_2His_2 ZF motif on coordinating zinc [290].

Finally, a version of the DNA-binding portion of GCN4 has been engineered to be metal-responsive [291]. Terpyridine-linked GCN4 dimers showed some binding to target DNA *in vitro* in the absence of metal. On binding copper or zinc, the terpyridine linker undergoes a conformational change, which induces a conformational change in the peptide dimer. This leads to an increase in DNA binding.

1.4 Scope of this thesis

Coiled coils are used frequently in nature to mediate protein-protein interactions. The relative simplicity of coiled-coil sequences and structures makes them prime targets for the design of artificial PIDs in which properties such as the strength and specificity of the interaction can be controlled. The aim of the work described in this thesis was to design and characterise a set of *de novo* coiled coils intended for use as constitutive PIDs *in vivo*.

Chapter 3 describes the design of a number of homotetrameric coiled coils and their biophysical and crystallographic characterisation. Sequence determinants that specify oligomeric state, including the identities of the *d* and *g* residues, are also explored. This is built upon in Chapter 4, where a collection of A_2B_2 heterotetramers are described. These heterotetramers, which contain two copies each of two different peptides, have different thermal stabilities, depending on the identity of the core residues.

In Chapter 5, a sub-set of these heterotetramers are examined in *E. coli* for their ability to act as artificial PIDs in a living organism. This is done using a transcription repression assay where the *de novo* coiled coils mediate an interaction between monomeric proteins that require oligomerisation in order to form functional DBDs. The work in this chapter shows that simple peptide sequences are able to locate their interaction partners in a complex cellular environment. Therefore, though not necessarily optimised for use *in vivo*, the coiled coils can be used successfully as artificial constitutive PIDs in organisms. Furthermore, a set of semi-synthetic ATFs, that consist of designed PIDs and a natural DBD, are produced.

Finally, Chapter 6 describes preliminary results towards the design of an ABCD heterotetramer, where all four constituent peptides have a different sequence. Moreover, following the successful design of a number of constitutive coiled coil-based PIDs, this chapter also discusses preliminary work relating to the design of ligand-binding tetrameric coiled coils with the intention that, in the future, the available constitutive protein-protein interactions may be made into inducible protein-protein interactions. Such components would be invaluable in the design of inducible ATFs.

Chapter 2: Methods and Materials

2.1 Coiled-coil design

All designed peptide sequences are listed in Table 8-1.

2.1.1 Homotetramer design

2.1.1.1 Design of CC-Tet derivatives

The sequence of the peptide CC-Tet-KE was based on the homotetrameric coiled coil CC-Tet [208]. CC-Tet-KE was designed by swapping the identities of the charged residues at the *e* and *g* positions in CC-Tet. The sequence is otherwise identical, including a C-terminal AG mass tag used for peptide identification.

The CC-Tet-KE-N peptides were designed as mutants of CC-Tet-KE that contained Asn residues at the *d* and *e* positions of the third heptad (positions 20 and 21). 4- and 4.5-heptad versions were designed. The 4-heptad version (CC-Tet-KE-N-4) retained the C-terminal AG mass tag. In the 4.5-heptad version (CC-Tet-KE-N-4.5), this mass tag was removed, and the peptide was extended C-terminally with an additional *gabc* motif (sequence: KLAA), representing approximately half of the full *gabcdef* heptad (sequence: KLAAIE*f*).

2.1.1.2 Design of Glu-Lys stabilised homomers

The peptides 1-LI-EK, 1-LI-KE, 2-LI-EK, 2-LI-KE, 3-LI-EK, 3-LI-KE, 4-LI-EK and 4-LI-KE were designed to contain different configurations of Glu-Lys pairs at positions deemed to be close enough for the formation of interhelical ionic interactions from the crystal structure of CC-Tet (PDB ID: 3R4A) [208]. Average inter-residue distances were determined from the .pdb coordinate file for CC-Tet using a script implemented in Python.

All eight peptides were designed in *c*-register (*i.e.* excepting capping Gly residues at the C and N termini, the peptide sequence begins at a *c* position) and contained *a*=Leu/*d*=Ile cores. Peptides 1-LI-EK and 1-LI-KE contained *e*/*g* ionic pairs where in 1-LI-EK *g*=Glu, *e*=Lys and in 1-LI-KE *g*=Lys, *e*=Glu. Peptides 2-LI-EK and 2-LI-KE contained *c*/*e* ionic pairs where in 2-LI-EK *c*=Glu, *e*=Lys and in 2-LI-KE

c=Lys, *e*=Glu. Peptides 3-LI-EK and 3-LI-KE contained *g*/*b* ionic pairs where in 3-LI-EK *g*=Glu, *b*=Lys and in 3-LI-KE *g*=Lys, *b*=Glu. Peptides 4-LI-EK and 4-LI-KE contained *b*/*c* ionic pairs where in 4-LI-EK *b*=Glu, *c*=Lys and in 4-LI-KE *b*=Lys, *c*=Glu. Where *e* or *g* positions were not Glu or Lys, these positions were made Gln. Where *b* or *c* positions were not Glu or Lys, these positions were made Ala.

The peptides 1-LV-EK, 1-LV-KE, 2-LV-EK, 2-LV-KE, 3-LV-EK and 3-LV-KE are based on 1-LI-EK, 1-LI-KE, 2-LI-EK, 2-LI-KE, 3-LI-EK and 3-LI-KE, respectively, except they contained *a*=Leu/*d*=Val cores.

The *f* positions of all 14 homomeric peptides were populated by Gln (or Lys, where required to promote solubility) and a single Trp at position *f*19 to provide a chromophore. Helix-capping Gly residues were added to N and C termini.

2.1.1.3 Design of Ala@g *homomers*

The peptides 2-LIA-EK and 2-LIA-KE were designed in *c*-register and contained *a*=Leu/*d*=Ile cores. The peptides contained *c*/*e* ionic pairs where in 2-LIA-EK *c*=Glu, *e*=Lys and in 2-LIA-KE *c*=Lys, *e*=Glu. All *g* and *b* positions were occupied by Ala. Both peptides contained Gln at *f* positions, except at *f*19 where Trp was used to provide a chromophore. Helix-capping Gly residues were added to N and C termini.

2.1.1.4 Design of ELAEIK

ELAEIK was designed by Drew Thompson using a protocol that had previously been used to design water-soluble homomeric coiled-coil barrels with oligomeric states of pentamer to heptamer [43]. The protocol was modified for the design of A_2B_2 heterotetrameric coiled coils. While no successful heterotetramers were designed using this method, it did yield a number of homomeric coiled coils, including the homotetramer ELAEIK.

Heptad positions *a*, *d*, *e*, and *g* were populated by all combinations of the residues Ala, Glu, Ile, Lys, Leu, Asn, Gln, Arg, Ser and Val to calculate all possible "*gade*" motifs. *gade* sequences that could be overlapped to give two type II *gabcde* repeats were assessed for their ability to form pairs of heterotypic interactions using the bZIP scores for all the possible interactions between that set of *gade* sequences [292]. The best score for the undesired products (*i.e.* homotypic interactions) was subtracted from the worst score for the desired products (*i.e.* two

different heterotypic interactions) to give a Δ-score for that pair of *gabcde* motifs. The following filtering was applied to the resulting pairs: core *a* and *d* positions must be Ile, Leu or Val; no Ile, Leu or Val at *c* and *b* positions; *e* and *g* positions must be polar (Glu, Lys, Gln, Asn, Ser or Thr). Sequence pairs with a Δ-score of >15 were considered. Full peptide sequences were designed in *c*-register. The *f* positions were populated by Lys and a single Trp at position *f*19. Helix-capping Gly residues were added to N and C termini.

2.1.2 ABAB Heterotetramer design

2.1.2.1 Design of Glu-Lys stabilised heteromers

The peptides 1-LI-A, 1-LI-B, 2-LI-A, 2-LI-B, 3-LI-A and 3-LI-B contained *a*=Leu/*d*=Ile cores while the peptides 1-LV-A, 1-LV-B, 2-LV-A, 2-LV-B, 3-LV-A and 3-LV-B contained *a*=Leu/*d*=Val cores. Oppositely charged Glu and Lys residues were arranged in various configurations. Peptides 1-LI-A, 1-LI-B, 1-LV-A and 1-LV-B contained *e*/*g* ionic pairs where in 1-LI-A and 1-LV-A *e*=*g*=Glu and in 1-LI-B and 1-LV-B *e*=*g*=Lys. Peptides 2-LI-A, 2-LI-B, 2-LV-A and 2-LV-B contained *c*/*e* ionic pairs where in 2-LI-A and 2-LV-A *c*=*e*=Glu and in 2-LI-B and 2-LV-B *c*=*e*=Lys. Peptides 3-LI-A, 3-LI-B, 3-LV-A and 3-LV-B contained *b*/*g* ionic pairs where in 3-LI-A and 3-LV-A *b*=*g*=Glu and in 3-LI-B and 3-LV-B *b*=*g*=Lys. Where *e* or *g* positions were not occupied by Glu or Lys, these positions were Gln. Where *b* or *c* positions were not occupied by Glu or Lys, these positions were Ala. The *f* positions were populated with Gln residues in B (basic) peptides or Lys in A (acidic) peptides to promote solubility and aid purification. A single Trp residue was placed at position *f*19 to provide a chromophore. Helix-capping Gly residues were added to N and C termini.

The peptides 1-LI-A-*g* and 1-LI-B-*g* were identical to 1-LI-A and 1-LI-B, respectively, except they were designed in *g*-register and the Trp chromophore was at *f*22.

2.1.2.2 Design of Ala@g heteromers

The peptides 2-LIA-A and 2-LIA-B contained *a*=Leu/*d*=Ile cores while the peptides 2-LVA-A and 2-LVA-B contained *a*=Leu/*d*=Val cores. All peptides contained *c*/*e* ionic pairs. In 2-LIA-A and 2-LVA-A *c*=*e*=Glu and in 2-LIA-B and 2-LVA-B *c*=*e*=Lys. All *g* and *b* positions were occupied by Ala. All peptides contained Gln at *f* positions

except at *f*19, which was Trp to provide a chromophore. Helix-capping Gly residues were added to N and C termini.

2.1.3 ABCD Heterotetramer design

The sequences and resulting ABCD heteromer combinations of ABCD peptides 1–8 were generated computationally using a script implemented in Python. All 4-heptad permutations (without repetition) of the heptads ELAAxEx, KLAAxKx, ELAAxKx and KLAAxEx (*gabcdef*) were generated such that the resulting 24 peptides contained each heptad once. Then, all possible 4-peptide permutations (with repetition) of the 24 peptides were generated, representing all possible homotetramers and ABAB, AABC and ABCD heterotetramers. All tetramers were scored and sorted based on whether *e-g′* interactions between equivalent heptads in adjacent helices would result in attractive (E-K) or repulsive (E-E or K-K) ionic interactions. ABCD heterotetramers where all 16 possible interhelical ionic interactions were satisfied and ABAB and AABC heterotetramers where 13 or more ionic interactions were satisfied were extracted and stored. Peptides appearing in the "favourable" ABAB and AABC heterotetramers were cross referenced with the "favourable" ABCD heterotetramers. Where peptide pairs or triplets considered likely to interact to form ABAB or AABC heteromers appeared within ABCD heteromer quadruplets, these quadruplets were discarded.

The cores of the peptides appearing in the remaining ABCD heteromers were then designed in order to minimise homomerisation of these peptides. Where favourable ionic interactions would occur in the homomer the core *d* positions were Val. Where repulsive interactions would occur, *d* positions were Ile. Final peptides sequences were generated in *c* register. All *b* and *c* positions were Ala, and all *f* positions were Gln. Helix-capping Gly residues were added to N and C termini. Models of the resulting ABCD heterotetramer sequences were generated in ISAMBARD then scored using Bristol University Docking Engine (BUDE) [293,294]. Assemblies achieving better initial BUDE scores than a cut-off of -100 were kept and were then subject to parameter optimisation in ISAMBARD, and final BUDE scores for optimised models were obtained. Models were subject to 30 rounds of parameter optimisation for chain separation (distance between individual helical axes) and ΦC_α (rotation of helices relative to central axis) using a genetic algorithm.

Finally, C-terminal mass/chromophore tags were manually designed and added to the ABCD peptides 5–8 following the pattern GxΦ, where x is a small amino acid

(Gly, Ala, Ser, Asn) and Φ is a chromophore (Trp, Tyr). The ABCD peptides 1–4 did not contain mass/chromophore tags and were designed to contain Trp chromophores at a single *f* position, *f*19.

2.1.4 Tetramer core mutant design

2.1.4.1 Design of CC-Tet core mutant

The peptide CC-Tet-IA was designed as a core mutant of the homotetrameric coiled coil CC-Tet [208] where two Ile residues were mutated to Ala at positions *d*13 and *d*20. The sequence was otherwise identical to the parent, CC-Tet.

2.1.4.2 Design of 1-LI-AB core mutant

The peptides 1-LI-A* and 1-LI-B* were designed as core mutants of the acidic and basic peptides 1-LI-A, and 1-LI-B, respectively. Both peptides contained Leu to Ala and Ile to Ala mutations in their central heptads (positions *a*14 and *d*17). The sequences were otherwise identical to the parent peptides.

2.1.4.3 Design of M2-like ELAEIK core mutants

The peptides ELAEIK-M2-A*a*, ELAEIK-M2-A*d* and ELAEIK-M2-A*a*-HI were based on the sequence for the peptide ELAEIK (Section 2.1.1.4) and the sequence of the 4-helix bundle region of the M2 protein from Influenza B, which binds the drug amantadine. The amantadine-contacting residues were identified from X-ray crystallographic and solution phase NMR structures of the M2 4-helix bundle solved in the presence of amantadine [271,272]. This gave the amantadine binding motif: VxxASxxAxxH. The peptides ELAEIK-M2-A*a* and ELAEIK-M2-A*d* contained this motif superimposed onto the parent ELAEIK sequence, with the motif designed to start at an *a* or *d* position, respectively (*a*7, *d*10). The peptide ELAEIK-M2-A*a*-HI was identical in sequence to ELAEIK-M2-A*a* except it contained a His to Ile mutation at position *a*17.

2.2 Peptide synthesis and purification

2.2.1 Automated solid-phase peptide synthesis

Peptides were synthesised on a 0.1 mmol scale by solid-phase peptide synthesis on ChemMatrix Rink amide resin (PCAS BioMatrix) or Novabiochem Rink amide resin (Merck) on a Liberty Blue automated microwave peptide synthesiser (CEM) using Fmoc-protected amino acids (supplied by Novabiochem or Cambridge

Reagents) at 0.2 M in dimethylformamide (DMF, Cambridge Reagents). Deprotection was performed with 20 % (v/v) morpholine (Alfa Aesar) in DMF or with 20 % morpholine and 5 % formic acid in DMF for peptides ABCD7 and ABCD8 to suppress aspartimide formation. Coupling was performed using 0.5 M 6-Chloro-1-hydroxybenzotriazole (Cambridge Reagents) in DMF as the activator and 1.0 M N,N′-Diisopropylcarbodiimide (DIC, Acros Organics) in DMF as the activator base. Following synthesis, peptides were N-terminally acetylated for 20 min at room temperature using approximately 1 % (v/v) acetic anhydride and approximately 1 % (v/v) pyridine in 1:1 DMF/dichloromethane (DCM, Sigma Aldrich) or 100 % DMF. Peptides were simultaneously cleaved from the resin and side chain deprotected for 2 h at room temperature using a mixture of 5 % (v/v) H_2O and 5 % (v/v) triisopropylsaline (TIPS, Acros Organics) in trifluoroacetic acid (TFA, Acros Organics). Resin was removed by filtration and crude peptides were precipitated in cold diethyl ether (Honeywell Research Chemicals) then isolated by centrifugation at 4 °C. Peptides were re-suspended in 1:1 ultrapure Milli-Q water/acetonitrile then lyophilised prior to purification.

2.2.2 Reversed-phase HPLC purification

Peptides were purified by reversed-phase high-performance liquid chromatography (HPLC) using C18 reversed phase columns (150 x 10 mm, 100 Å pore size, Phenomenex). A flow rate of 3 $mL.min^{-1}$ was used. Different linear gradients of buffer A (ultrapure Milli-Q water containing 0.1 % (v/v) TFA) and buffer B (acetonitrile containing 0.1 % (v/v) TFA) were used over different time periods, depending on the peptide. Typically, homomeric peptides and basic peptides were purified with a linear gradient of 20–80 % buffer B over 30 min at room temperature. Acidic peptides were purified with a linear gradient of 30–80 % buffer B over 50 min at 50 °C (using a water bath or column oven). Columns were equilibrated at the starting buffer conditions and washed with 95 % buffer B between gradients. Products were collected manually as fractions of approximately 0.25–1 mL.

Following peptide mass determination (see Section 2.2.3), fraction purity was confirmed by reversed-phase HPLC using analytical C18 reversed-phase columns (100 x 4.6 mm, 100 Å pore size, Phenomenex) with linear gradients of buffer A and B. A flow rate of 1 $mL.min^{-1}$ was used. Representative analytical HPLC traces for all peptides are shown in Section 8.2. Selected fractions were pooled and lyophilised prior to biophysical characterisation.

2.2.3 MALDI-TOF mass spectrometry

Peptide masses were confirmed by MALDI-TOF (matrix-assisted laser desorption/ionisation time-of-flight) mass spectrometry with an ultrafleXtreme II mass spectrometer in positive-ion reflector mode (Bruker, UK). Samples were air-dried onto a ground-steel target plate with α-cyano-4-hydroxycinnamic acid matrix (α-CHCA, Fluka Analytical). Representative mass spectra for all peptides are shown in Section 8.2. Theoretical masses are quoted as average masses and were determined using Peptide Synthetics' online Peptide Mass Calculator (Peptide Protein Research Ltd., http://www.peptidesynthetics.co.uk/tools/)

2.3 Biophysical characterisation

2.3.1 General

All biophysical characterisation was performed in phosphate buffered saline (PBS; 8.2 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.4) unless otherwise stated. Chemicals and solvents were obtained from Fisher Scientific or Sigma Aldrich unless otherwise stated.

2.3.2 Peptide concentration determination

Peptides were dissolved in ultrapure Milli-Q water and concentrations were measured by UV absorbance using a NanoDrop 2000 spectrophotometer (Thermo Scientific). For all peptides, concentration was determined using the extinction coefficient for the Trp chromophore at 278 nm (ε_{278}(Trp) = 5579 $M^{-1}.cm^{-1}$) except peptides ABCD5 and ABCD7 where concentration was determined using the extinction coefficient for the Tyr chromophore at 274 nm (ε_{274}(Tyr) = 1420 $M^{-1}.cm^{-1}$).

2.3.3 Circular dichroism spectroscopy

Circular dichroism (CD) spectroscopy was performed using either a JASCO J-810 or a JASCO J-815 spectropolarimeter with a Peltier temperature controller (Jasco, UK). Peptide solutions were prepared at varying concentrations in buffer by dilution of stock solutions in ultrapure Milli-Q water followed by addition of a 10X PBS stock or 2X sodium phosphate (± NaCl) stock to give a 1X buffer concentration. Samples were analysed in quartz cuvettes with variable pathlengths depending on peptide concentration (10 mm: 1–8 μM total peptide concentration, 2000–4000 μL total sample volume; 5 mm: 5–30 μM total peptide concentration, 1000–2000 μL total

sample volume; 1 mm: 50–200 µM total peptide concentration, 200–400 µL total sample volume; 0.1 mm: 750–1000 µM total peptide concentration, 50–100 µL total sample volume). Full spectra were measured between 190 and 260 nm with a 1 nm step size, 100 $nm.min^{-1}$ scanning speed, 1 nm bandwidth and 1 second response time except where very low peptide concentrations (1–2 µM) were used, in which case a scanning speed of 50 nm/min was used and the remaining parameters were kept the same. Spectra were measured at 5 °C or 20 °C unless otherwise stated. Variable temperature experiments were performed by heating and cooling samples 5–95–5 °C at a rate of 40 $°C.h^{-1}$ whilst monitoring CD at 222 nm at 0.5 °C intervals. For peptide 1-LI-B, full spectra (see above for parameters) were measured at 5 °C intervals while heating 5–95 °C. Data was buffer subtracted then CD (mdeg) was converted to mean residue ellipticity (MRE, $deg.cm^2.dmol^{-1}.res^{-1}$) by normalising for peptide concentration, number of amide bonds and cuvette pathlength, according to Equation 2-1. Where triplicate measurements were taken these were averaged and errors were calculated as one s.d. from the mean. Fraction helix (%) was calculated using Equation 2-2 [295]. Representative spectra and thermal denaturation measurements for all peptides are shown in Sections 8.3 and 8.4.

$$MRE = \frac{CD}{10 \times c \times l \times n}$$

Equation 2-1 Equation for the conversion of CD (mdeg) to MRE ($deg.cm^2.dmol^{-1}.res^{-1}$). *c*, molar peptide concentration; *l*, sample path length in cm; *n*, number of amide bonds in sample (including C-terminal amide following cleavage from Rink amide resin).

$$\text{Fraction helix (\%)} = 100 \times \frac{MRE_{222} - MRE_{coil}}{-42500 \times \left(1 - {}^{3}/_{n}\right) - MRE_{coil}}$$

Equation 2-2 Equation for calculating fraction helix (%) from MRE at 222 nm (MRE_{222}, $deg.cm^2.dmol^{-1}.res^{-1}$). MRE_{coil} = 640-45T; T, temperature (°C); *n*, number of amide bonds in sample (including C-terminal amide).

2.3.4 Sedimentation velocity analytical ultracentrifugation

Sedimentation velocity (SV) experiments were conducted at 20 °C in a Beckman-Optima XL-I or XL-A analytical ultracentrifuge using an An-50 or An-60 Ti rotor. Solutions were prepared in buffer at a total peptide concentration of 140 µM except ABCD5678 (which contains peptides with a mix of Trp and Tyr chromophores)

where a total peptide concentration of 220 μM was used in order to achieve $A_{280} \approx$ 0.95, which is generally required for successful SV data fitting. Samples were prepared either at a volume of 305 μL and loaded into an SV cell with a 12 mm graphite-filled centrepiece and quartz windows, or at a volume of 410 μL and loaded into an SV cell with a 12 mm aluminium centrepiece and sapphire windows. The reference channels of the graphite and aluminium centrepieces were loaded with 320 μL or 420 μL of buffer, respectively. All measurements were performed in 1X PBS (pH 7.4; density, ρ = 1.0054 g/cm2; viscosity, η = 0.01002 P), 10 mM sodium phosphate (pH 7.4; ρ = 0.9996 g/cm2; η = 0.01005 P) or 10 mM sodium phosphate with 1 M NaCl (pH 7.4; ρ = 1.0399 g/cm2; η = 0.01101 P). Following temperature equilibration to 20 °C for approximately 1.5 h at 3 krpm, samples were spun at 50 krpm. A total of 120 absorbance scans at 280 nm over a radial range of approximately 5.8 to 7.3 cm were measured at 5 min intervals. Data were fitted to a continuous c(s) distribution model using Sedfit, at 99% confidence level [296]. The baseline, meniscus, frictional ratio (f/f_0), and systematic time-invariant and radial-invariant noise were fitted. The partial specific volume ($\bar{v}$) for each peptide/peptide combination was calculated using Sedfit [296]. The buffer densities were calculated using SEDNTERP. Continuous c(s) distributions and residuals for all peptides/peptide combinations are shown in Section 8.5. For all SV experiments, residuals are shown below the c(s) distribution as a rectangular greyscale bitmap where the shade of grey indicates the magnitude of the difference between the data and the fit (mid-grey indicates residual value ≈ 0, *i.e.* agreement between data and fit; lighter shades indicate positive residual values and darker shades indicate negative residual values, *i.e.* deviations between data and fit). Each residuals bitmap displays the residuals for every scan, stacked from top to bottom, across the entire radial range included in the data processing, where the left and right edges of the bitmap represent the data boundaries at the meniscus and cell bottom, respectively. A homogenous grey bitmap indicates a good fit (and little noise).

2.3.5 Sedimentation equilibrium analytical ultracentrifugation

Sedimentation equilibrium (SE) experiments were performed at 20 °C in a Beckman-Optima XL-I or XL-A analytical ultracentrifuge using an An-50 Ti rotor and 12 mm six-channel epon-charcoal equilibrium cells with quartz windows. Solutions were prepared in buffer at a total peptide concentration of 70 μM except ABCD5678 where a total peptide concentration of 110 μM was used in order to

achieve $A_{280} \approx 0.5$, which is generally required for successful SE data fitting. All measurements were performed in buffers as described in Section 2.3.4. Samples were spun at speeds ranging from 15–42 krpm. Absorbance scans at 280 nm were measured at 3 or 4 krpm intervals following an initial equilibration period of 8 h and a second equilibration period of 1 h. Data sets were initially fitted to a single, ideal species model using Ultrascan II (http://www.ultrascan.uthscsa. edu/). 99 % confidence limits were calculated using Monte Carlo analysis of the obtained fits. The $\bar{v}$ for each peptide/peptide combination and the buffer densities were calculated using Sedfit [296].

SE experiments were performed on individual acidic and basic peptides at 20 °C in a Beckman-Optima XL-I or XL-A analytical ultracentrifuge using an An-60 Ti rotor. Solutions at a total peptide concentration of 70 µM were prepared in 1X PBS at a volume of 110 µL and loaded into an SV cell with a 12 mm graphite-filled centrepiece and quartz windows. The reference channels were loaded with 120 µL of buffer and samples were spun at speeds ranging from 45 to 60 krpm. Absorbance scans at 280 nm were measured at 3 krpm intervals following an initial equilibration period of 8 h and a second equilibration period of 1 h. Data sets were fitted as described above. All SE data is shown in Section 8.5, where representative data, fitted curves and residuals at three different rotor speeds are plotted in different colours.

2.3.6 Peptide crystallography

Peptides were crystallised at 20 °C using the sitting drop method. Solutions of peptides 2-LI-EK, 3-LI-EK and ELAEIK were prepared at 3.0 mM in unbuffered deionised water. Screening was carried out using the standard commercial screens JCSG-*plus*™, Morpheus®, Structure Screen 1 & 2 and PACT *premier*™ (Molecular Dimensions). Screens were prepared in 96-well MRC plates using an Oryx8000 Protein Crystallisation Robot (Douglas Instruments) with reservoir volumes of 50 µL. Droplets contained 0.3 µL peptide solution equilibrated with 0.3 µL of reservoir solution. Final crystallisation conditions for all peptides can be found in Table 8-2.

Single crystals were cryoprotected by soaking in 25 % (v/v) glycerol in ultrapure Milli-Q water and reservoir solution, prior to loop-mounting and plunge-freezing in N_2(l). Data collection was carried out under cryogenic conditions at wavelengths of 0.93 to 0.98 Å on beamline i03, i04 or i04-1 at Diamond Light Source (Didcot, UK).

Data indexing and integration were carried out in MOSFLM [297] and scaling was carried out in AIMLESS [298]. Initial phases for the peptides ELAEIK and 2-LI-EK were determined by molecular replacement with Phaser-MR [299] using full or partial poly-alanine parallel tetrameric coiled-coil models (as predicted by the Matthews coefficient [300]) generated in ISAMBARD or CC-Builder as the search models [293,301]. The phase for peptide 3-LI-EK was generated using ARCIMBOLDO [302]. AIMLESS, Phaser-MR and ARCIMBOLDO were all used as implemented in CCP4 [303]. Final models were obtained through iterative rounds of model building in COOT and refinement in PHENIX Refine [304,305]. Data collection and refinement statistics can be found in Table 8-3, Table 8-4 and Table 8-5.

All protein structure images were generated in PyMol (http://www.pymol.org).

2.3.7 DPH-binding assay

Binding of the hydrophobic dye 1,6-Diphenyl-1,3,5-hexatriene (DPH) to the peptides CC-Tet-IA and CC-Tet was assayed by measuring the fluorescence of samples containing DPH and different concentrations of peptides. DPH fluoresces strongly at 452 nm when present in an apolar environment, such as a lipid bilayer or a protein's hydrophobic core, where it becomes fluorescently polarized.

A DPH stock solution was prepared in 100 % dimethyl sulfoxide (DMSO) at a concentration of approximately 1 mM. DPH concentration was verified using absorbance at 350 nm (ε_{350}(DPH) = 88000 $M^{-1}.cm^{-1}$). Subsequently, a 10 µM stock of DPH in 1:1 DMSO/ultrapure water was prepared. Peptide stocks were prepared at 80 and 800 µM in ultrapure Milli-Q water.

Samples containing 1 µM DPH and varying peptide concentrations were prepared in 1X PBS in black Nunc 96-Well Flat Bottom plates (Thermo Fisher Scientific). Peptide concentrations were 0, 10, 20, 50, 75, 100, 150, 200, 250 and 500 µM. Samples were prepared in triplicate and incubated in the dark for 2 h at RT with gentle rocking. DPH fluorescence was measured on a CLARIOstar microplate reader (BMG LABTECH) with excitation at 350 nm and emission measured between 380 and 600 nm. DPH fluorescence at 452 nm was normalised to maximal fluorescence then normalised triplicate experiments were averaged. Coiled coil concentration (c) was determined as peptide concentration/oligomeric state. DPH fluorescence at 452 nm/maximal fluorescence at 452 nm measured for that peptide (F452/Fmax452) was plotted against c.

Buffer	Components
50X TAE	2.0 M Tris, 1.0 M acetate, 0.05 M EDTA
5X TBE	0.45 M Tris, 0.44 M borate, 0.01 M EDTA
10X PBS	100 mM Na_2HPO_4, 18.0 mM K_2HPO_4, 1.37 M NaCl, 27 mM KCl, pH 7.4
10X Tris-Gly	250 mM Tris, 2 M glycine
10X PAGE stacking buffer	1.25 M Tris, pH 6.8
2X PAGE resolving buffer	0.75 Tris, pH 8.3
2X SDS loading buffer	100 mM Tris, 4 % (w/v) SDS, 22.5 %(v/v) glycerol, 2.5 $mg.mL^{-1}$ bromophenol blue, 200 mM DTT
1X Tris-Gly-SDS PAGE running buffer	25 mM Tris, 200 mM glycine, 0.05% (w/v) SDS
1X Western blot transfer buffer	25 mM Tris, 200 mM glycine, 10 % (v/v) methanol
1X Western blot wash buffer	10 mM Na_2HPO_4, 1.8 mM K_2HPO_4, 137 mM NaCl, 2.7 mM KCl, 0.05 % (v/v) Tween® 20

Table 2-1 Stock buffer components. All buffers are used at 1X dilution unless otherwise stated. SDS, sodium dodecyl sulfate; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid.

2.4 Molecular biology

2.4.1 Buffers, solutions and cell culture medias

Chemicals were obtained from Sigma-Aldrich, Fisher Scientific or Bio-Rad unless otherwise stated. All buffers, buffer components and components of cell culture media were prepared using distilled water (Type II, Elix, Millipore) unless otherwise stated.

2.4.1.1 Buffers and solutions

Buffer components are listed in Table 2-1. Where appropriate, solutions were pH-adjusted using a Jenway 3150 pH meter.

2.4.1.2 Cell culture medias

All components were sterilised prior to use by autoclaving at 121 °C for 15 min or by filter-sterilisation using 0.22 µm filters (Sartorius Stedim Biotech). Following autoclaving/filtration, all components were handled under sterile conditions.

Lysogeny broth (LB) media was prepared from pre-weighed sachets and LB agar was prepared from pre-weighed capsules (both from MP Biomedicals).

M9 minimal media (42.3 mM Na_2HPO_4, 21.6 mM KH_2PO_4, 18.7 mM NH_4Cl, 8.56 mM NaCl) was supplemented with 0.2 % (w/v) casamino acids, 10 mM $CaCl_2$, 0.25 % (v/v) glycerol, 2 mM $MgSO_4$, 2 µg.mL^{-1} thiamine. Casamino acids (MP Biomedicals) were prepared as a 20 % (w/v) solution. $CaCl_2$ was prepared as a 0.1 M solution. Glycerol was prepared as a 50 % (v/v) solution. $MgSO_4$ was prepared as a 1 M solution. Thiamine was prepared as a 20 mg.mL^{-1} solution and filter-sterilised.

Where antibiotic selection was required, ampicillin, chloramphenicol and/or kanamycin were used at final concentrations of 100 µg.mL^{-1}, 25 µg.mL^{-1} and 50 µg.mL^{-1}, respectively. All antibiotics were prepared as 1000X stocks and filter-sterilised. Chloramphenicol solutions were prepared in 100 % ethanol.

Where appropriate, media was supplemented with 0.2 % (w/v) glucose to suppress expression from the P_{araBAD} promoter or 0.0002–0.2 % (w/v) arabinose to induce expression from this promoter. Glucose solutions were prepared as a 20 % (w/v) stock and filter-sterilised. Arabinose solutions were prepared as 20 % (w/v) or 2.0 % (w/v) stocks and filter-sterilised.

2.4.2 Cloning and plasmid preparation

All cloning took place in XL1-Blue *Escherichia coli* cells (*rec*A1, *end*A1, *gyr*A96(nal^R), *thi*-, *hsd*R17(r_K^-, m_K^+), *sup*E44, *rel*A1, *lac*, [F′, *pro*A$^+$B$^+$, *lacI*^{q}ZΔM15, ::Tn10(tet^R)]). Cells were grown on LB agar in static incubators or in LB media with shaking at 250 rpm at 37 °C. All restriction enzymes were provided by New England Biolabs unless otherwise stated. All restriction digests were carried out at 37 °C unless otherwise stated. All enzymes were used in the recommended buffers provided by the supplier unless otherwise stated. Where appropriate, enzymatic reactions were diluted with ultrapure Milli-Q water. Oligonucleotides were provided by Eurofins Genomics. Peptide genes and promoter DNAs were purchased as GeneArt Strings (Thermo Fisher Scientific). Sanger sequencing was provided by the Eurofins Genomics TubeSeq service.

2.4.2.1 Preparation of chemically-competent E. coli *cells*

E. coli cells were streaked onto un-supplemented LB agar plates and incubated overnight at 37 °C. The following day, colonies were selected for incubation overnight in un-supplemented LB media at 37 °C. The following day, the overnight cultures were diluted 100-fold into 50 mL fresh un-supplemented LB media and

grown to OD_{600} = 0.3–0.8. Cells were then isolated by centrifugation (10 min, 3600 rpm, 4 °C), resuspended in 20 mL cold (4 °C) 0.1 M $CaCl_2$ and incubated on ice for 20 min. Cells were then isolated again by centrifugation, resuspended in 3 mL cold 0.1 M $CaCl_2$ with 2 mL cold 50 % glycerol and incubated on ice for 1 h. Following this incubation, 0.5 mL aliquots were transferred to pre-chilled 1.5 mL tubes and flash-frozen in N_2(l) prior to storage at -80 °C. Following preparation, competent cells were handled on ice.

2.4.2.2 Transformation into competent E. coli *cells*

Where the DNA to be transformed was a purified plasmid product, 0.5 µL DNA was used. Where the DNA to be transformed was a ligation product, the entire 20–30 µL ligation reaction was used.

The DNA was pre-chilled in 1.5 mL tubes while competent cells were thawed on ice. Then, DNA was mixed with 100 µL competent cells and incubated on ice for 20–40 min. Cells were subjected to heat shock at 42 °C for 90 sec then returned to ice for 60 sec. Following the addition of 400 µL LB media, cells were incubated at 37 °C with shaking for 30-60 min. Following this recovery period, cells were plated on pre-dried agar plates (dried for approximately 1 h, 37 °C). Where one plasmid was transformed, or two plasmids were co-transformed, 150 µL of cells were plated. Where three or more plasmids were co-transformed, or a ligation product was transformed, cells were recovered by centrifugation, resuspended in approximately 150–200 µL media then plated. Cells were spread using 10–20 ColiRollers™ Plating Beads (Novagen) and incubated inverted at 37 °C overnight. Following incubation, parafilm-wrapped agar plates were stored at 4 °C for up to 2 weeks.

2.4.2.3 Purification of plasmid DNA from E. coli *cells*

Plasmids were purified from 5 mL overnight cultures of transformed XL1-Blue *E. coli* cells using a QIAprep® Spin Miniprep kit (QIAGEN) according to the manufacturer's instructions.

Where a larger amount of DNA was required, DNA was purified from 100 mL overnight cultures of transformed XL1-Blue *E. coli* cells using a HiSpeed® Plasmid Midi kit (QIAGEN) according to the manufacturer's instructions up to and including the DNA precipitation step (3.5 mL isopropanol, 5 min, RT). The DNA was then pelleted (60 min, 5000 rpm, 4°C) and washed with 5 mL 70 % (v/v) ethanol. The

DNA was then pelleted again (30–60 min, 5000 rpm, 4°C), air dried, and resuspended in 100-200 µL TE buffer (QIAGEN).

All purified DNA was stored at -20 °C. DNA concentrations were determined by UV absorbance using a DeNovix microvolume DS-11 spectrophotometer (ε_{260}(dsDNA) = 50 $M^{-1}.cm^{-1}$).

2.4.2.4 Agarose gel electrophoresis

A solution of 1–2 % (w/v) molten agarose was prepared in 1X TAE. Subsequently, 0.005 % (v/v) ethidium bromide was added to the solution then gels were poured in approximately 12.5 x 17 cm gel trays and allowed to set for 1 h prior to use. Gels were run in 1X TAE containing 0.005 % (v/v) ethidium bromide at 120 V for 1–2 h. Ethidium bromide-stained DNA was visualised by UV illumination using either a UV transilluminator or a GelDoc XR+ gel documentation system (BioRad). DNA samples were prepared in 5X GelPilot® DNA Loading Dye (QIAGEN) or 6X Purple Gel Loading Dye (NEB).

2.4.2.5 Non-denaturing polyacrylamide gel electrophoresis

Non-denaturing polyacrylamide gel electrophoresis (PAGE) was used to verify plasmid products prior to sanger sequencing. Restriction digest products were analysed on 7.5 % acrylamide gels and compared to DNA standards of known size to confirm the presence of the correct restricted fragments.

A solution containing 7.5 % (w/v) acrylamide, 1X TBE, 0.05 % (w/v) ammonium persulfate and 0.1 % (v/v) tetramethylethylenediamine was prepared in deionised water. Ammonium persulfate was prepared as a 10 % (w/v) stock solution. A stock solution of 30 % (w/v) acrylamide was used (37.5:1 bisacrylamide, Severn Biotech). Gels were poured between 1.5 mm glass plates and allowed to polymerise to 1 h prior to use. Gels were run in 1X TBE at 120 V for 45–60 min then stained for 10 min in a solution of 0.005 % (v/v) ethidium bromide. Stained gels were imaged as outlined above. DNA samples were prepared as outlined above.

2.4.2.6 Polymerase chain reaction

All polymerase chain reactions (PCRs) were performed in a Techgene thermal cycler (Techne) using Pfu polymerase (Promega). PCR reactions were typically performed in a volume of 50 µL containing 20 ng template, 1 µM PCR-grade

deoxynucleoside triphosphates (dNTPs, Invitrogen) and 0.4 µM primer(s) in 1X Pfu polymerase buffer. All reactions included a single 5 min denaturation cycle at 98 °C (after which 1 µL Pfu was added) followed by 30 cycles of denaturation (1 min, 94 °C), annealing (1 min, variable temperature) and extension (variable time, 72 °C). Extension time was determined as approximately 30 sec.kbp^{-1}. All reactions were completed with a final extension cycle at 72 °C.

2.4.2.7 Plasmid construction

The plasmids pVRb-LacUV5-SFGFP, pVRb-O1O1-LacUV5-SFGFP, pBAD-LacI-WT, pBAD-LacI-L251A-WTtet, pBAD-LacI-L251, pBAD-LacI-L251A, pVRc-LacI-L251A and pBAD-LacI-L251A-ccDi were provided by Abigail Smith. The plasmid pE-SUMOpro-Kan-LbCpf1 was provided by Kara van Aelst. All plasmids used in this work are listed in Table 8-6 and representative plasmid maps are shown in Section 8.7. All primers are listed in Table 2-2. Primers were resuspended in ultrapure Milli-Q water at 100 µM then prepared for use as 10 µM stocks. All stocks were stored at -20 °C.

Peptide gene design. DNA sequences for A1 (1-LI-A), A2 (2-LV-A), A3 (3-LV-A), B1 (1-LI-B), B2 (2-LV-B) and B3 (3-LV-B) were manually designed to reduce repetitive sequences and optimise for *E. coli* codon usage (Table 2-3). DNA sequences also contained 5′ XbaI and 3′ Acc65I restriction sites and flanking 5′ and 3′ buffer sequences. Synthesised DNA was resuspended in ultrapure Milli-Q water at a concentration of 20 ng.µL^{-1} and stored at -20 °C.

pBAD-LacI-L251A-A/B and pVRc-LacI-L251A-A/B plasmids. The plasmids pVRc-LacI-L251A-A1, pVRc-LacI-L251A-A2, pVRc-LacI-L251A-A3, pVRc-LacI-L251A-B1, pBAD-LacI-L251A-B1, pBAD-LacI-L251A-B2, pBAD-LacI-L251A-B3 and pBAD-LacI-L251A-A1 were prepared by inserting the peptide genes into pBAD and pVRc plasmids carrying a truncated version of the Lac repressor (residues 1–332) lacking its C-terminal tetramerisation domain (residues 333-360) and containing a point mutation, L251A. These plasmids also contain XbaI and Acc65I restriction sites for the introduction of novel oligomerisation domains at the Lac repressor C terminus. Insert DNA was digested with XbaI and Acc65I in NEBuffer 3 and products were purified from the restriction digest mixture using a QIAEX II gel extraction kit (QIAGEN) following instructions for desalting and concentrating DNA solutions. Alternately, insert DNA was extracted from existing plasmids by restriction digest, fragments were resolved by agarose gel electrophoresis then

No.	Sequence	Length (bases)
1	AGACCGCTTCTGCGTTC	17
2	GCATGACTGGTGGACAGC	18
3	*GGCATACTCTGCGACATCG	19
4	*TCGTTGCTGATTGGCGTTGCCACC	24
5	*AGATCTCGAGCTCGGATCC	19
6	*GCTAGCCATACCATGATGATG	21
7	*CCCGGTCTAGAGGGTAGTG	19
8	GATCAGAGCTAGCCTGCAGGACTCAGAAGTCAATC	35
9	GACTACATCTAGACCCATACCTCCAATCTGTTCGCGG	37
10	*GGGAGTTAGCGGCTATCG	18
11	*ACTCGGCTGCGGCCGCTTCTTTTTCAATCGCTGCCAG	37
12	*GGAAGTTAGCAGCCATCAAAC	21
13	*ACTTGGCAGCCGCCGCCTTTTGTTTAATCGCGGCCAATTTC	41
14	*CGGGATCGCACATGAGCTGTCTTCGG	26
15	*CGGGATATCACATGAGCTGTCTTCG	25
16	*CCGTTAACCACCATCAAACAG	21
17	CTAGCTCGCCATGGTTAATTCC	22
18	AGTCATAGACTAGTGCCACAGCTAACAC	28
19	GTAGCCGACTAGTTGTCATAATTG	24
20	GGGCTAACAGGAGGAATTAACC	22

Table 2-2 List of PCR and sequencing primers used in this work. * indicates primer was 5′ phosphorylated.

appropriate fragments were purified using a QIAquick gel extraction kit (QIAGEN) or using a QIAEX II gel extraction kit (QIAGEN) following instructions for agarose gel extraction. Plasmid DNA was digested as above then treated with 3 µL calf intestinal alkaline phosphatase (Promega) which was added directly to the restriction digest reaction and diluted with ultrapure Milli-Q water to maintain a glycerol concentration <5 % (v/v). Plasmid fragments were then resolved by agarose gel electrophoresis and the appropriate fragment was purified using a QIAquick gel extraction kit (QIAGEN). Fragments were ligated using 1 µL T4 DNA ligase (Promega) in a 20 µL reaction at 4 or 20 °C for 4–16 h. Ligation products were transformed into XL1-Blue *E. coli*. Colonies containing successful ligations were selected and grown overnight and resulting plasmids were purified. Successful insertions were confirmed by restriction digest (XbaI/Acc65I in NEBuffer 3) and sequencing using primers 1 and 2.

pBAD-LacI-(L251A)-ΔHTH plasmids. The plasmids pBAD-LacI-ΔHTH and pBAD-LacI-L251A-ΔHTH were prepared using PCR to delete residues 14–60 (inclusive) from the Lac repressor gene in the plasmids pBAD-LacI-L251 (does not contain L251A point mutation) and pBAD-LacI-L251A. Primers 3 and 4 were used with an annealing temperature of 50 °C and an extension time of 10 min. PCR

Gene	Sequence
A1	AGAAAAACCACCCTGGCGCCCGGTCTAGAGGGCAGC**GGTGCCATTGAGAAAG** **AGCTGGCAGCGATTGAAAAAGAATTGGCCGCAATCGAGTGGGAGTTAGCGGC** **TATCGAAAAGGAACTGGCTGGT**TAGGTACCAAGCTTGGCTGTTTTGGCGGA
A2	AGAAAAACCACCCTGGCGCCCGCTCTAGAGGGCAGC**GGTGAAGTTGAAAAGC** **AGTTAGCCGAGGTCGAGAAACAATTGGCAGAAGTAGAGTGGCAGCTGGCGGA** **GGTGGAGAAGCAACTGGCTGGT**TAGGTACCAAGCTTGGCTGTTTTGGCGGA
A3	AGAAAAACCACCCTGGCGCCCGCTCTAGAGGGCAGC**GGTGCTGTTCAGAAAG** **AACTTGAAGCCGTGCAGAAGGAACTGGAGGCCGTCCAGTGGGAGCTGGAAGC** **GGTGCAAAAAGAGCTTGAGGGC**TAGGTACCAAGCTTGGCTGTTTTGGCGGA
B1	AGAAAAACCACCCTGGCGCCCGCTCTAGAGGGTTCT**GGCGCAATTAAGCAGA** **AATTGGCCGCGATTAAACAAAAGCTGGCGGCTATCAAGTGGAAGTTAGCAGC** **CATCAAACAGAAACTGGCTGGC**TAGGTACCAAGCTTGGCTGTTTTGGCGGA
B2	AGAAAAACCACCCTGGCGCCCGCTCTAGAGGGTTCT**GGCAAGGTGAAGCAAC** **AGTTAGCTAAAGTCAAACAGCAATTGGCAAAGGTTAAATGGCAACTCGCGAA** **AGTAAAGCAGCAGCTGGCCGGC**TAGGTACCAAGCTTGGCTGTTTTGGCGGA
B3	AGAAAAACCACCCTGGCGCCCGCTCTAGAGGGCAGC**GGCGCTGTTCAACAGA** **AACTGAAAGCCGTCCAGCAAAAACTCAAGGCAGTGCAATGGAAGCTGAAGGC** **GGTTCAGCAGAAGCTTAAAGGT**TAGGTACCAAGCTTGGCTGTTTTGGCGGA

Table 2-3 DNA sequences for A1, A2, A3, B1, B2 and B3 genes with peptide-coding region in bold and XbaI (T:CTAGA) and Acc65I (G:GTACC) restriction sites underlined. Restriction enzyme cut site is indicated by :. All sequences are 155 bp.

products were treated with DpnI (1 µL added directly to PCR reaction), resolved by agarose gel electrophoresis, then purified using a QIAquick gel extraction kit (QIAGEN). The linear, blunt-end products were then ligated using 2 µL T4 DNA ligase in a 20 µL reaction at 15 °C for 16 h. Ligation products were transformed into XL1-Blue *E. coli* then plasmids were verified as described above except restriction digests were performed with HindIII/NcoI-HF in NEBuffer 2. The gene for peptide B1 was then inserted into these plasmids as described for the pBAD-LacI-L251A-A/B plasmids.

pBAD-6H-(T7-Xpress) plasmids. The plasmids pBAD-6H-T7-Xpress and pBAD-6H were prepared using PCR to delete the Lac repressor gene or the Lac repressor gene and N-terminal tags, respectively, from the plasmid pBAD-LacI-L251A. The plasmids pBAD-6H-T7-Xpress-B1 and pBAD-6H-B1 were prepared in the same way from the plasmid pBAD-LacI-L251A-B1. For pBAD-6H-T7-Xpress and pBAD-6H-T7-Xpress-B1, primers 5 and 7 were used. For pBAD-6H and pBAD-6H-B1, primers 6 and 7 were used. An annealing temperature of 50 °C and an extension time of 10 min was used. PCR products were then processed as described for the pBAD-LacI-(L251A)-ΔHTH plasmids except sequencing was performed with

primer 1 only. The genes for peptide B2 and B3 were then inserted into pBAD-6H as described for the pBAD-LacI-L251A-A/B plasmids.

pBAD-6H-SUMO plasmids. The plasmids pBAD-6H-SUMO, pBAD-6H-SUMO-B1 and pBAD-6H-SUMO-B2 were prepared by inserting the gene for SUMO into the plasmids pBAD-6H, pBAD-6H-B1 and pBAD-6H-B2, respectively. The SUMO gene was PCR-amplified from the plasmid pE-SUMOpro-Kan-LbCpf1. Primers 8 and 9 were used to introduce a 5′ NheI and 3′ XbaI restriction sites adjacent to the SUMO gene to give PCR product NheI-SUMO-XbaI. An annealing temperature of 55 °C and an extension time of 1 min was used. The PCR products were treated with DpnI, resolved by gel electrophoresis then purified with a QIAquick gel extraction kit (QIAGEN). Inserts and plasmids were digested with NheI/XbaI in NEBuffer 2 then processed, purified and ligated as described for the pBAD-LacI-L251A-A/B plasmids. Plasmids were verified as described for the pBAD-LacI-(L251A)-ΔHTH plasmids except sequencing was performed with primer 1 only. The gene for peptide B3 was inserted into pBAD-6H-SUMO to generate pBAD-6H-SUMO-B3, as described for the pBAD-LacI-L251A-A/B plasmids.

A1/B1 L14A-I17A Point mutations. Site directed mutagenesis was used to introduce simultaneous L14A and I17A point mutations into A1 and B1 genes in the plasmids pBAD-LacI-L251A-B1, pBAD-6H-SUMO-B1 and pVRc-LacI-L251A-A1. Primers 10 and 11 were used for the A1 gene with an annealing temperature of 55 °C and an extension time of 10 min. Primers 12 and 13 were used for the B1 gene with an annealing temperature of 50 °C and an extension time of 10 min. Blunt-end PCR products were then processed, purified and ligated as described for pBAD-LacI-(L251A)-ΔHTH plasmids.

LacI Y282A/D point mutations. Site directed mutagenesis was used to introduce Y282A or Y282D point mutations into the Lac repressor gene in the plasmids pBAD-LacI-L251A-WTtet, pBAD-LacI-L251A, pBAD-LacI-L251A-ccDi, pBAD-LacI-L251A-B1, pVRc-LacI-L251A and pVRc-LacI-L251A-A1. PCR was performed with primers 14 and 16 to introduce the Y282A mutation with an annealing temperature of 55-56 °C and an extension time of 6-10 min. PCR was performed with primers 15 and 16 to introduce the Y282D mutation with an annealing temperature of 55 °C and an extension time of 10 min. Blunt-end PCR products were then processed, purified and ligated as described for pBAD-LacI-(L251A)-ΔHTH plasmids. Alternately, the LacI-L251A-Y282A gene was extracted from existing plasmids by restriction digest, fragments were resolved by agarose gel

Promoter	Sequence
pro1	AGTCATAGACTAGTGC**CACAGCTAACACCACGTCGTCCCTATCTGCTG CCCTAGGTCTATGAGTGGTTGCTGGATAACTTTACGGGCATGCATAAG GCTCGGTATCTATATTCAGGGAGACCACAACGGTTTCCCTCTACAAAT AATTTTGTTTAACTTT**TAACAGGAGGAATTAACCATGGCGAGCTAG
pro5	AGTCATAGACTAGTGC**CACAGCTAACACCACGTCGTCCCTATCTGCTG CCCTAGGTCTATGAGTGGTTGCTGGATAACTTTACGGGCATGCATAAG GCTCGTAGGATATATTCAGGGAGACCACAACGGTTTCCCTCTACAAAT AATTTTGTTTAACTTT**TAACAGGAGGAATTAACCATGGCGAGCTAG

Table 2-4 DNA sequences of constitutive promoters following PCR to introduce SpeI site. The promoter region is in bold and SpeI (A:CTAGT) and NcoI (C:CATGG) restriction sites are underlined. Restriction enzyme cut site is indicated by :. All sequences are 190 bp.

electrophoresis then appropriate fragments were purified using a QIAquick gel extraction kit (QIAGEN). The insert was then ligated into the appropriate backbone.

Constitutive promoter plasmids. *Pro1* and *pro5* constitutive promoter DNA was designed using sequences from Davis *et al* with flanking buffer DNA and a 3′ NcoI restriction site [306]. A 5′ SpeI restriction site was introduced into the promoter DNA using PCR mutation with primers 17 and 18 with an annealing temperature of 57 °C and an extension time of 1 min (Table 2-4). PCR products were then purified using a QIAquick PCR purification kit (QIAGEN). Using primers 19 and 20 the P_{araBAD} promoter was removed from and a SpeI restriction site was introduced into the plasmids pBAD-LacI-L251A, pBAD-LacI-L251A-ccDi, pBAD-6H-SUMO, pBAD-6H-SUMO-B1, pVRc-LacI-L251A, pVRc-LacI-L251A-A1, pVRc-LacI-L251A-Y282A and pVRc-LacI-L251A-Y282A-A1. These PCR products were resolved by agarose gel electrophoresis and purified using a QIAquick gel extraction kit (QIAGEN). Plasmids and inserts were then digested with SpeI/NcoI in Cut Smart buffer. Inserts were purified using a QIAquick PCR purification kit (QIAGEN) and plasmids were processed and purified as described for the pBAD-LacI-L251A-A/B plasmids. Inserts and plasmids were ligated and verified as described for the pBAD-LacI-L251A-A/B plasmids except restriction digests were performed with NdeI/NheI in Cut Smart buffer and sequencing was performed with primers 1 and 3.

2.4.3 GFP-repression assay

TB28 *E. coli* cells (F^-, λ^-, *ilvG*-, *rfb-50*, *rph-1*, Δ*lacIZYA*) [307] were transformed with a pVRb reporter plasmid carrying superfolder green fluorescent protein (GFP) [308] and pBAD and/or pVRc sample plasmids carrying the proteins to be investigated

for their ability to repress GFP expression. Following antibiotic selection on LB agar plates, colonies were selected and grown overnight at 37 °C in 5 mL supplemented M9 media. The following morning, 10 mL fresh supplemented M9 media was inoculated with 50 µL of overnight culture and induced with varying arabinose concentrations (0.0–0.2 % (w/v)). Cultures were grown at 37 °C until OD_{600} = approximately 0.4–0.6. OD_{600} values were recorded to three significant figures using a Lambda Bio UV/Vis spectrometer (PerkinElmer). 5 mL of each culture was pelleted (10 min, 5000 rpm, 4 °C) then re-suspended in 250 µL 1X PBS. Two 100 µL aliquots of each sample were loaded into consecutive wells of a black 96 well polypropylene microplate. GFP fluorescence of each 100 µL sample was measured using a FlexStation Microplate Reader (Molecular Devices) with excitation at 470 nm and emission at 510 nm. GFP/OD_{600} was calculated for each 100 µL sample then these duplicate measurements were averaged. Next, the four technical replicates were averaged. Mean values were plotted on a bar chart with error bars as one s.d. from the mean. Fold-repression values were calculated relative to the GFP-only control unless otherwise stated.

2.4.4 Western blotting

Western blotting was performed on TB28 *E. coli* whole cell lysates following GFP assays (co-transformed with all assay plasmids) or on TB28 *E. coli* whole cell lysates of cells transformed separately with plasmids carrying each protein of interest. Where transformed separately, cells were grown under “assay-like” conditions in 10 mL supplemented M9 minimal media at 37 °C until OD_{600} = 0.4–0.6. Cells were induced with 0.0–0.2 % arabinose prior to growth.

Following growth, 1–3 mL of each culture was pelleted. Pellets were re-suspended in 2X SDS loading buffer according to Equation 2-3 for induced cells (0.0002–0.2 % arabinose) and according to Equation 2-4 for non-induced cells (0.0 % arabinose).

$$\text{Vol. loading buffer} = \frac{\text{Vol. pelleted} \times OD_{600}}{10}$$

Equation 2-3 Equation for determining the volume of 2X SDS loading buffer with which to resuspend cells induced with arabinose. Vol., volume (µL).

$$\text{Vol. loading buffer} = \frac{(\text{Vol. pelleted}/3) \times OD_{600}}{10}$$

Equation 2-4 Equation for determining the volume of 2X SDS loading buffer with which to resuspend uninduced cells. Vol., volume (µL).

Samples were denatured at 90–95 °C for 10 min then resolved on 4–15 % precast MiniPROTEAN TGX gels (BioRad) in 1X Tris-Gly-SDS at 175 V for 35 min. Resolved proteins were then transferred to 0.45 µm Immoblion®-P PVDF (polyvinylidene fluoride) Membrane (Millipore) at 4 °C at 230 mA in 1X Western blot transfer buffer for 2 h. Following transfer, membranes were fixed in methanol then blocked overnight at room temperature in 10–15 mL 10 % (w/v) skimmed milk in 1X PBS. The following day, membranes were probed with primary anti-His antibody (mouse, Abcam) at a final concentration of 0.2 $\mu g.mL^{-1}$ in 10 mL 10 % (w/v) skimmed milk in 1X PBS, then secondary anti-mouse IgG HRP-conjugated antibody (goat, Abcam) at a final concentration of 0.8 $\mu g.mL^{-1}$ in 10 mL 10 % (w/v) skimmed milk in 1X PBS at room temperature for 2h, with five 5 min washes with 1X Western blot wash buffer between antibody incubations. After a final 5 min wash with 1X PBS, detection was carried out using either BM Chemiluminescence Western Blotting Substrate (POD, Roche) or Amersham ECL Western Blotting Detection Reagent (GE Healthcare). Membranes were imaged using an Amersham Imager 600 (GE Healthcare) in chemiluminescence mode using manually set exposure times (10–40 sec for induced samples) Alternatively, Amersham Hyperfilm™ MP autoradiography film (GE Healthcare) was exposed to membranes for approximately 10–20 sec for induced samples and approximately 30–40 min for non-induced samples. Exposed autoradiography films were developed using a Curix 60 X-ray film developer (Agfa HealthCare). Digitised blot images were processed using ImageJ to quantify band density as a percentage of all quantified bands [309]. Percentage densities were calculated as relative densities compared to a reference band then averaged. Error bars were plotted as one s.d. from the mean. P values were calculated from T scores using an online P value from T score calculator (Social Science Statistics, http://www.socscistatistics.com/pvalues/tdistribution.aspx).

Chapter 3: The design and characterisation of homotetrameric coiled coils

3.1 Chapter introduction

Tetrameric coiled coils make up less than 4 % of all known coiled coil structures [214]. Conversely, more than 86 % of known coiled-coil structures form parallel or antiparallel dimers. As such, it is perhaps unsurprising that many *de novo* coiled-coil design attempts have focussed on dimeric targets [207,214]. While the study and design of dimers has increased our understanding of protein sequence-structure relationships and led to useful applications [310-313], coiled coils with larger oligomeric states are also important design targets. For example, tetramers offer many advantages over components with lower oligomeric states when the aim is to produce modular protein components that may be of use in synthetic biology, nanotechnology or materials science. As introduced in Chapter 1, tetramers are particularly useful in these contexts because: (1) tetramers display more binary-like characteristics than lower oligomeric states making them more useful as switches; (2) they can co-localise more proteins of interest; and (3) they are more stable than lower order coiled coils of comparable length, making them more tolerant to modifications that expand their function. In this work, the *de novo* coiled coils are primarily intended for use as PIDs in artificial transcription factors. Many naturally occurring transcription factors rely on tetramerisation for function, including the Lac repressor in bacteria and p53 in humans [314,315].

To date, most attempts at designing novel tetrameric coiled coils have centred around the redesign of naturally occurring coiled coils, for example through the redesign of their cores or through the optimisation of core-flanking ionic residues [187,205,206,223,316]. Furthermore, there has been much work in the past on designing 4-helix bundles, which do not necessarily contain the extensive knobs-into-holes (KIH) interactions that characterise coiled coils [317,318]. These have used both computational and empirical approaches [42,234,319,320]. While truly *de novo* tetrameric coiled coils are generally rare, studies of this nature have allowed rules to be elucidated that have enabled the design of a totally *de novo* homotetrameric coiled coil, CC-Tet [208].

CC-Tet was previously designed as part of the coiled coil basis set, a set of *de novo* coiled coils intended for use in synthetic biology applications [208]. The set, which also includes a dimer and trimers, was designed according to established relationships between the identities of the *a* and *d* residues and the oligomeric state of coiled coils [187]. For example, placing Leu at *a* and Ile at *d* leads to tetramers. As well as implementing these core-design rules, CC-Di, -Tri and -Tet were also designed under the assumption that lower-order coiled coils form Type-N interfaces where the residues at the *a* and *d* positions primarily contribute to KIH packing and oligomeric state specification (Figure 3-1a) [204,208]. Furthermore, their heptads follow the classic **HPPHPPP** repeat pattern (**H**, hydrophobic; **P**, polar) where the *a* and *d* residues form the hydrophobic seam through which the amphipathic helices interact. The core-flanking *e* and *g* positions largely contribute to partner selection. Together, these residues form a single *gade* motif. However, as introduced in Chapter 1, higher-order coiled coils can also contain two hydrophobic seams where each helix interacts with its two neighbours *via* separate hydrophobic seams [202]. Depending on the extent of overlap between the seams these coiled coils are designated Type-I, -II or -III [43,203,204]. Furthermore, the inter-seam angle helps to determine oligomeric state. While theoretically optimal for pentamer formation [43], the angle between the seams in Type-II interfaces should also be conducive with tetramer formation: the inter-seam angle in an idealised Type-II interface is 103 ° and the internal angle of a square (representing a tetramer) is 90 °, giving a discrepancy of just 13 ° (Figure 3-1c) [204]. Because both Type-N and Type-II interfaces appear to be able to form tetrameric coiled coils, both approaches to describing these coiled-coil structures must be considered. Furthermore, in Type-N coiled coils, typically only the *e* and *g* residues take part in interhelical interactions, such as ionic interactions, that help to specify the coiled-coil structure (Figure 3-1b). However, in Type-II coiled coils the *e* and *g* residues are increasingly involved in KIH interactions and the other core-flanking residues are positioned closer in space. Therefore, residues at the more-peripheral *b* and *c* positions may also take part in interhelical interactions and, thus, a wider range of interhelical interactions are potentially available to tetrameric coiled coils (Figure 3-1d).

This chapter describes the design and characterisation of an expanded set of *de novo* homotetramers. These coiled coils are intended for use as PIDs *in vivo*, particularly in the design of artificial transcription factors. However, they may also find use in the fields of nanotechnology and materials science [263,321-323].

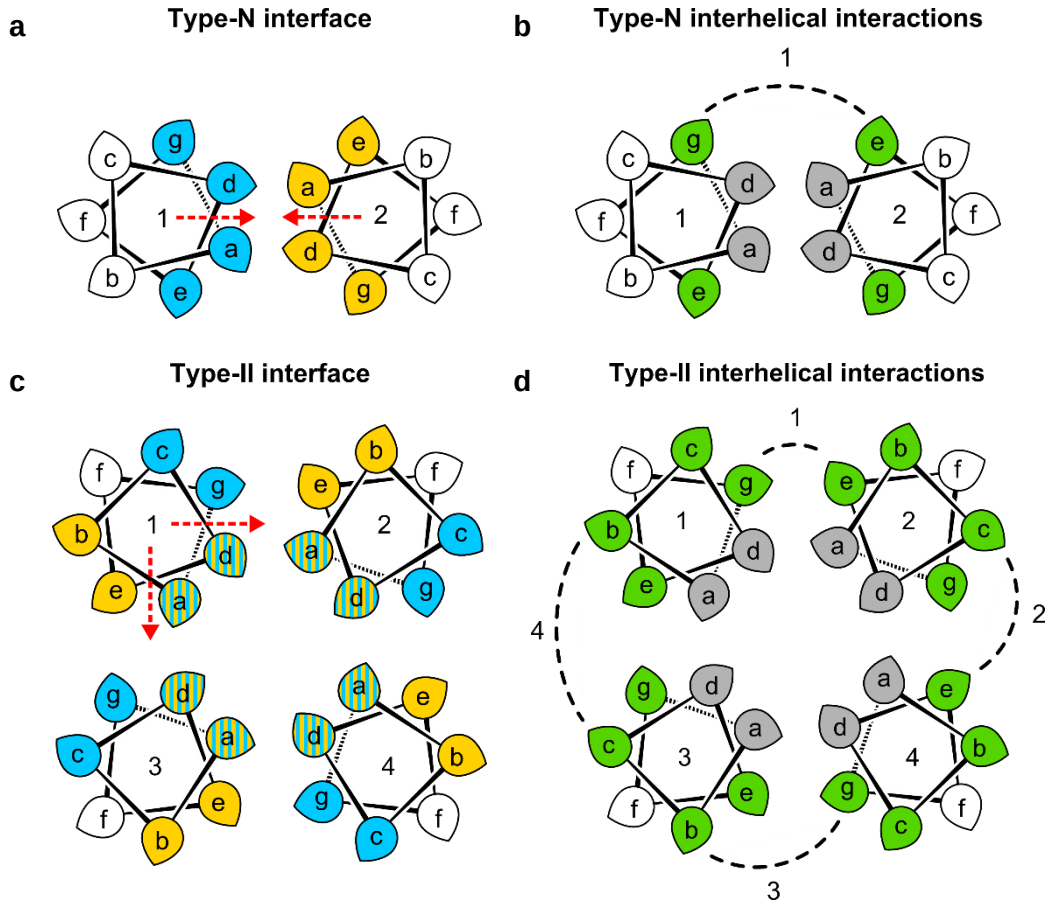


Figure 3-1 Type-N and Type-II coiled-coil interfaces and potential interhelical interactions. Helical wheels for a dimeric coiled coil demonstrating (a) a Type-N interface involving the *gade* (blue/yellow) heptad positions and (b) the potential interhelical interactions that may occur between core-flanking residues. Helical wheels for a tetrameric coiled coil demonstrating (c) a Type-II interface involving *abcdeg* residues as overlapping *deab* (blue) and *cdga* (yellow) motifs and (d) the potential interhelical interactions that may occur between core-flanking residues. Red arrows indicate approximate orientation of the hydrophobic seams. Potential interhelical interactions: 1, *e*-*g*′; 2, *c*-*e*′; 3, *b*-*g*′; 4, *b*-*c*′.

Furthermore, they represent a step towards heterotetrameric coiled coils, which will ultimately be useful for building increasingly complex components for synthetic biology.

3.2 CC-Tet is not robust to mutation

3.2.1 A CC-Tet charge-swapped variant is not tetrameric

The coiled coils in the basis set contain different core residues that specify dimeric, trimeric and tetrameric oligomeric states [208]. Otherwise, their sequences are very similar to each other. For example, all of the designs contain Glu at the *g* positions and Lys at the *e* positions. These oppositely charged residues can form interhelical

```
                  g abcdefg abcdefg abcdefg abcdef
CC-Tet:     Ac – G E LAAIKQE LAAIKKE LAAIKWE LAAIKQ GAG – NH2
CC-Tet-KE:  Ac – G K LAAIEQK LAAIEKK LAAIEWK LAAIEQ GAG – NH2
```

Figure 3-2 Peptide sequences for CC-Tet and CC-Tet-KE. Both peptides were in *g*-register and were N-terminally acetylated and C-terminally amidated. Ac, acetyl.

ionic interactions. Altering the residues at these positions could expand the basis set and increase the number of available components for use in synthetic biology. Furthermore, it would allow for the determination of the contribution of these residues to oligomeric state. Therefore, a charge-swapped version of CC-Tet was made. That is, the residues at the *g* ang *e* positions were swapped such that *g*=Lys and *e*=Glu to generate the peptide CC-Tet-KE (Figure 3-2). This peptide was otherwise identical to CC-Tet including at the core positions, which were *a*=Leu and *d*=Ile; the *c* and *b* positions, which were Ala; and the *f* positions, which were Gln to promote helicity, Lys to promote solubility or Trp to provide a chromophore for detection and concentration determination. This peptide was made by solid-phase peptide synthesis, purified by HPLC and confirmed by mass spectrometry then investigated using various biophysical techniques.

Far UV circular dichroism (CD) spectroscopy was used to probe the folded state of the peptide. This technique is widely used to study protein secondary structures, stabilities and interactions [324-326]. CD is the difference in absorption of right and left circularly polarised light by a chiral molecule and CD spectroscopy involves measuring CD as a function of wavelength. In the far UV region (190-250 nm) the chromophore is the amide bond, which gives signal when located in a regularly folded chiral environment, *i.e.* when the bonds adopt regular φ/ψ angles within secondary structure elements. The CD spectra of α helices have a characteristic shape with two minima at 208 and 222 nm and a maximum at 193 nm [327]. A mean residue ellipticity at 222 nm (MRE_{222}) value of approximately -36000 $deg.cm^2.dmol^{-1}.res^{-1}$ is indicative of a fully α-helical structure.

When measured at 10 μM peptide concentration both CC-Tet and CC-Tet-KE had CD spectra with minima at 208 and 222 nm (Figure 3-3a). This showed that CC-Tet-KE, like CC-Tet, adopted a highly α-helical structure.

The thermal stability of CC-Tet-KE was also investigated by monitoring MRE_{222} while heating the peptide from 5 to 95 °C (Figure 3-3b). As the temperature increased there was an increase in MRE_{222} with both CC-Tet and CC-Tet-KE,

consistent with the loss of α-helical character and partial unfolding. With CC-Tet-KE, part of a sigmoidal unfolding curve could also be seen and the MRE_{222} at 95 °C was much higher than for CC-Tet. Therefore, CC-Tet-KE was less thermally stable than CC-Tet, although the thermal midpoint of unfolding (T_M, the temperature at which the sample is 50 % unfolded) could not be determined for either peptide.

Next, CC-Tet-KE was investigated using analytical ultracentrifugation (AUC) to determine its solution-phase molecular weight. Both sedimentation velocity (SV) and sedimentation equilibrium (SE) experiments were performed [328]. In SV experiments, solutes are subject to high centrifugal force and the resulting sedimentation is monitored over time using absorbance. This provides information about the solute's size and shape and about sample homo-/heterogeneity. In SE experiments, solutes are spun at lower speeds and for longer times than in SV experiments. This allows an equilibrium between sedimentation and back diffusion to establish. At equilibrium, the radial distribution of solute concentration provides information about solute mass and, for associating species, association constants.

In both SV and SE experiments, the predicted molecular weights for CC-Tet-KE corresponded to a trimeric oligomeric state (Figure 3-3c and d, Figure 8-77). Therefore, although CC-Tet-KE formed an α-helical oligomer, it was not a tetramer like the parent, CC-Tet [208].

The sequences differed only at the *e* and *g* positions therefore it appears that the identities of the residues here play some role in oligomeric state specification in these peptides. Although CC-Tet was designed with a traditional Type-N heptad repeat, its crystal structure has features of a Type-II coiled interface: some residues at *g* and *e*, as well as the *a* and *d* positions, act as knobs in KIH interactions, as determined using SOCKET [201,208]. Therefore, while the *g*=Glu/*e*=Lys arrangement may be conducive to KIH formation in CC-Tet, the charge-swapped version may not be. So, without these additional stabilising KIH interactions, the weakened structure "defaults" to a trimer. In CC-Tri and CC-Di, only the *a* and *d* positions act as knobs in KIH interactions so the identities of the residues at the *e* and *g* positions should not affect the structure. Therefore, it is predicted that charge-swapped versions of CC-Tri and CC-Di would retain the oligomeric state of the parent.

Furthermore, Glu may be accommodated at the *g* position in CC-Tet because it is relatively small. Lys, however, is a much larger residue and may be poorly

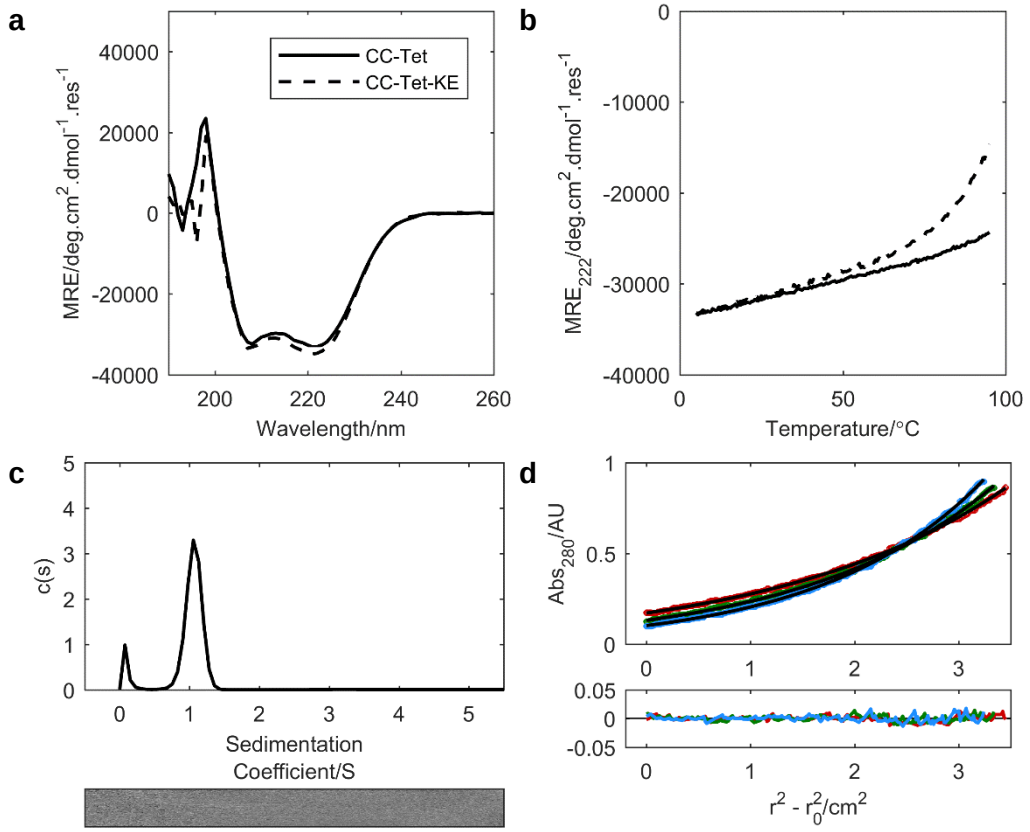


Figure 3-3 Biophysical characterisation of the CC-Tet charge-swapped version, CC-Tet-KE. (a) CD spectra at 5 °C of CC-Tet and CC-Tet-KE. (b) Temperature-dependent CD measurements monitoring MRE_{222} between 5 and 95 °C for CC-Tet and CC-Tet-KE, key as in (a). Peptide concentrations for CD spectroscopy were 10 µM. (c) Sedimentation velocity c(s) distributions (top) and residuals (bottom) for CC-Tet-KE returning a Mw of 10.5 kDa (3.1 x monomer mass). (d) Sedimentation equilibrium data at 30, 33, and 36 krpm (red, green and blue circles) and fits (black lines) (top) and residuals (same colours) (bottom) for CC-Tet-KE returning a Mw of 9.6 kDa (2.8 x monomer mass). All measurements were performed in PBS (pH 7.4). Mw, molecular weight.

accommodated at this position due to the direction of the C_α-C_β bond vector that, in a tetramer, points towards the adjacent helix (conversely, at the *e* position, this bond vector points out into solution). Therefore, due to steric clashing, placing Lys at the *g* position may actively force the coiled coil into a lower oligomeric state where the *g* positions are more solvent-exposed and the core flanking *e* and *g* residues on adjacent helices are further apart in space.

3.2.2 A buried hydrogen bonding network cannot recover tetramerisation

Because CC-Tet-KE was trimeric, further versions of this peptide were made to determine whether the tetrameric state could be restored. Given that the X-ray crystal structure of CC-Tet has some Type-II-character, *i.e.* it contained some non-

core residues that acted as knobs, CC-Tet-KE was redesigned to make the interfaces more Type-II-like. The interfaces between adjacent helices in the tetramer were treated as dimeric interactions, allowing for the implementation of design rules known to promote dimer formation. Promoting this dimer-like interaction in the context of CC-Tet-KE may encourage the peptide to form a Type-II coiled coil and possibly a tetramer (though Type-II interfaces can also form larger oligomers [43,203]).

Placing Asn at *a* sites of coiled-coil sequences has been shown to promote dimer formation [208,229,329-331]. It is believed that this is due to the formation of specific interhelical *a*-to-*a* hydrogen bonds that are only satisfied in the dimer [332-334]. Otherwise, burying this polar residue destabilises the hydrophobic core and disfavours alternate oligomeric states. For example, the basis set homodimer, CC-Di, contains Asn at *a*17 (Figure 3-4b) [208]. When this position is Ile, in keeping with the other *a* sites in CC-Di, this peptide forms a trimer [208].

This dimer-specifying rule was introduced into the dimer-like interfaces of CC-Tet-KE (Figure 3-4c). Asn was introduced at *d*20 and *e*21, replacing Ile and Glu residues, respectively, to give the peptide CC-Tet-KE-N-4 (Figure 3-4a). These positions are equivalent to *a* positions when the *deab* and *cdga* interfaces are considered as *gade*-like motifs, the aim being that the Asn residues might form interhelical *d*-to-*e* hydrogen bonds. An additional version of this peptide was made with a C-terminal extension of an additional half-heptad. This peptide was designated CC-Tet-KE-N-4.5. For this pair of peptides, the “N” suffix denotes that the peptides contained Asn while the “4” or “4.5” suffixes denote the number of heptads in the peptides.

CD spectroscopy showed that CC-Tet-KE-N-4 was predominantly unfolded with a fraction helix value of just 24 % (Figure 3-5a). The longer peptide, CC-Tet-KE-N-4.5, was more folded and had a characteristically α-helical spectrum with a fraction helix value of 83 %. It was also somewhat less folded than CC-Tet-KE. Furthermore, temperature-dependent CD experiments revealed CC-Tet-KE-N-4.5 was much less thermally stable than CC-Tet-KE with a T_M value of 55 °C (Figure 3-5b). However, the thermal denaturation was not reversible and a higher MRE_{222} was observed on returning the sample to 5 °C. This was most likely due to loss of peptide through precipitation. The T_M of CC-Tet-KE-N-4 at this peptide concentration was too low to measure. Thus, the introduction of two polar Asn residues, one at a buried *d* position and one at a more peripheral *e* position, was

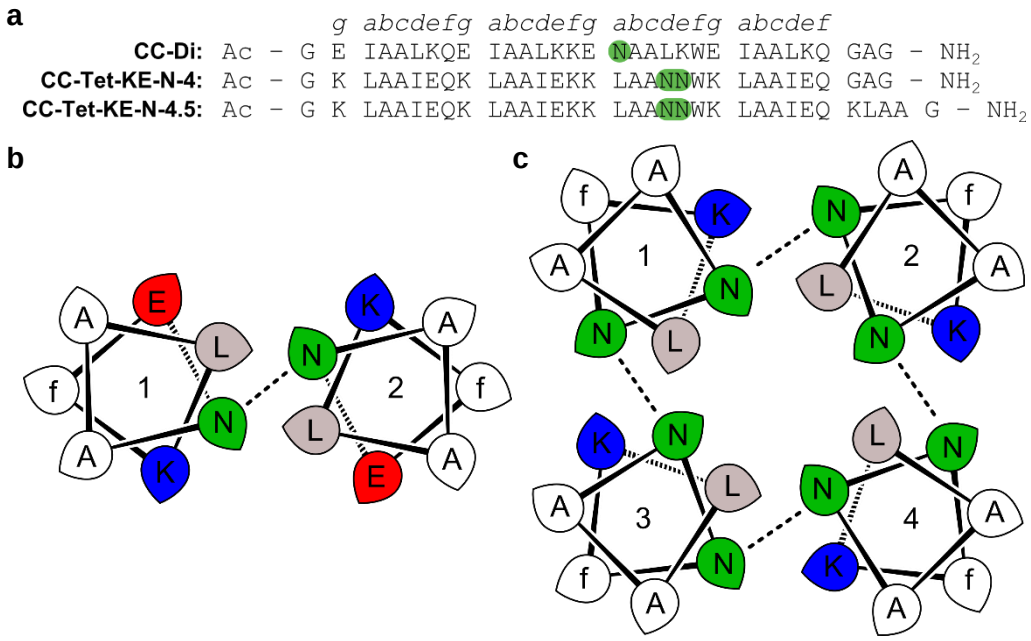


Figure 3-4 Buried Asn residues in dimeric and tetrameric coiled-coil interfaces. (a) Peptide sequences for CC-Di [208], CC-Tet-KE-N-4 and CC-Tet-KE-N-4.5. All peptides were in *g*-register and were N-terminally acetylated and C-terminally amidated. (b) CC-Di helical wheel with helices numbered 1 and 2 showing Asn at *a* positions. (c) CC-Tet-KE-N helical wheel with helices numbered 1-4 showing Asn at *d* and *e* positions (which are equivalent to *a* positions within the *deab* and *cdga gade*-like motifs). Dashed black lines indicate possible Asn-Asn interactions.

clearly very destabilising to CC-Tet-KE. Extending the length of the peptide increased its thermal stability and appeared to be able to overcome some, though not all, of this destabilisation [225].

As CC-Tet-KE-N-4 was unfolded, AUC experiments were performed for CC-Tet-KE-N-4.5 only. In both SV and SE experiments, CC-Tet-KE-N-4.5 returned molecular weights corresponding to a trimeric oligomeric state (Figure 3-5c and d, Figure 8-78).

Therefore, applying the Asn@*a* rule to the dimer-like interfaces of a potential Type-II coiled coil did not encourage the assembly to form a higher oligomeric state. This may be because, although CC-Tet contains KIH interactions involving core flanking *e* and *g* positions (a characteristic of Type-II coiled coils), not all of the *e* and *g* positions form knobs in these interactions: most are hole residues. Therefore, CC-Tet, and perhaps other tetrameric coiled coils, represents an intermediate between the two types of interface.

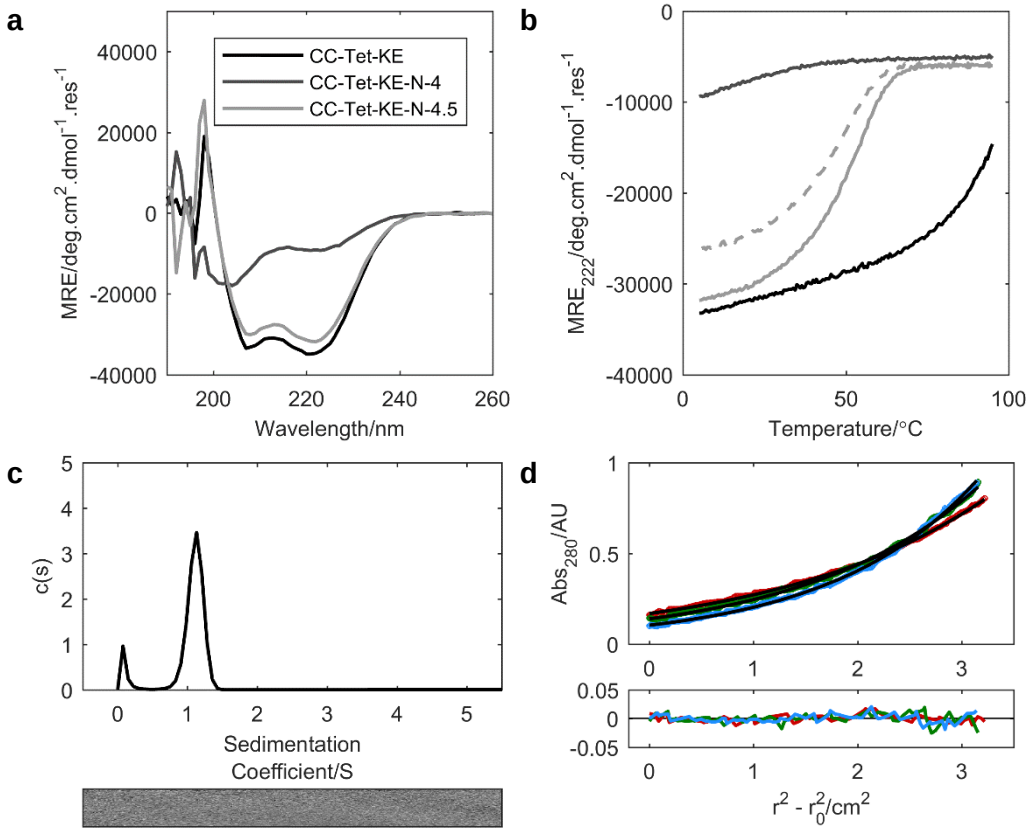


Figure 3-5 Biophysical characterisation of the CC-Tet-KE-N variants. (a) CD spectra at 5 °C of CC-Tet-KE, CC-Tet-KE-N-4 and CC-Tet-KE-N-4.5. (b) Temperature-dependent CD measurements monitoring MRE_{222} between 5 and 95 °C (solid lines) for CC-Tet-KE, CC-Tet-KE-N-4 and CC-Tet-KE-N-4.5 and between 95 and 5 °C (dashed line) for CC-Tet-KE-N-4.5, key as in (a). Peptide concentrations for CD spectroscopy were 10 µM. (c) Sedimentation velocity c(s) distributions (top) and residuals (bottom) for CC-Tet-KE-N-4.5 returning a Mw of 10.3 kDa (2.9 x monomer mass). (d) Sedimentation equilibrium data at 30, 33, and 36 krpm (red, green and blue circles) and fits (black lines) (top) and residuals (same colours) (bottom) for CC-Tet-KE-N-4.5 returning a Mw of 10.7 kDa (3.0 x monomer mass). All measurements were performed in PBS (pH 7.4).

Furthermore, in the absence of crystallographic data for CC-Tet-KE, it is not possible to determine what type of interface this peptide forms. It cannot be assumed that it forms a Type-II coiled coil and, as a trimer, it may in fact form a Type-I interface where the hydrophobic seams overlap by one residue (and therefore the *gade*-like motifs are *gade* and *cdga* and overlap by three residues). If this is the case, it appears that introducing Type-II sequence features to an otherwise Type-I interface is not sufficient to cause the structure to adopt the former. It may be interesting to introduce Asn at *d* and *e* positions in CC-Tet as well to see whether this maintains the tetrameric oligomeric state.

3.3 Design and characterisation of an updated homotetramer set

The tetrameric coiled coil CC-Tet is not robust to an *e*/*g* charged-residue swap. Thus, CC-Tet may not be the optimal starting point to develop an expanded set of tetrameric coiled coils for applications in synthetic biology. Therefore, an updated set of homotetramers was designed to determine whether other heptad arrangements would more robustly specify tetrameric coiled coils.

3.3.1 Leu/Ile core homotetramer design

The design of the updated homotetramer set was based on the observation that, in tetrameric coiled coils, the interhelical distances between the *e*/*g*, *c*/*e, b*/*g* and *b*/*c* positions may all allow salt bridges to form between charged residues at these positions [187,192,224,225]. The average distances between the C_α atoms of these positions in CC-Tet were determined from the crystal structure PDB coordinate file (Figure 3-6). While the *e*/*g* distance was by far the shortest, the *c*/*e*, *b*/*g* and *b*/*c* distances were all similar with values of 14-15 Å.

The average C_α-N_ζ and C_α-C_δ distances were also measured for Lys and Glu residues, respectively, at all *e*/*g* positions in CC-Tet to determine the length of these side chains. The average side chain length was 5.7 ± 0.5 Å for Lys and 3.3 ± 0.3 Å for Glu. While these side chains are a suitable length to form ionic interactions between *e*/*g* positions, they may not be able to bridge the *c*/*e*, *b*/*g* and *b*/*c* positions. However, the Glu and Lys side chains in the CC-Tet crystal structure adopt many different poses and many are not in an extended conformation. The maximal C_α-N_ζ and C_α-C_δ distances for Lys and Glu were 6.3 and 3.9 Å, respectively. Therefore, it is conceivable that fully extended Lys and Glu side chains could bridge at least the *c*-*e* distance to form salt bridges (assuming a maximum distance for salt bridging of 4 Å). With the *b*/*g* and *b*/*c* charge configurations, salt bridges may not form due to the slightly larger inter-residue distances. However, the charged residues may still promote the folding of the coiled coils through electrostatic steering, where charged residues do not necessarily form salt bridges but instead take part in long-range electrostatic interactions that increase the rate of association without affecting the rate of dissociation, leading to stronger interactions [335,336]. This phenomenon is likely to occur in all eight designs, even where salt bridges are unable to form.

a

Interaction	Distance (Å)
e/*g*	8.9 ± 1.6
c/*e*	14.2 ± 0.9
b/*g*	14.6 ± 0.6
b/*c*	14.8 ± 1.2

b

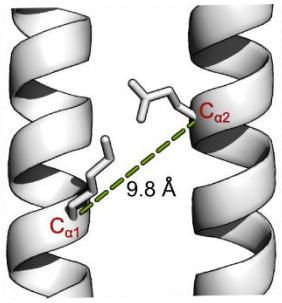
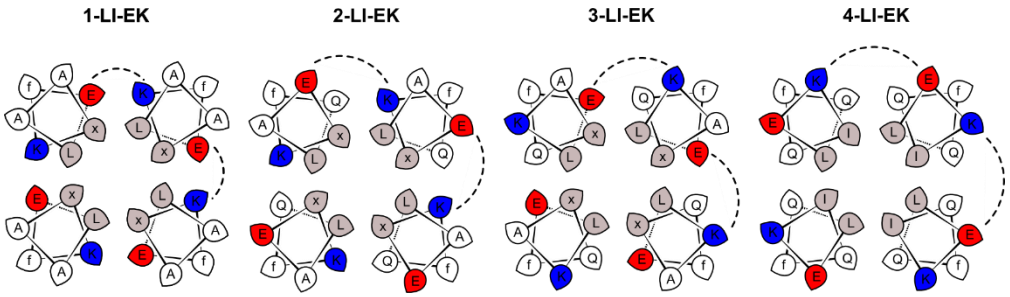


Figure 3-6 Average distances between heptad positions that could take part in interhelical ionic interactions in CC-Tet. (a) Distances measured as the mean distances (Å) between the C_α atoms of the residues of interest calculated from the PDB coordinate file for CC-Tet (PDB ID: 3R4A) [208]. Errors are one s.d. from the mean. *e*/*g* distances were calculated as g_i-e'_{i+5} interactions. *c*/*e* distances were calculated as e_i-c'_{i-2} interactions. *b*/*g* distances were calculated as b_i-g'_{i-2} interactions. *b*/*c* distances were calculated as b_i-c'_{i+1} interactions. Residues in the adjacent helix are indicated with *x*′ where *x* is any heptad position. (b) Example interaction between Lys and Glu residues in CC-Tet with C_α atoms labelled ($C_{\alpha 1}$ and $C_{\alpha 2}$, respectively) and inter-C_α distance show (green dashed line).

Therefore, the set of updated homotetramers was intended to test whether changing the heptad positions of the ionic residues would (1) still lead to the formation of parallel homotetrameric coiled coils; (2) allow for the formation of interhelical ionic interactions; and (3) lead to coiled coils with different thermal stabilities. For point 3, the strength of the ionic interaction should get weaker as the charges are moved further apart. If such ionic interactions contribute to coiled coil stability, they would be expected to become less significant as the charges are moved progressively to more-peripheral heptad positions. Therefore, coiled coils with charged residues at *b*/*c* positions would be predicted to be less stable than those with charged residues at *e*/*g* positions and those with charged residues at *c*/*e* or *b*/*g* would be expected to have intermediate thermal stabilities.

Eight new homotetramer peptide sequences were designed (Figure 3-7). The peptides had four heptad arrangements: xLAAIx*f*, QLAxIx*f*, xLxAIQ*f* or QLxxIQ*f*, written beginning at a *g* position, where x is Glu or Lys and *f* is the heptad *f* position. The peptides were designed with oppositely charged Glu and Lys residues at all of the *e*/*g*, *c*/*e* *b*/*g* or *b*/*c* positions. Furthermore, all were designed with an LI-core, *i.e.* Leu at *a* and Ile at *d*. Where *e* or *g* positions were not occupied by a charged residue, they were made Gln. This polar residue was included at these sites to ensure that the hydrophobic seam was limited to the *a* and *d* positions. Placing additional hydrophobic residues at these positions, such as Ala or Val, can lead to the formation of higher-order coiled coils such as hexamers [43,213] or to the formation

a



b

				cdefg		*abcdefg*		*abcdefg*		*abcdefg*		*ab*		
1-Lx-EK:	Ac	–	G	AxKfE		LAAxKfE		LAAxKWE		LAAxKfE		LA	G	– NH_2
1-Lx-KE:	Ac	–	G	AxEfK		LAAxEfK		LAAxEWK		LAAxEfK		LA	G	– NH_2
2-Lx-EK:	Ac	–	G	ExKfQ		LAExKfQ		LAExKWQ		LAExKfQ		LA	G	– NH_2
2-Lx-KE:	Ac	–	G	KxEfQ		LAKxEfQ		LAKxEWQ		LAKxEfQ		LA	G	– NH_2
3-Lx-EK:	Ac	–	G	AxQfE		LKAxQfE		LKAxQWE		LKAxQfE		LK	G	– NH_2
3-Lx-KE:	Ac	–	G	AxQfK		LEAxQfK		LEAxQWK		LEAxQfK		LE	G	– NH_2
4-LI-EK:	Ac	–	G	KIQfQ		LEKIQfQ		LEKIQfQ		LEKIQfQ		LE	G	– NH_2
4-LI-KE:	Ac	–	G	EIQfQ		LKEIQfQ		LKEIQfQ		LKEIQfQ		LK	G	– NH_2

x = I/V

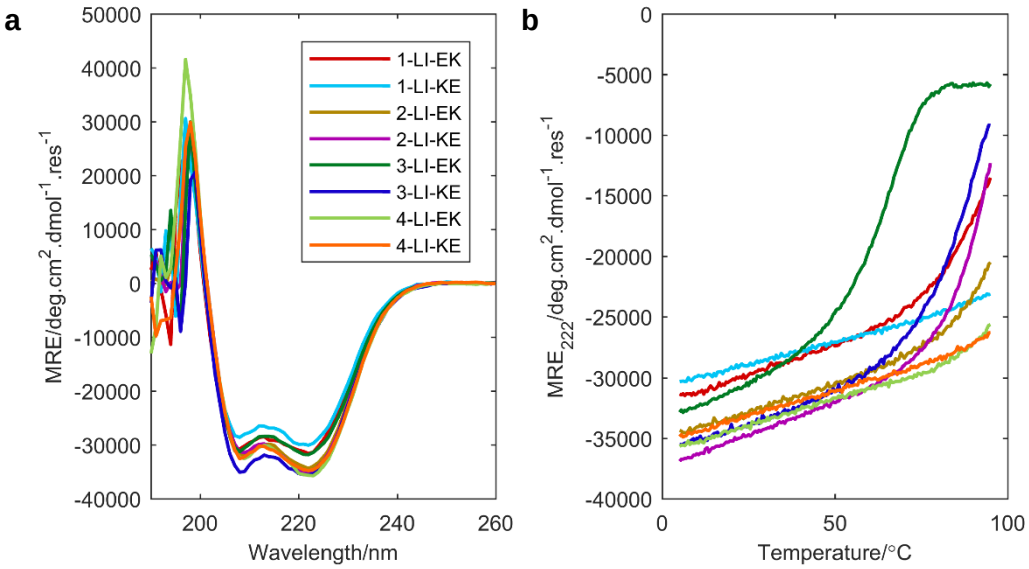
Figure 3-7 Design approach for homotetrameric coiled coils. (a) Tetrameric helical wheel representations for peptides 1-LI-EK (far left), 2-LI-EK (centre left) and 3-LI-EK (centre right) and 4-LI-EK (far right) demonstrating the different arrangements of charged residues at *e*/*g*, *c*/*e*, *b*/*g* or *b*/*c* positions, respectively. Dashed lines indicate potential Glu-Lys interactions. (b) Sequences of peptides. All peptides were in *c*-register and were N-terminally acetylated and C-terminally amidated. Core *d* positions were either all Ile or all Val (denoted as x) except in 4-LI-EK and 4-LI-KE where *d* positions were all Ile.

of antiparallel structures [205,222,337]. Where *b* or *c* positions were not occupied by charged residues, these positions were made Ala. Finally, *f* positions were occupied by Gln, Lys or Trp and helix-capping Gly residues were added to the N and C termini. All peptides were designed in *c*-register *i.e.* excluding the N-terminal Gly, the peptides began at a *c* position. The peptides were named for the location of the charged residues (1, *e*/*g*; 2, *c*/*e*; 3, *b*/*g*; 4, *b*/*c*), the identity of the core residues (LI, *a*=Leu/*d*=Ile) and the orientation of the charged Glu and Lys residues (EK, Glu appears first in the linear heptad sequence; KE, Lys appears first in the linear heptad sequence).

3.3.2 Leu/Ile core designs form homotetrameric coiled coils in solution

3.3.2.1 CD spectroscopy shows the updated designs form α-helical homomers

All of the peptides had characteristically α-helical CD spectra with high fraction helix values (Figure 3-8a and c). Furthermore, these α-helical species were all



c

Peptide	MRE_{222} [a]	Fraction helix (%)	T_M (°C)
1-LI-EK	-31603	82	> 95.0
1-LI-KE	-30056	78	> 95.0
2-LI-EK	-34375 ± 266	89	> 95.0
2-LI-KE	-35277 ± 1472	92	> 95.0
3-LI-EK	-31815 ± 817	83	67.5 ± 0.5
3-LI-KE	-35550 ± 520	92	89.8 ± 1.0
4-LI-EK	-35669	93	> 95.0
4-LI-KE	-34766	90	> 95.0

Figure 3-8 CD spectroscopy data for updated homotetramer set. (a) CD spectra at 5 °C of 1-LI-EK, 1-LI-KE, 2-LI-EK, 2-LI-KE, 3-LI-EK, 3-LI-KE, 4-LI-EK and 4-LI-KE. (b) Temperature-dependent CD measurements monitoring MRE_{222} between 5 and 95 °C for the above peptides, key as in (a). (c) MRE_{222}, fraction helix values and T_M values for the above peptides. [a] Units, $deg.cm^2.dmol^{-1}.res^{-1}$. Peptides were at 10 µM. All measurements were performed in PBS.

highly thermally stable and most had T_M values above the accessible range (Figure 3-8b and c). The least stable peptides were 3-LI-EK and 3-LI-KE, where the charged residues were at the *b* and *g* positions. These were expected to be amongst the least stable because the *b*/*g* residues were positioned relatively far apart in space. Furthermore, the C_α-C_β bond vector of the *b* site is oriented out into solution and was therefore not expected to be optimally positioned for the formation of ionic interactions (Figure 3-7a, centre right). For similar reasons, the peptides 4-LI-EK and 4-LI-KE were also expected to be relatively unstable. However, they were amongst the most stable.

3.3.2.2 AUC shows the updated designs form homotetramers in solution

In SV and SE experiments the majority of the peptides gave molecular weights corresponding to tetramers (Figure 3-9, Figure 8-80–Figure 8-87). The peptides also all displayed a single peak in their c(s) distributions indicating that they each formed a single species in solution.

The exception was 1-LI-KE, which in both SE and SV experiments returned molecular weights of 3.5 x monomer mass. It was therefore not clear whether this peptide formed a trimer, tetramer or an interconverting mixture of the two. This is in contrast to CC-Tet-KE, which differs only in register but unambiguously formed a trimer.

Interestingly, when the charged residues were placed at the *c*/*e*, *b*/*g* and *c*/*b* positions, all the designs formed tetramers in solution, regardless of the order of the Glu/Lys residues: charge-swapped variants of these peptides could be made without affecting the oligomeric state of the resulting assembly. Conversely, while 1-LI-EK and CC-Tet did form tetramers, neither 1-LI-KE nor CC-Tet-KE could reliably form tetramers in solution. Therefore, the *c*/*e*, *b*/*g* and *c*/*b* heptad arrangements offer an improvement over the original xLAAIx*f* heptad for the design of tetramers.

3.3.3 Crystallographic characterisation of Leu/Ile core homotetramers

Crystal structures were solved for 2-LI-EK and 3-LI-EK at 1.7 and 1.1 Å resolution, respectively (Figure 3-10a). Both peptides crystallised as parallel homotetrameric coiled coils (Figure 3-10b). Furthermore, Glu/Lys side chains in close proximity could be observed in both structures, as well as in the crystal structure for CC-Tet, implying the formation of ionic interactions. The inter-side chain distances ($N_{\zeta(Lys)}$-$C_{\delta(Glu)}$) were measured for all Glu/Lys pairs. Surprisingly, the shortest average distances were in the 2-LI-EK structure (Figure 3-10a). This was probably due to the location of some of the charged residues within the unstructured termini of CC-Tet. Despite the larger Glu-Lys distances in CC-Tet, a higher proportion of these pairs appeared to form salt bridges, *i.e.* they had an $N_{\zeta(Lys)}$-$C_{\delta(Glu)}$ distance ≤ 4.0 Å (Figure 3-10a). The fewest salt bridges were observed in 3-LI-EK where just 6.25 % of Glu/Lys pairs formed a salt bridge. This corresponds to just one salt bridge in the whole tetramer. 3-LI-EK also had the largest average $N_{\zeta(Lys)}$-$C_{\delta(Glu)}$

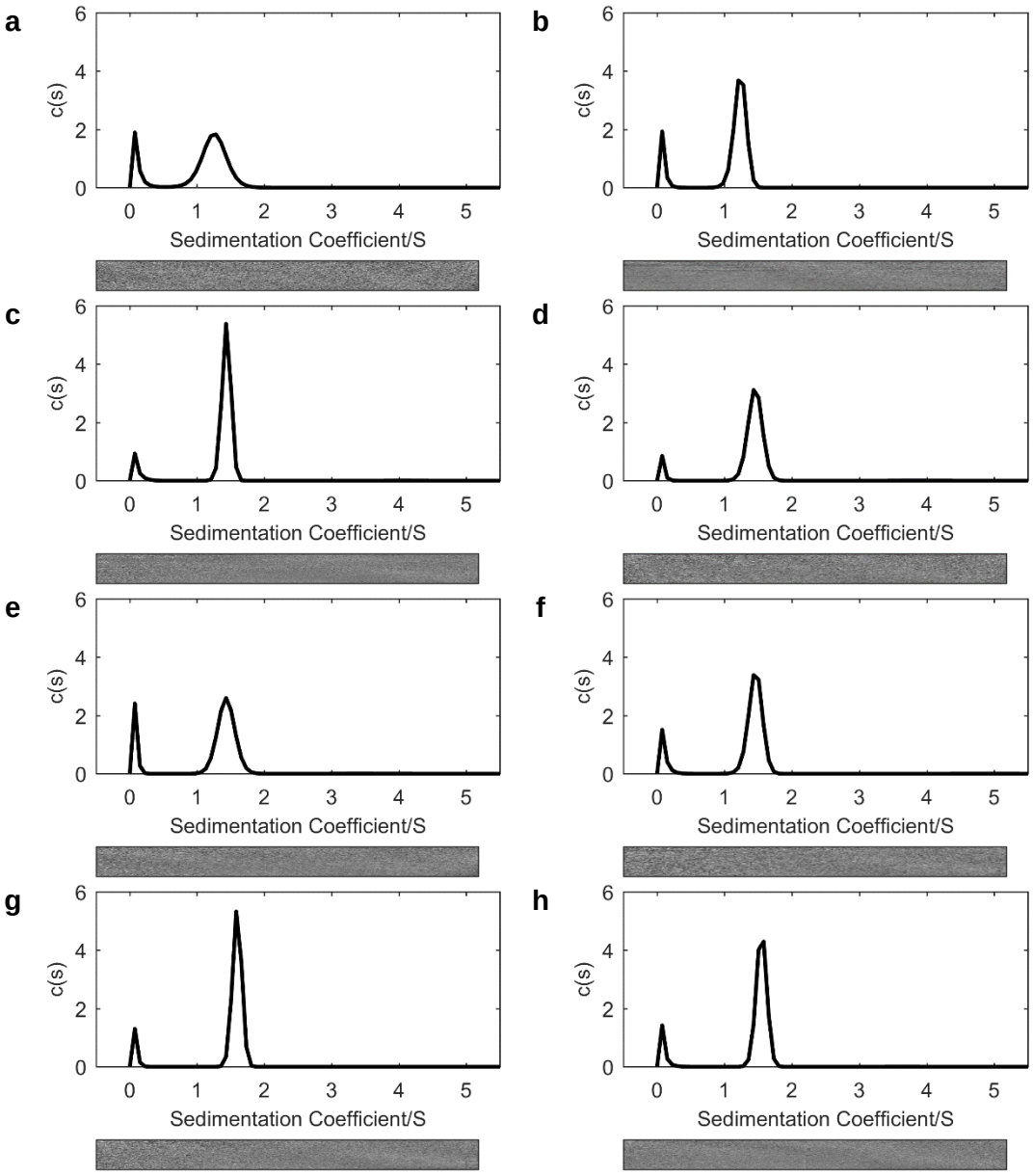


Figure 3-9 Sedimentation velocity data for updated homotetramer designs. c(s) distributions (top) and residuals for: (a) 1-LI-EK returning a Mw of 12.9 kDa (4.0 x monomer mass); (b) 1-LI-KE returning a Mw of 11.3 kDa (3.5 x monomer mass); (c) 2-LI-EK returning a Mw of 13.4 kDa (3.9 x monomer mass); (d) 2-LI-KE returning a Mw of 14.4 kDa (4.2 x monomer mass); (e) 3-LI-EK returning a Mw of 14.5 kDa (4.2 x monomer mass); (f) 3-LI-KE returning a Mw of 13.1 kDa (3.8 x monomer mass); (g) 4-LI-EK returning a Mw of 15.3 kDa (4.1 x monomer mass); (h) 4-LI-KE returning a Mw of 15.4 kDa (4.2 x monomer mass). All measurements were performed in PBS (pH 7.4).

distance. This may partially explain why 3-LI-EK and 3-LI-KE were less stable than the other designs.

Therefore, only the *e*/*g* and *c*/*e* charge arrangements appear to be conducive with the formation of interhelical salt bridges in tetramers, as predicted from the inter-

a

Peptide	Space group	Resolution (Å)	R-work/ R-free	Average $N_{\zeta(Lys)}$-$C_{\delta(Glu)}$ distance (Å)	Salt bridges (%) [a]
CC-Tet	I222	2.07	0.209/0.270	5.6 ± 2.7	50.0
2-LI-EK	C222	1.70	0.189/0.195	4.5 ± 1.0	37.5
3-LI-EK	$P12_11$	1.11	0.187/0.208	7.0 ± 2.8	6.25

b

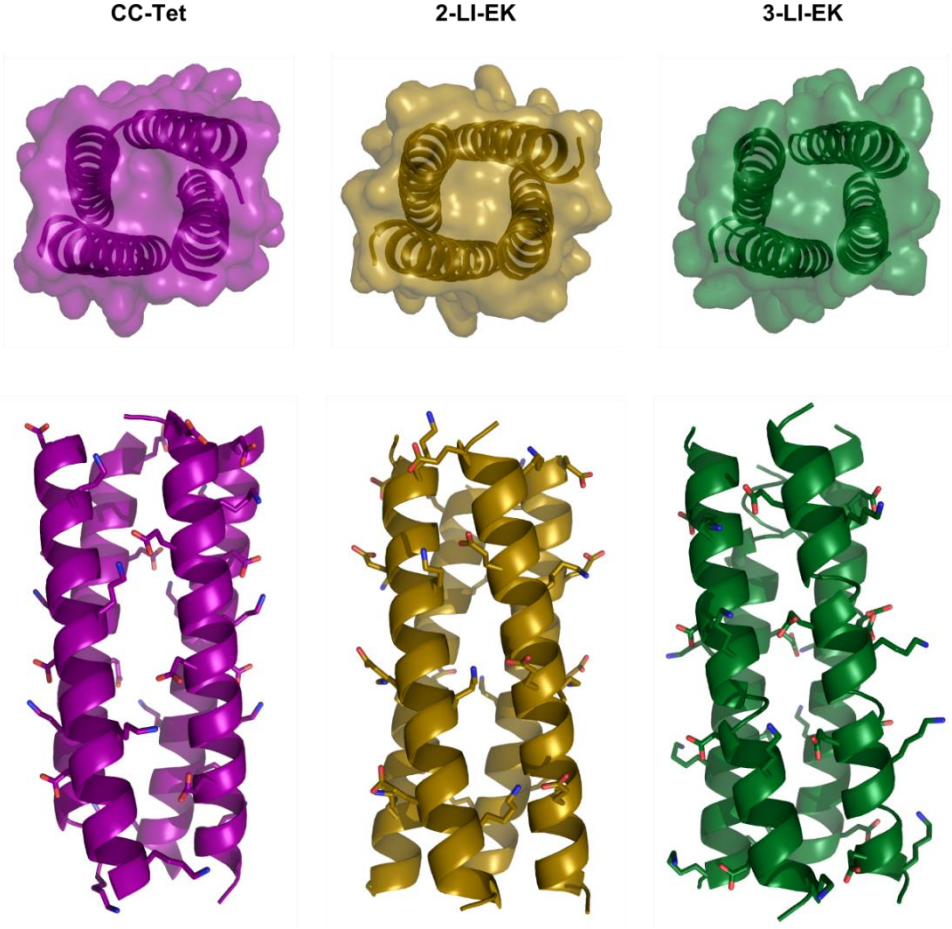


Figure 3-10 Crystallographic characterisation of updated homotetramer designs. (a) Collection and refinement statistics and Glu-Lys interaction analysis for CC-Tet (structure reproduced from reference [208]), 2-LI-EK and 3-LI-EK. See Table 8-3 and Table 8-4 for full statistics. [a] Percentage of *e*/*g*, *e*/*c* or *b*/*g* Glu/Lys pairs that engage in interhelical salt bridges (*i.e*. $N_{\zeta(Lys)}$-$C_{\delta(Glu)}$ distance ≤ 4.0 Å). (b) Crystal structures for: Left, CC-Tet (PBD ID: 3R4A [208,213]); centre, 2-LI-EK; right, 3-LI-EK. Structures shown from N termini as cartoons/surfaces (top) and from sides as cartoons with Glu/Lys shown as sticks (bottom).

residue distances measured in CC-Tet (Figure 3-6a). However, in both tetramers, not all of the Glu/Lys pairs took part in salt bridges. Furthermore, the temperature factors for these side chains were relatively similar to those of the other solvent-exposed residues that were not expected to take part in specific interactions, especially when Ala residues – which have short, inflexible side chains – were excluded (Figure 3-11b).

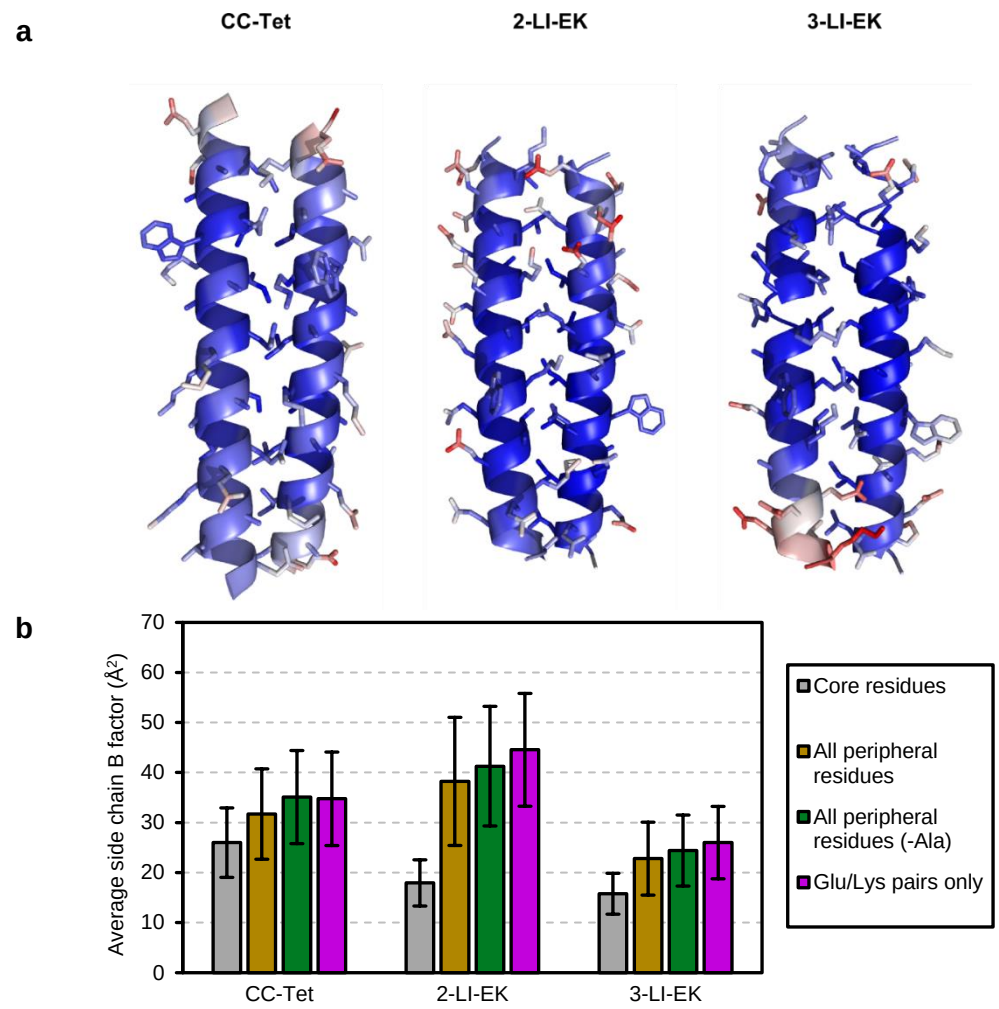


Figure 3-11 B factor analysis of homotetramers. (a) Crystal structures of homotetramers coloured by temperature (B) factors. Left: CC-Tet (PDB ID: 3R4A), min. B factor (blue) = 16.7 Å^2, max. B factor (red) = 70.0 Å^2. Centre: 2-LI-EK, min. B factor (blue) = 10.8 Å^2, max. B factor (red) = 84.6 Å^2. Right: 3-LI-EK, min. B factor (blue) = 10.0 Å^2, max. B factor (red) = 49.1 Å^2. (b) Average side chain B factors for CC-Tet, 2-LI-K and 3-LI-EK for core residues (*a*, *d*), all peripheral residues (*b*, *c*, *e*, *f*, *g*), for all peripheral residues except alanines (-Ala) and for Glu/Lys pairs only (*e*/*g*, *c*/*e* or *b*/*g*, respectively).

In crystallography, temperature factors (or B factors) represent uncertainty in the location of an atom in the crystallographic model. This increases with disorder or motion of that atom. The locations of more-mobile atoms are less certain, and they therefore have higher B factors. Generally, atoms in the main chain, in buried side chains and side chains involved in non-covalent interactions will have low B factors. The B factors of the Glu/Lys side chains in CC-Tet, 2-LI-EK and 3-LI-EK implied that these side chains were relatively mobile and therefore that any interactions between them were dynamic.

a **CC-Tet**

Sequence	ELAAIKQELAAIKKELAAIKWELAAIKQ
Register	gabcdefgabcdefgabcdefgabcdef
Helix 1 (W)	--------Z--XZ--Z---Z--Z-----
Helix 2 (X)	----Y--YW--Y--YW--Y--Y------
Helix 3 (Y)	-----X--X--ZX--X---X--X-----
Helix 4 (Z)	----W--WY--W--WY--W---------

b **2-LI-EK**

Sequence	EIKQQLAEIKQQLAEIKWQLAEIKQQLA
Register	cdefgabcdefgabcdefgabcdefgab
Helix 1 (W)	----XZ--X--XZ--X--XZ--XZ----
Helix 2 (X)	-----W--YW--W--YW--W--Y-----
Helix 3 (Y)	----ZX--Z--ZX--Z--ZX--ZX----
Helix 4 (Z)	-----Y--WY--Y--WY--Y--W-----

c **3-LI-EK**

Sequence	AIQQELKAIQQELKAIQWELKAIQQELK
Register	cdefgabcdefgabcdefgabcdefgab
Helix 1 (W)	----YZ--Y--YZ--Y--Y---Y-----
Helix 2 (X)	----ZY--ZY-ZY--Z--ZY--Z-----
Helix 3 (Y)	-----W--XW-XW--XW--W--XW----
Helix 4 (Z)	-----X--WX--X--WX--X---X----

Figure 3-12 SOCKET output for updated homotetramers. (a) CC-Tet. (b) 2-LI-EK. (c) 3-LI-EK. Positions of knob residues in one helix (W, X, Y or Z) are indicated by a letter. The identity of the letter indicates which helix that knob residue interacts with. Knob cut-off = 7.0 Å.

The structures for 2-LI-EK and 3-LI-EK were analysed using SOCKET to determine which residues take part in KIH interactions [201]. CC-Tet was also analysed for comparison (Figure 3-12) [208]. There were more KIH interactions in 2-LI-EK and 3-LI-EK than in CC-Tet. Furthermore, more of the *e* and *g* positions were knobs in 2-LI-EK and 3-LI-KE than in CC-Tet. These additional KIH interactions might have been expected to increase the stability of the coiled coils. However, as shown by temperature-dependent CD measurements, both assemblies were less stable than CC-Tet (Figure 3-3a and Figure 3-8a). It is interesting to note that in all three structures *a*, *d*, *e*, and *g* residues act as knobs in KIH interactions, which is characteristic of Type-II coiled coils. However, in all of the structures only some of these residues are knobs. Therefore, while the tetramers have Type-II character, they do not have completely Type-II interfaces and may therefore represent an intermediate between Type-N and Type-II coiled coils.

It remains to be seen whether additional KIH interactions and interhelical ionic interactions also formed in the 4-LI-EK or 4-LI-KE tetramers, where the charged residues were placed at the *b*/*c* positions. Additionally, further investigation is

required to determine whether the salt bridges observed in the CC-Tet and 2-LI-EK structures also form in solution and whether they contribute to coiled-coil stability.

3.4 Solution-phase characterisation of Leu/Val core variants

The design of the expanded set of homotetrameric coiled coils described above means that synthetic biologists wishing to use *de novo* designed homotetramers as PIDs *in vivo* or *in vitro* now have a larger number of designs to choose from. Furthermore, the presented designs have a range of thermal stabilities allowing synthetic biologists to choose the coiled coil that is most suitable for their application. However, relative to most natural proteins from mesophilic organisms, all of the designs are still hyperthermally stable. In order to be useful in applications where less stable coiled coils are required, the homotetramer set must be expanded to include coiled coils with lower thermal stabilities. This could be achieved in many ways such as by decreasing the length of the peptides [191,338]. Here, however, the identities of the core residues were changed. Specifically, Ile at *d* was replaced with Val.

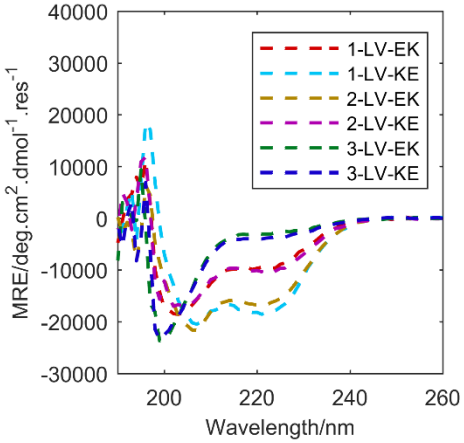
Val, like Ile, is a β-branched aliphatic amino acid that should have similar core packing properties to Ile. The side chains of the two amino acids are very similar, differing only in a methylene group ($-CH_2$). The absence of this methylene in Val makes the residue overall less hydrophobic than Ile so it should form less stable hydrophobic cores. Indeed, replacing Ile with Val at the *a* positions of homo- and heterodimeric coiled coils results in less stable assemblies [339,340].

6 peptides based on 1-LI-EK, 1-LI-KE, 2-LI-EK, 2-LI-KE, 3-LI-EK and 3-LI-KE were designed that contained Val at *d* instead of Ile. The sequences were otherwise identical to the LI-core parents. This gave the LV-core (LV, *a*=Leu/*d*=Val) peptides 1-LV-EK, 1-LV-KE, 2-LV-EK, 2-LV-KE, 3-LV-EK and 3-LV-KE, respectively.

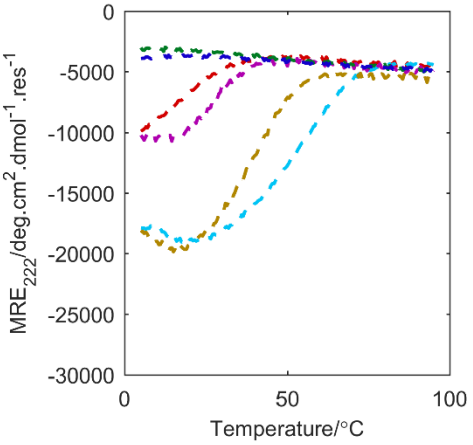
3.4.1 Leu/Val core variants are much less stable than Leu/Ile core homotetramers

By CD spectroscopy, all six LV-core peptides were much less folded than the corresponding LI-core parents (Figure 3-13a and c). 3-LV-EK and 3-LV-KE were almost entirely unfolded. 1-LV-KE and 2-LV-EK were the most folded and α helical. However, the minimum at approximately 208 nm had a more negative MRE indicating that as well as folded material, there was also unfolded peptide present.

a



b



c

Peptide	MRE_{222} [a]	Fraction helix (%)	T_M (°C)
1-LV-EK	-9605	26	< 20.0
1-LV-KE	-18272	48	55.5
2-LV-EK	-16737 ± 1219	44	41.0 ± 0.9
2-LV-KE	-10497 ± 263	28	26.8 ± 2.1
3-LV-EK	-3006	9	< 20.0
3-LV-KE	-3832	11	< 20.0

Figure 3-13 CD spectroscopy data for LV-core versions of homotetramers. (a) CD spectra at 5 °C of 1-LV-EK, 1-LV-KE, 2-LV-EK, 2-LV-KE, 3-LV-EK and 3-LV-KE. (b) Temperature-dependent CD measurements monitoring MRE_{222} between 5 and 95 °C for the above peptides, key as in (a). (c) MRE_{222}, fraction helix values and T_M values for the above peptides. [a] Units, $deg.cm^2.dmol^{-1}.res^{-1}$. Peptides were at 10 µM. All measurements were performed in PBS (pH 7.4).

Furthermore, the LV-core peptides were much less thermally stable than the LI-core homotetramers (Figure 3-13b and c). Indeed, where the T_M values of the LI-core homotetramers were often too high to measure, some of the LV-core peptides had T_Ms that were too low to measure. 3-LV-EK and 3-LV-KE were particularly unstable and did not show any change in MRE_{222} with increasing temperature. The most stable peptides were 1-LV-KE and 2-LV-EK. 1-LV-KE had a T_M value of 55.5 °C, which was only 12 °C lower than the T_M of the least stable LI-core homomer, 3-LI-EK (Figure 3-8a and c). However, while 3-LI-EK had a typical sigmoidal unfolding transition, 1-LI-KE displayed cold denaturation where the peptide initially became more folded when heated to around 20 °C then unfolded as the temperature was increased further. This phenomenon was also observed with 2-LV-EK and 2-LV-KE. Cold denaturation can be an indication of exposed hydrophobic residues, which further shows that the peptides were poorly folded. This is because, at low temperatures, ordered cages of water molecules surround exposed hydrophobic residues and form hydrogen bonds to the peptide,

preventing it from forming the hydrogen bonds required to stabilise its secondary structure [341,342]. As temperature is increased and the cages are disrupted, helices can form. These helices then unfold again as temperature is increased further and the intramolecular hydrogen bonds are broken.

3.4.2 Leu/Val core variants do not form tetramers in solution

By AUC, the LV-core homomers had a range of oligomeric states from monomers potentially up to tetramers (Figure 3-14, Figure 8-88–Figure 8-91). However, for most of the peptides, the oligomeric state assignment was ambiguous. For example, in both SE and SV experiments, 1-LV-KE fitted to molecular weights corresponding to oligomeric states between dimer and trimer. This may reflect the fact that the peptides were quite unfolded: there was likely to be a mixture of unfolded peptide monomers and folded oligomers present. The experiments could also be performed at higher peptide concentrations, which should increase the amount of oligomer present. Furthermore, in the case of 2-LV-KE, the oligomer peak in the SV c(s) distribution has a broad tail (Figure 3-14c). This may be an indication that this peptide formed a number of different oligomeric states.

Finally, using SV, 3-LV-KE fitted to a molecular weight corresponding to a monomer (Figure 3-14d). Given that the SV experiments were performed by centrifuging at 50 krpm and that a solute with a mass this small was not expected to sediment at this speed, this result needs to be verified, for example by performing SE experiments at increased speed.

Placing Val residues at *d* does not appear to be a viable approach to expanding the set of homotetramers because the new peptide designs were generally poorly folded and did not form tetramers. Rather than producing homotetramers that were marginally less stable than the LI-core homotetramers, it produced peptides that were generally impractically unstable and unable to adopt a unique structure. They are therefore unlikely to be useful in synthetic biology applications. Therefore, if a further expanded homotetramer set is indeed required, other designs should be considered such as those that contain a mixture of Ile and Val residues at the *d* positions. In this way, the stability of the coiled coil might be tuned by increasing the Val/Ile ratio in the peptide core without complete loss of the tetramer promoting Ile residues. Alternatively, increasing the lengths of some of the LV-core peptides might make them more folded and therefore more useful. However, it is also possible that the *a*=Leu/*d*=Val core arrangement simply is not optimal for tetramer

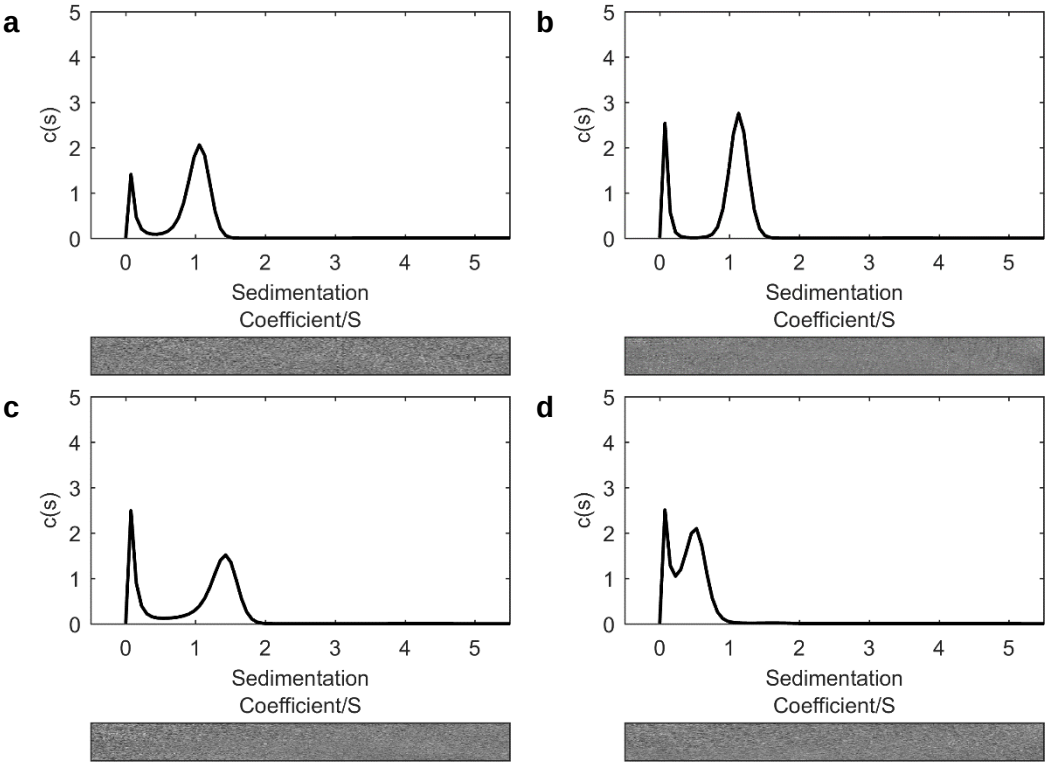


Figure 3-14 Sedimentation velocity data for LV-core homomer designs. c(s) distributions (top) and residuals for: (a) 1-LV-KE returning a Mw of 8.4 kDa (2.6 x monomer mass); (b) 2-LV-EK returning a Mw of 9.7 kDa (2.8 x monomer mass); (c) 2-LV-KE returning a Mw of 11.9 kDa (3.5 x monomer mass); (d) 3-LV-KE returning a Mw of 3.9 kDa (1.1 x monomer mass). All measurements were performed in PBS (pH 7.4).

formation, despite the β-branched side chain of Val. This may be why the LV-core peptides failed to reliably form tetramers. Indeed, a version of GCN4-p1 that contains *a*=Leu and *d*=Val also fails to adopt a unique oligomeric state which further implies it was not a suitable choice for the *d* positions [187].

The frequency at which Ile, Leu and Val occur at the core *a*/*d* positions of parallel tetrameric coiled coils relative to the frequency that they occur in all proteins was analysed using the CC+ coiled coil database [343]. This revealed that Val actually occurs more frequently at the *a* positions than at the *d* positions (Figure 3-15). Therefore, peptides with *a*=Val/*d*=Ile cores might be better at specifying tetramers.

3.5 The effect of non-polar residues at *g* in homotetramers

The inclusion of Leu at *a* positions and Ile at *d* positions of coiled coils was a fairly reliable specifier of parallel tetrameric structures in the homotetramers described in section 3.3. As introduced in Chapter 1, the inclusion of the β-branched Ile at the

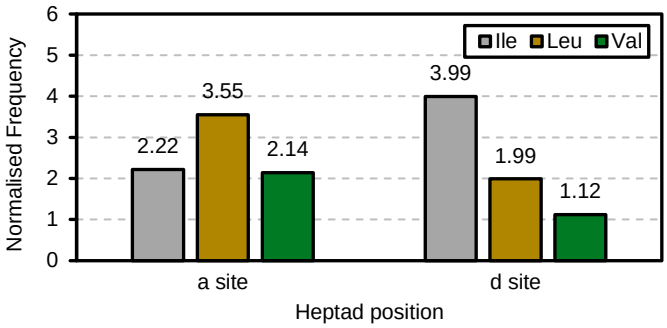


Figure 3-15 Normalised frequencies of Ile, Leu and Val residues at *a* and *d* positions in tetrameric coiled coils. Data generated from a set of 46 parallel homo- and heterotetrameric canonical coiled coil sequences of length ≥14 residues that were extracted from the CC+ coiled coil database [343]. Amino acid frequencies were normalised to the frequency of that amino acid in the SWISS-PROT database [347].

d positions precludes the formation of dimeric structures due to steric clashes [187]. However, while the Leu/Ile core arrangement may not be compatible with lower oligomeric states, it may be tolerated in higher oligomeric states such as pentamers and above – the so-called α-barrels [202,344]. For example, expansion of the hydrophobic seam of CC-Tet through the inclusion of hydrophobic residues at the core flanking *e* positions yielded a hexameric coiled coil, CC-Hex, with a Leu/Ile/Ala core (heptad ELKAIA*f*) [213]. Furthermore, placing Ala at all *e* and *g* positions of the GCN4-p1 leucine zipper dimer led to the formation of an unusual staggered heptameric coiled coil [345]. Sequences from a library of GCN4 *e*/*g* position mutants that were found to form higher order oligomers also contained Ala at some *e*/*g* positions [346]. Additionally, a blunt-ended *de novo* heptameric coiled coil that was designed computationally contained Ala at all *e* and *g* positions [43].

However, there are also cases where placing Ala at core flanking positions does not lead to α-barrel formation, but instead leads to the formation of antiparallel tetramer structures. For example, a charge-swapped version of CC-Hex (heptad KLEAIA*f*) formed an antiparallel tetramer (C. Wood, personal communication). Additionally, placing Ala at just the *g* positions of GCN4 also led to the formation of an antiparallel tetramer similar to rop-like Ala-coils [205,348]. The Ala-coil is a type of helical bundle that contains Ala (or other small, hydrophobic residues) at every 7th residue in addition to an *a*/*d* hydrophobic core. When these Ala residues coincide with *e* positions they are called ferritin-like Ala-coils. This group includes the Lac repressor tetramerisation domain [190]. When the Ala residues fall at the *g* positions they are called rop-like Ala-coils [349]. Both types of Ala-coil are characterised by antiparallel orientations that bring together the Ala residues at two of the four

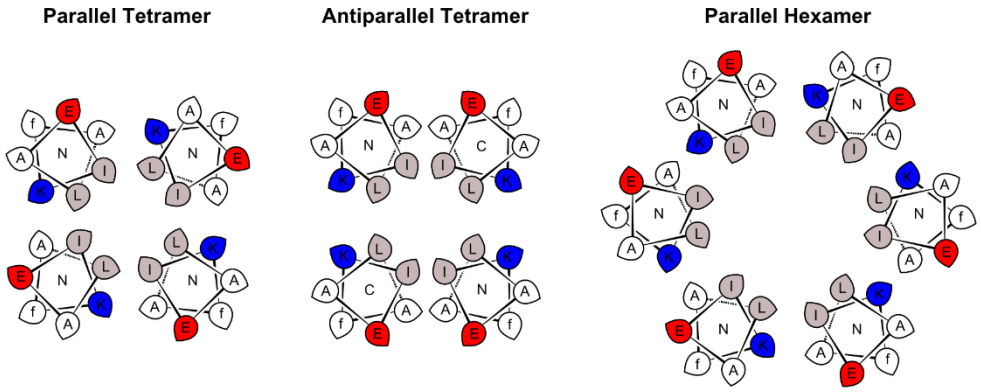
interhelical interfaces. This results in close interhelical contacts at these Ala-containing interfaces such that the tetramers adopt an oblate structure when viewed from the termini.

In terms of sequence, the antiparallel tetrameric charge-swapped variant of CC-Hex, CC-Hex-KE, is similar to the ferritin-like Ala-coils. To determine whether peptides with sequences similar to rop-like Ala-coils would form parallel α-barrels (like CC-Hex), antiparallel tetramers (like CC-Hex-KE) or indeed just parallel tetramers, two peptides were designed that contained Ala at all of the *g* positions (Figure 3-16). These peptides, 2-LIA-EK and 2-LIA-KE, both contained Leu at *a*, Ile at *d* and Ala at *g*, forming an LIA-core. Charged Glu or Lys residues were placed at the *c*/*e* positions in both orientations (EK, *c*=Glu/*e*=Lys; KE, *c*=Lys/*e*=Glu). As in the peptides 2-LI-EK and 2-LI-KE, these Ala@*g* peptides may be able to form *c*-*e′* interhelical ionic interactions. All *b* positions were also Ala and *f* positions were Gln, Lys or Trp.

CD spectroscopy revealed that both 2-LIA-EK and 2-LIA-KE formed α-helical structures with fraction helix values of 86 and 79 %, respectively (Figure 3-17a). Both peptides were also highly thermally stable with T_M values above the measurable range (Figure 3-17b). The peptides were also slightly more stable than the peptides 2-LI-EK and 2-LI-KE (Figure 3-8a). The sequences of 2-LI-EK and 2-LI-KE are almost identical to those of 2-LIA-EK and 2-LIE-KE, except 2-LI-EK and 2-LI-KE contain Gln at *g* rather than Ala.

When investigated using AUC, in SV and SE experiments both Ala@*g* peptides fitted to molecular weights corresponding to tetramers (Figure 3-17c and d, Figure 8-92, Figure 8-93). In SV experiments, both peptides appeared to form a single species.

Therefore, it appears that the Ala@*g* peptides do not form α-barrels in solution. Unlike the Ala@*e* peptides, CC-Hex and CC-Hex-KE, where the oligomeric state changed from hexamer to tetramer on swapping the charged residues, the Ala@*g* heptad arrangement appears to be compatible only with tetramer formation. However, whether these tetramers are in the parallel or antiparallel orientation remains to be seen. X-ray crystal structures will be needed to determine the helix orientation and to determine whether, if they are indeed forming antiparallel tetramers, the peptides adopt a true Ala-coil fold.




```
                    cdefg abcdefg abcdefg abcdefg ab
2-LIA-EK:  Ac – G EIKfA LAEIKfA LAEIKWA LAEIKfA LA G – NH2
2-LIA-KE:  Ac – G KIEfA LAKIEfA LAKIEWA LAKIEfA LA G – NH2
```

Figure 3-16 Helical wheels and sequences for Ala@*g* peptides. Top: helical wheel diagrams for three proposed structures that could be adopted by 2-LIA-EK. Left: parallel tetramer; centre: antiparallel tetramer; right: α-barrel, parallel hexamer shown as example. Bottom: sequences of peptides 2-LIA-EK and 2-LIA-KE. All peptides were in *c*-register and were N-terminally acetylated and C-terminally amidated.

Both parallel and antiparallel arrangements may be feasible with the 2-LIA-EK and 2-LIA-KE peptides. Parallel tetramers may be possible due to the small, moderately hydrophobic nature of Ala. This residue is often placed at solvent-exposed heptad positions in coiled coils (such as *b* or *c*) without affecting solubility or oligomeric state. However, due to the positioning of Ala at core-flanking *g* positions, it may have become part of an expanded *adg* hydrophobic seam. They may therefore have had more of an effect on core packing, and helix orientation, than Ala residues at more peripheral positions.

In CC-Hex-KE, Glu/Glu and Lys/Lys pairs are brought together at the antiparallel interfaces. Evidently the subsequent charge repulsion is not sufficient to prevent the structure from forming. If 2-LIA-EK and 2-LIA-KE are forming antiparallel structures, these would also contain repulsive Glu/Glu and Lys/Lys interactions at the interhelical interfaces. Conversely, parallel structures would contain productive ionic interactions between the E/K pairs at the *c*/*e* positions. If the Ala@*g* peptides are found to adopt antiparallel structures, optimisation of the locations of the charged residues could be performed to relieve the repulsive interactions.

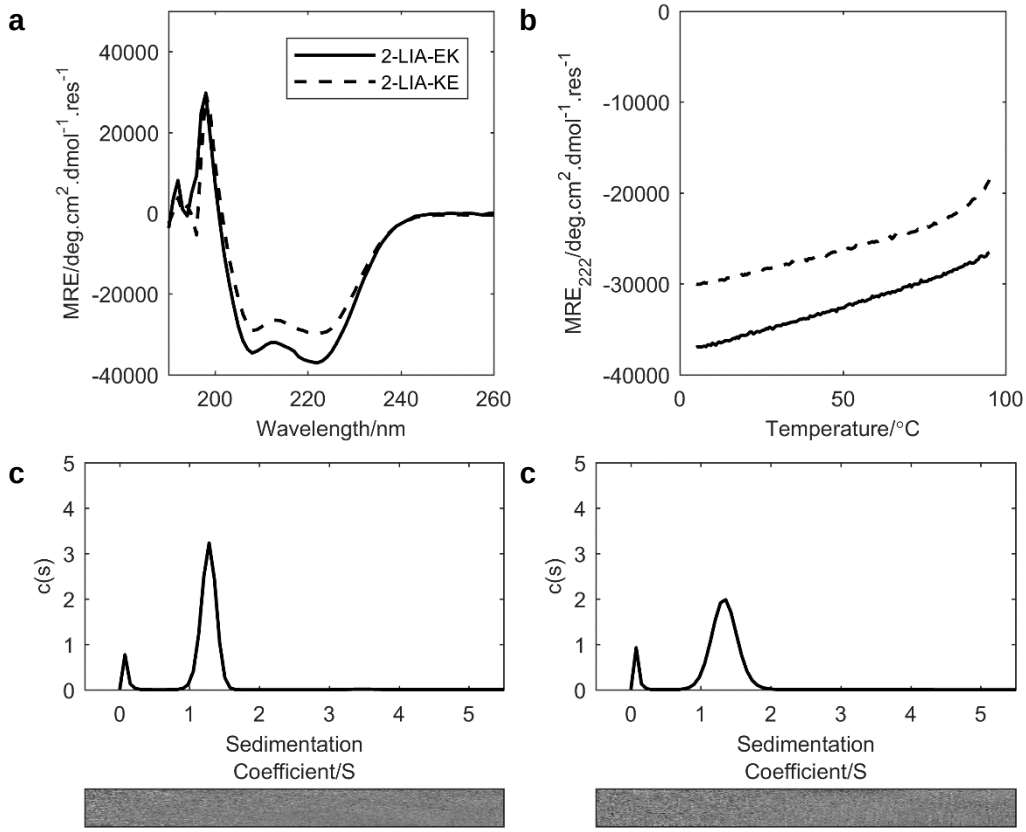


Figure 3-17 Biophysical characterisation of the Ala@*g* peptides 2-LIA-EK and 2-LIA-KE. (a) CD spectra at 5 °C of 2-LIA-EK and 2-LIA-KE. (b) Temperature-dependent CD measurements monitoring MRE_{222} between 5 and 95 °C for 2-LIA-EK and 2-LIA-KE, key as in (a). Peptide concentrations for CD spectroscopy were 10 µM. (c) Sedimentation velocity c(s) distribution (top) and residuals (bottom) for 2-LIA-EK returning a Mw of 12.5 kDa (3.8 x monomer mass). (d) Sedimentation velocity c(s) distribution (top) and residuals (bottom) for 2-LIA-KE returning a Mw of 13.3 kDa (4.1 x monomer mass). All measurements were performed in PBS (pH 7.4).

3.6 Chapter conclusions

To expand the available set of well-characterised, *de novo* tetrameric coiled coils, which previously contained only CC-Tet, a number of peptides have been designed (Figure 3-18). Of these designs, those that contain LI-cores (*a*=Leu/*d*=Ile) reliably form tetramers in solution and in their crystal structures. Tetramers form when charged residues are arranged at the *e*/*g*, *c*/*e*, *b*/*g* and *b*/*c* positions. While the formation of interhelical ionic interactions between charged residues at some *e*/*g* and *c*/*e* positions can be observed in the crystal structures, it remains to be seen whether these interactions form in solution and contribute to coiled-coil stability. Furthermore, tetrameric coiled coils are observed both when polar (Gln) and mildly hydrophobic (Ala) residues are placed at the core-flanking heptad positions. It was

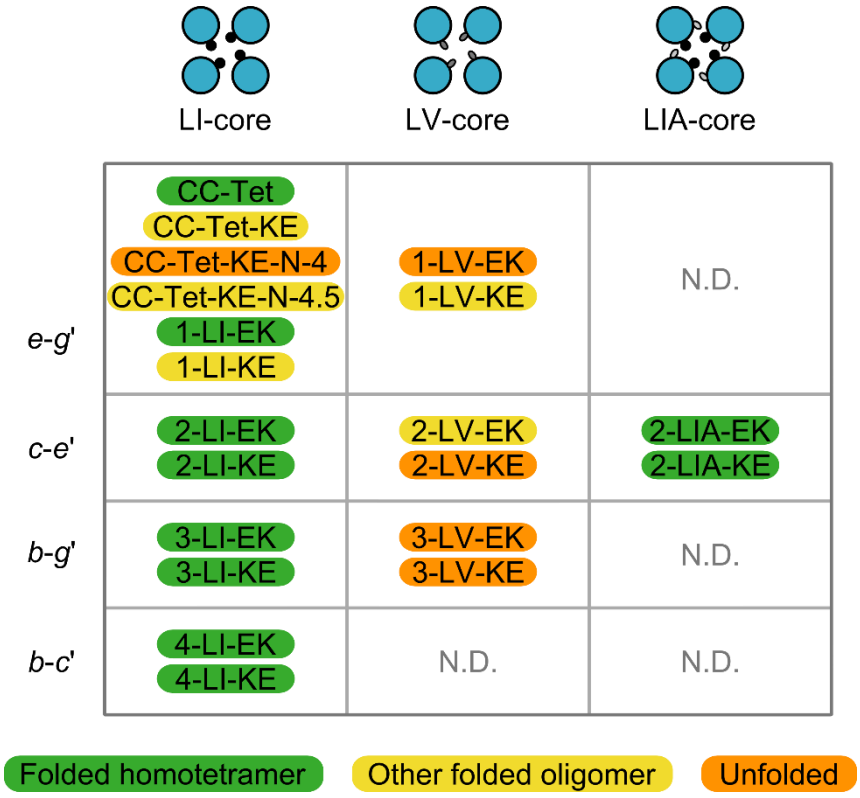


Figure 3-18 Summary of designs discussed in Chapter 3. Successful designs, *i.e.* those that form homotetrameric coiled coils, are highlighted in green. Designs that formed other oligomeric states or were unfolded are highlighted in yellow and orange, respectively. Those designs with *a*=Leu and *d*=Ile were generally the most successful and were tolerant to many different arrangements or charged residues.

generally held that polar residues would be required at these positions to prevent the coiled coils from forming larger α-barrel assemblies, but Ala is also well tolerated at the *g* positions. It is possible that introducing increasingly hydrophobic residues here (*e.g.* Val) could lead to higher-order assemblies. One exception amongst the LI-core designs was when *e*=Glu and *g*=Lys. Peptides designed in both *c*- and *g*-register with this arrangement of charged residues do not form tetramers.

Furthermore, peptides that contain LV-cores (*a*=Leu/*d*=Val) do not form stable tetramers. While it was deemed necessary to enhance the tetramer set with designs that had a greater range of stabilities, the LI-core designs have T_M values spanning < 70 °C to > 95 °C. This range may in fact be suitable for many applications. Alternatively, other approaches to destabilising the coiled coils may be taken such as truncating the sequences of the LI-core designs or considering mixed LI/V-cores.

While the structures of the novel tetramers have been characterised, it would also be helpful to characterise other properties that are of consideration if the coiled coils are to be used for *in vivo* synthetic biology applications. For example: what are their dissociation constants? Are the designs orthogonal to each other? That is, can the coiled coils be used together without cross interacting? Do they still fold *in vivo*? And, perhaps more importantly, are they tolerated *in vivo*?

As well as being tools for synthetic biology in their own right, these homotetramers also represent a step towards more complex tetramer designs such as A_2B_2 heterotetramers and ultimately ABCD heterotetramers. They may even be of use for designing coiled coils with dynamic behaviours that are capable of switching in response to chemical modifications or ligands.

Chapter 4: The design and characterisation of A_2B_2 heterotetramers

4.1 Chapter introduction

Following the successful design of a number of homotetrameric coiled coils, efforts were turned towards A_2B_2 heterotetramers. These contain two copies each of two different components, A and B, which are presumed to be arranged in an alternating "chequerboard" pattern. As artificial PIDs, heteromeric assemblies offer a more attractive target than homomeric ones as they can be used to mediate the interaction between different protein domains rather than simply co-localising multiple copies of the same protein. Therefore, multiple protein functions can be combined in a single assembly, allowing increasingly elaborate functions to be performed. For example, in the context of ATFs, multiple DBDs with different sequence specificities could be combined *via* the heterotetramer to facilitate DNA looping when the protein binds multiple DNA sites. Furthermore, as discussed in Chapters 1 and 3, tetramers and other higher order oligomers have increased response sensitivity.

In the past, novel heterotetrameric coiled coils have generally been achieved by mutating naturally occurring homotetrameric coiled coils to convert them from single-peptide to dual-peptide systems. For example, a portion of the Lac repressor's antiparallel coiled-coil tetramerisation domain, Lac21, and the model parallel tetrameric coiled coil, GCN4-pLI, have been converted into heterotetramers by introducing residues with charged side chains at various core-flanking positions to produce two oppositely charged peptides [192,224-226,337]. These peptides do not fold into homomeric coiled coils due to repulsion from the charged residues, but they are able to form heteromeric coiled coils where charge complementation relieves the repulsion. Aside from the introduced charged residues, the sequences of these coiled coils are generally very similar to those of the naturally derived homomeric parents and, therefore, few truly *de novo* heterotetrameric coiled coils have been reported.

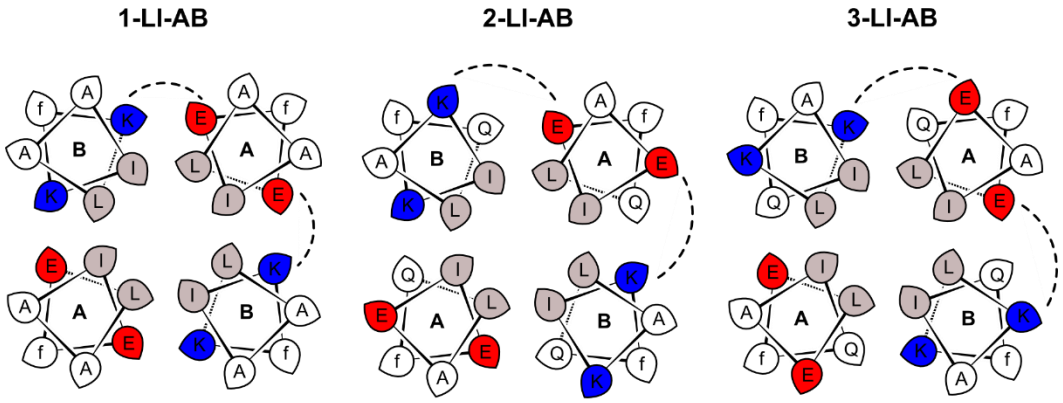
Here, heterotetrameric coiled coils were achieved by designing sets of overall acidic (-ve) and basic (+ve) peptides containing strategically placed charged residues and tetramer-specifying core residues. These peptides were anticipated to form repulsive charge interactions in the homomeric state, disfavouring these species, while repulsion would be relieved when acidic and basic peptides were paired up to form A_2B_2 heterotetramers.

Through this methodical analysis of the various combinations of core residues and flanking charged residues, a set of *de novo* heterotetrameric coiled coils has been designed and characterised. The individual peptides vary widely in their fraction helix values, thermal stability and oligomeric state. Where excessive peptide homomerisation occurs, such species can be weakened by altering the core residues, without impeding heterotetramer formation. When the peptides are combined in acidic/basic pairs, the resulting heterotetramers have a range of thermal stabilities, which appear to be determined predominantly by the identity of the core residues.

4.2 Design and characterisation of Leu/Ile core heterotetramers

4.2.1 Leu/Ile core heterotetramer design

The acidic (A) and basic (B) peptides 1-LI-A, 2-LI-A, 3-LI-A, 1-LI-B, 2-LI-B and 3-LI-B were based on the homotetrameric coiled coil pairs 1-LI-EK/1-LI-KE, 2-LI-EK/2-LI-KE and 3-LI-EK/3-LI-KE, respectively (see Chapter 3). However, rather than containing both Glu and Lys at the core-flanking positions in a single peptide, the acidic and basic peptides were designed to contain either all Glu give overall negatively charged peptides or all Lys to give overall positively charged peptides (Figure 4-1). As in the above homomers, the charged residues were placed at the *e*/*g*, *c*/*e* or *b*/*g* positions and the peptides were also designed to contain LI-cores (*a*=Leu/*d*=Ile) to promote tetramer formation. Where *e* or *g* positions were not a charged residue, these positions were populated by polar Gln to disfavour the formation of higher oligomeric states. Where *b* or *c* positions were not a charged residue, these positions were made Ala to promote helix formation. Thus, this set of six peptides differed in the identity and location of the charged residues, but their sequences were otherwise similar. All six peptides were made in *c*-register. Versions of 1-LI-A and 1-LI-B were also designed in *g*-register and are referred to as 1-LI-A-*g* and 1-LI-B-*g*. The peptides were named for the location



		cdefg	abcdefg	abcdefg	abcdefg	ab		
1-LI-A:	Ac – G	AIEKE	LAAIEKE	LAAIEWE	LAAIEKE	LA	G	– NH_2
2-LI-A:	Ac – G	EIEKQ	LAEIEKQ	LAEIEWQ	LAEIEKQ	LA	G	– NH_2
3-LI-A:	Ac – G	AIQKE	LEAIQKE	LEAIQWE	LEAIQKE	LE	G	– NH_2
1-LI-B:	Ac – G	AIKQK	LAAIKQK	LAAIKWK	LAAIKQK	LA	G	– NH_2
2-LI-B:	Ac – G	KIKQQ	LAKIKQQ	LAKIKWQ	LAKIKQQ	LA	G	– NH_2
3-LI-B:	Ac – G	AIQQK	LKAIQQK	LKAIQWK	LKAIQQK	LK	G	– NH_2

Figure 4-1 A_2B_2 heterotetramer design. Top: helical wheel representations for the heterotetramers 1-LI-AB (left), 2-LI-AB (centre) and 3-LI-AB (right) where dashed lines indicate potential ionic interactions between charged residues. Bottom: sequences for the constituent acidic and basic peptides 1-LI-A, 2-LI-A, 3-LI-A, 1-LI-B, 2-LI-B and 3-LI-B. All peptides were in *c*-register and were N-terminally acetylated and C-terminally amidated. A, acidic; B, basic; Ac, acetyl.

of the charged residues (1, *e*/*g*; 2, *c*/*e*; 3, *b*/*g*), the identity of the core residues (LI, *a*=Leu/*d*=Ile) and the identity of the charged residues (A, acidic, Glu; B, basic, Lys).

When combined in A/B pairs the peptides were anticipated to form the heterotetramers 1-LI-AB (and 1-LI-AB-*g*), 2-LI-AB and 3-LI-AB in which the oppositely charged residues would take part in interhelical *e*-*g*′, *c*-*e*′ or *b*-*g*′ ionic interactions, respectively, where the prime symbol (′) indicates that the residue is located in the adjacent helix (Figure 4-1).

The individual peptides were first characterised separately and then in the various A/B pairings.

4.2.2 Some LI-core peptides show off-target homomerisation

4.2.2.1 CD spectroscopy shows the individual peptides fold to different extents

CD spectra were measured for the individual peptides to investigate whether they formed any homomeric species. When investigated alone at 10 μM each, the peptides 1-LI-A, 1-LI-B, 1-LI-A-*g* and 1-LI-B-*g*, which contained charged residues

at the *e*/*g* positions, were relatively unfolded (Figure 4-2a, Figure 4-4b). Furthermore, the fraction helix values for these peptides were all similar (Figure 4-2d). The 1-LI-B-*g* peptide was slightly more folded than 1-LI-B and so the register of the peptides may have had some effect on how folded the homomers were.

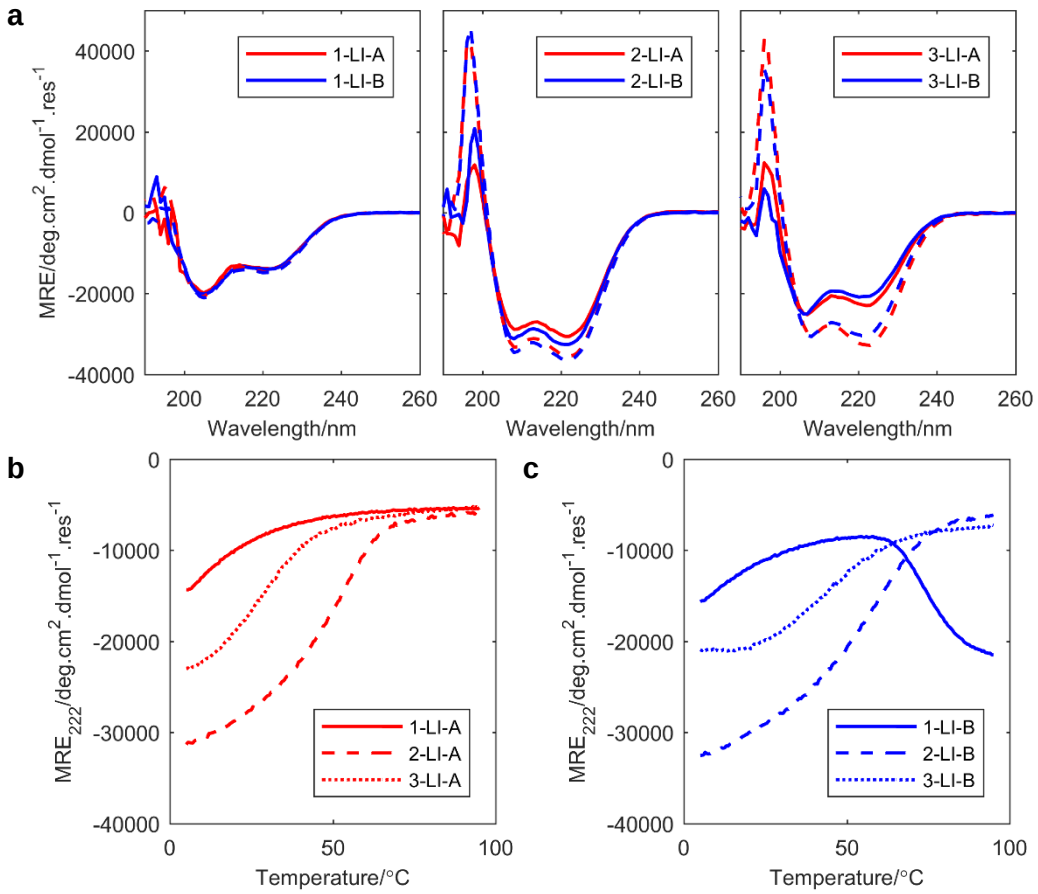
The peptides 2-LI-A and 2-LI-B were more folded than the *e*/*g* peptides and adopted highly α-helical structures with high fraction helix values (Figure 4-2d). Conversely, the peptides 3-LI-A and 3-LI-B had intermediate fraction helix values (Figure 4-2d).

The CD signals for peptides 1-LI-A and 1-LI-B were found to be concentration-independent, *i.e.* their CD spectra did not change as the peptide concentration was increased from 10 to 100 μM (Figure 4-2a). Conversely, the CD signals for 2-LI-A, 2-LI-B, 3-LI-A and 3-LI-B were concentration-dependent and the peptides became more folded as their concentration was increased. For example, at 100 μM 3-LI-A and 3-LI-B reached fraction helix values of 86 and 80 %, respectively.

The temperature-dependent behaviour of the individual peptides was also investigated by heating the samples from 5 to 95 °C while monitoring MRE_{222} (Figure 4-2b and c). The peptides all became fully unfolded as temperature was increased except 1-LI-B, which appeared to become more folded above approximately 60 °C. There was a concern that, on heating, this peptide was forming some β-structure, which can be aggregation-prone. However, when CD spectra were taken at 5 °C intervals during the course of a variable temperature experiment, the peptide appeared to gain α-helical structure rather than β structure (Figure 8-48). While this result was certainly unusual, the peptide in question is unlikely to be used in applications at elevated temperatures, and so this temperature-dependent phenomenon was not of concern.

Disregarding the behaviour of the basic peptide above 60 °C, the 1-LI-A and 1-LI-B peptides had T_M values below the accessible temperature range and were almost fully unfolded at the physiologically relevant temperature of 37 °C (Figure 4-2b and c). The remaining peptides 2-LI-A, 2-LI-B, 3-LI-A and 3-LI-B had a range of T_M values (Figure 4-2d).

The relatively small amount of folding observed for 1-LI-A, 1-LI-B, 1-LI-A-*g* and 1-LI-B-*g* may be explained by the location of their charged residues at the *e* and *g* positions. If these peptides did oligomerise, the charged residues would be close



d

Peptide	MRE$_{222}$ [a]	Fraction helix (%)	T_M (°C)
1-LI-A	-13956 ± 481	37	< 20.0
1-LI-A-*g*	-13367	36	< 20.0
2-LI-A	-30678 ± 2137	80	53.3 ± 0.3
3-LI-A	-23018 ± 990	61	28.2 ± 0.3
1-LI-B	-13936 ± 458	37	< 20.0
1-LI-B-*g*	-15486	41	< 20.0
2-LI-B	-32634 ± 1037	85	59.7 ± 0.3
3-LI-B	-20749 ± 1345	55	47.0 ± 2.6

Figure 4-2 CD spectroscopy data for acidic and basic LI-core peptides. (a) CD spectra at 5 °C of 1-LI-A and 1-LI-B (left), 1-LI-A and 1-LI-B (centre) and 1-LI-A and 1-LI-B (right). Peptides were at 10 (solid lines) or 100 µM (dashed lines). Temperature-dependent CD measurements monitoring MRE$_{222}$ between 5 and 95 °C for (b) 1-LI-A, 2-LI-A and 3-LI-A and (c) 1-LI-B, 2-LI-B and 3-LI-B. Peptides were at 10 µM. (d) MRE$_{222}$, fraction helix values and T_M values for the above peptides at 10 µM. [a] Units, deg.cm^2.dmol^{-1}.res^{-1}. All measurements were performed in PBS (pH 7.4).

to the interface, providing a strong repulsive force that prevents folding of that species. In 2-LI-A and 2-LI-B, which have charged residues at the *e* and *c* positions, these residues would be slightly further apart, resulting in weaker repulsion and allowing these peptides to homomerise.

Finally, in 3-LI-A and 3-LI-B the charged residues at the *b* and *g* positions are not only further apart, but they also point away from each other due to the trajectory of the C_α-C_β bond vector. These peptides were therefore expected to experience the weakest repulsion between the charged residues, which would potentially result in highly folded homomers. Unexpectedly, however, these peptides in fact had intermediate fraction helix and T_M values.

4.2.2.2 AUC shows some of the peptides form oligomers

Due to the extent of folding observed with the peptides 2-LI-A, 2-LI-B, 3-LI-A and 3-LI-B the peptides were analysed using AUC to determine the oligomeric states of these off-target homomers (Figure 4-3, Figure 8-95–Figure 8-100).

The peptides were investigated using both SV and SE experiments (Figure 4-3b, c, e and f, Figure 8-95–Figure 8-100). With the exception of 3-LI-A, there was agreement between the two techniques. The peptides 2-LI-A, 2-LI-B and 3-LI-B all returned molecular weights corresponding to trimers. In SV experiments, 3-LI-A was determined to have a molecular weight of 13.0 kDa, corresponding approximately to a tetramer. However, in SE experiments, the molecular weight was determined to be 11.4 kDa, corresponding approximately to a trimer.

1-LI-A and 1-LI-B were investigated using SE experiments at high speeds as these peptides were anticipated to be monomers and were therefore not expected to sediment at the available rotor speeds (max. 60 krpm) during SV experiments (Figure 4-3a and d). Both peptides returned molecular weights corresponding to monomers. The strong repulsion provided by placing charged residues at the core-proximal *e* and *g* positions appeared to be sufficient to prevent these peptides from oligomerising.

The peptides 2-LI-A, 2-LI-B and 3-LI-B may have formed trimers, despite having tetramer-promoting LI-cores, due to repulsion by the charged residues also: in a tetrameric coiled coil, the charged residues may have been positioned too close to each other leading to strong repulsion that forces the coiled coil into a lower oligomeric state where the repulsion was relieved. Therefore, the low thermal stabilities of the homomeric coiled coils relative to the heteromers may have been due to non-optimal core arrangements and to the repulsive forces between the charged residues. Conversely, 3-LI-A may have been able to adopt a tetrameric oligomeric state because the charged residues at the *b* and *g* positions are

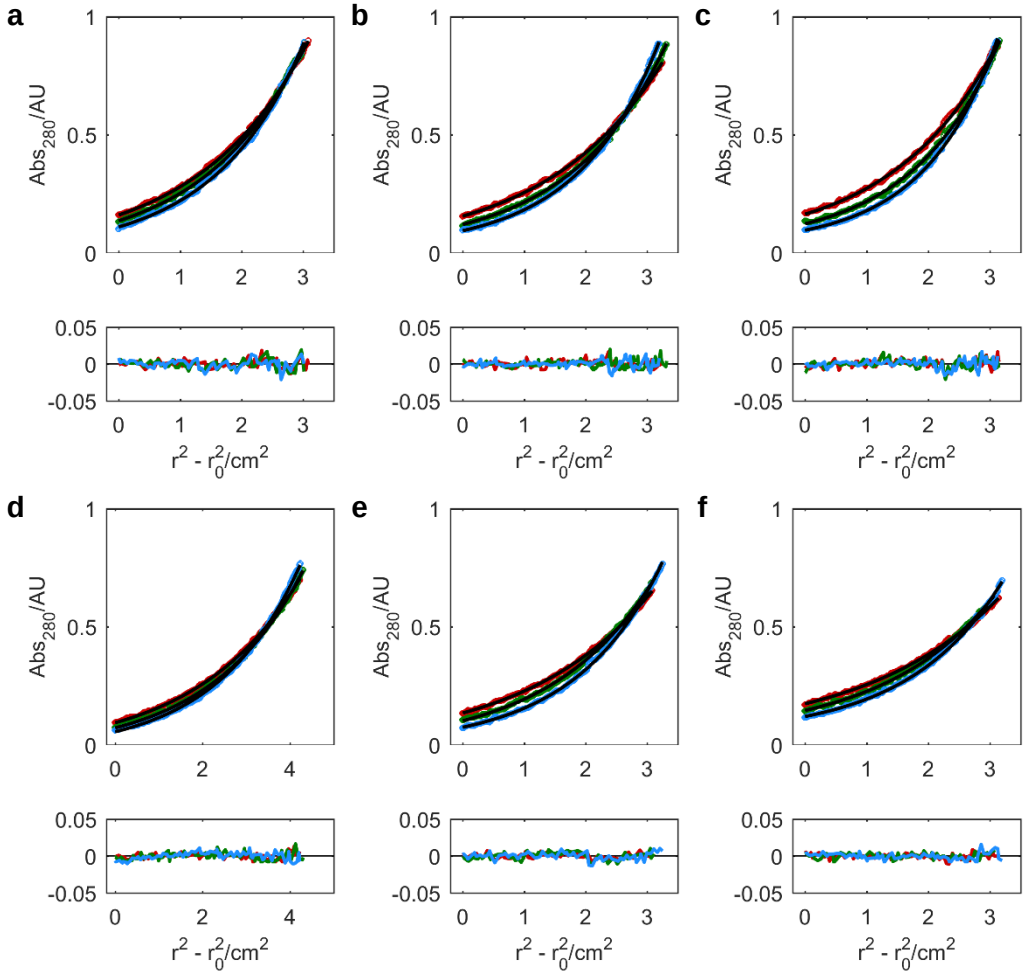


Figure 4-3 Sedimentation equilibrium data for acidic and basic LI-core peptides. Data (red, green and blue circles) and fits (black lines) (top) and residuals (same colours) (bottom) for the peptides: (a) 1-LI-A returning a Mw of 3.6 kDa (1.1 x monomer mass); (b) 2-LI-A returning a Mw of 10.3 kDa (3.0 x monomer mass); (c) 3-LI-A returning a Mw of 11.4 kDa (3.3 x monomer mass); (d) 1-LI-B returning a Mw of 3.2 kDa (1.0 x monomer mass); (e) 2-LI-B returning a Mw of 11.1 kDa (3.2 x monomer mass); (f) 3-LI-B returning a Mw of 10.4 kDa (3.0 x monomer mass). All measurements were performed in PBS (pH 7.4).

positioned further apart in space and are less optimally orientated for interaction in a folded coiled coil. Therefore, the repulsion may have been weaker in this coiled coil, allowing it to form a tetramer, albeit a marginally stable one.

4.2.3 Leu/Ile core designs form heterotetrameric coiled coils in solution

4.2.3.1 CD spectroscopy shows the acidic and basic peptides interact to form α-helical assemblies

When the LI-core peptides were mixed at equimolar concentrations in A/B pairs they all formed highly α-helical heteromeric species (Figure 4-4). The 1-LI-AB,

2-LI-AB and 3-LI-AB pairs had low MRE_{222} values with fraction helix values of 87, 88 and 90 %, respectively (Figure 4-4e). The 1-LI-AB-*g* pair was less folded with a fraction helix value of 80 %. Therefore, it appears that the *c*-register is more optimal for the 1-LI-AB heteromer.

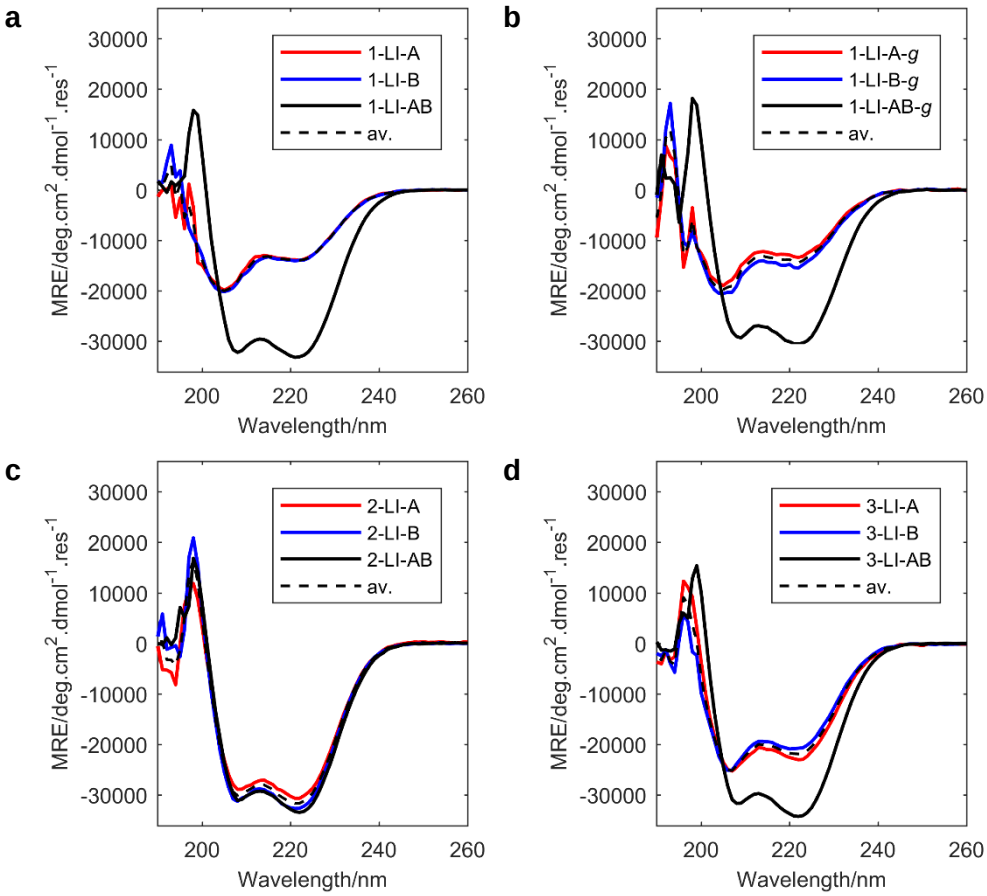
With the exception of the 2-LI-AB pair, in which the A and B monomers were themselves folded into highly α-helical structures, the A/B mixed spectra all showed much greater α-helical character than the constituent peptides, or the average of the spectra of the two peptides alone. This indicated that, on mixing, the peptides were interacting to form a heteromeric species rather than a mix of non-interacting homomeric structures. Because 2-LI-A and 2-LI-B were themselves highly folded, it was not obvious from the CD spectra alone whether these peptides were also interacting to form a heteromeric species or if the observed α-helicity in the mixed spectra was simply due to the individual homomers.

4.2.3.2 The Leu/Ile core heterotetramers are hyperthermally stable

The temperature-dependent behaviour of the heteromers was investigated. At a peptide concentration of 10 μM each, the LI-core heteromers were all hyperthermally stable when heated to 95 °C and did not display unfolding transitions (Figure 4-5). When the peptide concentration was reduced to 1 μM each, the heteromers still did not display unfolding transitions.

All of the heteromers were much more stable than their constituent peptides. This provided further evidence that the peptides interacted to form heteromeric assemblies. For example, 2-LI-A and 2-LI-B, which formed homomers that were difficult to distinguish from the heteromer, had much lower T_M values than the species that formed when the peptides were mixed (Figure 4-2b and c, Figure 4-5).

While full sigmoidal unfolding transitions were not observed within the measurable temperature range, the MRE_{222} did change as the heteromers were heated. For example, for 1-LI-AB at 10 μM, the fraction helix value decreased from 86 % at 5 °C to 61 % at 95 °C, representing an approximately 30 % reduction in the amount of folding. 2-LI-AB and 3-LI-AB showed similar decreases in the amount of folding, becoming 30 and 28 % less folded at 95 °C compared to 5 °C, respectively. The observed loss of α-helicity may have represented fraying of the α helices where



e

Peptide	MRE_{222} [a]	Fraction helix (%)	T_M (°C)
1-LI-AB	-33101 ± 832	87	> 95.0
1-LI-AB-*g*	-30517	80	> 95.0
2-LI-AB	-33459 ± 112	88	> 95.0
3-LI-AB	-34262 ± 1451	90	> 95.0

Figure 4-4 CD spectra for LI-core A and B peptides alone and mixed at equimolar concentrations. (a) 1-LI-A and 1-LI-B. (b) 1-LI-A-*g* and 1-LI-B-*g*. (c) 2-LI-A and 2-LI-B. (d) 3-LI-A and 3-LI-B. Peptides were at 10 µM each and spectra were measured at 5 °C. av., average of the A and B spectra measured alone. (e) MRE_{222}, fraction helix values and T_M values for the above heteromers at 10 µM each. [a] Units, $deg.cm^2.dmol^{-1}.res^{-1}$. All measurements were performed in PBS (pH 7.4).

the ends underwent local unwinding without affecting the global coiled-coil structure. This is observed in many coiled-coil systems.

The temperature-dependent experiments could be performed in the presence of a chemical denaturant, such as guanidinium hydrochloride, in order to access the unfolding transitions of the heteromers. The addition of a chemical denaturant may shift the T_M values to lower, more readily accessible temperatures.

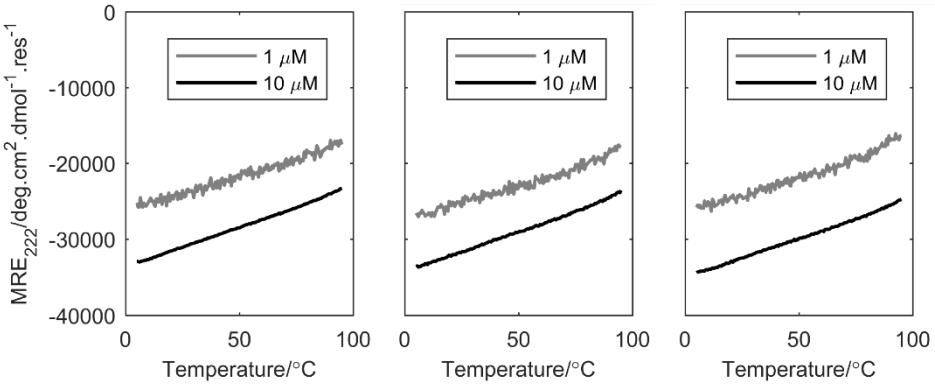


Figure 4-5 Temperature-dependent properties of LI-core heterotetramers. Temperature-dependent CD measurements monitoring MRE_{222} between 5 and 95 °C for 1-LI-AB (left), 2-LI-AB (centre) and 3-LI-AB (right). Peptides were at 1 or 10 µM each. All measurements were performed in PBS (pH 7.4).

While the absence of a sigmoidal unfolding transition can be indicative of hyperthermally stable coiled coils it can also be an indication that a protein does not adopt a unique structure. Instead these coiled coils may exist in molten globular states consisting of folded secondary structure elements (here, α helices) that do not form structured, well-packed cores [350]. In this case, changes in MRE_{222} with increasing temperature may represent rearrangements of the secondary structure elements rather than a loss of protein structure. Hydrophobic dyes such as ANS and its derivatives bind to the easily-accessible and poorly-structured cores of molten globule states and can be used to determine whether or not a protein adopts a unique structure [351]. Measuring whether these heteromers bind this dye may be necessary to determine whether they are indeed hyperthermally stable or are merely unable to adopt specific structures.

4.2.3.3 AUC shows the heteromers form tetramers

The solution-phase molecular weights of the heteromeric species were determined using AUC (Figure 4-6, Figure 8-113–Figure 8-116). In both SV and SE experiments, all the heteromers fitted to molecular weights corresponding to tetramers.

The SV experiments also revealed that the 1-LI-AB, 1-LI-AB-*g* and 2-LI-AB heteromers formed discreet species in solution as indicated by the single, defined peaks observed in their c(s) distributions (Figure 4-6a and b). Conversely, two peaks were observed in the c(s) distribution of 3-LI-AB (Figure 4-6c). The larger of these peaks, which represented the more-prevalent species in solution, was found to have a molecular weight of 14.0 kDa, which was close to the predicted 3-LI-AB

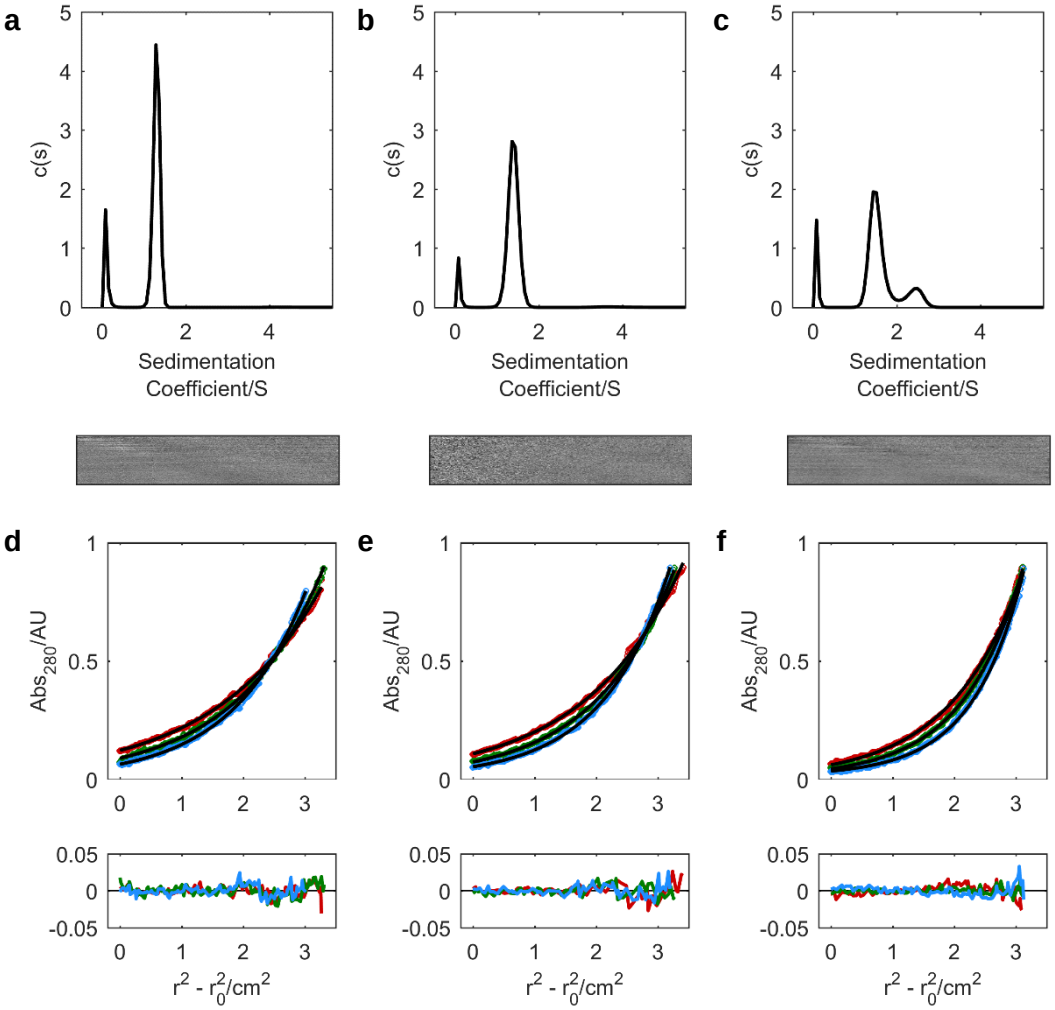


Figure 4-6 AUC data for the heteromers 1-LI-AB, 2-LI-AB and 3-LI-AB. Sedimentation velocity c(s) distributions (top) and residuals (bottom) for: (a) 1-LI-AB returning a Mw of 12.8 kDa (3.9 x mean monomer mass); (b) 2-LI-AB returning a Mw of 13.7 kDa (3.9 x mean monomer mass); (c) 3-LI-AB returning a Mw of 14.0 and 28.4 kDa (4.0 and 8.2 x mean monomer mass). Sedimentation equilibrium data at 30, 33, and 36 krpm (red, green and blue circles) and fits (black lines) (top) and residuals (same colours) (bottom) for: (d) 1-LI-AB returning a Mw of 13.2 kDa (4.1 x mean monomer mass); (e) 2-LI-AB returning a Mw of 13.6 kDa (3.9 x mean monomer mass); (f) 3-LI-AB returning a Mw of 14.2 kDa (4.1 x mean monomer mass). All measurements were performed in PBS (pH 7.4).

heterotetramer mass of 13.9 kDa. The second, smaller peak was found to have a molecular weight of 28.4 kDa corresponding to an octameric oligomeric state. It is interesting that the larger oligomeric state is a multiple of four. This could indicate that the tetramers dimerise.

The SE data for 3-LI-AB were initially analysed by fitting to a single ideal species model, which gave a predicted molecular weight of 14.2 kDa (Figure 4-6f). Subsequently, the data were fitted to a two-species model and to a tetramer-

octamer equilibrium model, however these gave poorer quality fits. It is possible that the octameric species was only present at the slightly higher concentrations used for SV experiments (140 µM total peptide concentration) and was therefore not prevalent in the SE experimental sample (70 µM total peptide concentration). In the future, performing SV experiments at different concentrations may help determine if the observed octamerisation is concentration-dependent: the ratio between the two peaks observed in the c(s) distribution would change if the proportions of the two species changed.

4.3 Design and characterisation of Leu/Val core heterotetramers

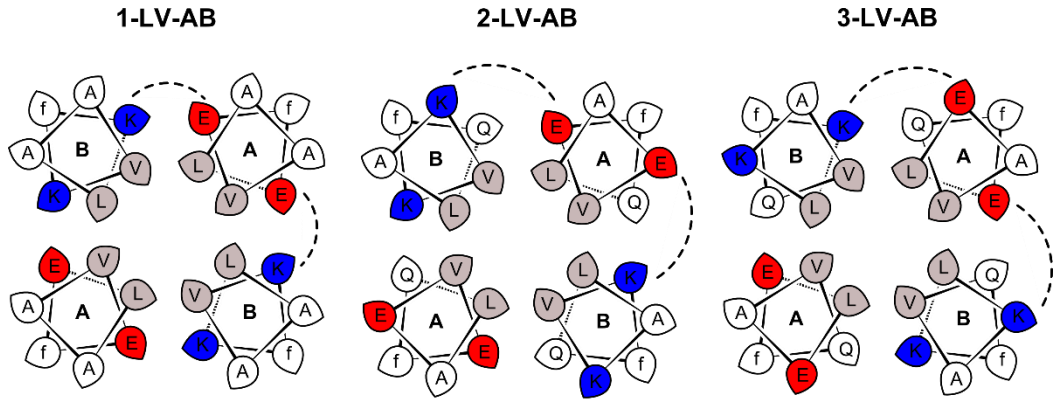
4.3.1 Leu/Val core heterotetramer design

Ideally, the constituent peptides that make up a heteromeric coiled coil should themselves be unstructured. This is because homomeric interactions might lead to unwanted cross interactions between functional domains if the coiled coils were to be used as PIDs *in vivo*.

Given that a number of the LI-core peptides described above formed α-helical homomers when investigated alone, the peptides were redesigned to limit this off-target oligomerisation. This was achieved by changing the residues at the *d* positions in the peptides from Ile to Val to produce the peptides 1-LV-A, 1-LV-B, 2-LV-A, 2-LV-B, 3-LV-A, and 3-LV-B (Figure 4-7). The peptide sequences were otherwise identical to the LI-core parent peptides and follow the same naming scheme (LV, *a*=Leu/*d*=Val). Like Ile, Val is a β-branched amino acid and so may also promote tetramer formation when included at the *d* positions. However, Val is less hydrophobic than Ile, so should produce less-stable coiled coils [340]. Both the off-target homomers and the heteromers were anticipated to be destabilised by the inclusion of Val.

4.3.2 The Leu/Val core peptides are monomeric in solution

The LV-core acidic and basic peptides were all highly unfolded when analysed alone using CD spectroscopy (Figure 4-8a). With the exception of 2-LV-B, the peptides did not gain α-helical character on increasing the peptide concentration and the CD spectra at 10 and 100 µM were essentially identical. The fraction helix values for 2-LV-B at 10 and 100 µM were 19 and 25 %, respectively. All LV-core



cdefg abcdefg abcdefg abcdefg ab

1-LV-A: Ac – G AVEKE LAAVEKE LAAVEWE LAAVEKE LA G – NH_2
2-LV-A: Ac – G EVEKQ LAEVEKQ LAEVEWQ LAEVEKQ LA G – NH_2
3-LV-A: Ac – G AVQKE LEAVQKE LEAVQWE LEAVQKE LE G – NH_2

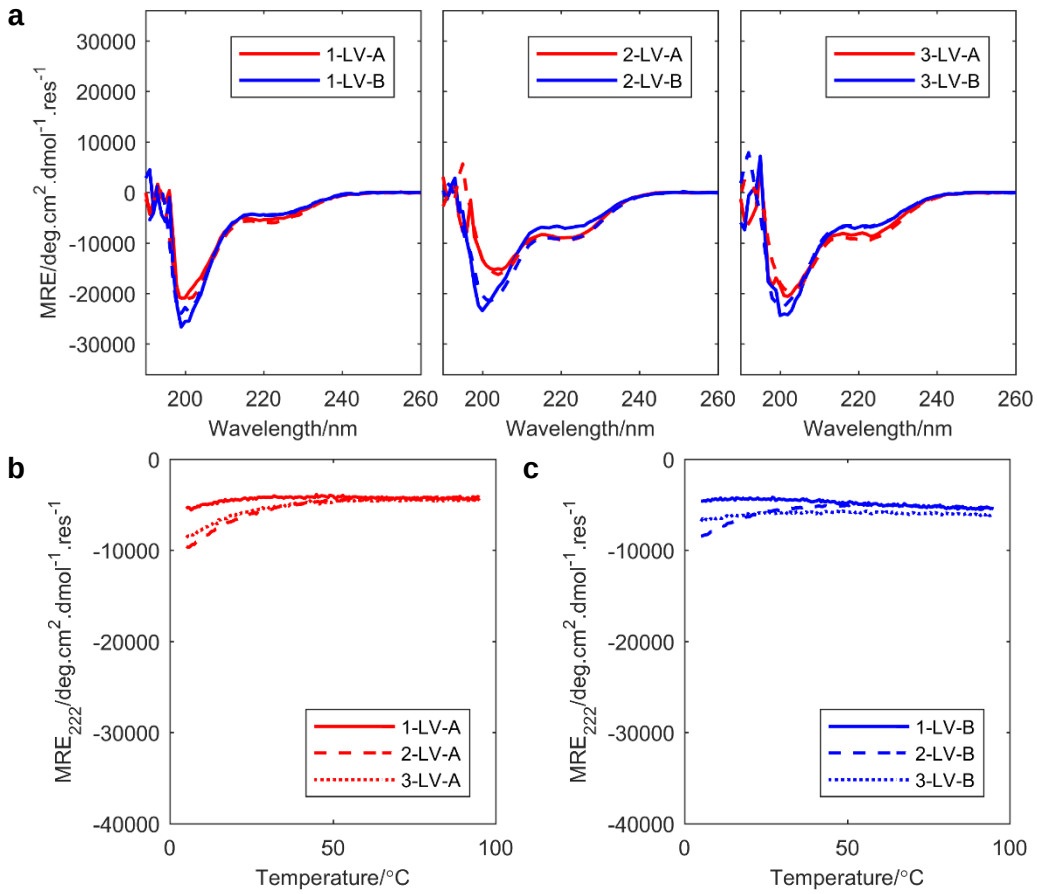
1-LV-B: Ac – G AVKQK LAAVKQK LAAVKWK LAAVKQK LA G – NH_2
2-LV-B: Ac – G KVKQQ LAKVKQQ LAKVKWQ LAKVKQQ LA G – NH_2
3-LV-B: Ac – G AVQQK LKAVQQK LKAVQWK LKAVQQK LK G – NH_2

Figure 4-7 LV-core A_2B_2 heterotetramer design. Top: helical wheels for the heterotetramers 1-LV-AB (left), 2-LV-AB (centre) and 3-LV-AB (right) where dashed lines indicate potential ionic interactions between charged residues. Bottom: sequences for the constituent acidic and basic peptides 1-LV-A, 2-LV-A, 3-LV-A, 1-LV-B, 2-LV-B and 3-LV-B. All peptides were in *c*-register and were N-terminally acetylated and C-terminally amidated. A, acidic; B, basic; Ac, acetyl.

peptides were less folded than the corresponding LI-core parents. The three acidic peptides appeared to be slightly more folded than their basic partners.

The temperature-dependent properties of the individual peptides were also investigated, and all were found to have very low thermal stabilities at 10 μM (Figure 4-8b and c). It was not possible to extract T_M values from the data as they fell far below the measurable temperature range.

SE experiments were performed on all LV-core acidic and basic peptides and all were found to be monomeric in solution which was likely due to a combination of charge repulsion and the weakened LV-cores (Figure 4-9, Figure 8-101–Figure 8-106).



d

Peptide	MRE_{222} [a]	Fraction helix (%)	T_M (°C)
1-LV-A	-5340 ± 536	15	< 20.0
2-LV-A	-8920 ± 262	24	< 20.0
3-LV-A	-8243 ± 222	22	< 20.0
1-LV-B	-4574 ± 242	13	< 20.0
2-LV-B	-6997 ± 637	19	< 20.0
3-LV-B	-6584 ± 621	18	< 20.0

Figure 4-8 CD spectroscopy data for acidic and basic LV-core peptides. (a) CD spectra at 5 °C of 1-LV-A and 1-LV-B (left), 2-LV-A and 2-LV-B (centre) and 3-LV-A and 3-LV-B (right). (b) Peptides were at 10 (solid lines) or 100 µM (dashed lines). Temperature-dependent CD measurements monitoring MRE_{222} between 5 and 95 °C for (b) 1-LV-A, 2-LV-A and 3-LV-A and (c) 1-LV-B, 2-LV-B and 3-LV-B. Peptides were at 10 µM. (d) MRE_{222}, fraction helicites and T_M values for the peptides at 10 µM. [a] Units, $deg.cm^2.dmol^{-1}.res^{-1}$. All measurements were performed in PBS (pH 7.4).

4.3.3 Leu/Val core designs form heterotetrameric coiled coils in solution

4.3.3.1 CD spectroscopy shows the acidic and basic peptides interact to form α-helical assemblies

When the LV-core peptides were mixed in their cognate pairs at equimolar concentrations, the CD spectra showed that they all formed highly α-helical species

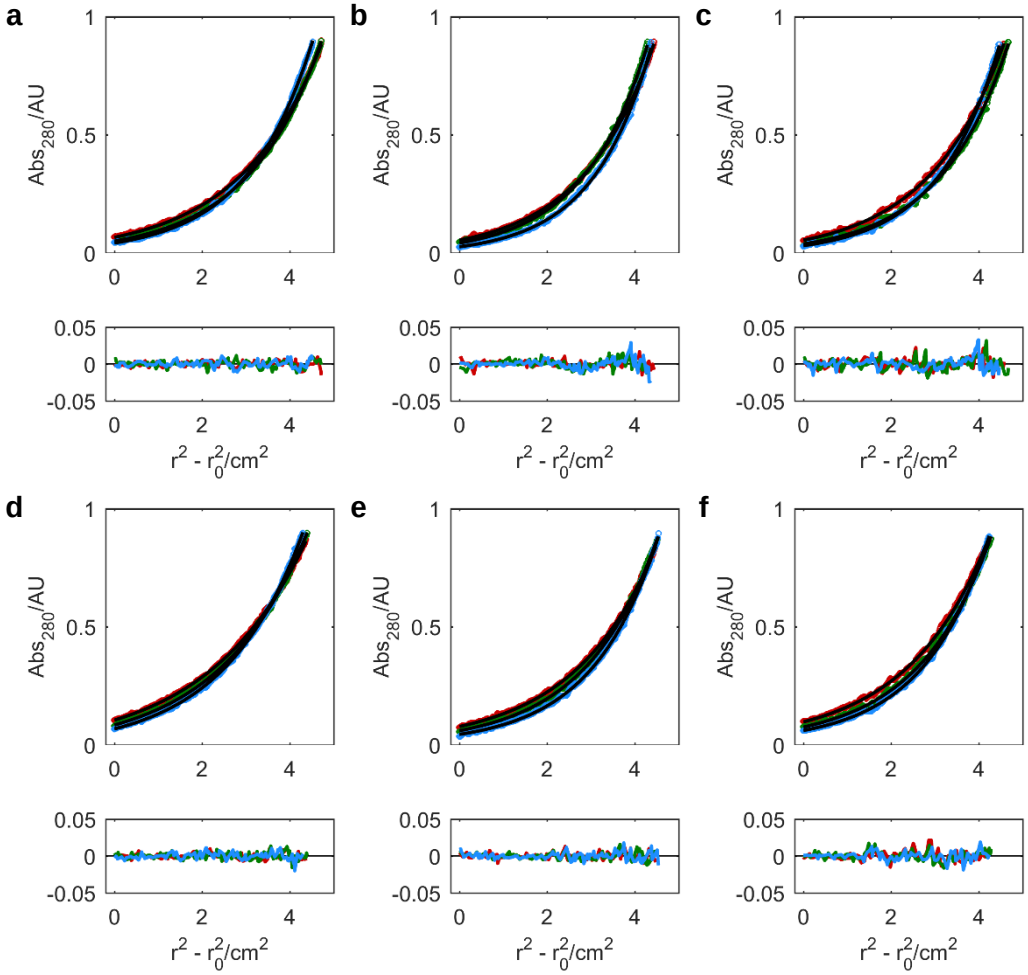
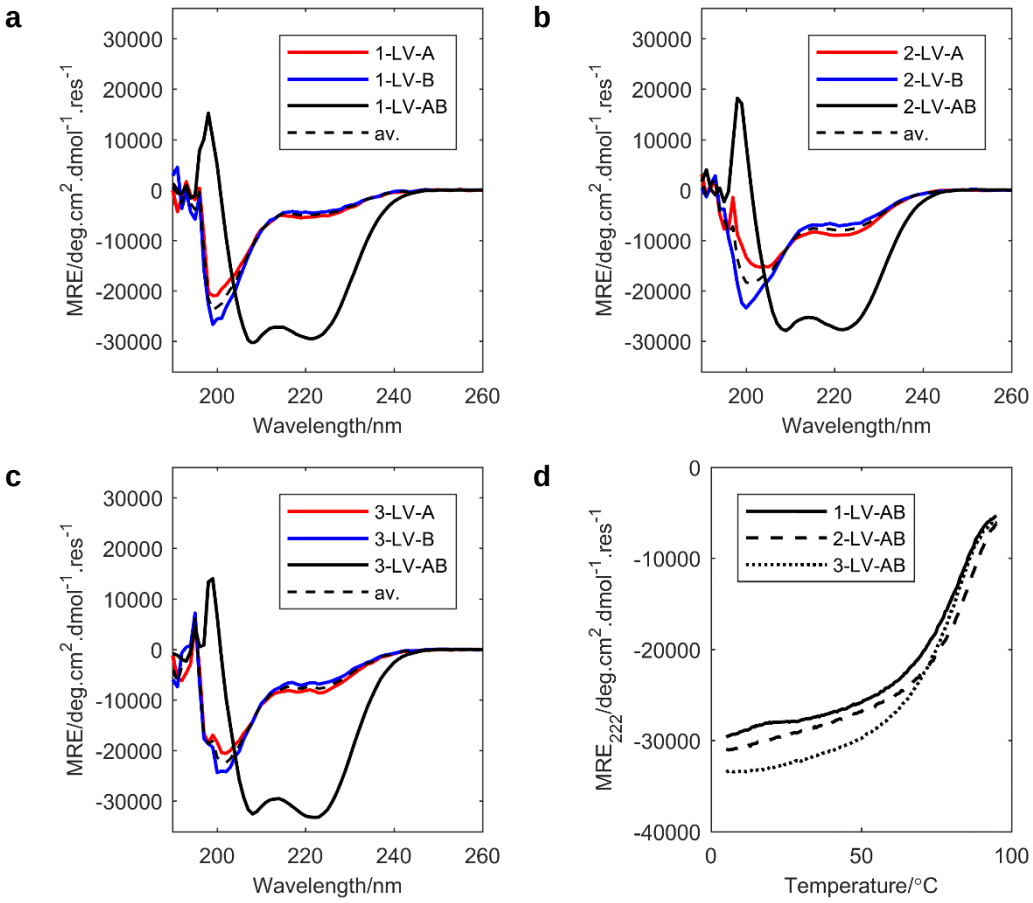


Figure 4-9 Sedimentation equilibrium data at 54, 57 and 60 krpm (red, green and blue circles) and fits (black lines) (top) and residuals (same colours) (bottom) for the peptides: (a) 1-LV-A returning a Mw of 3.3 kDa (1.0 x monomer mass); (b) 2-LV-A returning a Mw of 3.6 kDa (1.1 x monomer mass); (c) 3-LV-A returning a Mw of 3.4 kDa (1.0 x monomer mass); (d) 1-LV-B returning a Mw of 3.2 kDa (1.1 x monomer mass); (e) 2-LV-B returning a Mw of 3.6 kDa (1.1 x monomer mass); (f) 3-LV-B returning a Mw of 3.7 kDa (1.1 x monomer mass). All measurements were performed in PBS (pH 7.4).

(Figure 4-10). The mixed spectra demonstrated significantly more folding than the averages of the spectra for the individual peptides, showing that these peptides also interacted to form heteromeric coiled coils (Figure 4-10a-c).

These heteromers, 1-LV-AB, 2-LV-AB and 3-LV-AB, had fraction helix values of 77, 73 and 87 %, respectively, at 10 µM (Figure 4-10e). As anticipated due to the lower hydrophobicity of Val, these were lower than the fraction helix values for the corresponding LI-core heteromers, 1-LI-AB, 2-LI-AB and 3-LI-AB, at the same concentration. When the peptide concentration was increased to 100 µM, the



e

Peptide	MRE_{222} [a]	Fraction helix (%)	T_M (°C)
1-LV-AB	-29451 ± 1753	77	83.2 ± 1.5
2-LV-AB	-27719 ± 284	73	86.2 ± 1.4
3-LV-AB	-33245 ± 669	87	82.2 ± 0.8

Figure 4-10 CD spectroscopy data for LV-core A and B peptides alone and mixed at equimolar concentrations. CD spectra at 5 °C for: (a) 1-LV-A and 1-LV-B; (b) 2-LV-A and 2-LV-B; (c) 3-LV-A and 3-LV-B. Peptides were at 10 μM each. av., average of the A and B spectra measured alone. (d) Temperature-dependent CD measurements monitoring MRE_{222} between 5 and 95 °C for LV-core heteromers 1-LI-AB, 2-LV-AB and 3-LV-AB. Peptides were at 10 μM each. (e) MRE_{222}, fraction helicities and T_M values for the heteromers at 10 μM each. [a] Units, $deg.cm^2.dmol^{-1}.res^{-1}$. All measurements were performed in PBS (pH 7.4).

fraction helix values for the LV-core heteromers increased to 81, 80 and 88 %, respectively.

Temperature-dependent CD measurements revealed that the LV-core heteromers were also highly thermally stable, although they were all less stable than the corresponding LI-core heteromers (Figure 4-10d). The LV-core heteromers displayed sigmoidal unfolding transitions demonstrating a cooperative unfolding

reaction and T_M values between 80 and 90 °C were extracted for all three (Figure 4-10e).

4.3.3.2 AUC shows the heteromers form tetramers

SV and SE experiments showed that all three LV-core heteromers formed monodisperse tetramers demonstrating that the Ile to Val core substitutions have not compromised the formation of heterotetramers (Figure 4-11, Figure 8-117–Figure 8-121).

This is in contrast to the LV-core homomers discussed in Chapter 3, where the introduction of the Leu/Val core led to coiled coils with various different oligomeric states. That LV-core heteromers form tetramers despite this core also being compatible with other oligomeric states may be due to the intrinsic 2-fold symmetry of the A_2B_2 heterotetramer designs. In a heteromer, the highly charged A and B monomers are unlikely to be arranged adjacent to other monomers with the same charge, but they may be arranged in an alternating “chequerboard” pattern that relieves the charge repulsion. The resulting assembly must therefore contain an even number of components. Heterodimers are unlikely to form due to the presence of β-branched Val at the *d* positions – which would cause steric clashes in dimers – and higher oligomeric states such as hexamers are unlikely due to the polar residues at the *e* and *g* positions. Therefore, a tetramer may in fact be the only oligomeric state available to these assemblies.

The LV-core heterotetramers offer many improvements over the LI-core heterotetramers in terms of both monomer and coiled-coil behaviour: the acidic and basic peptides are highly unfolded monomers in isolation that, when mixed, also interact to form heterotetramers. These heterotetramers demonstrate cooperative unfolding transitions indicating that they adopt a specific conformation and, while they are slightly less thermally stable than the LI-core parents, all have T_M values above 80 °C so they are therefore more thermally stable than most natural proteins.

4.4 Mixed Leu/Ile/Val core heterotetramers have intermediate stabilities

Given that LI-core heterotetramers were hyperthermally stable with T_M values far outside of the accessible temperature range, and LV-core heterotetramers had

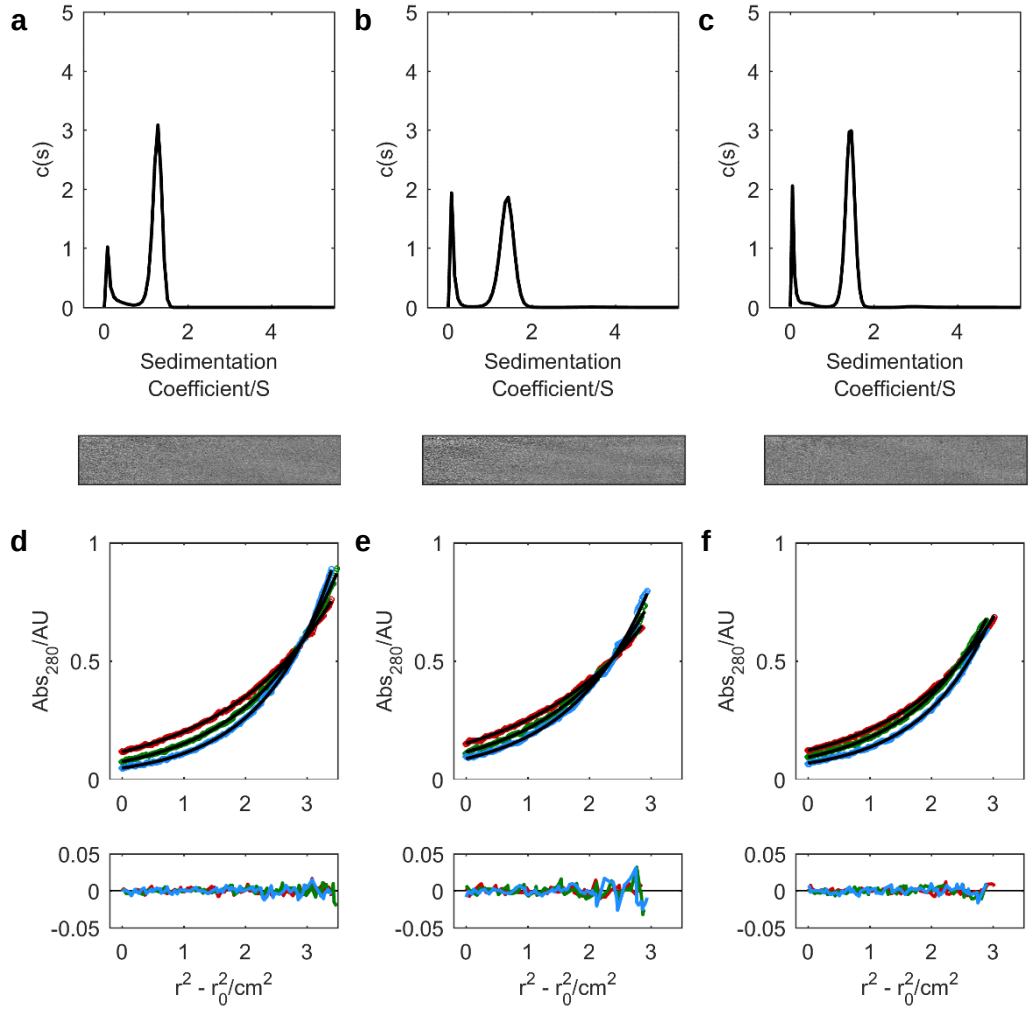


Figure 4-11 AUC data for the heteromers 1-LV-AB, 2-LV-AB and 3-LV-AB. Sedimentation velocity c(s) distributions (top) and residuals (bottom) for: (a) 1-LV-AB returning a predicted Mw of 12.6 kDa (4.0 x mean monomer mass); (b) 2-LV-AB returning a predicted Mw of 13.1 kDa (4.1 x mean monomer mass); (c) 3-LV-AB returning a predicted Mw of 12.0 kDa (3.8 x mean monomer mass). Sedimentation equilibrium data (red, green and blue circles) and fits (black lines) (top) and residuals (same colours) (bottom) for: (d) 1-LV-AB returning a predicted Mw of 12.1 kDa (3.8 x mean monomer mass); (e) 2-LV-AB returning a predicted Mw of 12.7 kDa (4.0 x mean monomer mass); (f) 3-LV-AB returning a predicted Mw of 11.7 kDa (3.7 x mean monomer mass). All measurements were performed in PBS (pH 7.4).

relatively lower T_M values of 80–90 °C, coiled coils with mixed Leu/Ile/Val cores were anticipated to have intermediate stabilities. These mixed-core heterotetramers were generated by mixing A and B peptides with LI or LV cores to generate six additional combinations: 1-LI-A/1-LV-B, 1-LV-A/1-LI-B, 2-LI-A/2-LV-B, 2-LV-A/2-LI-B, 3-LI-A/3-LV-B and 3-LV-A/3-LI-B. The resulting heteromers would have mixed cores but a single charge configuration, *i.e.* charged residues at only *e*/*g*, *c*/*e* or *b*/*g* positions.

The CD spectra of these mixed-core combinations showed that they all formed highly α-helical structures (Figure 4-12a, c and e). SE and SV experiments showed that all were tetrameric in solution (Figure 8-122–Figure 8-127).

The temperature-dependent behaviours of the mixtures were also measured, and they were indeed found to have thermal stabilities between those of the LI- and LV-core heterotetramers (Figure 4-12b, d and f). All of the mixed-core heterotetramers had T_M values above the accessible range. Nonetheless, the beginnings of the unfolding transitions could be observed, which was a promising indication that the coiled coils formed specific structures, albeit highly stable ones.

The effect was most pronounced with the *e*/*g* set (consisting of 1-LI-AB, 1-LI-A/1-LV-B, 1-LV-A/1-LI-B and 1-LV-AB, Figure 4-12b) where more of the unfolding transition could be observed. However, the effect became less obvious for the *e*/*c* and *b*/*g* sets where the mixed-core heterotetramers appeared to become more stable (Figure 4-12d and f). In the future, reducing the peptide concentration could make the unfolding transition accessible and potentially allow the T_M values to be measured.

The above mixed-core heterotetramers were made by combining peptides where the charged residues were at the same positions so that they only differed at the core residues. Hybrid heteromers, where the charged residues were located at different positions, were also investigated (Figure 4-13). For example, 1-LI-A/2-LV-B, as well as containing a mixed Leu/Ile/Val core, contained *e*-*g*′ interactions at two of the four interhelical interfaces, and *e*-*c*′ interactions at the remaining two interfaces (Figure 4-13a, far left).

The resulting hybrid heteromers were all highly α-helical and were all found to be tetrameric using SE and SV (Figure 4-13b, Figure 8-128–Figure 8-131). They were also highly thermally stable with T_M values outside of the accessible range (Figure 4-13c). However, like the initial mixed-core heterotetramers described above, they all displayed the beginnings of unfolding transitions. These heterotetramers would also benefit from investigation at lower peptide concentrations.

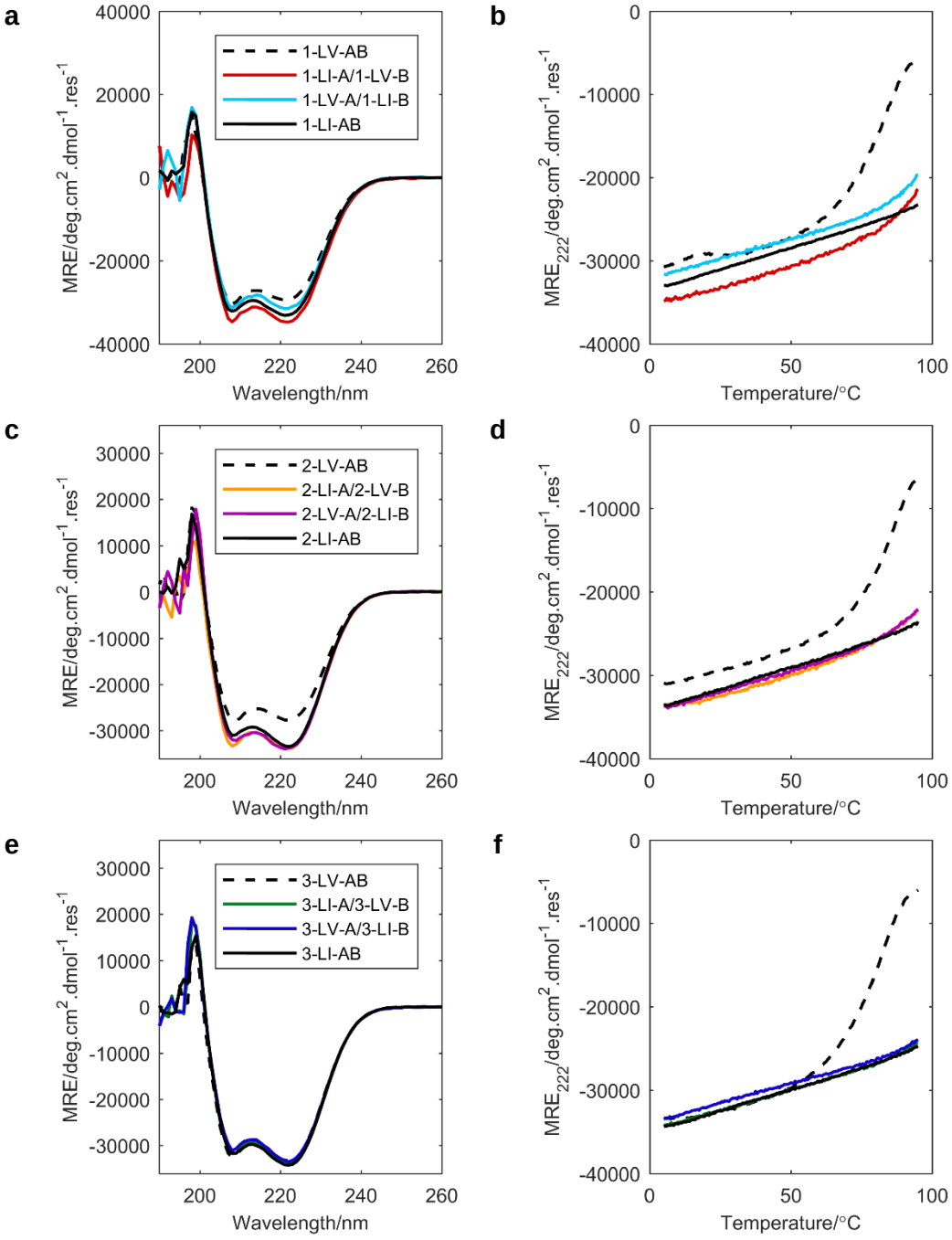


Figure 4-12 CD spectroscopy data for LI- LV and mixed-core heteromers. CD spectra at 5 °C for: (a) 1-LI-AB, 1-LV-AB, 1-LI-A/1-LV-B and 1-LV-A/1-LI-B; (c) 2-LI-AB, 2-LV-AB, 2-LI-A/2-LV-B and 2-LV-A/2-LI-B; (e) 3-LI-AB, 3-LV-AB, 3-LI-A/3-LV-B and 3-LV-A/3-LI-B. Temperature-dependent CD measurements monitoring MRE_{222} between 5 and 95 °C for the same heteromers. Keys as in (a), (c) and (e). Peptides were at 10 µM each. All measurements were performed in PBS (pH 7.4).

While the nature of the ionic interactions may contribute to the differences in stability observed within this group, the differences in stability between this group

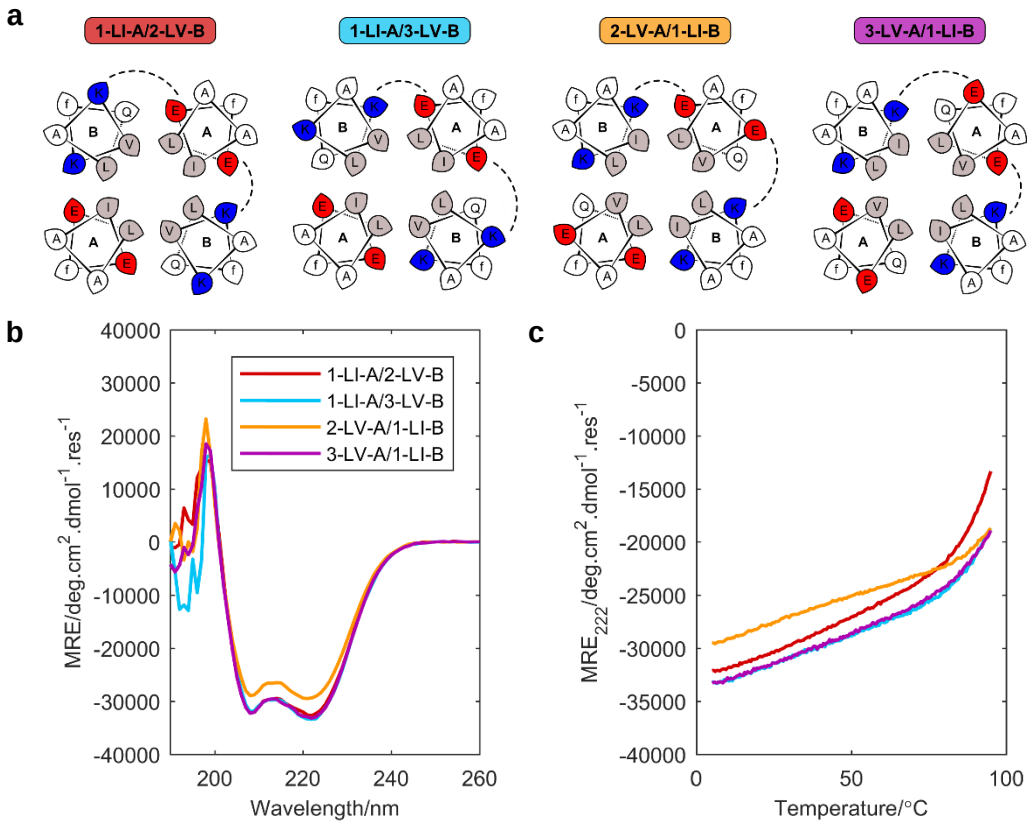


Figure 4-13 CD spectroscopy data for mixed-core heteromers containing mixed *e-g'*/*c-e'* and *e-g'*/*b-g'* ionic interactions. (a) Helical wheels for 1-LI-A/2-LV-B, 1-LI-A/3-LV-B, 2-LV-A/1-LI-B and 3-LV-A/1-LI-B. Dashed lines indicate potential interhelical ionic interactions. (b) CD spectra at 5 °C for the same heteromers. (c) Temperature-dependent CD measurements monitoring MRE_{222} between 5 and 95 °C for the same heteromers, key as in (b). Peptides were at 10 µM. All measurements were performed in PBS (pH 7.4).

of coiled coils and the LI- and LV-core coiled coils is probably driven largely by the identities of the core residues.

4.5 Factors contributing to heterotetramer stability

4.5.1 Glu-Lys interhelical interactions contribute to 1-LV-AB stability

Previous analyses of *de novo* heterodimeric coiled coils, such the peptide Velcro ACID-p1/BASE-p1 system [188] and the EE/KK peptide system [352], have provided conflicting results regarding the contribution of interhelical ionic interactions to heteromeric coiled-coil stability. These systems consist of oppositely charged, unfolded peptides that interact to form heteromeric coiled coils when mixed. It is unclear whether the charged residues in such systems contribute to heteromer stability (1) through the formation of specific interhelical ionic interactions; or (2)

through charge repulsion in the individual peptides that is relieved in the heteromeric state. Previous analysis of designed heterotetrameric coiled coils suggests that specific interhelical ionic interactions do contribute to heterotetramer stability [226].

Investigations into the contribution of interhelical ionic interactions to coiled-coil stability are often performed by measuring the coiled-coil properties under different pH or ionic strengths. If ionic interactions do indeed contribute to stability, the coiled coil would decrease in stability when the charged residues were screened, for example under high salt conditions, and increase in stability when charged residues were exposed, for example under low/no salt conditions.

Therefore, the heterotetramer 1-LV-AB was investigated under buffered conditions at varying ionic strengths to assess the contribution of the interhelical Glu-Lys salt bridges to coiled-coil stability, and to determine whether the design was robust to changing conditions. Specifically, CD spectroscopy and AUC were performed in 10 mM sodium phosphate buffer (pH 7.4) with 0.0 or 1.0 M NaCl (Figure 4-14, Figure 4-15).

At concentrations of either 1 or 5 µM, 1-LV-AB was more folded in the absence of salt than in its presence (Figure 4-14a). At 5 µM peptide concentration, 1-LV-AB had fraction helix values of 77 and 68 % in the presence of 0.0 and 1.0 M NaCl, respectively. At 1 µm, these values were 69 and 52 %. Furthermore, the temperature-dependent CD measurements revealed that 1-LV-AB had dramatically different T_M values under the different salt conditions (Figure 4-14b). At 5 µM peptide concentration and with no salt 1-LV-AB had a T_M above the measurable range but with 1.0 M NaCl the T_M was 77.5 °C. At 1 µM peptide concentration the T_M obtained from the sample with no salt was also above the measurable range whereas in the presence of 1.0 M NaCl it was 60.0 °C. While the unavailability of the T_M values in the absence of salt meant that ΔT_M values could not be determined, these values must have been > 17.5 and > 35 °C at 5 and 1 µM, respectively, indicating that coiled-coil stability was significantly decreased by electrostatic screening in buffer with increased ionic strength. Therefore, the charged residues do appear to contribute to the stability of this heterotetramer through the formation of interhelical salt bridges.

When the peptides 1-LV-A and 1-LV-B were investigated alone using CD spectroscopy they were both highly unfolded under high and low salt conditions

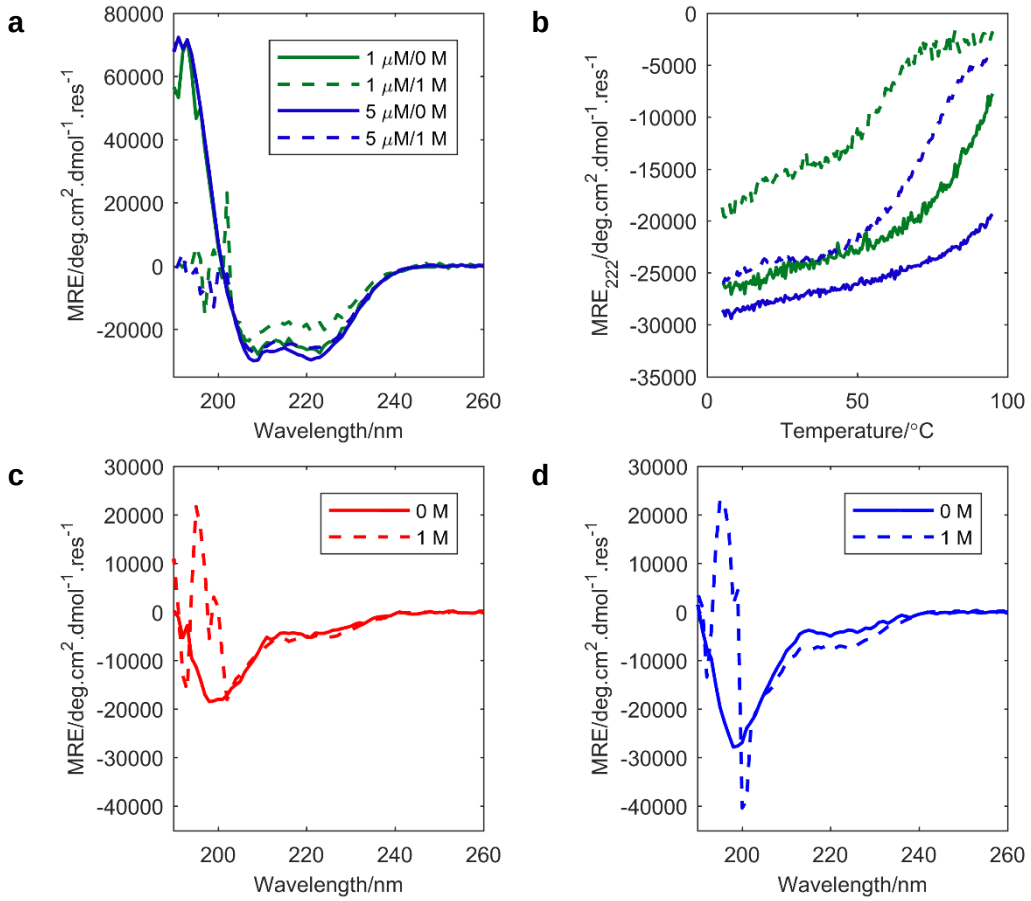


Figure 4-14 CD spectroscopy of 1-LV-A and 1-LV-B alone and mixed at equimolar concentrations in the presence of 0.0 or 1.0 M NaCl. (a) CD spectra at 5 °C for heterotetramer 1-LV-AB. (b) Temperature-dependent CD measurements monitoring MRE_{222} between 5 and 95 °C for heterotetramer 1-LV-AB, key as in (a). Peptides were present at 1 or 5 µM each. CD spectra at 5 °C of (c) 1-LV-A and (d) 1-LV-B. Peptides were at 5 µM. All measurements were performed in 10 mM sodium phosphate buffer (pH 7.4) in the presence of 0.0 (solid lines) or 1.0 M NaCl (dashed lines).

(Figure 4-14c and d). 1-LV-B became slightly more folded in the presence of 1.0 M NaCl.

In SV and SE experiments, 1-LV-AB was found to be tetrameric under both salt conditions (Figure 4-15, Figure 8-120, Figure 8-121). However, in the presence of 1.0 M NaCl, the peak representing the heterotetramer in the c(s) distribution was slightly broader than the peak observed in the absence of salt (Figure 4-15a and b). Furthermore, under the high salt conditions, the predicted molecular weights were slightly lower than the actual molecular weight with both AUC techniques. These observations may be related to the fact that 1-LV-AB was less well folded under the high salt conditions and therefore the presence of increased amounts of unfolded monomers may have influenced the AUC fits.

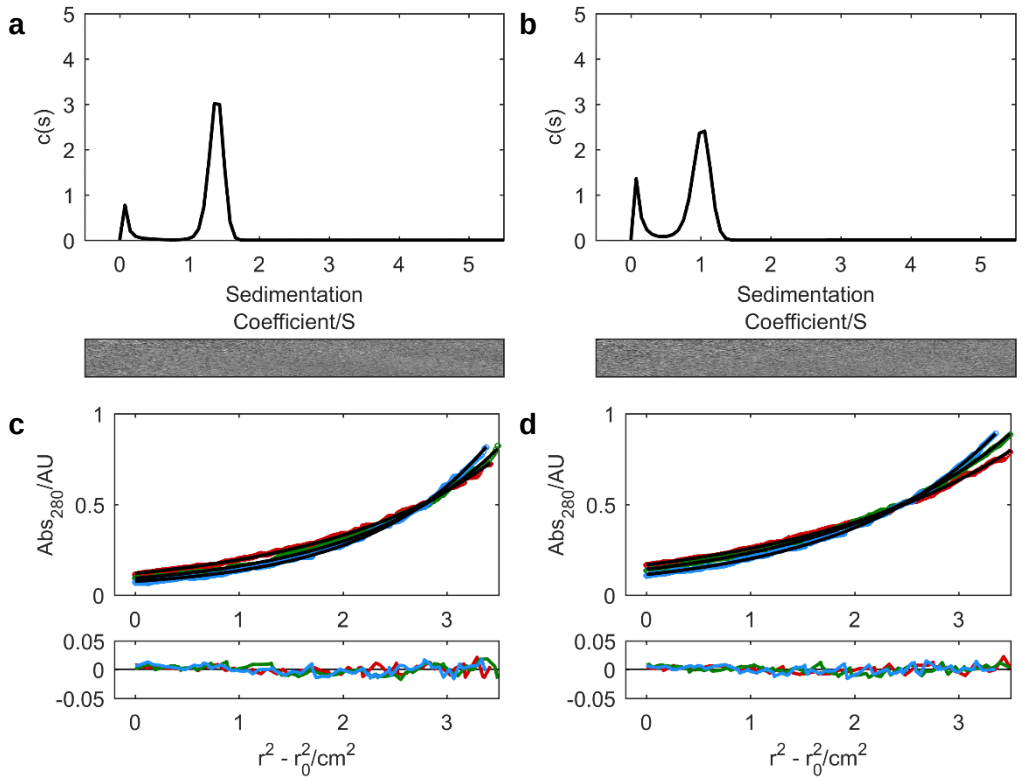


Figure 4-15 AUC data for 1-LV-AB in the presence of 0.0 or 1.0 M NaCl. Sedimentation velocity c(s) distributions (top) and residuals (bottom) for 1-LV-AB with (a) 0.0 or (b) 1.0 M NaCl returning Mw of 13.1 kDa (4.1 x mean monomer mass) and 12.0 kDa (3.8 x mean monomer mass), respectively. Sedimentation equilibrium data at 30, 33, and 36 krpm (red, green and blue circles) and fits (black lines) (top) and residuals (same colours) (bottom) for 1-LV-AB with (c) 0.0 or (d) 1.0 M NaCl returning Mw of 12.7 kDa (4.0 x mean monomer mass) and 11.7 kDa (3.7 x mean monomer mass), respectively. All measurements were performed in 10 mM sodium phosphate buffer (pH 7.4) in the presence of 0.0 or 1.0 M NaCl

Given that the oligomeric state of 1-LV-AB did not change with increasing ionic strength it was likely that the presence of the interhelical ionic interactions did not influence coiled-coil specificity. That is, formation of the salt bridges did not have a role in specifying the oligomeric state of the coiled coil, which was probably dictated primarily by the identity of the core residues as well as by the need to position the charged peptides in an alternating arrangement.

In summary, the charged residues in the 1-LV-A and 1-LV-B peptides appear to be able to form interhelical ionic interactions in the heterotetramer 1-LV-AB which contribute to coiled-coil stability but not specificity. The absolute energetic contribution of the salt bridges could not be quantified due to the high thermal stability of the coiled coil in the absence of salt, even at very low peptide concentrations. Furthermore, on masking the charged residues with high salt,

1-LV-B becomes slightly more folded, presumably because Lys-Lys repulsion is relieved. This may imply that relief from the repulsive homomeric interactions on formation of the heteromeric species also provides some driving force for heterotetramer formation.

It would be interesting to investigate whether the other LV-core peptides are stabilised under high salt conditions. It would also be useful to know whether the charged residues in the other LV-core heterotetramers also contribute to coiled-coil stability and, if so, which configuration of charged residues (*e*/*g*, *c*/*e* or *b*/*g*) provides the greatest increase in thermal stability.

Finally, though interactions between charged residues can be observed in the crystal structures of homomeric coiled coils, it can be unclear whether they form in solution or contribute to coiled-coil stability. Therefore, it may be interesting to investigate the homotetrameric coiled coils described in Chapter 3, for example 1-LV-EK and 1-LV-KE, in a manner analogous to the one described above.

4.5.2 Heterotetramers are more stable than the equivalent homotetramers

Intriguingly, the heterotetrameric coiled coils were consistently more highly folded and more thermally stable than the equivalent homomers with the same core residues and the same configuration of charged residues (Figure 4-16).

The heteromers and homomers were investigated using CD spectroscopy at a total peptide concentration of 20 μM (Figure 4-16). For some of the coiled coil sets the differences were very pronounced. For example, the homomers 3-LV-EK and 3-LV-KE had fraction helix values of just 10 and 11 %, respectively, while the equivalent LV-core heteromer, 3-LV-AB, had a fraction helix value of 87 % (Figure 4-16e). In all cases the LV-core heteromers were also more thermally stable than the corresponding homomers (Figure 4-16b, d and f). Furthermore, although all of the LI-core homomers and heteromers formed highly α-helical assemblies, the homomers were also less thermally stable than the heteromers.

Homomers and heteromers were expected to have similar stabilities given that they contained the same core residues, are the same length, and contain the same numbers of charged residues so have the same potential to form interhelical ionic interactions. Why this was not the case is unclear, but it may provide further indication that, on formation of a heterotetramer, relief from the repulsive

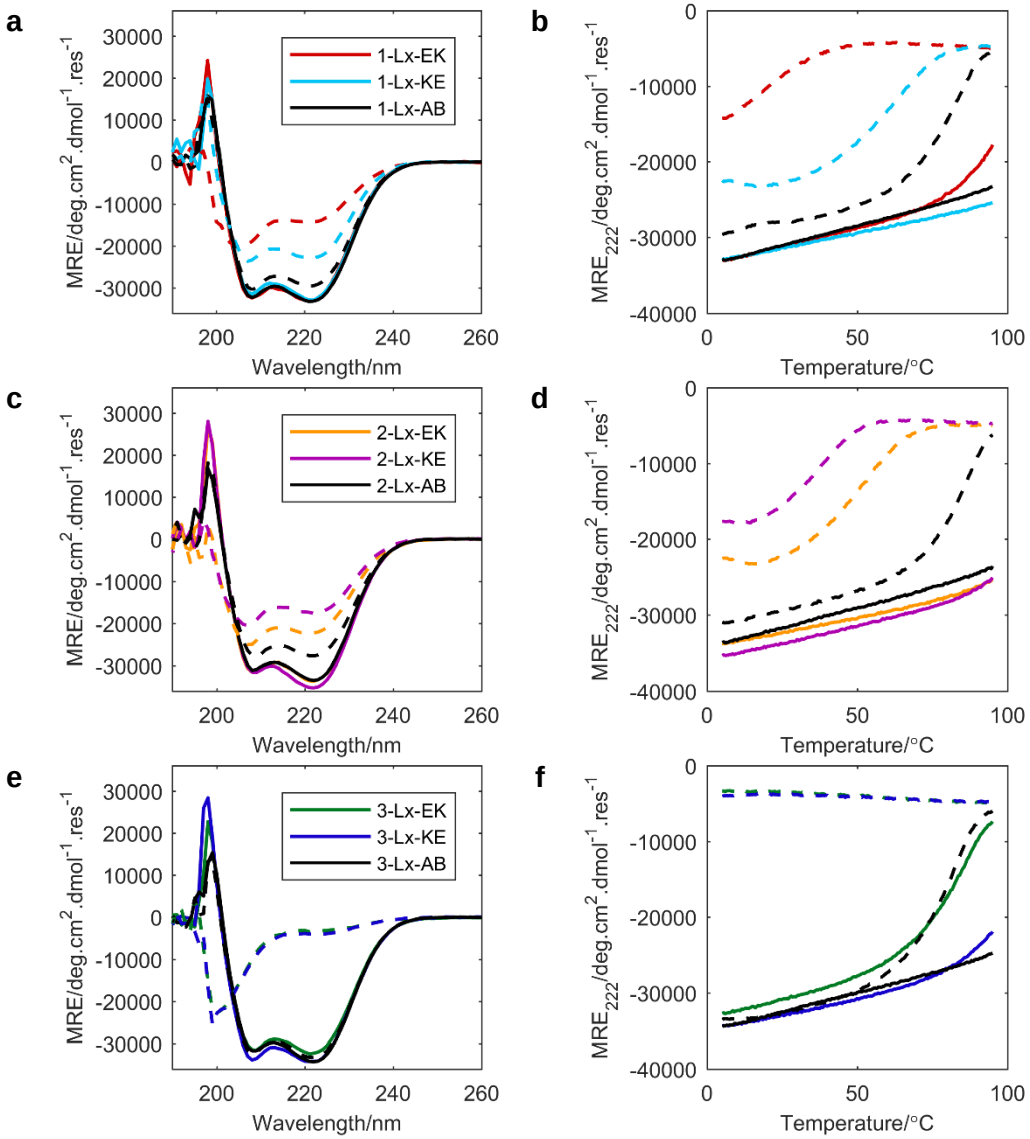


Figure 4-16 CD spectroscopy data for homo- and heterotetramers. CD spectra at 5 °C for: (a) 1-LI-EK, 1-LV-EK, 1-LI-KE, 1-LV-KE, 1-LI-AB and 1-LV-AB; (c) 2-LI-EK, 2-LV-EK, 2-LI-KE, 2-LV-KE, 2-LI-AB and 2-LV-AB; (e) 3-LI-EK, 3-LV-EK, 3-LI-KE, 3-LV-KE, 3-LI-AB and 3-LV-AB. Temperature-dependent CD measurements monitoring MRE_{222} between 5 and 95 °C for the same peptides/heteromers: (b), (d) and (f). Keys as in (a), (c) and (e). Solid lines, LI-core designs; dashed lines, LV-core designs. The total peptide concentration was 20 μM. All measurements were performed in PBS (pH 7.4).

interactions present in the individual acidic and basic components provides an additional driving force for heteromer folding that is not present in the homomers.

4.6 The effect of non-polar residues at *g* in heterotetramers

As introduced in the Chapter 3, placing non-polar Ala residues at core-flanking *e* and *g* positions has previously been found to lead to the formation of higher oligomeric states such as hexamers and heptamers [213,345] or to the formation of antiparallel structures [222,223]. When Ala was placed at the *g* positions of the homotetrameric peptides 2-LI-EK and 2-LI-KE the resulting peptides, 2-LIA-EK and 2-LIA-KE, still formed homotetramers (Section 3.5). Whether they were forming parallel tetramers or antiparallel rop-like Ala-coil structures was not determined [348].

To investigate the effect of introducing core-flanking Ala residues to heterotetramers, Ala was placed at the *g* positions of the peptides 2-LI-A, 2-LI-B, 2-LV-A and 2-LV-B to produce the peptides 2-LIA-A, 2-LIA-B, 2-LVA-A and 2-LVA-B, respectively. These peptides were combined to make the A/B heteromers 2-LIA-AB and 2-LVA-AB. As with the homomers, removing the polar Gln residues from the *g* positions could have resulted in coiled coils with higher oligomeric states, antiparallel tetramers or parallel tetramers.

*4.6.1 Ala@*g *peptides may form higher order homomers*

When the acidic and basic peptides were investigated alone at 10 µM using CD spectroscopy, 2-LIA-A and 2-LIA-B were highly folded into α-helical structures with fraction helix values of 86 and 79 %, respectively (Figure 4-17a and b). These values were similar to the fraction helix values for the parent peptides 2-LI-A (80 %) and 2-LI-B (85 %) (Figure 4-2d). The LV-core peptides 2-LVA-A and 2-LVA-B were less folded at the same concentration with fraction helix values of 44 and 23 %, respectively (Figure 4-17a, Figure 4-18b). The fraction helix value for 2-LVA-B was also similar to that of the parent 2-LV-B (19 %), while 2-LVA-A was more folded than its parent 2-LV-A which had a fraction helix value of 24 % (Figure 4-8d).

The Ala@*g* peptides also demonstrated a concentration-dependent increase in folding (Figure 4-17a and b). At 100 µM peptide concentration, 2-LIA-A and 2-LIA-B were highly α-helical with fraction helix values of 92 and 87 %, respectively. At the higher concentration folding of 2-LVA-A and 2-LVA-B also increased to 77 and 46 %, respectively.

The temperature-dependent behaviour of the peptides was investigated. At 10 µM 2-LIA-A had a T_M value of 74 °C, however 2-LIA-B was much more stable than its

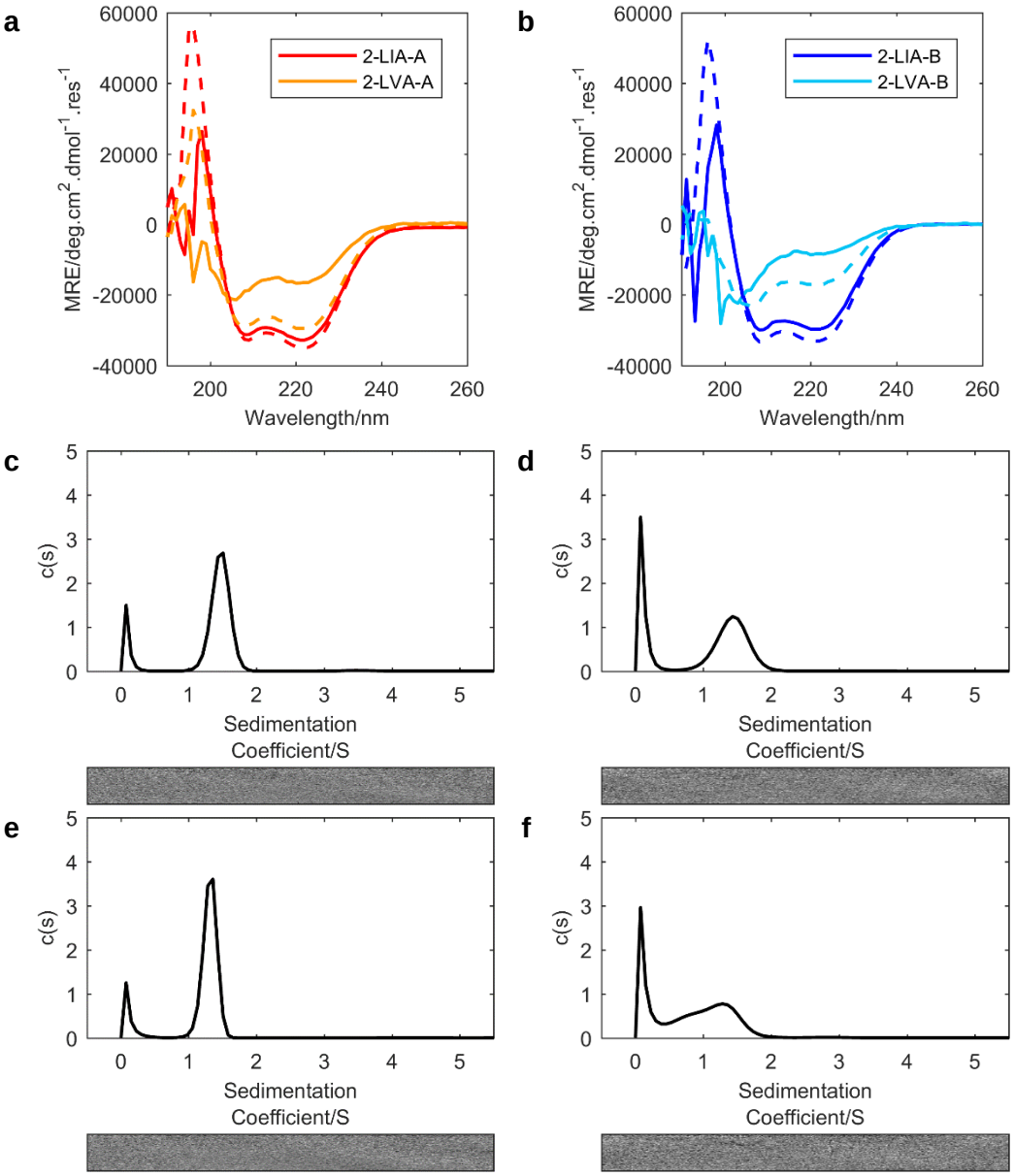


Figure 4-17 Biophysical characterisation of acidic and basic Ala@*g* peptides. CD spectra at 5 °C for (a) 2-LIA-A and 2-LVA-A and (b) 2-LIA-B and 2-LVA-B. Peptides were at 10 (solid lines) or 100 µM (dashed lines). Sedimentation velocity c(s) distributions (top) and residuals (bottom) for: (c) 2-LIA-A returning a Mw of 14.2 kDa (4.4 x monomer mass); (d) 2-LVA-A returning a Mw of 13.6 kDa (4.3 x monomer mass); (e) 2-LIA-B returning a Mw of 14.5 kDa (4.5 x monomer mass); (f) 2-LVA-B returning a Mw of 12.7 kDa (4.0 x monomer mass). All measurements were performed in PBS (pH 7.4).

acidic partner with a T_M value above the accessible range. At the same concentration, 2-LVA-A had a much lower T_M of 28 °C. Conversely, 2-LVA-B had a T_M below the measurable range.

When the peptides were analysed using SE and SV these experiments returned relatively ambiguous oligomeric state assignments (Figure 4-17c-f, Figure 8-109–

Figure 8-112). However, the peptides generally appeared to form oligomeric states of tetramers and larger. This was in contrast to the parent LI- and LV-core peptides, which formed trimers and monomers, respectively.

Furthermore, while single, sharp oligomer peaks were observed in the c(s) distributions of 2-LIA-A and 2-LIA-B, the peaks observed for 2-LVA-A and 2-LVA-B were much broader. This may indicate that the peptides formed a range of close oligomeric states that could not be resolved as separate peaks. In the c(s) distribution for 2-LVA-B particularly there appears to be a small, overlapping peak below 1 S, which may correspond to a smaller species.

2-LIA-B had a single, clear peak in its c(s) distribution corresponding to an oligomeric state of 4.5. The predicted oligomeric state from SE for this peptide was also 4.5. This peptide may form a mix of tetramers and pentamers and, therefore, widening the hydrophobic seam through the addition of alanine may have allowed this peptide to access higher oligomeric states.

*4.6.2 Ala@*g *peptides interact to form heterotetramers*

When the acidic and basic peptides were mixed at equimolar concentrations the resulting 2-LIA-AB and 2-LVA-AB pairs were highly α-helical with fraction helix values of 90 and 89 %, respectively (Figure 4-18a and b). The spectra of the mixtures were also more highly α-helical than the averages of the individual peptide spectra, indicating the peptides interacted. Both A/B heteromers were highly thermally stable, with T_M values above the accessible range (Figure 4-18c and d). Like the LI-core heterotetramers discussed in Section 4.2.3.2, these Ala@*g* heteromers did not display even the beginnings of an unfolding transition.

In both SE and SV experiments the 2-LIA-AB and 2-LVA-AB heteromers returned molecular weighs corresponding to tetramers in solution (Figure 4-19, Figure 8-132, Figure 8-133). The c(s) distributions indicated that each heteromer formed a single species in solution (Figure 4-19a and b). Therefore, like the homotetramers 2-LIA-EK and 2-LIA-KE, these heteromers assemblies also appear to form tetramers in solution despite having sequences that could be compatible with the formation of higher oligomeric states.

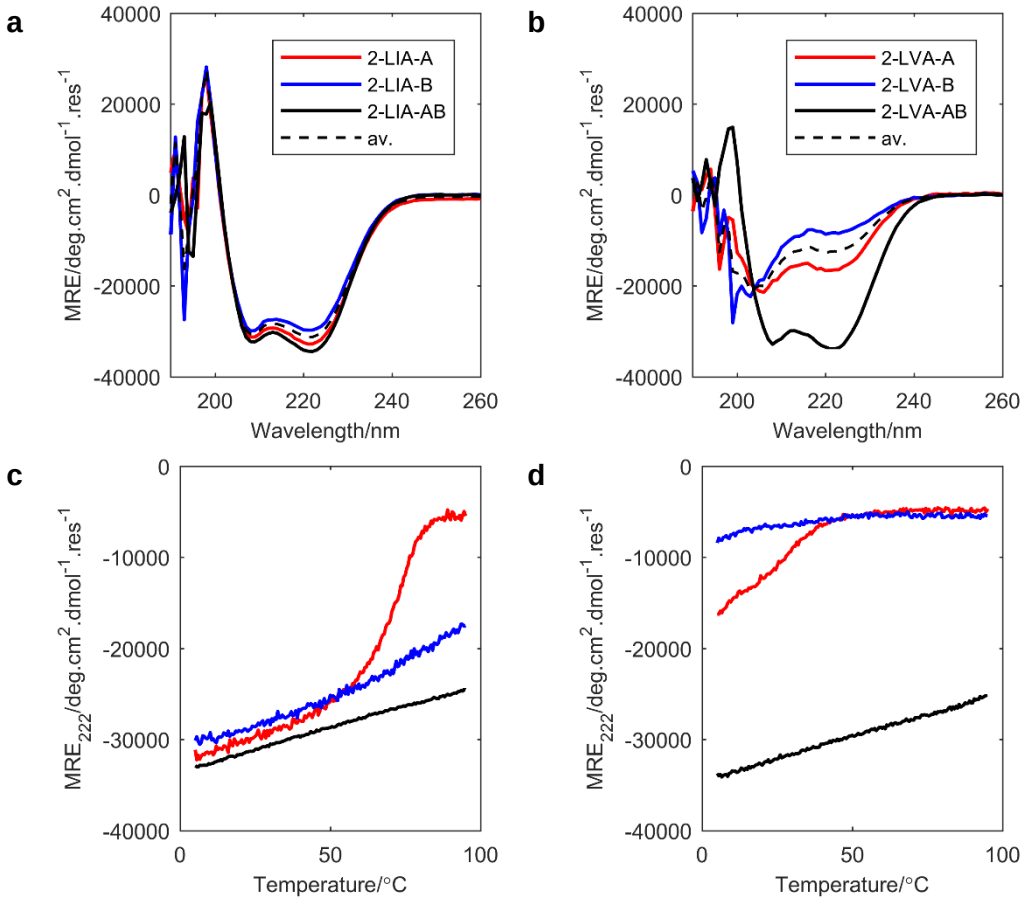


Figure 4-18 CD spectroscopy data for Ala@*g* peptides alone and mixed at equimolar concentrations. CD spectra at 5 °C for (a) 2-LIA-A and 2-LIA-B and (b) 2-LVA-A and 2-LVA-B. Temperature-dependent CD measurements monitoring MRE_{222} between 5 and 95 °C for the same peptides/heteromers (c), (d). Keys as in (a) and (b). Peptides were at 10 µM each. All measurements were performed in PBS (pH 7.4).

While no crystallographic data could be obtained for the heteromers to verify their structures, it is possible that these Ala@*g* heteromers may be forming antiparallel coiled coils, due to the similarity of their sequences to rop-like Ala-coils. In the absence of structural data, techniques such as disulphide exchange with Cys-labelled peptides could provide information on helix orientation. For example, A and B peptides labelled with Cys at the same terminus could interact to form disulphide-linked A-A, B-B and A-B dimers if the peptides were oriented in the parallel orientation. In the antiparallel orientation, only the A-A and B-B dimers would be observed.

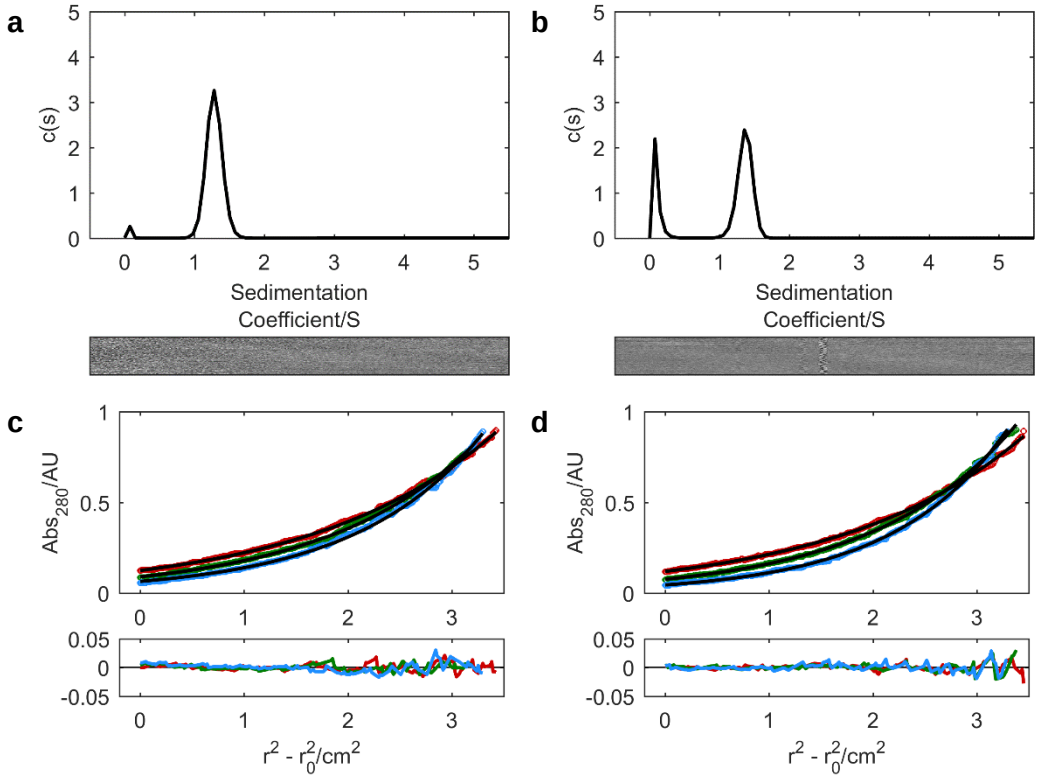


Figure 4-19 AUC data for the Ala@*g* heteromers. Sedimentation velocity c(s) distributions (top) and residuals (bottom) for: (a) 2-LIA-AB returning a Mw of 13.4 kDa (4.1 x mean monomer mass); (b) 2-LVA-AB returning a Mw of 12.9 kDa (4.1 x mean monomer mass). Sedimentation equilibrium data (red, green and blue circles) and fits (black lines) (top) and residuals (same colours) (bottom) for: (c) 2-LIA-AB returning a Mw of 12.8 kDa (4.0 x mean monomer mass); (d) 2-LVA-AB returning a Mw of 12.4 kDa (3.9 x mean monomer mass). All measurements were performed in PBS (pH 7.4).

4.7 Chapter conclusions

A set of acidic and basic peptides with a range of different core- and charged-residue configurations has been designed (Figure 4-20). The *in vitro* properties of the peptides differed in various ways. For example, the peptides 2-LI-A, 2-LI-B, 2-LIA-A and 2-LIA-B all form highly folded homomeric α-helical structures while other peptides, such as 1-LV-A and 1-LV-B, were almost entirely unfolded. The peptides also formed a number of different oligomeric states ranging from monomers to tetramers and perhaps higher.

The peptides could be combined in A/B pairs to make a set of A_2B_2 heterotetrameric assemblies, which also had differing properties. Primarily, the heterotetramers were found to differ in their thermal stabilities, which appeared to be largely dictated by the identities of the core residues: Leu/Ile core

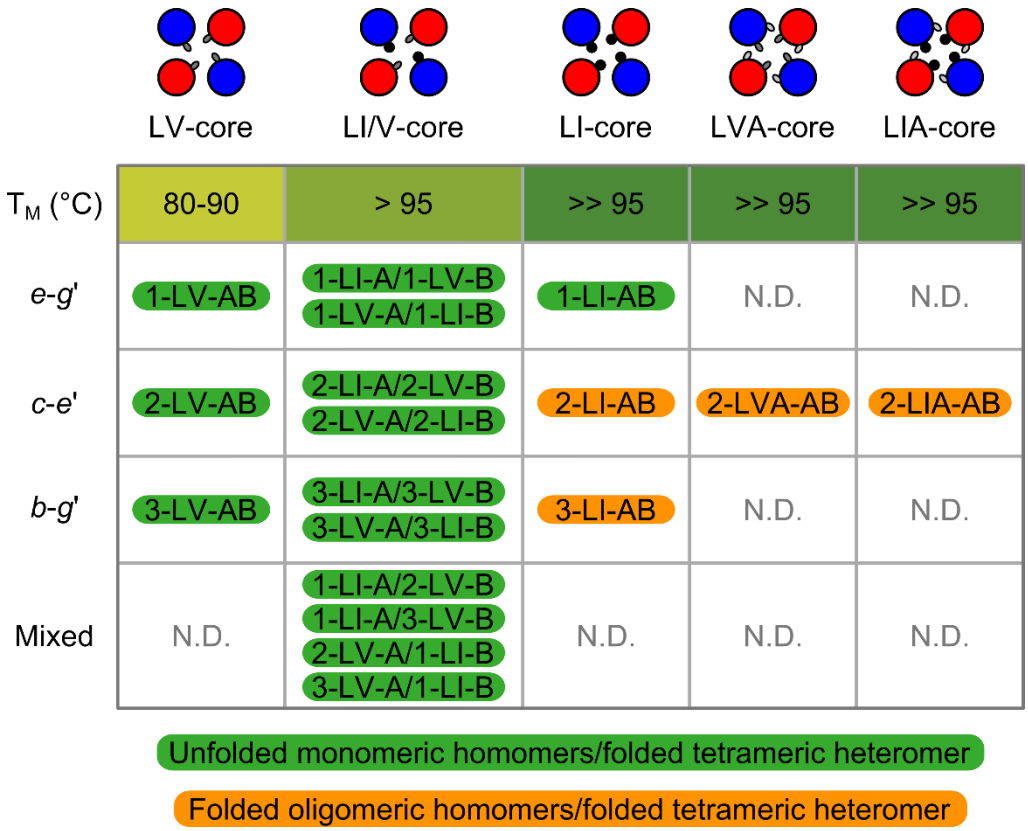


Figure 4-20 Summary of designs discussed in Chapter 4. Successful designs, *i.e.* those consisting of peptides that form unfolded monomers alone and interact to form heterotetrameric coiled coils, are highlighted in green. Designs where constituent peptides form folded homomeric species are highlighted in orange. Thermal stabilities for each core arrangement are indicated. LV-core designs were the least stable, mixed LI/V-core designs had intermediate stabilities, and LI-core designs and those with extended hydrophobic seams (Ala@*g*) were the most stable.

heterotetramers were the most stable; Leu/Val cores were generally the least stable; and mixed Leu/Ile/Val cores gave intermediate stabilities. All, or at least part, of a cooperative unfolding transition could be observed for the LV- and mixed-core heterotetramers. However, the LI-core heterotetramers were much more stable and did not display an unfolding transition. The absence of a cooperative unfolding transition can be an indication of a molten globule state, however, given that the other heterotetramers demonstrated cooperative unfolding, it seems likely that the LI-core heterotetramers also form unique structures, albeit incredibly thermally stable ones.

Having the ability to tune the stability of the heterotetrameric coiled coils through varying the core residues is useful as it allows coiled coils with specific properties to be designed for different applications. For example, in the context of transcriptional regulators, using a highly stable coiled coil as an interaction domain

to recruit RNAP to a gene of interest would drive strong expression of that gene; a coiled coil with a lower stability would drive a lower level of gene expression. Thus, the level of gene expression could be tuned by changing the coiled-coil domain. However, even the least stable heteromeric coiled coils discussed in this chapter are still hyperthermally stable when compared to most natural proteins from mesophilic organisms so the subtle differences in the coiled coil stabilities may be lost in a cellular context. Therefore, should a wider range of coiled-coil behaviours be required, the heterotetramers could be further destabilised by decreasing their length or by making further core substitutions to introduce even smaller hydrophobic residues such as alanine [174,191,225]. However, the latter approach has previously been found to lead to antiparallel conformations in some coiled coils and should therefore be approached with caution [189].

Heterotetramers consisting of peptides that are themselves unfolded are generally considered to be the most useful. This is because, should the peptides be used as artificial protein-protein interactions to mediate the interactions between other protein domains, they would not cause off target homo-oligomerisation of their fusion partners. In this regard, the most successful designs of the heterotetramers described here are 1-LI-AB, 1-LV-AB, 2-LV-AB and 3-LV-AB, as well as the hybrid heterotetramers that can be made by combining any of the unfolded peptides that comprise these coiled coils.

While the success of these designs must still be verified through solving their crystal structures, the biophysical data indicates that they do adopt the anticipated structures *in vitro*. It is therefore now possible to begin investigating how these heterotetramers, and their constituent peptides, behave inside living organisms as well as to begin introducing features into the coiled coils that may allow them to perform more-complex behaviours such as conformational switching.

Chapter 5: Characterisation of novel tetrameric coiled coils in *E. coli*

5.1 Chapter introduction

Following their design and characterisation *in vitro*, the *in vivo* behaviour of the *de novo* coiled coils was investigated to determine whether they were suitable to use as PIDs in ATFs. Challenges included whether the short peptides would be able to find their partners in the complex cellular environment; whether they would cross react with endogenous proteins, given the prevalence of coiled-coil sequences in nature; and whether the highly charged peptides would have toxic effects on the cell. A subset of A_2B_2 heterotetramers was chosen from the suite of designed tetramers discussed in this thesis. Heterotetramers were chosen over homotetramers due to their increased complexity: as they contain multiple components they can be used to make more-complex ATFs with more inputs available to increase the achievable level of control.

The peptides 1-LI-A, 2-LV-A, 3-LV-A, 1-LI-B, 2-LV-B and 3-LV-B were selected for further analysis in *E. coli* because these peptides were unfolded monomers in isolation but also interacted to form α-helical heterotetramers when mixed in acidic/basic pairs. Furthermore, while they were all highly thermally stable, the heterotetramers exhibited a range of T_M values. They also represented the three main charge configurations (*e*/*g*, *c*/*e* and *b*/*g*). For clarity and brevity these peptides are referred to as A1, A2, A3, B1, B2 and B3, respectively, throughout this chapter.

To determine whether the acidic and basic peptides were able to interact in *E. coli*, a transcription repression assay based on the Lac repressor was used (Figure 5-1). This assay has previously been used to demonstrate the *in vivo* interaction of a set of *de novo* heterodimeric coiled coils [173,174]. These heterodimers were designed to have different stabilities and their *in vitro*-measured K_Ds correlated with the level of transcription repression they achieved: those with the strongest interaction *in vitro* also gave the highest level of repression *in vivo*.

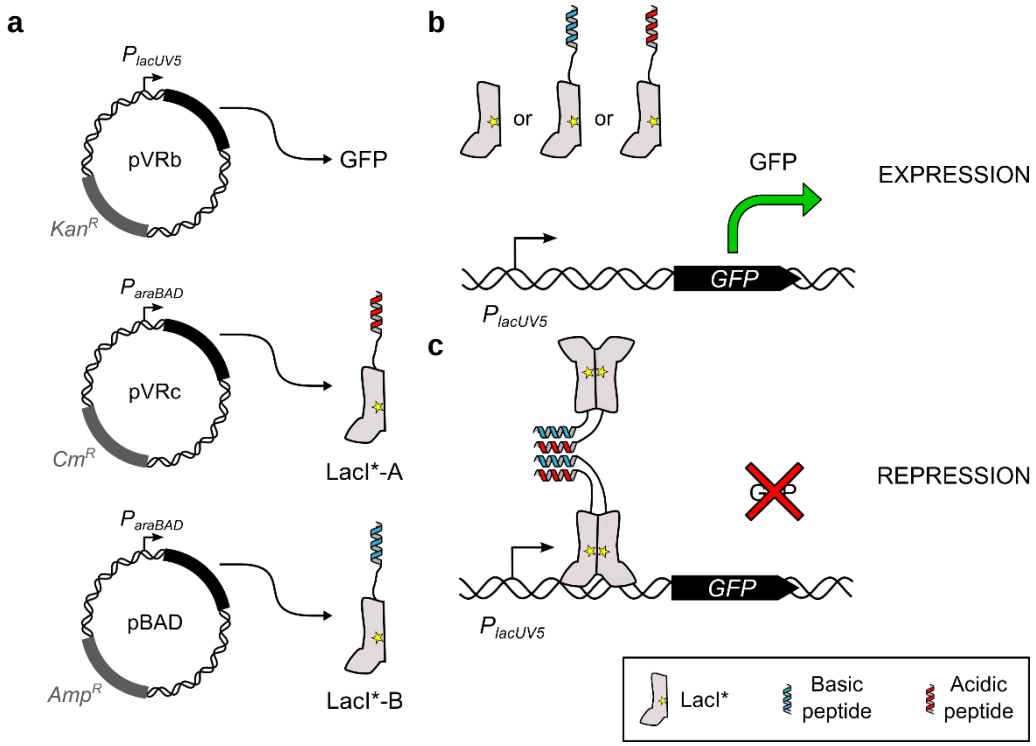


Figure 5-1 Experimental approach for the transcription repression assay. (a) The LacI*-peptide fusions and GFP reporter were expressed from separate plasmids with different antibiotic resistance genes. (b) When neither peptide (LacI*) or either the acidic (A) or basic (B) peptide (LacI*-A, LacI*-B) were present, LacI* should not form functional DNA-binding modules and so the reporter gene should not be repressed. (c) When both the acidic and basic peptide fusions are present, a LacI*-A/LacI*-B complex should form *via* the coiled coil leading to functional DNA-binding modules and repression of the reporter gene.

The Lac repressor, LacI, is a bacterial transcriptional repressor that is involved in controlling the expression of the *lac* operon genes, which are involved in lactose metabolism in the absence of the preferred carbon source, glucose [113]. Lac repressor monomers interact *via* a central dimerisation domain, bringing together the N-terminal helix-turn-helix DBDs to form a dimeric DNA-binding module that can bind to one of three operator sites, O_1, O_2 or O_3. These are spaced at different distances from the *lac* promoter and have different sequences, therefore different affinities for the Lac repressor [114,353]. Binding of LacI to the high-affinity O_1 operator, which overlaps the *lac* promoter, is essential for repression [354]. However, LacI in fact forms a dimer of dimers *via* a C-terminal tetramerisation domain [353]. Tetrameric LacI therefore contains two DNA-binding dimers and can bind two operators simultaneously, looping the intervening DNA [115,315]. This brings about a stronger level of repression by increasing the local concentration of the repressor at the promoter [116,355,356]. Transcription repression by the Lac repressor can be relieved

by inducers such as allolactose and IPTG, which cause a conformational change that prevents LacI from recognising operator DNA [117,118].

In the transcription repression assay, the acidic and basic coiled-coil peptides were fused to the C terminus of a version of the Lac repressor, replacing the wild-type (WT) tetramerisation domain. This version of Lac repressor, LacI*, also contained a single amino acid substitution (L251A) in the dimerisation domain that interrupts the dimerisation interface, weakening the interaction between the monomers [357]. As a result, the LacI* monomers were expected to form productive DNA-binding dimers only when brought together *via* some exogenous protein-protein interaction. If the interaction between the exogenous proteins (here, the coiled-coil peptides) was productive, LacI* dimers would bind to the O_1 *lac* operator site in a pVRb reporter plasmid causing repression of the reporter gene, GFP [308]. The reporter plasmid contained a single operator therefore looping was not expected to occur. Furthermore, the GFP reporter gene was expressed from a mutated version of the *lac* promoter, P_{lacUV5}, which contains mutations in the -10 box that increase expression from this promoter [358]. This promoter is also unaffected by CAP, which usually activates transcription from the WT *lac* promoter in the absence of glucose [106,359]. To prevent cross talk between the *de novo* constructs and the endogenous WT *lac* operon, all transcription repression assays were performed in TB28 *E. coli* cells, which do not contain the *lac* operon (*ΔlacIZYA*) [307].

The acidic and basic LacI*-peptide fusion proteins were expressed from different plasmids: either a pVRc plasmid, which contained a Cm^R gene, or a pBAD plasmid, which contained an Amp^R gene. While the components could have been expressed from the same plasmid, placing them on separate plasmids allowed greater flexibility for studying different combinations of the fusion proteins.

Both fusion proteins were expressed from the bacterial arabinose-inducible promoter, P_{araBAD} [360]. This promoter is part of the *araBAD* operon and is involved in arabinose metabolism. It is regulated by the AraC protein, which both represses and activates transcription [361]. In the absence of arabinose, an AraC dimer binds to two DNA sites, one proximal to the promoter ($araI_1$) and one distal to the promoter ($araO_2$), looping the intervening DNA [362,363]. This causes repression by preventing RNAP and CAP (which also activates transcription from this promoter) from accessing the DNA. On binding arabinose, the AraC dimer undergoes a conformational change that alters the orientation of its DBDs, causing it to bind at two adjacent, promoter-proximal DNA sites, $araI_1$ and $araI_2$ [363]. This opens up the

promoter allowing RNAP (and CAP, under low-glucose conditions) to bind. AraC may also actively recruit RNAP to the promoter when it is in the arabinose-bound state [364]. Arabinose can therefore be used to induce the expression of a gene that is expressed from the P_{araBAD} promoter. In these experiments, increasing the arabinose concentration should increase the expression level of the acidic and basic components.

While there are many other methods for studying the interactions of proteins *in vivo*, this approach has the benefit of demonstrating that the peptides interact while also generating ready to use semi-artificial transcription factors consisting of a designed protein-protein interaction and a natural DNA-binding module.

When studied in *E. coli* using this transcription repression-based interaction assay, all the peptides interacted in the anticipated heterotetramer pairs. Additionally, western blotting was used to investigate the effect of fusing *de novo* peptides to naturally occurring proteins. This revealed that the designed coiled-coil sequences generally led to a reduction in expression level. Therefore, while the *de novo* peptides still interacted as designed, they were not necessarily optimal for use in living organisms.

5.2 *De novo* coiled coils assemble *in vivo*

5.2.1 Acidic and basic peptides interact in E. coli

Initially, to determine whether they were able to interact in *E. coli*, both the A1 and B1 peptides were fused to LacI* (Lac repressor that contained the L251A substitution and was C-terminally truncated to remove the WT tetramerisation domain) *via* a Gly-Ser-Gly linker, generating the proteins LacI*-A1 and LacI*-B1 (Figure 5-2a, Figure 8-156). The peptide gene sequences were manually optimised for *E. coli* codon usage [365] and to reduce repetition to minimise the risk of recombination. The proteins were also fused to an N-terminal 6 His tag, a T7 epitope tag and an Xpress™ epitope tag. The acidic and basic proteins were placed on separate pVRc or pBAD plasmids, respectively, and were expressed from the arabinose-inducible promoter, P_{araBAD}. However, it is important to note that these experiments were performed without arabinose induction. While arabinose binding to AraC is required to achieve maximal gene expression from P_{araBAD}, a small amount of basal gene expression from this promoter is observed even in the absence of arabinose [360,366].

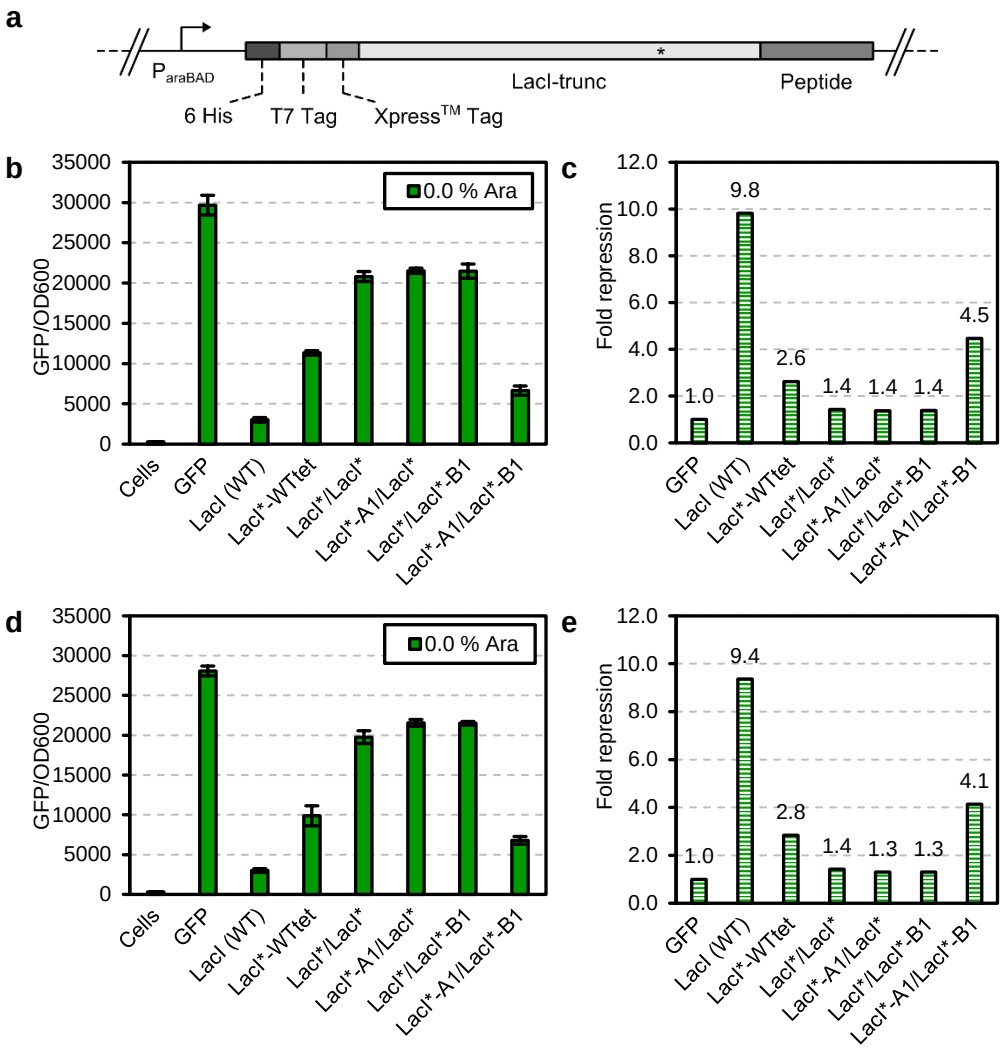


Figure 5-2 Transcription repression with the A1/B1 *de novo* coiled coil with induction by 0.0 % arabinose. (a) Peptides were fused C-terminally to LacI*, which was fused N-terminally to 6 His, T7 and Xpress tags. The construct was expressed from the arabinose-inducible P_{araBAD} promoter. The * indicates the approximate location of L251A substitution. DNA architectures were identical in both the pVRc and pBAD plasmids. (b) Transcription repression assay and (c) fold repression values relative to the GFP-only control where LacI*-A1 was expressed from pVRc and LacI*-B1 was expressed from pBAD. (d) Transcription repression assay and (e) fold repression values relative to the GFP-only control where LacI*-A1 was expressed from pBAD and LacI*-B1 was expressed from pVRc. LacI (WT), wild-type Lac repressor; LacI*-WTtet, Lac repressor with L251A substitution and wild-type tetramerisation domain; LacI*, Lac repressor with L251A substitution and no coiled-coil domain. Error bars are one s.d. from the mean, n=4.

When LacI*-A1 and LacI*-B1 were co-expressed, a decrease in the level of GFP was observed relative to the controls where just LacI* or just one of the LacI*-peptide fusions was expressed (Figure 5-2b). A 4.5-fold increase in repression relative to the GFP-only control was observed, indicating an interaction between

the acidic and basic peptides and the formation of a complex that repressed GFP expression (Figure 5-2c). The fact that the coiled-coil interaction could still be observed when the fusion proteins were expressed at basal levels, which were expected to be very low, demonstrates the strength of the interaction between the peptides. Furthermore, though the strength of repression achieved by the complex containing the *de novo* A1/B1 coiled coil was lower than that achieved by the completely WT Lac repressor, LacI (WT), it was greater than that achieved by LacI* that still contained the WT tetramerisation domain (LacI*-WTtet).

To investigate whether swapping the plasmids influenced the strength of repression, the transcription repression assay was also performed with the LacI*-A1 and LacI*-B1 proteins placed on the pBAD and pVRc plasmids, respectively (Figure 5-2d). The plasmids had different origins of replication resulting in a slightly higher copy number of pBAD (origin of replication: pBR322) relative to pVRc (origin of replication: p15A). This may have resulted in slightly higher expression of the protein encoded on the pBAD plasmid, which may have affected the results. However, the results were similar to those with the opposite plasmid arrangement with the complex giving 4.1-fold repression relative to the GFP-only control (Figure 5-2d and e). As the identity of the plasmid appeared to have little effect on the strength of repression, all of the remaining experiments were performed with the acidic peptide fusions expressed from the pVRc plasmid and basic peptide fusions expressed from the pBAD plasmid.

Furthermore, to determine whether other *de novo* coiled coils with lower *in vitro* thermal stabilities were also able to interact in *E. coli*, the transcription repression assay was performed with the A2/B2 and A3/B3 heterotetramers (Figure 5-3a and c). These coiled coils gave fold repression values of 4.6 and 2.6 relative to the GFP-only control indicating that these peptides also interacted in *E. coli* (Figure 5-3b and d). Additionally, when acidic/basic cross reactions were investigated using the A2/B1 and A1/B2 pairs, these heteromers also formed and gave fold repression values of 3.5 and 4.7, respectively (Figure 5-3e). A low level of repression was observed with the A3/B3 coiled coil, which had a relatively low thermal stability *in vitro*, and an intermediate level of repression was observed with A2/B1, which had an intermediate stability *in vitro*. However, there did not appear to be a robust correlation between the *in vitro* stability and the *in vivo* level of repression for the other pairs. This may have been due to other confounding factors

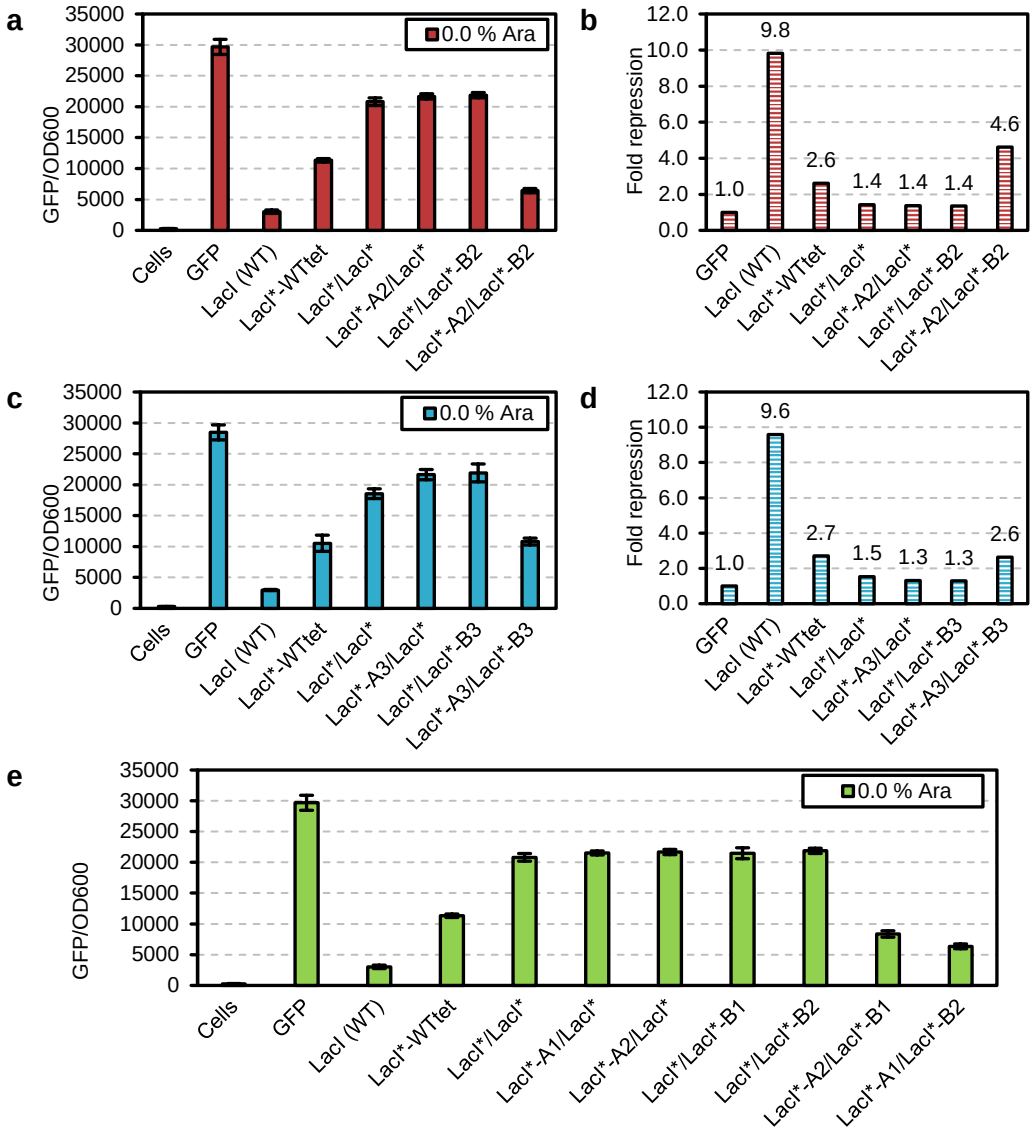


Figure 5-3 Transcription repression with the A2/B2, A3/B3 A2/B1 and A1/B2 *de novo* coiled coils with induction by 0.0 % arabinose. (a) Transcription repression assay and (b) fold repression values relative to the GFP-only control for LacI*-A2/LacI*-B2. (c) Transcription repression assay and (d) fold repression values relative to the GFP-only control for LacI*-A3/LacI*-B3. (e) Transcription repression assay with LacI*-A2/LacI*-B1 and LacI*-A1/LacI*-B2. Error bars are one s.d. from the mean, n=4.

that would have altered the observed amount of repression, such as variable cellular levels of the fusion proteins.

Both LacI*-A1/LacI*-B1 and LacI*-A2/LacI*-B2 appeared to repress GFP expression more strongly than LacI*-WTtet while LacI*-A3/LacI*-B3 showed similar levels of repression to LacI*-WTtet. The Lac repressor tetramerisation domain is itself a tetrameric coiled coil, albeit an antiparallel one [367]. The difference in the strength of repression with the natural and designed coiled coils may reflect the

longer length of the designed sequences – 30 amino acids *versus* 22 amino acids – which can contribute to the stability of coiled coils [225]. Additionally, *de novo* proteins are often hyperthermally stable relative to natural proteins [3,31,42,43,368]. In the context of coiled coils, this may reflect the fact that designed sequences are often idealised, primarily containing amino acids that favour α helix formation. Conversely, natural sequences have been subject to a multitude of selective pressures and their sequences are generally more varied in composition.

5.2.2 A second operator site enhances repression

In the above assays a single O_1 *lac* operator was present in the GFP reporter plasmids. However, as introduced above, the WT *lac* operon contains three separate operator sites, O_1, O_2 and O_3 [114]. While binding of Lac repressor to the O_1 operator, which is located near the promoter, is always necessary for repression, binding of a Lac repressor tetramer simultaneously to two operators has the effect of tethering Lac repressor at the DNA, preventing it from diffusing away and increasing its local concentration. This enhances the level of repression that can be achieved. A reporter plasmid containing a second O_1 operator site 92 bp upstream of the first O_1 site (the same distance between O_1 and O_3 in the *lac* operon) was used to measure whether the presence of a second operator led to an enhancement in the level of repression by the LacI*-A1/LacI*-B1 complex (Figure 5-4a). Because the A1/B1 coiled coil should form a tetramer, the resulting LacI*-A1/LacI*-B1 complex should also be tetrameric with two DNA-binding dimers, like the WT Lac repressor.

When investigated with induction by 0.0 % arabinose, with a single O_1 operator the LacI*-A1/LacI*-B1 complex gave 5.0-fold repression relative to the GFP-only control (Figure 5-4b). With two operator sites this value was 7.1, therefore the complex showed a 1.4-fold enhancement in repression on adding the second operator (Figure 5-4c). Conversely, the wild-type Lac repressor gave a much greater 6.5-fold enhancement. However, it is more meaningful to compare the LacI*-A1/LacI*-B1 complex to the LacI*-WTtet sample, given that both contain the L251A substitution so rely primarily on the tetramerisation domain for both dimerisation and tetramerisation. The LacI*-WTtet complex gave a 2.0-fold enhancement.

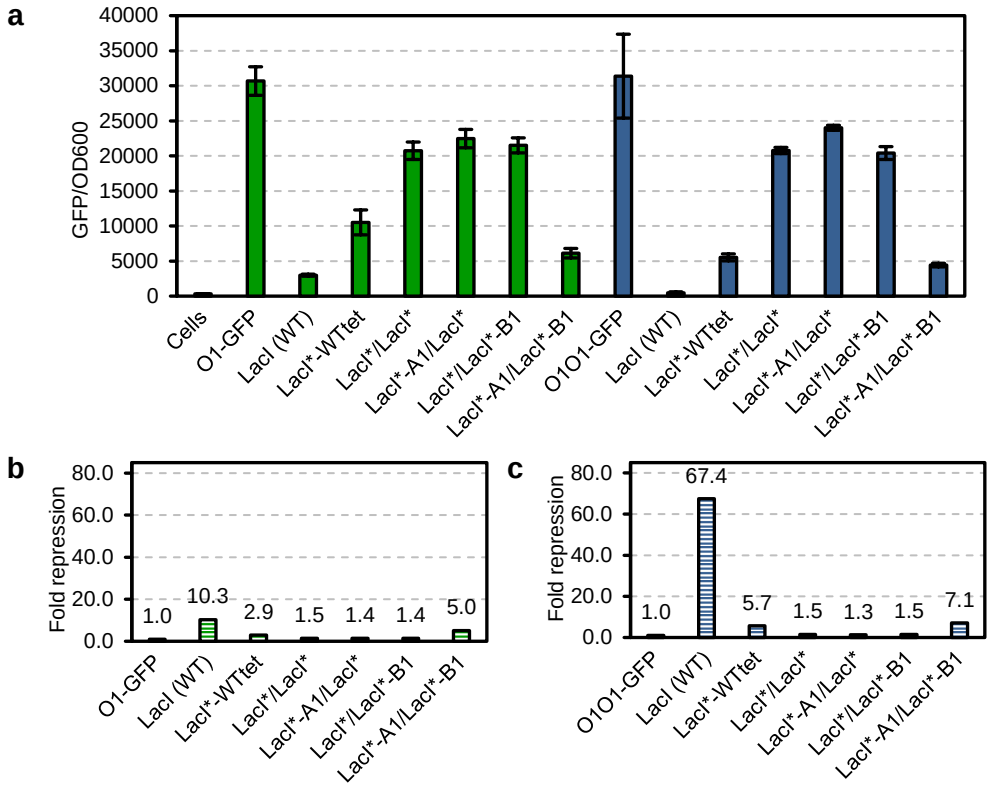


Figure 5-4 Transcription repression with the A1/B1 *de novo* coiled coil with one or two O_1 operator site with induction by 0.0 % arabinose. (a) Transcription repression assay (one O_1 operator, green; two O_1 operators, blue) and fold repression values relative to the GFP-only control with (b) one or (c) two O_1 operator sites. Error bars are one s.d. from the mean, n=4.

The LacI*-A1/LacI*-B1 complex achieved a higher overall level of repression than LacI*-WTtet. It also gave a small enhancement in the strength of repression in the presence of two *lac* operator sites, which was similar to that observed with LacI*-WTtet. However, the enhancement in repression observed for LacI*-A1/LacI*-B1 on adding the second O_1 site did not necessarily show that the complex was able to loop DNA: the parallel tetramer may not have held the dimeric DNA-binding modules in the correct conformation to bind two DNA sites simultaneously and the second operator may simply have maintained the complexes in the vicinity of the DNA for longer, leading to the observed increase in repression. Further investigation of the complex is needed to determine whether it can in fact loop DNA, for example by comparing it to a complex with just one DNA-binding dimer, such as a dimer of LacI* monomers. Alternately, *in vitro* techniques such as gel shift assays with DNA substrates of different lengths or electron microscopy could be employed [115,369].

5.2.3 Weakening the Lac repressor dimerisation interface improves repressor properties

In the above experiments, all of the acidic/basic pairings demonstrated GFP repression showing that they were able to interact in *E. coli* when fused to LacI*. However, while the LacI* system was sufficient to demonstrate the interaction between these peptides, it is not necessarily optimal for more extensive analysis or for use in ATFs. This is because the LacI*/LacI*, LacI*-A/LacI* and LacI*/LacI*-B controls all gave fold repression values of around 1.4, showing that residual dimerisation between the LacI* monomers was occurring (Figure 5-2c, Figure 5-3b and d). Therefore, while the L251A substitution did somewhat weaken the dimerisation interface in the Lac repressor, the interaction was not entirely abolished and the LacI* monomers were still able to dimerise, forming DNA-binding dimers that could repress GFP expression regardless of whether they were fused to the *de novo* peptides.

Furthermore, when the LacI*-A1/LacI*-B1 transcription repression assay was performed with induction by 0.2 % arabinose high levels of repression were observed both in the controls where neither or just one peptide was present, and when the acidic and basic peptide were present (Figure 5-5a). With induction, the LacI*-A1/LacI*-B1 complex repressed only around 1.3-fold more strongly than any of the controls (Figure 5-5b). However, without induction, this complex repressed at least 3-fold more strongly than the controls (Figure 5-2). Thus, at increasing protein concentrations the LacI* residual dimerisation increases and begins to overwhelm the contribution of the *de novo* coiled coil to oligomerisation.

To improve the properties of the repressors, further substitutions were made to the Lac repressor dimerisation interface. A tyrosine in this interface, Y282, was selected for mutagenesis as substitutions at this position often produce a monomeric phenotype [370-372]. Y282A and Y282D substitutions were introduced into LacI* resulting in the double L251A/Y282A and L251A/Y282D variants $LacI^{AA}$ and $LacI^{AD}$ (Figure 5-6). The Y282A substitution was proposed to weaken the interface by interrupting the shape complementarity due to the small size of alanine relative to tyrosine. The Y282D substitution was proposed to weaken the interface by introducing repulsive charges into the interface that would actively prevent the monomers from dimerising unless they were brought together by an interaction domain that was strong enough to overcome the repulsion.

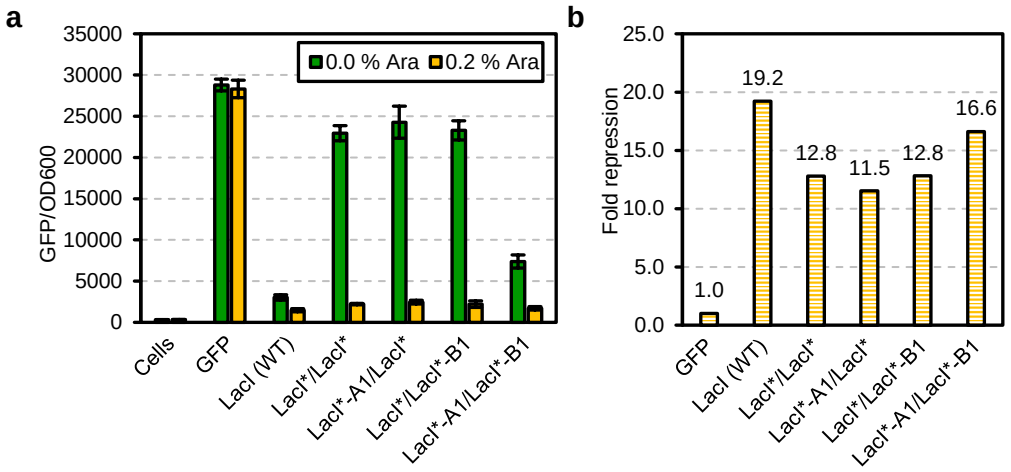


Figure 5-5 Transcription repression with the A1/B1 *de novo* coiled coil with induction by 0.0 and 0.2 % arabinose. (a) Transcription repression assay and (b) fold repression values relative to the GFP-only control with induction by 0.2 % arabinose. Error bars are one s.d. from the mean, n=4.

The LacI^{AA} and LacI^{AD} variants were tested with a simple model coiled-coil system before implementation with the heterotetramers. The variants were fused to the *de novo* homodimeric coiled coil, CC-Di to make the proteins LacI^{AA}-ccDi and LacI^{AD}-ccDi [208]. The Y282A and Y282D substitutions were also made in LacI*-WTtet to make the proteins LacI^{AA}-WTtet and LacI^{AD}-WTtet.

The new proteins were initially examined in the transcription repression assay with induction by 0.0 and 0.2 % arabinose (Figure 5-7a). Unlike LacI*, the LacI^{AA} and LacI^{AD} proteins did not display any residual dimerisation at 0.0 % arabinose and showed similar levels of GFP to the GFP-only control. Furthermore, the LacI^{AA}-WTtet and LacI^{AD}-WTtet proteins showed no repression at 0.0 % arabinose; at this arabinose concentration, LacI*-WTtet showed some repression activity. Finally, LacI^{AA}-ccDi repressed around 1.6-fold more strongly than Lac^{AA} alone and LacI^{AD}-ccDi did not show any repression at this arabinose concentration.

With induction by 0.2 % arabinose, LacI* displayed extensive residual dimerisation and had a fold repression value of 11.3. With LacI^{AA}, the residual dimerisation was roughly halved with a fold repression value of 6.2 at the same arabinose concentration. Therefore, though the additional Y282A substitution further weakened the dimerisation interface, it also did not entirely eliminate the residual dimerisation.

LacI^{AD} showed little residual dimerisation at 0.2 % arabinose. However, at the same arabinose concentration, LacI^{AD}-ccDi repressed only around 1.2-fold more strongly

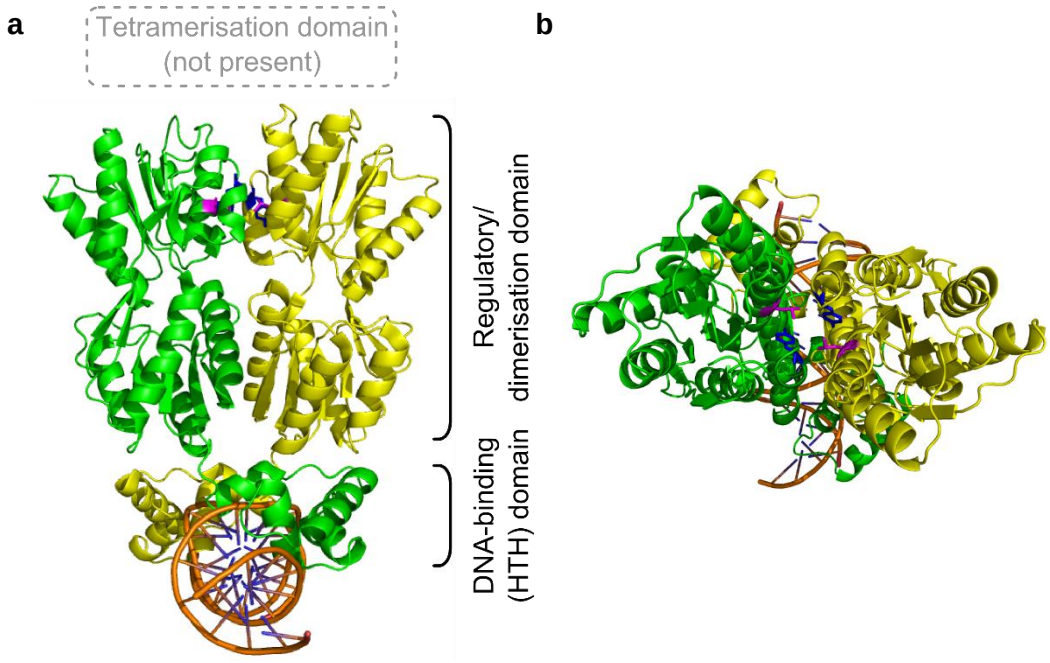


Figure 5-6 Crystal structure of a C-terminally truncated Lac repressor dimer (PDB ID: 1EFA [373]) showing the domain structure and the locations of L251 and Y282. The structure is presented from (a) the side, with DNA-binding and regulatory domains labelled, and (b) the top. The tetramerisation domain (not present in this structure) is located at the top of the structure in (a). LacI monomers, green and yellow; DNA, orange/purple; L251, magenta; Y282, blue. Image generated using PyMol.

than LacI^{AD} alone. The poor repression activity displayed by LacI^{AD}-ccDi may be due to the aspartates causing excessive disruption to the local conformation of the dimerisation interface, such that the parallel homodimeric coiled coil CC-Di could no-longer mediate the interaction between the monomers. Conversely, the WT tetramerisation domain, an antiparallel tetramer, may still have been able to mediate this interaction in LacI^{AD}-WTtet, which did demonstrate relatively strong repression at 0.2 % arabinose.

Western blotting was performed on a sample of *E. coli* cells taken during the assay using an antibody specific for the 6 His tag that is present on all of the Lac repressor derivatives. All of the proteins were expressed at relatively similar levels (Figure 5-7b). Therefore, the differences in repression were not due to different levels of expression or degradation of the proteins.

Due to poor repression by LacI^{AD}-ccDi, the LacI^{AA} proteins were selected for further analysis at 0.02, 0.002 and 0.0002 % arabinose (Figure 5-7c). With increasing arabinose concentrations, and therefore increasing levels of the *de novo*

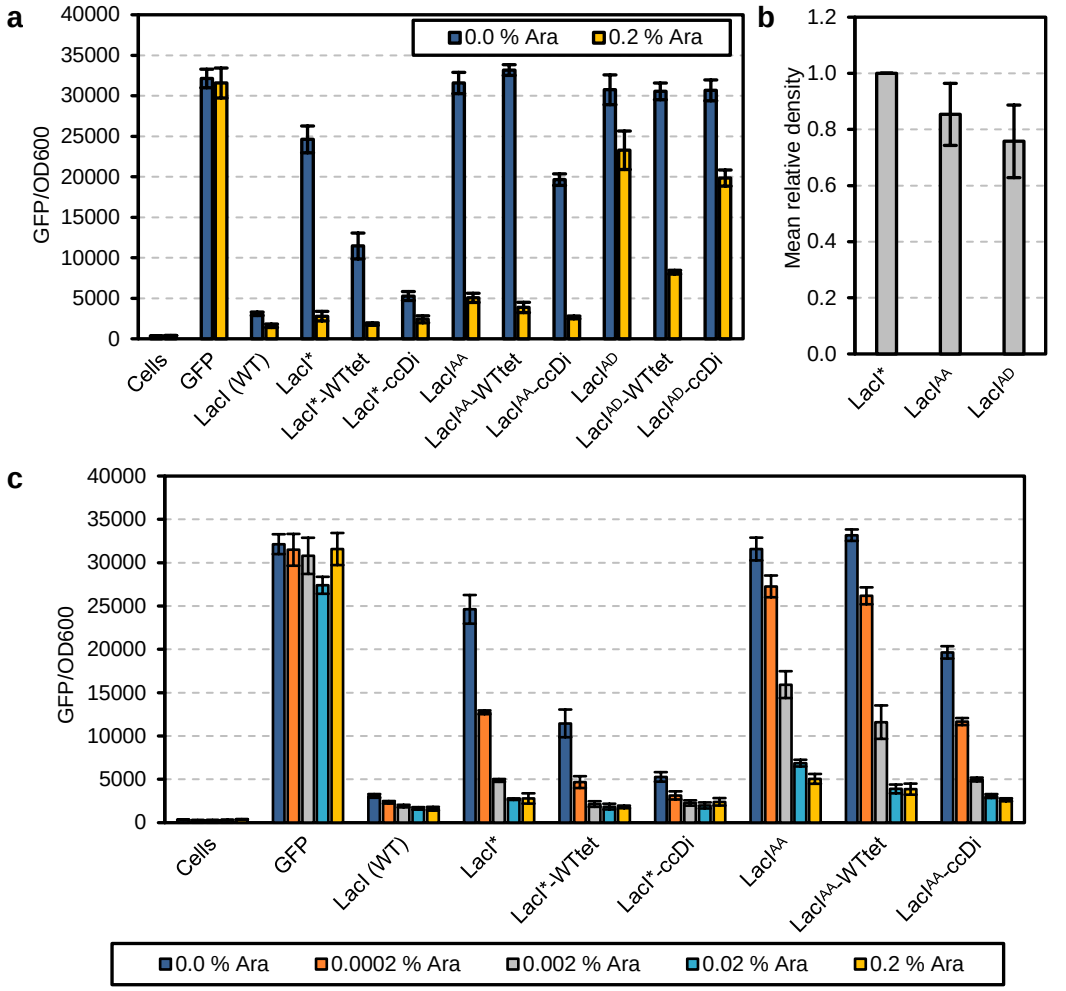


Figure 5-7 Transcription repression by the CC-Di *de novo* coiled coil fused to LacI* containing additional Y282 substitutions with induction by different arabinose concentrations. (a) Transcription repression assay with LacI*-ccDi, LacI^{AA}-ccDi and LacI^{AD}-ccDi with induction by 0.0 and 0.2 % arabinose, n=4. (b) Expression levels of LacI^{AA} and LacI^{AD} variants relative to LacI*, n=3, representative gel image in Figure 5-20a. (c) Transcription repression assay with LacI*-ccDi and LacI^{AA}-ccDi with induction by 0.0, 0.0002, 0.002, 0.02 and 0.2 % arabinose, n=4. Error bars are one s.d. from the mean.

transcription factors, the level of repression caused by LacI*-ccDi and LacI^{AA}-ccDi was also increased.

Above 0.002 % arabinose LacI*-ccDi consistently gave fold repression values of 13–14, potentially indicating that the maximal achievable level of repression for this protein had been reached (Table 5-1). The fold repression values for LacI^{AA}-ccDi, however, continued to increase up to 0.2 %. Furthermore, LacI*-ccDi induced with 0.2 % arabinose repressed GFP expression only 2.2-fold more strongly than un-induced LacI*-ccDi, whereas Lac^{AA}-ccDi induced with 0.2 % arabinose repressed 7.3-fold more strongly than un-induced LacI^{AA}-ccDi. Therefore, the LacI^{AA} system,

	0.0 % Ara	0.0002 % Ara	0.002 % Ara	0.02 % Ara	0.2 % Ara
GFP	1.0	1.0	1.0	1.0	1.0
LacI*	1.3	2.5	6.3	10.1	11.3
LacI*-WTtet	2.8	6.7	14.3	14.7	17.4
LacI*-ccDi	6.1	10.0	13.6	13.8	13.1
LacI^{AA}	1.0	1.2	1.9	4.0	6.2
LacI^{AA}-WTtet	1.0	1.2	2.7	7.0	8.2
LacI^{AA}-ccDi	1.6	2.7	6.2	8.9	12.0

Table 5-1 Fold repression values for LacI* proteins and LacI^{AA} proteins relative to the GFP-only control at different arabinose concentrations.

unlike the LacI* system, could essentially be “turned off” because in the absence of arabinose the level of repression was much lower than in the LacI* system. Furthermore, by altering the arabinose concentration the repression level of LacI^{AA}-ccDi could be tuned over a wider range than LacI*-ccDi.

To ensure that the LacI^{AA} system behaved similarly with other coiled coils, it was tested with the A1/B1 heterotetramer. Both the acidic and basic peptides were fused to LacI^{AA} and investigated with induction by 0.0 and 0.2 % arabinose (Figure 5-8). With 0.0 % arabinose the controls with neither or just one of the coiled-coil peptides showed no repression (Figure 5-8b, top). When both LacI^{AA}-A1 and LacI^{AA}-B1 were present there was a small decrease in GFP, showing that the LacI^{AA}-A1/LacI^{AA}-B1 complex was able to cause some repression even in the absence of arabinose.

When induced with 0.2 % arabinose, some residual dimerisation was observed in the controls however this was less extensive than that observed with LacI* (Figure 5-8b, bottom). Around half the amount of residual dimerisation was observed with the LacI^{AA} system compared with the LacI* system (Figure 5-5b). Also, there is a larger difference between the levels of repression observed with 0.0 and 0.2 % arabinose with LacI^{AA}-A1/LacI^{AA}-B1 compared to LacI*-A1/LacI*-B1. Therefore, LacI^{AA}-A1/LacI^{AA}-B1, like LacI^{AA}-ccDi can be switched off and the repression level can also be tuned more finely.

Overall, the LacI^{AA} system provides another tool for studying the interactions between exogenous proteins in *E. coli* and also makes LacI^{AA}-ccDi and LacI^{AA}-A1/LacI^{AA}-B1 more useful as semi-artificial transcription factors than the LacI* versions.

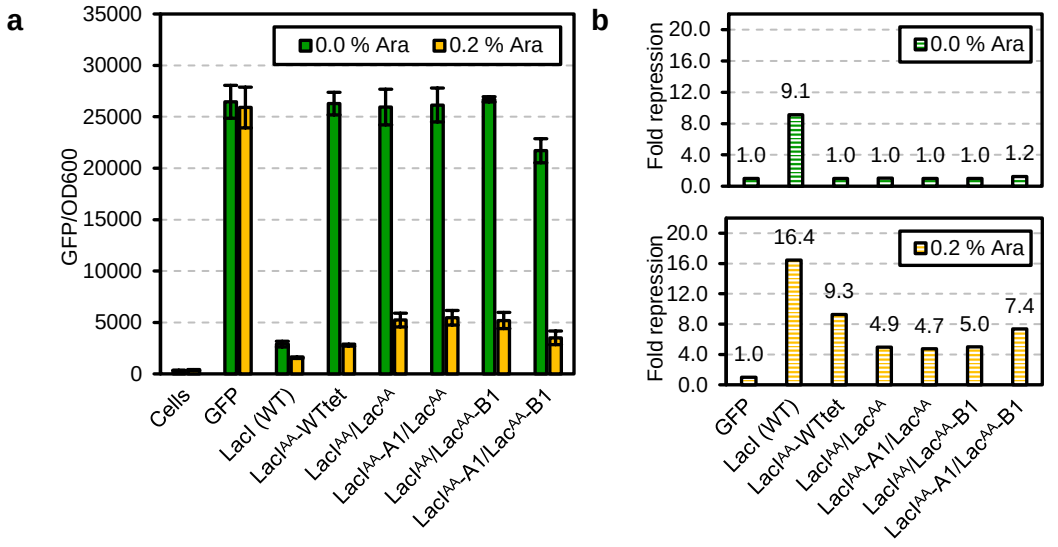


Figure 5-8 Transcription repression with the A1/B1 *de novo* coiled coil with LacI^{AA} with induction by 0.0 and 0.2 % arabinose. (a) Transcription repression assay and (b) fold repression values relative to the GFP-only control with 0.0% (top) 0.2 % (bottom) arabinose. Error bars are one s.d. from the mean, n=4.

5.3 Demonstrating tetramerisation

The systems introduced above were able to demonstrate that the *de novo* designed peptides interact in the anticipated pairings in *E. coli*. However, these versions of the assay do not distinguish between dimers and tetramers: either oligomeric state would have formed complexes that contained at least one dimeric DNA-binding module and so could have brought about GFP repression. Therefore, to obtain more information from the assay, the assay was modified such that functional DNA-binding dimers could only form when tetramerisation (or higher-order oligomerisation) occurred. To achieve this, the proteins fused to the basic peptides were changed to non-DNA-binding proteins. Initially a DNA-binding-deficient variant of Lac repressor was used, followed by a small solubility tag, SUMO. Finally, the effect of expressing the basic peptides alone, without folded protein domains fused to them, was investigated. In the resulting complexes, the DNA-binding dimers should only have formed between the LacI* monomers that were still fused to the acidic peptides. Because LacI^{AA} was generated after these experiments were performed, the LacI* variant is used throughout this section.

5.3.1 Lac repressor helix-turn-helix deletions

A variant of the Lac repressor that could not bind DNA was generated by removing the entire DNA-binding helix-turn-helix (HTH) domain (residues 14-60, inclusive,

Figure 5-6). This HTH deletion was made in the Lac repressor both with and without the L251A substitution to generate the proteins LacI*-ΔHTH and LacI-ΔHTH, respectively. Then, the basic peptide B1 was fused to the C termini in place of the WT tetramerisation domain to generate the proteins Lac*-ΔHTH-B1 and LacI-ΔHTH-B1 (Figure 5-9a). When co-expressed with LacI*-A1 these proteins were anticipated to form a complex that contained one Lac repressor dimer that could bind DNA and one that could not (Figure 5-9b).

The LacI-ΔHTH and LacI*-ΔHTH proteins were first investigated alone to ensure that DNA-binding had been eliminated (Figure 5-9c). Neither of the ΔHTH variants displayed any ability to repress GFP, even with induction by 0.2 % arabinose. LacI* showed strong repression with induction by 0.2 % arabinose due to the residual dimerisation discussed in Section 5.2.3. The LacI-Trunc protein (WT dimerisation domain, C-terminally truncated to remove the tetramerisation domain) showed strong repression at both 0.0 and 0.2 % arabinose. The dimerisation interface in this protein was intact and so DNA-binding dimers could form in the absence of an exogenous PID.

The LacI-ΔHTH-B1 and LacI*-ΔHTH-B1 proteins were then investigated in combination with LacI*-A1 (Figure 5-9d and e). With both complexes, repression was observed at 0.2 % arabinose, presumably due to residual dimerisation of the LacI*-A1 proteins. Neither the LacI*-A1/ LacI-ΔHTH-B1 or the LacI*-A1/ LacI*-ΔHTH-B1 complex demonstrated GFP repression at 0.0 % arabinose. This may have been an indication that the A1/B1 heterotetramer was in fact not forming a tetramer in *E. coli*. However, given the extensive *in vitro* data showing that this coiled coil is a heterotetramer in solution, this result may have been due to a limitation of the assay.

Specifically, it was initially assumed that two Lac repressor dimers would form, one between the monomers attached to the acidic peptides and one between the monomers attached to the basic peptides. However, it is also possible, and perhaps more likely, that the dimers could have formed between one monomer attached to an acidic peptide and one monomer attached to a basic peptide. In this assay, this would have resulted in hybrid Lac repressor dimers that contained one monomer that could bind DNA and one that could not, producing a dimer that overall could not bind DNA and could not repress GFP expression (Figure 5-9f). In other words, rather than the monomers that were opposite each other in the coiled coil interacting, the monomers that are adjacent to each other interacted. This

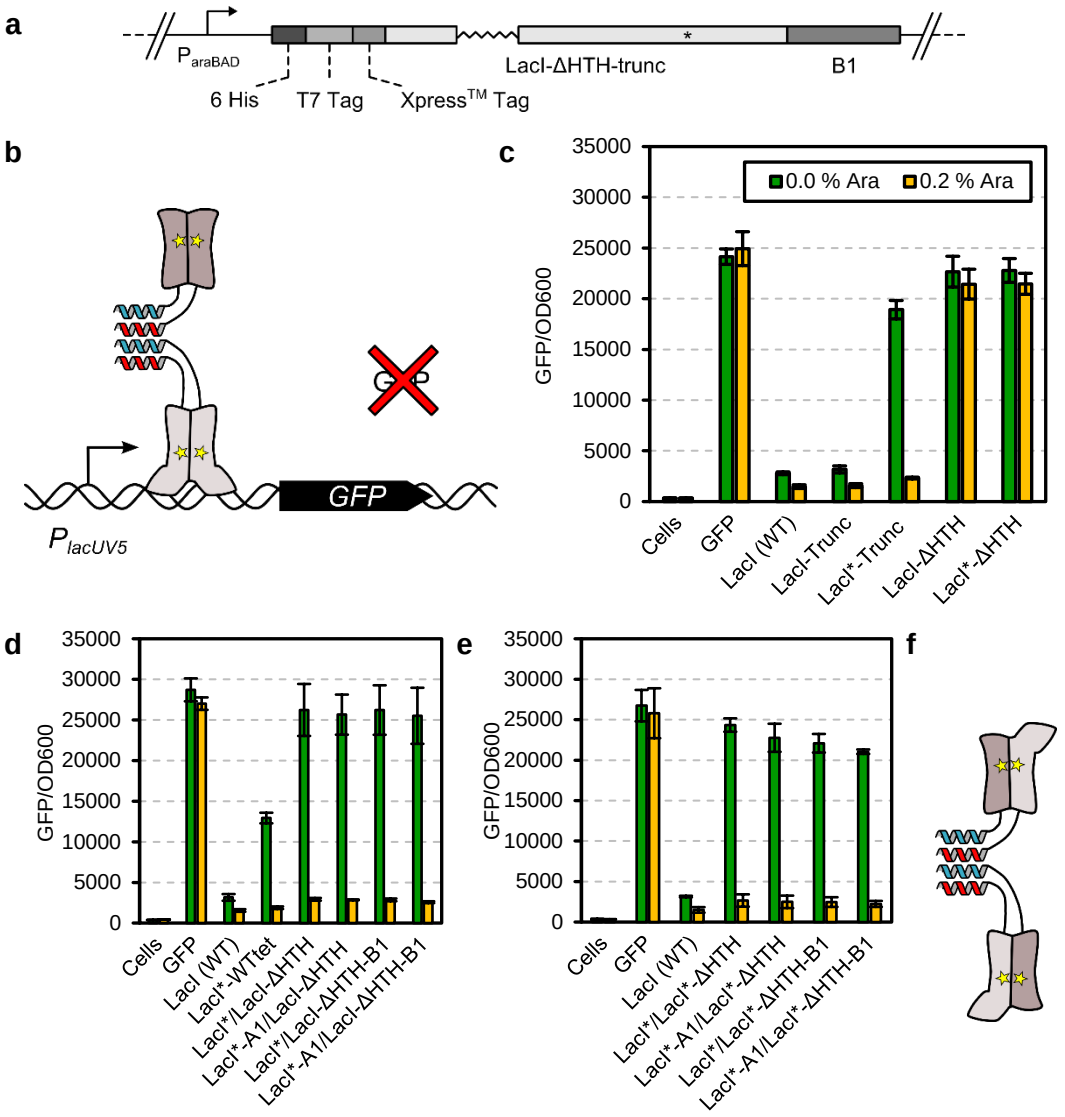


Figure 5-9 Transcription repression with ΔHTH variants of the Lac repressor with induction by 0.0 and 0.2 % arabinose. (a) B1 peptides were fused C-terminally to LacI-ΔHTH or LacI*-ΔHTH, which were fused N-terminally to 6 His, T7 and Xpress tags and the constructs were expressed from the P_{araBAD} promoter. The * indicates the approximate position of the L251A substitution. (b) Initially proposed model for DNA-binding of LacI*-A1/LacI-ΔHTH-B1 and LacI*-A1/LacI*-ΔHTH-B1 where light grey proteins represent Lac repressor with intact HTH domains and dark grey proteins represent ΔHTH variants. (c) Repression assay with LacI-ΔHTH and LacI*-ΔHTH proteins. (d) Repression assay with LacI*-A1/LacI-ΔHTH-B. (e) Repression assay with LacI*-A1/LacI*-ΔHTH-B. Keys as in (c). (f) Updated model for complex formation containing hybrid Lac repressor dimers. Error bars are one s.d. from the mean, n=4.

model for complex formation would prevent linker concatenation and is therefore the most probable arrangement.

5.3.2 SUMO fusions

Next, in order to produce a complex that should have contained only a single, dimeric DNA-binding module, the Lac repressor proteins were removed from the basic peptides entirely. Initially, the basic peptides B1, B2 and B3 were fused *via* a Gly-Ser-Gly linker to the C terminus of yeast SUMO (Smt3, small ubiquitin-related modifier), which is a small, monomeric protein often used as a solubility tag for protein expression [374]. SUMO was also fused to an N-terminal 6 His tag. This gave the proteins SUMO-B1, SUMO-B2 and SUMO-B3 (Figure 5-10a). When co-expressed with the corresponding LacI*-A proteins, the resulting complexes were anticipated to contain a single dimeric Lac repressor DNA-binding module, two separate, non-interacting SUMO domains and a linking heterotetrameric coiled coil (Figure 5-10b).

The LacI*-A1/SUMO-B1 complex was examined with induction by 0.0, 0.0002 and 0.002 % arabinose (Figure 5-10c). Low arabinose concentrations were used to limit residual dimerisation of the LacI*-A1 protein. With 0.0 % arabinose, the complex gave 2.0-fold repression relative to the GFP-only control (Figure 5-10d). This value increased as the arabinose was increased. The level of repression in the controls with neither or just one peptide also increased as the arabinose concentration was increased, presumably due to residual dimerisation of LacI* and LacI*-A1. However, at every inducer concentration there was a stronger level of repression with LacI*-A1/SUMO-B1 compared to the controls. The fold repression values of LacI*-A1/SUMO-B1 relative to the LacI*/SUMO (no-coiled coil) controls at each arabinose concentration were similar (Figure 5-10e). Therefore, even when there was residual dimerisation of the LacI* proteins, this did not overwhelm the signal and the A1/B1 interaction could still clearly be observed. This is in contrast to the LacI*-A1/LacI*-B1 system where the signal from the coiled coil was overwhelmed when expression of the proteins was induced with high arabinose concentrations (Figure 5-5). This may be because the arabinose concentrations used here are much lower.

The LacI*-A2/SUMO-B2 and LacI*-A3/SUMO-B3 complexes were also investigated with induction by 0.0 % arabinose (Figure 5-11a and b). These complexes gave fold repression values of 1.4 and 1.3, respectively, relative to the GFP-only control.

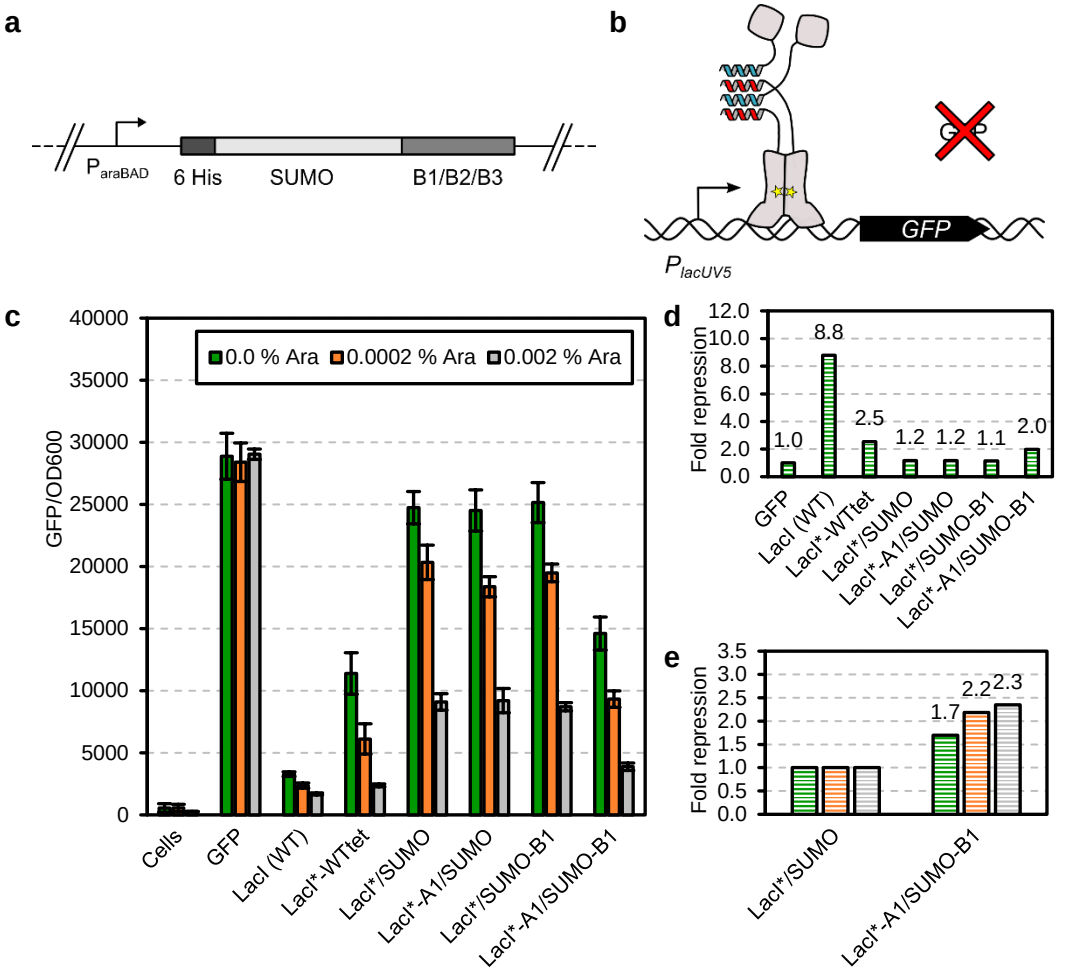


Figure 5-10 Transcription repression with the LacI*-A1/SUMO-B1 complex with induction by various arabinose concentrations. (a) Basic peptides were fused C-terminally to SUMO, which was fused N-terminally to 6 His, constructs were expressed from P_{araBAD}. (b) Proposed model for DNA-binding of LacI*-A/SUMO-B complexes where light grey squares represent SUMO. (c) Transcription repression assay with LacI*-A1/SUMO-B1 with induction by 0.0, 0.0002 and 0.002 % arabinose, (d) fold repression values relative to the GFP-only control with induction by 0.0 % arabinose and (e) fold repression values relative to LacI*/SUMO (no-coiled coil) control with induction by 0.0, 0.0002 and 0.002 % arabinose, key as in (c). Error bars are one s.d. from the mean, n=4.

The strengths of repression achieved by the LacI*-A/SUMO-B complexes were much lower than those achieved by the LacI*-A/LacI*-B complexes. There may have been many reasons for this. For example, the SUMO domains may have sterically hindered the formation of the coiled coil: the LacI* and SUMO proteins were both fused to the N termini of the peptides, so all these protein domains would have projected from the same end of the coiled coil, potentially forming a crowded environment that hinders the coiled-coil interaction. Introducing longer linkers between the coiled-coil peptides and their fusion partners may relieve this. This

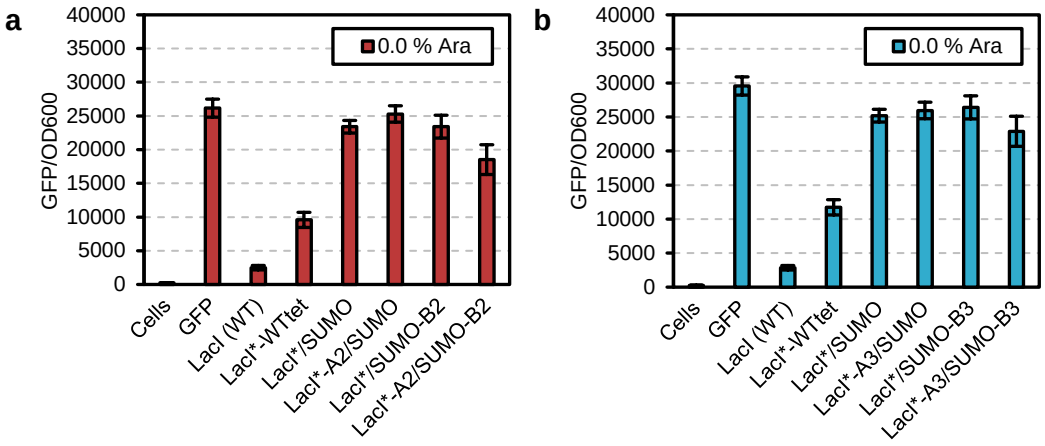


Figure 5-11 Transcription repression with the LacI*-A2/SUMO-B2 and LacI*-A3/SUMO-B3 complexes with induction by 0.0 % arabinose. Transcription repression assays with (a) LacI*-A2/SUMO-B2 and (b) LacI*-A3/SUMO-B3 with induction by 0.0 % arabinose. Error bars are one s.d. from the mean, n=4.

crowding could also have forced the coiled coils into a less-optimal antiparallel conformation where the strength of the interaction was weaker.

Alternatively, the SUMO-B proteins may have been present at a lower concentration than the LacI*-B proteins, either because they were expressed less strongly or degraded more rapidly, resulting in fewer LacI*-A/SUMO-B complexes available to repress GFP expression.

Finally, because only the acidic peptides were fused to the DNA-binding protein, the concentration of active DBDs would have been lower in the cell overall. Furthermore, because the LacI*-A/SUMO-B complexes contained only a single dimeric DNA-binding module, the local concentration of these dimers at the DNA would also have been lower. Therefore, on dissociation from the DNA, rebinding may have been less likely with the LacI*-A/SUMO-B complexes. Removing LacI* from the basic peptides would have roughly halved the total number of DBDs, which would be expected to also halve the observed amount of repression. However, less than half the amount of GFP repression was observed with the LacI*-A/SUMO-B complexes compared to the LacI*-A/LacI*-B complexes containing the same coiled coil. Therefore, it may have been a combination of factors that led to the observed decrease in repression.

5.3.3 Defining the minimal basic peptide

As well as fusing the *de novo* peptides to other protein domains in order to co-localise those proteins, at times it would also be useful to be able to express a single, un-tagged coiled-coil peptide. For example, a single peptide designed to interact with a second peptide could be used to disrupt a complex between the second peptide and a third peptide. This would trigger disassembly of the complex so that proteins attached to the second and third peptides are no-longer co-localised. Therefore, though SUMO appears to be a suitable fusion partner and is itself a very small protein that should not interfere with coiled-coil assembly, it would be useful to determine whether single *de novo* coiled-coil peptides can be expressed and are functional *in vivo*.

To investigate the effect of expressing minimal coiled-coil peptides, the basic peptides B1, B2 and B3 were fused to N-terminal 6 His, T7 and Xpress tags, to make the 6H-Tags-B proteins, or to only an N-terminal 6 His tag, to make the 6H-B proteins (Figure 5-12a and b). The T7 and Xpress tags were included initially for consistency with the previously described fusion proteins while the 6 His tag was included to enable future detection of the proteins by western blotting. When co-expressed with the LacI*-A proteins, these basic components were anticipated to take part in the formation of complexes containing a single dimeric DNA-binding module that is held together, or “pinned”, by the tetrameric coiled coil (Figure 5-12c).

When 6H-B1 and 6H-Tags-B1 were investigated with LacI*-A1 with induction by 0.0 % arabinose, the resulting complexes gave fold repression values of 1.6 and 1.5, respectively (Figure 5-12d). Because these values are similar and the T7 and Xpress tags appeared to do little to improve the strength of repression, the remaining experiments in this section used only the LacI*-A/6H-B system.

The LacI*-A1/6H-B1 complex was also examined with induction by 0.0002 and 0.002 % arabinose (Figure 5-13a). As with the LacI*-A1/SUMO-B1 complex, the strength of repression by LacI*-A1/6H-B1 also increased as the arabinose concentration was increased (Figure 5-13b). Furthermore, while the level of repression with the controls also increased as the arabinose concentration was increased, the level of repression with LacI*-A1/6H-B1 was higher than in the controls at every inducer concentration. Also, the fold repression values of LacI*-A1/6H-B1 relative to the LacI*/6H (no-coiled coil) controls at each arabinose

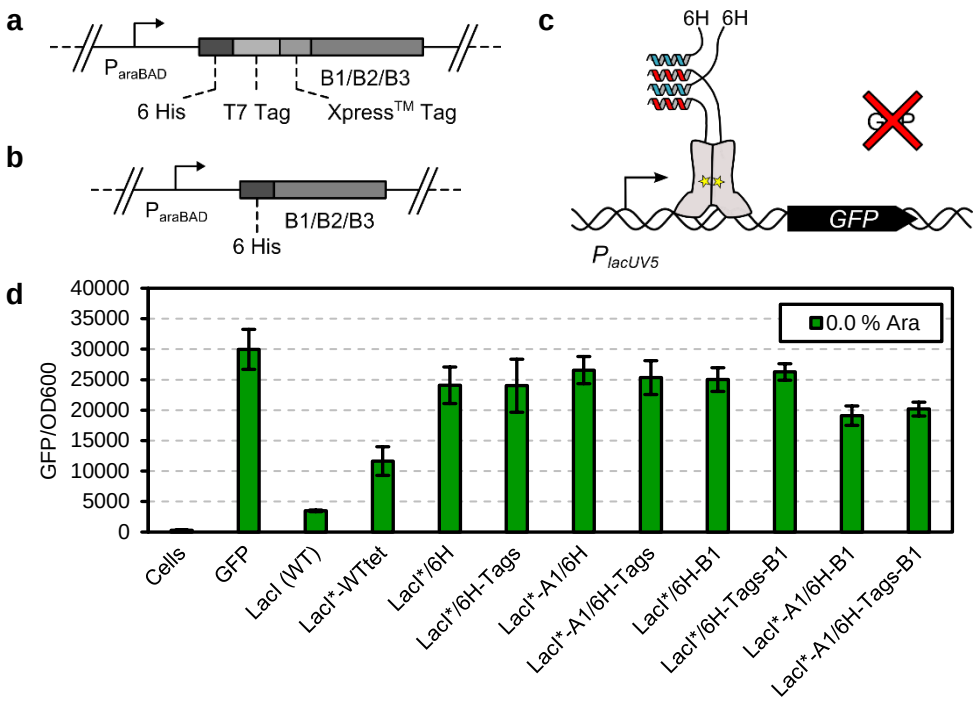


Figure 5-12 Transcription repression with LacI*-A1/6H-Tags-B1 and LacI*-A1/6H-B1 complexes with induction by 0.0 % arabinose. Basic peptides were fused C-terminally to (a) 6 His, T7 and Xpress tags or (b) to just a 6 His tag and the constructs were expressed from the P_{araBAD} promoter. (c) Proposed model for DNA-binding of LacI*-A/6H-Tags-B or LacI*-A/6H-B complexes. (d) Transcription repression assay with LacI*-A1/6H-Tags-B1 and LacI*-A1/6H-B1 with induction by 0.0 % arabinose. Error bars are one s.d. from the mean, n=4.

concentration were similar, again showing that the residual dimerisation of LacI* was not overwhelming the signal from the A1/B1 coiled-coil interaction at these arabinose concentrations (Figure 5-13c). Therefore, as with the LacI*-A1/SUMO-B1 complex, though the residual dimerisation of LacI*-A1 was at least partially responsible for the observed increase in repression as the inducer concentration was increased, some of the repression must also have been due to the formation of the heterotetrameric coiled coil.

When the LacI*-A2/6H-B2 and LacI*-A3/6H-B3 complexes were investigated with induction by 0.0 % arabinose, the fold repression values relative to the GFP-only controls were 1.2 and 1.1, respectively, which were similar to the values in the controls (Figure 5-13d and e). Therefore, the A2/B2 and A3/B3 heterotetramers did not appear to interact in the LacI*/6H system when studied at 0.0 % arabinose.

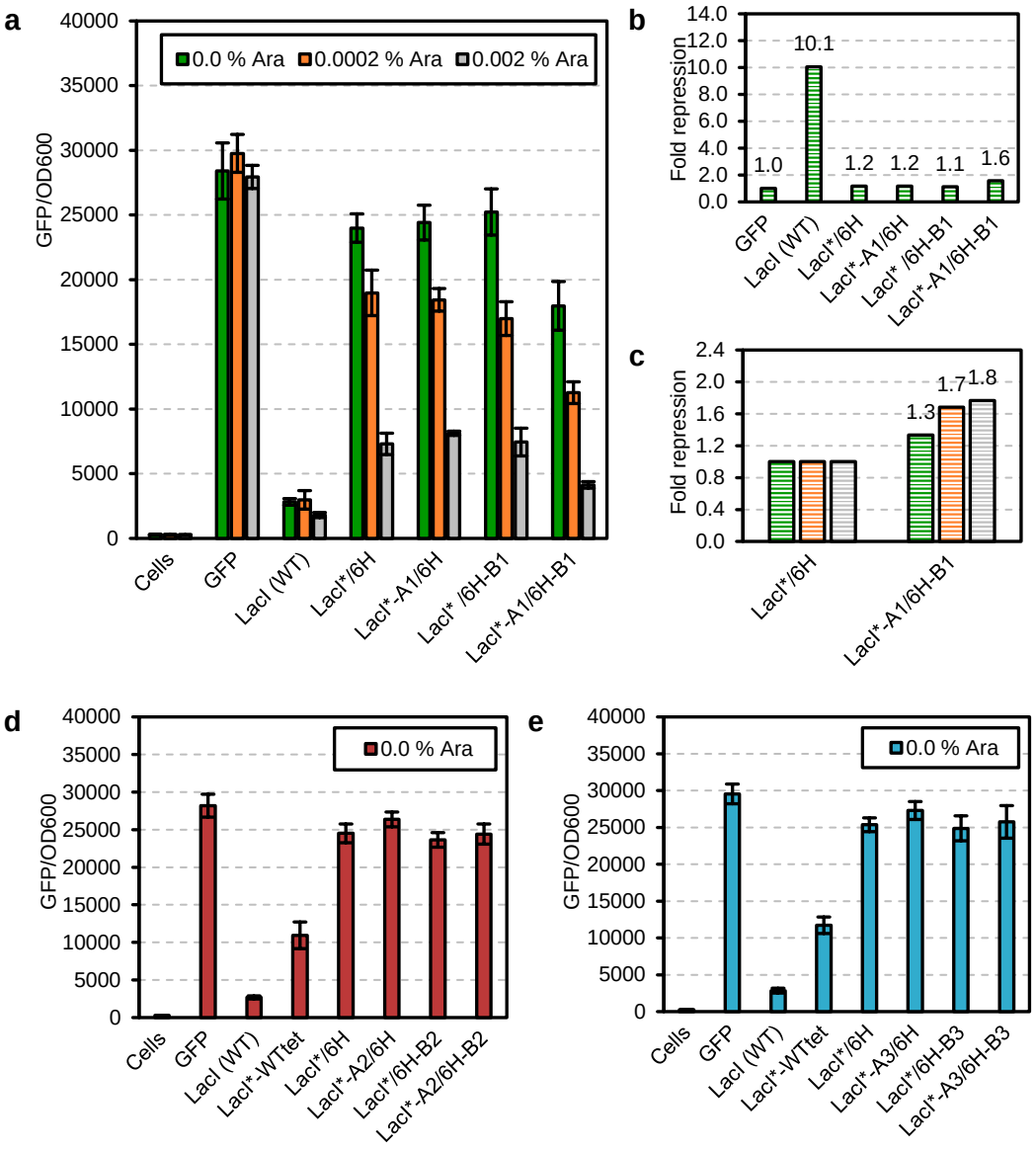


Figure 5-13 Transcription repression with A1/B1, A2/B2 and A3/B3 *de novo* coiled coils in LacI*-A/6H-Tags-B and LacI*-A/6H-B complexes with induction by different arabinose concentrations. (a) Transcription repression assay with LacI*-A1/6H-B1 with induction by 0.0, 0.0002 and 0.002 % arabinose, (b) fold repression values relative to GFP-only control with induction by 0.0 % arabinose and (c) fold repression values relative to LacI*/6H (no-coiled coil) control with induction by 0.0, 0.0002 and 0.002 % arabinose, key as in (a). Repression assays with (d) LacI*-A2/6H-B2 and (e) LacI*-A3/6H-B3 with induction by 0.0 % arabinose. Error bars are one s.d. from the mean, n=4.

Like the LacI*-A/SUMO-B complexes, these LacI*-A/6H complexes showed lower GFP-repression activity than the LacI*-A/LacI*-B complexes, which again could have been due to a combination of factors. In the LacI*-A/6H-B complexes, steric hinderance from the fusion partners of the basic peptides should not have been a problem but the loss of the second dimeric DNA-binding module from the system

may have reduced the level of repression by reducing both the local and global DNA-binding module concentration. Furthermore, protein degradation may have been a more significant issue for the 6H-B proteins than for the SUMO-B proteins. The 6H-B components are highly basic and, when not part of a heterotetrameric complex, highly unfolded. These proteins may have therefore been targets for rapid degradation in *E. coli*. Therefore, increasing the expression levels of these components may have increased the achievable level of repression.

5.4 Tuning the level of repression

In the experiments described above, the LacI*-A/SUMO-B and LacI*-A/6H-B complexes repressed GFP expression less strongly than the LacI*-A/LacI*-B complexes. One reason for this may have been decreased cellular concentrations of the basic component. Therefore, increasing the level of the 6H-B and SUMO-B components should lead to stronger repression. It would be useful for these complexes to be able to repress more strongly as it would provide a clearer indication that the acidic and basic peptides interact as designed in these complexes. A simple way to increase the protein concentration would be to increase the concentration of arabinose used to induce the cells. However, because both the acidic and basic components are expressed from the same P_{araBAD} promoter this would also increase the concentration of the LacI*-A components, complicating the matter with the residual dimerisation of LacI*. Therefore, if the control of the acidic and basic components was decoupled by expressing them from different promoters the concentrations of the SUMO-B or 6H-B components could be increased independently while the LacI*-A components are kept at a constant, low level where they should not exhibit residual dimerisation.

To this end, a library of constitutive bacterial promoters designed by Davis *et al* for synthetic biology applications was utilised [306]. This library of variable-strength promoters was generated by mutagenesis of the -10 and -35 boxes within an insulated promoter cassette that spans the region -105 to +55 from the transcription start site. The introduced insulator sequences make the promoters largely insensitive to changes in upstream and downstream sequences. A number of these promoters were selected for investigation.

5.4.1 Expressing LacI proteins from constitutive promoters*

A subset of promoters with low relative strengths was selected from the library and was initially tested with LacI*-ccDi to determine whether any of the promoters expressed this protein at a similar level to the un-induced P_{araBAD} promoter, or at least at a low level where residual dimerisation was minimal (Figure 5-14).

The promoters *pro1* and *pro5*, which Davis *et al* report to have relative strengths of 0.009 and 0.05, respectively, were introduced into pBAD plasmids containing the genes for the LacI* proteins, replacing the P_{araBAD} promoter and the gene for the AraC transcription factor. While these were the promoters with the weakest relative strengths available, they still appeared to be much stronger than un-induced P_{araBAD} and strong repression was observed when LacI*-ccDi was expressed from both promoters (Figure 5-14b). Fold repression values relative to the GFP-only control of 14.4 and 25.9 where observed when LacI*-ccDi was expressed from *pro1* and *pro5*, respectively, which were higher than the fold repression value achieved by LacI (WT) (Figure 5-14c). However, the LacI* controls also strongly repressed GFP expression when expressed from *pro1* and *pro5* indicating that residual dimerisation was occurring. The fold repression values relative to LacI* expressed from the same promoter for pro1-LacI*-ccDi and pro5-LacI*-ccDi were just 1.4 and 1.1 (Figure 5-14d). Therefore, as the level of expression was increased by changing the promoter, the residual dimerisation of LacI* began to overwhelm the signal such that the presence of the *de novo* coiled coil became redundant. The observation that the level of expression of LacI*-ccDi from *pro1* and *pro5*, measured as high levels of GFP repression, was much higher than from un-induced P_{araBAD} was perhaps not surprising given that the level of expression from P_{araBAD} in the absence of arabinose (*i.e.* when AraC is in the repressive state) was expected to be very low.

While the level of residual dimerisation was high when LacI*-ccDi was expressed from *pro1* or *pro5*, it was possible that a complex that was more highly degraded would be more suitable for use with the constitutive promoters as the high expression might be balanced by the high degradation (see Section 5.5.2). Therefore, the LacI*-A1/SUMO-B1 complex was selected because the LacI*-A1 protein was found to be present at lower cellular levels than LacI* and the other acidic components (Figure 5-19). The *pro1* promoter (the promoter with the weakest relative strength) was introduced to the pVRc plasmid containing the gene for LacI*-A1, replacing P_{araBAD} and *araC*. When LacI*-A1 and SUMO-B1 were

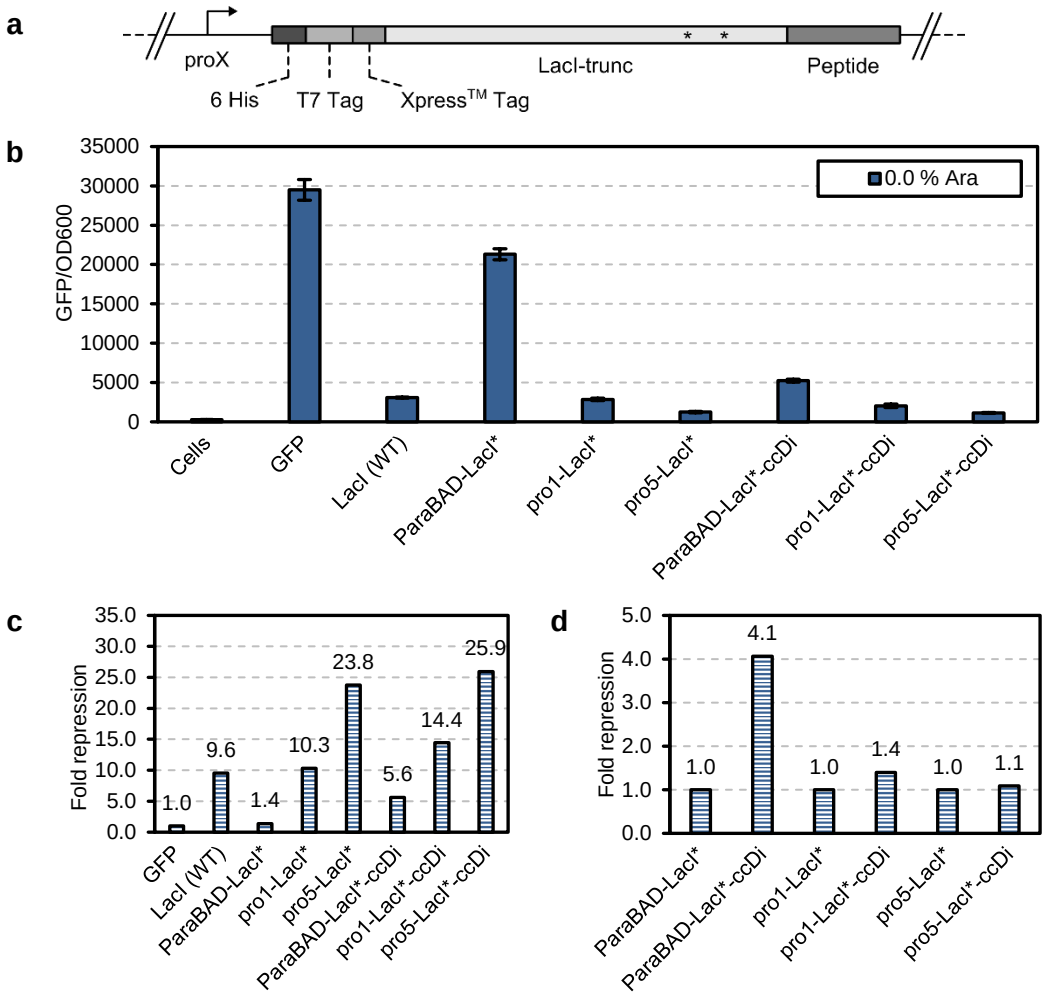


Figure 5-14 Transcription repression with LacI*-ccDi when expressed from different promoters with induction by 0.0 % arabinose. (a) CC-Di was fused C-terminally to LacI* (and LacI^{AA}), which was fused N-terminally to 6 His, T7 and Xpress tags and the construct was expressed from P_{araBAD}, *pro1* or *pro5*. The * indicate the approximate locations of L251A and Y282A substitutions. (b) Transcription repression assay with LacI*-ccDi. Fold repression values relative to (c) the GFP-only control and (d) LacI* (no-coiled coil control) for each promoter. Error bars are one s.d. from the mean, n=4.

co-expressed with induction by 0.0 and 0.2 % arabinose, high repression was observed (Figure 5-15a). However, residual dimerisation was also observed in the controls with pro1-LacI*/SUMO, pro1-LacI*-A1/SUMO and pro1-LacI*/SUMO-B1 and high fold repression values were observed, even with induction by 0.0 % arabinose (Figure 5-15b). Furthermore, the fold repression value of pro1-LacI*-A1/SUMO-B1 relative to LacI*/SUMO was only 1.4 (Figure 5-15c). This was similar to the value observed for LacI*-ccDi expressed from *pro1*.

Therefore, it seemed that residual dimerisation was unavoidable when the LacI* proteins were expressed from these constitutive promoters and it was unlikely that

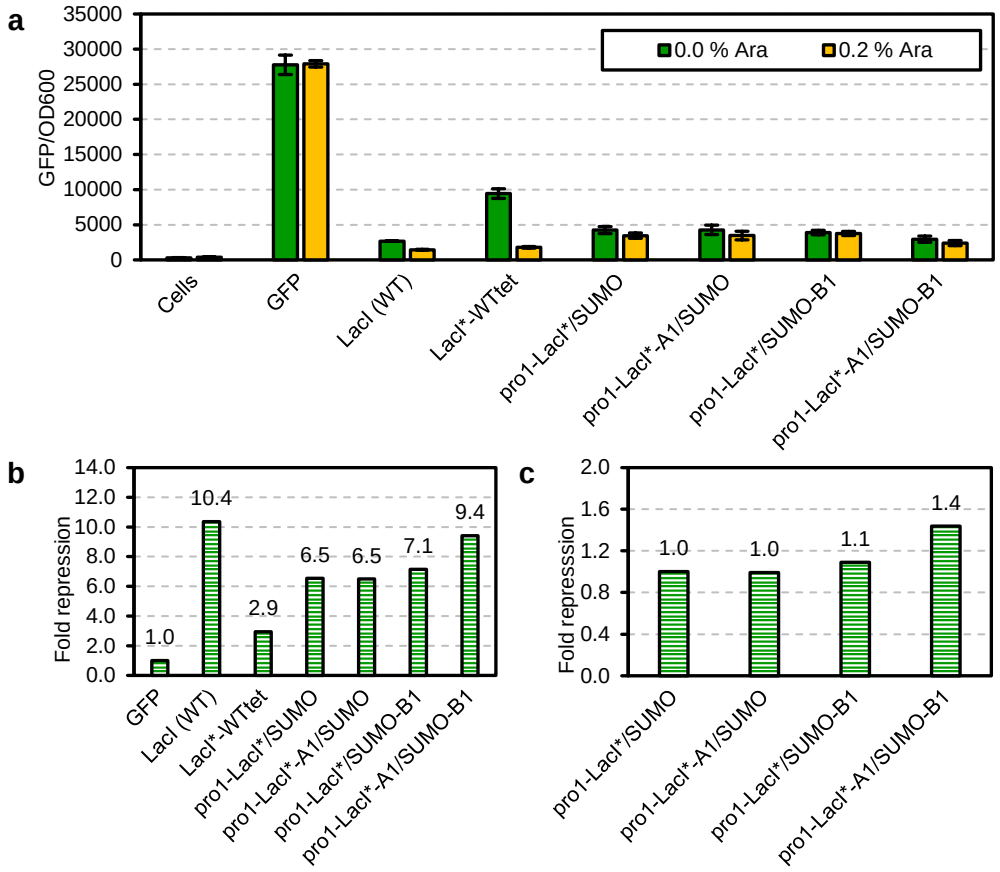


Figure 5-15 Transcription repression with LacI*-A1/SUMO-B1 when LacI* was expressed from *pro1* and SUMO was expressed from P_{araBAD} with induction by 0.0 and 0.2 % arabinose. (a) Transcription repression assay with LacI*-A1/SUMO-B1. Fold repression values relative to (b) the GFP-only control and (c) the LacI*/SUMO (no-coiled coil control) with induction by 0.0 % arabinose. Error bars are one s.d. from the mean, n=4.

weaker constitutive promoters would be available. Rather than attempting to alter the expression level further with other promoters or, for example, by introducing degradation tags to the proteins, the approach that was taken was to limit the residual dimerisation by using the LacI^{AA} variant introduced in Section 5.2.3.

5.4.2 Expressing LacI^{AA} proteins from constitutive promoters

To determine whether LacI^{AA}-A/SUMO-B and LacI^{AA}-A/6H-B complexes could form, the LacI^{AA}-A1 protein was initially tested with SUMO-B1 and 6H-B1 where all proteins were expressed from P_{araBAD} (Figure 5-16a and c). With induction by 0.0 % arabinose, both the LacI^{AA}-A1/SUMO-B1 and LacI^{AA}-A1/6H-B1 complexes showed little repression with fold repression values relative to the GFP-only control of 1.1 and 1.2, respectively (Figure 5-16b and d). As observed previously, with induction by 0.2 % arabinose the LacI^{AA} proteins displayed some residual dimerisation, albeit

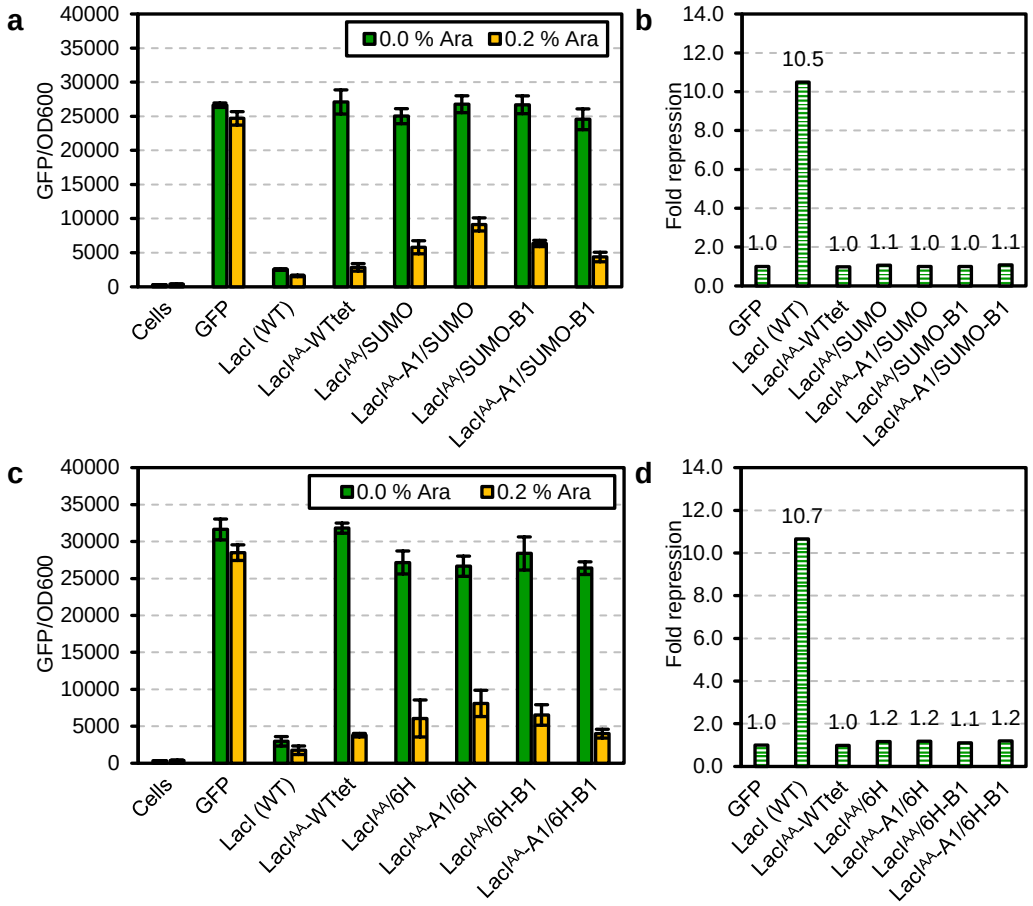


Figure 5-16 Transcription repression with LacI^{AA}-A1/SUMO-B1 and LacI^{AA}-A1/6H-B1 when all proteins were expressed from P_{araBAD} with induction by 0.0 and 0.2 % arabinose. (a) Transcription repression assay and (b) fold repression values relative to the GFP-only control with induction by 0.0 % arabinose with LacI^{AA}-A1/SUMO-B1. (c) Transcription repression assay and (d) fold repression values relative to the GFP-only control with induction by 0.0 % arabinose with LacI^{AA}-A1/6H-B1. Error bars are one s.d. from the mean, n=4.

less than the LacI* proteins under the same conditions. The level of repression increased 5.6-fold for LacI^{AA}-A1/SUMO-B1 and 6.6-fold for LacI^{AA}-A1/6H-B1 on inducing with 0.2 % arabinose compared to the non-induced state.

Consequently, the LacI^{AA}-A1/SUMO-B1 and LacI^{AA}-A1/6H-B1 complexes were investigated where LacI^{AA}-A1 was expressed from *pro1* and SUMO-B1 or 6H-B1 were expressed from P_{araBAD}. The desired outcome was for the LacI^{AA} proteins to undergo no, or at least limited, residual dimerisation when expressed from *pro1* and for the level of repression to increase as the arabinose concentration, and therefore the SUMO-B1 or 6H-B1 concentration, was increased.

When the pro1-LacI^{AA}-A1/SUMO-B1 complex was investigated, at all arabinose concentrations high levels of GFP were observed in the controls where neither or

just one coiled-coil peptide was present, indicating that residual dimerisation was not occurring when the double L251A/Y272A dimerisation variants were expressed from *pro1* (Figure 5-17a). The levels of the LacI^{AA} proteins in these samples was presumably constant. This could be confirmed in the future with western blotting. Furthermore, when both components of pro1-LacI^{AA}-A1/SUMO-B1 were present, the strength of repression increased as the arabinose concentration was increased up to 0.2 % (Figure 5-17b and c).

The fold repression relative to the LacI^{AA}/SUMO (no-coiled coil) also increased as the arabinose concentration was increased, and the values were similar to the fold repression values relative to the GFP-only control (Figure 5-17d). This further highlighted that little, if any, residual dimerisation of LacI^{AA} was occurring. Furthermore, because the level of repression increased as the arabinose concentration was increased, the complex is also a useful tool for controlling the expression of a gene of interest because the level of repression can be finely and reliably tuned *via* the inducer concentration.

The level of repression achieved by the pro1-LacI^{AA}-A1/SUMO-B1 complex with induction by 0.2 % arabinose in this experiment was comparable to the level of repression achieved by LacI*-A1/LacI*-B1 with induction by 0.0 % arabinose (Figure 5-2b and c). Therefore, relatively strong levels of repression can also be achieved by the LacI^{AA}-A1/SUMO-B1 complex, demonstrating unambiguously that this complex is capable of forming and repressing GFP expression in *E. coli* and that the A1/B1 coiled coil must therefore form a tetramer.

Furthermore, the fold increase in repression by pro1-LacI^{AA}-A1/SUMO-B1 on inducing with 0.2 % arabinose relative to the level of repression with 0.0 % arabinose was 2.8. While the corresponding values for LacI*-A1/LacI*-B1 and LacI^{AA}-A1/SUMO-B1 when all components were expressed from the P_{araBAD} promoter were 4.3 and 5.6, respectively, the 2.8-fold increase when LacI^{AA}-A1 was expressed from *pro1* could be attributed solely to the formation of the A1/B1 heterotetrameric coiled coil rather than to the aberrant dimerisation of LacI^{AA}.

The pro1-LacI^{AA}-A1/6H-B1 complex was also investigated with induction by 0.0 and 0.2 % arabinose (Figure 5-18a). Again, at both arabinose concentrations a high level of GFP was measured when neither or just one coiled-coil peptide was present showing that residual dimerisation was not occurring and that repression required the presence of both coiled-coil peptides. Also, the level of repression with

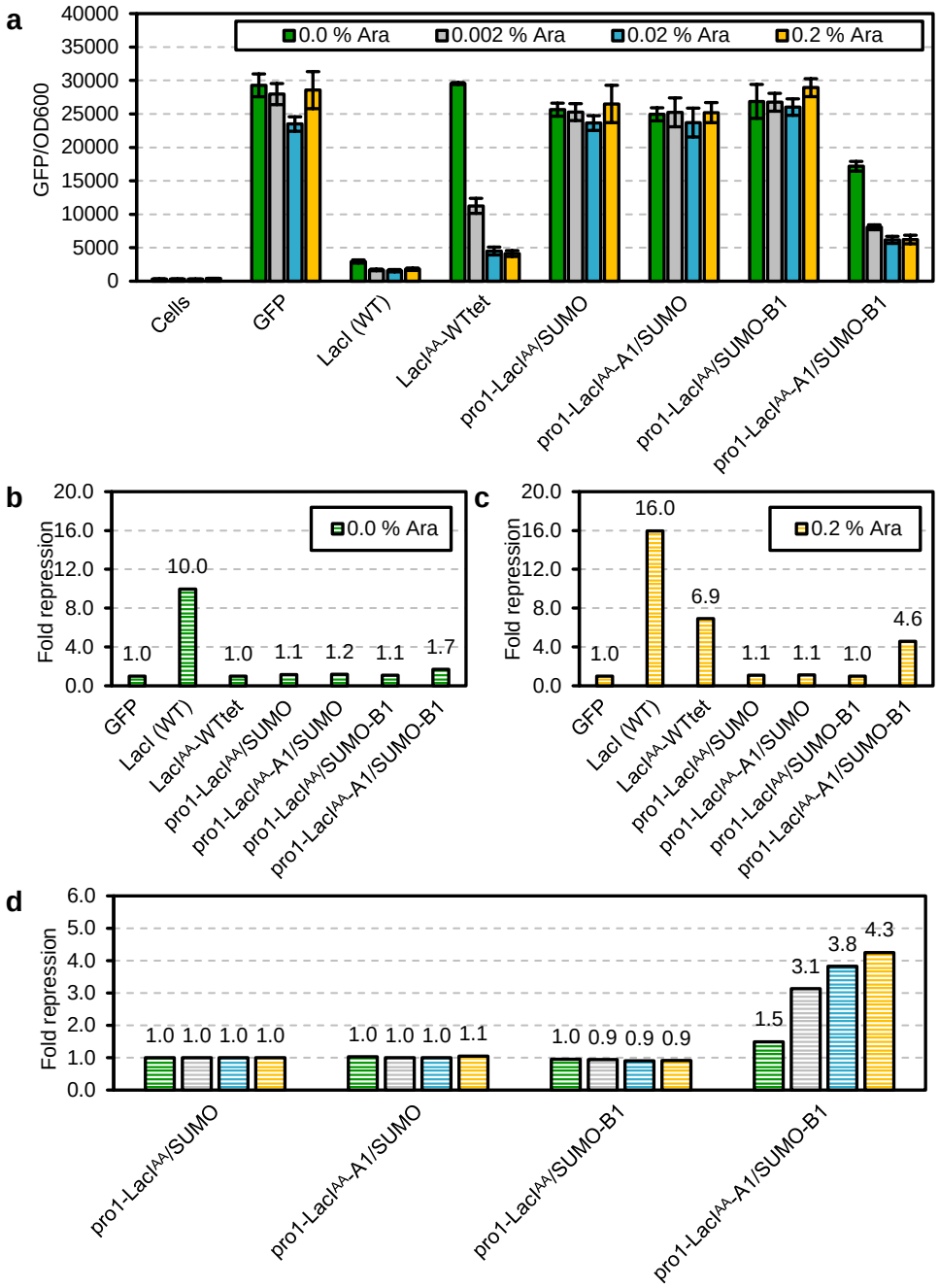


Figure 5-17 Transcription repression with LacI^{AA}-A1/SUMO-B1 where LacI^{AA}-A1 was expressed from *pro1* and SUMO-B1 was expressed from P_{araBAD} with induction by 0.0, 0.002, 0.02 and 0.2 % arabinose. (a) Transcription repression assay and fold repression values relative to the GFP-only control with induction by (b) 0.0 and (c) 0.2 % arabinose. (d) Fold repression values relative to the LacI^{AA}/SUMO control. Key as in (a). Error bars are one s.d. from the mean, n=4.

the pro1-LacI^{AA}-A1/6H-B1 complex increased as the arabinose concentration was increased so, again, the level of repression could be tuned (Figure 5-18b-d).

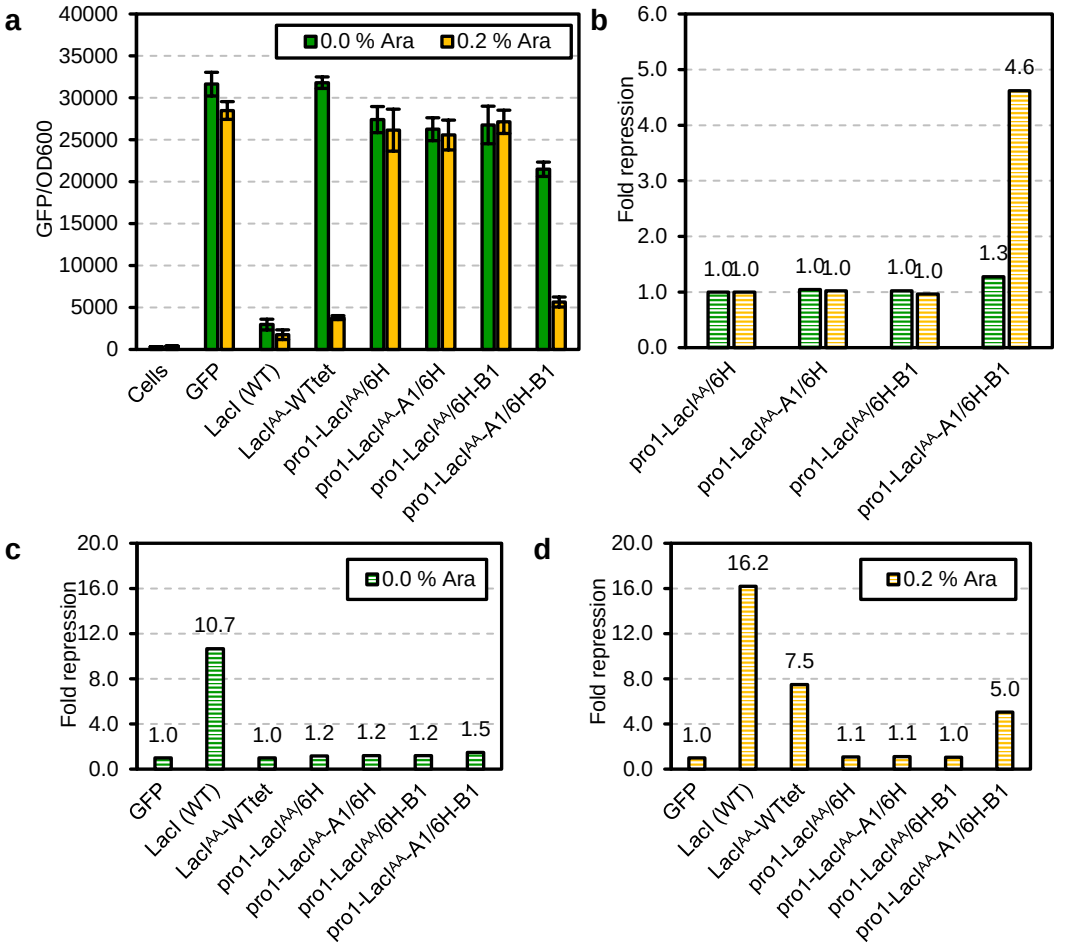


Figure 5-18 Transcription repression with LacI^{AA}-A1/6H-B1 where LacI^{AA}-A1 was expressed from *pro1* and 6H-B1 was expressed from P_{araBAD} with induction by 0.0 and 0.2 % arabinose. (a) Transcription repression assay. (b) Fold repression values relative to the LacI^{AA}/6H control with induction by 0.0 and 0.2 % arabinose, key as in (a). Fold repression values relative to the GFP-only control with induction by (c) 0.0 and (d) 0.2 % arabinose. Error bars are one s.d. from the mean, n=4.

Furthermore, the LacI^{AA}-A1/6H-B1 complex was capable of strong repression, providing a clear indication that the A1/B1 pair form a tetramer in *E. coli* and that minimal coiled-coil peptides can be expressed at useful levels and can be used as "pinning" parts to direct co-localisation of protein domains.

The levels of repression with induction by 0.2 % arabinose achieved by the LacI^{AA}-A1/SUMO-B1 and LacI^{AA}-A1/6H-B1 complexes were similar to each other (4.6 and 5.0 relative to the GFP-only control, respectively). Furthermore, as the arabinose concentration was increased, the strength of repression achieved by the LacI^{AA}-A1/SUMO-B1 complex appeared to plateau and the level of repression may have been approaching a maximum achievable value. This value may have been defined by the level of expression of LacI^{AA}-A1 from *pro1* whereby the level of

repression by both complexes at some point becomes limited by the availability of LacI^{AA}-A1 and therefore by the availability of DNA-binding modules.

It would be interesting to now test A2/B2 and A3/B3 with the updated assays as these pairs showed only weak repression with the LacI*-A/SUMO-B assay where all components were expressed from the P_{araBAD} promoter (Figure 5-11) and no repression with the original LacI*-A/6H-B assay (Figure 5-13d and e).

Finally, an alternative approach to decoupling the expression of the two components of the heteromeric complex was to change the promoter from which the basic component was expressed [375]. When SUMO-B1 was expressed from *pro1*, it appeared to be expressed at a slightly higher (though still useful) level compared to when it was expressed from un-induced P_{araBAD}, measured as an increase in the level of repression relative to the GFP-only control (data not shown). Therefore, the *pro1* promoter may be a useful tool in the construction of rudimentary genetic circuits using these semi-artificial transcription factors.

5.5 Expression levels of *de novo* coiled coils

The *de novo* coiled coil-based transcription factors developed in this chapter are able to repress transcription and therefore they must themselves be expressed *in vivo*. However, it remains to be seen how well they are expressed and how rapidly they are degraded. For example, the *de novo* peptides discussed here are generally highly charged and highly unfolded and may therefore be targets for degradation *in vivo*. Alternatively, they may be poorly expressed. Both would lead to a decrease in the level of repression brought about by the components because they would be unable to reach high cellular concentrations.

Western blotting was performed to determine whether appending *de novo* peptide sequences to Lac repressor and SUMO had an effect on their concentration in cells. Blotting was performed using an antibody specific for the 6 His tag that was present at the N terminus of all of the proteins. The cells used for blotting were either harvested following growth for a transcription repression assay (and were therefore co-transformed with multiple plasmids) or were harvested after growth under the same conditions used in transcription repression assays (which were transformed with a single plasmid). The latter was performed, for example, where proteins of interest had the same mass and therefore would not have been resolved by gel electrophoresis. While assays were often performed without

induction by arabinose, blotting was generally performed on cells that had been induced with arabinose as this aided protein detection.

While the experiments that follow offer some initial information about the cellular levels of the proteins, it is important to note that they were performed without a loading control and therefore further investigation is needed. However, the preliminary results suggest that such further investigation would indeed be fruitful.

5.5.1 De novo *coiled-coil sequences change the expression level of proteins fused to them*

The LacI*-peptide fusions were investigated with induction by 0.02 % arabinose to determine the effect of appending the *de novo* heterotetramer peptides to this protein. All *de novo* heterotetramer peptides appeared to lead to a decrease in the level of LacI* (Figure 5-19). The proteins were expressed from the same plasmids as in the transcription repression assays, *i.e*. acidic components were expressed from the pVRc plasmid and basic components, LacI (WT) and LacI*-WTtet were expressed from the pBAD plasmid. When LacI* was expressed from the pVRc plasmid, it was present at a lower level than when it was expressed from the pBAD plasmid, possibly reflecting the different copy numbers of the plasmids (Figure 5-19a).

The acidic peptides A1, A2 and A3 all led to a decrease in the level of LacI* when fused to the C terminus, with A1 producing the largest decrease (Figure 5-19b). The basic peptides B1, B2 and B3 also all led to a small decrease in the level of LacI* when fused to the C terminus and all had similar mean relative densities (Figure 5-19c).

While the proteins were quantified at 0.02 % arabinose (Figure 5-19d), the proteins were also detectable at 0.0 % arabinose when very long exposures were used (Figure 5-19e). With increasing exposure times, non-specific bands, which were consistent across all samples, could also be observed in the blots between the 35 and 40 kDa markers.

The LacI*-ccDi, LacI^{AA}-ccDi and LacI^{AD}-ccDi samples were blotted following cell growth for a repression assay with induction by 0.2 % arabinose (Figure 5-20a). Here, all proteins were expressed from the pBAD plasmid. LacI* was at a higher level than LacI (WT) but its level decreased progressively as it was fused to the WT lac tetramerisation domain or CC-Di (Figure 5-20b and c).

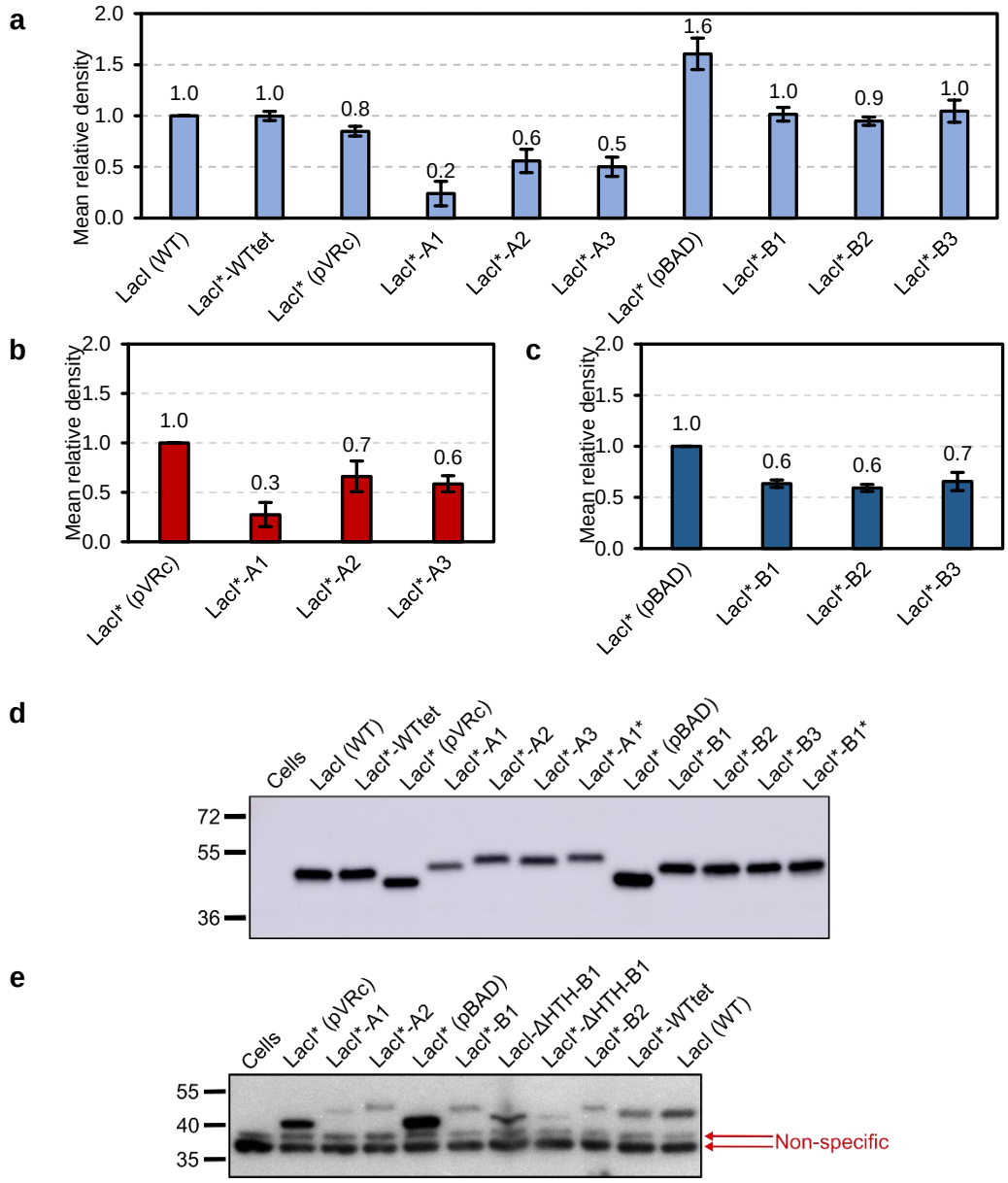


Figure 5-19 Western blotting analysis of LacI*-peptide fusion proteins. (a) Mean relative densities of detected bands relative to LacI (WT), measured following induction by 0.02 % arabinose, n=3. (b) Mean relative densities of LacI* fused to acidic peptides A1, A2 and A3, relative to LacI* where all proteins were expressed from pVRc, n=3. (c) Mean relative densities of LacI* fused to basic peptides B1, B2 and B3, relative to LacI* where all proteins were expressed from pBAD, n=3. Representative membranes for samples induced with (d) 0.02 and (e) 0.0 % arabinose. Non-specific bands are indicated. Error bars are one s.d. from the mean.

The LacI^{AA} and LacI^{AD} variants were present at lower levels than both LacI (WT) and LacI* (Figure 5-20b) and the levels of these proteins also decreased when the coiled-coil sequences were fused to their C termini. When compared to their respective no-coiled coil control (LacI*, LacI^{AA} or LacI^{AD}), both the Lac repressor tetramerisation domain and CC-Di appeared to be similarly destabilising for all

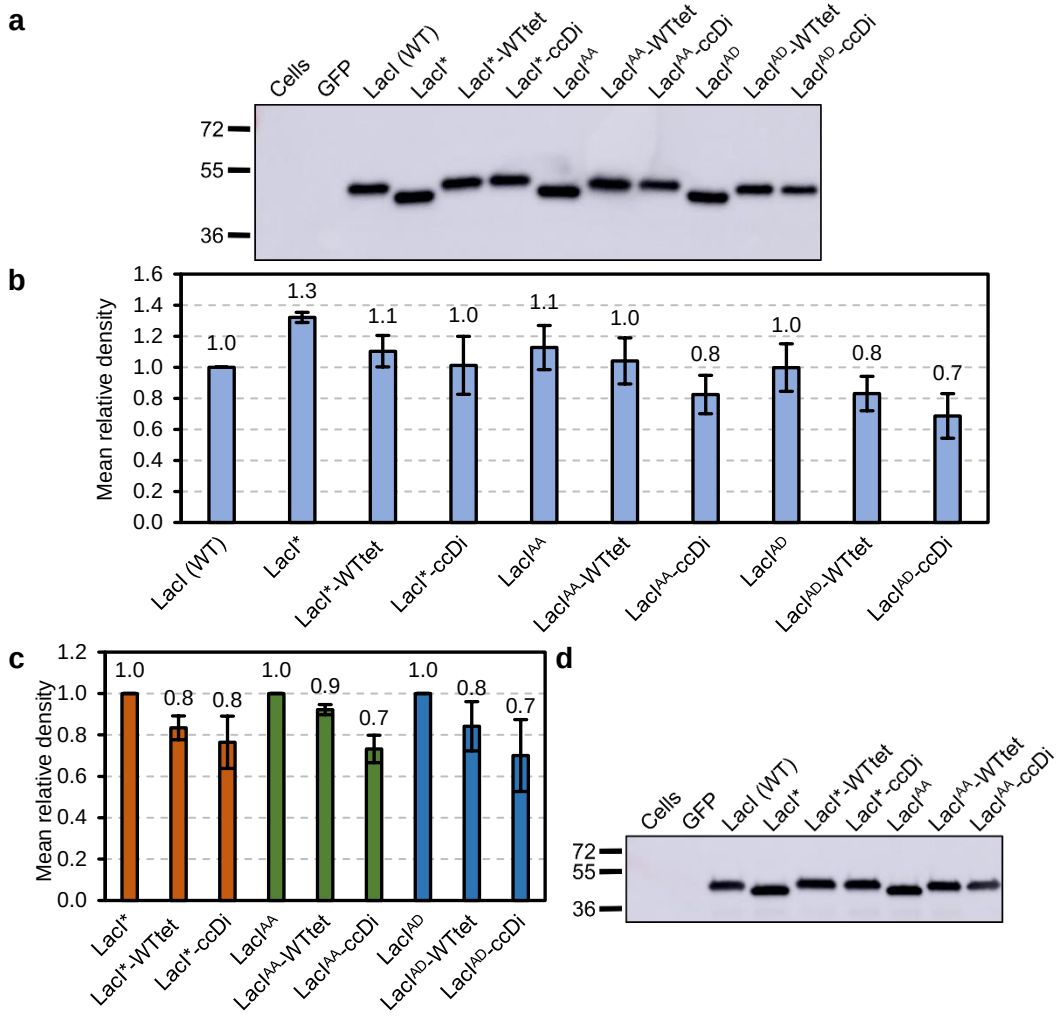


Figure 5-20 Western blotting analysis of LacI*-ccDi, LacI^{AA}-ccDi and LacI^{AD}-ccDi proteins. (a) Representative membrane for samples induced with 0.2 % arabinose. (b) Mean relative densities of detected bands relative to LacI (WT), measured following induction by 0.2 % arabinose, n=3. (c) Mean relative densities of LacI*, LacI^{AA} or LacI^{AD} fused to WT tetramerisation domain or CC-Di, relative to LacI*, LacI^{AA} or LacI^{AD}, respectively, n=3. (d) Representative membrane for samples induced with 0.02 % arabinose, n=3. Error bars are one s.d. from the mean.

variants (Figure 5-20c). The samples were also blotted following induction with 0.02 % arabinose (Figure 5-20d).

LacI^{AA} fused to the heterotetrameric peptides A1 and B1 was investigated following induction with 0.02 % arabinose (Figure 5-21a). Like LacI, LacI^{AA} was also present at a lower level when expressed from the pVRc plasmid compared to when it was expressed from pBAD (Figure 5-21b). The acidic peptide A1 led to much lower levels of both LacI* and LacI^{AA} (Figure 5-21c) while the basic peptide B1 led to small decreases in the levels of these proteins (Figure 5-21d).

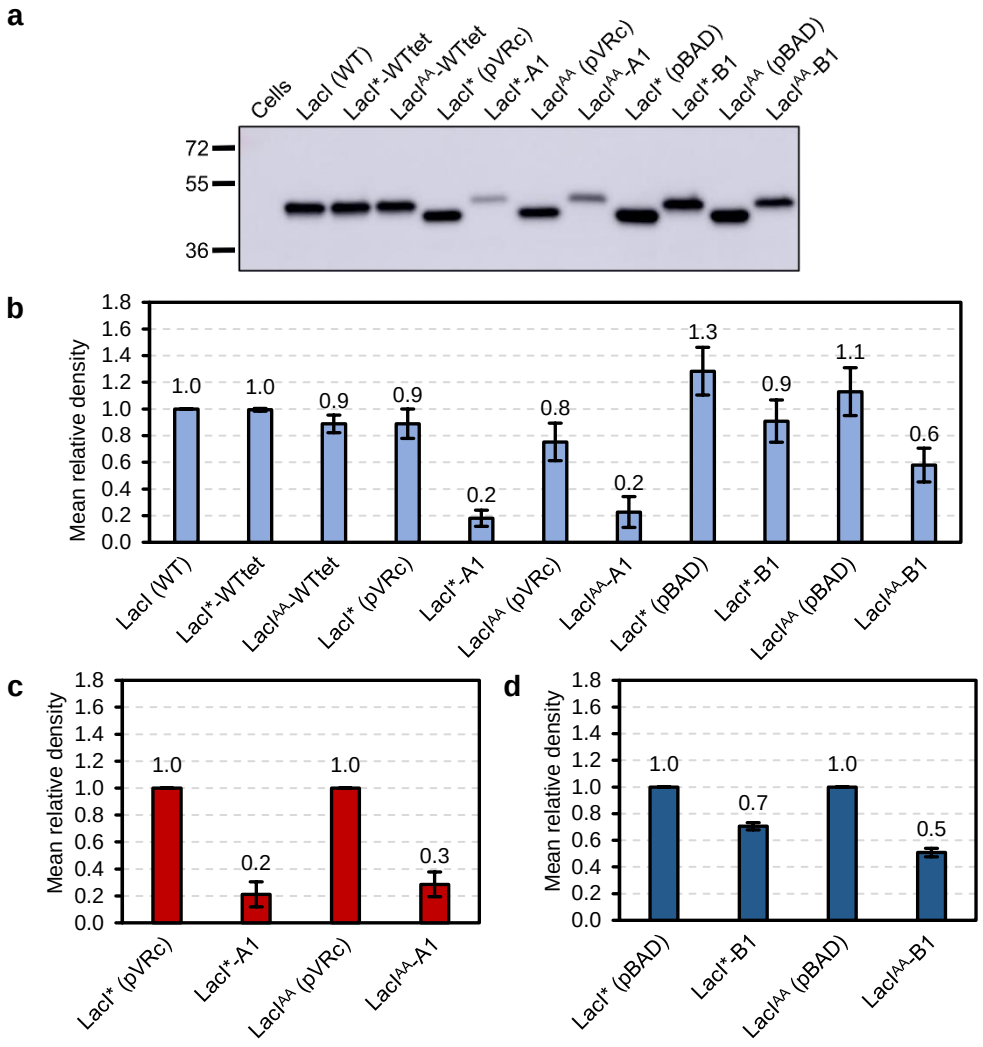


Figure 5-21 Western blotting analysis of LacI^{AA}-peptide fusion proteins. (a) Representative membrane for samples induced with 0.02 % arabinose. (b) Mean relative densities of detected bands relative to LacI (WT), measured following induction by 0.02 % arabinose, n=3. (c) Mean relative densities of LacI*-A1 and LacI^{AA}-A1 relative to LacI* and LacI^{AA}, respectively, n=3. (d) Mean relative densities of LacI* and LacI^{AA} fused to basic peptide B1 relative to LacI* and LacI^{AA}, respectively, n=3. Error bars are one s.d. from the mean.

When investigated with induction with 0.2 % arabinose, the ΔHTH variants LacI-ΔHTH and LacI*-ΔHTH were both present at much lower levels than the versions with intact HTH domains, LacI and LacI* (Figure 5-22a and e). The addition of the basic peptide B1 to the C terminus of these proteins decreased their cellular level further to a level 10-fold lower than the level of LacI (WT). When the mean density of the bands was calculated relative to LacI-ΔHTH and LacI*-ΔHTH, B1 appeared to destabilise both proteins to a similar extent (Figure 5-22b and c).

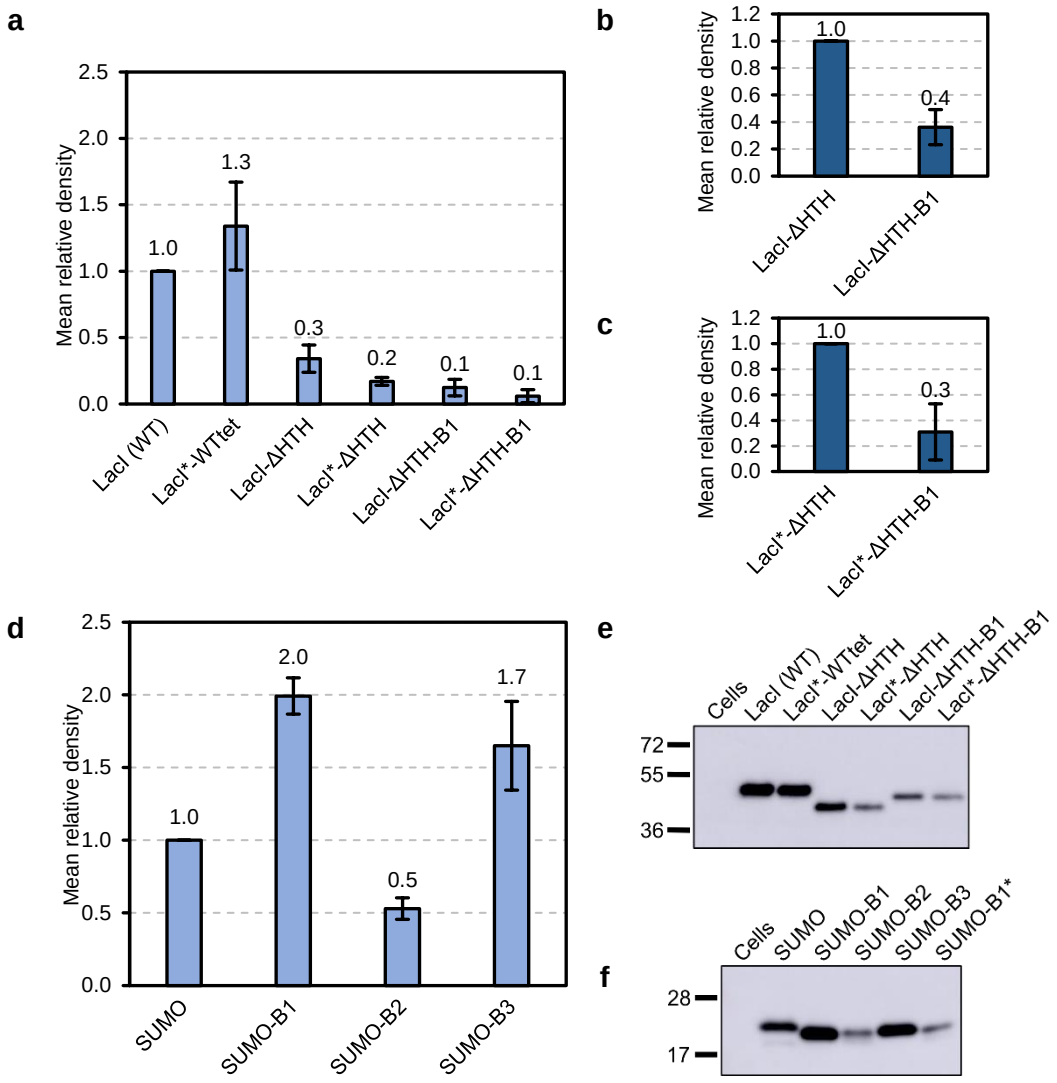


Figure 5-22 Western blotting analysis of LacI-ΔHTH-peptide and SUMO-peptide fusion proteins. (a) Mean relative densities of detected bands for LacI-ΔHTH proteins relative to LacI (WT), measured following induction by 0.2 % arabinose, n=3. (b) Mean relative density of LacI-ΔHTH-B1 relative to LacI-ΔHTH, n=3. (c) Mean relative density of LacI*-ΔHTH-B1 relative to LacI*-ΔHTH, n=3. (d) Mean relative densities of detected bands for SUMO proteins relative to SUMO (no-coiled coil control), measured following induction by 0.02 % arabinose, n=3. Representative membranes for (e) LacI-ΔHTH samples induced with 0.2 % arabinose and (f) SUMO samples induced with 0.02 % arabinose. Error bars are one s.d. from the mean.

The SUMO-B fusion proteins were investigated with induction by 0.02 % arabinose (Figure 5-22d and f). The B2 peptide decreased the level of SUMO relative to the SUMO only (no-coiled coil control) following the pattern observed for LacI*. However, unlike with LacI*, the peptides B1 and B3 appeared to increase the level of SUMO.

In summary, appending coiled-coil sequences to the C termini of the investigated proteins generally led to a decrease in the level of those proteins in the cell, as detected by western blotting. This may have been due to increased degradation or to lower levels of protein synthesis for these proteins. For example, when the acidic and basic peptides are not interacting with their partners to form a heterotetrameric coiled coil, they are fully unfolded. This may make them a target for cellular unfolded protein stress responses, leading to the degradation of whatever protein they are fused to [376-379].

The heterotetramers can be compared to intrinsically disordered proteins: the acidic and basic peptides are highly disordered alone, but they undergo a disorder-to-order transition and gain structure on interacting with their partner. Intrinsically disordered proteins are believed to be very widespread in nature and more than 20 % of essential proteins in *E. coli* are predicted to contain an intrinsically disordered region [380,381]. Furthermore, though often involved in important cellular functions, these proteins tend to be tightly regulated and maintained at low levels. This mitigates the problems associated with accumulating unfolded proteins in the cell such as aggregation and other non-specific, off-target interactions [382,383]. The components of the novel transcription factors described here may be treated in the same way as these endogenous intrinsically disordered proteins.

However, though the components of the heterotetramers were expected to be disordered when alone, both the WT Lac repressor tetramerisation domain and CC-Di, were expected to be folded as homomers. Yet these coiled coils also led to lower levels of the fusion partner. As homomers, these coiled coils do not rely on a partner for folding so, provided they are present at a high enough concentration, there should be little unfolded peptide present. Therefore, there may be some other feature of the *de novo* coiled coils, not just their structure, that influences their stability in the cell.

For example, the peptide A1 led to much lower protein levels than the other acidic peptides A2 and A3, which may imply the decrease in protein levels was influenced by the peptide sequence. For example, the A1 peptide sequence may have contained a protease recognition sequence, which would have led to a decrease in the protein level through increased degradation. The degradation could also have occurred at the mRNA level rather than the protein level.

Alternatively, rather than degradation being the source of the low protein levels, the LacI*-A1 protein may instead have been synthesised less efficiently than the other proteins. The LacI*-A1 protein runs at a slightly lower molecular weight than expected and, though this could be due to the high localised charge on the acidic peptide causing the protein to run aberrantly, it may have also been an indication of a specific truncated product. Mass spectrometry could be used to determine whether the protein is indeed synthesised correctly and further investigation using techniques such as qPCR might help determine whether this truncation, if present, occurs at the transcription, translation or post-translation level.

The changes in protein level may be context-dependent. For example, the B1 and B3 peptides were found to destabilise the Lac repressor derivatives but stabilise SUMO. The peptides may induce misfolding of some of the proteins but not others, leading to increased degradation. However, the function of the Lac repressor derivatives did not appear to be excessively perturbed – they were still able to cause gene repression – so their structures may have only been partially disrupted.

Finally, while CC-Di has an overall charge of +1 with a pI of approximately 8.4, the acidic and basic components of the heterotetramers are highly charged. For example, A1 has an overall charge of -5 and a pI of approximately 4.3 and B1 has an overall charge of +8 and a pI of approximately 10.9. Given that these pI values are far from the cytoplasmic pH of *E. coli* (approximately 7.2-7.8 [384,385]) the peptides are likely to retain their extreme charge inside the cell, making them prime targets for non-specific interactions with other, charged cellular components including proteins. For the basic peptides this could also include DNA, RNA and ribosomes [386]. While non-specific interactions would surely interfere with their *in vivo* functions, it remains to be seen what effect high localised charge has on protein stability, if indeed there is one given that localised charge can be an important aspect of some protein functions, such as specific and non-specific interactions between DNA-binding proteins and DNA.

5.5.2 Heteromeric peptides have a chaperoning effect over their partners

An observation that indicated the unfolded nature of the acidic and basic components of the heterotetramers may have contributed to the decreased stability of the fusion proteins was that the co-expression of acidic and basic partner-tagged proteins appeared to somewhat protect both proteins from degradation. That is, when LacI*-A1 and SUMO-B1 were co-expressed, each protein was present at a

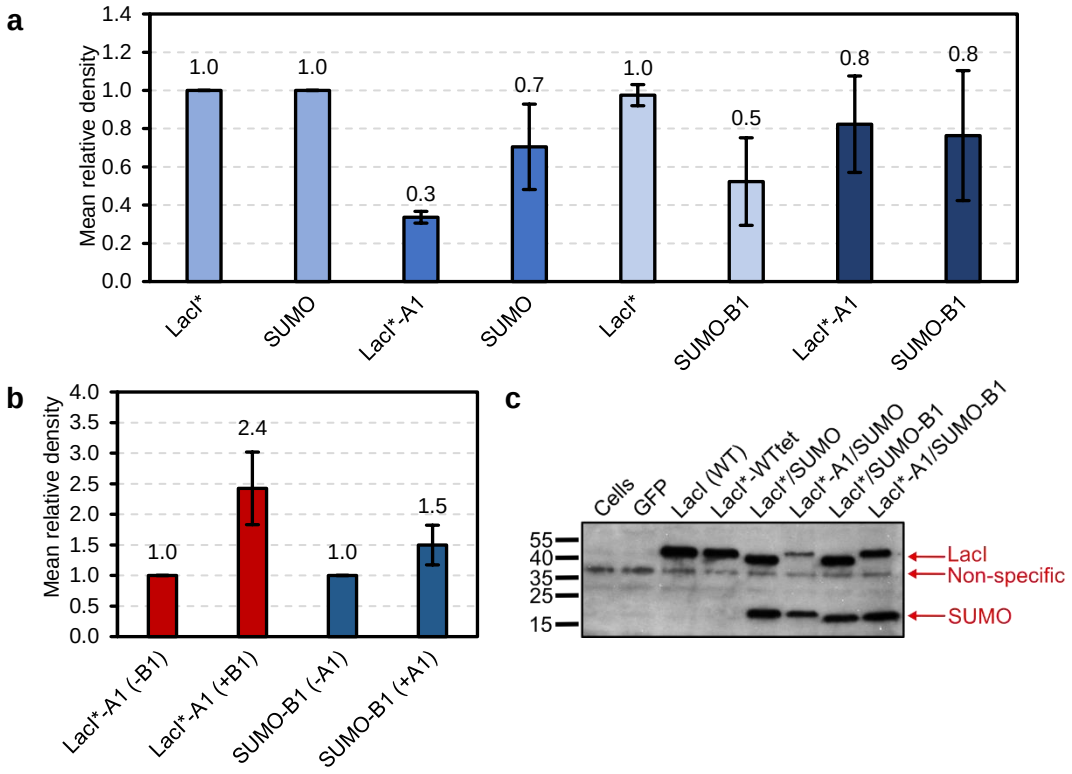


Figure 5-23 Western blotting analysis of LacI*-A1 and SUMO-B1 proteins following cell growth for a repression assay with induction by 0.002 % arabinose. Cells were co-transformed with pVRc plasmids encoding LacI* or LacI*-A1 and pBAD plasmids encoding LacI* or LacI*-B1 (and a pVRb plasmid encoding the reporter gene). (a) Mean relative densities of detected bands for proteins relative to LacI (WT), measured following induction by 0.002 % arabinose where co-expressed proteins are indicated by different shades of blue, n=4. (b) Mean relative densities of LacI*-A1 in the presence of SUMO-B1 relative to the level of LacI*-A1 on the presence of SUMO, and SUMO-B1 in the presence of LacI*-A1 relative to the level of SUMO-B1 in the presence of LacI*, n=4. (c) Representative membrane for LacI*-A1 and SUMO-B1 samples induced with 0.002 % arabinose, positions of LacI derivatives and SUMO proteins indicated. Non-specific bands are indicated. Error bars are one s.d. from the mean.

higher level compared to when either LacI*-A1 or SUMO-B1 were expressed with the corresponding SUMO or LacI* no-coiled coil control. This phenomenon was measured by blotting cells harvested post-repression assay with induction by 0.002 % arabinose (Figure 5-23a and c).

When LacI*-A1 was co-expressed with SUMO-B1 the mean relative density relative to LacI* increased to 0.8 from 0.2 when LacI*-A1 was co-expressed with SUMO alone (Figure 5-23a). Similarly, when SUMO-B1 was co-expressed with LacI*-A1 the mean relative density relative to SUMO increased to 0.8 from 0.5 when co-expressed with LacI*.

This effect was more obvious when the level of LacI*-A1 in the presence of SUMO-B1 was normalised relative to the level of LacI*-A1 in the presence of SUMO (*i.e.* in the absence of B1) and when the level of SUMO-B1 in the presence of LacI*-A1 was normalised relative to the level of SUMO-B1 in the presence of LacI* (*i.e*. in the absence of A1) (Figure 5-23b). When analysed in this way, the level of LacI*-A1 increased 2.4-fold in the presence of B1 and the level of SUMO-B1 increased 1.5-fold in the presence of A1. Hence, it appears that the A1 and B1 peptides have a chaperoning effect over each other where, on folding into the obligate A1/B1 heterotetramer, they protect each other from degradation. This implies that the decreased levels of the proteins fused to the *de novo* sequences may be due to increased degradation of these proteins rather than decreased synthesis. However, there are certainly more questions to answer here and the destabilising/stabilising effects of *de novo* protein sequences, as well as their influences over the proteome at large, remain ripe for investigation.

5.6 Chapter conclusions

When fused to various other proteins the *de novo* designed heterotetrameric coiled coils A1/B1, A2/B2 and A3/B3 have been found to interact in *E. coli* to form complexes that bring about repression of a reporter gene, GFP (Figure 5-24). The strongest levels of repression are achieved when both the acidic and the basic components are attached to the DNA-binding protein, LacI*, where the resulting complexes are presumably tetrameric and contain two dimeric DNA-binding modules. However, this variant of the Lac repressor was also found to display residual dimerisation at elevated concentrations so alternative variants were considered. When fused to the Lac repressor variant $LacI^{AA}$ the heterotetramer A1/B1 is still able to repress GFP expression and the $LacI^{AA}$ proteins exhibit reduced residual dimerisation.

The heterotetramers are also able to repress GFP expression when the basic peptides are fused to the non-DNA-binding protein SUMO or to just a 6 His tag, albeit with lower strength than when both peptides are fused to a DBD. The level of repression can be improved by increasing the arabinose concentration, but this also leads to increased residual dimerisation of the LacI*-A and $LacI^{AA}$-A proteins. Therefore, alternative means of controlling the expression of the components were investigated. When $LacI^{AA}$-A1 is expressed from a low-level constitutive promoter, *pro1*, and SUMO-B1 or 6H-B1 are expressed from P_{araBAD}, the level of the basic

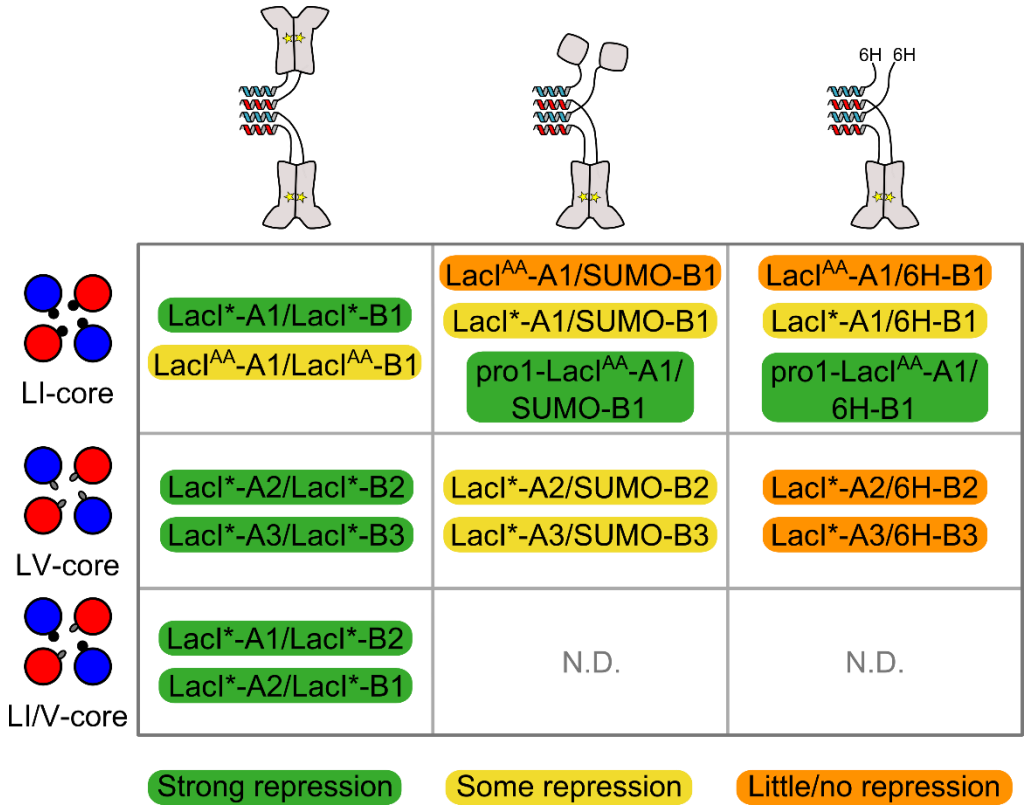


Figure 5-24 Summary of constructs discussed in Chapter 5. The constructs that were able to achieve relatively strong levels of repression are highlighted in green, while those that achieved a small amount of repression, or no repression are highlighted in yellow and orange, respectively.

components can be increased independently by increasing the arabinose concentration. This also leads to an increase in repression but avoids the residual dimerisation. As well as demonstrating that the *de novo* coiled coils fold as expected in *E. coli*, this work has generated a number of well characterised, semi-artificial transcription factors that are ready to be used to control the expression of more-useful genes of interest.

Furthermore, regulatory components that are compatible with the transcription factors have been identified. These could now be used to design genetic circuits based on coiled coils with increasingly complex behaviours, such as switching or strand displacement. For example, in a triggered-disassembly system (like the one introduced in Section 5.3.3), where a single peptide disrupts a complex between a second and third peptide by interacting strongly with one of those peptides, the disrupting component could be expressed from an inducible promoter such as P_{araBAD} while the other two components could be expressed from constitutive promoters like *pro1*. Thus, the interaction between the second and third peptides,

and co-localisation of any proteins they are fused to, could be disrupted in an inducible manner.

Western blotting revealed that appending *de novo* coiled-coil sequences to naturally occurring proteins impacts the level of those proteins in *E. coli*. Generally, the *de novo* peptides lead to decreased protein levels. Though the reasons for this remain unclear, the unfolded nature of the acidic and basic peptides is likely to be a key contributing factor [377]. Therefore, though the peptides do interact as anticipated, their sequences are not necessarily optimised yet for *in vivo* applications.

While an initial analysis of the *in vivo* behaviour of *de novo* coiled coils has been performed, many questions remain. For example, do the *de novo* peptides have off-target interactions with other cellular components? Is the behaviour of the *de novo* complexes homogenous across the population, or is there cell-cell variability in the level of repression achieved by the complexes? How do the characteristics of the tetramers differ from those of, for example, dimers or coiled coils with higher oligomeric states? And what are the effects of expressing these, and other, semi- or fully-artificial proteins on the endogenous *E. coli* proteome?

The work presented in this chapter represents a first step into the arena of protein design in the cell, that is, designing completely novel proteins that will behave predictably and effectively inside living organisms, not just in the test tube. Just as a toolkit of routine techniques has been refined for the characterisation of *de novo* coiled coils *in vitro*, so too should an equivalent toolkit be developed for characterising new designs *in vivo*. Such a toolkit might include pull-down assays or co-immunoprecipitation to analyse protein-protein interactions; quantitative western blotting to determine how effectively proteins are expressed; bioinformatics to detect the presence of unwanted sequence features such as protease sites; and proteomics techniques such as Stable Isotope Labelling with Amino acids in Cell culture (SILAC) for measuring the effect of *de novo* proteins on endogenous protein levels. In time, this will allow for the determination of rules to guide the design of proteins that are more and more suited for use in living systems, both in *E. coli* and beyond.

Chapter 6: Towards increasingly complex tetrameric coiled coils

6.1 Chapter introduction

Thus far, a set of homo- and heterotetrameric coiled coils has been designed and characterised *in vitro*. Select A_2B_2 heterotetramers have also been demonstrated to interact as designed in *E. coli*. These designs can therefore be used as artificial PIDs *in vivo*. However, there are still targets within tetrameric coiled coil and 4-helix bundle design space that have not been explored here, including ABCD tetramers and dimeric helix-loop-helix or helix-turn-helix structures.

Furthermore, the A_2B_2 protein-protein interactions are currently constitutive: provided the two components are co-expressed at sufficient concentrations they will interact. Therefore, the only means of controlling the interaction is at the transcription/translation level by selectively activating or repressing the expression of the heteromer components. However, control at this level is generally slow. Methods for promoting or blocking protein-protein interactions at the post-translational level, for example through ligand binding or post-translational chemical modifications, would be very useful in the context of ATFs. Ligand binding is a particularly alluring, albeit highly challenging, prospect.

6.2 Towards ABCD heterotetramers

Discussed here are preliminary results towards designing ABCD heterotetramers, where all four constituent peptides have different sequences. As well as being an unexplored target in *de novo* coiled-coil design, these coiled coils would represent a highly useful addition to the set of artificial PIDs.

6.2.1 ABCD heterotetramer design

Using a design strategy modified from that of a previously characterised heterotrimeric coiled coil [387,388], the sequences and ABCD heterotetramer combinations of the peptides 1–8 were generated computationally (Figure 6-1). In

all peptides, *e* and *g* positions were Glu or Lys with charge patterning designed to achieve the maximum number of productive ionic interactions in the ABCD species. Attractive interactions in the alternative unwanted A_2B_2 and A_2BC states were minimised. All *a* positions were Leu to promote tetramer formation while all *d* positions were β-branched Ile or Val. The identity of the *d* residues depended on the nature of the ionic interactions in the heptad in question: where interactions were predicted to be attractive in the homomer (Glu/Lys), *d* positions were Val to destabilise that heptad; where homomer interactions were predicted to be repulsive (Glu/Glu or Lys/Lys), *d* positions were Ile to promote a tetrameric oligomeric state. Therefore, the homomers would contain both all-Ile and all-Val (which are more destabilising) *d*-layers in their cores. The ABCD heterotetramers would contain mixed Ile/Val *d*-layers.

ABCD heterotetramers where all possible ionic interactions were made (16 Glu-Lys interactions in total) were filtered to remove those that contained peptide combinations judged likely to also engage in A_2B_2 or A_2BC interactions (based on predicted e_i-g_{i+2}' ionic interactions). Parametric models of the 120 ABCD heterotetramers remaining after this initial filtering were generated using ISAMBARD and scored using BUDE [293,294]. Heterotetramers were further filtered by their initial BUDE scores and to remove redundant sequences. The remaining three ABCD heterotetramers were then subject to parametric optimisation in ISAMBARD and scored again. The final generated ABCD heterotetramers contained 8 different peptides (peptides 1–8) in the combinations 1/2/3/4, 3/4/7/8 and 5/6/7/8 (Figure 6-2). The 1/2/3/4 and 5/6/7/8 combinations were selected for study and are referred to as 1234 and 5678, respectively. The 3/4/7/8 combination was not pursued due to time constraints.

Peptides 5–8 were given C-terminal mass tags to aid peptide identification. These mass tags follow the pattern GGxψ, where x is a small polar residue (Ser or Asp) and ψ is a chromophore (Trp or Tyr).

6.2.2 ABCD designs form heteromers in solution

6.2.2.1 CD spectroscopy shows ABCD designs form α-helical homomers and heteromers

Peptides 1–8 were investigated using CD spectroscopy alone and in pair, triplet and quadruplet (ABCD) combinations (Figure 6-3). Individually, the peptides showed a range of behaviours: some were relatively highly folded, such as

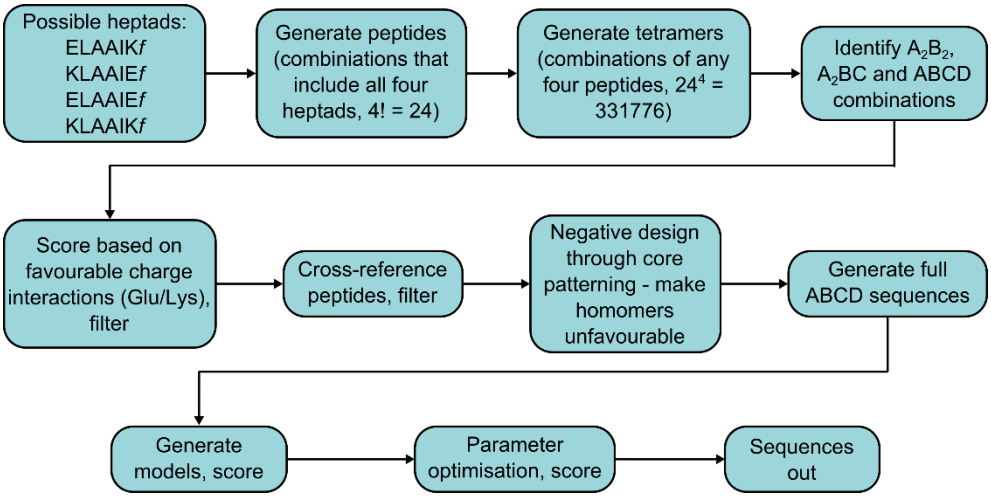


Figure 6-1 Schematic of the design strategy for ABCD heterotetramers.

peptides 1 and 6, while others were very unfolded, such as peptides 3 and 8 (Figure 6-3a and b). When investigated in the ABCD combinations, both 1234 and 5678 adopted α-helical structures with fraction helix values of 68 and 69 %, respectively. In both cases, the measured spectra for the ABCD combinations had greater α-helical character than the average of the spectra for the four individual peptides. This indicated that the peptides were interacting to form heteromeric species rather than existing as non-interacting homomeric species. However, it was not possible to determine whether all four constituent peptides were represented in the heteromers or if they were made up of a subset of the peptides. Given that peptides 3, 4, 7 and 8 were each relatively unfolded, the 3478 heteromer should also be investigated in the future.

Temperature-dependent CD measurements for the 1234 and 5678 combinations showed that both underwent reversible, cooperative unfolding transitions with similar T_M values of 72 and 71 °C, respectively (Figure 6-3c).

The off-target AB and ABC interactions were also investigated, *i.e.* the unwanted interactions between peptide pairs and triplets (Figure 6-3d, Figure 8-73–Figure 8-76). In most cases there appeared to be some unwanted interaction between these peptide combinations. For example, peptides 2 and 4 interacted to form a heteromer with a fraction helix value of 75 % – more folded than either ABCD heteromer (Figure 6-3d). For all other off-target combinations of peptides 1–4 the measured CD spectra had more negative MRE_{222} values than the averages of the spectra for the individual peptides (Figure 6-3e).

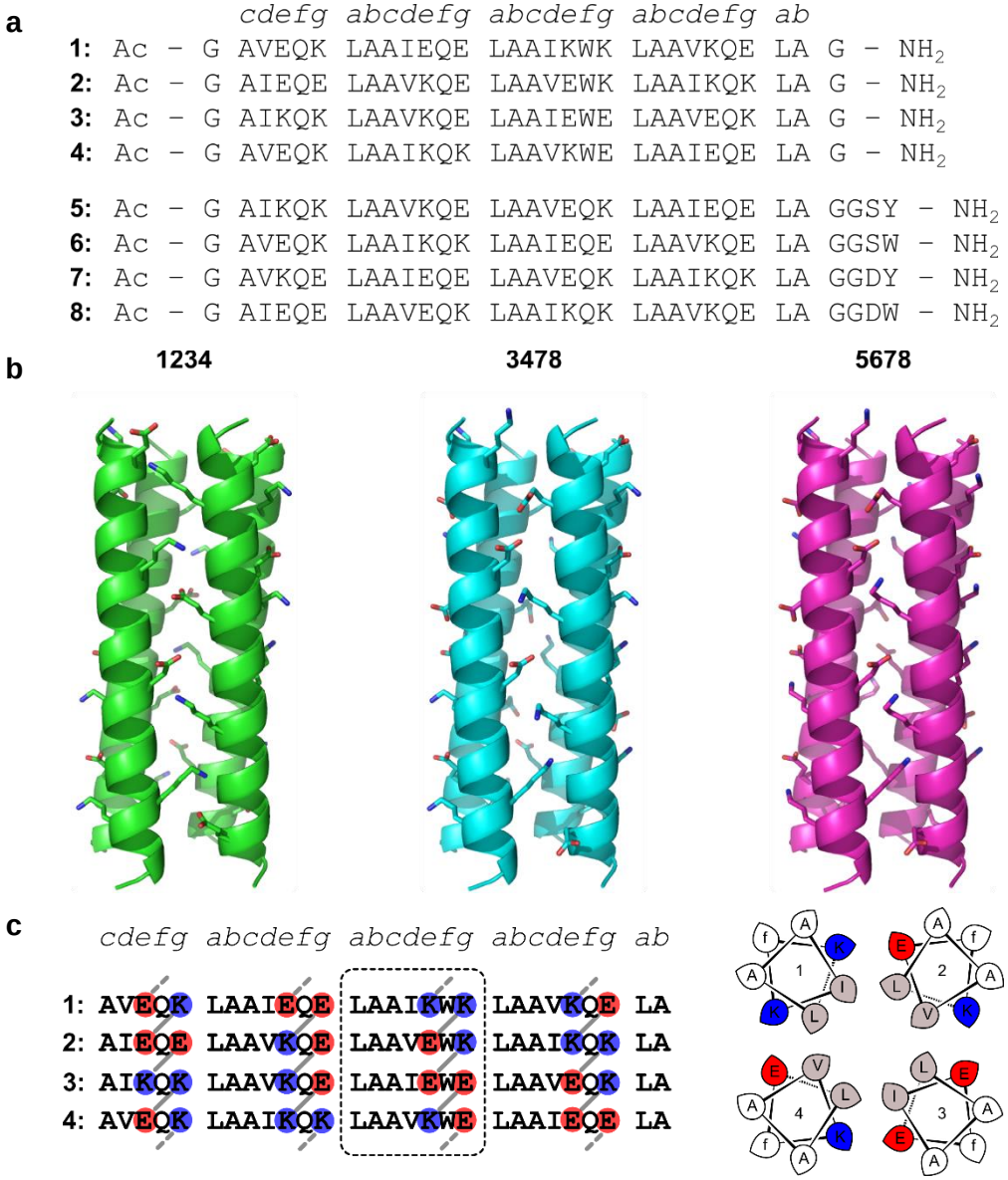


Figure 6-2 *De novo* designed ABCD heterotetramers. (a) Sequences of peptides 1–8. All peptides were N-terminally acetylated and C-terminally amidated. (b) Molecular models of heterotetramers 1234, 3478 and 5678 generated and optimised in ISAMBARD [293]. (c) Potential Glu-Lys ionic interactions in heterotetramer 1234 (left) and helical wheel showing one representative heptad (right). Interhelical ionic interactions are represented as grey lines. Peptide 1 interacts with 4, indicated by dashed grey lines. The heptad represented on the helical wheel is highlighted by the box.

Given that the peptides cross-interact extensively, it is not clear whether the ABCD combinations form a single heteromeric species (the ABCD heteromer) or a mix of several off-target heteromers and homomers. Only one unfolding transition was observed for each ABCD combination implying that only one species was present. However, if all of the species had similar T_M values it would not be possible to distinguish the separate transitions.

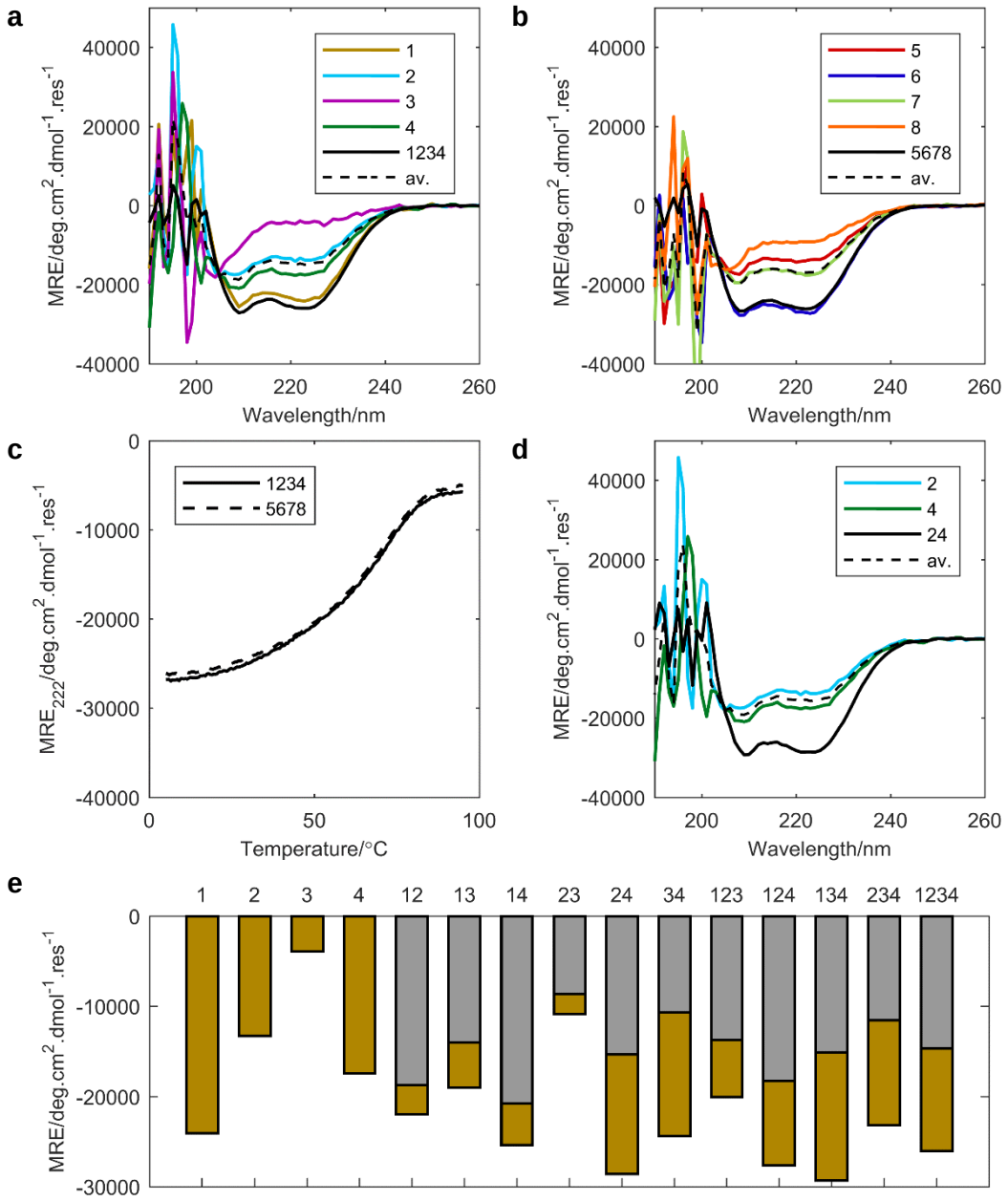


Figure 6-3 CD spectroscopy data for ABCD heteromers. (a) CD spectra at 20 °C of peptides 1–4. (b) CD spectra at 20 °C of peptides 5–8. (c) Temperature-dependent CD measurements monitoring MRE_{222} between 5 and 95 °C for heteromers 1234 and 5678. (d) CD spectra at 20 °C of peptides 1 and 2. (e) Bar graph showing MRE_{222} values of all combinations of peptides 1–4 (yellow) overlaid with the average MRE_{222} values of the individual peptides in those combinations (grey). Peptides were at 2 μM. All measurements were performed in PBS (pH 7.4).

The appearance of the off-target heteromers, despite the inclusion of potentially destabilising repulsive interactions in these heteromers may be explained in that the method used to generate the ABCD heterotetramers only considered e_i-g_{i+2}′ ionic interactions, *i.e.* those where an *e* residue interacts with a *g* residue in the same heptad of the adjacent helix (Figure 6-2c). However, the flexible nature of

Lys means e_i-g_{i-5}′ interactions are also possible, were an *e* residue interacts with the *g* residue in the previous heptad of the adjacent helix.

Revisiting the design procedure may be necessary to eliminate the unwanted "failure modes" more robustly, for example by introducing scoring penalties for these possible e_i-g_{i-5}′ interactions. Furthermore, while the optimal ionic interactions were selected with the assumption that the heterotetramers would adopt parallel orientations, the peptides may in fact have adopted antiparallel conformations that relieve repulsive ionic interactions.

6.2.2.2 AUC shows the ABCD designs do not form tetramers in solution

When the 1234 and 5678 heteromers were investigated by AUC, in both SE and SV experiments the heteromers fitted to molecular weights corresponding to trimers (Figure 6-4, Figure 8-135, Figure 8-136). In SV, both heteromers showed a single peak in their c(s) distributions (Figure 6-4b and d). However, it was not clear whether the ABCD combinations formed a single trimeric species or a mix of several trimers with very similar (or identical) molecular weights.

The trimeric oligomeric states may have been due to the presence of Val at *d*. While the LV-core A_2B_2 designs reliably formed heterotetramers, the LV-core homomers formed a range of oligomeric states and placing Val at *d* did not appear to promote tetramer formation (Chapters 3 and 4). Placing Ile at all *d* positions of peptides 1–8 may better promote tetramerisation but is also likely to increase the off-target homomerisation of these peptides.

While the strategy described here was not suitable for the design of ABCD heterotetramers, one side effect of the protocol was the generation of 576 novel A_2B_2 heteromer pairs. While filtering would be required to select the designs where the Glu-Lys interactions are optimised, all of the resulting pairs should be charge neutral, unlike the previously described A_2B_2 heterotetramers, which consist of overall highly acidic and basic peptides. As discussed in Chapter 5, such designs might be more suitable for use *in vivo*. Furthermore, given that LV-cores are compatible with A_2B_2 heterotetramer formation, it should be simpler to destabilise the homomers without affecting the heteromer oligomeric state.

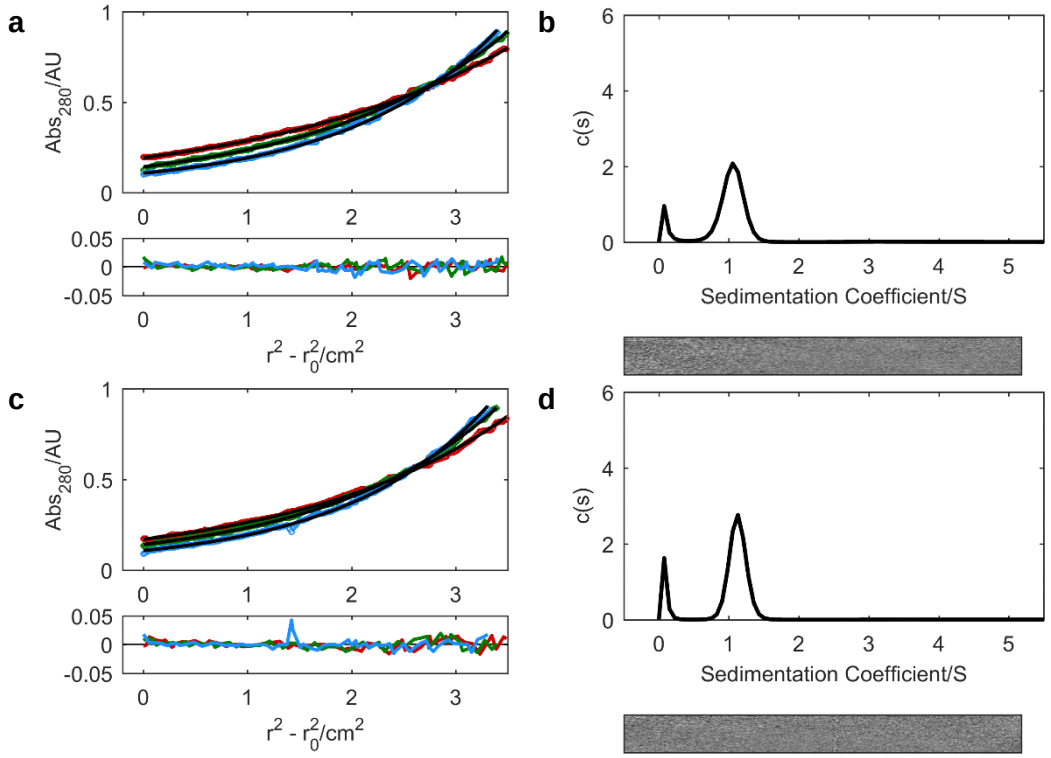


Figure 6-4 AUC data for ABCD heteromers. (a) Sedimentation equilibrium data at 30, 34, and 38 krpm (red, green and blue circles) and fits (black lines) (top) and residuals (same colours) (bottom) for 1234 returning a Mw of 9.1 kDa (2.8 x mean monomer mass). (b) Sedimentation velocity c(s) distribution (top) and residuals (bottom) for 1234 returning a Mw of 9.7 kDa (2.9 x mean monomer mass). (c) Sedimentation equilibrium data at 30, 33, and 36 krpm (red, green and blue circles) and fits (black lines) (top) and residuals (same colours) (bottom) for 5678 returning a Mw of 10.2 kDa (2.9 x mean monomer mass). (d) Sedimentation velocity c(s) distribution (top) and residuals (bottom) for 5678 a predicted Mw of 10.9 kDa (3.1 x mean monomer mass). All measurements were performed in PBS (pH 7.4).

6.3 Towards ligand-binding coiled coils

Although coiled coils are usually involved in simple protein-protein interactions, some coiled coils also demonstrate more-complex behaviours, notably ligand binding. There are many varied examples of small molecule binding to natural and designed coiled coils and helical bundles. This includes dye binding to α-barrels [261-264], drug and heme binding to 4-helix bundles [238,266,279,284] and metal binding [288,289]. Furthermore, binding can modify the structural properties of coiled coils by changing the oligomeric state or by inducing folding [275,282]. Therefore, attempts were made to introduce ligand binding into various coiled coils discussed in this thesis with the aim of using small molecules to alter their structural properties.

In nature, ligand binding occurs in many transcription factors with binding of specific ligands causing allosteric changes to the protein's structure that ultimately alter its mode of activity. There are many examples of these inducible transcription factors, including LacI, AraC and TetR [118,363,389]. These systems allow cells to sense small molecules in their environments and alter their behaviour accordingly. They have also been exploited widely by molecular and synthetic biologists as they allow gene expression to be switched on or off simply by adding small molecules to growth media.

Novel ligand-inducible transcription factors present a very appealing design target: transcription factors that can be readily engineered to respond to non-natural small molecules of interest would be invaluable in the design of biosensors [390]. Attempts have been made to alter the ligand specificity of naturally-occurring transcription factors and to convert existing binding proteins into transcription factors through protein fusions [29,159,391,392]. However, using rational protein design to approach this challenge remains largely unexplored.

Discussed here are initial attempts at introducing ligand binding into *de novo* designed coiled coils. The approaches taken are: (1) grafting of specific ligand binding motifs from other helical proteins into *de novo* tetramers; (2) introducing internal cavities that may be occupied by non-specific hydrophobic ligands to stabilise the otherwise destabilised structures; and (3) introducing destabilising core mutations to provide a starting point for high throughput selection for peptide sequences that undergo ligand-induced folding. The focus has been on promoting coiled-coil interactions using small molecules, although disruption of interactions using ligands also presents a powerful and rapid method for controlling the assembly state of coiled coil-based PIDs.

6.3.1 Binding motifs from natural proteins

An amantadine binding motif identified from the influenza A virus M2 proton channel was grafted into a *de novo* tetrameric coiled coil in an attempt to confer amantadine binding on this coiled coil.

The M2 proton channel is a transmembrane homotetrameric helical bundle involved in endosome acidification and viral release into the cytoplasm [267-269]. Amantadine is an antiviral drug previously used in the treatment of influenza that binds to the M2 transmembrane channel as well as at other secondary locations

[270-272,393-395]. Channel-lining residues proposed to contact amantadine were identified from X-ray crystal and NMR structures and introduced into a *de novo* homotetrameric coiled coil. The resulting peptides were intended to be water-soluble versions of the transmembrane M2 protein.

6.3.1.1 Design of M2-like homotetrameric coiled coils

The core and amantadine-contacting residues were identified from crystal and NMR structures of M2TM (a 25 residue transmembrane portion of M2 [396,397]) solved in the presence of amantadine [271,272]. Residues Val27, Ala30, Ser31 and Gly34 surround the drug and the 4-helix bundle core also contains a His residue involved in proton sensing, His37 (Figure 6-5a). As Gly is destabilising to α helices, Gly34 was replaced with Ala. The crystal structure of M2TM was solved for a variant with Ala at this position [271]. The resulting motif, VxxASxxGxxH (where x is any residue), was superimposed onto the sequence of ELAEIK, a homotetrameric coiled coil that was previously designed using a computational method. Two versions were designed, ELAEIK-M2-*a* and ELAEIK-M2-*d* where the motif started at a core *a* or *d* position, respectively (Figure 6-5b). In ELAEIK-M2-*a*, the His residue fell at a core *d* position. As placing charged residues in hydrophobic cores can be destabilising, a version of this peptide was also made without the His residue, ELAEIK-M2-*a*-HI (Figure 6-5b). In this peptide the *d* position is Ile, as in the parent sequence.

Peptide models based on the ELAEIK crystal structure were generated using the mutagenesis tool in PyMol (Figure 6-5a). Cavity volumes were determined using the CASTp server using a probe radius of 1.4 Å [398]. The cavity volumes of ELAEIK-M2-*a*, ELAEIK-M2-*a*-HI and ELAEIK-M2-*d* were calculated as approximately 236, 195 and 96 Å^3, respectively. The volume of one molecule of amantadine (Figure 6-6d) was determined to be approximately 228 Å^3, using Equation 6-1. While this volume is slightly smaller than the cavity volume of ELAEIK-M2-*a*, it is unlikely that amantadine would bind to any of the peptides without significant structural rearrangements to enlarge the cavity.

$$\text{Vol}\left(\text{cm}^3\right) = \frac{\text{Molar volume}}{\text{N}}$$

Equation 6-1 Equation for determining the volume of one molecule. Vol, volume of one molecule, cm^3; Molar volume, cm^3.mol^{-1}; N, Avogadro's number, mol^{-1}.

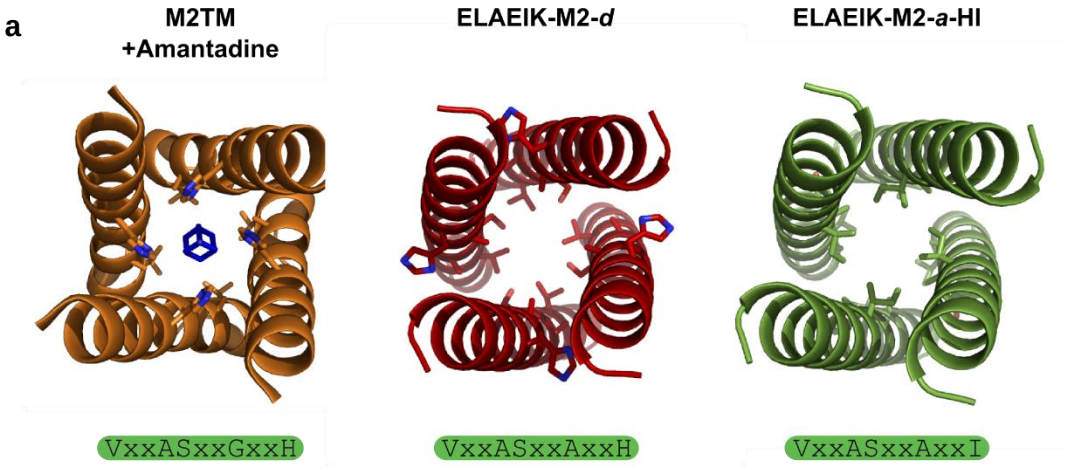


Figure 6-5 Design of M2-like peptides. (a) NMR structure of M2TM bound to amantadine (left, PDB ID: 2KQT) and models of ELAEIK-M2-*d* (centre) and ELAEIK-M2-*a*-HI (right). Models were generated from the ELAEIK crystal structure using the mutagenesis tool in PyMol. Amantadine-binding motifs are shown below each structure. (b) Sequences of M2TM [397] and M2-like peptides. All peptides were N-terminally acetylated and C-terminally amidated. Amantadine-binding motifs are highlighted in green.

6.3.1.2 The M2-like peptides are highly unfolded

The M2-like peptides and the parent, ELAEIK, were investigated using CD spectroscopy (Figure 6-6a). ELAEIK formed a well folded α-helical structure with a T_M of 84 °C (Figure 8-22). ELAEIK also formed a tetramer both in solution and in the crystal structure (Figure 8-94, Table 8-5). Conversely, all three M2-like peptides were entirely unfolded at 10 µM. When the peptide concentrations were increased to 750-1000 µM, the peptides still showed no folding (Figure 6-6b). In the presence of 1 mM amantadine, neither ELAEIK-M-*a* nor ELAEIK-M2-*a*-HI showed an increase in folding.

The peptides showed no folding at all in the presence and absence of ligand and were deemed too unstable to pursue further. In contrast, the M2 channel folds into a transmembrane 4-helix bundle even in the absence of amantadine [271]. Increasing the stabilities by increasing the lengths of the M2-like peptides may improve the system. However, it is also possible that the introduced VxxASxxAxxH motif simply

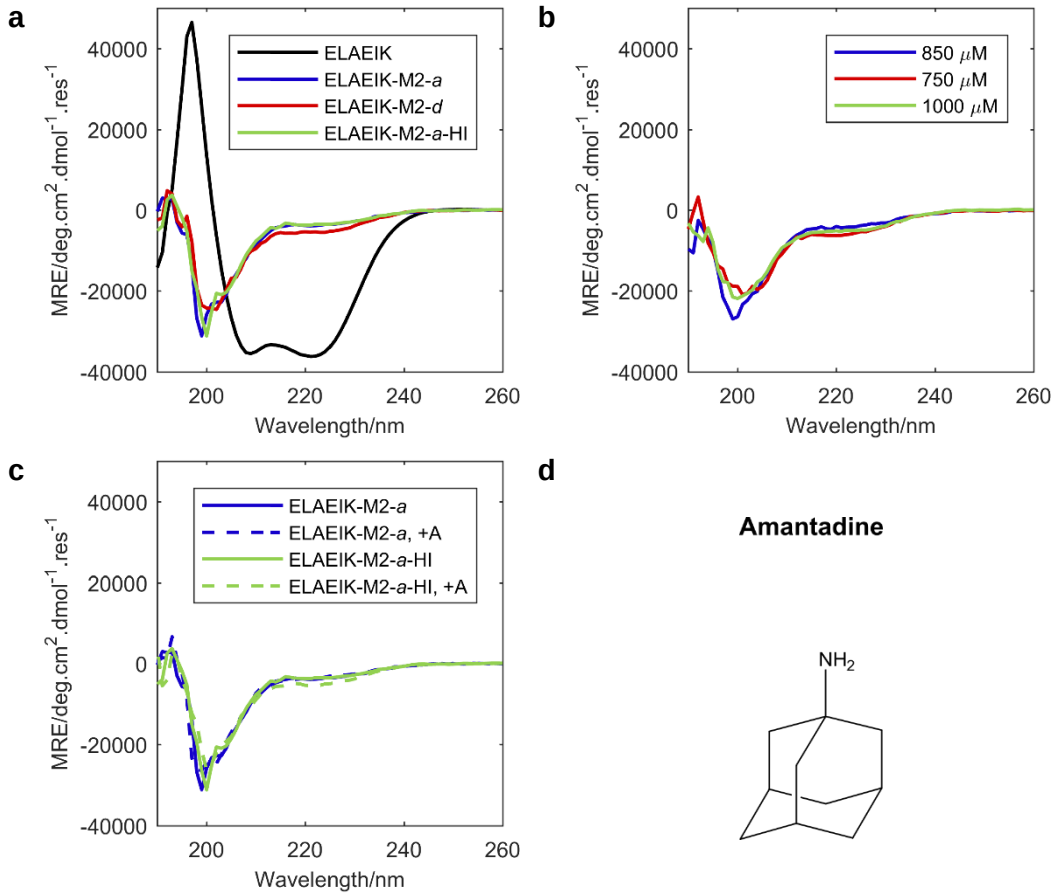


Figure 6-6 CD spectroscopy of M2-like peptides. (a) CD spectra at 5 °C of M2-like peptides at a peptide concentration of 100 µM. (b) CD spectra at 5 °C of M2-like peptides at variable peptide concentrations. Key colours as in (a). (c) CD spectra at 5 °C of peptides ELAEIK-M2-*a* and ELAEIK-M2-*a*-HI at a peptide concentration of 100 µM in the presence (+A) and absence of 1 mM amantadine. (d) Structural formula of amantadine. All measurements were performed in PBS (pH 7.4).

isn't compatible with coiled coil formation: while the M2 protein forms a 4-helix bundle, it does not form a coiled coil. Even a visual inspection of the structures shows that M2 and the ELAEIK-based models have quite different structures and that the residues identified as forming contacts with amantadine are not positioned in the same locations (Figure 6-5a). Therefore, more care is needed to design tetrameric coiled coils with amantadine binding sites. Additionally, less complex analogues could be pursued such as adamantane, which does not contain any polar substituents [277].

6.3.2 Non-specific hydrophobic ligands and chemical rescue of structure

Attempts were also made to introduce hydrophobic cavities into a coiled coil. These cavities were proposed to potentially accommodate hydrophobic ligands. This has

been achieved in the past in trimers [275], tetramers [276,277] and covalently templated 3-helix bundles [278] by replacing larger core hydrophobic residues such as Ile and Leu with smaller residues such as Gly and Ala. The resulting hydrophobic cavities can be occupied by small hydrophobic molecules such as benzene [275,276], adamantane [277] and hydrophobic dyes [278]. The bound hydrophobic molecules essentially replace the lost hydrophobic moieties and stabilise the otherwise destabilised, cavity-containing structures. This is akin to the phenomena of "chemical rescue of structure" and "chemical rescue of function" by small molecule complementation that have been demonstrated with various enzymes [399-403], receptors [404,405] and in an ATF [256]. Chemical rescue of structure generally involves substituting a bulky residue for a smaller one such that the protein structure is perturbed. The structure is then restored by adding some small molecule that binds to the protein, replacing the lost side chain. These mutations are often made in interaction interfaces such that the interaction can only occur in the presence of that small molecule. Inducible interactions of this nature would be invaluable in the design of ATFs because protein-protein interactions (and subsequent transcription repression or activation) could be induced rapidly on the addition of a small molecule. Similarly, chemical rescue of function can be achieved by mutating key catalytic residues in enzymes then replacing the functionality of the lost side chain by adding an analogous small molecule.

To investigate whether the structure of a coiled coil could be perturbed by mutation then restored by small molecule binding, substitutions were made in the core of the homotetrameric coiled coil CC-Tet [208]. The substitutions were proposed to create a cavity that could accommodate the environmentally-sensitive dye, DPH (Figure 6-7b). DPH fluoresces strongly at 452 nm when present in a hydrophobic environment and as such it has previously been used to probe lipid bilayers [406-408]. This property also makes it useful for probing cavities and channels in the hydrophobic cores of proteins and so, more recently, it has also been used to investigate the binding properties of hydrophobic pores in α-helical barrels [264]. While DPH was found to bind to the pores of the *de novo* barrels CC-Pent, CC-Hex2 and CC-Hept, it did not bind to CC-Tet, presumably because its tightly packed core does not contain any channels or cavities [208,264]. Therefore, core substitutions were made in CC-Tet that were proposed to enable DPH binding.

Two *d* Ile residues in CC-Tet were changed to Ala to produce the peptide CC-Tet-IA (Figure 6-7a). This was anticipated to result in a roughly dumbbell-

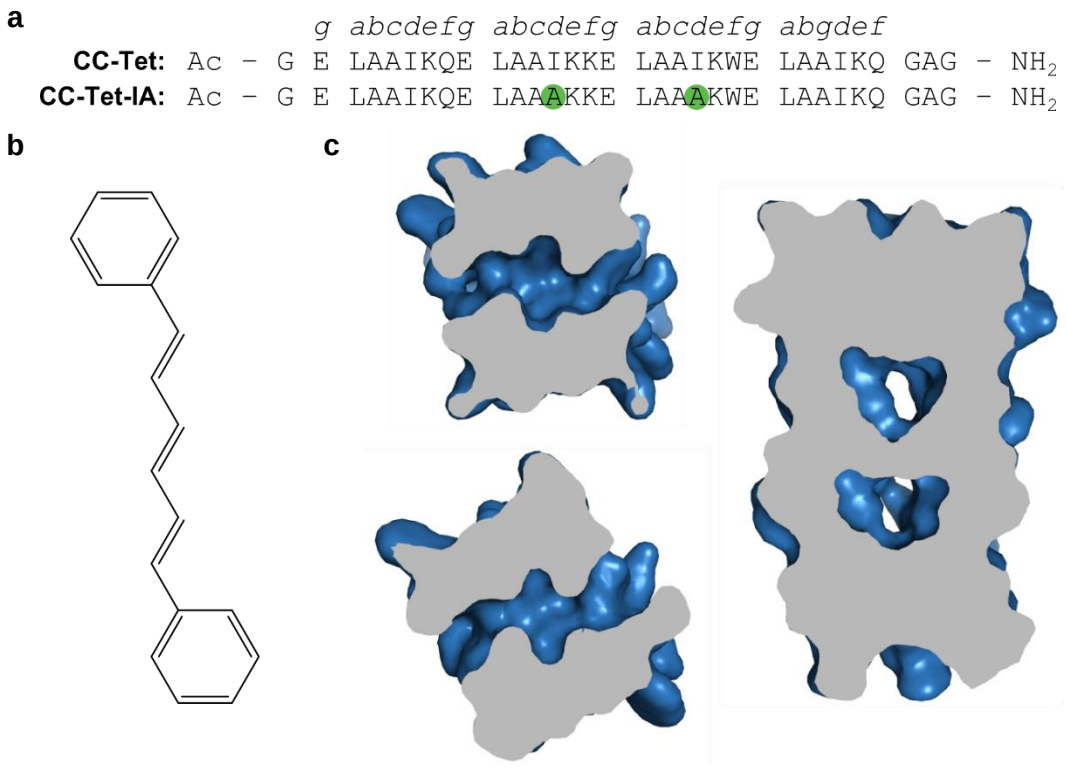


Figure 6-7 Design of DPH-binding homotetramer. (a) Peptide sequences of CC-Tet [208] and CC-Tet-IA. Both peptides were N-terminally acetylated and C-terminally amidated. Ile-Ala substitutions at *d*13 and *d*20 are highlighted in green. (b) Structural formula of DPH, an environmentally sensitive dye. (c) Molecular models of CC-Tet-IA showing the N-terminal (left, upper) and C-terminal (left, lower) cavities and the whole structure from the side (right). Structures are shown from the N-terminus or with the N-terminus at the top. Models were generated using the mutagenesis tool in PyMol using the crystallographic model of CC-Tet (PDB ID: 3R4A [208,213]).

shaped cavity in the tetramer core that may have accommodated DPH, where the aromatic rings of DPH would occupy one lobe each.

A molecular model of CC-Tet-IA was generated from the crystallographic model of CC-Tet (PDB ID: 3R4A [208,213]) using the mutagenesis tool in PyMol. The resulting model contains two separate cavities that form channels to the protein surface at two of the four interhelical interfaces in the coiled coil (Figure 6-7c). Given that these cavities were not linked in the model it is unlikely that DPH could occupy them without structural rearrangements that allow the aliphatic portion of the molecule to be accommodated. It may be necessary to make further substitutions to make a continuous channel. For example, the Leu at *a*17 could be changed to a smaller hydrophobic residue such as Val. However, all additional changes that introduce more small residues to the core are likely to further destabilise the coiled coil. Therefore, even if there was space for DPH to be accommodated, binding

would need to be very thermodynamically favourable in order to rescue the structure of the coiled coil.

When investigated with CD spectroscopy CC-Tet-IA formed an α-helical structure, like the parent peptide, CC-Tet (Figure 6-8a). Despite containing core Ala residues, which were expected to destabilise the structure, CC-Tet-IA was in fact almost as highly folded as CC-Tet. However, it was much less thermally stable than CC-Tet, with a T_M value of just 39 °C (Figure 6-8b). CC-Tet has a T_M above the measurable range at 10 μM. Furthermore, AUC revealed that CC-Tet-IA had a molecular weight corresponding to a dimer in both SV and SE experiments (Figure 6-8c, Figure 8-79).

Therefore, rather than resulting in a destabilised tetramer, the Ile-Ala substitutions have resulted in a completely different structure – a folded, though marginally thermally stable, dimer. The collapse to a dimer may minimise the volume of the internal cavities, which are expected to be destabilising. This has previously been observed in a model antiparallel coiled coil system where the relative positions of Ala residues in the cores controlled the oligomeric state: where the Alas were present in the same core layer in the coiled-coil structure the peptides formed a dimer with a relatively small cavity [227].

Furthermore, it remains to be seen whether this is a parallel or an antiparallel dimer. Previously, coiled coils with core Alas have resulted in antiparallel structures because, again, they minimise the volume of the destabilising internal cavities [189].

It was possible that the addition of DPH could induce CC-Tet-IA to re-adopt the parallel tetrameric structure. Such “allosteric” switches have been observed in coiled coils before, for example in a system based in GCN4-p1 with a core Ala substitution [275]. The resulting cavity somewhat destabilised the dimeric state but could accommodate benzene when the coiled coil adopted a trimeric state. Benzene binding stabilised the trimeric coiled coil and so it underwent a conformational change from dimer to trimer on ligand binding. Therefore, DPH binding by CC-Tet-IA was investigated using a fluorescence-based DPH-binding assay. Increasing peptide concentrations were incubated with 1 μM DPH and investigated using fluorescence spectroscopy. As previously demonstrated, CC-Tet did not demonstrate DPH binding [264], and there was no significant increase in DHP fluorescence on increasing the peptide concentration. CC-Tet-IA also did not show a significant increase in DPH fluorescence, indicating that DPH did not

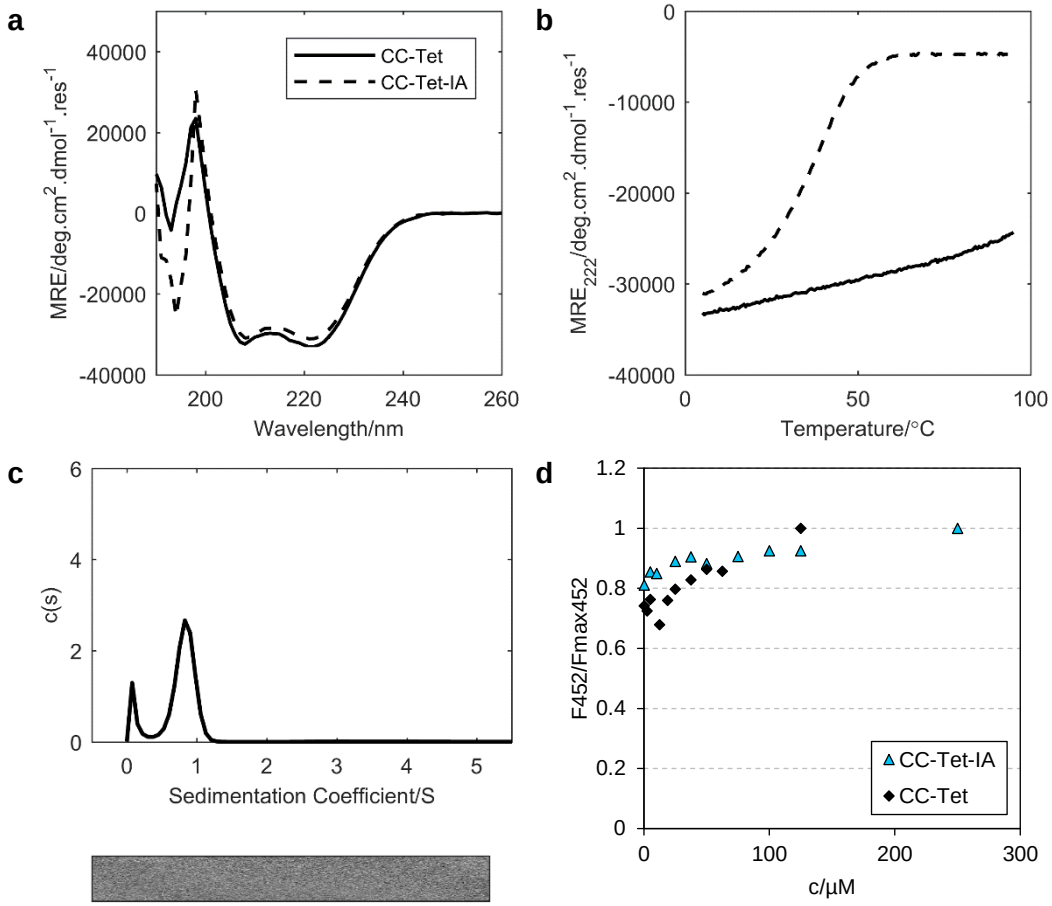


Figure 6-8 Biophysical characterisation of the CC-Tet core variant, CC-Tet-IA. (a) CD spectra at 5 °C of CC-Tet and CC-Tet-IA. (b) Temperature-dependent CD measurements monitoring MRE_{222} between 5 and 95 °C for CC-Tet and CC-Tet-IA. The peptide concentration for CD spectroscopy was 10 µM. (c) Sedimentation velocity c(s) distribution (top) and residuals (bottom) for CC-Tet-IA returning a predicted Mw of 6.5 kDa (2.0 x monomer mass). All biophysical measurements were performed in PBS (pH 7.4). (d) DPH-binding assay with CC-Tet and CC-Tet-IA. Varying concentrations of CC-Tet and CC-Tet-IA were incubated with 1 µM DPH then DPH fluorescence at 452 nm was measured. The assay was performed in PBS at 20 °C. c, coiled coil concentration (peptide concentration/oligomeric state).

bind to this peptide either. Therefore, the presence of the cavity-forming core substitutions at *d*13 and *d*20 was not sufficient to introduce hydrophobic dye binding activity into this peptide. Instead, the peptide formed a folded dimer that could not be encouraged to re-adopt a tetrameric structure by DPH binding.

6.3.3 In vivo *selection strategies for generating ligand-binding coiled coils*

A final approach to introducing cavities into the cores of coiled coils involved placing Ala at the *a* and *d* residues in the central heptad of a heterotetrameric coiled coil. Positioning Ala at two consecutive core positions was proposed to produce a

cavity directly in the centre of the coiled coil. As such, adopting an alternative conformation, such as an antiparallel structure, would not have led to a decrease in cavity volume and so the coiled coil should be globally destabilised (*i.e.* all possible structures are destabilised) and exist in a partially unfolded state. This is in contrast to CC-Tet-IA discussed above in which core Ala substitutions destabilised the tetrameric conformation but did not globally destabilise the peptide. Therefore, the peptide simply adopted a different structure rather than a partially unfolded structure, as intended.

Positions *a*14 and *d*17 in the peptides 1-LI-A and 1-LI-B (described in Chapter 4) were both changed to Ala to make the peptides 1-LI-A* and 1-LI-B*, respectively (Figure 6-9a). A molecular model of the heterotetramer was generated using CC Builder 2.0 because a crystal structure was unavailable for the 1-LI-AB parent [301]. The model revealed a cavity in the centre of the coiled coil (Figure 6-9b). The cavity was determined to have a volume of 51.8 $Å^3$ using the CASTp server with a probe radius of 1.4 Å. A cavity of this size would not be able to accommodate very large molecules. Even small hydrophobic molecules like benzene, which has a molecular volume of approximately 150 $Å^3$ according to Equation 6-1, would be too large. However, the peptides provide a proof of principle that coiled coils can be globally destabilised in this way, and the cavity volume could be enlarged in the future if necessary.

When 1-LI-A* and 1-LI-B* were investigated alone, both had similar CD spectra to the parent peptides 1-LI-A and 1-LI-B, indicating little folding (Figure 6-10a). When the peptides were mixed at 10 µM each, the CD spectrum looked similar to the spectra of the individual peptides and to the average of the individual spectra, showing the peptides did not interact at this concentration. When the peptide concentrations were increased to 100 µM there was a small increase in MRE_{222} and the fraction helix value increased from 39 % at the lower concentration to 53 % at the higher concentration (Figure 6-10a). Therefore, the peptides were still able to interact to some extent, but the interaction was greatly weakened. This is in contrast to the parent heterotetramer 1-LI-AB, which was essentially fully folded at 10 µM (Figure 6-10a).

The 1-LI-A* and 1-LI-B* peptides were also investigated using AUC, both alone and together. SE experiments of the individual peptides showed that they were monomeric in solution, like the parent peptides (Figure 8-107, Figure 8-108). SV experiments of the peptides together gave a predicted molecular weight of 4.3 kDa

a

```
                    cdefg abcdefg abcdefg abcdefg ab
1-LI-A :  Ac – G AIEKE LAAIEKE LAAIEWE LAAIEKE LA G – NH2
1-LI-A* : Ac – G AIEKE LAAIEKE AAAAEWE LAAIEKE LA G – NH2
1-LI-B :  Ac – G AIKQK LAAIKQK LAAIKWK LAAIKQK LA G – NH2
1-LI-B* : Ac – G AIKQK LAAIKQK AAAAKWK LAAIKQK LA G – NH2
```

b

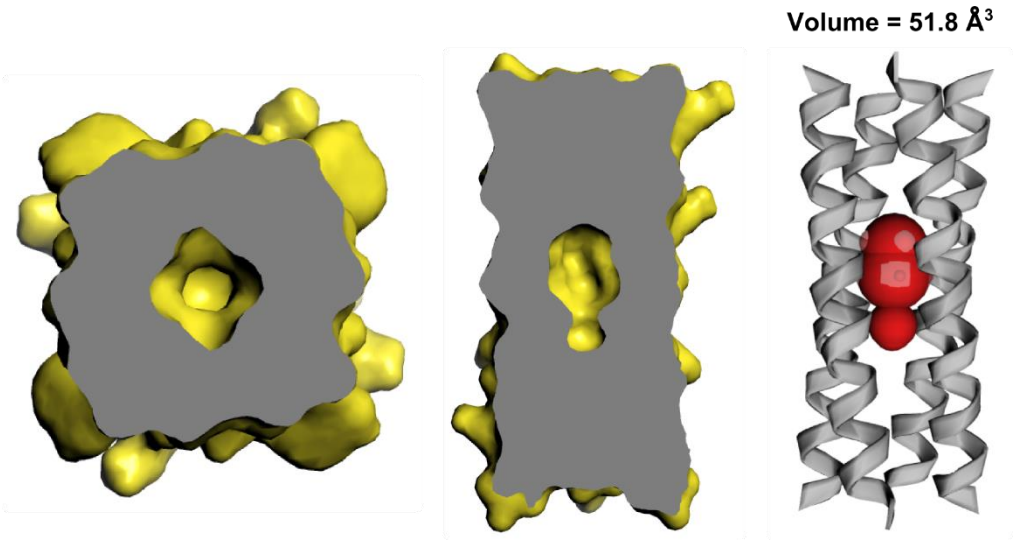


Figure 6-9 Design of a heterotetramer core mutant. (a) Peptide sequences for 1-LIA, 1-LI-B, 1-LI-A* and 1-LI-B*. All peptides were N-terminally acetylated and C-terminally amidated. Alanine substitutions are highlighted in green. (b) Molecular model for the 1-LI-AB* heterotetramer shown from the N-terminus (left), from the side with the N-terminus at the top (centre) and from the side with the N-terminus at the top with the cavity highlighted (right). The cavity volume is shown above the model. Models were generated in CC Builder 2.0 [301] using the default parameters for CC-Tet (radius, 6.8 Å; pitch, 213 Å; interface angle, 22.1 °). The cavity volume was calculated using the CASTp server using a probe radius of 1.4 Å [398].

(Figure 6-10c, Figure 8-134). This corresponded to an oligomeric state of 1.4, showing that there was lots of unfolded monomer present.

Unlike CC-Tet-IA, where only the tetrameric state was destabilised, this heteromer appears to be globally destabilised; the core substitutions have not caused the coiled coil to adopt an alternate structure and the peptides are instead mostly unfolded. This is likely due to the central location of the Ala substitutions that would create a similar cavity in both parallel and antiparallel structures. However, it is interesting that the peptides did not form, for example, a dimer in which the cavity volume would have been smaller, as was the case in CC-Tet-IA. Therefore, it remains to be seen whether the peptides will still form a tetramer when the concentration is greatly increased to the point where the peptides form a well folded heteromer.

Globally destabilised coiled coils, like the heteromer described above, may present a starting point for developing coiled coils that can undergo chemical rescue of structure. A coiled coil that is truly destabilised may be better than those that adopt alternative, metastable conformations as these other conformations represent competing states that may themselves be stabilised by ligand binding instead of the desired oligomeric state.

One possible approach to achieving chemical rescue of structure in this destabilised coiled coil is through high throughput screening and selection for ligand binding that induces the interaction between the peptides. Here, the *de novo* coiled coils could potentially undergo coupled folding and binding, as is observed with many intrinsically disordered proteins [409,410]. Alternatively, a partially destabilised protein could undergo ligand-induced stabilisation [411]. Destabilised or unfolded protein are generally degraded more rapidly *in vivo* [377], while stabilised proteins should have greater longevity. Ligand-stabilised mutants have been identified for many natural proteins including *E. coli* dihydrofolate reductase [412], human FKBP12 [413] and estrogen receptor [414], and a fluorescent protein from a freshwater eel, *Anguilla japonica* [415].

There are two main ways that coiled coils displaying ligand-induced folding or stabilisation could be generated: (1) libraries of small molecules could be screened for binding to a cavity containing coiled coil target; or (2) libraries of peptide sequences could be screened for binding to a ligand of interest.

Protein complementation assays use interacting exogenous components to bring together fragments of a reporter protein to restore its activity [416]. Such assays could be used to screen for peptide-small molecule combinations that rescue coiled-coil folding or stability and thus facilitate the interaction between the components of the split system (Figure 6-11a). These assays include bacterial two-hybrid assays like the Lac repressor-based transcription repression assay described in Chapter 5. Other two-hybrid assays based on transcription activation have also been described [109,167]. A transcription activation assay using a chemiluminescent reporter has previously been used to identify small molecules that can mediate the interaction between DNA and a mutant zinc finger deficient in DNA binding [256]. Alternatively, systems that involve the reconstitution of split fluorescent proteins [417,418] or enzymes could be used [419-422]. For example, a split dihydrofolate reductase system has previously been used to select optimal heterodimer sequences from an *in vivo* library of coiled-coil sequences [209,210]. A split T7 RNA polymerase has

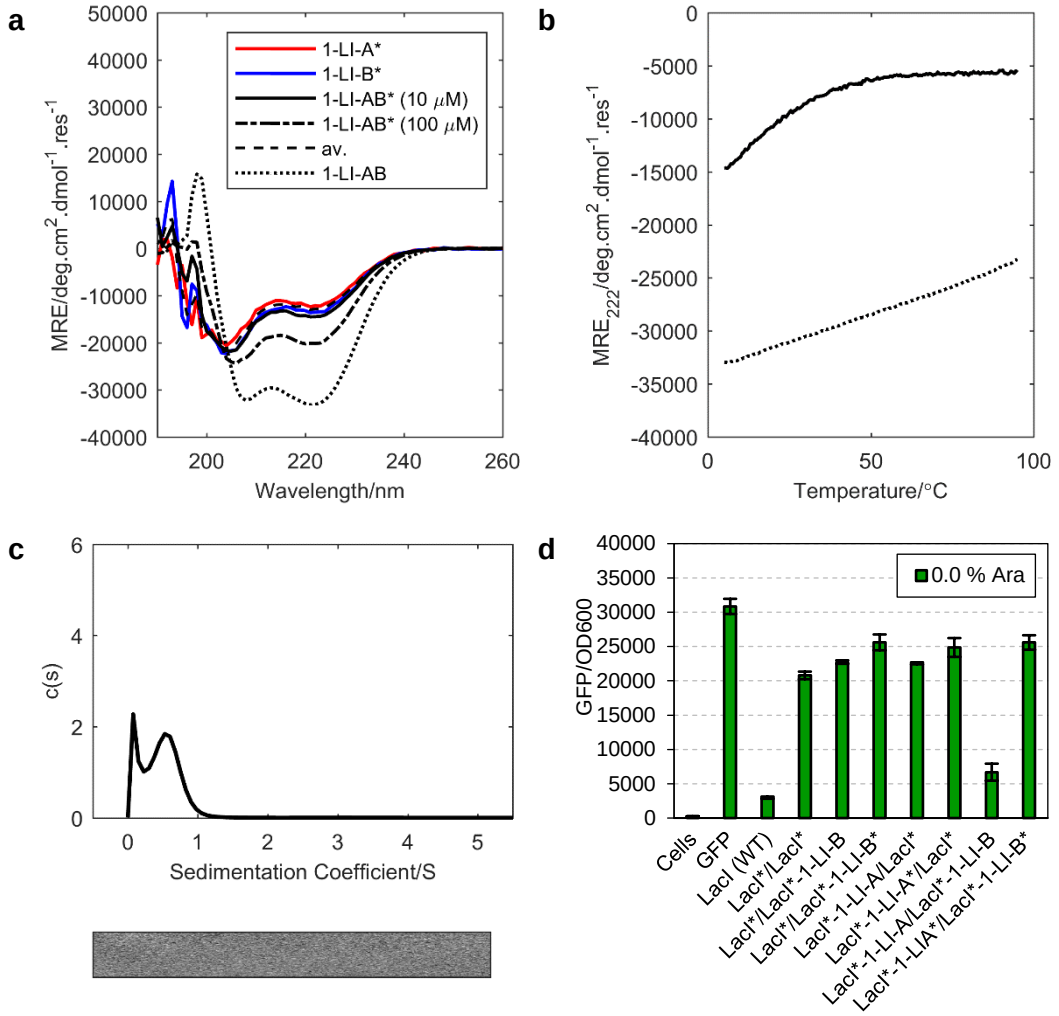


Figure 6-10 *In vitro* and *in vivo* characterisation of the heterotetramer core mutant. (a) CD spectra at 5 °C of peptides 1-LI-A* and 1-LI-B* and heteromers 1-LI-AB and 1-LI-AB*. (c) Temperature-dependent CD measurements monitoring MRE_{222} between 5 and 95 °C for heteromers 1-LI-AB and 1-LI-AB*. Peptides were at 10 µM, except where indicated. (c) Sedimentation velocity c(s) distribution (top) and residuals (bottom) for 1-LI-AB* returning a predicted Mw of 4.3 kDa (1.4 x mean monomer mass). All biophysical measurements were performed in PBS (pH 7.4). (d) Transcription repression assay for LacI*-1-LI-A/LacI*-1-LI-B and LacI*-1-LI-A*/LacI*-1-LI-B* with induction by 0.0% arabinose. Error bars are one s.d. from the mean, n=4.

also been proposed as a tool for selecting protein-protein and protein-ligand-protein interactions [423]. Other approaches to screening protein-protein interactions and protein folding that do not involve protein complementation include FRET-based assays and proteolysis assays [424,425]. Variants of the latter have been used to select stably folded variants of ubiquitin [426], *de novo* designed mini proteins [427] and ligand-stabilised variants of human FKBP12 (FK506 binding protein) [413]. Ligand-stabilised variants of computationally designed steroid-binding proteins have also been selected using transcription and fluorescence-based assays [123,428].

To demonstrate that the coiled coils described here may also be amenable to *in vivo* screening or selection, the 1-LI-A* and 1-LI-B* peptides were examined in the transcription repression assay introduced in Chapter 5 (Figure 6-10d). When neither or just one peptide was present, no repression of the GFP reporter was observed. The same was observed with the parent peptides, 1-LI-A and 1-LI-B. When the acidic and basic LacI*-peptide fusions were co-expressed, the 1-LI-A and 1-LI-B peptides interacted, with a fold repression value of 4.6. Conversely, the 1-LI-A* and 1-LI-B* peptides did not interact, and the fold repression value was the same as in the no coiled coil-control. Therefore, the core mutant peptides did not interact in *E. coli*.

This LacI*-1-LI-A*/LacI*-1-LI-B* system could potentially now be used to screen a library of small molecules for those that can rescue the heterotetramer interaction. However, as discussed above, the cavity in this heteromer was predicted to be very small and therefore the availability of small molecules that could occupy the cavity is likely to be quite limited.

The alternative approach is to screen a library of peptide sequences for binding to a small molecule of interest. A DNA library encoding peptides with a subset of amino acids at select core positions could be generated, transformed into *E. coli* and screened for ligand-dependent folding using a high throughput method such as fluorescence-activated cell sorting (FACS) (Figure 6-11b and c). This approach would not only allow greater freedom for the user to define the ligand properties (*e.g.* pharmacological activity), but it also expands the range of ligands that could be bound by the peptides. This is because, rather than simply containing hydrophobic cavities that can bind hydrophobic molecules, the coiled coils could contain cavities with more diverse chemical properties due to the introduction of residues with other chemical functionalities. For example, polar or charged amino acids could make specific interactions with complementary ligands.

Such *in vivo* selection strategies may, in principle, provide a means for generating coiled coils that undergo chemical rescue of structure. However, it is likely that the A_2B_2 heterotetramers described above are not the optimal starting point for such endeavours because the coiled coil-ligand interaction would involve five different components. The likelihood of all components colliding at the same time is very low. Therefore, other coiled coil or 4-helix bundle arrangements may need to be considered. For example, dimeric 4-helix bundles consisting of two helix-loop-helix (HLH) monomers, or single chain 4-helix bundles may be better options (Figure

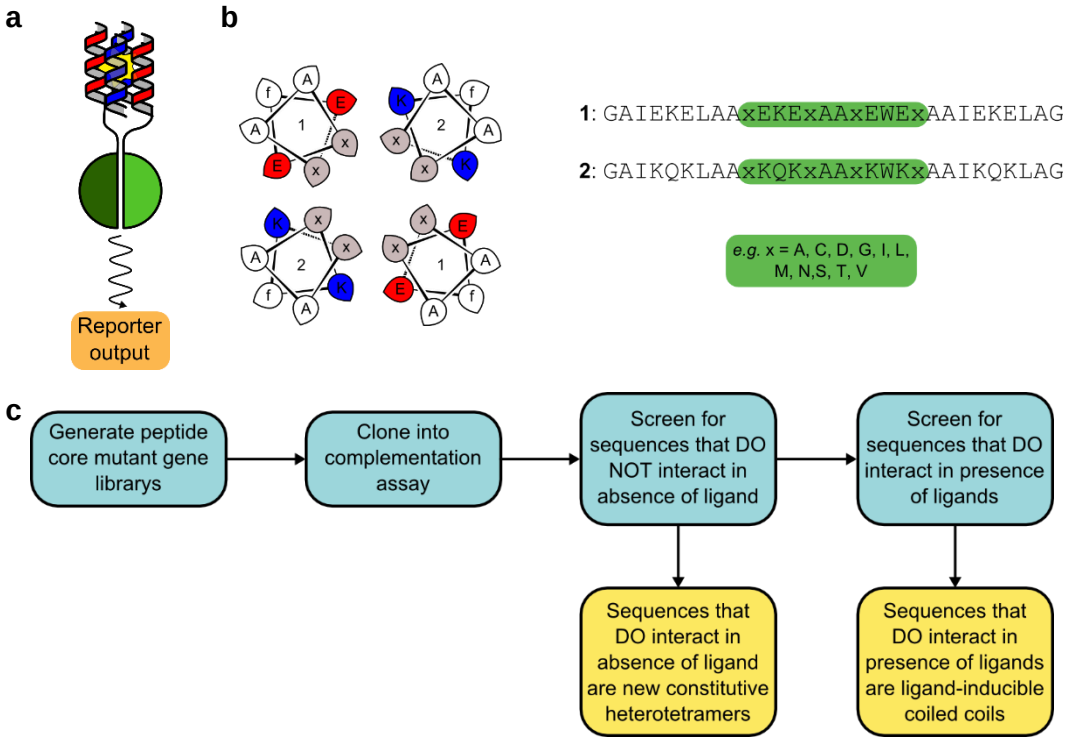


Figure 6-11 Design strategy for heterotetrameric core mutants that bind small molecules. (a) Schematic of ligand-binding heterotetrameric coiled coil in a reporter system. The ligand (yellow) promotes the interaction between the coiled-coil components (red and blue) leading to complementation of the components of the reporter system (green). Reporter outputs could include fluorescence, enzymatic activity, transcriptional regulation. (b) Helical wheel (left) and sequences of potential peptide core mutant library. Amino acids that may be placed at the selected core positions would be chosen for chemical diversity. (c) Schematic outlining *in vivo* selection steps and possible outcomes.

6-12). These systems would reduce the number of components that need to come together to give a productive interaction. They would also allow cavities with greater chemical diversity to be generated because the sequences of the individual helices, particularly the residues lining the cavity, can be changed independently. Furthermore, parallel and antiparallel conformations can be explored relatively simply and reliably because the peptides are constrained in one conformation by the linkers. Finally, the HLH dimers may allow for the design of components with prearranged binding sites where the monomers are folded prior to ligand binding, and ligand binding merely bridges the two components, facilitating the interaction.

These alternative 4-helix bundle folds represent novel design targets that must first be achieved before more-complex properties can be considered. However, once proteins that adopt the desired structures have been designed and characterised,

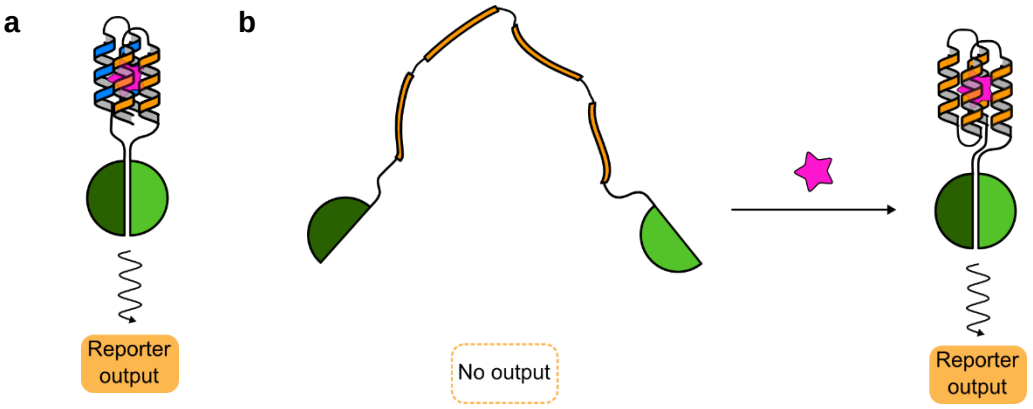


Figure 6-12 Ligand induced 4-helix bundles. (a) A dimer of helix-loop-helix monomers. The interaction between the helical monomers (blue and yellow) is mediated by a small molecule (pink). (b) A single-chain 4-helix bundle (yellow) that exists as an intrinsically disordered protein in the absence of the ligand and is induced to fold by small molecule binding. In both (a) and (b), small molecule binding could be selected for *in vivo* using various protein complementation assays.

they should also be amenable to high throughput *in vivo* screening for small molecule-induced folding.

6.4 Chapter conclusions

Here, very preliminary steps have been made towards the design of novel ABCD heterotetrameric coiled coils and 4-helix bundles that undergo ligand-induced folding.

Firstly, attempts have been made to design ABCD heterotetramers using a computational method that optimised charge interactions. While the peptides that were generated did appear to show some interaction, they did not form the desired assemblies. The method could be improved by better considering the off-target charge interactions that could occur. The search space could also be increased by screening peptides that can contain any combination of all four heptads, rather than containing all four heptads in a different order. This would generate 256 initial peptides, rather than 24, and would result in over 4 billion possible tetramer combinations. While this would significantly increase the computational power required to perform the screening, it may increase the likelihood of finding a successful design.

Attempts were also made to design coiled coils that bind small-molecule ligands through grafting a naturally occurring binding site from a 4-helix bundle onto a

homotetrameric coiled-coil sequence. However, the structures of the chosen amantadine-binding 4-helix bundle and the *de novo* coiled coil were not compatible, and the resulting peptides were excessively destabilised by the introduced binding motif.

Subsequently, steps were made towards the design of coiled coils that could potentially undergo ligand-induced folding in a manner analogous to intrinsically disordered proteins or proteins that undergo chemical rescue of structure. These steps included demonstrating that coiled coils can be globally destabilised through core mutations and that these destabilised coiled coils may be amenable to high throughput *in vivo* screening experiments. However, for future experiments it is expected that other 4-helix bundle structures will need to be explored such as dimers of HLH motifs or single-chain designs.

Rationally designing ligand binding sites in proteins is by no means a simple task. Various computational methods have previously been used to introduce small molecule binding sites into proteins [428-431]. However, *in vivo* selection and directed evolution has also provided a powerful means of introducing novel ligand binding properties into proteins. Indeed, work by Hecht and co-workers has shown that small molecules bind readily to binary patterned 4-helix bundles, both with and without cavities [211,212]. While these experiments demonstrate the suitability and versatility of 4-helix bundles as ligand-binding proteins, they do not consider the effects of ligand binding on protein stability nor whether small molecule binding can elicit 4-helix bundle folding.

If achieved, ligand-inducible 4-helix bundles would be highly useful in the design of novel transcription factors that respond to a small molecule of interest. This type of protein-protein interaction would also be particularly useful in the design of biosensors, where binding of an environmental small molecule leads to a conformational change in the coiled coil, which in turn leads to a readily detectable output [432]. Implementation of a high-throughput method for screening a large library of 4-helix bundles for ligand-induced folding would provide a means of rapidly and flexibly identifying proteins that respond to novel ligands.

Finally, there are other approaches for post-translationally modulating protein structure and activity that are not explored here, including light-mediated conformational changes and chemical modifications. Light-induced conformational changes have been introduced into individual peptides and coiled coils through the

inclusion of residues and modifications that undergo photoinduced isomerisation [433-437]. However, such designs cannot be readily expressed *in vivo* and must instead be introduced into cells, for example, *via* transfection [438]. This limits their use as *in vivo* switchable protein-protein interaction systems.

Post-translational chemical modifications include reduction and oxidation of cysteines and methionines. For example, a peptide has been designed that undergoes a conformational change from monomeric helical hairpin to homodimeric coiled coil on reduction of an inter helical disulphide bond [439]. Additionally, homodimers that undergo dissociation [440] and a peptide that undergoes an α-β conformational switch on methionine oxidation [441] have also been described. A key post-translational chemical modification, however, is phosphorylation of serine and threonine residues. Coiled coils that respond to enzymatic phosphorylation and dephosphorylation have been designed and demonstrated *in vitro*. Depending on the location of the phosphorylatable residue [442], phosphorylation can be both stabilising [443] and destabilising [444-446] to coiled-coil structures, making it a useful means of controlling peptide association. Such switches could be implemented *in vivo* to control the association of larger proteins.

Chapter 7: Conclusions and future work

7.1 Overall conclusions

Through the work described in this thesis, a set of homo- and heterotetrameric coiled coils has been designed and characterised both *in vitro* and *in vivo*. The results of the characterisation demonstrate that the *de novo* coiled coils are useful as artificial protein-protein interaction domains for synthetic biology applications inside cells.

Firstly, a number of homotetramers have been designed by varying the arrangement of charged glutamate and lysine residues in peptides with Leu/Ile cores (*i.e. a*=leucine and *d*=isoleucine). The resulting seven homotetramers have different thermal stabilities, however this is not dictated by the locations of the charged residues as initially proposed. Furthermore, changing the identity of the *d* positions to valines, which was predicted to give less-stable homotetramers, in fact leads to a loss of structural specificity and the Leu/Val core peptides do not form tetramers within the investigated concentration range. Therefore, this approach cannot be used to create additional homotetramers with a wider range of thermal stabilities.

Using design principles derived from the homotetramers, sets of acidic (A) and basic (B) peptides have also been designed. When these peptides are mixed in A/B pairs they assemble into A_2B_2 heterotetramers. The thermal stabilities of these heterotetramers depend on the core residues in the constituent peptides: heterotetramers with Leu/Ile cores are the most stable; those with Leu/Val cores are the least stable; and those with mixed Leu/Ile/Val cores have intermediate stabilities. Therefore, unlike the homotetramers, the identities of the core residues can be used to modulate the stabilities of the heterotetramers without adversely affecting their oligomeric states.

A selection of the *in vitro*-characterised heterotetramers have also been investigated in *E. coli* to determine their suitability as PIDs. All three of the investigated heterotetramers interact within cells when studied using a

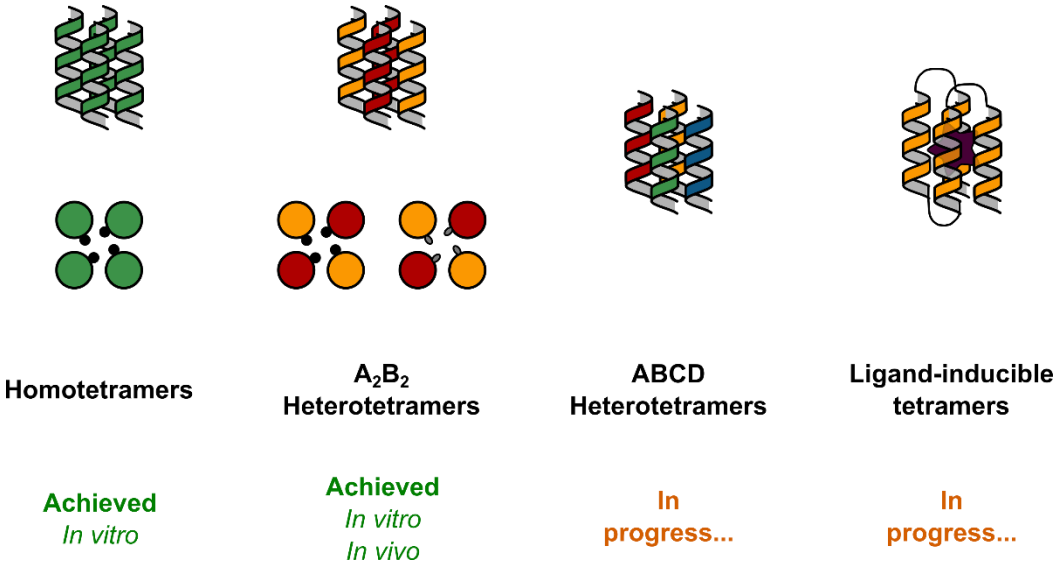


Figure 7-1 Graphical summary. Homotetramers have been designed and characterised *in vitro*. A_2B_2 heterotetramers have been designed and characterised both *in vitro* and *in vivo*. Their stability can be tuned by varying the core residues. Preliminary steps have been made towards the design of ABCD heterotetramers and ligand-inducible coiled coils.

transcription repression assay based on oligomerisation mutants of the Lac repressor. Some of the heterotetramer complexes are also able to form when the basic peptides are fused to a small solubility tag, or when they are expressed as "minimal" coiled-coil peptides fused only to a six-residue identification tag.

Furthermore, when the basic components are expressed from an inducible promoter, their expression level can be modulated by altering the inducer concentration. The expression level of the basic component determines the resulting level of transcription repression. Thus, the level of repression can be tuned over a reasonable range. Therefore, not only does this work yield a number of semi-artificial transcriptional repressors, consisting of a designed PID and an engineered DBD, it also identifies further regulatory components that these repressors are compatible with.

Preliminary work has also been carried out to investigate what effect the coiled-coil peptides have on the cellular concentrations of the proteins they were fused to. Generally, the peptides lead to decreased levels of the fusion proteins. This most likely occurs through increased degradation of the protein due to the unfolded nature of the peptides when not interacting with their cognate partner.

Additional coiled-coil tetramer designs have been pursued *in vitro*. Firstly, the effects of placing alanine at *g* positions have been investigated. Expanding the

hydrophobic seam of a coiled coil in this way has previously led to the formation of parallel [206] and antiparallel [205] tetramers and higher order α-barrels [213]. In the work described here, the peptides form tetramers, but the orientation of the helices remains to be determined. ABCD heterotetramers, where all four peptides are different, have also been explored. The ABCD combinations do show heteromeric interactions, but extensive cross-interactions are also observed between the constituent peptides. Moreover, the designs do not form tetramers.

Finally, a number of core mutations have been made in existing homo- and heterotetramers in an attempt to confer ligand-binding behaviour on these coiled coils. While none of the peptides bind the tested small molecules, the globally-destabilised heterotetramer mutant may provide a starting point for high-throughput selection studies to identify coiled coils that undergo ligand-induced folding or ligand-induced stabilisation. Such coiled coils would be useful as inducible protein-protein interaction domains in the design of ATFs.

7.2 Future work

7.2.1 Improved coiled coil designs

As well as the tetramers described above, there are a number of other designs that would make useful additions to the suite of coiled coil-based artificial PIDs. For example, charge neutral A_2B_2 heterotetramers could be designed to consist of peptides with a net neutral charge instead of highly acidic and basic peptides. As discussed in Chapter 5, these may have improved *in vivo* properties due to reduced non-specific interactions with other highly charged cellular components, such as negatively charged ribosomes. Designs for such coiled coils may be readily achieved using a modified version of the script used to generate the ABCD heterotetramer designs described in Chapter 6.

Furthermore, and also as discussed in Chapter 6, ligand-binding or ligand-inducible coiled coils present a highly desirable target. High-throughput screening assays, such as protein complementation assays or proteolysis-susceptibility assays, could be used to select coiled coils that do not fold or are rapidly degraded in the absence of ligand but become folded or stabilised in the presence of a small molecule. Similar screening strategies could also be used to identify constitutive coiled-coil interactions with advantageous *in vivo* properties such as reduced off-target interactions or reduced susceptibility to proteases. Additionally, this method

could be used to select coiled coil-encoding DNA sequences with beneficial properties such as efficient transcription. This type of approach has previously been used to select constitutively-interacting heterodimeric coiled-coil peptides from two libraries of semi-randomised peptides [209].

Coiled coils that use different switching methods would also be useful in the design of inducible protein-protein interactions. For example, coiled coils designed to change conformation in response to reversible chemical modifications would provide a means of rapidly altering the association state of proteins fused to said coiled coil. Phosphorylation is a particularly appealing post-translational modification for controlling assembly in this way because phosphorylation and dephosphorylation are carried out by enzymes, so both could be readily performed *in vivo*. Furthermore, phosphorylation has previously been used to control both the assembly [444] and disassembly [443] of coiled coils. Alternatively, peptides could be designed to perform strand displacement or strand exchange. For example, displacement has been demonstrated in heterodimers consisting of peptides of different lengths [447,448]. However, *in vivo* this type of switching would need to be controlled at the transcriptional level and is therefore likely to occur on a slower time scale than chemical modifications or small-molecule binding.

7.2.2 Fully-artificial transcription factors

The focus of this thesis has been the design of new protein-protein interactions and these interactions have been tested using a natural DNA-binding protein. Therefore, all of the semi-artificial transcription factors that have been generated recognise a natural promoter sequence. This could lead to issues with cross talk. Improved ATFs could be designed by combining the *de novo* coiled-coil domains with engineered TALEs as the DBDs. In this way, orthogonal and fully-artificial transcription factor/promoter pairs could be designed to minimise off-target interactions in the host organism. Alternatively, the DBDs could be designed to target the ATF to any promoter of interest in an organism's genome.

Furthermore, this thesis has only focused on using protein-protein interactions in ATFs. However, there are many other scenarios where both constitutive and inducible interaction domains would be of use. For example, constitutive interactions could be used to co-localise enzymes of an enzymatic pathway in order to increase productivity [449]. Inducible interactions could be used in synthetic cell-surface receptors to transmit signals to the cell interior.

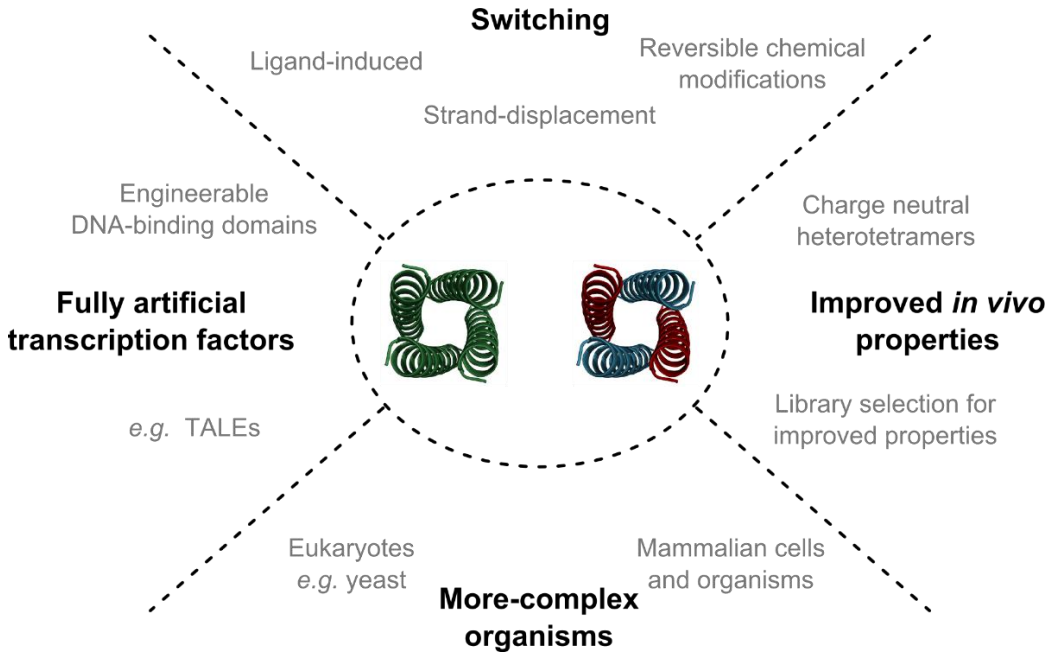


Figure 7-2 Future work. Avenues remaining to be explored include incorporating switching behaviours into designed protein-protein interaction domains; designing fully-artificial transcription factors with engineerable DNA-binding domains; improving the *in vivo* properties of the components; and using the components in organisms beyond *E. coli*.

7.2.3 More-complex organisms

As more designs are added to the protein-protein interaction, DNA-binding domain and ATF toolkit, increasingly sophisticated characterisation techniques will be needed to ensure the new components are suitable for use in organisms. As discussed in Chapter 5, these techniques may include quantitative Western blotting, pull-down or immunoprecipitation assays and proteomics. Such techniques would provide information on whether a component is present in a cell in reasonable quantities, whether it has undesirable cross interactions with other cellular components, and whether it has global effects on the cell other than those intended. These properties are likely to be organism-specific, so the characterisation may need to be performed anew as components are taken out of *E. coli* and introduced into increasingly complex organisms such as mammals. This may be particularly desirable given the interest in ATFs for treating diseases involving aberrant gene expression, most notably cancer.

There is much left to explore, but what is already apparent is the broad utility and versatility of ATFs and their constituent parts. In the future, the same components

used as parts of a synthetic genetic circuit in a bacterium may also find use as a therapeutic agent in humans, allowing these modest components to span all domains of life.

Chapter 8: Appendix

8.1 Peptide sequences

Sequences, registers, heptads and masses (Da) of all peptides discussed in this thesis are displayed in Table 8-1. All peptides are in either *c*-register or *g*-register. All peptides are N-terminally acetylated and C-terminally amidated. Where present, heptads are given as consensus motifs.

Peptides **6-20** and **26-45** are named for the locations of their charged Glu/Lys residues (1, *e*/*g*; 2, *c*/*e*; 3, *g*/*b*; 4, *b*/*c*) and the identities of their core residues (LI, *a*=Leu/*d*=Ile; LV, *a*=Leu/*d*=Val; LIA, *a*=Leu/*d*=Ile/*g*=Ala; LVA, *a*=Leu/*d*=Val/*g*=Ala). The designed homomeric peptides are further labelled for the order of Glu and Lys residues in their linear sequence (EK; KE) and peptide components of heteromeric assemblies are labelled as acidic (A, contain Glu at all charged positions) or basic (B, contain Lys at all charged positions).

Other peptides are named for their consensus heptad motif *e.g.* ELAEIK or using a historical name suffixed with identifiers that denote additional mutations to the parent sequence *e.g.* CC-Tet-KE is a version of CC-Tet [208] in which the Glu and Lys residues at *e*/*g* positions have been swapped.

CC-Tet was synthesised by Will Dawson [208].

No.	Peptide	Sequence	Register	Heptad (*gabcde*)	Mass (Da)
1	CC-Tet	Ac – G E LAAIKQE LAAIKKE LAAIKWE LAAIKQ GAG – NH_2	*g*	ELAAIK	3374.993
2	CC-Tet-KE	Ac – G K LAAIEQK LAAIEKK LAAIEWK LAAIEQ GAG – NH_2	*g*	KLAAIE	3374.993
3	C-Tet-KE-N-4	Ac – G K LAAIEQK LAAIEKK LAANNWK LAAIEQ GAG – NH_2	*g*	KLAAIE	3360.926
4	CC-Tet-KE-N-4.5	Ac – G K LAAIEQK LAAIEKK LAANNWK LAAIEQ KLAA G – NH_2	*g*	KLAAIE	3616.286
5	CC-Tet-IA	Ac – G E LAAIKQE LAAAKKE LAAAKWE LAAIKQ GAG – NH_2	*g*	ELAAIK	3290.832
6	1-LI-EK	Ac – G AIKKE LAAIKKE LAAIKWE LAAIKKE LA G – NH_2	*c*	ELAAIK	3246.950
7	1-LI-KE	Ac – G AIEQK LAAIEQK LAAIEWK LAAIEQK LA G – NH_2	*c*	KLAAIE	3246.820
8	2-LI-EK	Ac – G EIKQQ LAEIKQQ LAEIKWQ LAEIKQQ LA G – NH_2	*c*	QLAEIK	3475.027
9	2-LI-KE	Ac – G KIEQQ LAKIEQQ LAKIEWQ LAKIEQQ LA G – NH_2	*c*	QLAKIE	3475.027
10	3-LI-EK	Ac – G AIQQE LKAIQQE LKAIQWE LKAIQQE LK G – NH_2	*c*	ELKAIE	3475.027
11	3-LI-KE	Ac – G AIQQK LEAIQQK LEAIQWK LEAIQQK LE G – NH_2	*c*	KLEAIQ	3475.027
12	4-LI-EK	Ac – G KIQKQ LEKIQKQ LEKIQWQ LEKIQKQ LE G – NH_2	*c*	QLEKIQ	3703.364
13	4-LI-KE	Ac – G EIQKQ LKEIQKQ LKEIQWQ LKEIQKQ LK G – NH_2	*c*	QLKEIQ	3703.364
14	1-LV-EK	Ac – G AVKQE LAAVKQE LAAVKWE LAAVKQE LA G – NH_2	*c*	ELAAVK	3190.712
15	1-LV-KE	Ac – G AVEQK LAAVEQK LAAVEWK LAAVEQK LA G – NH_2	*c*	KLAAVE	3190.712
16	2-LV-EK	Ac – G EVKQQ LAEVKQQ LAEVKWQ LAEVKQQ LA G – NH_2	*c*	QLAEVK	3418.919
17	2-LV-KE	Ac – G KVEQQ LAKVEQQ LAKVEWQ LAKVEQQ LA G – NH_2	*c*	QLAKVE	3418.919
18	3-LV-EK	Ac – G AVQQE LKAVQQE LKAVQWE LKAVQQE LK G – NH_2	*c*	ELKAVQ	3418.919
19	3-LV-KE	Ac – G AVQQK LEAVQQK LEAVQWK LEAVQQK LE G – NH_2	*c*	KLEAVQ	3418.919
20	2-LIA-EK	Ac – G EIKQA LAEIKQA LAEIKWA LAEIKQA LA G – NH_2	*c*	ALAEIK	3246.820

No.	Peptide	Sequence	Register	Heptad (*gabcde*)	Mass (Da)
21	2-LIA-KE	Ac – G KIEQA LAKIEQA LAKIEWA LAKIEQA LA G – NH_2	*c*	ALAKIE	3246.820
22	ELAEIK	Ac – G EIKKE LAEIKKE LAEIKWE LAEIKKE LA G – NH_2	*c*	ELAEIK	3479.096
23	ELAEIK-M2-*a*	Ac – G EIKKE VAEASKE AAEHKWE LAEIKKE LA G – NH_2	*c*	-	3363.793
24	ELAEIK-M2-*d*	Ac – G EIKKE LAEVKKA SAEAKWH LAEIKKE LA G – NH_2	*c*	-	3346.896
25	ELAEIK-M2-*a*-HI	Ac – G EIKKE VAEASKE AAEIKWE LAEIKKE LA G – NH_2	*c*	-	3339.812
26	1-LI-A	Ac – G AIEKE LAAIEKE LAAIEWE LAAIEKE LA G – NH_2	*c*	ELAAIE	3250.715
27	1-LI-A-*g*	Ac – G E LAAIEKE LAAIEKE LAAIEWE LAAIEK G – NH_2	*g*	ELAAIE	3250.715
28	1-LI-A*	Ac – G AIEKE LAAIEKE AAAAEWE LAAIEKE LA G – NH_2	*c*	ELAAIE	3166.554
29	2-LI-A	Ac – G EIEKQ LAEIEKQ LAEIEWQ LAEIEKQ LA G – NH_2	*c*	QLAEIE	3478.922
30	3-LI-A	Ac – G AIQKE LEAIQKE LEAIQWE LEAIQKE LE G – NH_2	*c*	ELEAIQ	3478.922
31	1-LV-A	Ac – G AVEKE LAAVEKE LAAVEWE LAAVEKE LA G – NH_2	*c*	ELAAVE	3194.607
32	2-LV-A	Ac – G EVEKQ LAEVEKQ LAEVEWQ LAEVEKQ LA G – NH_2	*c*	QLAEVE	3422.814
33	3-LV-A	Ac – G AVQKE LEAVQKE LEAVQWE LEAVQKE LE G – NH_2	*c*	ELEAVQ	3422.814
34	1-LI-B	Ac – G AIKQK LAAIKQK LAAIKWK LAAIKQK LA G – NH_2	*c*	KLAAIK	3243.054
35	1-LI-B-*g*	Ac – G K LAAIKQK LAAIKQK LAAIKWK LAAIKQ G – NH_2	*g*	KLAAIK	3243.054
36	1-LI-B*	Ac – G AIKQK LAAIKQK AAAAKWK LAAIKQK LA G – NH_2	*c*	KLAAIK	3158.893
37	2-LI-B	Ac – G KIKQQ LAKIKQQ LAKIKWQ LAKIKQQ LA G – NH_2	*c*	QLAKIK	3471.261
38	3-LI-B	Ac – G AIQQK LKAIQQK LKAIQWK LKAIQQK LK G – NH_2	*c*	KLKAIQ	3471.261
39	1-LV-B	Ac – G AVKQK LAAVKQK LAAVKWK LAAVKQK LA G – NH_2	*c*	KLAAVK	3186.946
40	2-LV-B	Ac – G KVKQQ LAKVKQQ LAKVKWQ LAKVKQQ LA G – NH_2	*c*	QLAKVK	3415.153

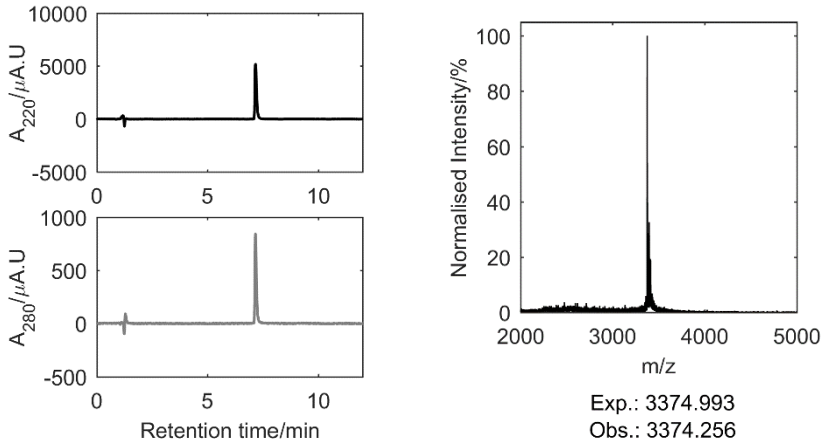
No.	Peptide	Sequence	Register	Heptad (*gabcde*)	Mass (Da)
41	3-LV-B	Ac – G AVQQK LKAVQQK LKAVQWK LKAVQQK LK G – NH_2	*c*	KLKAVQ	3415.153
42	2-LIA-A	Ac – G EIEKA LAEIEKA LAEIEWA LAEIEKA LA G – NH_2	*c*	ALAEIE	3250.715
43	2-LVA-A	Ac – G EVEKA LAEVEKA LAEVEWA LAEVEKA LA G – NH_2	*c*	ALAEVE	3194.607
44	2-LIA-B	Ac – G KIKQA LAKIKQA LAKIKWA LAKIKQA LA G – NH_2	*c*	ALAKIK	3243.054
45	2-LVA-B	Ac – G KVKQA LAKVKQA LAKVKWA LAKVKQA LA G – NH_2	*c*	ALAKVK	3186.946
46	ABCD 1	Ac – G AVEQK LAAIEQE LAAIKWK LAAVKQE LA G – NH_2	*c*	-	3218.766
47	ABCD 2	Ac – G AIEQE LAAVKQE LAAVEWK LAAIKQK LA G – NH_2	*c*	-	3218.766
48	ABCD 3	Ac – G AIKQK LAAVKQE LAAIEWE LAAVEQK LA G – NH_2	*c*	-	3218.766
49	ABCD 4	Ac – G AVEQK LAAIKQK LAAVKWE LAAIEQE LA G – NH_2	*c*	-	3218.766
50	ABCD 5	Ac – G AIKQK LAAVKQE LAAVEQK LAAIEQE LA GGSY – NH_2	*c*	-	3467.989
51	ABCD 6	Ac – G AVEQK LAAIKQK LAAIEQE LAAVKQE LA GGSW – NH_2	*c*	-	3491.026
52	ABCD 7	Ac – G AVKQE LAAIEQE LAAVEQK LAAIKQK LA GGDY – NH_2	*c*	-	3495.999
53	ABCD 8	Ac – G AIEQE LAAVEQK LAAIKQK LAAVKQE LA GGDW – NH_2	*c*	-	3519.036

Table 8-1 Sequences of all designed peptides with systematic names, registers, consensus heptads and masses (Da). All peptides were C-terminally acetylated (Ac) and N-terminally amidated (NH_2). All peptides were designed in either *c*- or *g*-resister. All consensus heptads are written starting at a *g* position, regardless of peptide register.

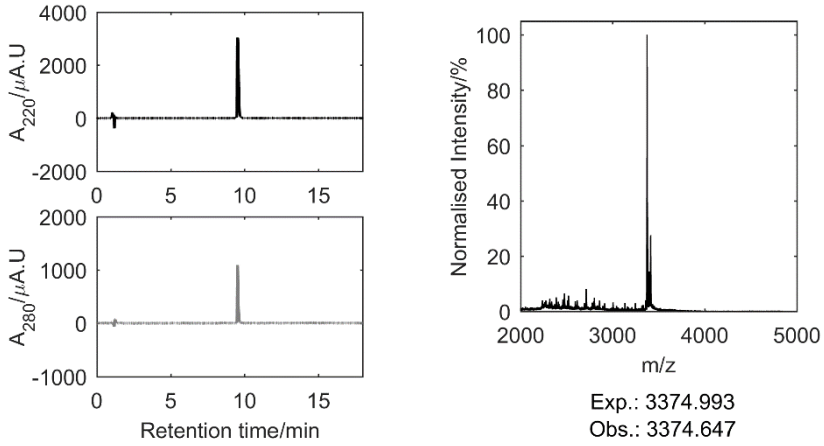
8.2 Analytical HPLC traces and mass spectra of designed peptides

This section contains representative analytical HPLC traces (left) monitoring absorbance at 220 nm (top, black) and 280 nm (bottom, gray) and representative mass spectra (right) for each peptide discussed in this thesis. Exp, expected mass (Da); obs, observed mass (Da); m/z, mass:charge ratio. Peptides are numbered as in Table 8-1.

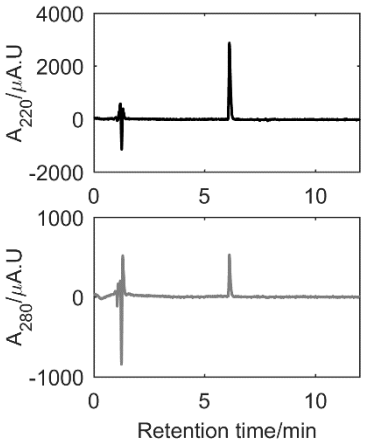
1

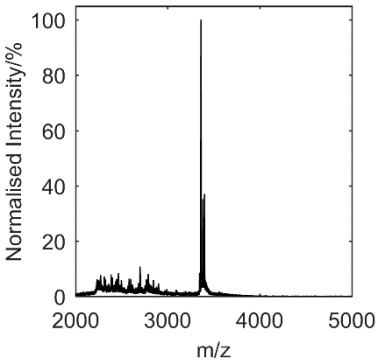


2



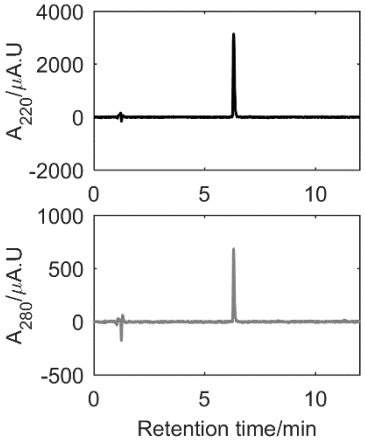
3

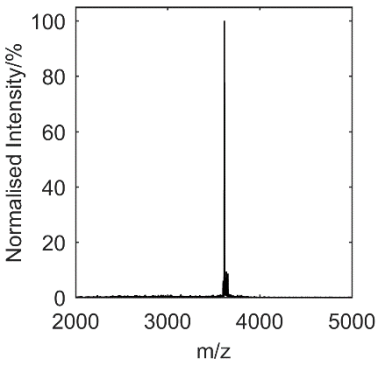




Exp.: 3360.926
Obs.: 3360.767

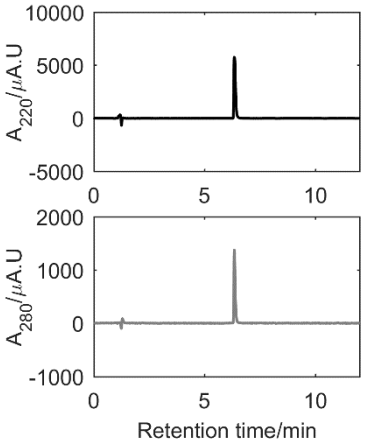
4

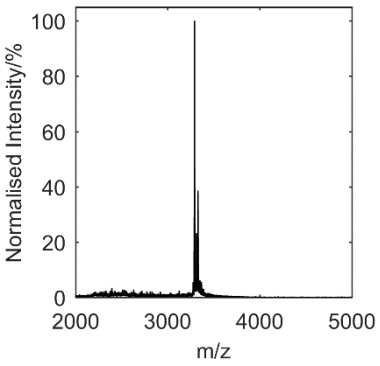




Exp.: 3616.286
Obs.: 3615.244

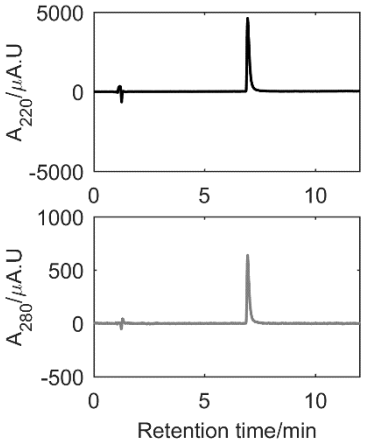
5

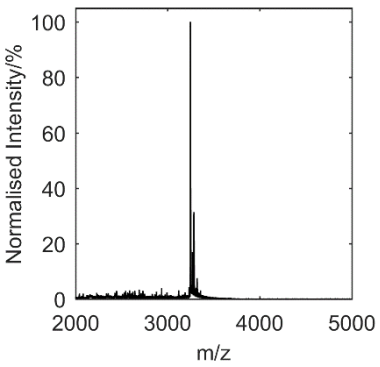




Exp.: 3290.832
Obs.: 3290.403

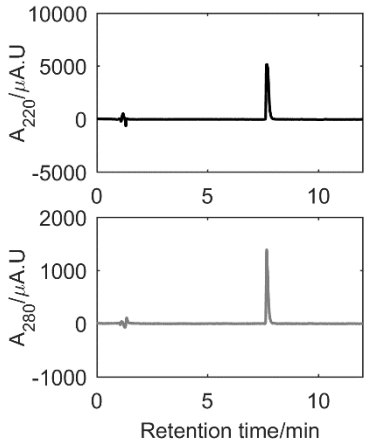
6

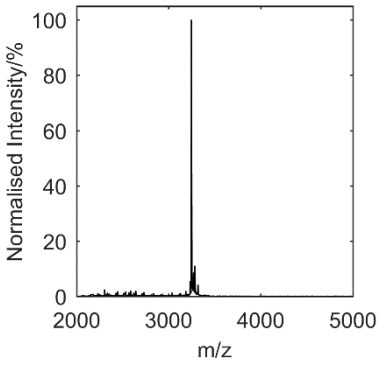




Exp.: 3246.95
Obs.: 3246.927

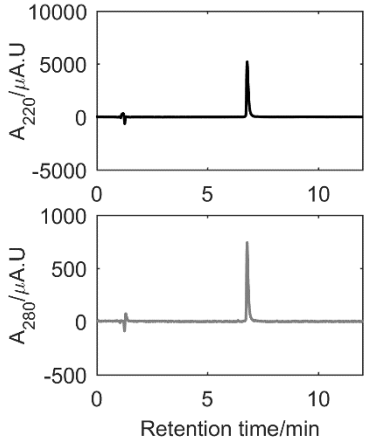
7

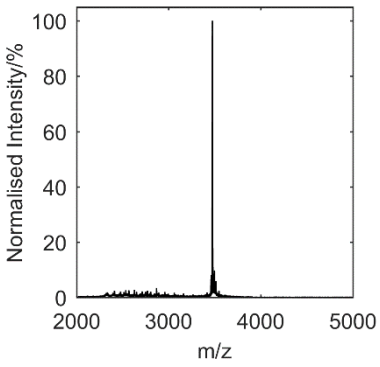




Exp.: 3246.82
Obs.: 3246.6

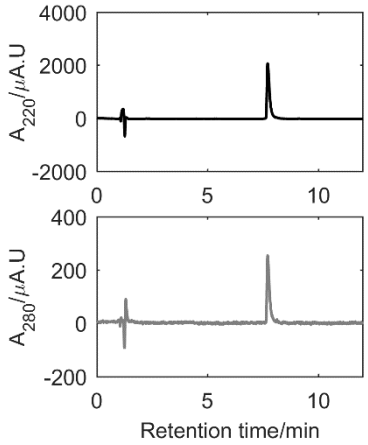
8

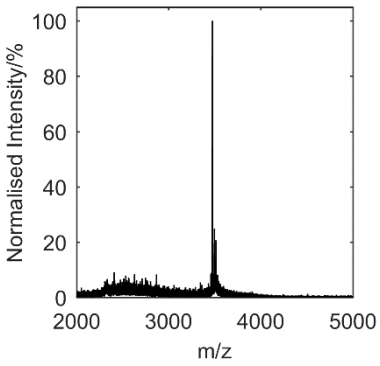




Exp.: 3475.027
Obs.: 3474.337

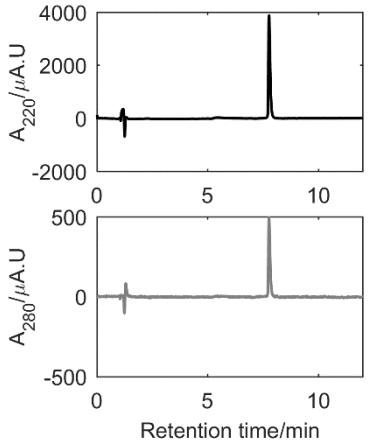
9

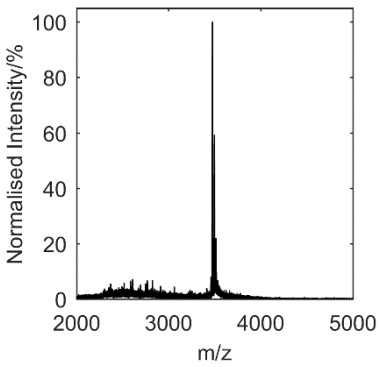




Exp.: 3475.027
Obs.: 3473.661

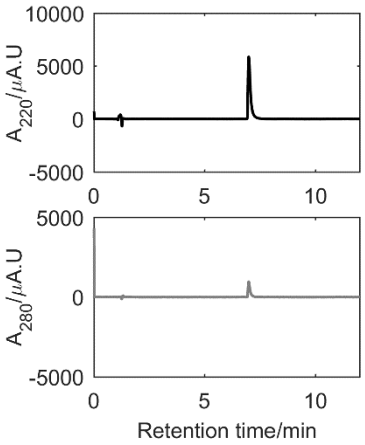
10

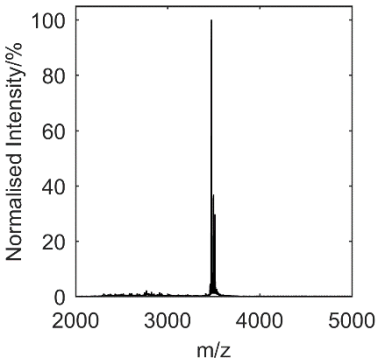




Exp.: 3475.027
Obs.: 3473.581

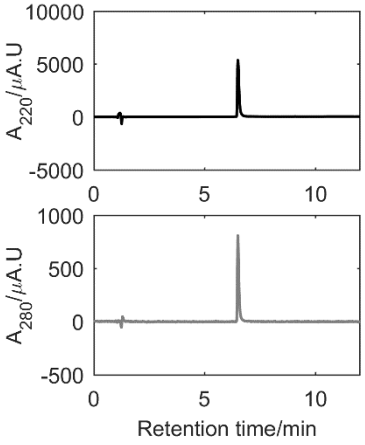
11

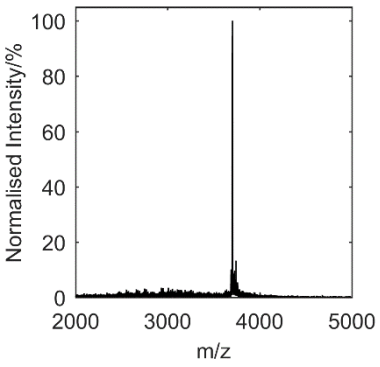




Exp.: 3475.027
Obs.: 3474.177

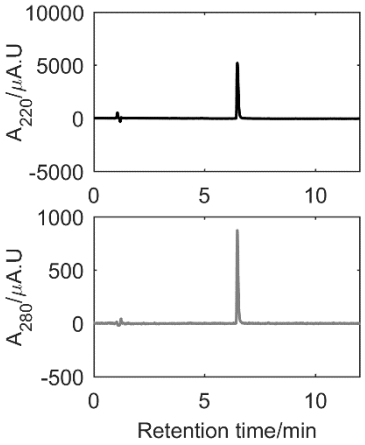
12

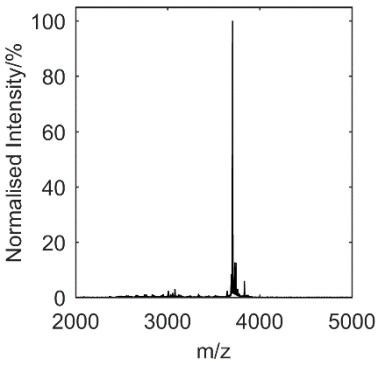




Exp.: 3703.364
Obs.: 3703.117

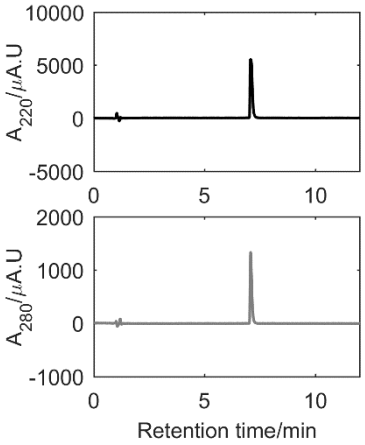
13

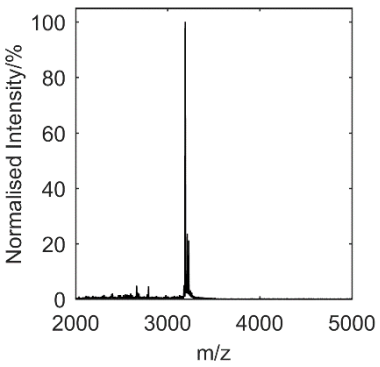




Exp.: 3703.364
Obs.: 3704.233

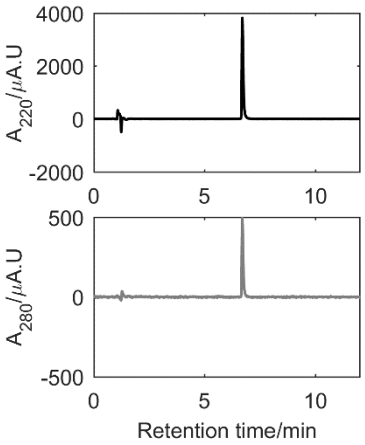
14

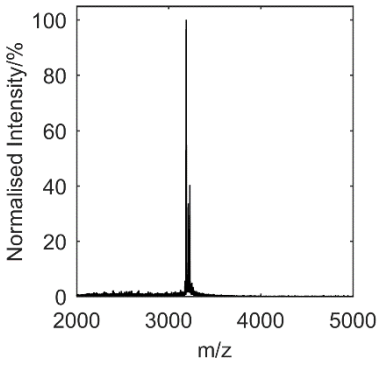




Exp.: 3190.712
Obs.: 3189.986

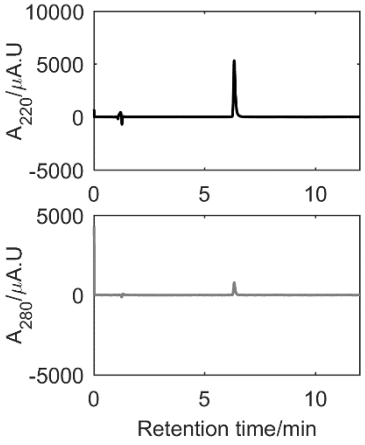
15

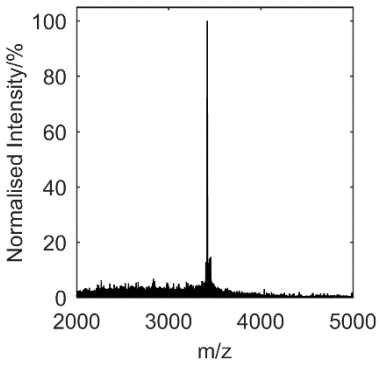




Exp.: 3190.712
Obs.: 3190.026

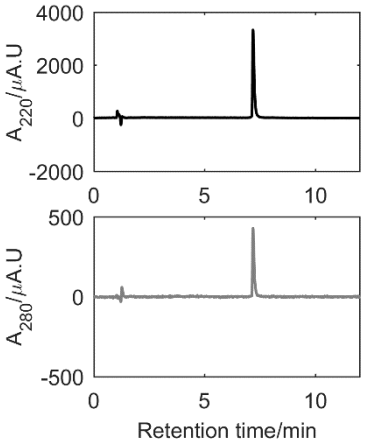
16

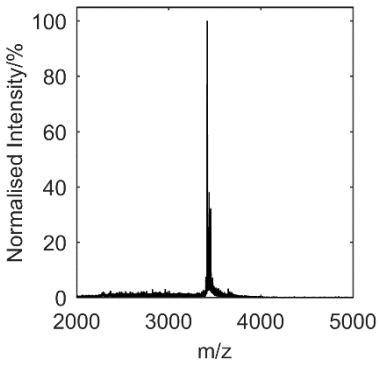




Exp.: 3418.919
Obs.: 3418.584

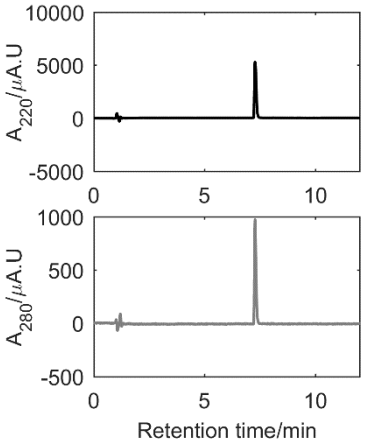
17

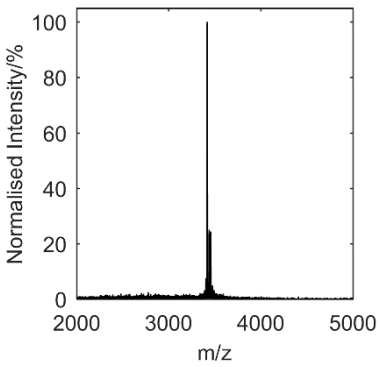




Exp.: 3418.919
Obs.: 3418.444

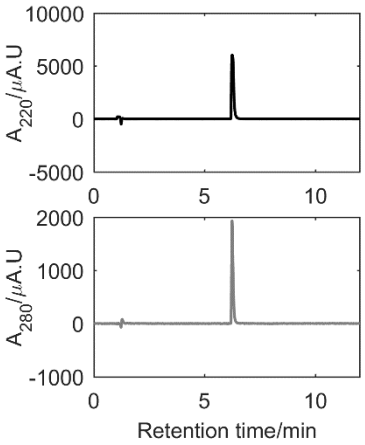
18

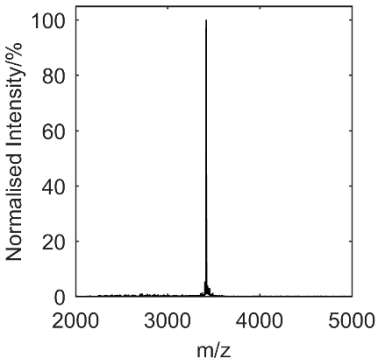




Exp.: 3418.919
Obs.: 3417.807

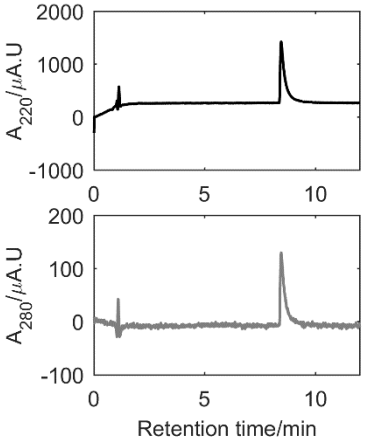
19

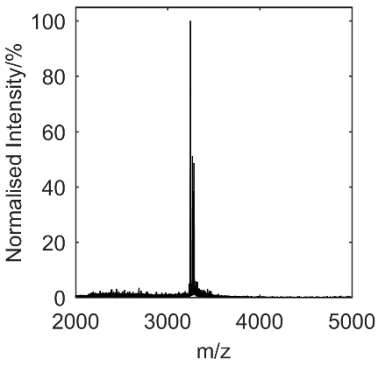




Exp.: 3418.919
Obs.: 3417.849

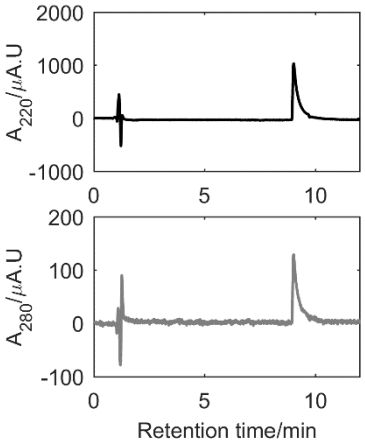
20

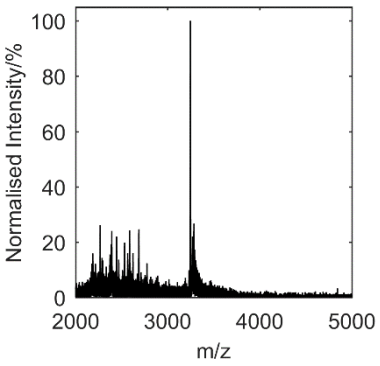




Exp.: 3246.82
Obs.: 3246.716

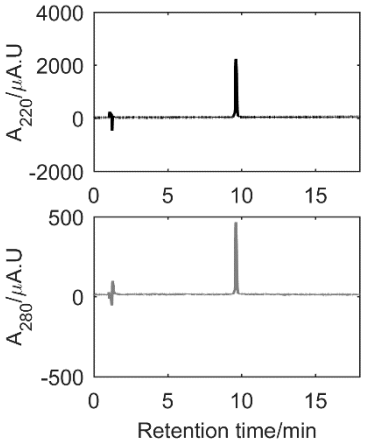
21

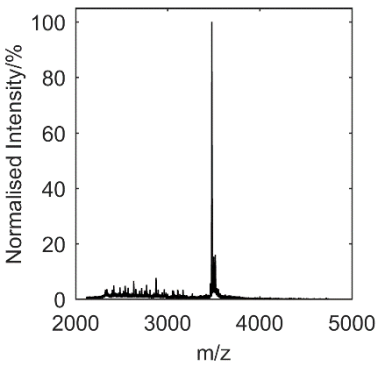




Exp.: 3246.82
Obs.: 3246.953

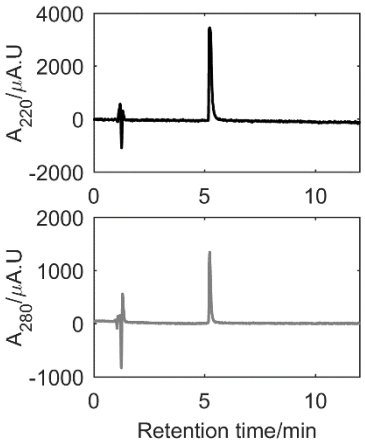
22

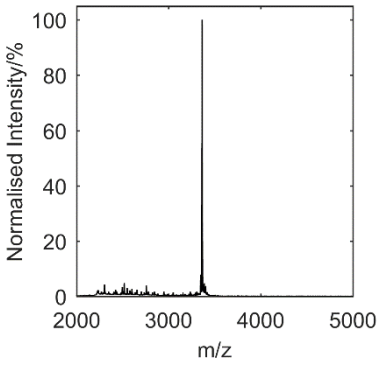




Exp.: 3479.096
Obs.: 3478.727

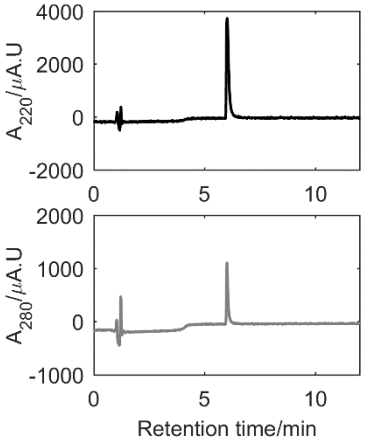
23

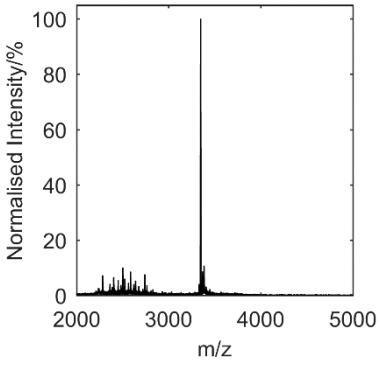




Exp.: 3363.793
Obs.: 3362.376

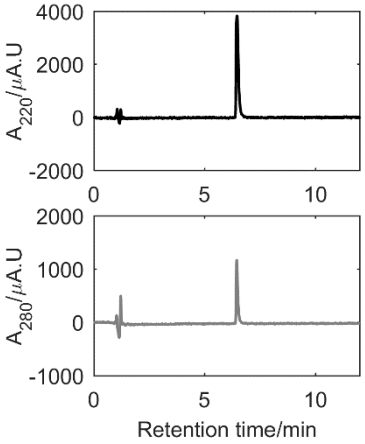
24

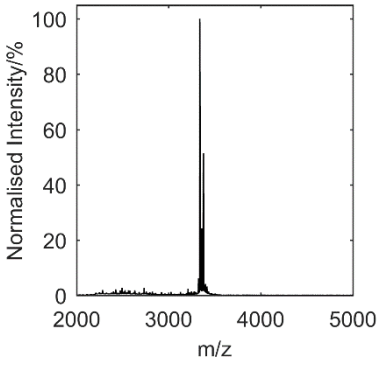




Exp.: 3346.896
Obs.: 3346.196

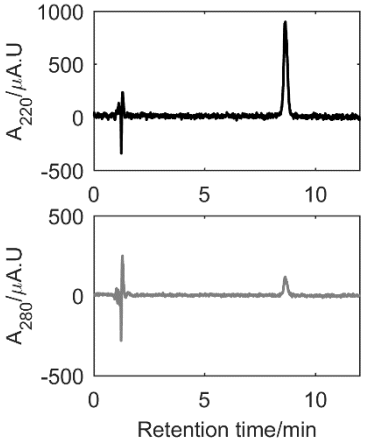
25

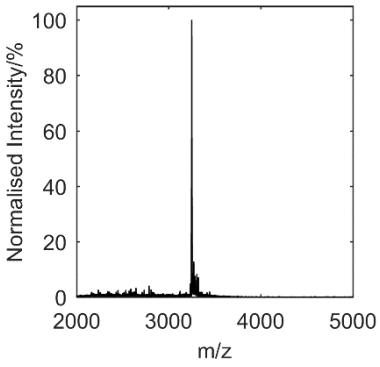




Exp.: 3339.812
Obs.: 3338.249

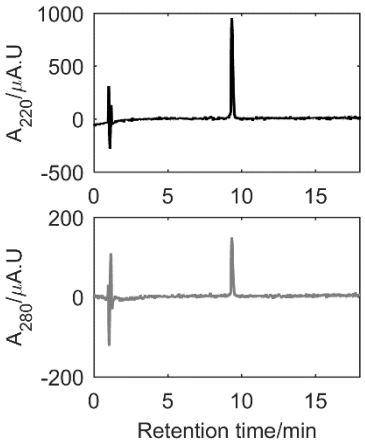
26

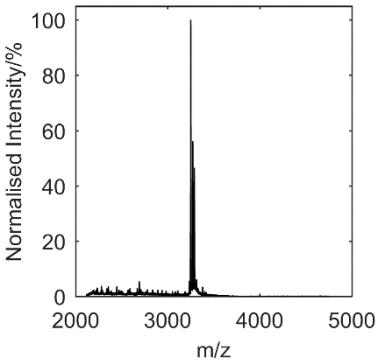




Exp.: 3250.715
Obs.: 3250.383

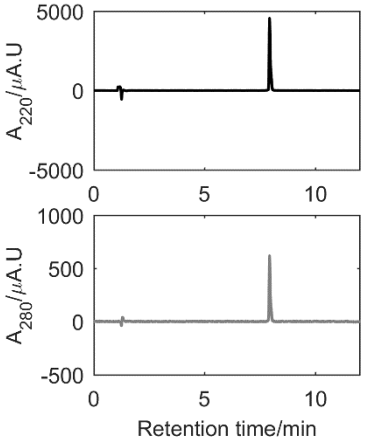
27

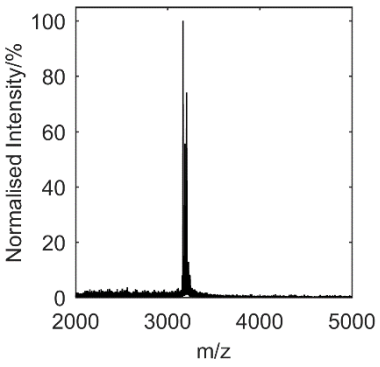




Exp.: 3250.715
Obs.: 3250.487

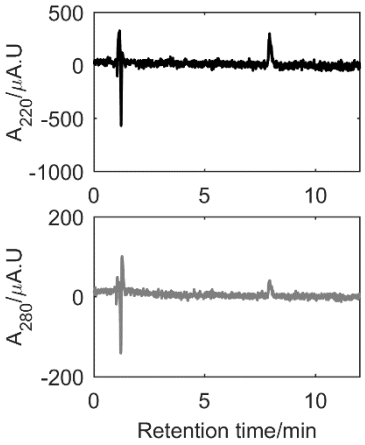
28

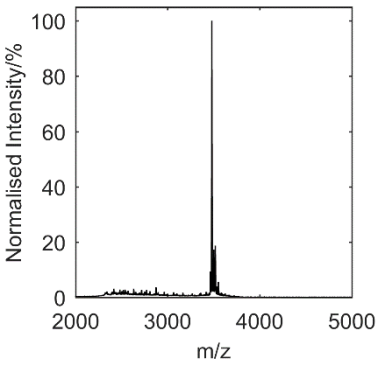




Exp.: 3166.554
Obs.: 3166.092

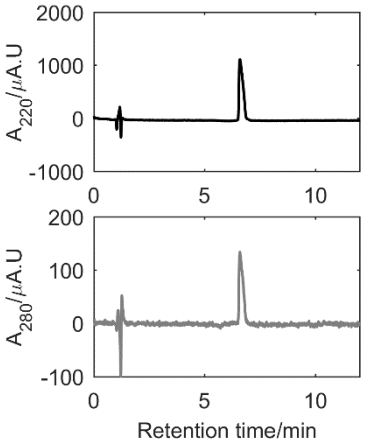
29

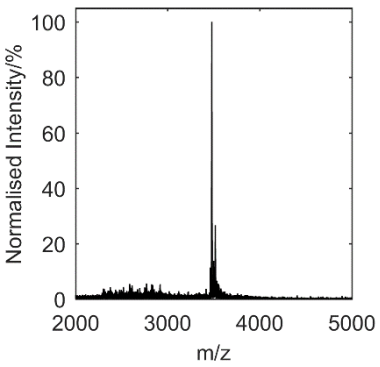




Exp.: 3478.922
Obs.: 3478.749

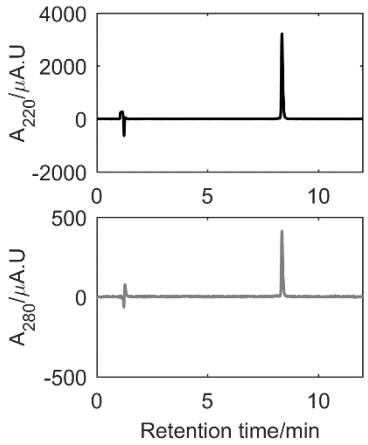
30

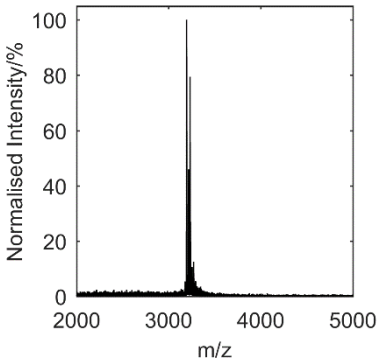




Exp.: 3478.922
Obs.: 3478.833

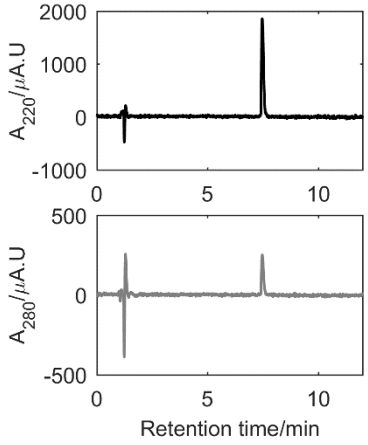
31

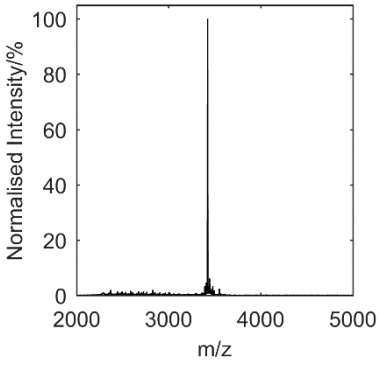




Exp.: 3194.607
Obs.: 3194.032

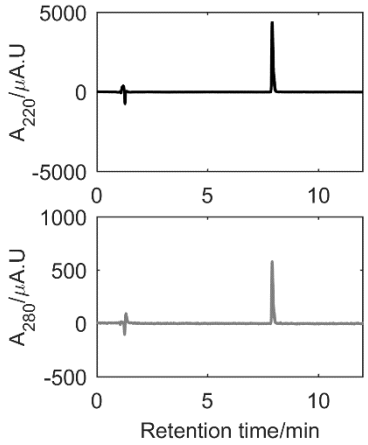
32

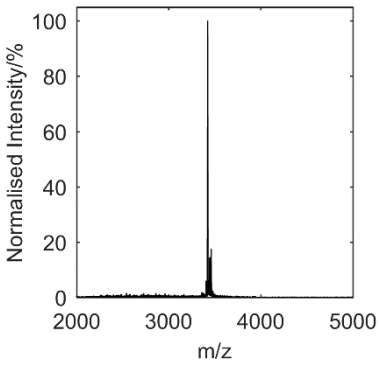




Exp.: 3422.814
Obs.: 3421.673

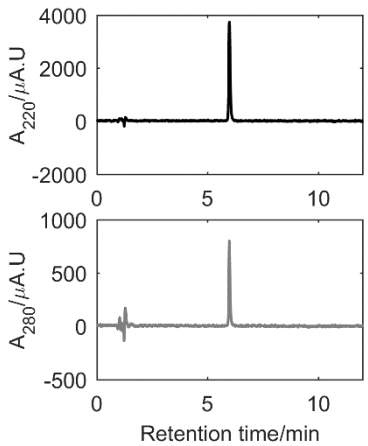
33

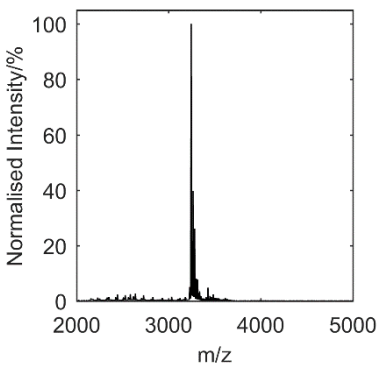




Exp.: 3422.814
Obs.: 3421.919

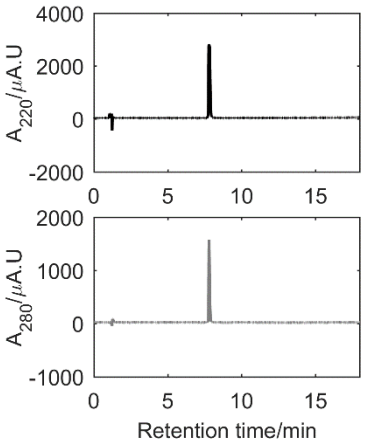
34

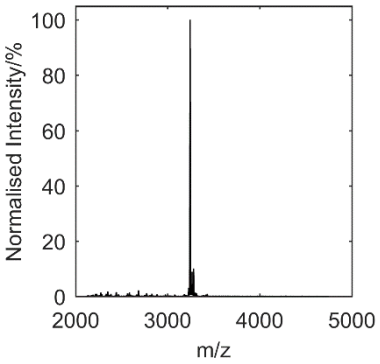




Exp.: 3243.054
Obs.: 3242.61

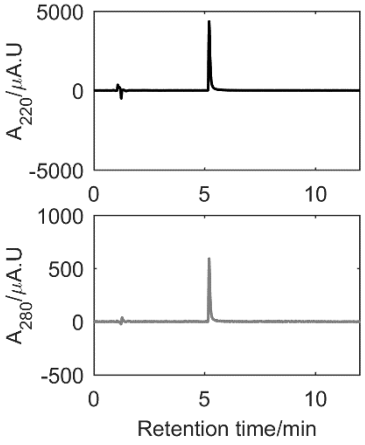
35

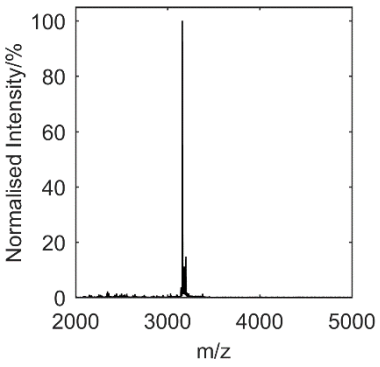




Exp.: 3243.054
Obs.: 3242.683

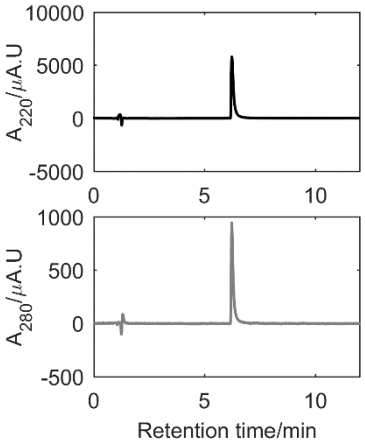
36

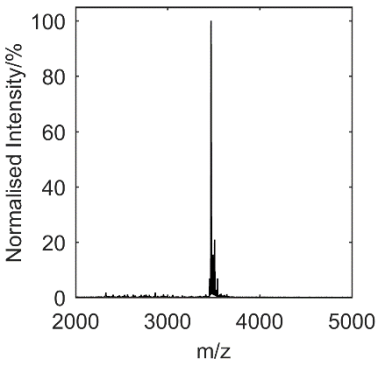




Exp.: 3158.893
Obs.: 3158.422

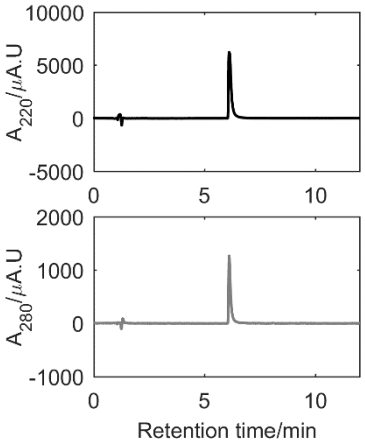
37

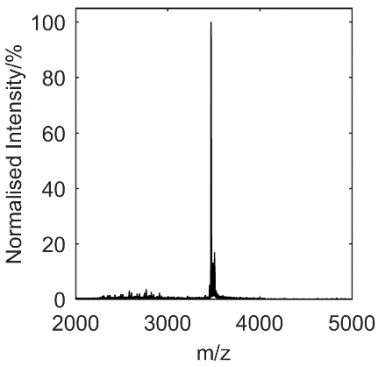




Exp.: 3471.261
Obs.: 3470.835

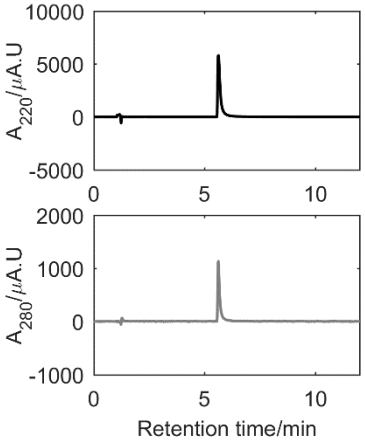
38

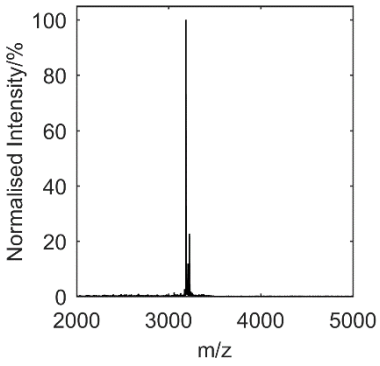




Exp.: 3471.261
Obs.: 3470.942

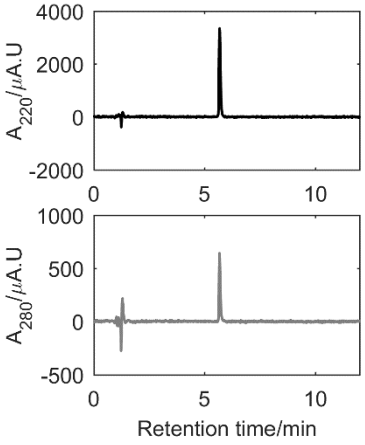
39

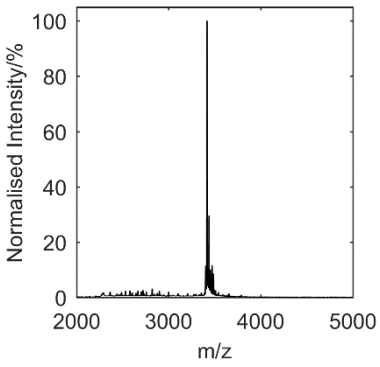




Exp.: 3186.946
Obs.: 3186.39

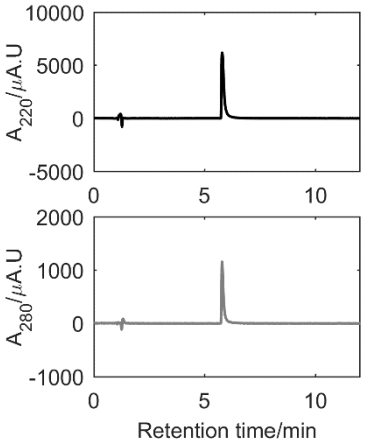
40

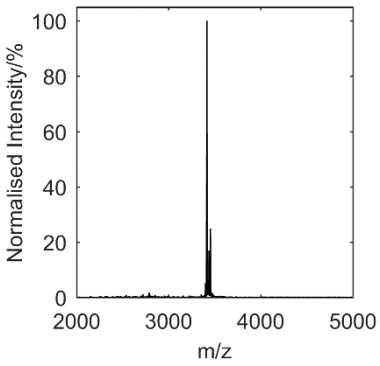




Exp.: 3415.153
Obs.: 3414.788

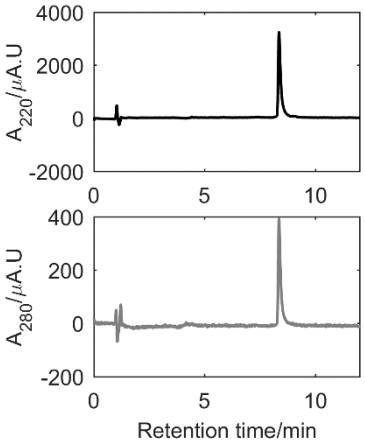
41

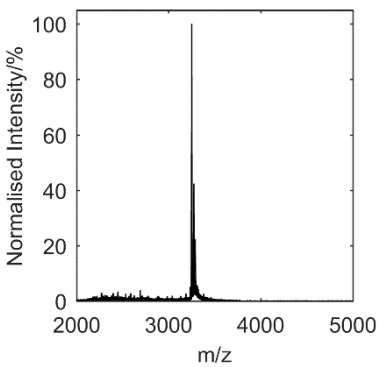




Exp.: 3415.153
Obs.: 3413.935

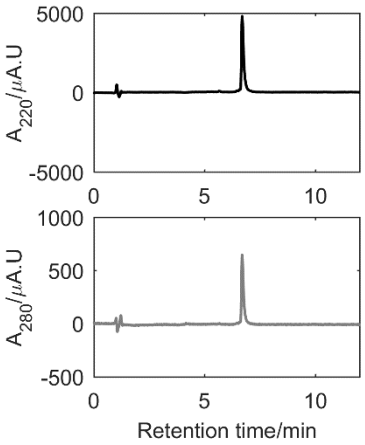
42

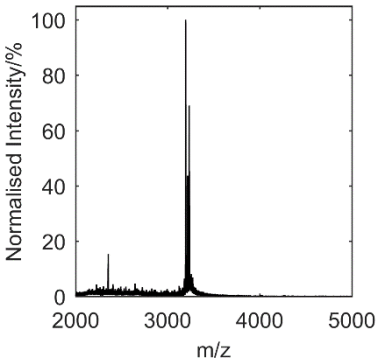




Exp.: 3250.715
Obs.: 3249.814

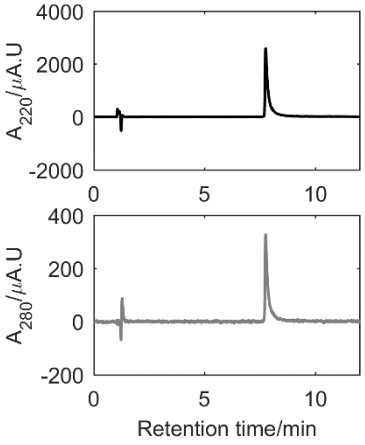
43

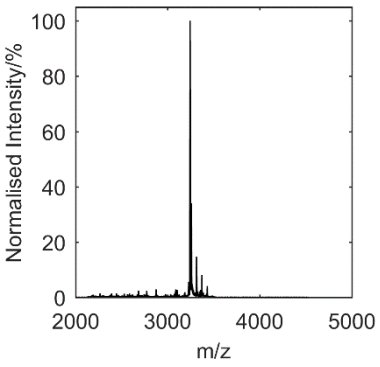




Exp.: 3194.607
Obs.: 3195.11

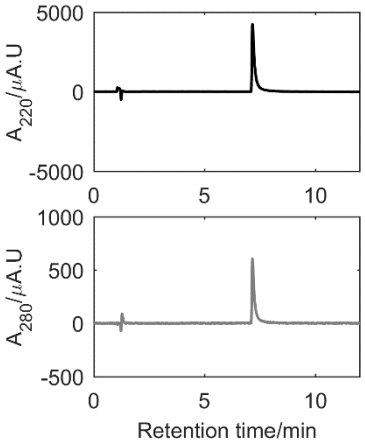
44

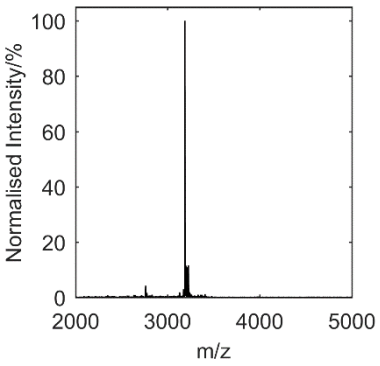




Exp.: 3243.054
Obs.: 3242.69

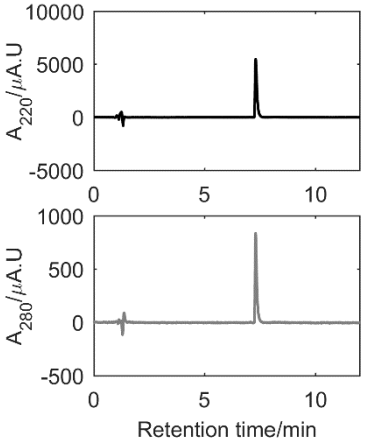
45

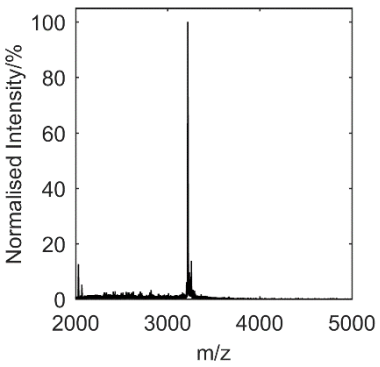




Exp.: 3186.946
Obs.: 3185.773

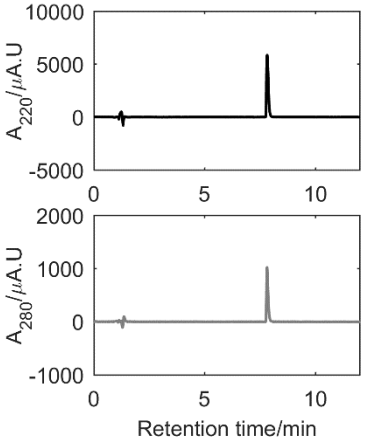
46

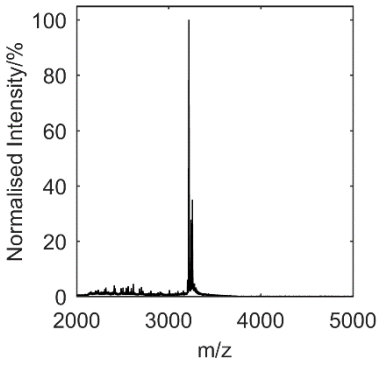




Exp.: 3218.766
Obs.: 3219.009

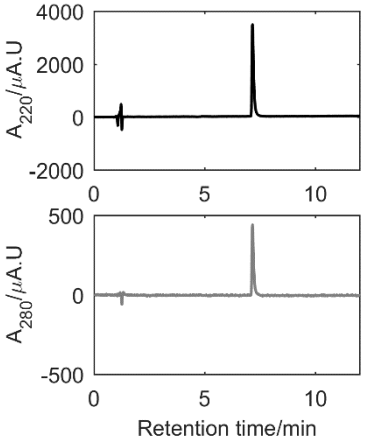
47

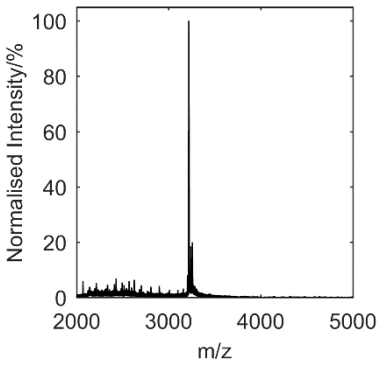




Exp.: 3218.766
Obs.: 3218.931

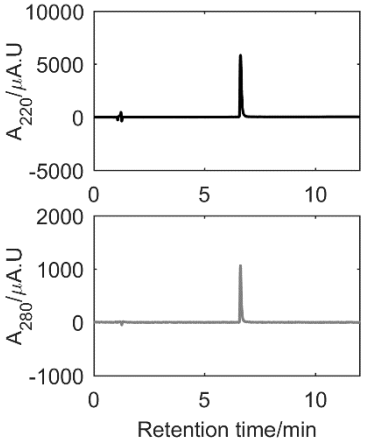
48

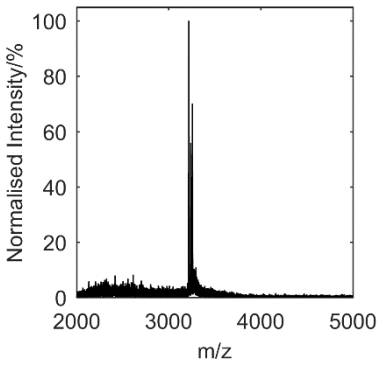




Exp.: 3218.766
Obs.: 3219.118

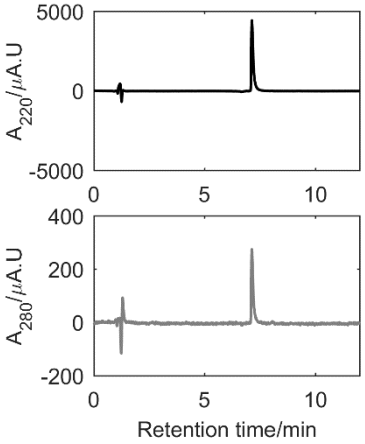
49

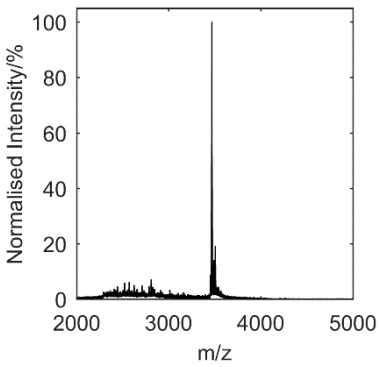




Exp.: 3218.766
Obs.: 3219.029

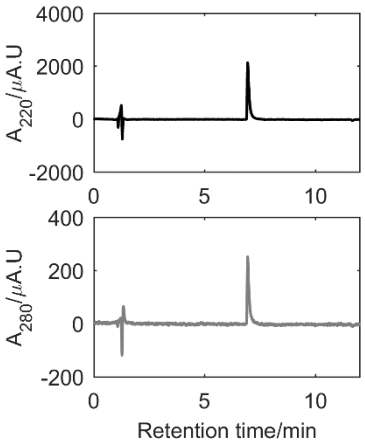
50

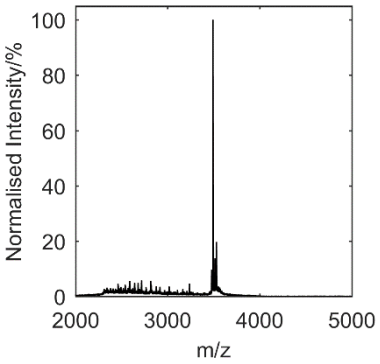




Exp.: 3467.989
Obs.: 3468.249

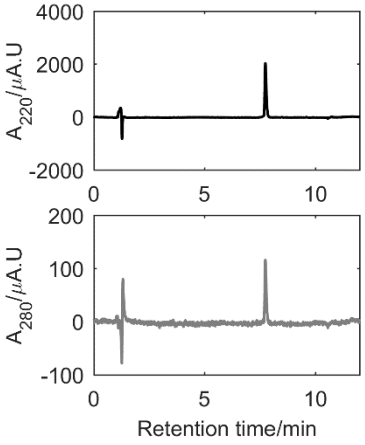
51

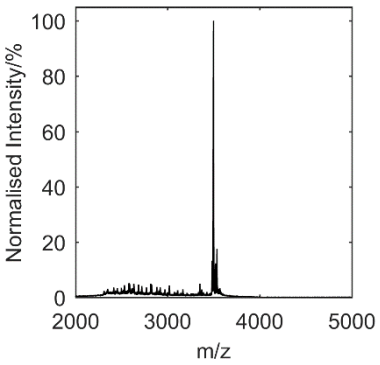




Exp.: 3491.026
Obs.: 3492.19

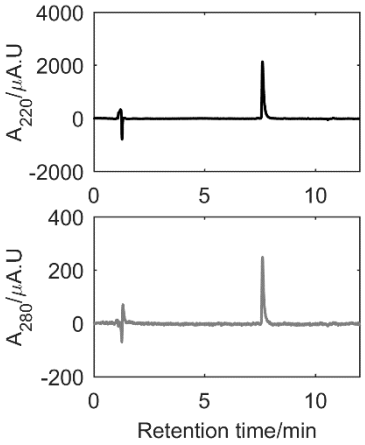
52

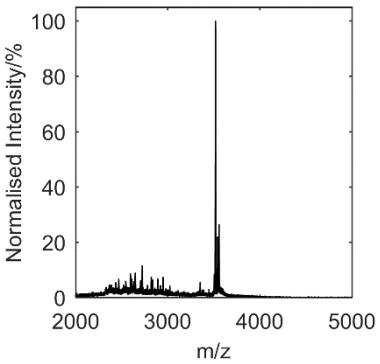




Exp.: 3495.999
Obs.: 3496.519

53





Exp.: 3519.036
Obs.: 3519.054

8.3 CD data of designed peptides

This section contains representative circular dichroism (CD) spectra (left or centre) and thermal denaturation experiments (right) for all peptides discussed in this thesis. Peptide concentrations are indicated in keys. All spectra were measured at 5 °C except those for peptides **46-53**, which were measured at 20 °C. Thermal denaturation experiments were performed by heating (5–95 °C) and cooling (95–5 °C) samples while monitoring MRE_{222}. All measurements were performed in 1X PBS (pH 7.4), unless otherwise stated. High tension (HT) plots are shown below all CD spectra and have that same keys as the associated spectra.

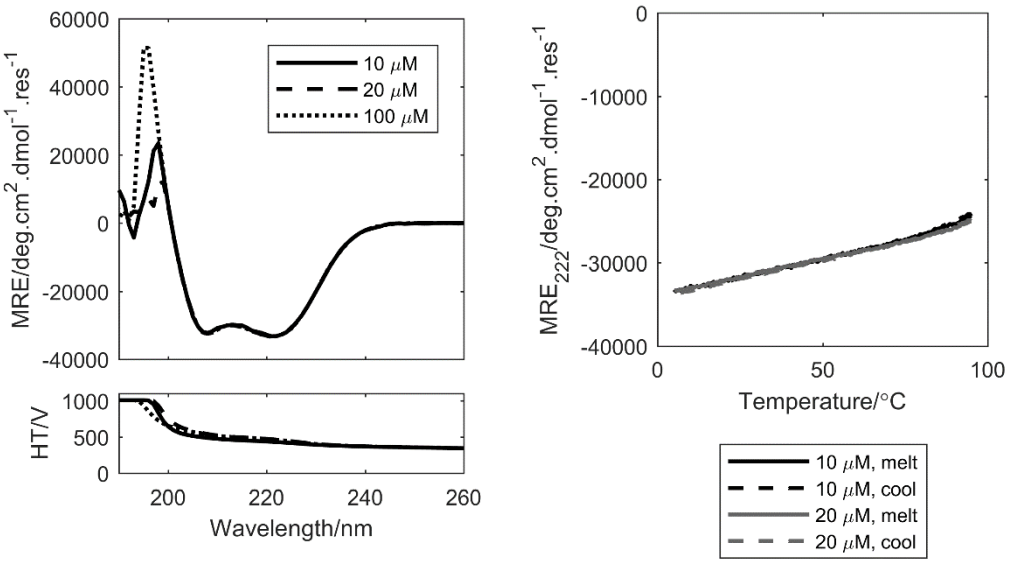


Figure 8-1 CD spectroscopy data for CC-Tet.

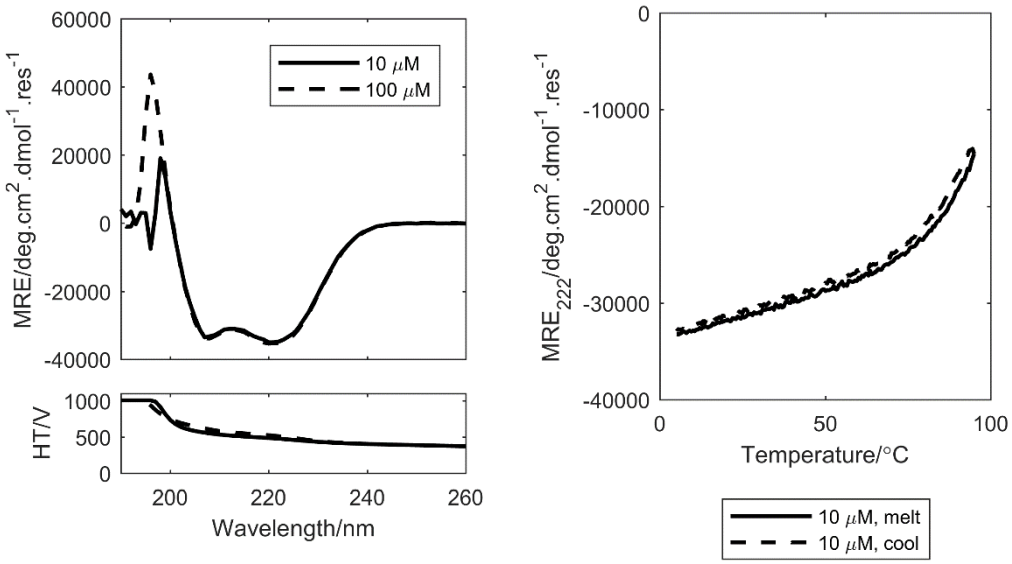


Figure 8-2 CD spectroscopy data for CC-Tet-KE.

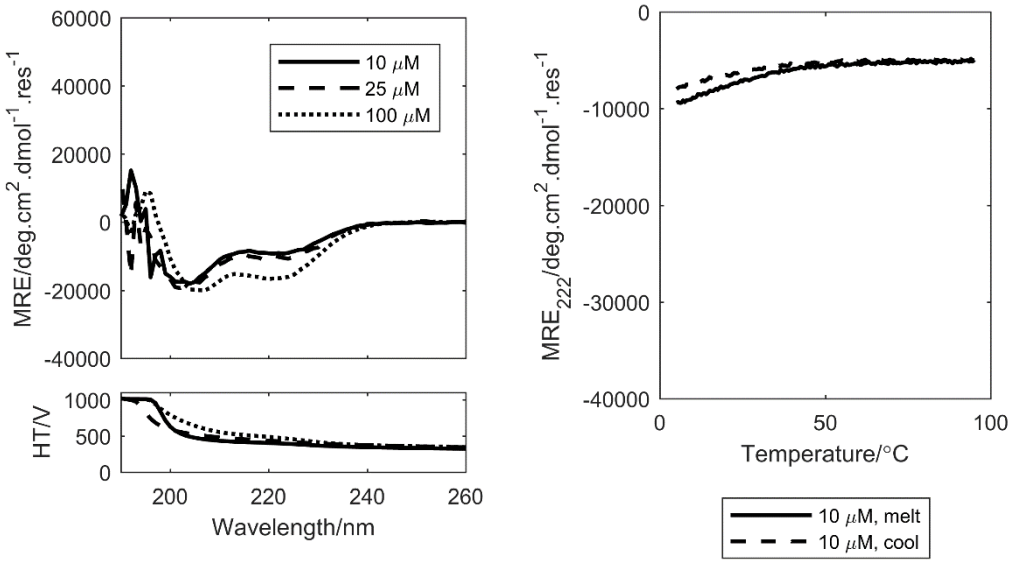


Figure 8-3 CD spectroscopy data for CC-Tet-KE-N-4.

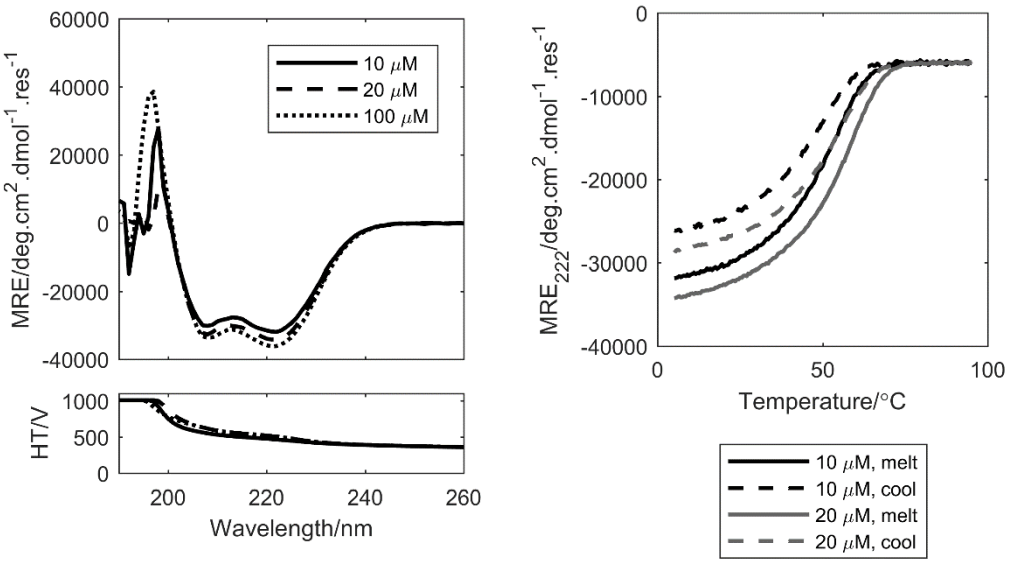


Figure 8-4 CD spectroscopy data for CC-Tet-KE-N-4.5.

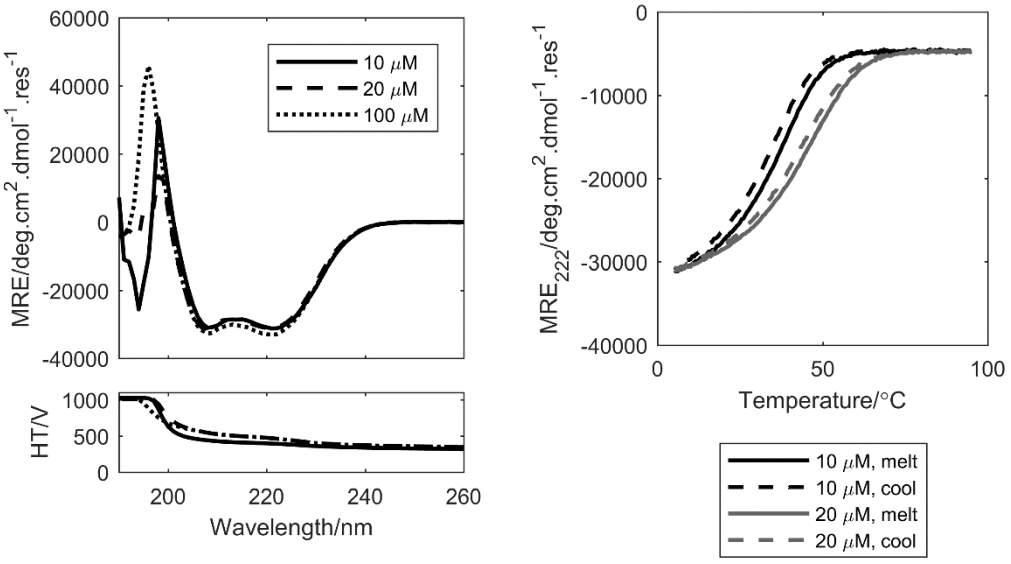


Figure 8-5 CD spectroscopy data for CC-Tet-IA.

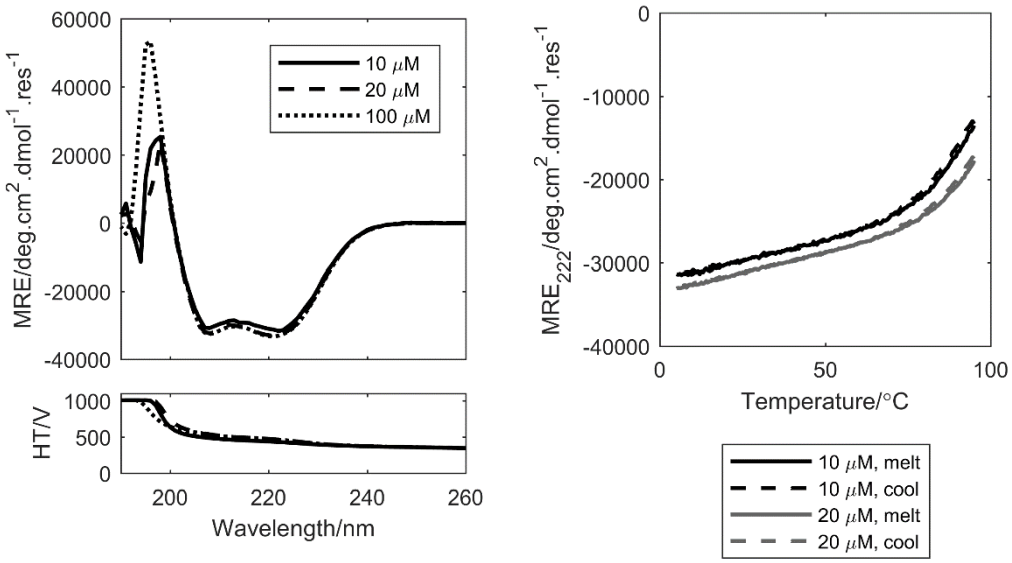


Figure 8-6 CD spectroscopy data for 1-LI-EK.

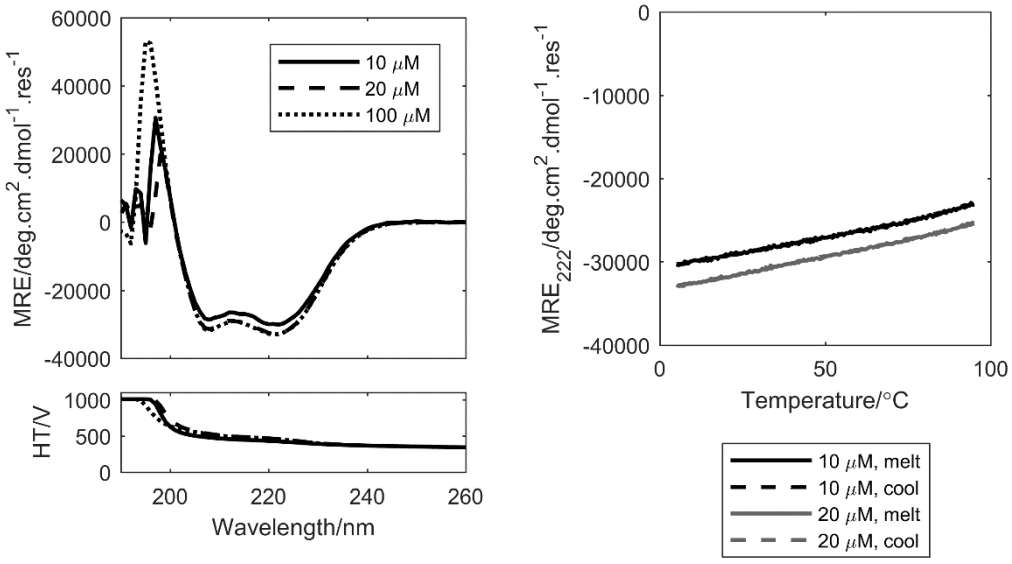


Figure 8-7 CD spectroscopy data for 1-LI-KE.

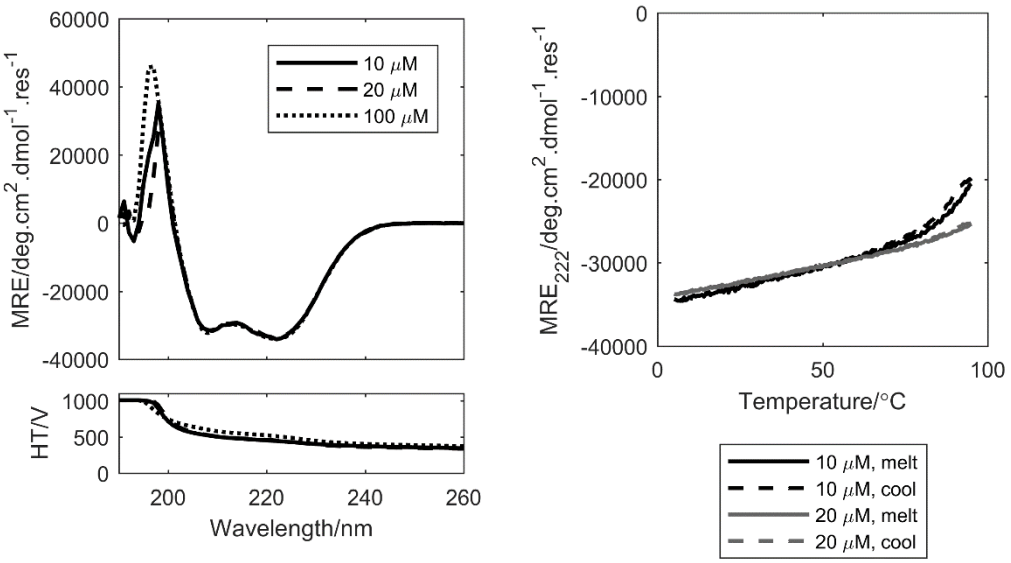


Figure 8-8 CD spectroscopy data for 2-LI-EK.

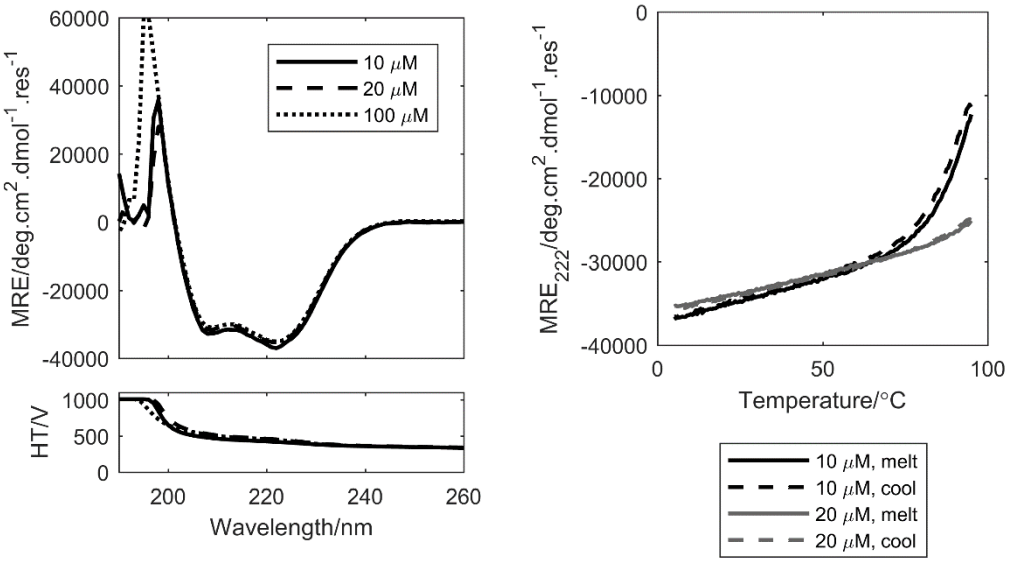


Figure 8-9 CD spectroscopy data for 2-LI-KE.

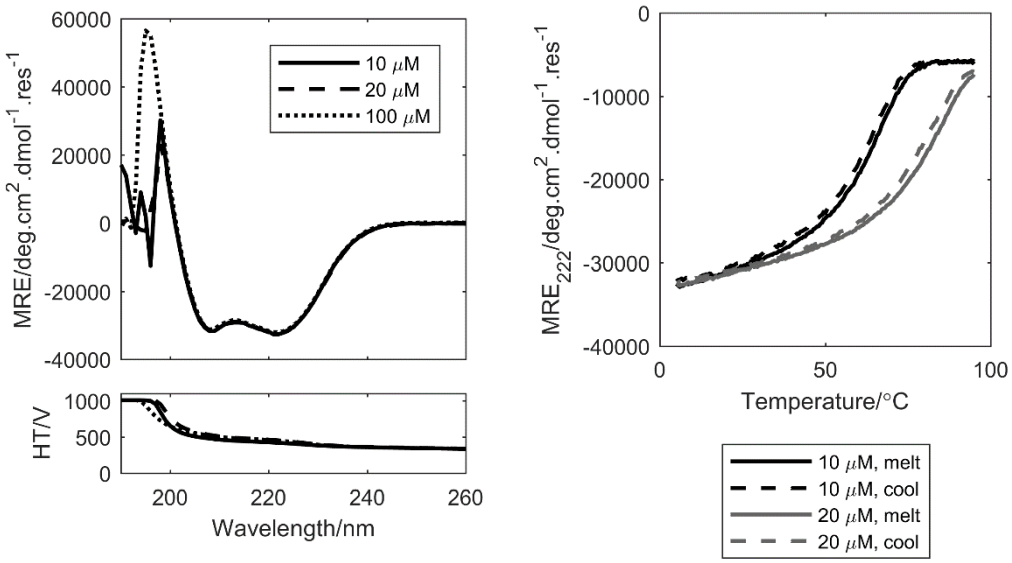


Figure 8-10 CD spectroscopy data for 3-LI-EK.

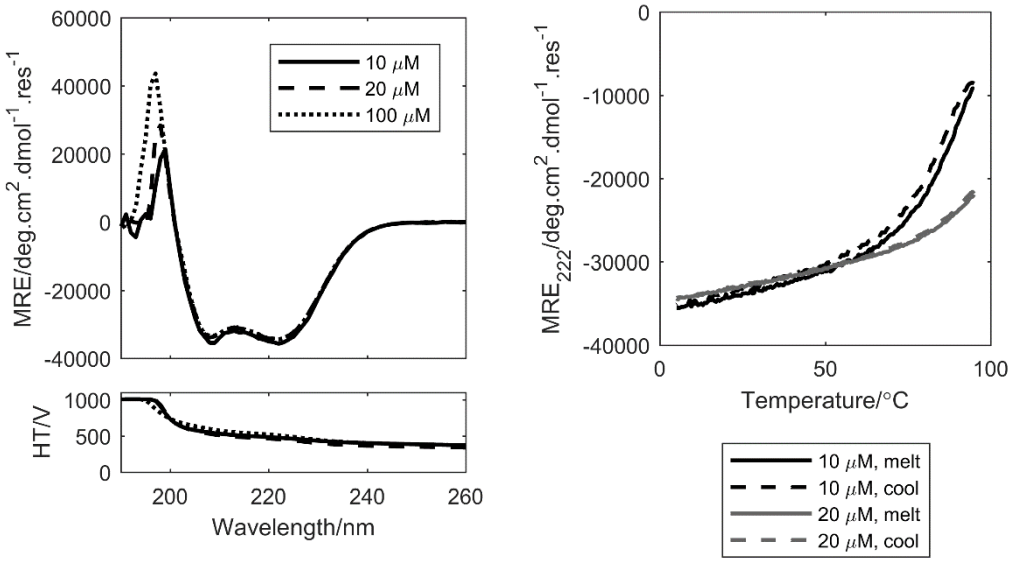


Figure 8-11 CD spectroscopy data for 3-LI-KE.

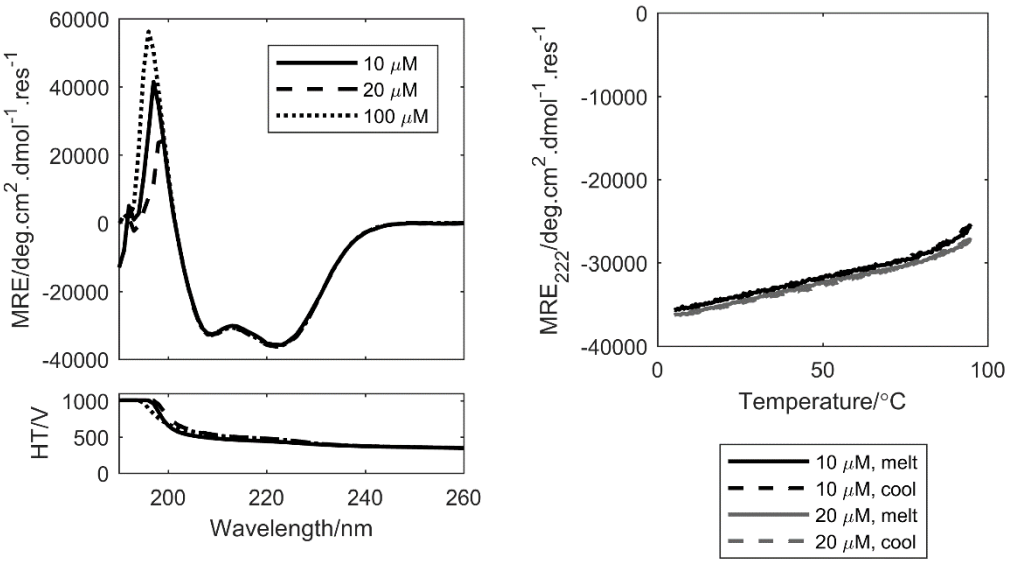


Figure 8-12 CD spectroscopy data for 4-LI-EK.

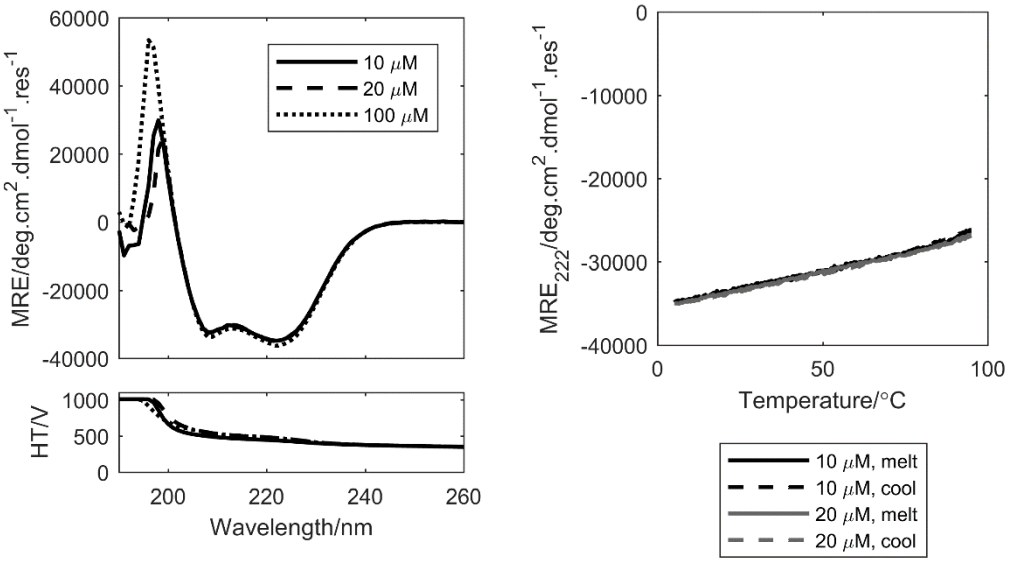


Figure 8-13 CD spectroscopy data for 4-LI-KE.

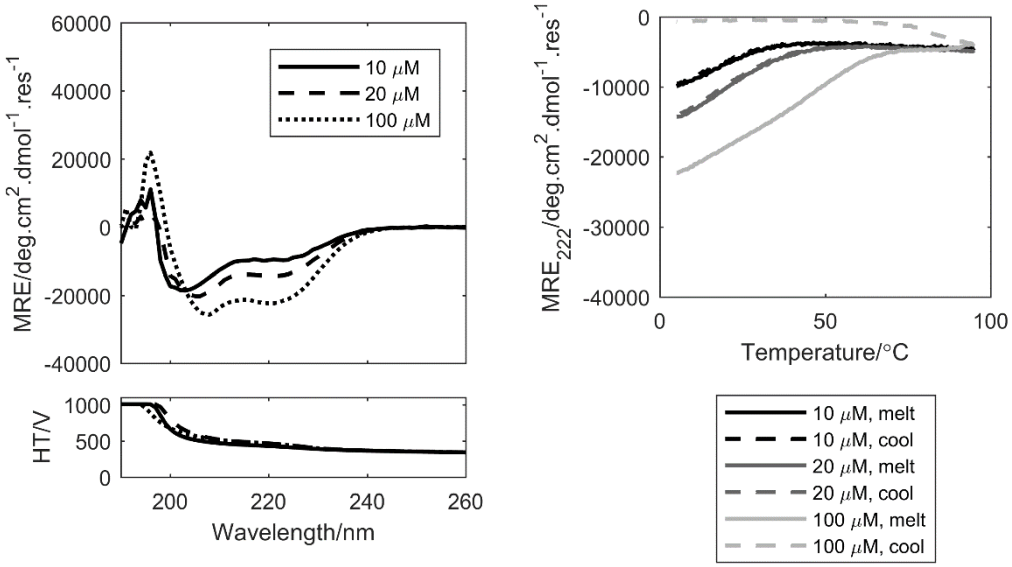


Figure 8-14 CD spectroscopy data for 1-LV-EK.

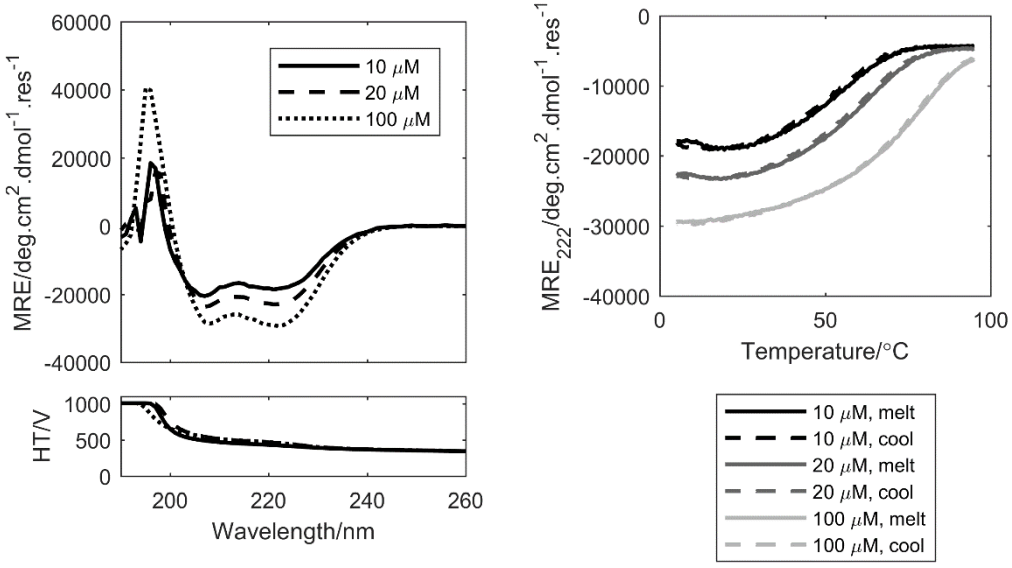


Figure 8-15 CD spectroscopy data for 1-LV-KE.

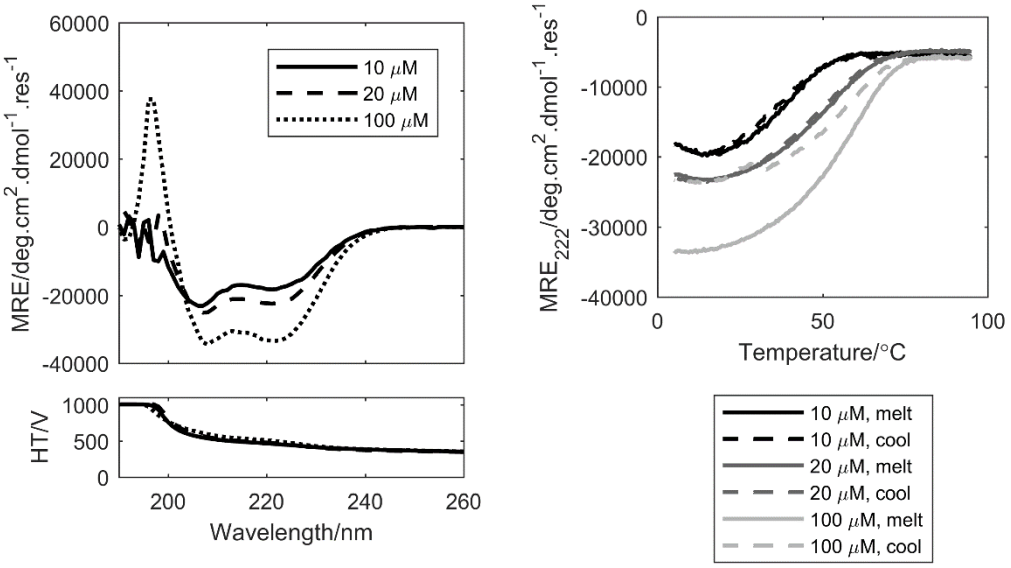


Figure 8-16 CD spectroscopy data for 2-LV-EK.

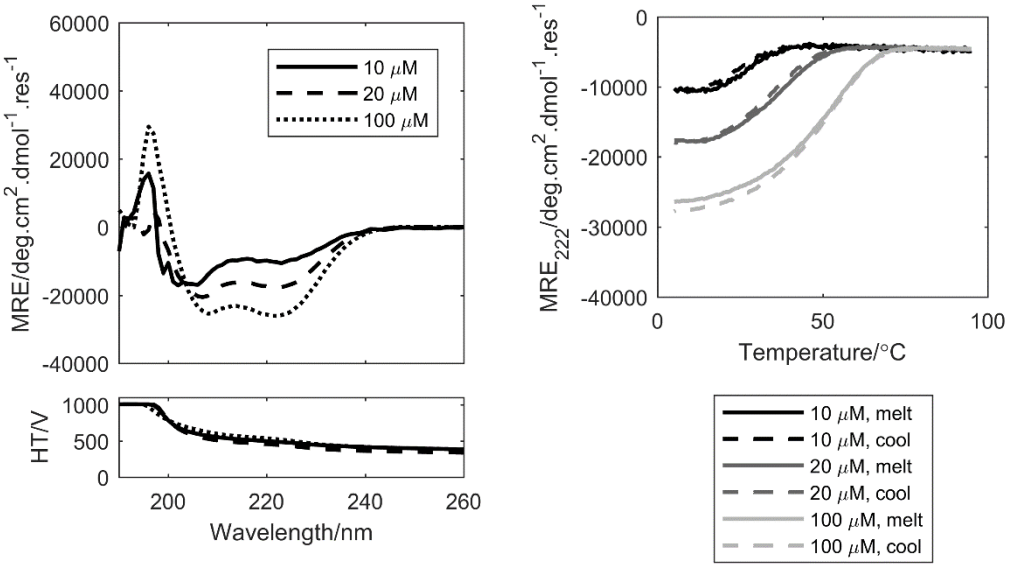


Figure 8-17 CD spectroscopy data for 2-LV-KE.

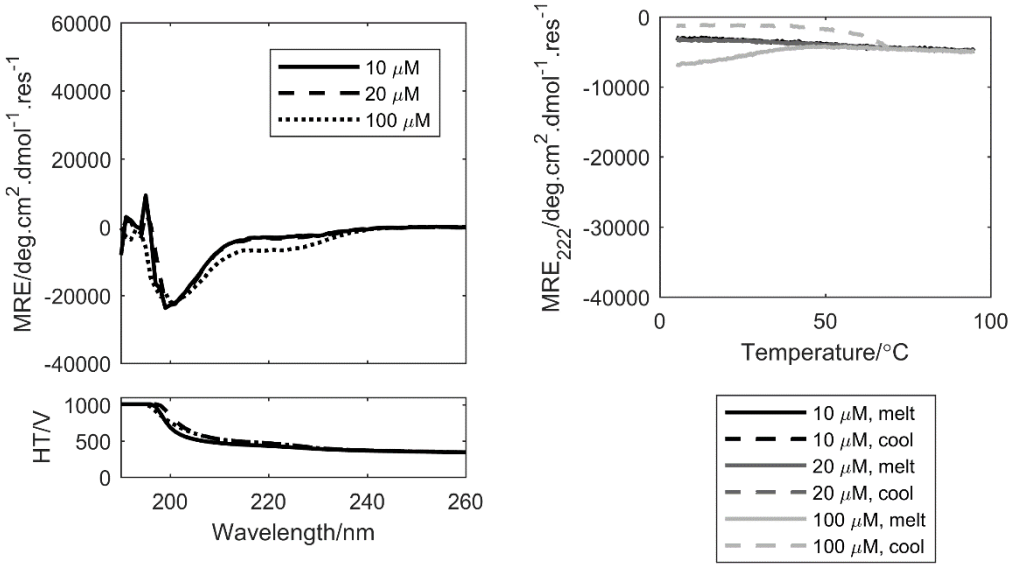


Figure 8-18 CD spectroscopy data for 3-LV-EK.

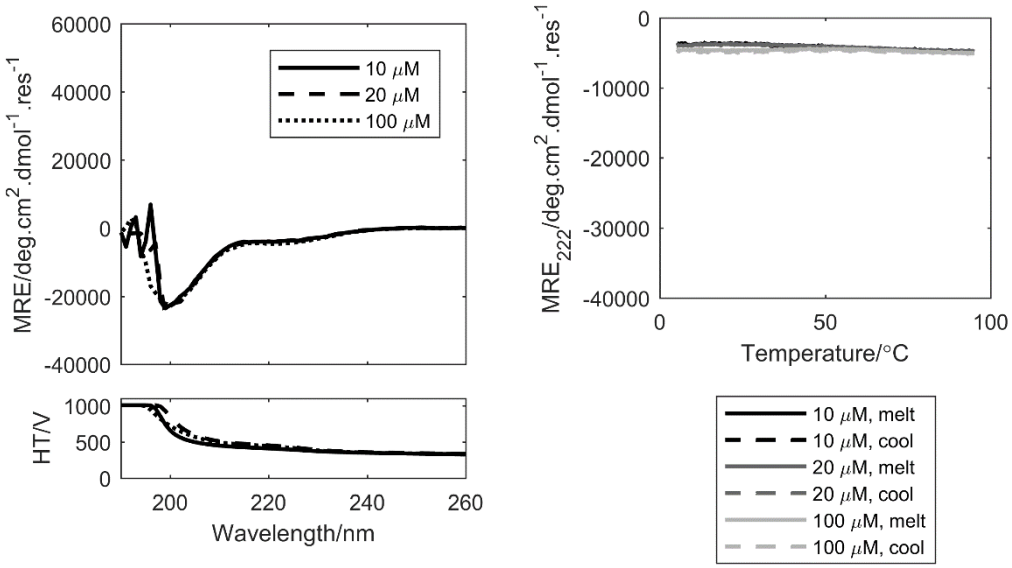


Figure 8-19 CD spectroscopy data for 3-LV-KE.

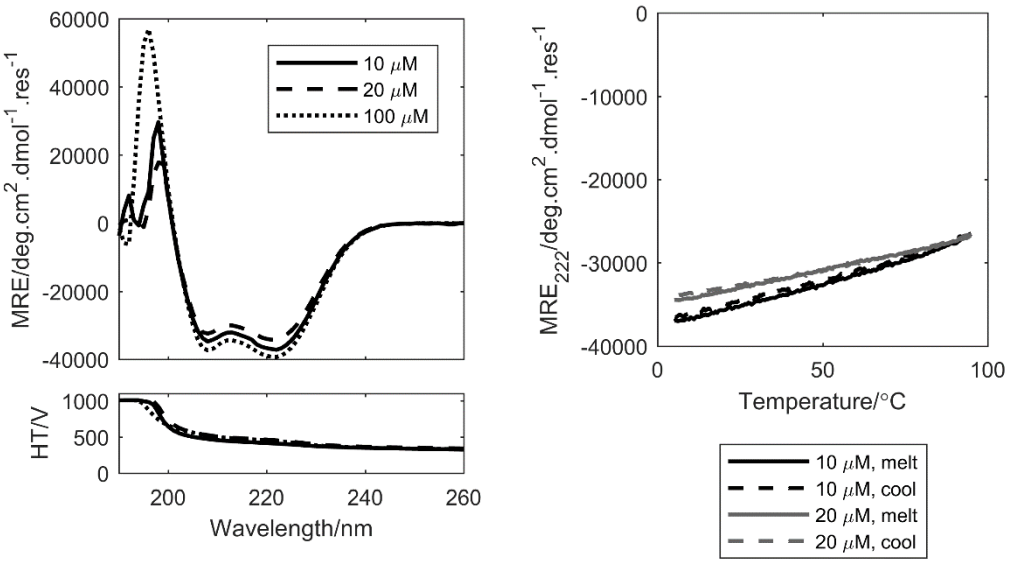


Figure 8-20 CD spectroscopy data for 2-LIA-EK.

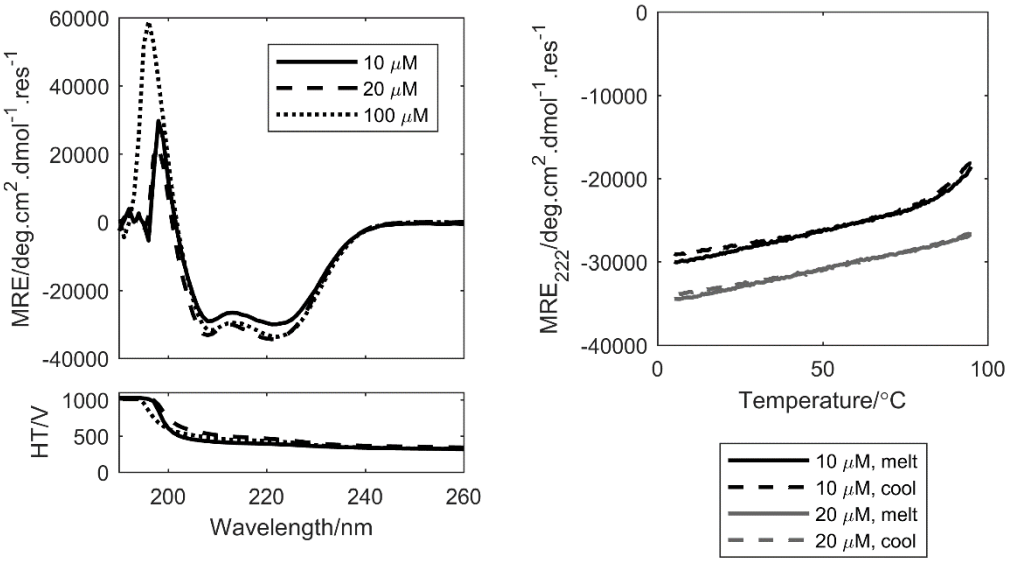


Figure 8-21 CD spectroscopy data for 2-LIA-KE.

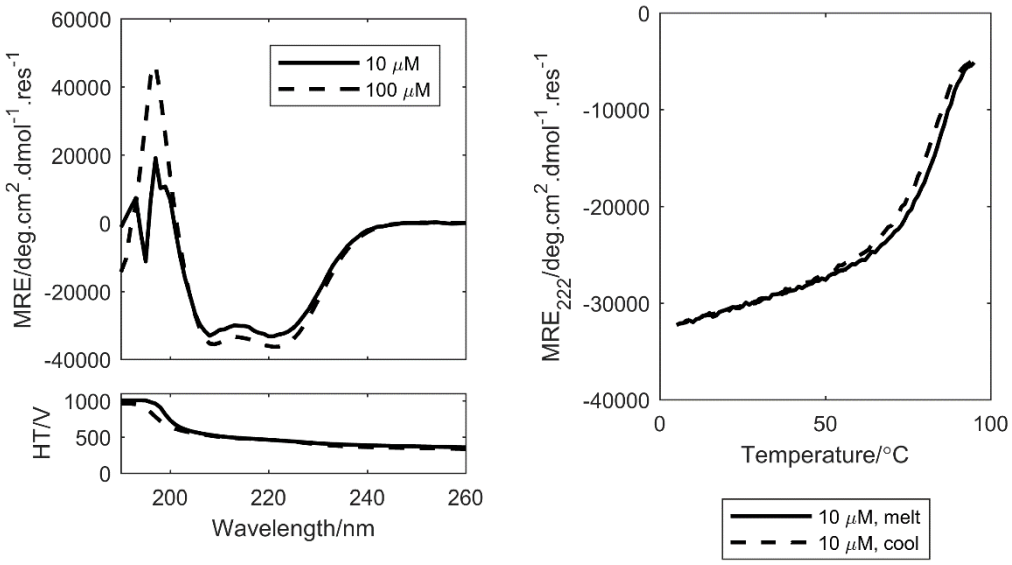


Figure 8-22 CD spectroscopy data for ELAEIK.

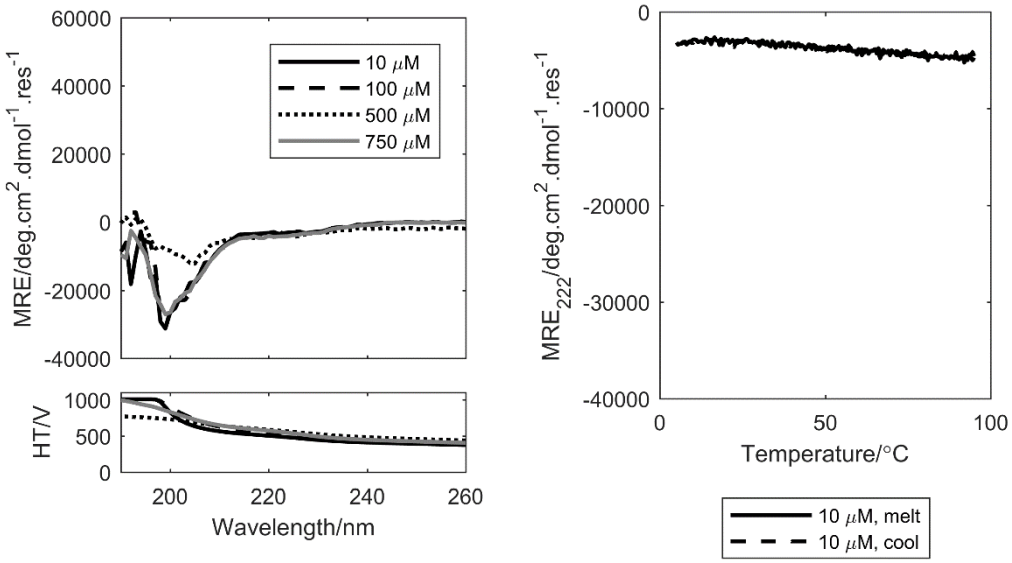


Figure 8-23 CD spectroscopy data for ELAEIK-M2-*a*.

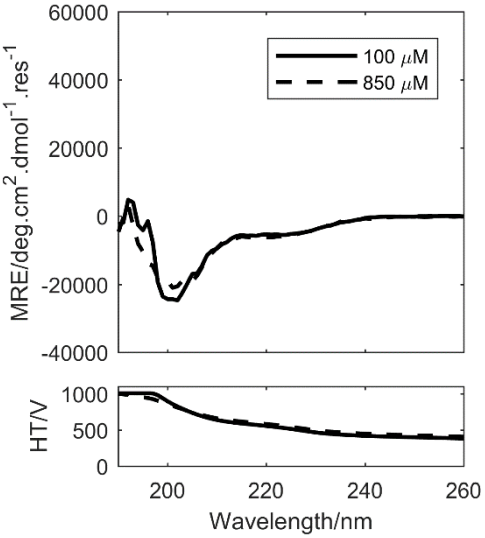


Figure 8-24 CD spectroscopy data for ELAEIK-M2-*d*.

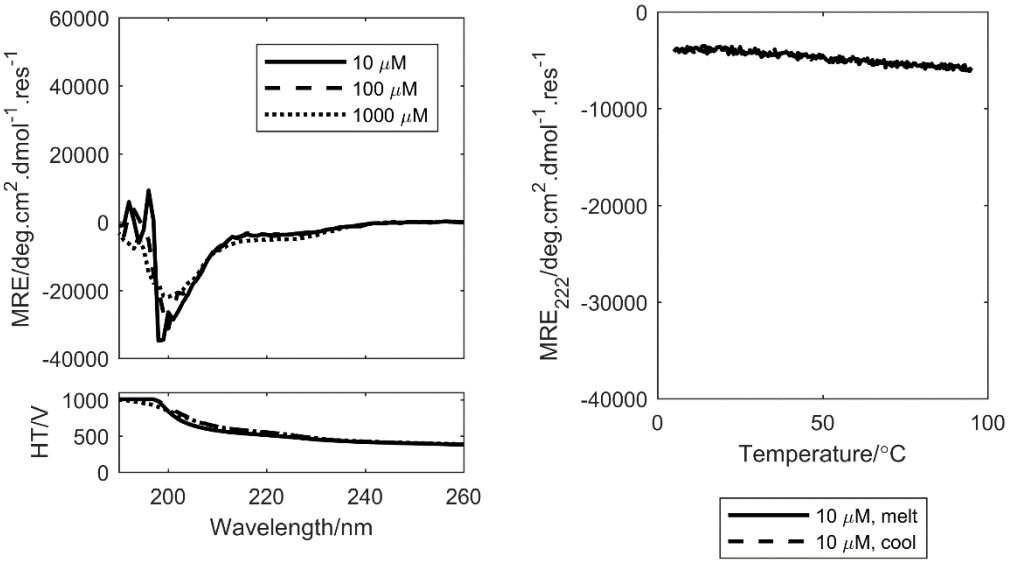


Figure 8-25 CD spectroscopy data for ELAEIK-M2-*a*-HI.

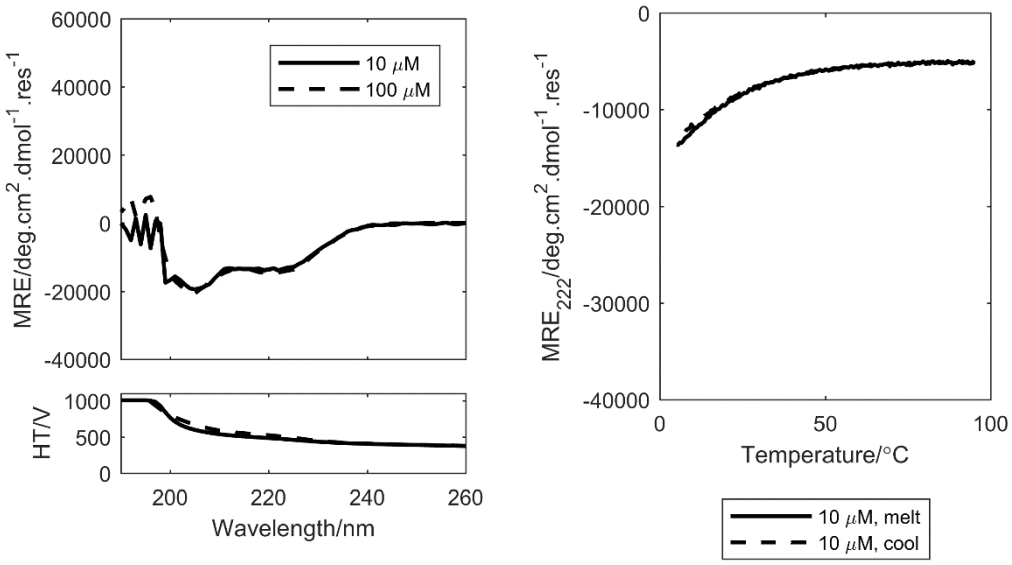


Figure 8-26 CD spectroscopy data for 1-LI-A.

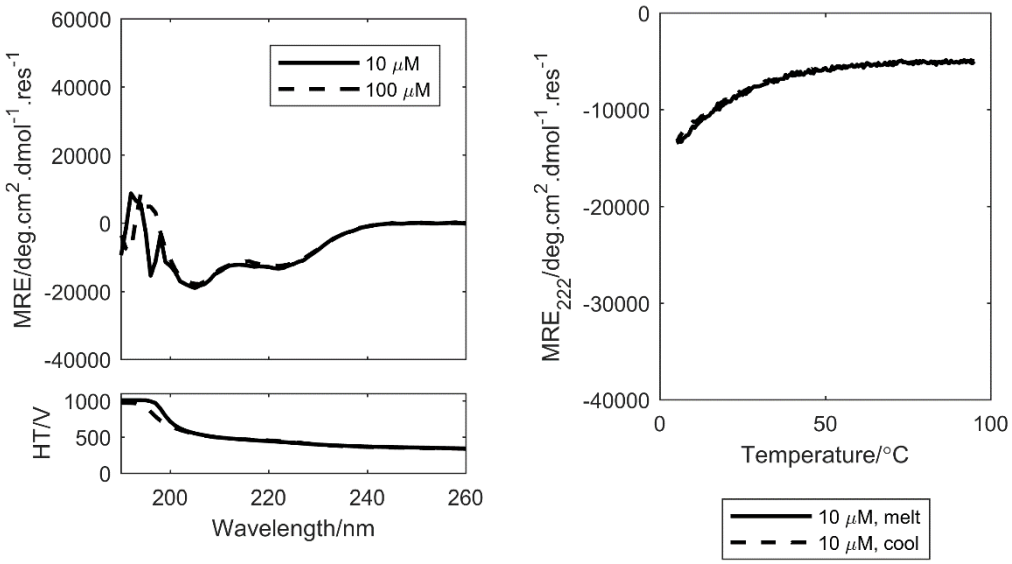


Figure 8-27 CD spectroscopy data for 1-LI-A-*g*.

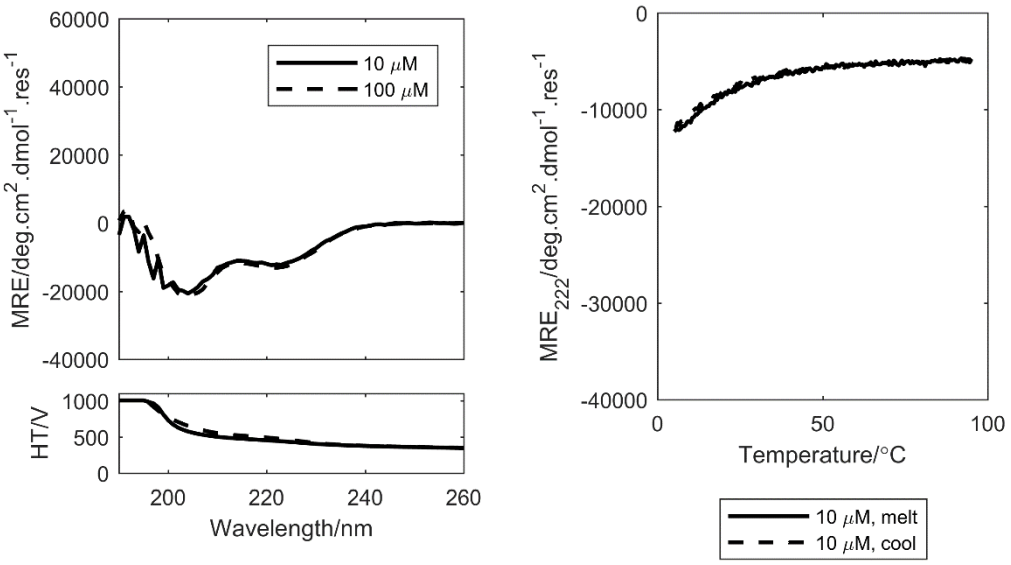


Figure 8-28 CD spectroscopy data for 1-LI-A*.

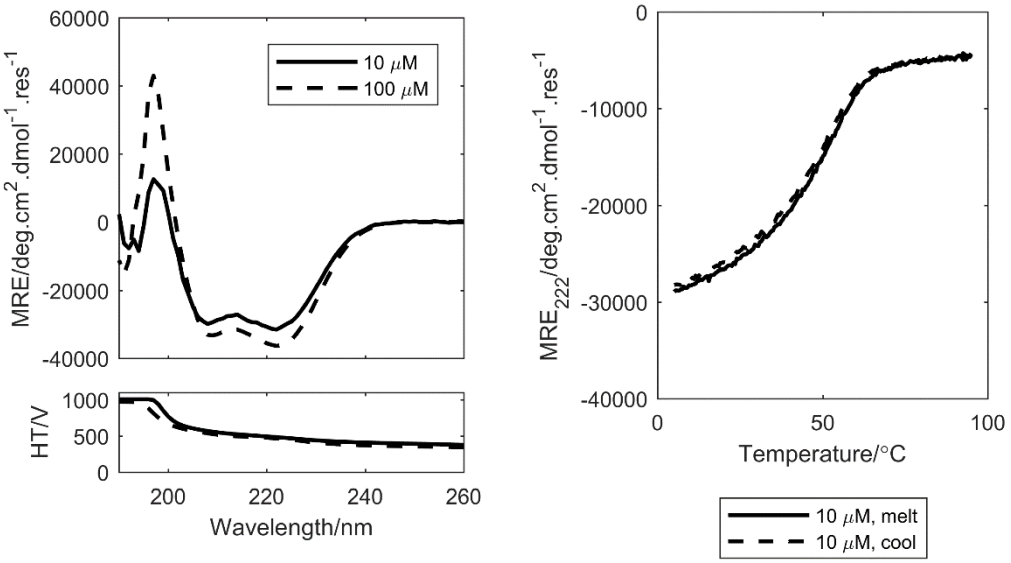


Figure 8-29 CD spectroscopy data for 2-LI-A.

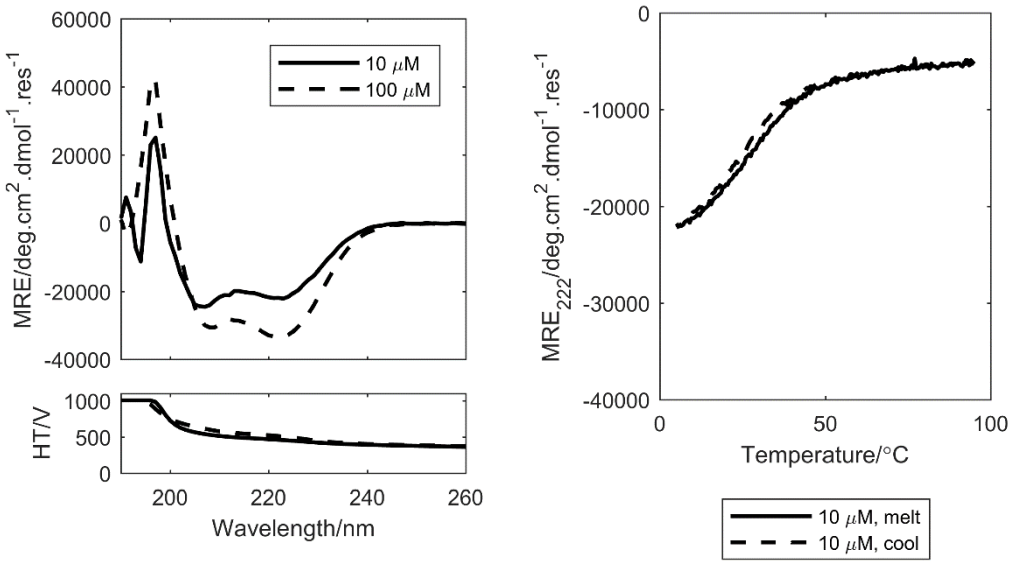


Figure 8-30 CD spectroscopy data for 3-LI-A.

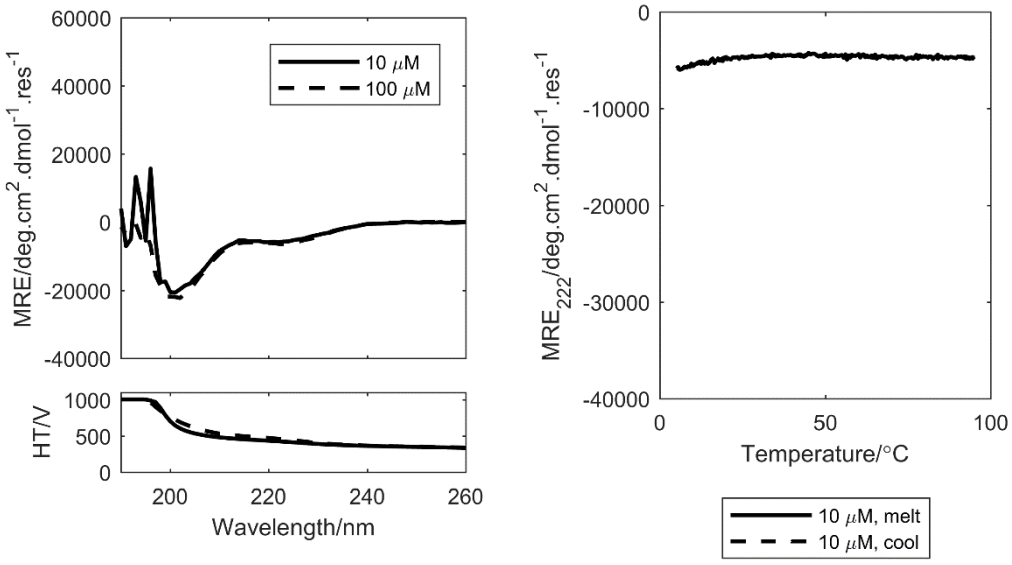


Figure 8-31 CD spectroscopy data for 1-LV-A.

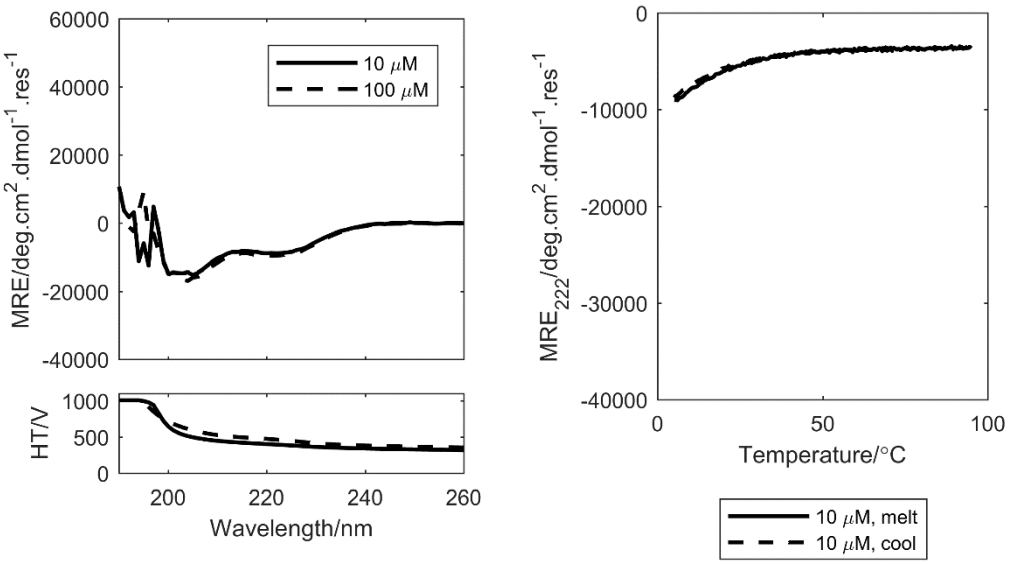


Figure 8-32 CD spectroscopy data for 2-LV-A.

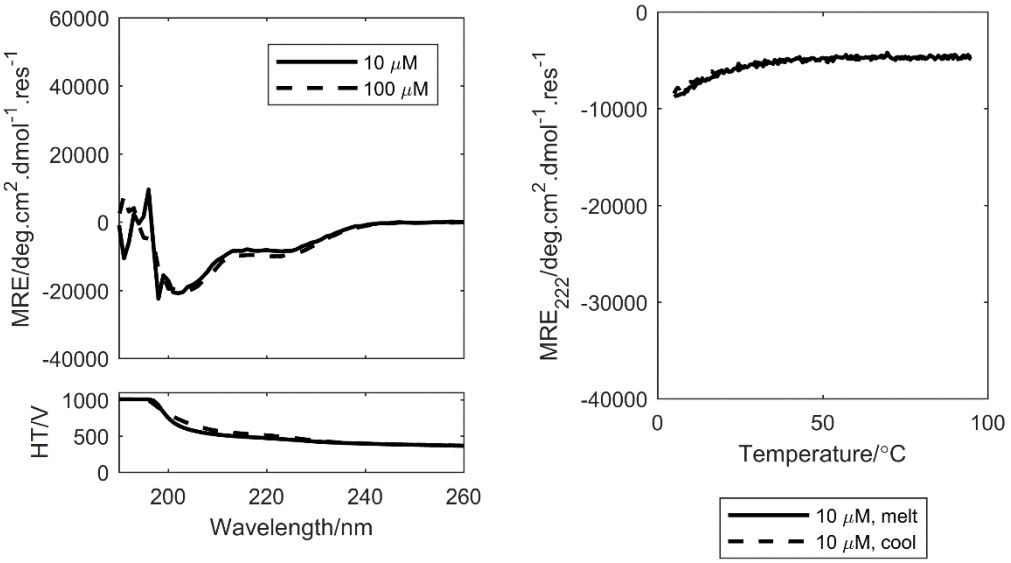


Figure 8-33 CD spectroscopy data for 3-LV-A.

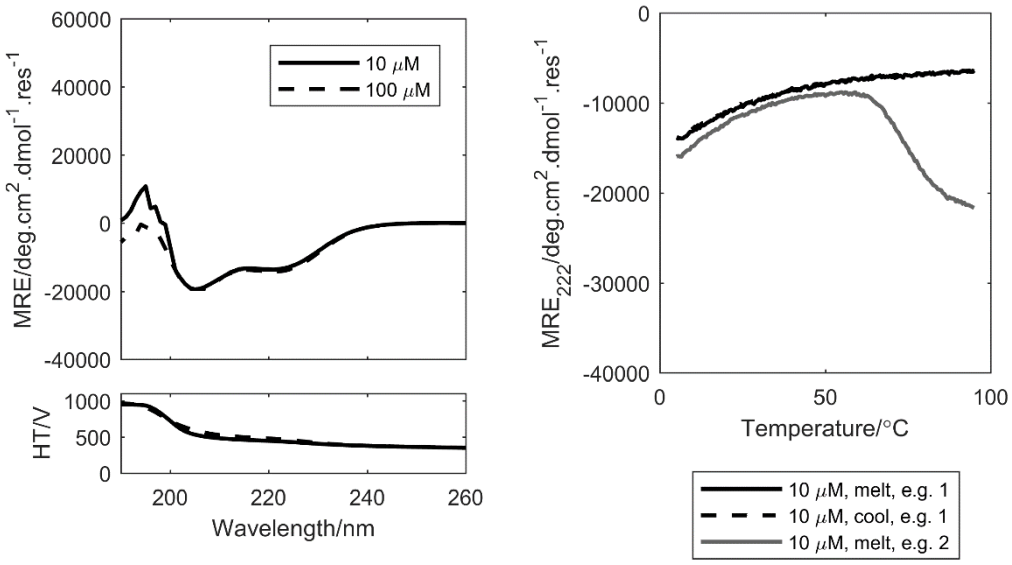


Figure 8-34 CD spectroscopy data for 1-LI-B.

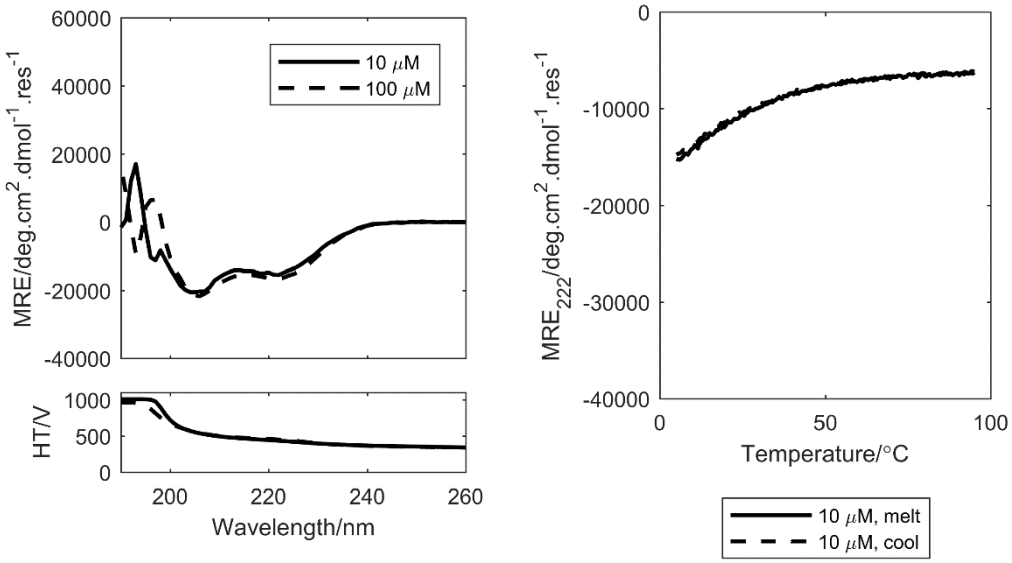


Figure 8-35 CD spectroscopy data for 1-LI-B-*g*.

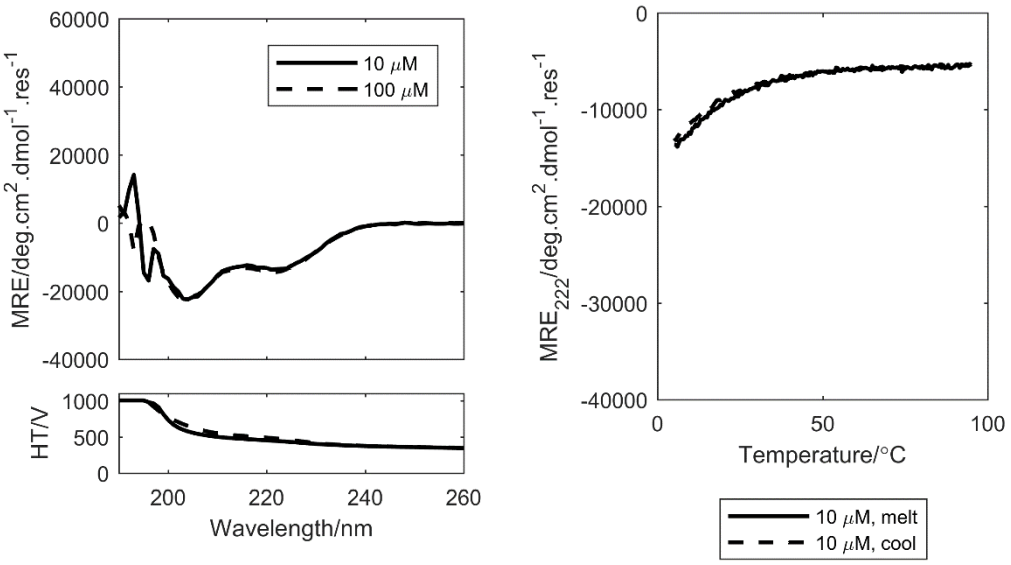


Figure 8-36 CD spectroscopy data for 1-LI-B*.

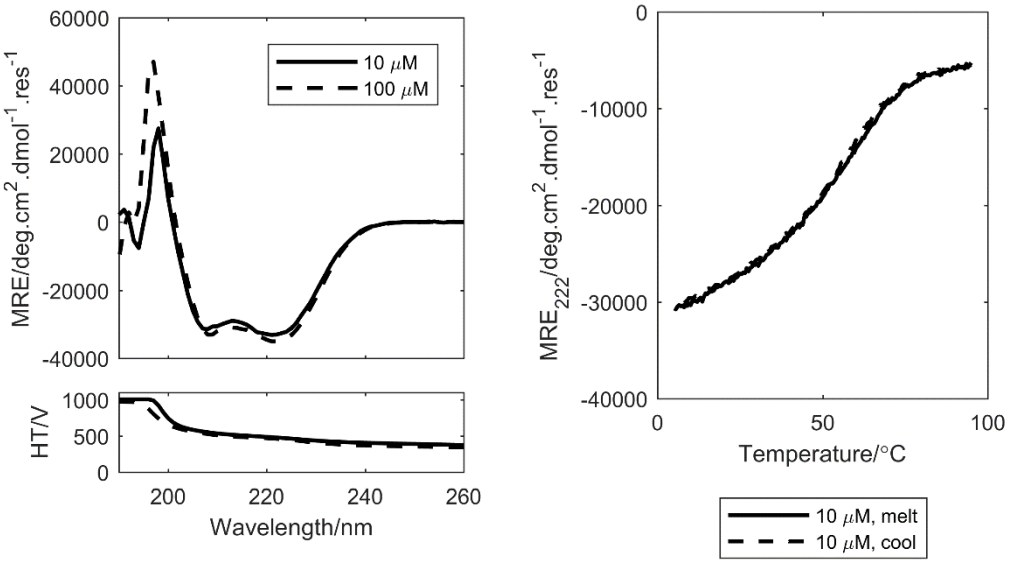


Figure 8-37 CD spectroscopy data for 2-LI-B.

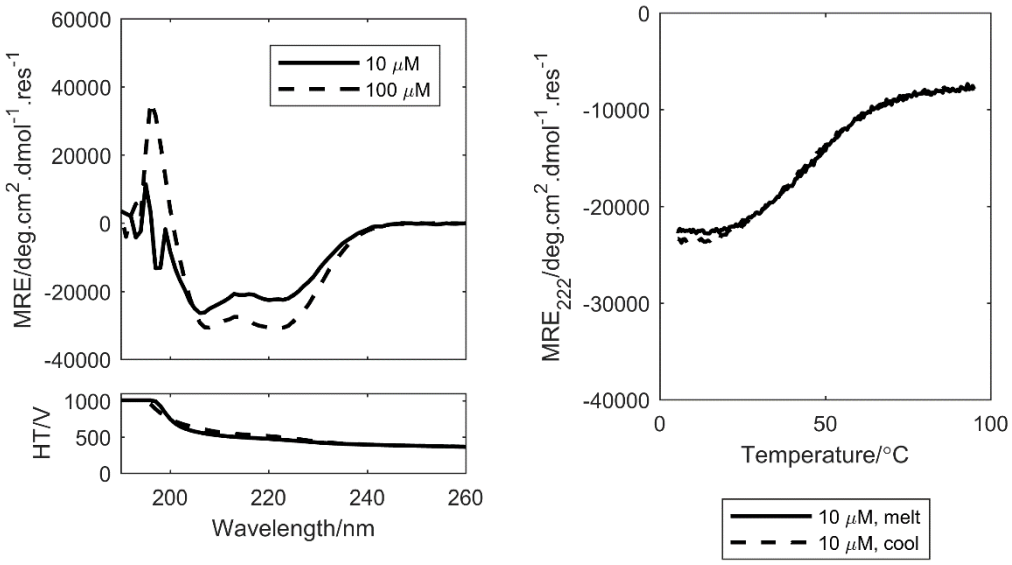


Figure 8-38 CD spectroscopy data for 3-LI-B.

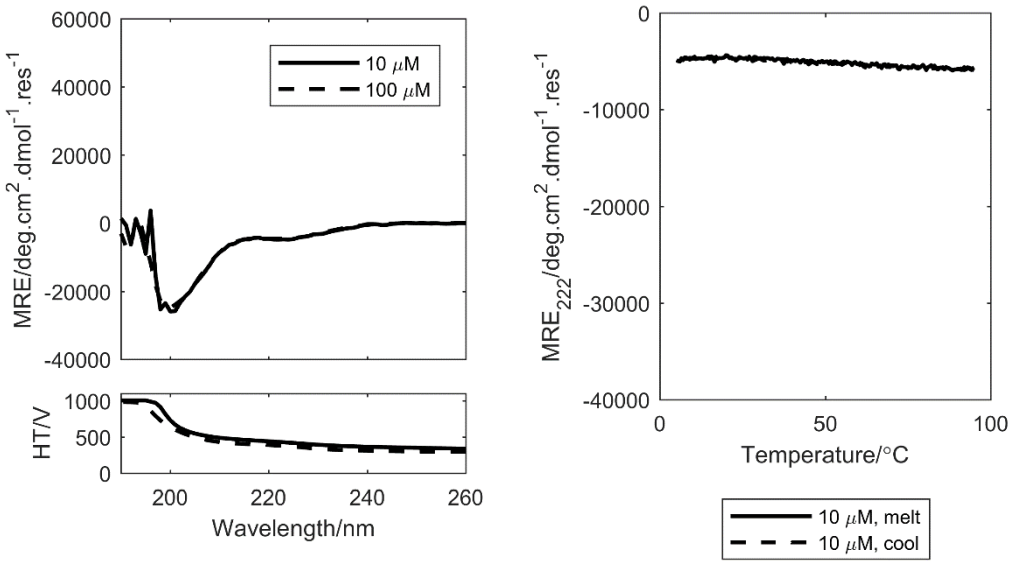


Figure 8-39 CD spectroscopy data for 1-LV-B.

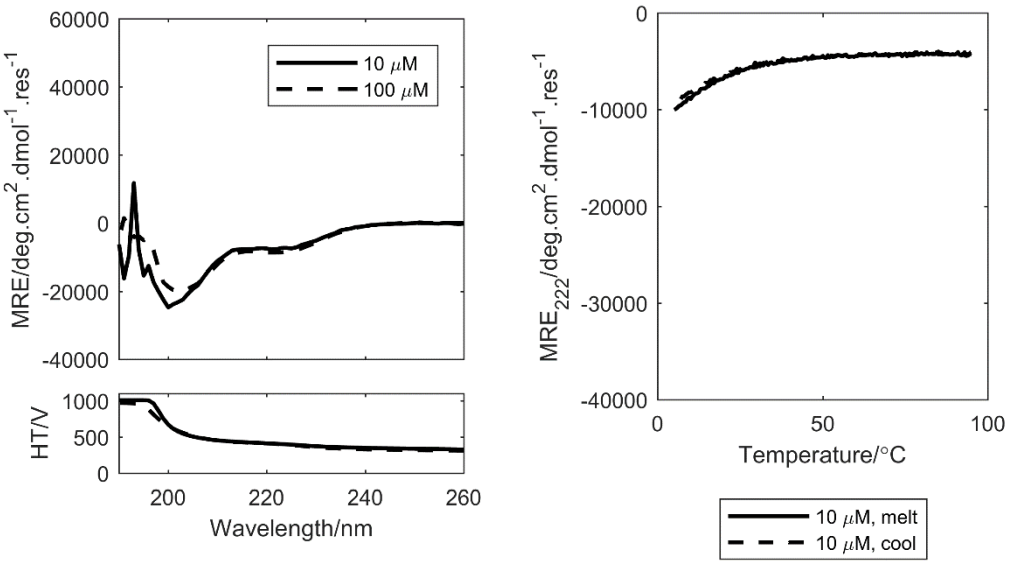


Figure 8-40 CD spectroscopy data for 2-LV-B.

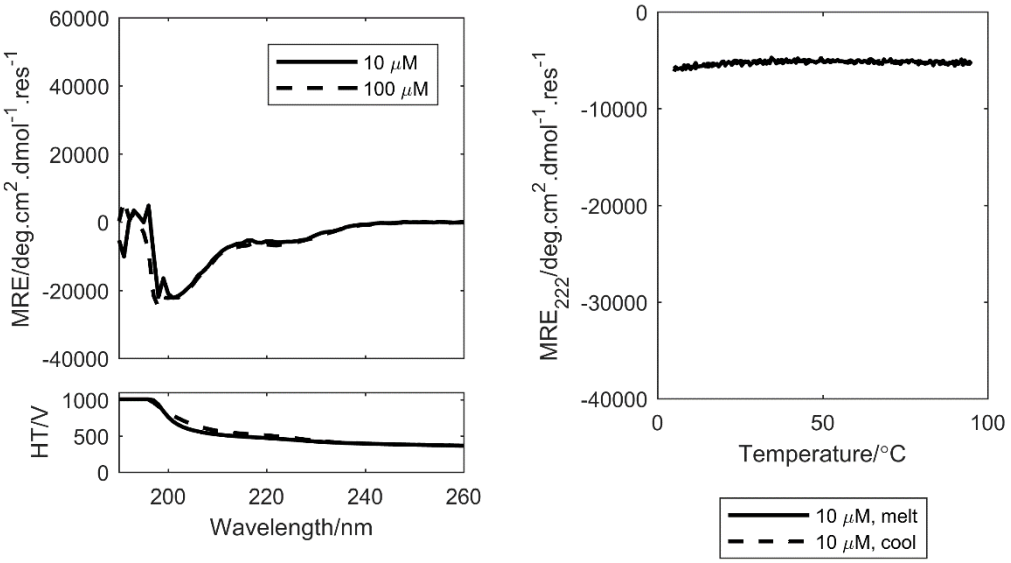


Figure 8-41 CD spectroscopy data for 3-LV-B.

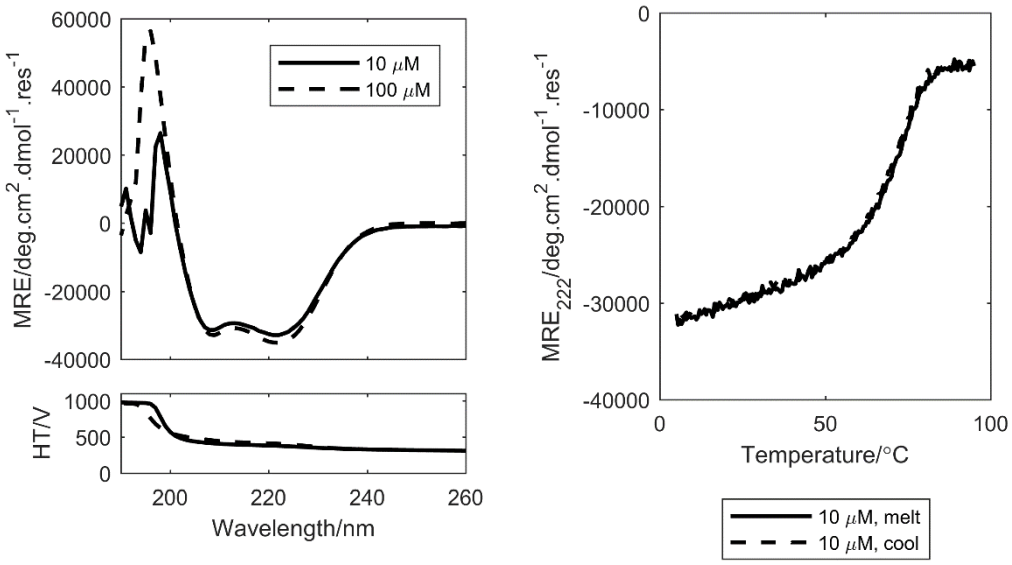


Figure 8-42 CD spectroscopy data for 2-LIA-A.

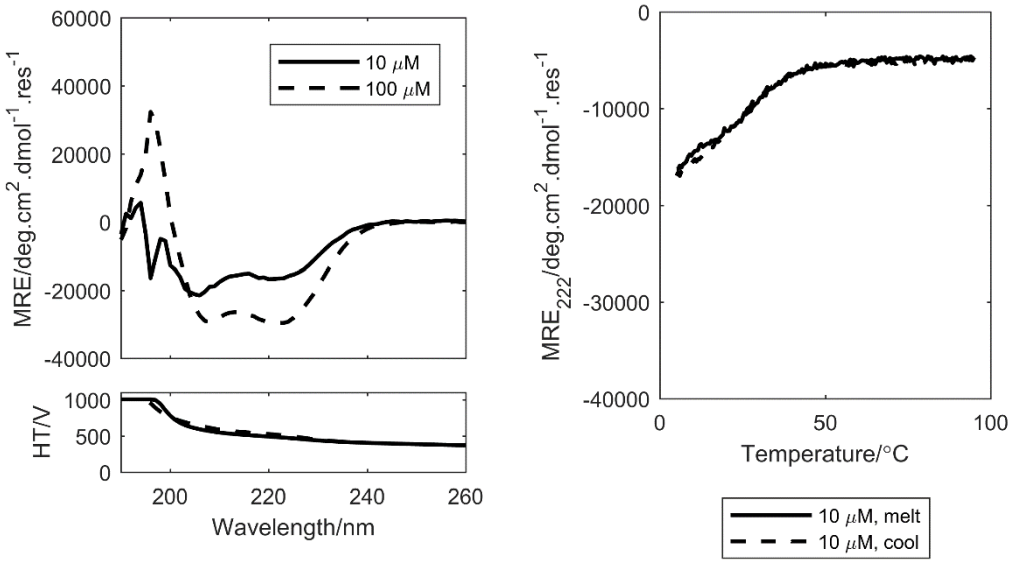


Figure 8-43 CD spectroscopy data for 2-LVA-A.

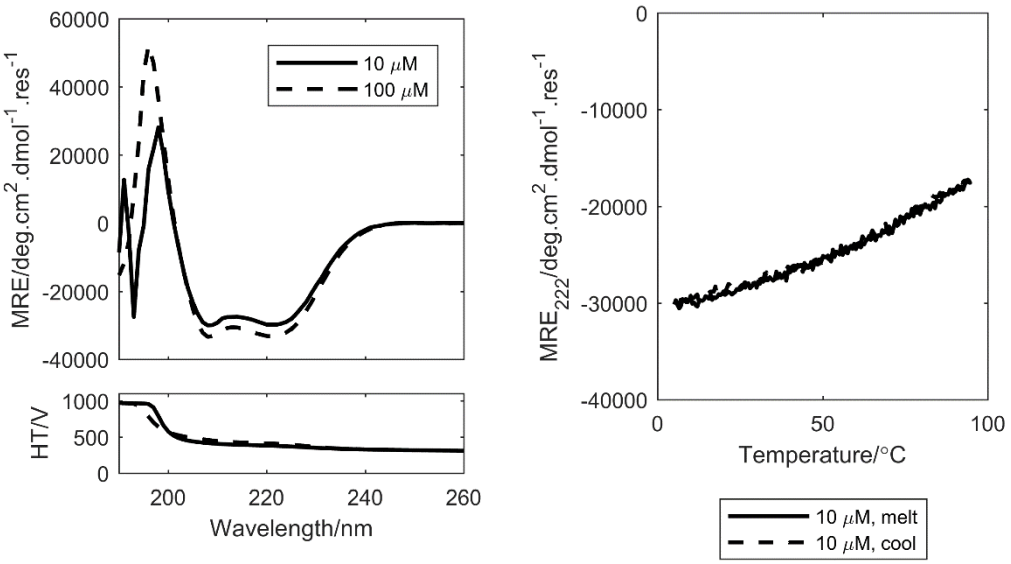


Figure 8-44 CD spectroscopy data for 2-LIA-B.

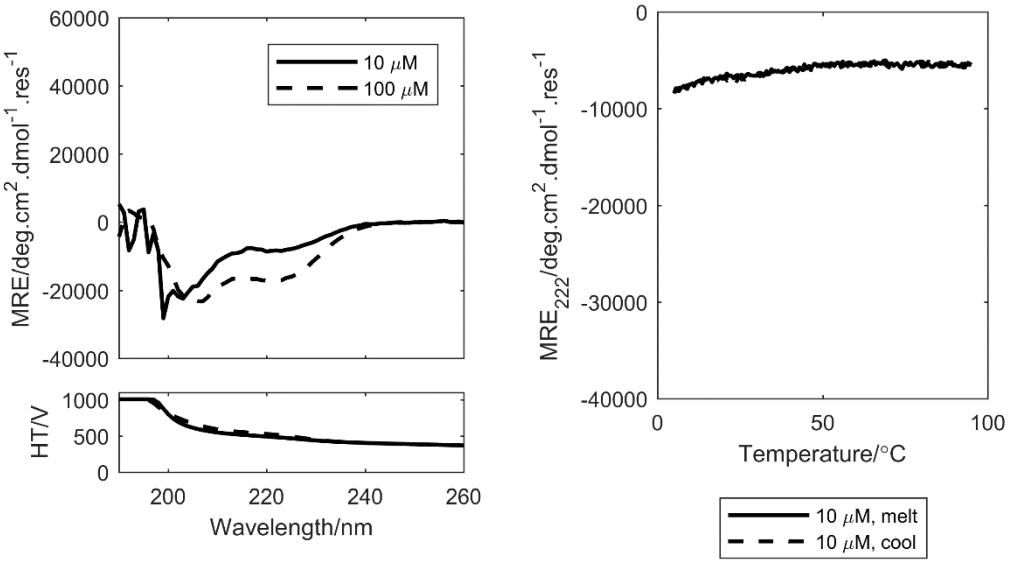


Figure 8-45 CD spectroscopy data for 2-LVA-B.

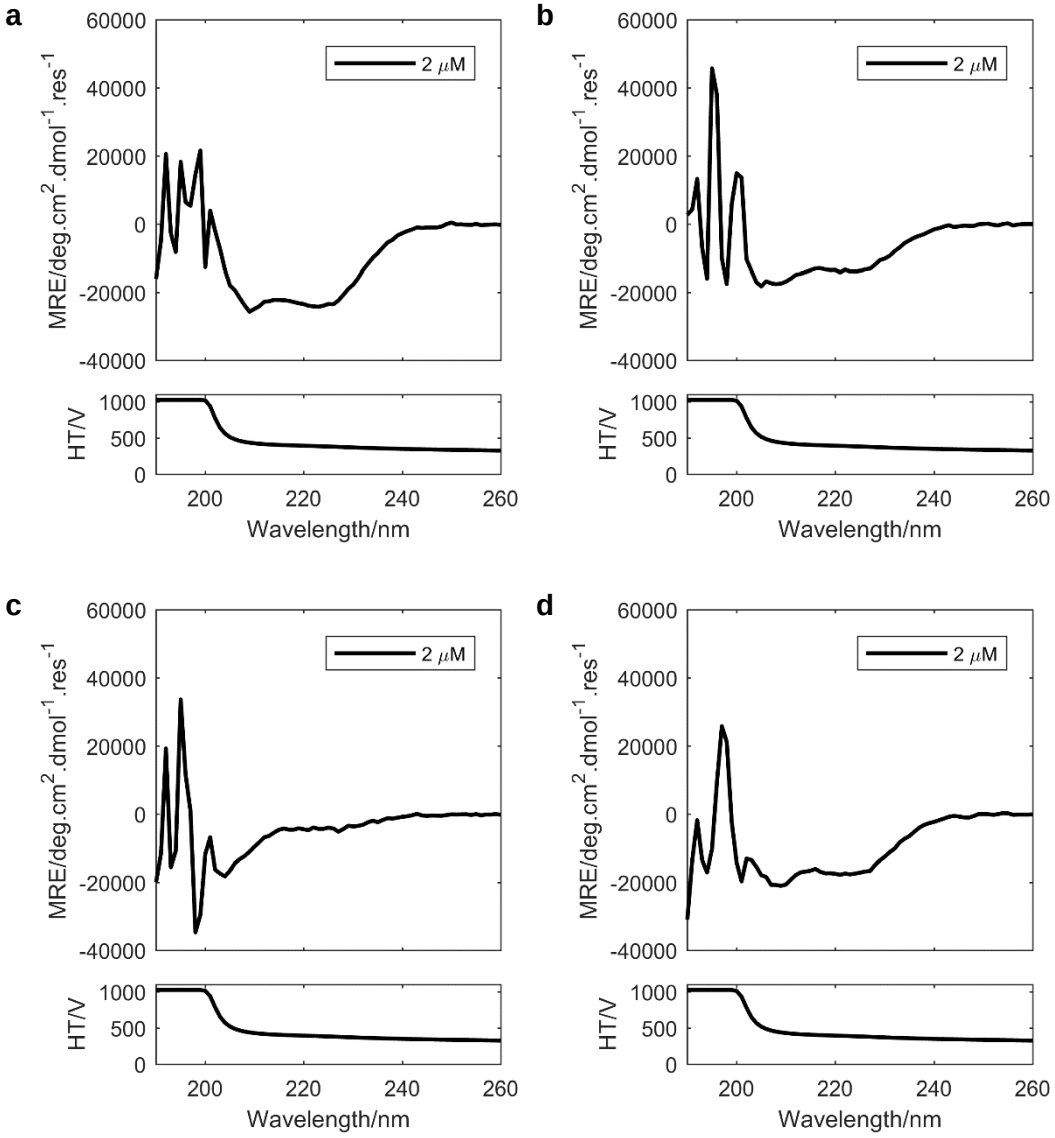


Figure 8-46 CD spectroscopy data for ABCD peptides: (a) 1; (b) 2; (c) 3; (d) 4.

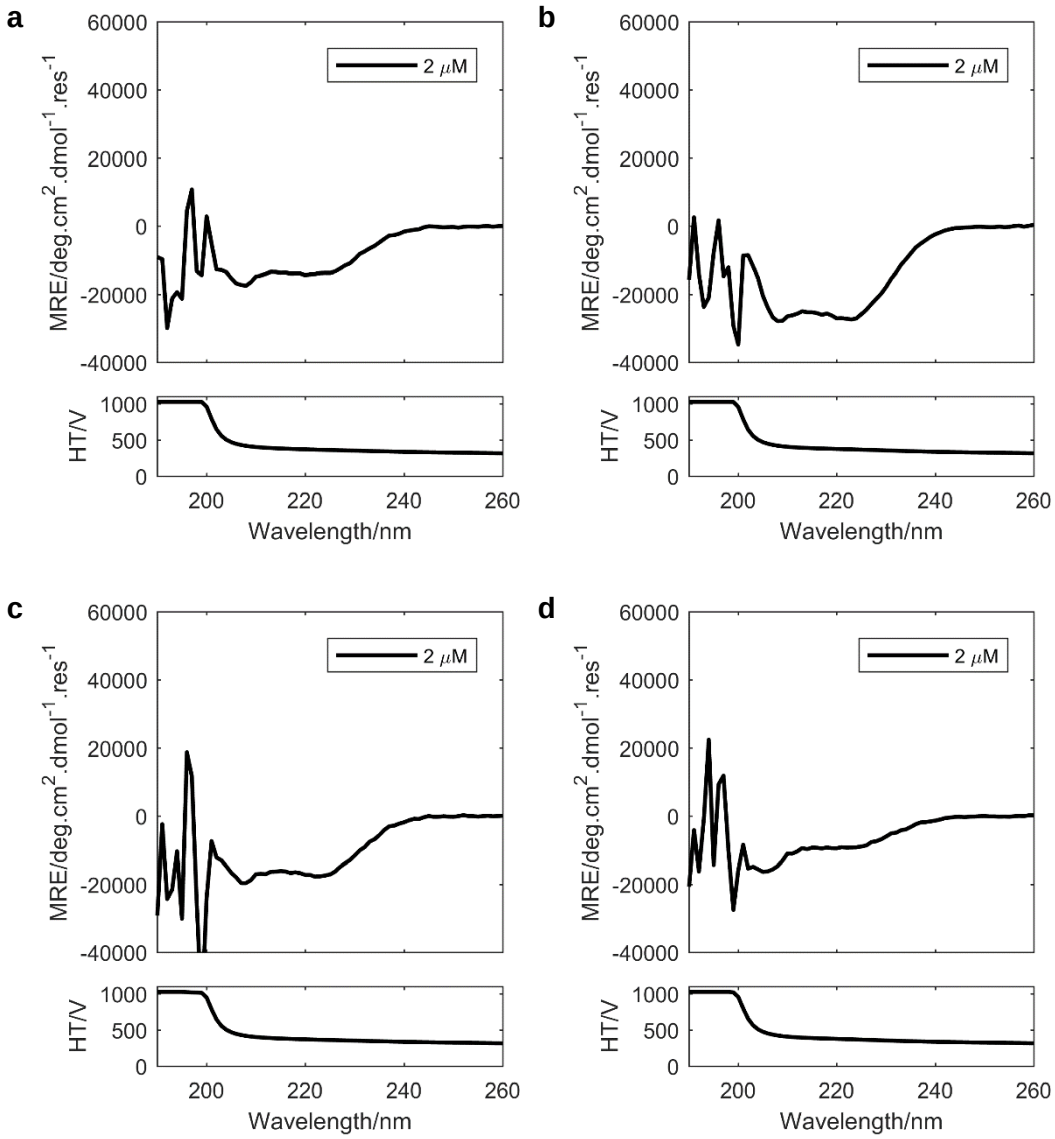


Figure 8-47 CD spectroscopy data for ABCD peptides: (a) 5; (b) 6; (c) 7; (d) 8.

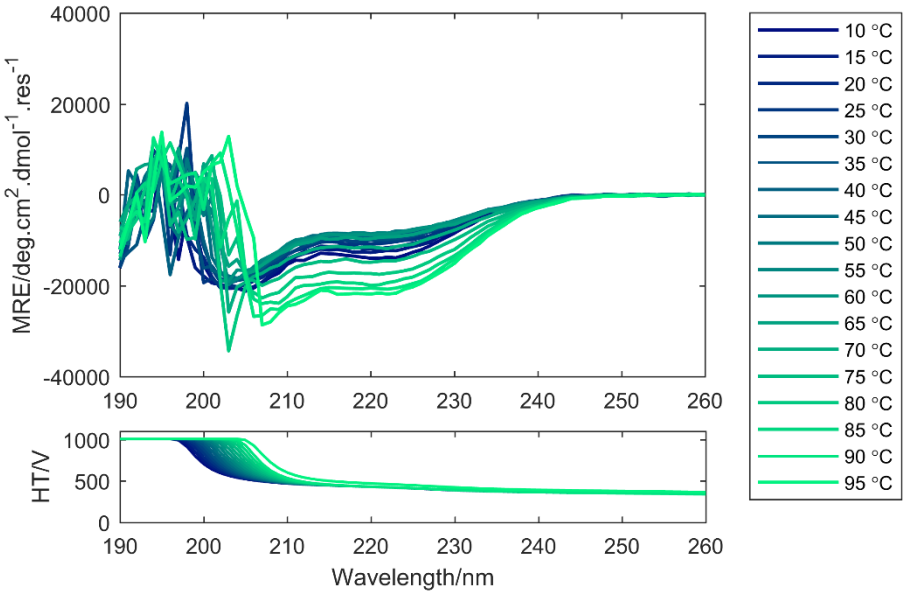


Figure 8-48 CD spectra for peptide **37** aka 1-LI-B measured at 5 °C intervals between 10 and 95 °C. Peptide concentration 10 µM.

8.4 CD Data of designed heteromers

This section contains representative CD spectra (left or centre) and thermal denaturation experiments (right) for all ABAB and ABCD heteromeric coiled coils discussed in this thesis. Peptide concentrations are indicated and each peptide in the mixture is present at the stated concentration. All ABAB spectra were measured at 5 °C and all ABCD spectra were measured at 20 °C. Thermal denaturation experiments were performed by heating (5–95 °C) and cooling samples (95–5 °C) while monitoring MRE_{222}. All measurements were performed in 1X PBS. High tension (HT) plots are shown below all CD spectra and have that same keys as the associated spectra.

8.4.1 ABAB heteromers

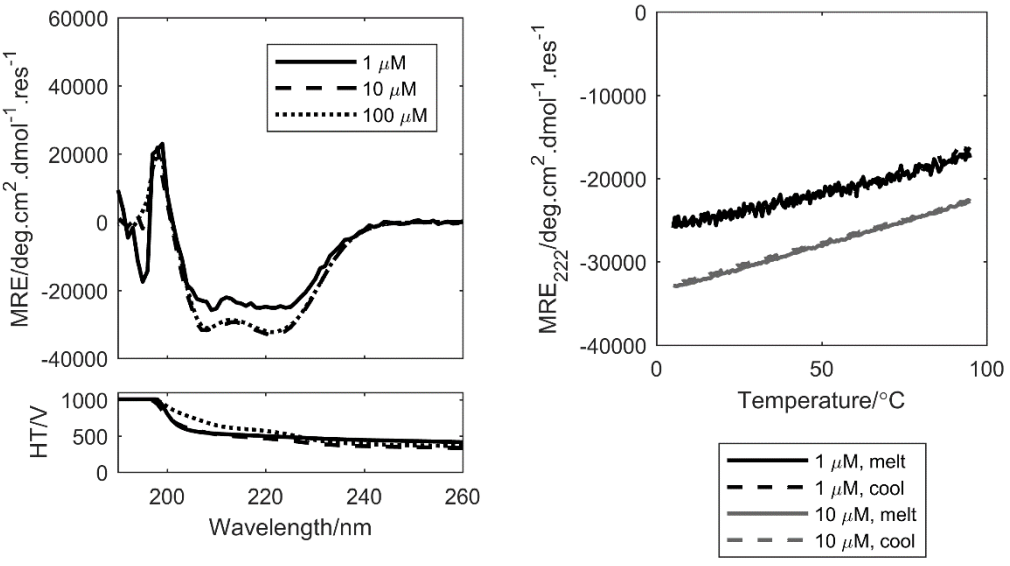


Figure 8-49 CD spectra (left) and temperature-dependent CD signal (right) for heteromer 1-LI-AB.

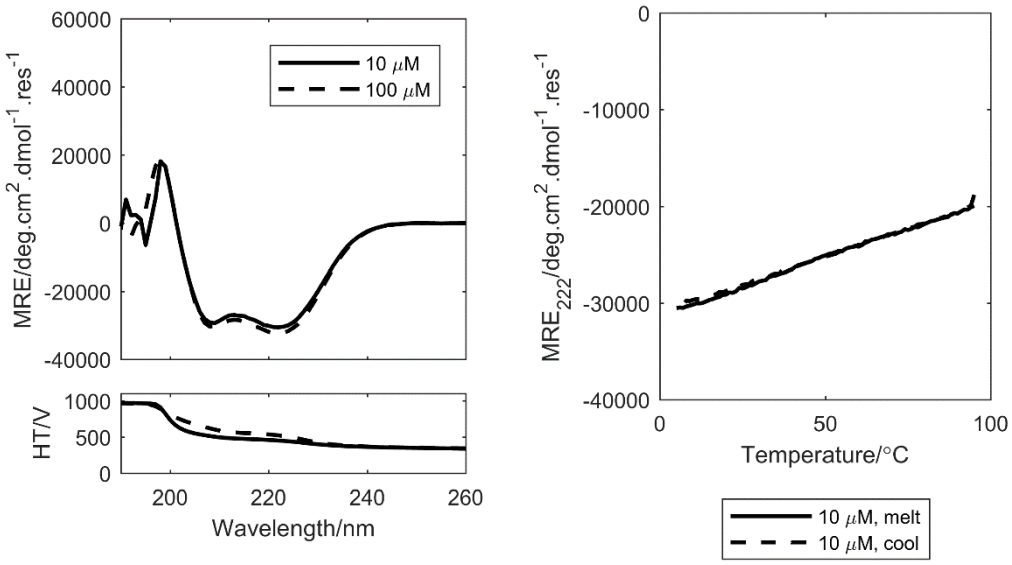


Figure 8-50 CD spectra (left) and temperature-dependent CD signal (right) for heteromer 1-LI-AB-*g*.

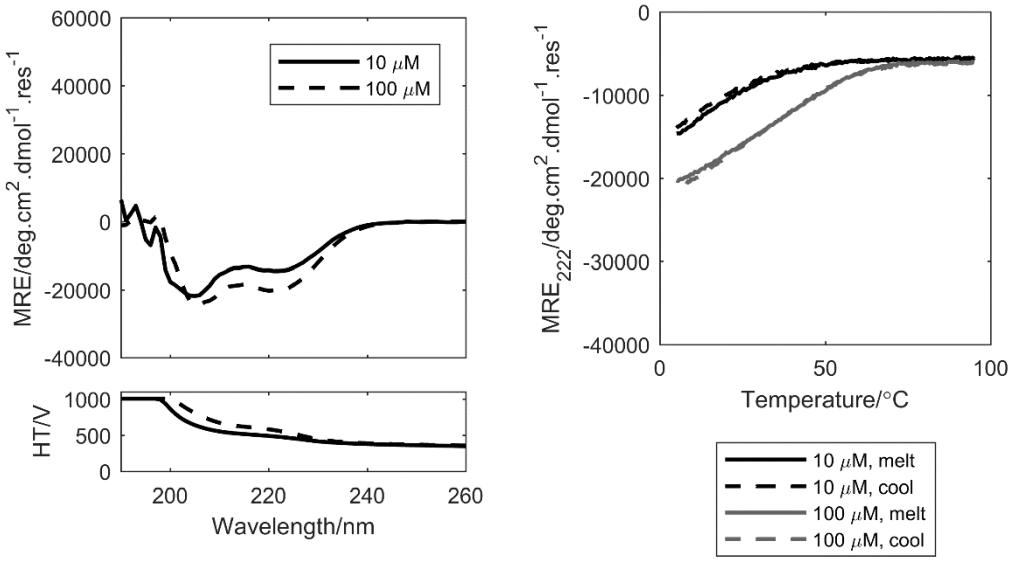


Figure 8-51 CD spectra (left) and temperature-dependent CD signal (right) for heteromer 1-LI-AB*.

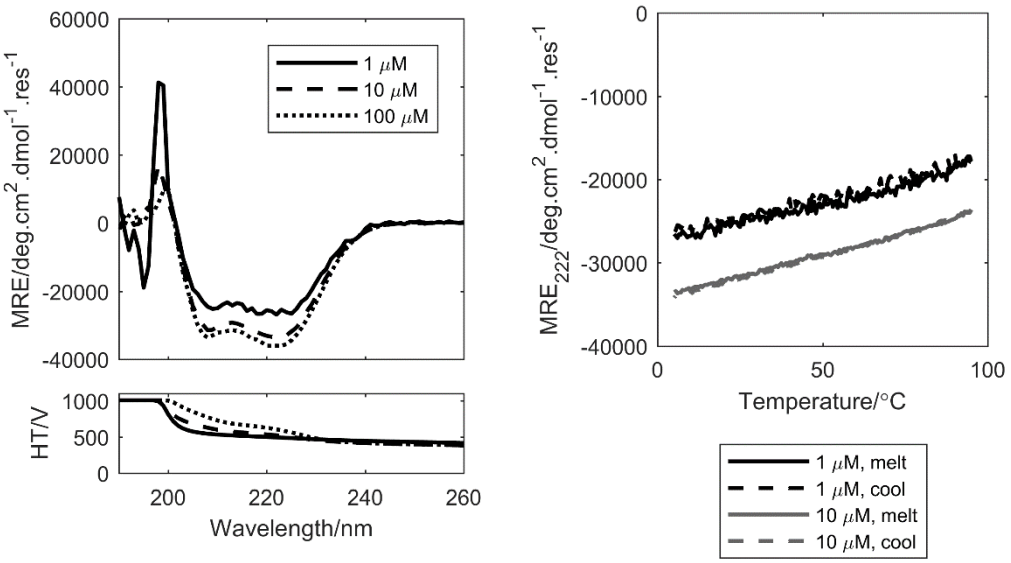


Figure 8-52 CD spectra (left) and temperature-dependent CD signal (right) for heteromer 2-LI-AB.

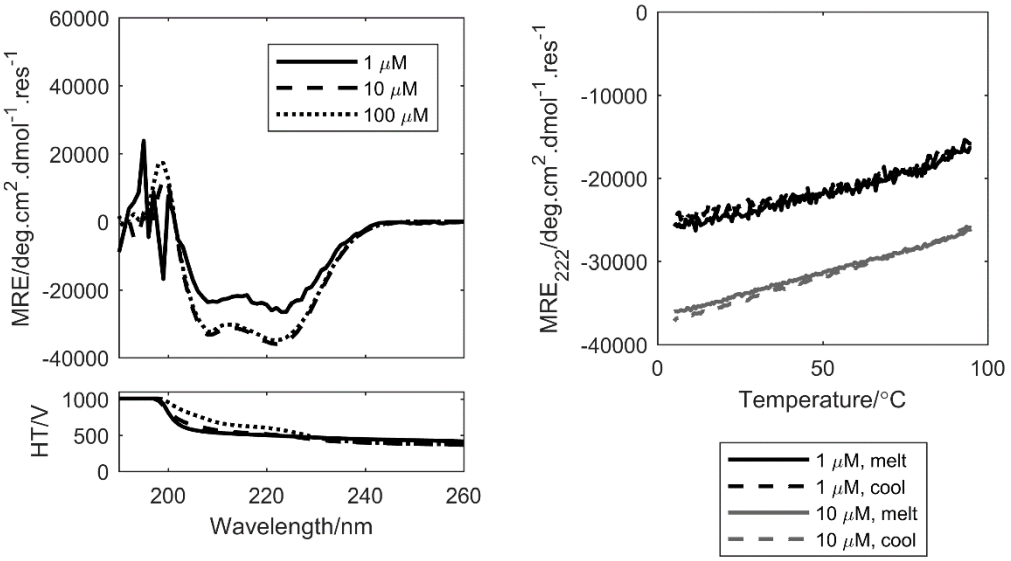


Figure 8-53 CD spectra (left) and temperature-dependent CD signal (right) for heteromer 3-LI-AB.

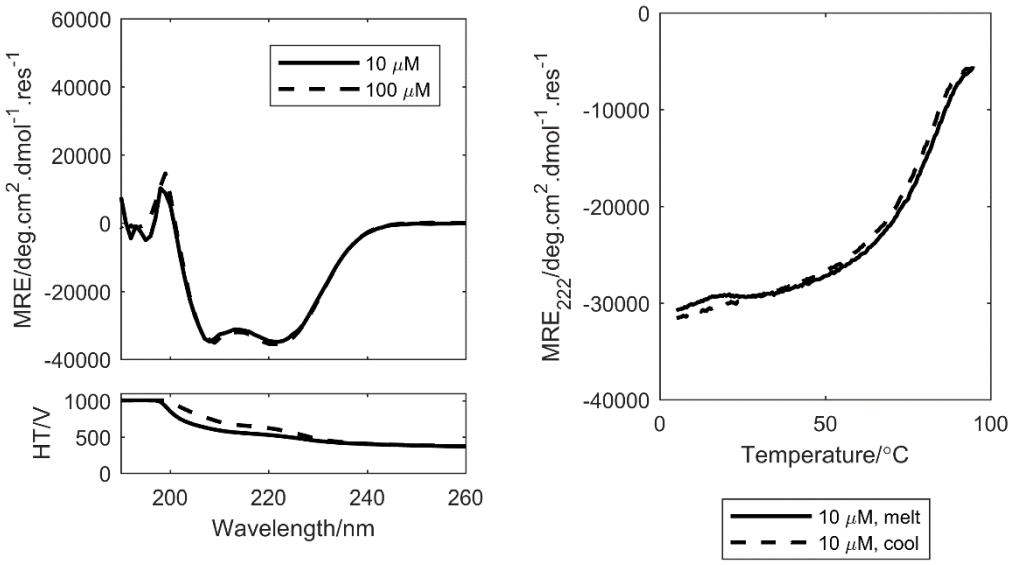


Figure 8-54 CD spectra (left) and temperature-dependent CD signal (right) for heteromer 1-LV-AB.

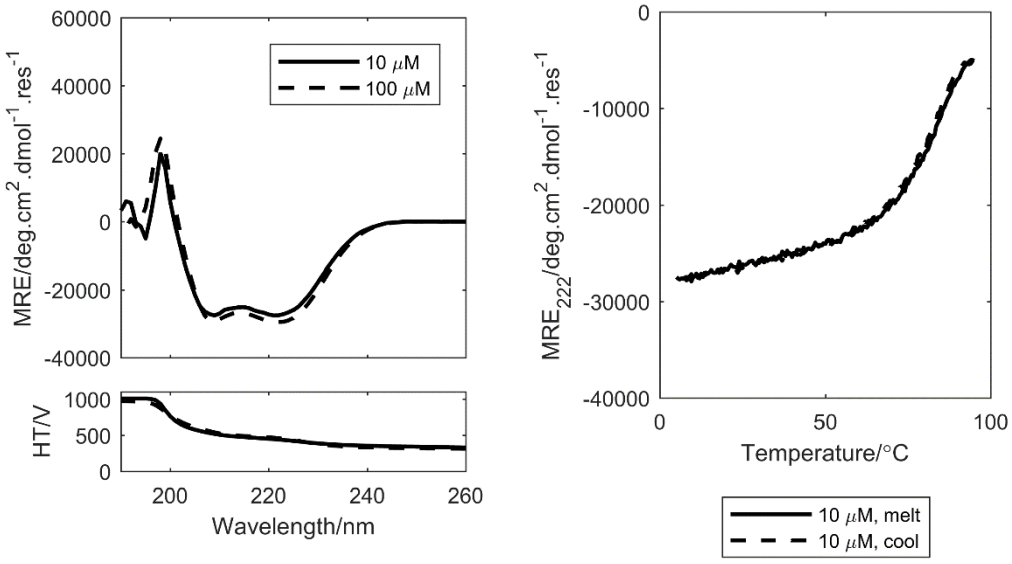


Figure 8-55 CD spectra (left) and temperature-dependent CD signal (right) for heteromer 2-LV-AB.

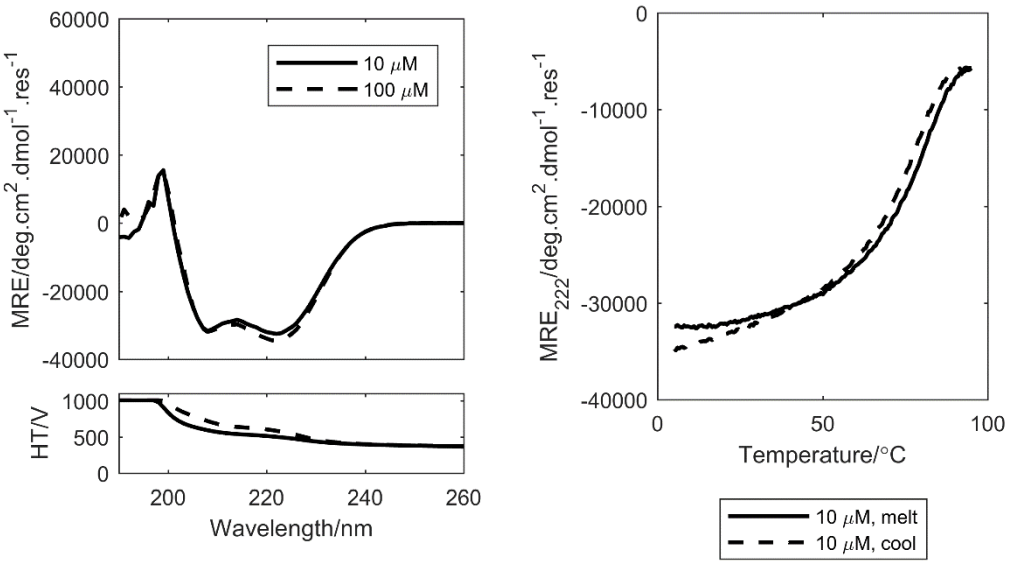


Figure 8-56 CD spectra (left) and temperature-dependent CD signal (right) for heteromer 3-LV-AB.

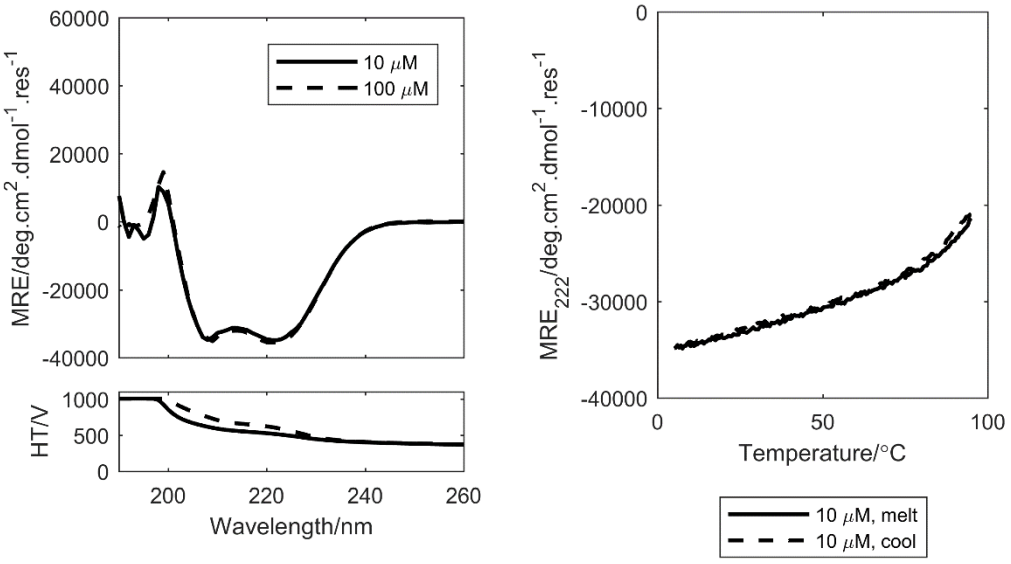


Figure 8-57 CD spectra (left) and temperature-dependent CD signal (right) for heteromer 1-LI-A/1-LV-B.

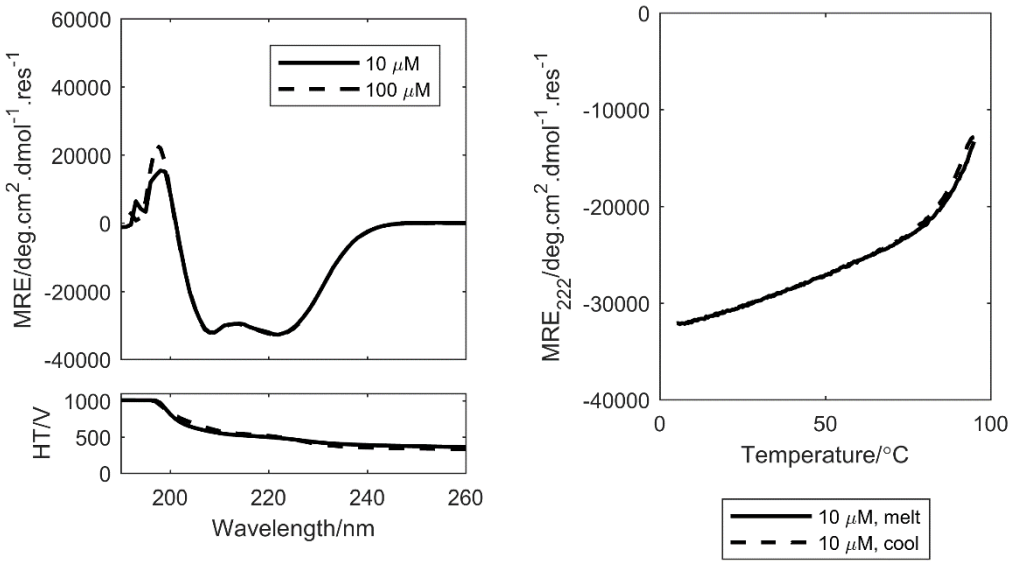


Figure 8-58 CD spectra (left) and temperature-dependent CD signal (right) for heteromer 1-L1-A/2-LV-B.

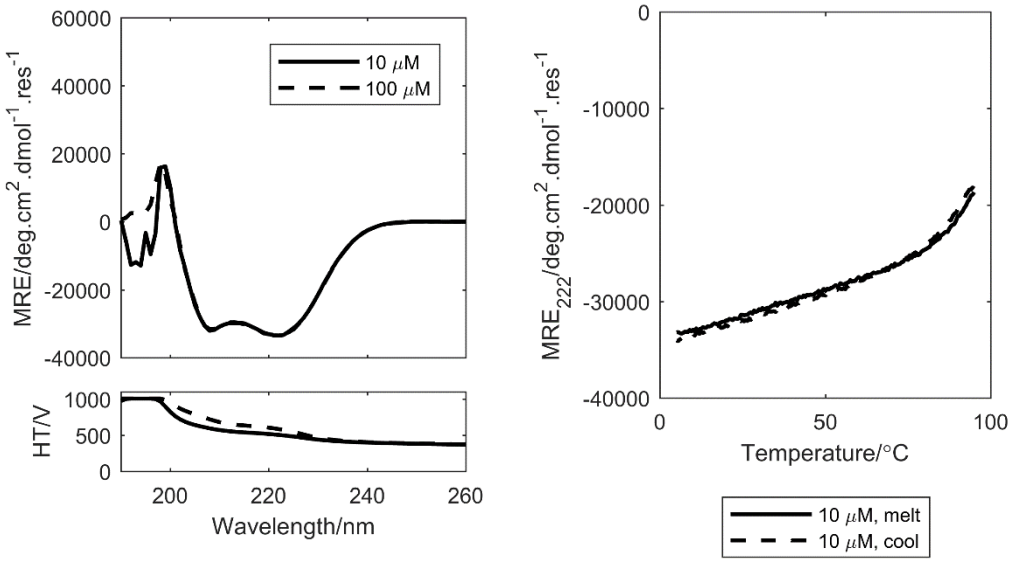


Figure 8-59 CD spectra (left) and temperature-dependent CD signal (right) for heteromer 1-LI-A/3-LV-B.

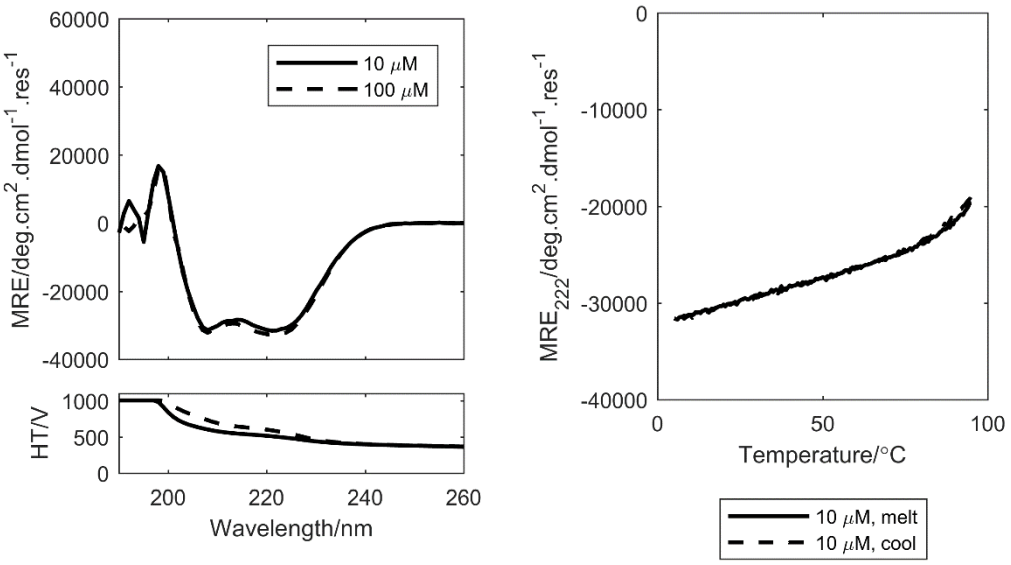


Figure 8-60 CD spectra (left) and temperature-dependent CD signal (right) for heteromer 1-LV-A/1-LI-B.

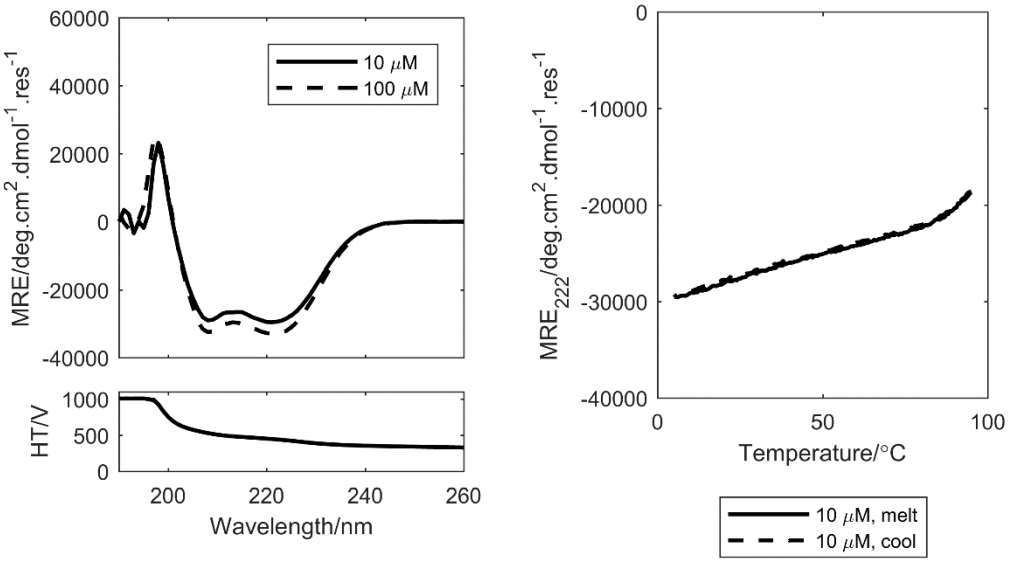


Figure 8-61 CD spectra (left) and temperature-dependent CD signal (right) for heteromer 2-LV-A/1-LI-B.

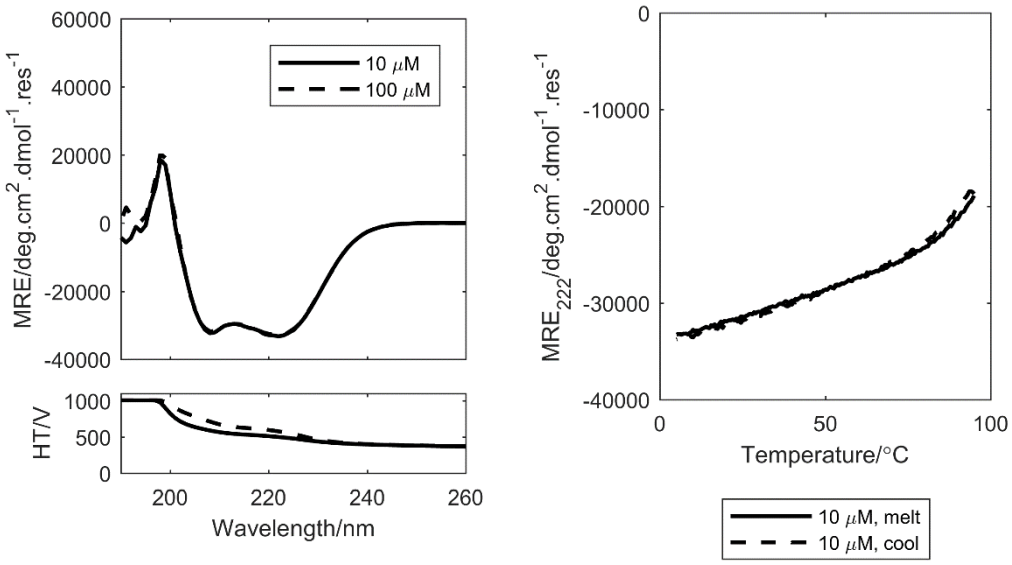


Figure 8-62 CD spectra (left) and temperature-dependent CD signal (right) for heteromer 3-LV-A/1-LI-B.

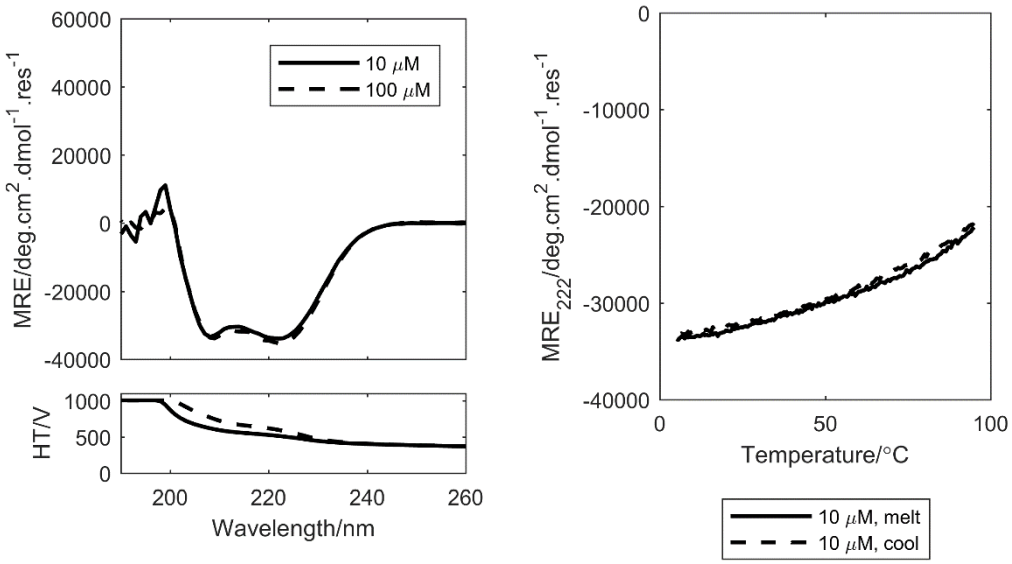


Figure 8-63 CD spectra (left) and temperature-dependent CD signal (right) for heteromer 2-LI-A/2-LV-B.

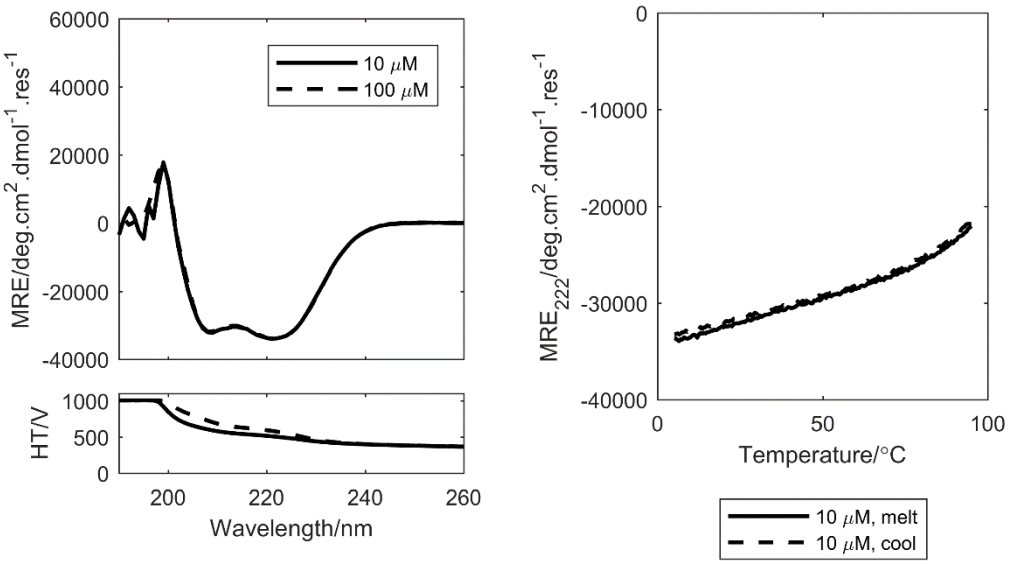


Figure 8-64 CD spectra (left) and temperature-dependent CD signal (right) for heteromer 2-LV-A/2-LI-B.

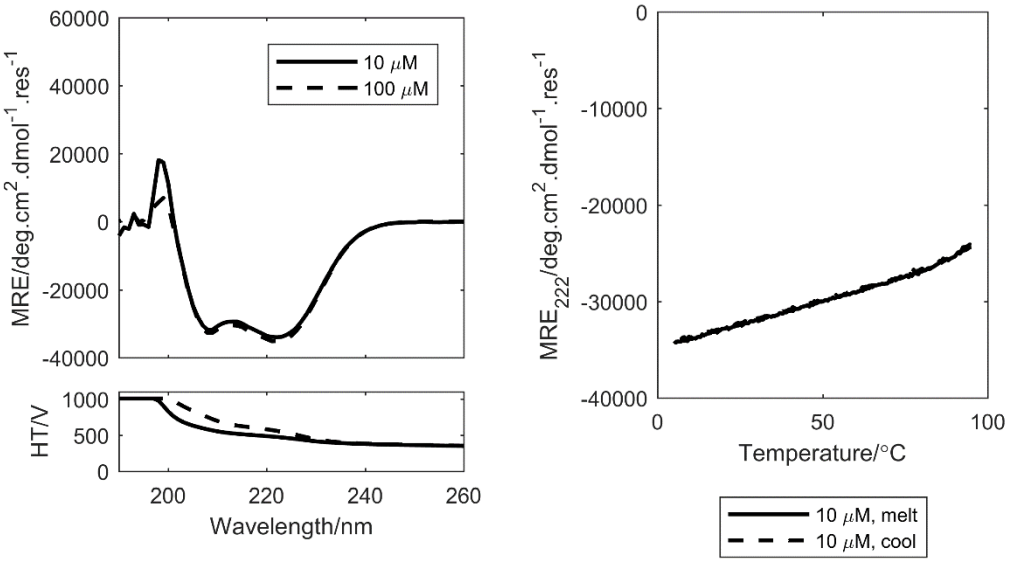


Figure 8-65 CD spectra (left) and temperature-dependent CD signal (right) for heteromer 3-LI-A/1-LV-B.

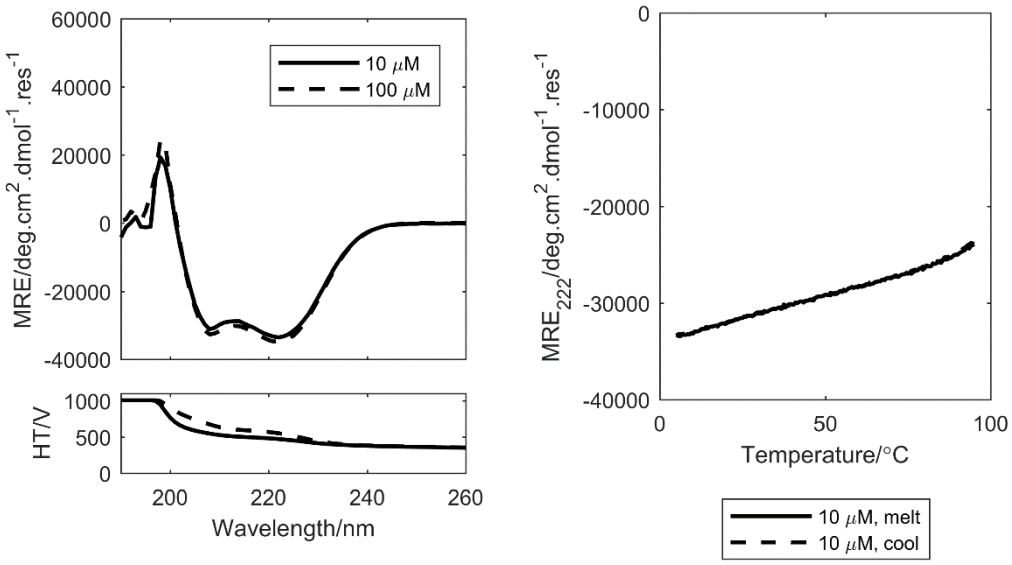


Figure 8-66 CD spectra (left) and temperature-dependent CD signal (right) for heteromer 3-LV-A/3-LI-B.

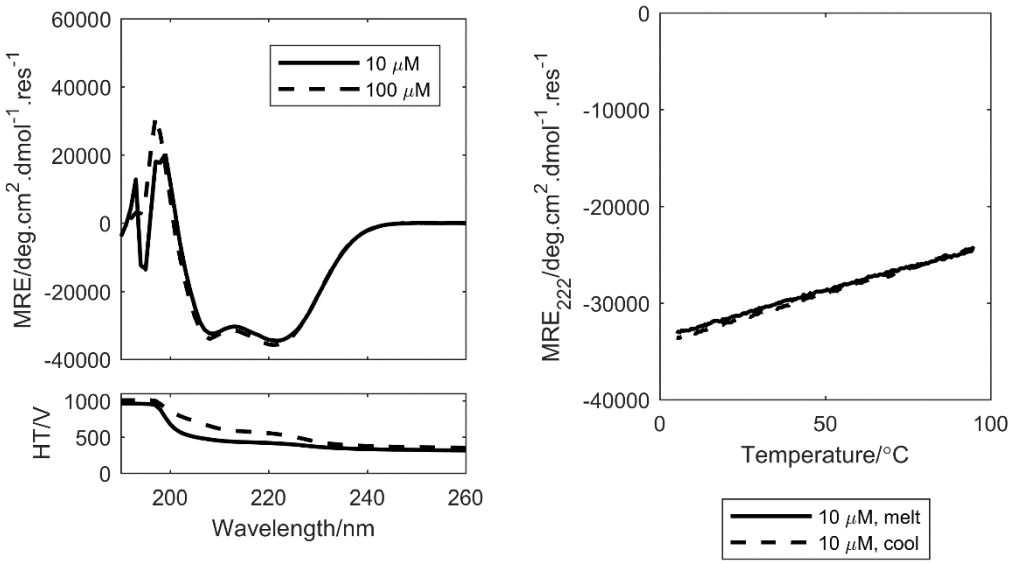


Figure 8-67 CD spectra (left) and temperature-dependent CD signal (right) for heteromer 2-LIA-AB.

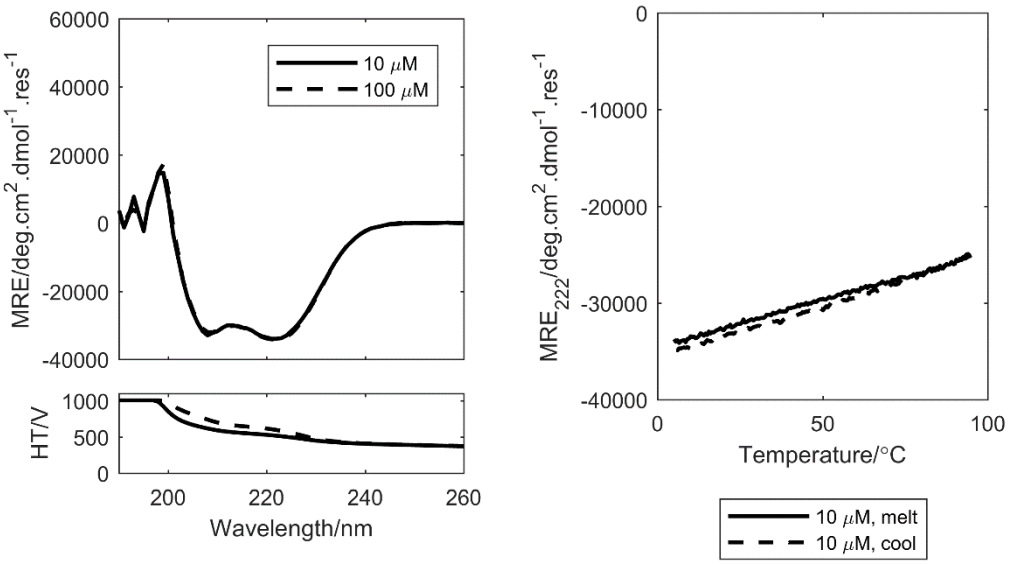


Figure 8-68 CD spectra (left) and temperature-dependent CD signal (right) for heteromer 2-LVA-AB.

8.4.2 1-LV-AB CD spectra at different salt concentrations

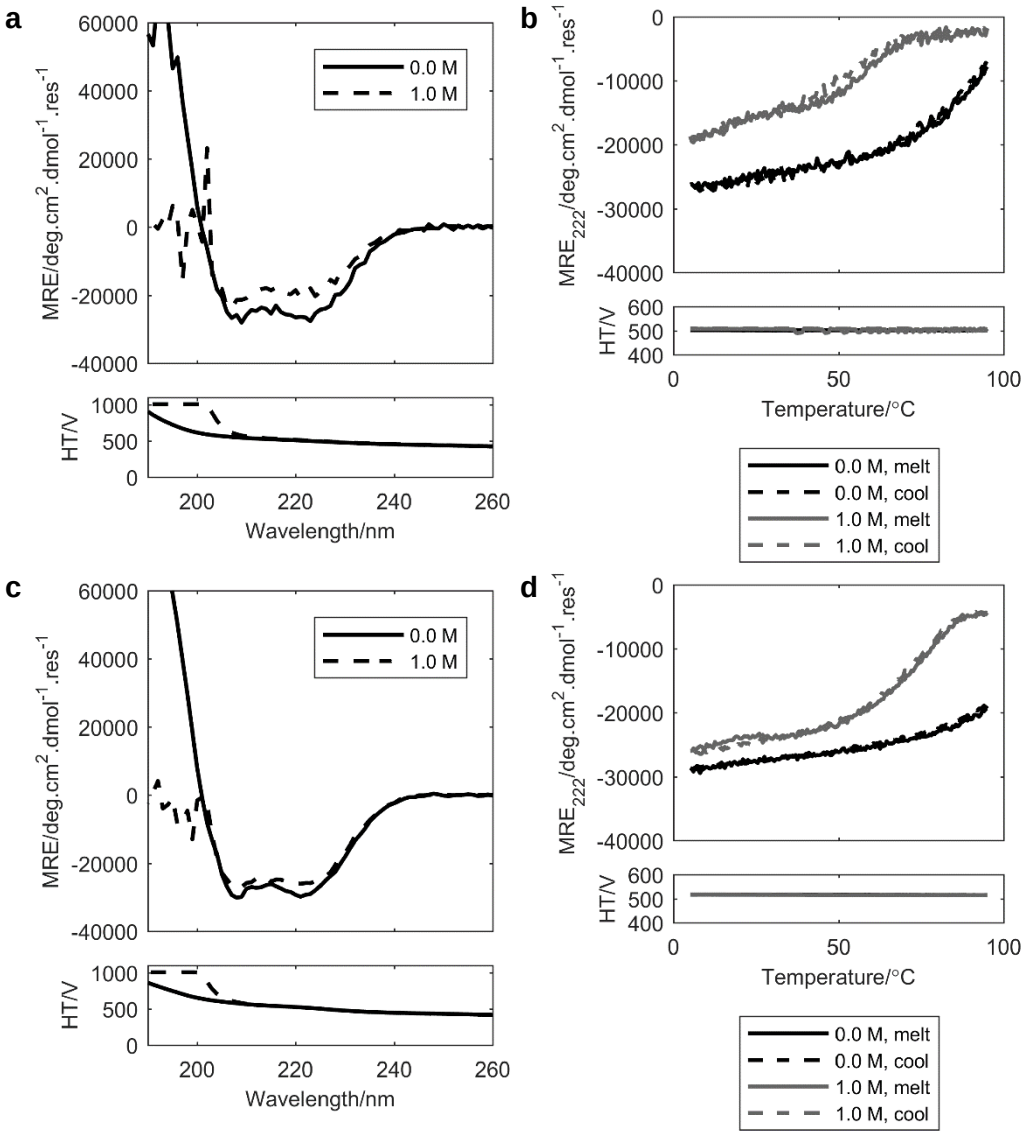


Figure 8-69 CD spectra (top) and HT traces (bottom) for heterotetramer 1-LV-AB where each peptide was present at (a) 1 µM or (c) 5 µM each. Thermal denaturation experiments (top) and temperature dependent HT traces (bottom) for heterotetramer 1-LV-AB where each peptide was present at 1 (b) or 5 (d) µM each. Measurements were performed in 10 mM sodium phosphate buffer in the presence of 0.0 or 1.0 M NaCl.

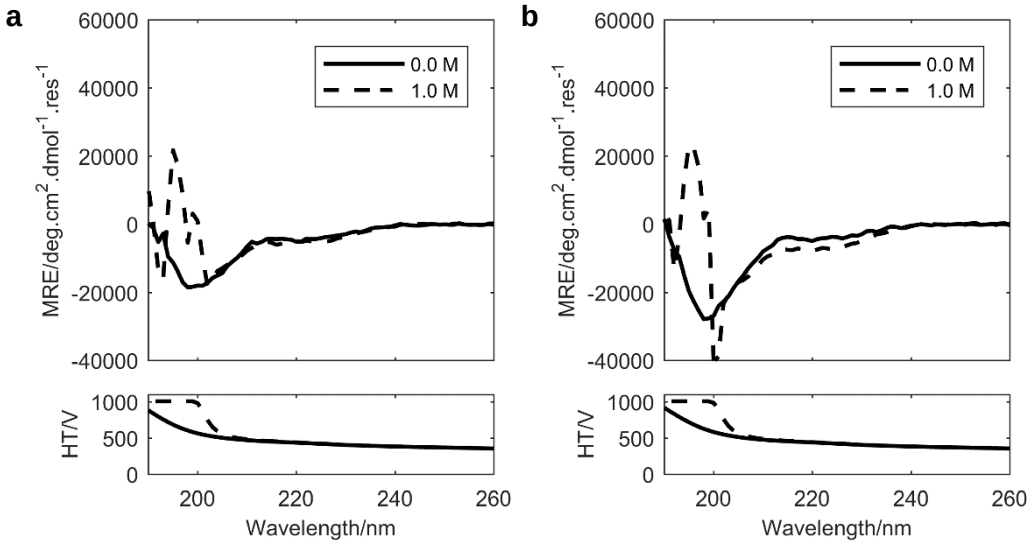


Figure 8-70 CD spectra (top) and HT traces (bottom) for (a) 1-LV-A and (b) 1-LV-B where each peptide was present at 5 μM. Measurements were performed in 10 mM sodium phosphate buffer in the presence of 0.0 or 1.0 M NaCl.

8.4.3 ABCD heteromers and cross interactions

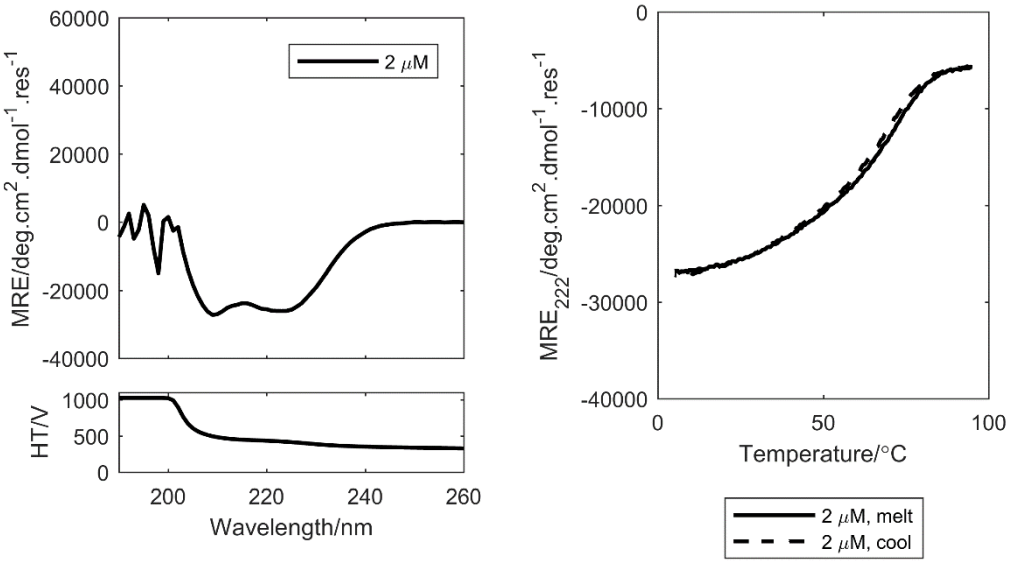


Figure 8-71 CD spectra at 20 °C (left) and temperature-dependent CD signal (right) for ABCD heteromer 1234.

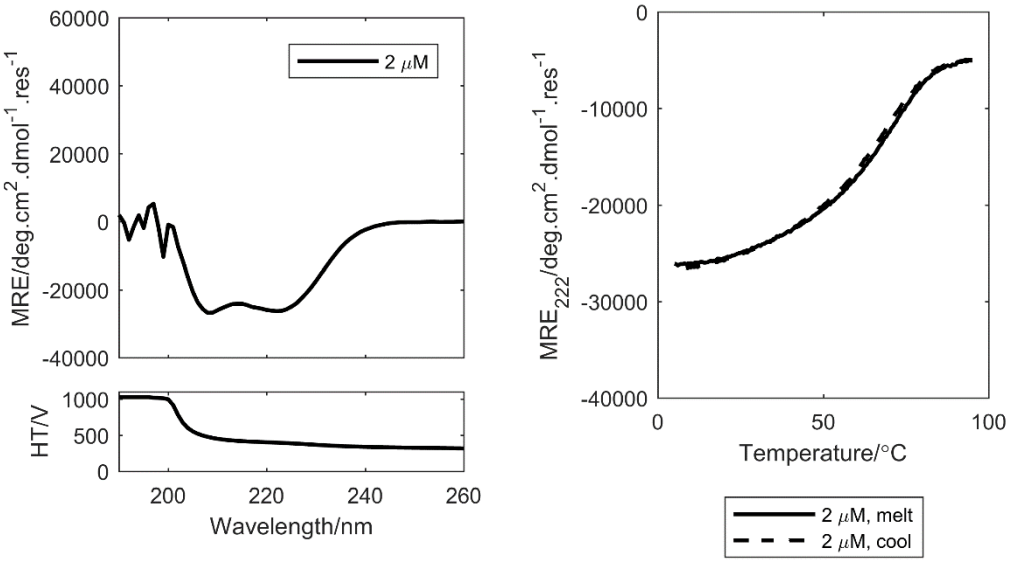


Figure 8-72 CD spectra at 20 °C (left) and temperature-dependent CD signal (right) for ABCD heteromer 5678.

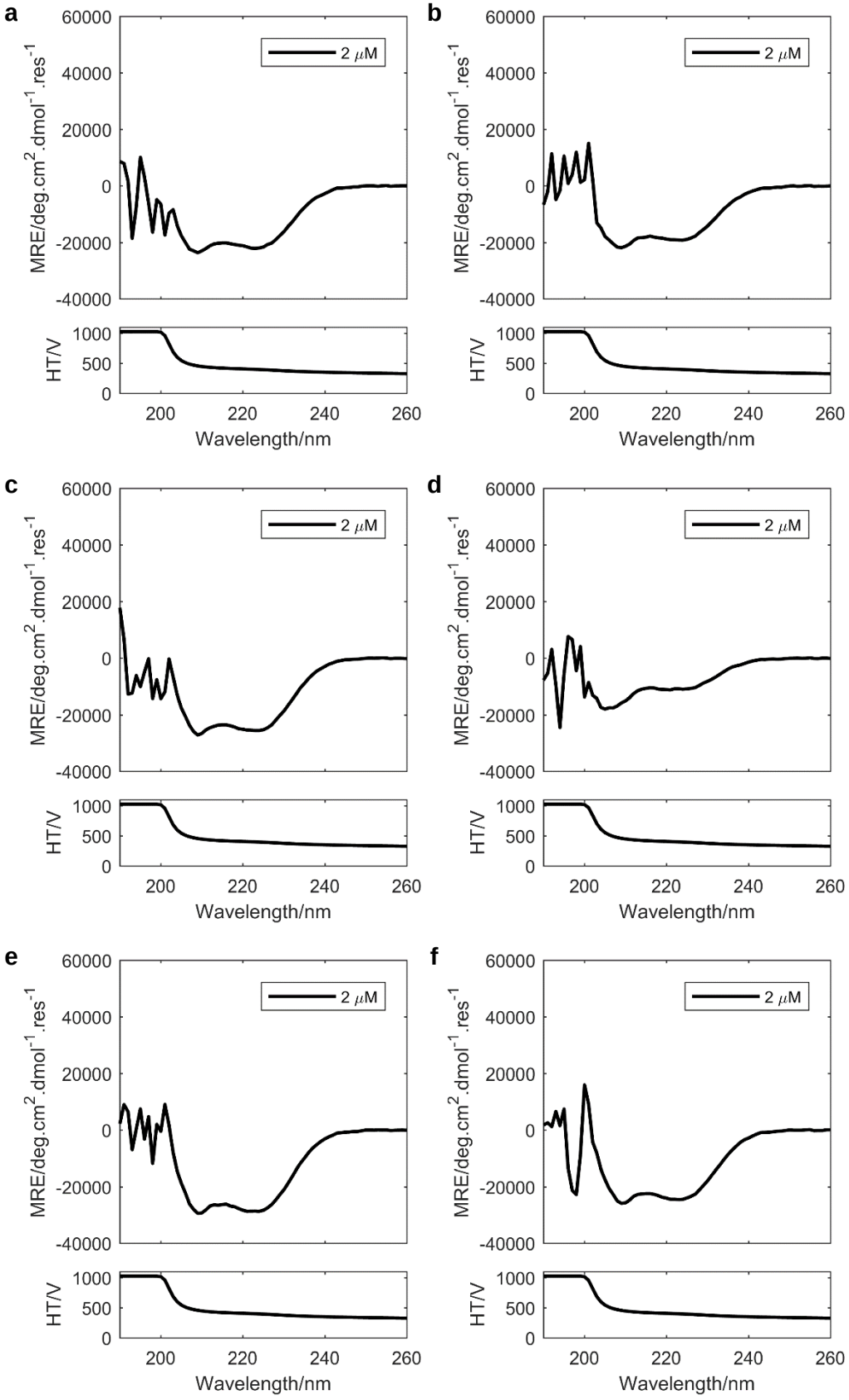


Figure 8-73 CD spectra at 20 °C for off-target ABCD heteromers 12 (a), 13 (b), 14 (c), 23 (d), 24 (e) and 34 (f).

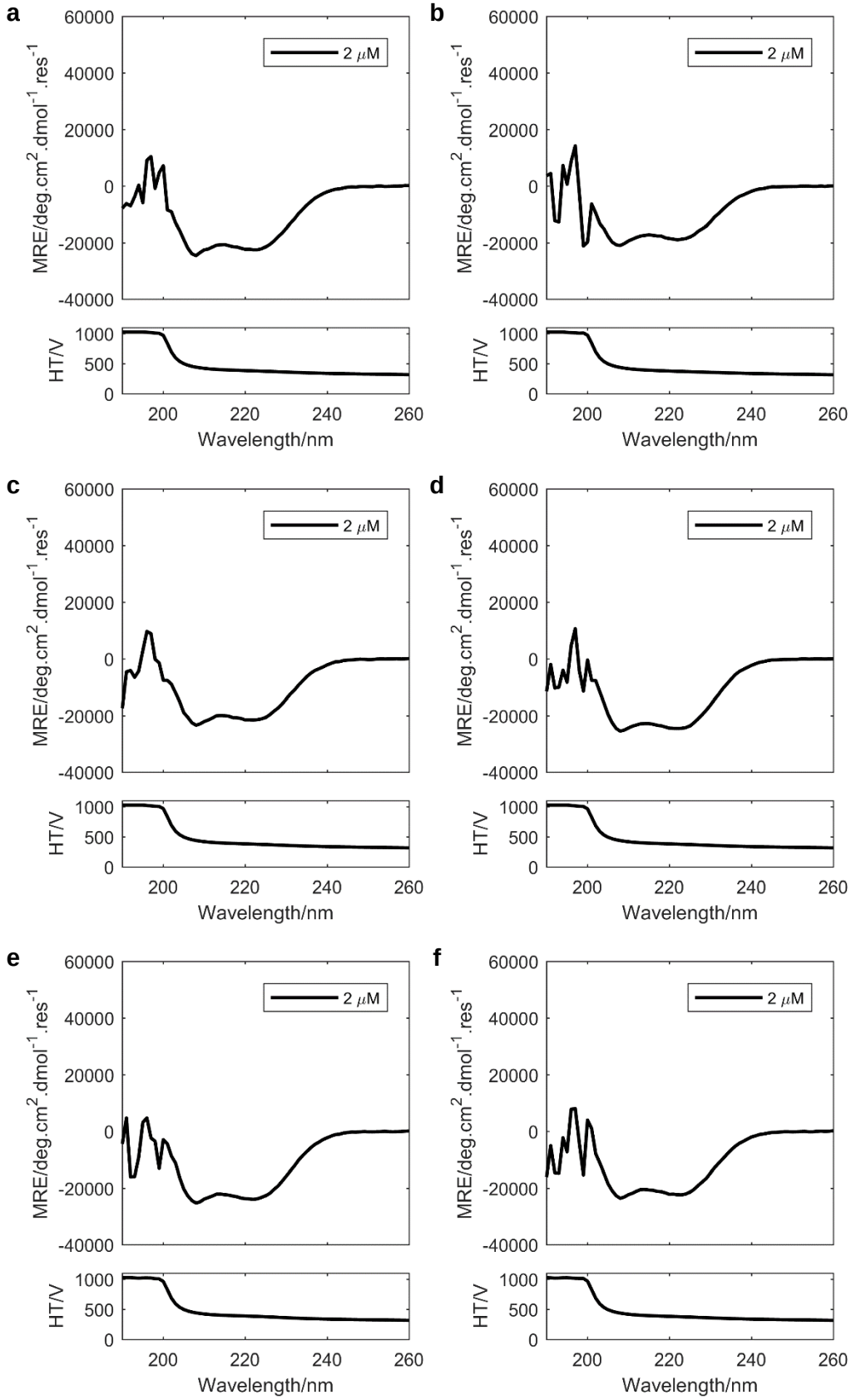


Figure 8-74 CD spectra at 20 °C for off-target ABCD heteromers 56 (a), 57 (b), 58 (c), 67 (d), 68 (e) and 78 (f).

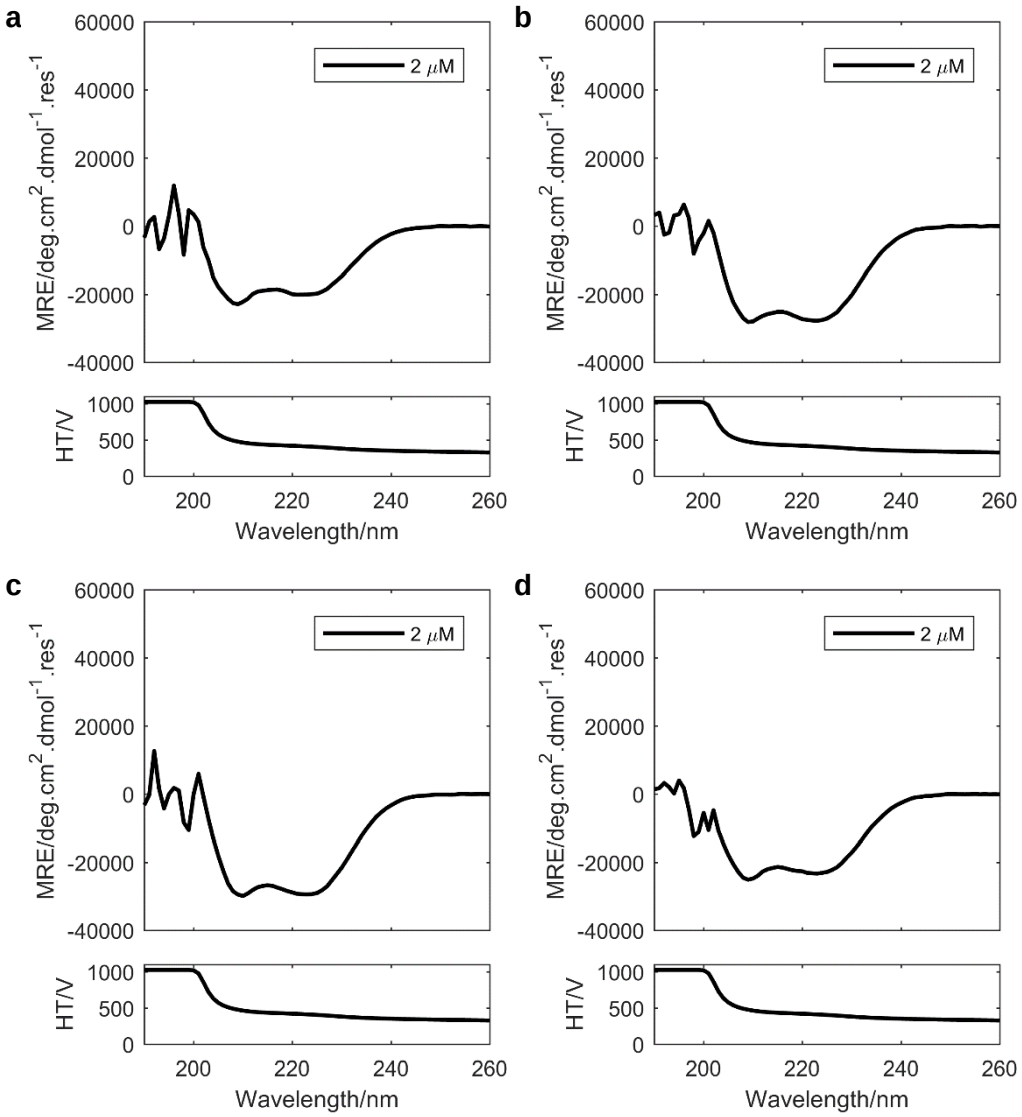


Figure 8-75 CD spectra at 20 °C for off-target ABCD heteromers 123 (a), 124 (b), 134 (c) and 234 (d).

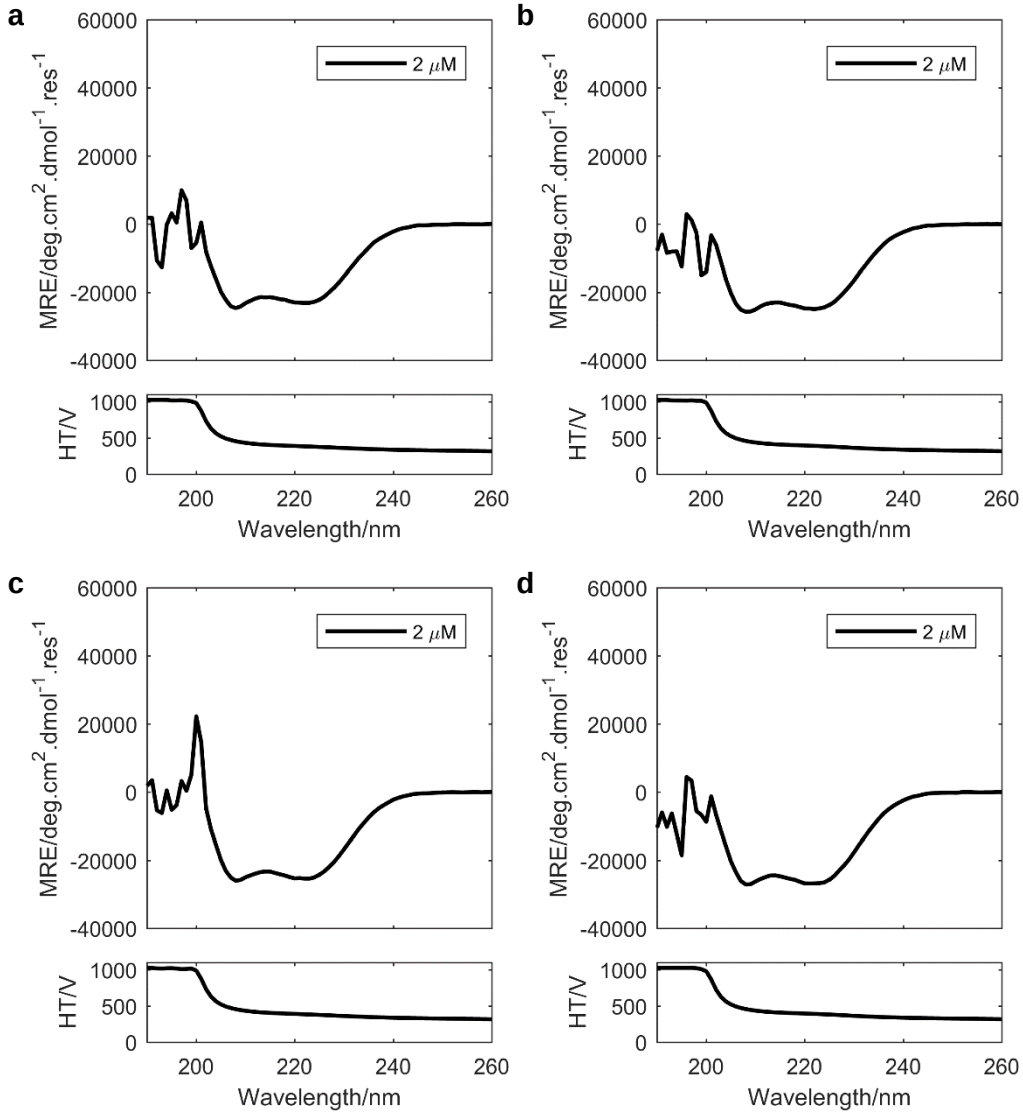


Figure 8-76 CD spectra at 20 °C for off-target ABCD heteromers 567 (a), 568 (b), 578 (c) and678 (d).

8.5 Analytical ultracentrifugation data

8.5.1 Homomers

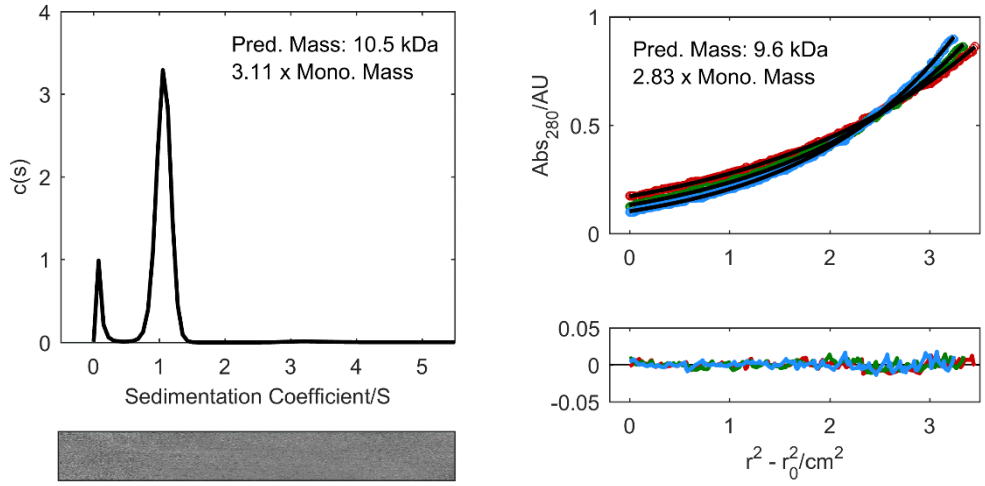


Figure 8-77 AUC data for CC-Tet-KE ($\bar{v}$ = 0.7696 $cm^3.g^{-1}$). Left: sedimentation velocity (SV) continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.06, $s_{20,w}$ =1.09, f/f_0 = 1.33 and Mw = 10491 Da (3.11 x monomer mass). Right, top: Sedimentation equilibrium (SE) data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 33 krpm (green) and 36 krpm (blue), returning Mw = 9561 Da (2.83 x monomer mass, 99 % confidence limits: 9502-9619 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

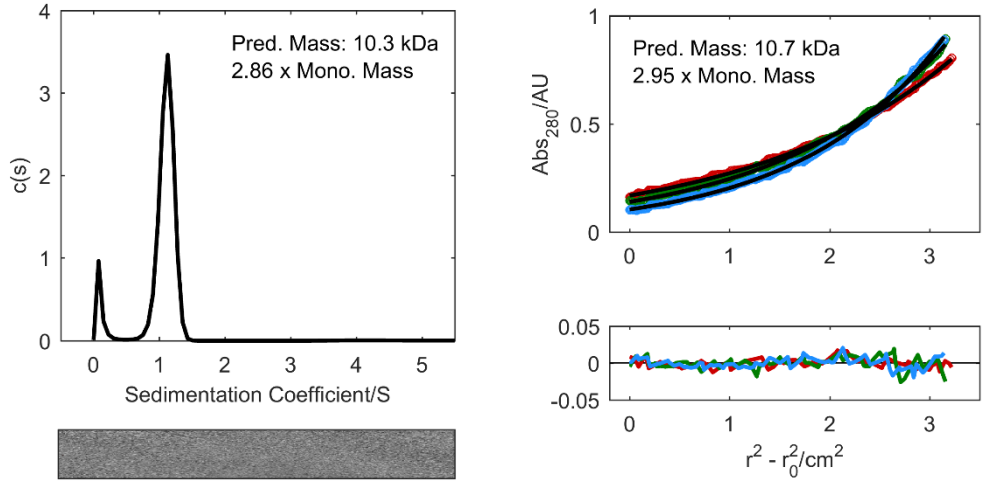


Figure 8-78 AUC data for CC-Tet-KE-N-4.5 ($\bar{v}$ = 0.7673 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.11, $s_{20,w}$ = 1.14, f/f_0 = 1.28 and Mw = 10349 Da (2.86 x monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 33 krpm (green) and 36 krpm (blue), returning Mw = 10650 Da (2.95 x monomer mass, 99 % confidence limits: 10581-10721 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

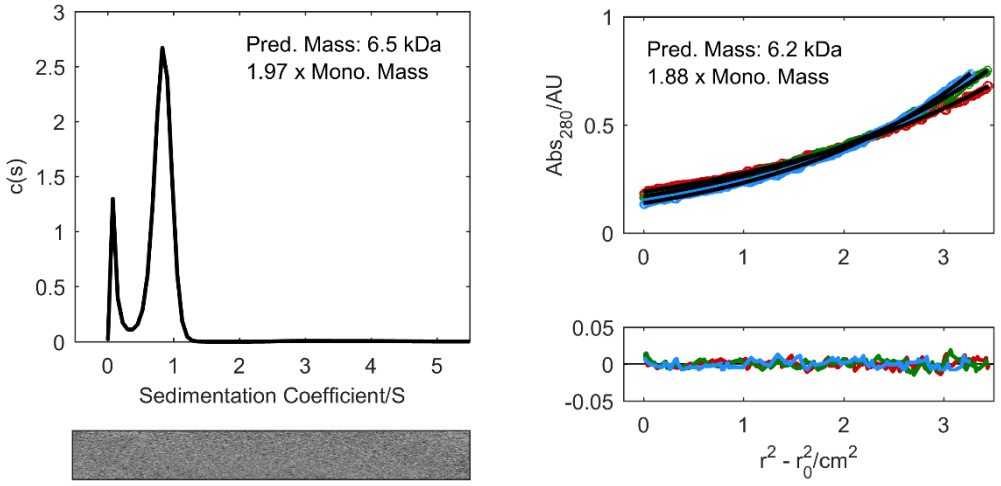


Figure 8-79 AUC data for CC-Tet-IA ($\bar{v}$ = 0.7595 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 0.83, $s_{20,w}$ = 0.85, f/f_0 = 1.30 and Mw = 6485 Da (1.97 x monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 33 krpm (red), 36 krpm (green) and 39 krpm (blue), returning Mw = 6188 Da (1.88 x monomer mass, 99 % confidence limits: 6137-6238 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

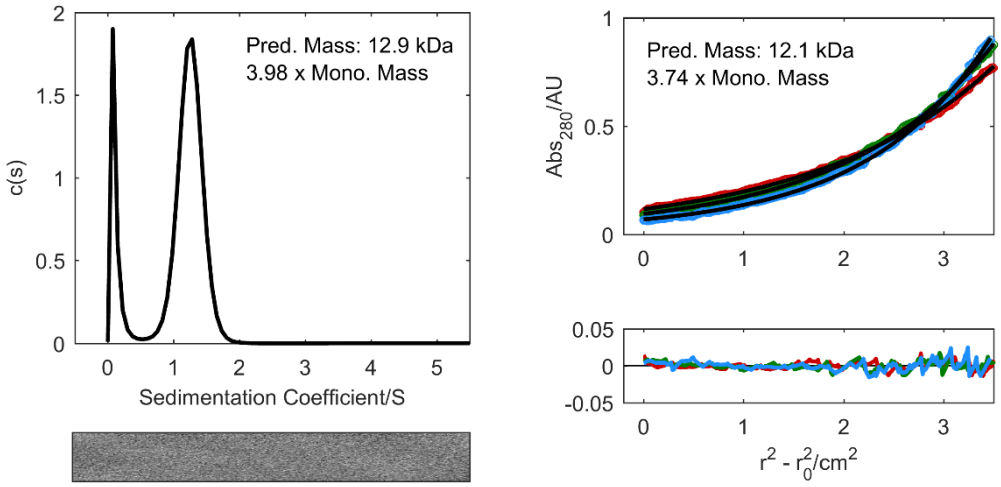


Figure 8-80 AUC data for 1-LI-EK ($\bar{v}$ = 0.7669 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.25, $s_{20,w}$ = 1.28, f/f_0 = 1.32 and Mw = 12931 Da (3.98 x monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 33 krpm (green) and 36 krpm (blue), returning Mw = 12140 Da (3.74 x monomer mass, 99 % confidence limits: 12091-12186 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

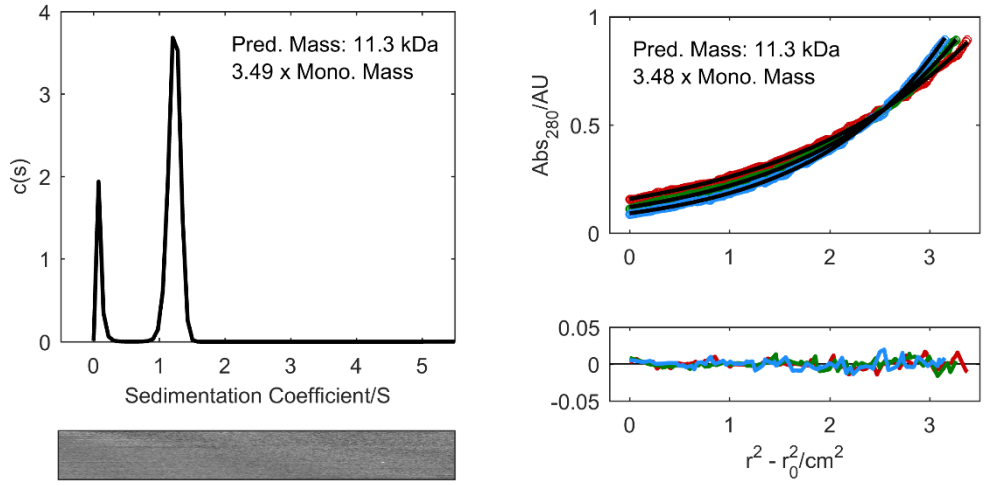


Figure 8-81 AUC data for 1-LI-KE ($\bar{v}$ = 0.7669 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.23, $s_{20,w}$ = 1.26, f/f_0 = 1.23 and Mw = 11328 Da (3.49 x monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 33 krpm (green) and 36 krpm (blue), returning Mw = 11310 Da (3.48 x monomer mass, 99 % confidence limits: 11253-11361 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

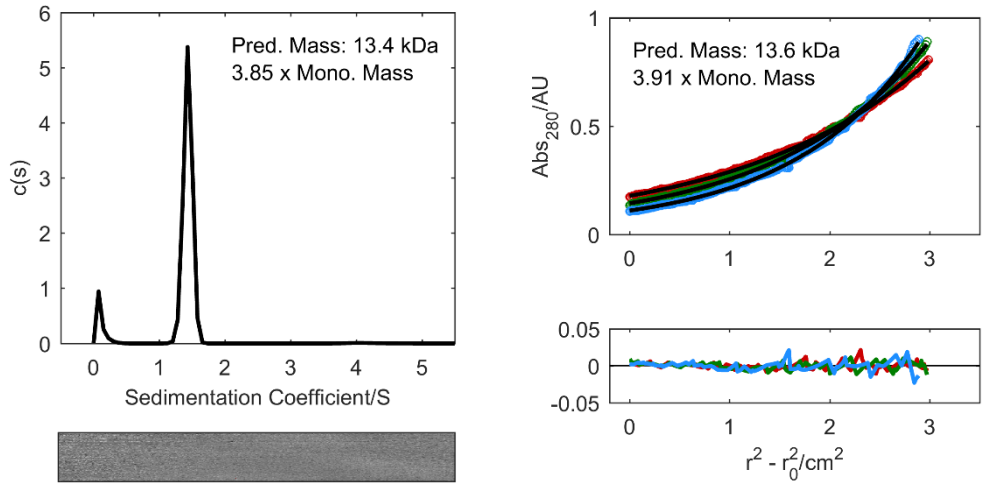


Figure 8-82 AUC data for 2-LI-EK ($\bar{v}$ = 0.7545 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.44, $s_{20,w}$ = 1.47, f/f_0 = 1.25 and Mw = 13381 Da (3.85 x monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 33 krpm (green) and 36 krpm (blue), returning Mw = 13600 Da (3.91 x monomer mass, 99 % confidence limits: 13506-13687 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

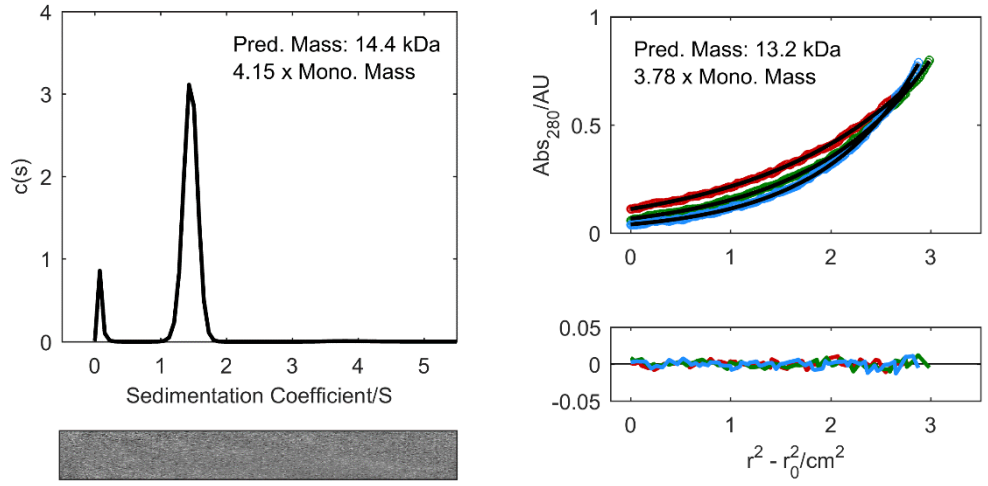


Figure 8-83 AUC data for 2-LI-KE ($\bar{v}$ = 0.7545 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.46, $s_{20,w}$ = 1.49, f/f_0 = 1.29 and Mw = 14405 Da (4.15 x monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 34 krpm (green) and 38 krpm (blue), returning Mw = 13150 Da (3.78 x monomer mass, 99 % confidence limits: 13092-13182 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

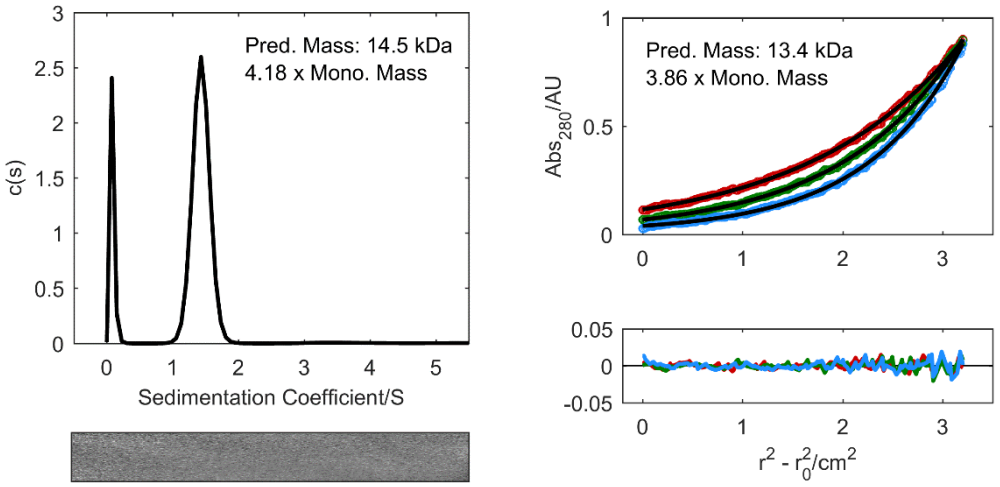


Figure 8-84 AUC data for 3-LI-EK ($\bar{v}$ = 0.7545 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.43, $s_{20,w}$ = 1.47, f/f_0 = 1.32 and Mw = 14523 Da (4.18 x monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 34 krpm (green) and 38 krpm (blue), returning Mw = 13420 Da (3.86 x monomer mass, 99 % confidence limits: 13338-13437 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

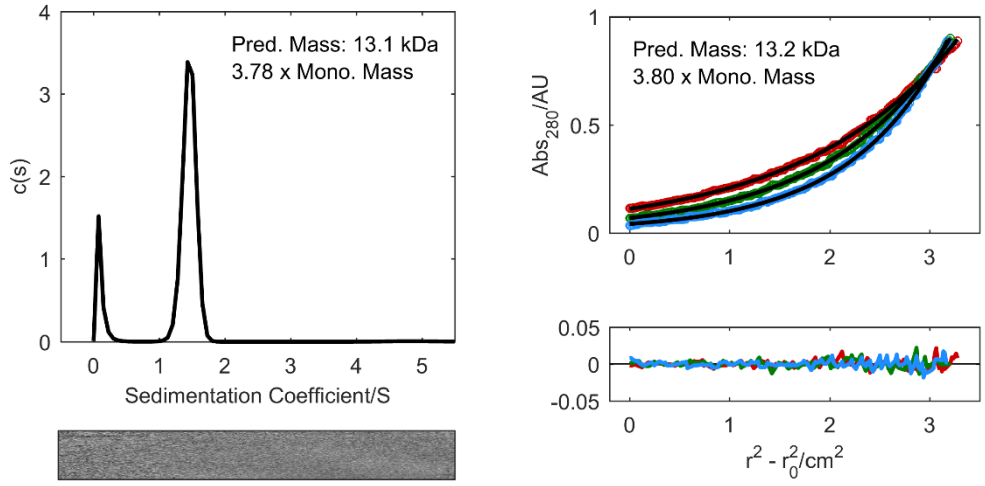


Figure 8-85 AUC data for 3-LI-KE ($\bar{v}$ = 0.7545 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.46, $s_{20,w}$ = 1.49, f/f_0 = 1.21 and Mw = 13119 Da (3.78 x monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 34 krpm (green) and 38 krpm (blue), returning Mw = 13220 Da (3.80 x monomer mass, 99 % confidence limits: 13190-13284 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

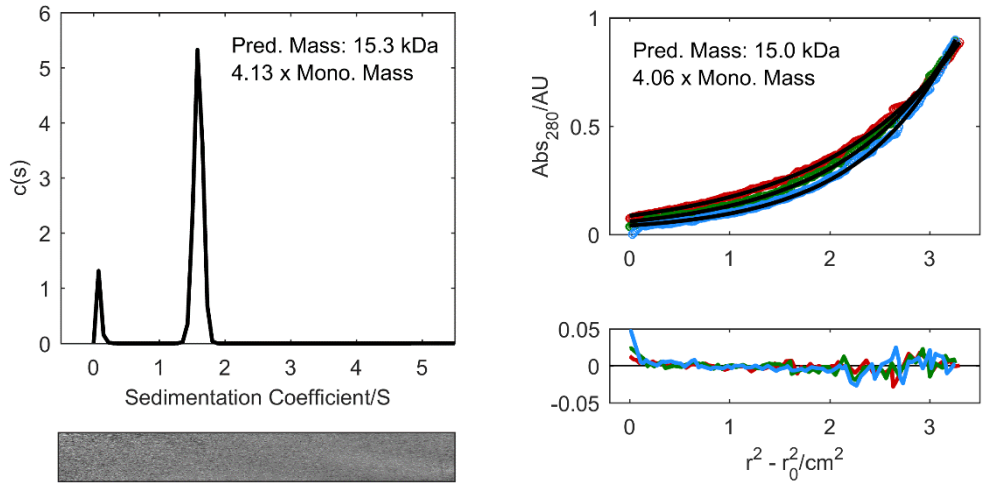


Figure 8-86 AUC data for 4-LI-EK ($\bar{v}$ = 0.7594 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.60, $s_{20,w}$ = 1.63 f/f_0 = 1.20 and Mw = 15289 Da (4.13 x monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 33 krpm (green) and 36 krpm (blue), returning Mw = 15040 Da (4.06 x monomer mass, 99 % confidence limits: 14990-15099 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

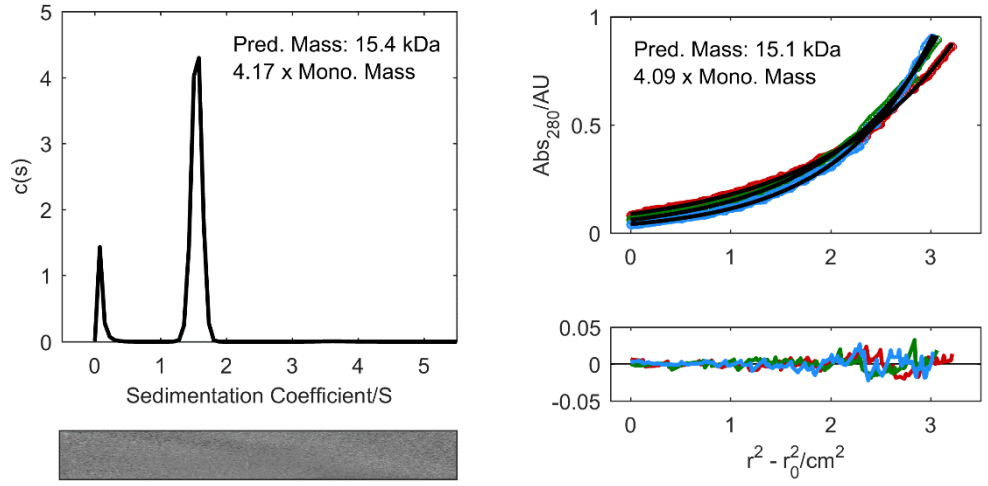


Figure 8-87 AUC data for 4-LI-KE ($\bar{v}$ = 0.7594 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.55, $s_{20,w}$ = 1.59, f/f_0 = 1.24 and Mw = 15433 Da (4.17 x monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 33 krpm (green) and 36 krpm (blue), returning Mw = 15140 Da (4.09 x monomer mass, 99 % confidence limits: 15074-15200 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

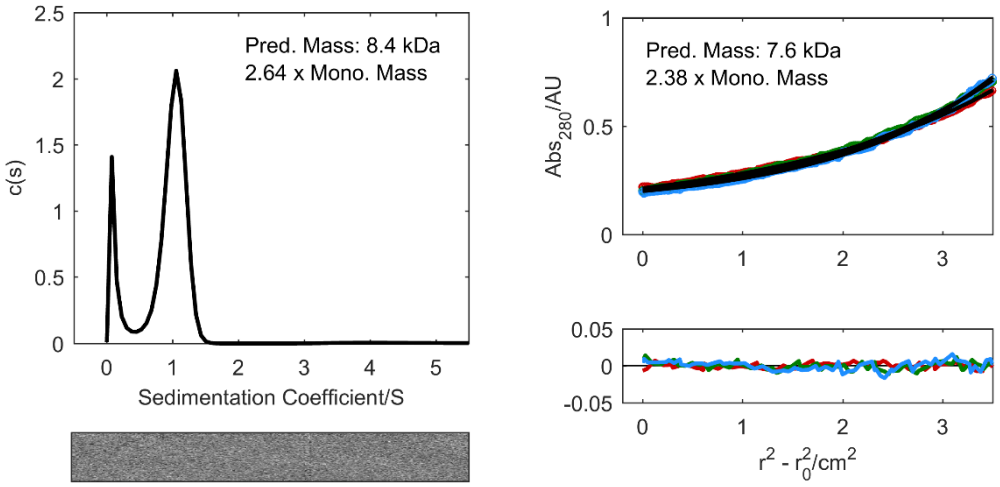


Figure 8-88 AUC data for 1-LV-KE ($\bar{v}$ = 0.7595 $cm^3.g^{-1}$). Left: sedimentation velocity continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.03, $s_{20,w}$ = 1.05, f/f_0 = 1.25 and Mw = 8410 Da (2.64 x monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 33 krpm (red), 36 krpm (green) and 39 krpm (blue), returning Mw = 7585 Da (2.38 x monomer mass, 99 % confidence limits: 7530-7639 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

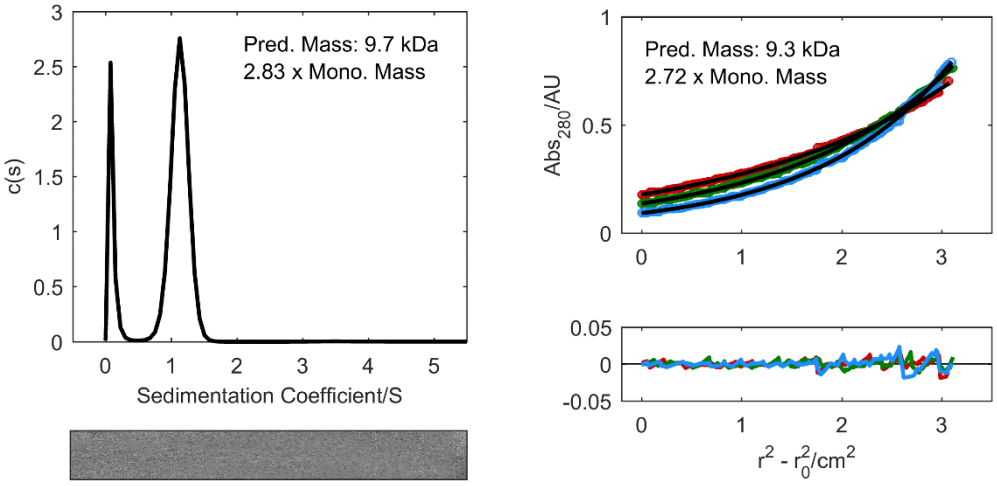


Figure 8-89 AUC data for 2-LV-EK ($\bar{v}$ = 0.7474 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.13, $s_{20,w}$ = 1.15, f/f_0 = 1.32 and Mw = 9662 Da (2.83 x monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 34 krpm (green) and 38 krpm (blue), returning Mw = 9285 Da (2.72 x monomer mass, 99 % confidence limits: 9189-9332 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

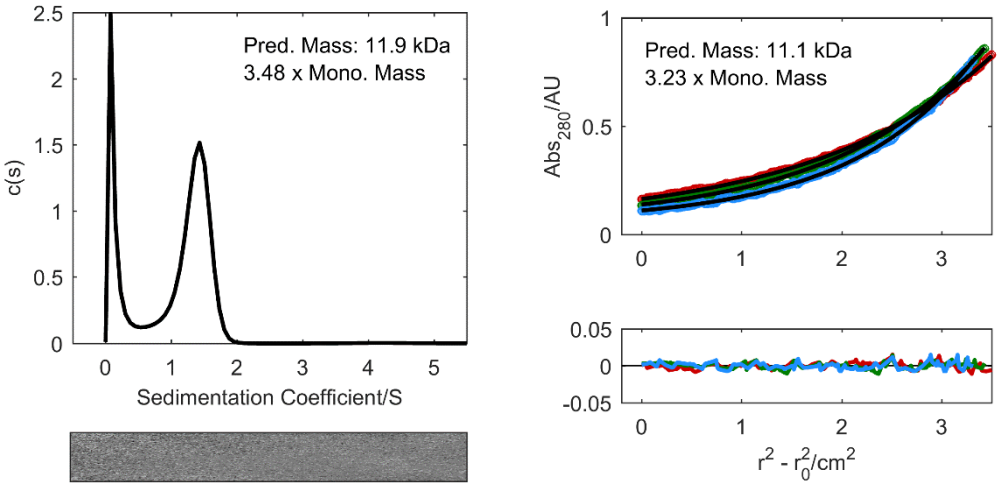


Figure 8-90 AUC data for 2-LV-KE ($\bar{v}$ = 0.7474 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.34, $s_{20,w}$ = 1.37, f/f_0 = 1.28 and Mw = 11889 Da (3.48 x monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 33 krpm (green) and 36 krpm (blue), returning Mw = 11060 Da (3.23 x monomer mass, 99 % confidence limits: 11019-11094 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

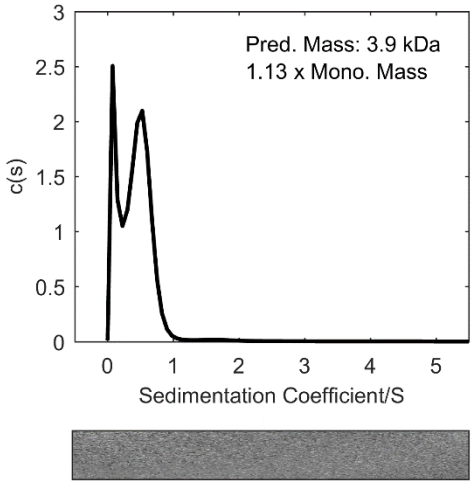


Figure 8-91 SV data for 3-LV-KE ($\bar{v}$ = 0.7474 $cm^3.g^{-1}$). Continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 0.51, $s_{20,w}$ = 0.52, f/f_0 = 1.58 and Mw = 3861 Da (1.13 x monomer mass). Predicted molecular weight is also indicated in figure inset. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

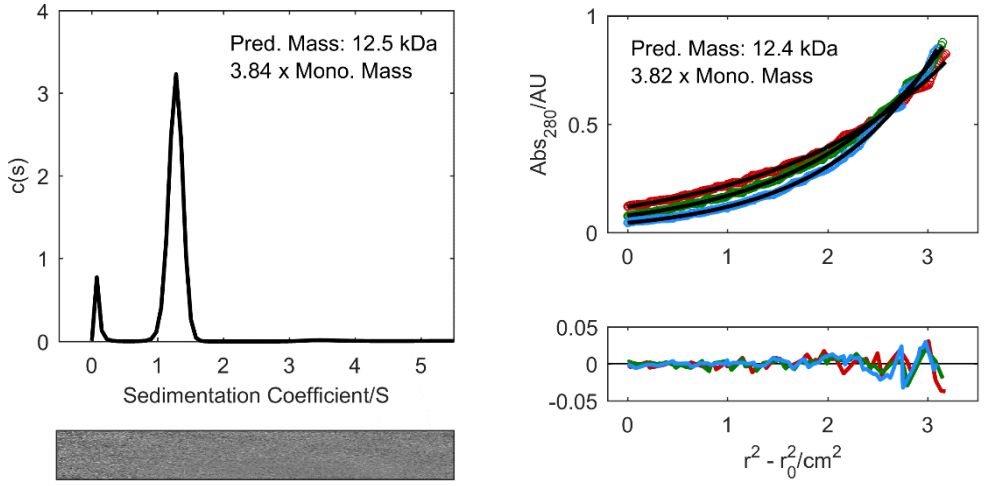


Figure 8-92 AUC data for 2-LIA-EK ($\bar{v}$ = 0.7669 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.27, $s_{20,w}$ = 1.30, f/f_0 = 1.26 and Mw = 12479 Da (3.84 x monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 34 krpm (green) and 38 krpm (blue), returning Mw = 12390 Da (3.82 x monomer mass, 99 % confidence limits: 12337-12492 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

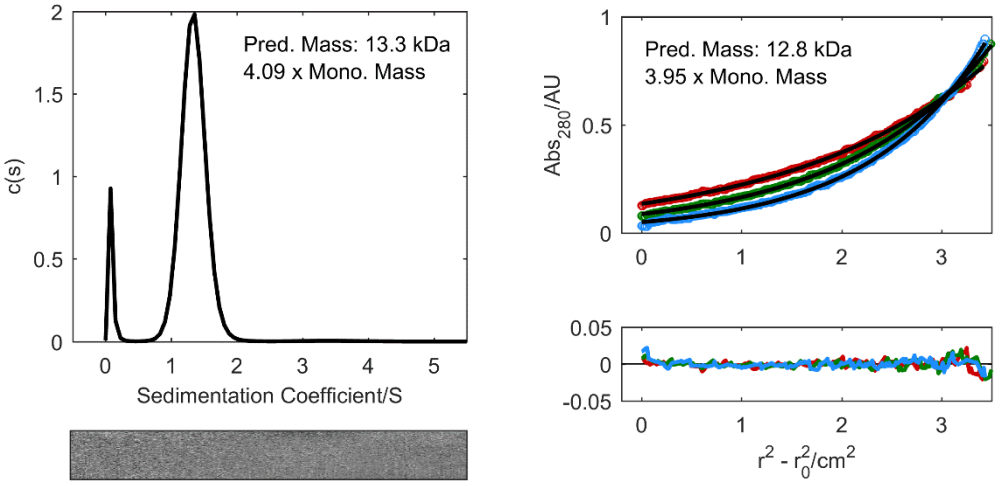


Figure 8-93 AUC data for 2-LIA-KE ($\bar{v}$ = 0.7669 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.34, $s_{20,w}$ = 1.37, f/f_0 = 1.25 and Mw = 13277 Da (4.09 x monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 33 krpm (red), 36 krpm (green) and 39 krpm (blue), returning Mw = 12810 Da (3.95 x monomer mass, 99 % confidence limits: 12773-12851 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

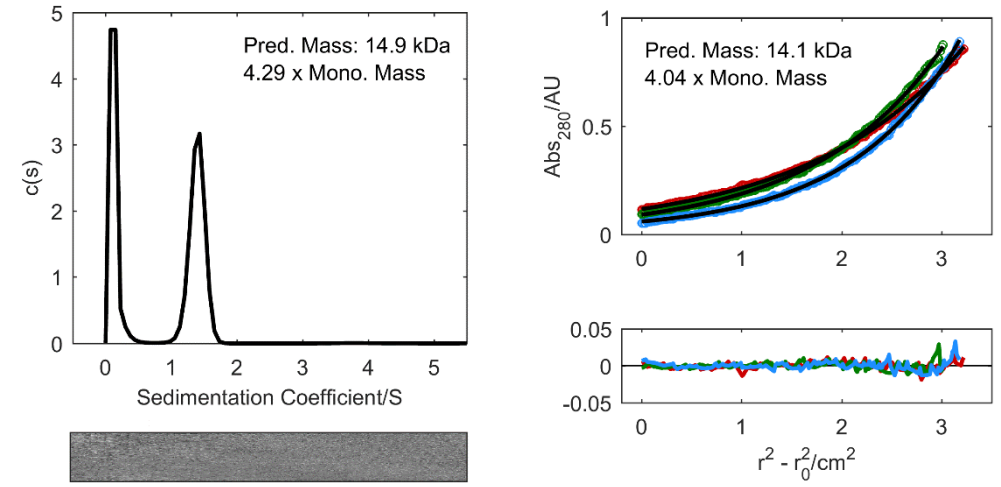


Figure 8-94 AUC data for ELAEIK ($\bar{v}$ = 0.7697 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.40, $s_{20,w}$ = 1.43, f/f_0 = 1.28 and Mw = 14924 Da (4.29 x monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 33 krpm (green) and 36 krpm (blue), returning Mw = 14060 Da (4.04 x monomer mass, 99 % confidence limits: 13992-14121 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

8.5.2 ABAB homomers

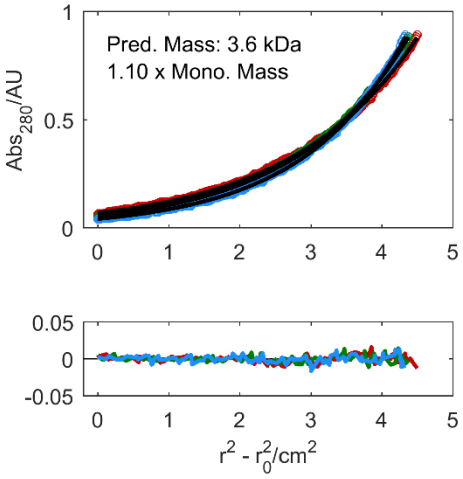


Figure 8-95 SE data for 1-LI-A ($\bar{v}$ = 0.7591 $cm^3.g^{-1}$). Data (circles) fitted to single-ideal species model curves (black lines) at 54 krpm (red), 57 krpm (green) and 60 krpm (blue), returning Mw = 3591 Da (1.10 x monomer mass, 99 % confidence limits: 3577-3605 Da). Bottom: residuals for the above fit, same colours. Predicted molecular weight is also indicated in figure inset. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

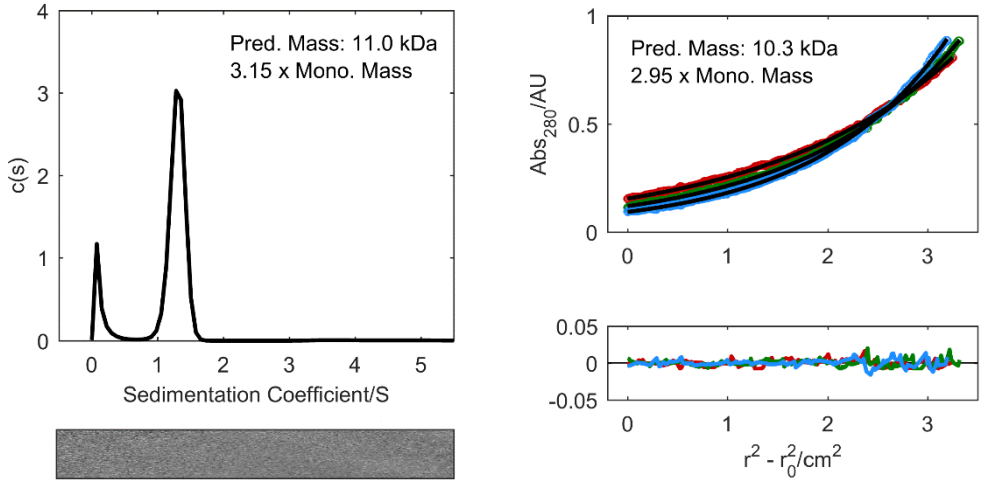


Figure 8-96 AUC data for 2-LI-A ($\bar{v}$ = 0.7473 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.30, $s_{20,w}$ = 1.33, f/f_0 = 1.25 and Mw = 10955 Da (3.15 x monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 33 krpm (green) and 36 krpm (blue), returning Mw = 10270 Da (2.95 x monomer mass, 99 % confidence limits: 10212-10331 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

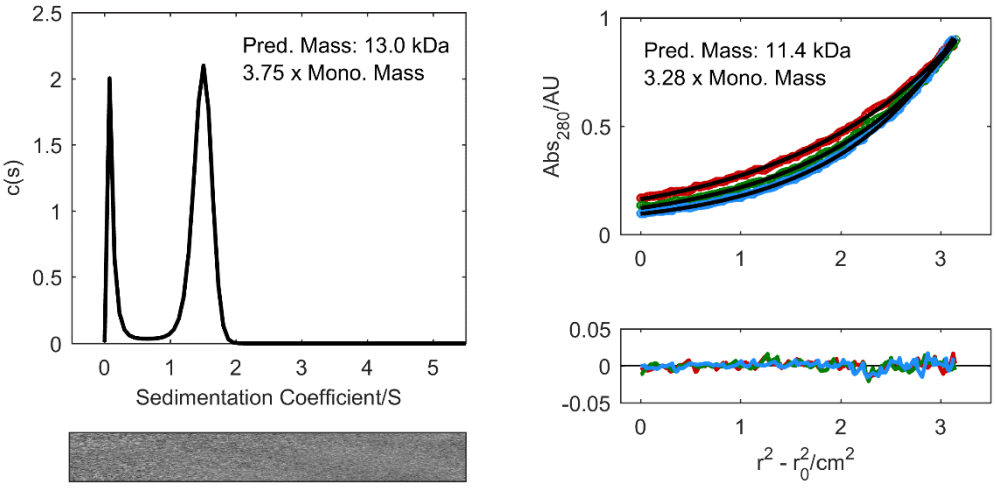


Figure 8-97 AUC data for 3-LI-A ($\bar{v}$ = 0.7473 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.47, $s_{20,w}$ = 1.50, f/f_0 = 1.24 and Mw = 13029 Da (3.75 x monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 33 krpm (green) and 36 krpm (blue), returning Mw = 11420 Da (3.28 x monomer mass, 99 % confidence limits: 11382-11466 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

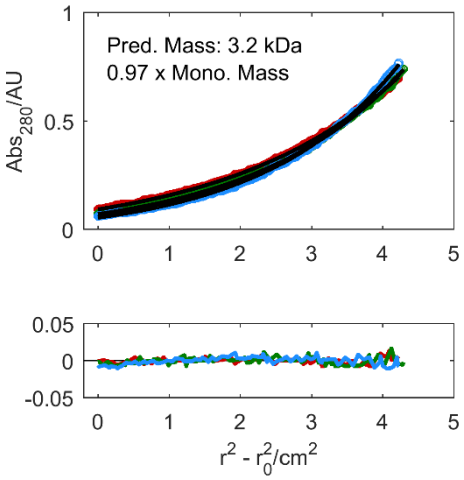


Figure 8-98 SE data for 1-LI-B ($\bar{v}$ = 0.7926 $cm^3.g^{-1}$). Data (circles) fitted to single-ideal species model curves (black lines) at 54 krpm (red), 57 krpm (green) and 60 krpm (blue), returning Mw = 3154 Da (0.97 x monomer mass, 99 % confidence limits: 3132-3177 Da). Bottom: residuals for the above fit, same colours. Predicted molecular weight is also indicated in figure inset. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

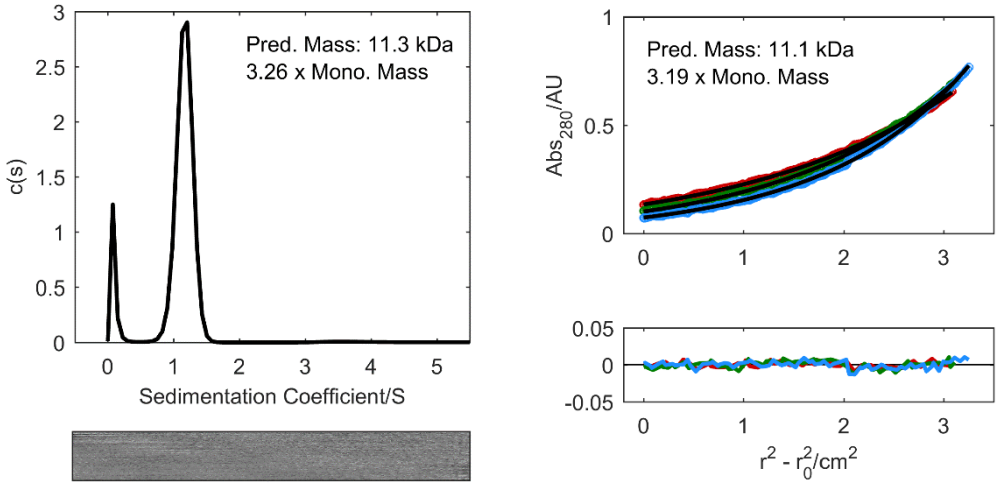


Figure 8-99 AUC data for 2-LI-B ($\bar{v}$ = 0.7785 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.17, $s_{20,w}$ = 1.20, f/f_0 = 1.22 and Mw = 11299 Da (2.36 x monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 33 krpm (green) and 36 krpm (blue), returning Mw = 11080 Da (3.19 x monomer mass, 99 % confidence limits: 10997-11169 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

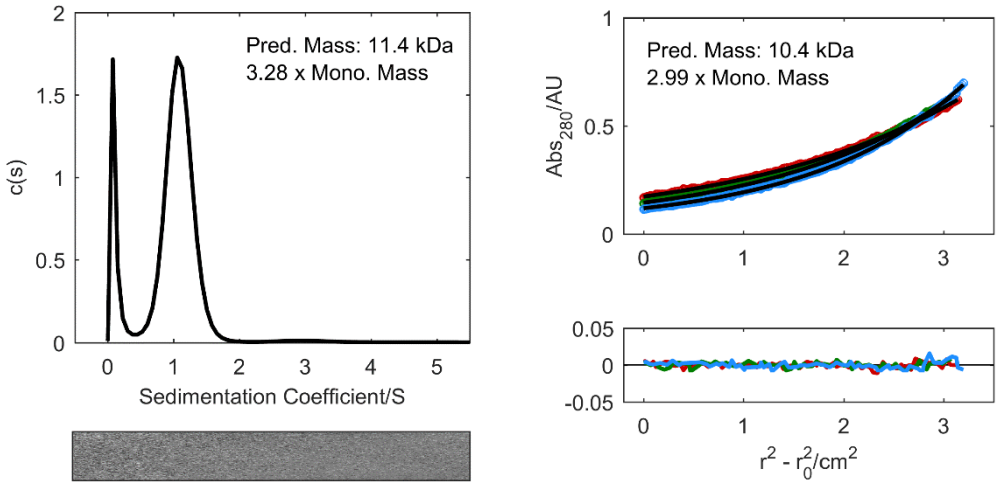


Figure 8-100 AUC data for 3-LI-B ($\bar{v}$ = 0.7785 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.08, $s_{20,w}$ = 1.11, f/f_0 = 1.32 and Mw = 11379 Da (3.28 x monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 33 krpm (green) and 36 krpm (blue), returning Mw = 10370 Da (2.99 x monomer mass, 99 % confidence limits: 10591-10771 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

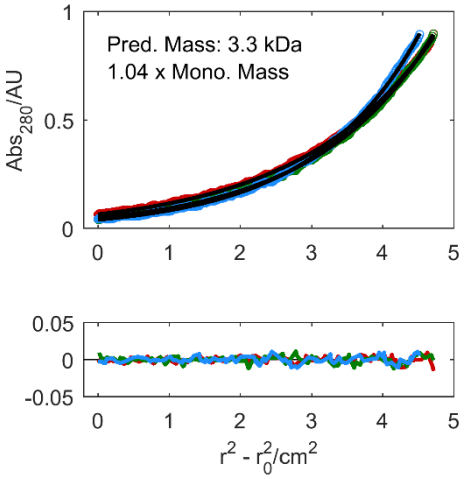


Figure 8-101 SE data for 1-LV-A ($\bar{v}$ = 0.7516 $cm^3.g^{-1}$). Data (circles) fitted to single-ideal species model curves (black lines) at 54 krpm (red), 57 krpm (green) and 60 krpm (blue), returning Mw = 3335 Da (1.04 x monomer mass, 99 % confidence limits: 3325-3345 Da). Bottom: residuals for the above fit, same colours. Predicted molecular weight is also indicated in figure inset. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

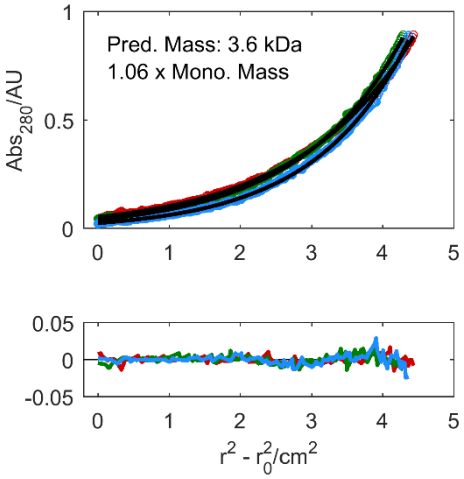


Figure 8-102 SE data for 2-LV-A ($\bar{v}$ = 0.7401 $cm^3.g^{-1}$). Data (circles) fitted to single-ideal species model curves (black lines) at 54 krpm (red), 57 krpm (green) and 60 krpm (blue), returning Mw = 3615 Da (1.06 x monomer mass, 99 % confidence limits: 3601-3630 Da). Bottom: residuals for the above fit, same colours. Predicted molecular weight is also indicated in figure inset. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

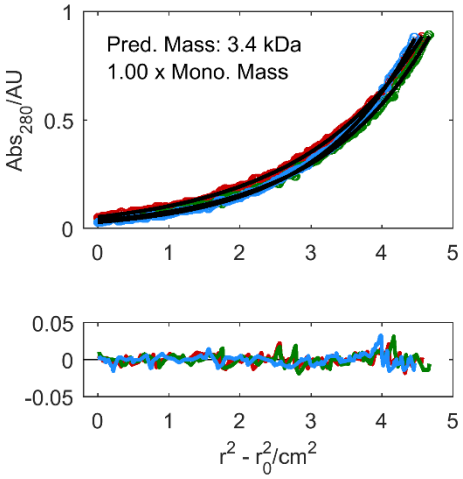


Figure 8-103 SE data for 3-LV-A ($\bar{v}$ = 0.7401 $cm^3.g^{-1}$). Data (circles) fitted to single-ideal species model curves (black lines) at 54 krpm (red), 57 krpm (green) and 60 krpm (blue), returning Mw = 3422 Da (1.00 x monomer mass, 99 % confidence limits: 3409-3435 Da). Bottom: residuals for the above fit, same colours. Predicted molecular weight is also indicated in figure inset. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

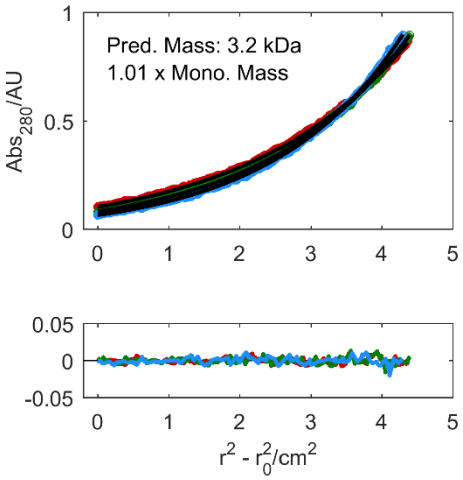


Figure 8-104 SE data for 1-LV-B ($\bar{v}$ = 0.7857 $cm^3.g^{-1}$). Data (circles) fitted to single-ideal species model curves (black lines) at 54 krpm (red), 57 krpm (green) and 60 krpm (blue), returning Mw = 3219 Da (1.01 x monomer mass, 99 % confidence limits: 3206-3231 Da). Bottom: residuals for the above fit, same colours. Predicted molecular weight is also indicated in figure inset. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

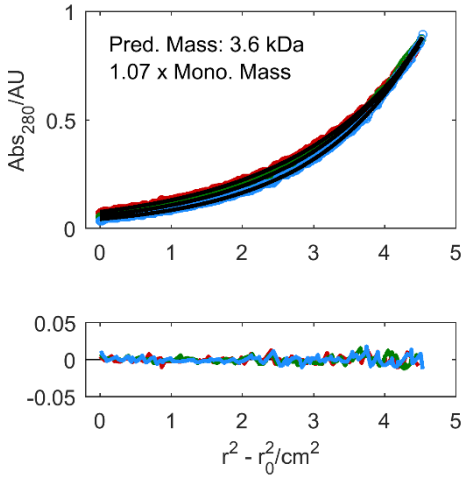


Figure 8-105 SE data for 2-LV-B ($\bar{v}$ = 0.7718 $cm^3.g^{-1}$). Data (circles) fitted to single-ideal species model curves (black lines) at 54 krpm (red), 57 krpm (green) and 60 krpm (blue), returning Mw = 3646 Da (1.07 x monomer mass, 99 % confidence limits: 3634-3658 Da). Bottom: residuals for the above fit, same colours. Predicted molecular weight is also indicated in figure inset. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

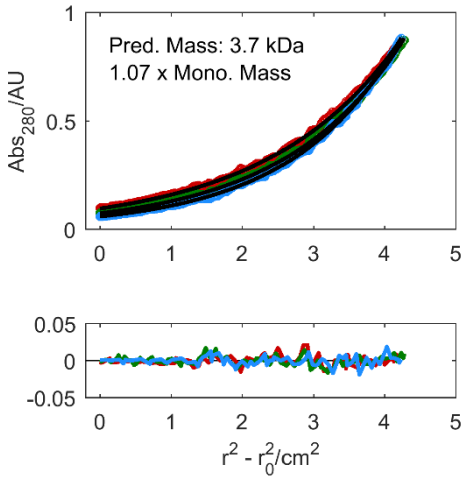


Figure 8-106 SE data for 3-LV-B ($\bar{v}$ = 0.7718 $cm^3.g^{-1}$). Data (circles) fitted to single-ideal species model curves (black lines) at 54 krpm (red), 57 krpm (green) and 60 krpm (blue), returning Mw = 3655 Da (1.07 x monomer mass, 95 % confidence limits: 3636-3675 Da). Bottom: residuals for the above fit, same colours. Predicted molecular weight is also indicated in figure inset. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

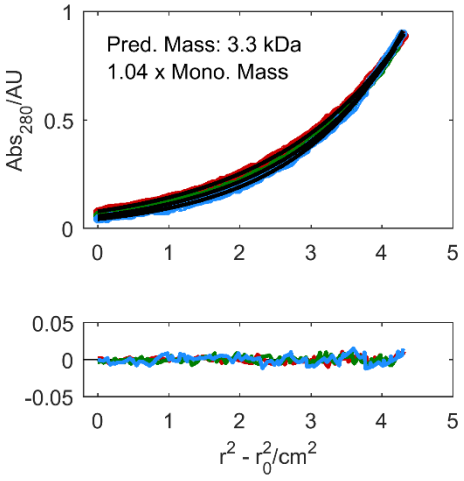


Figure 8-107 SE data for 1-LI-A* ($\bar{v}$ = 0.7481 $cm^3.g^{-1}$). Data (circles) fitted to single-ideal species model curves (black lines) at 54 krpm (red), 57 krpm (green) and 60 krpm (blue), returning Mw = 3289 Da (1.04 x monomer mass, 99 % confidence limits: 3279-3300 Da). Bottom: residuals for the above fit, same colours. Predicted molecular weight is also indicated in figure inset. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

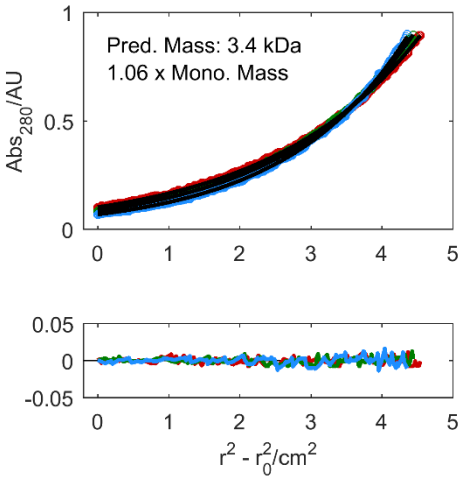


Figure 8-108 SE data for 1-LI-B* ($\bar{v}$ = 0.7825 $cm^3.g^{-1}$). Data (circles) fitted to single-ideal species model curves (black lines) at 54 krpm (red), 57 krpm (green) and 60 krpm (blue), returning Mw = 3364 Da (1.06 x monomer mass, 99 % confidence limits: 3353-3376 Da). Bottom: residuals for the above fit, same colours. Predicted molecular weight is also indicated in figure inset. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

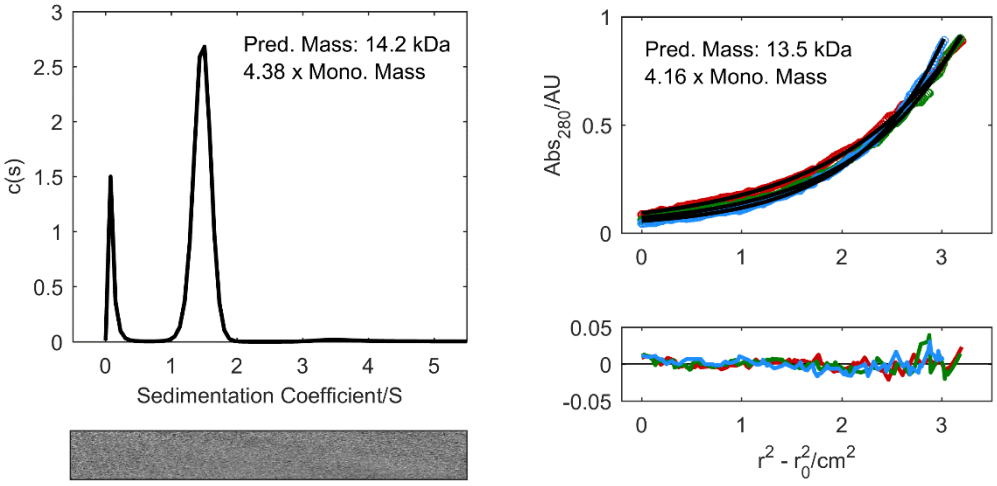


Figure 8-109 AUC data for 2-LIA-A ($\bar{v}$ = 0.7591 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.47, $s_{20,w}$ = 1.50, f/f_0 = 1.24 and Mw = 14245 kDa (4.38 x monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 33 krpm (green) and 36 krpm (blue), returning Mw = 13520 Da (4.16 x monomer mass, 99 % confidence limits: 13466-13582 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

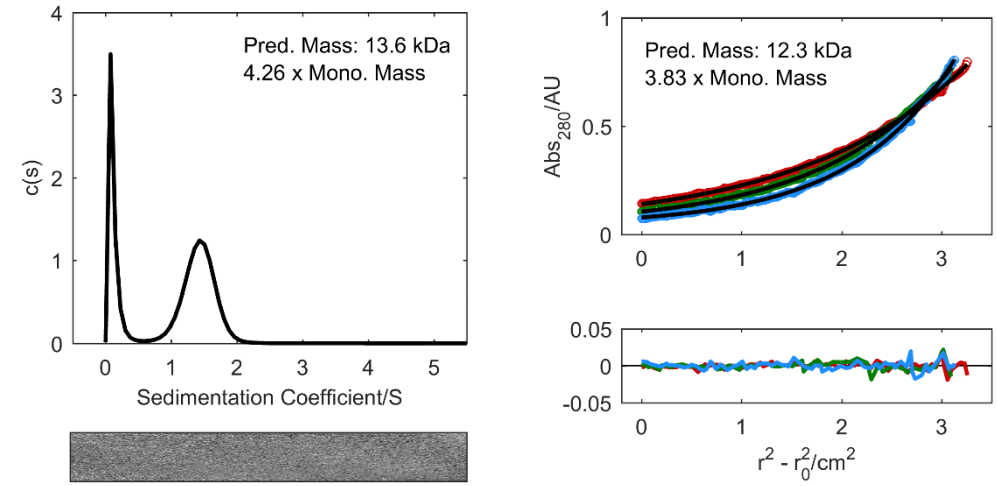


Figure 8-110 AUC data for 2-LVA-A ($\bar{v}$ = 0.7516 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.42, $s_{20,w}$ = 1.46, f/f_0 = 1.29 and Mw = 13605 Da (4.26 x monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 34 krpm (green) and 38 krpm (blue), returning Mw = 12250 Da (3.83 x monomer mass, 99 % confidence limits: 12190-12316 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

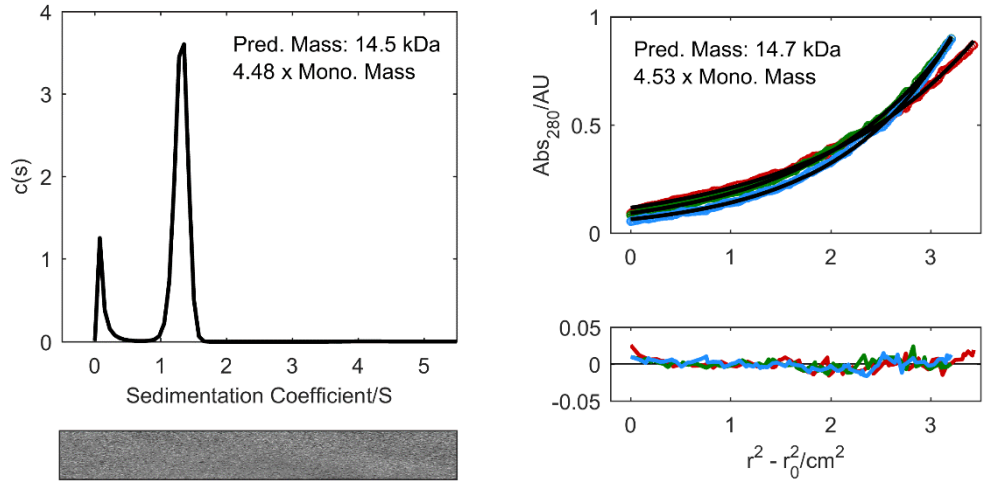


Figure 8-111 AUC data for 2-LIA-B ($\bar{v}$ = 0.7926 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.31, $s_{20,w}$ = 1.35, f/f_0 = 1.19 and Mw = 14542 Da (4.48 x monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 33 krpm (green) and 36 krpm (blue), returning Mw = 14680 Da (4.53 x monomer mass, 99 % confidence limits: 14608-14759 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

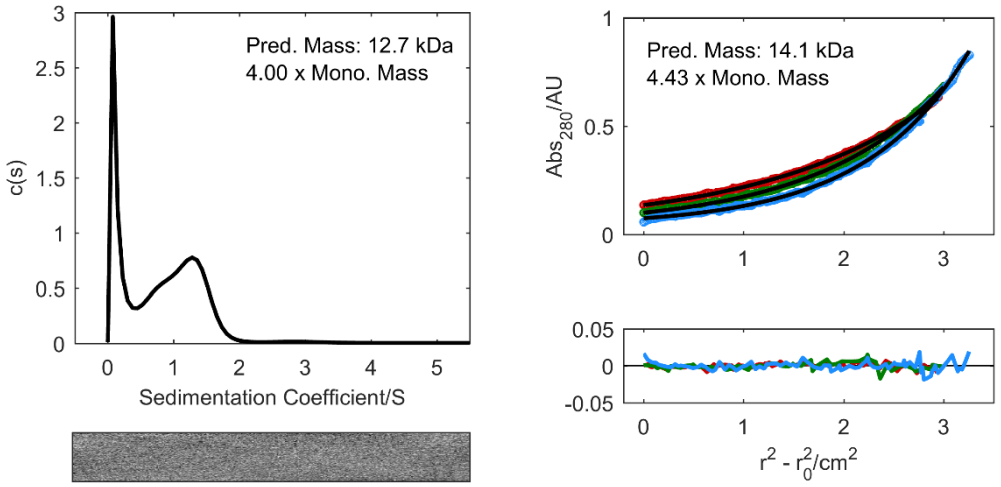


Figure 8-112 AUC data for 2-LVA-B ($\bar{v}$ = 0.7857 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.13, $s_{20,w}$ = 1.16, f/f_0 = 1.32 and Mw = 12749 Da (4.00 x monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 34 krpm (green) and 38 krpm (blue), returning Mw = 14110 Da (4.43 x monomer mass, 99 % confidence limits: 14044-14170 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

8.5.3 ABAB heteromers

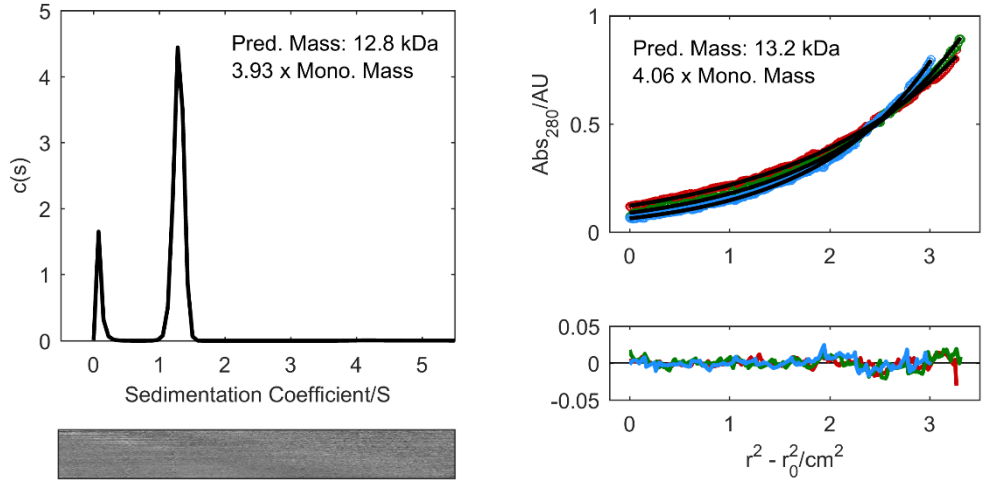


Figure 8-113 AUC data for 1-LI-AB ($\bar{v}$ = 0.7780 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.29, $s_{20,w}$ = 1.33, f/f_0 = 1.20 and Mw = 12772 Da (3.93 x mean monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 33 krpm (green) and 36 krpm (blue), returning Mw = 13190 Da (4.06 x mean monomer mass, 99 % confidence limits: 13107-13276 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

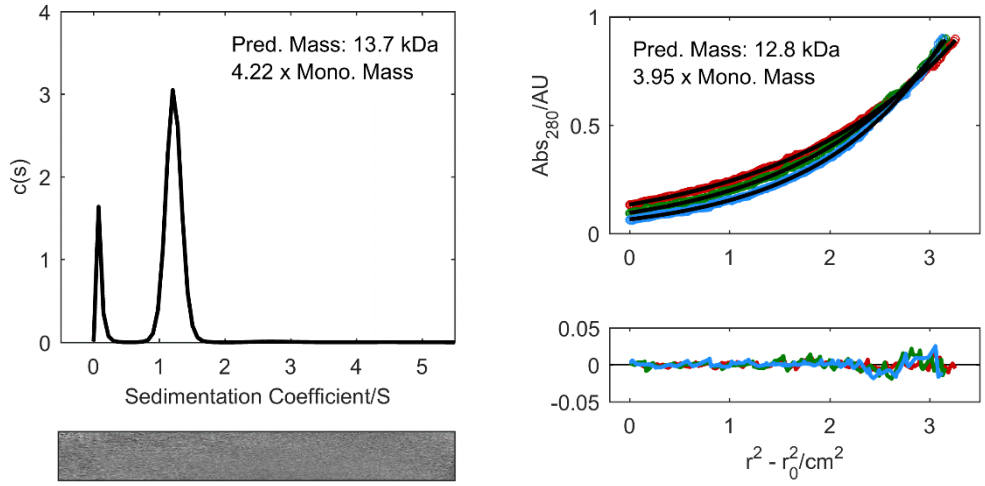


Figure 8-114 AUC data for 1-LI-AB-*g* ($\bar{v}$ = 0.7780 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.60, $s_{20,w}$ = 1.63, f/f_0 = 1.20 and Mw = 13698 Da (4.22 x mean monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 33 krpm (green) and 36 krpm (blue), returning Mw = 12820 Da (3.95 x mean monomer mass, 99 % confidence limits: 12757-12892 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

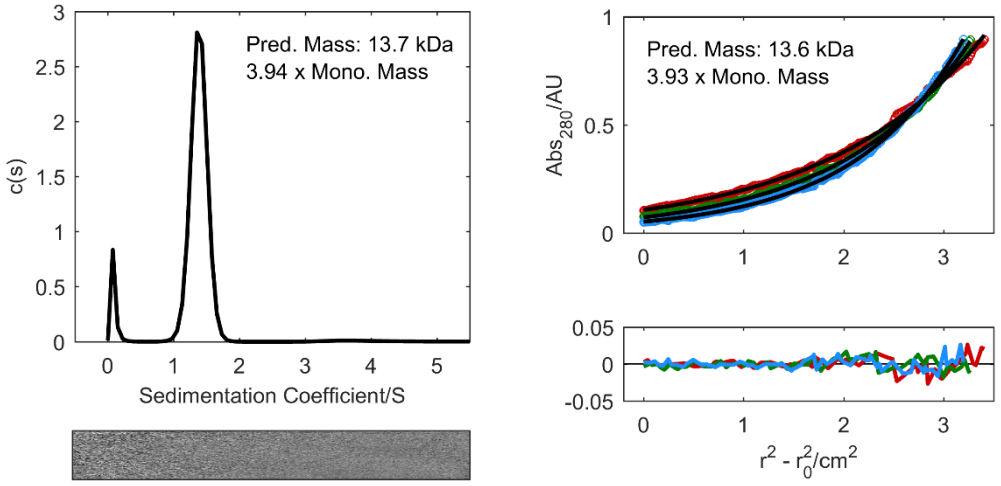


Figure 8-115 AUC data for 2-LI-AB ($\bar{v}$ = 0.7649 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.39, $s_{20,w}$ = 1.42, f/f_0 = 1.25 and Mw = 13697 Da (3.94 x mean monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 33 krpm (green) and 36 krpm (blue), returning Mw = 13640 Da (3.93 x mean monomer mass, 99 % confidence limits: 13583-13703 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

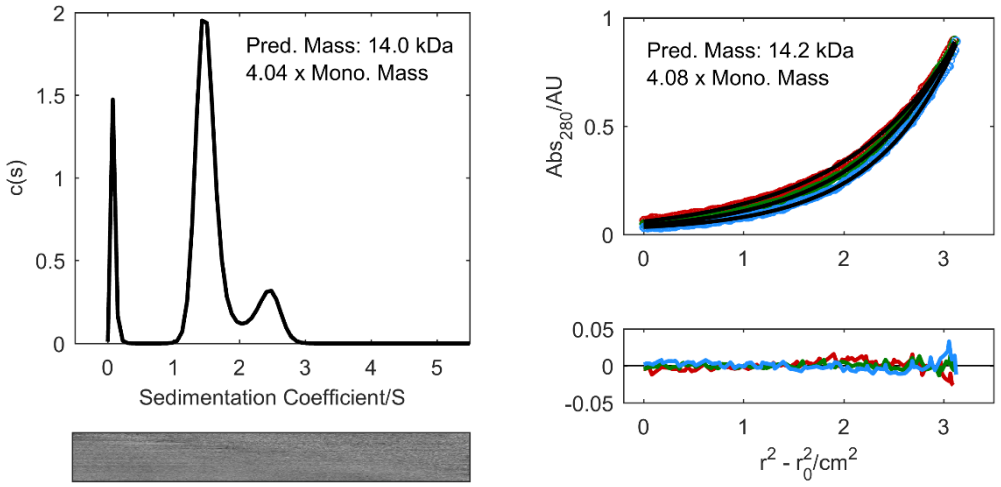


Figure 8-116 AUC data for 3-LI-AB ($\bar{v}$ = 0.7649 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning a major product with s = 1.51, $s_{20,w}$ = 1.55, f/f_0 = 1.16 and Mw = 14025 Da (4.04 x mean monomer mass) and a minor product with s = 2.42, $s_{20,w}$ = 2.48, f/f_0 = 1.16 and Mw = 28384 Da (8.17 x monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 33 krpm (green) and 36 krpm (blue), returning Mw = 14180 Da (4.08 x mean monomer mass, 99 % confidence limits: 14118-14252 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

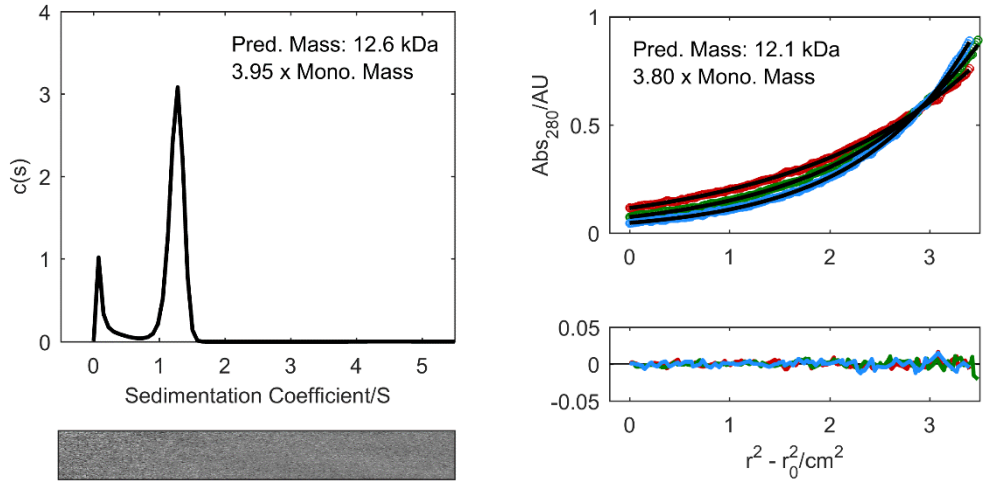


Figure 8-117 AUC data for 1-LV-AB ($\bar{v}$ = 0.7708 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.25, $s_{20,w}$ = 1.28, f/f_0 = 1.27 and Mw = 12589 Da (3.95 x mean monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 34 krpm (green) and 38 krpm (blue), returning Mw = 12110 Da (3.80 x mean monomer mass, 99 % confidence limits: 12058-12154 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

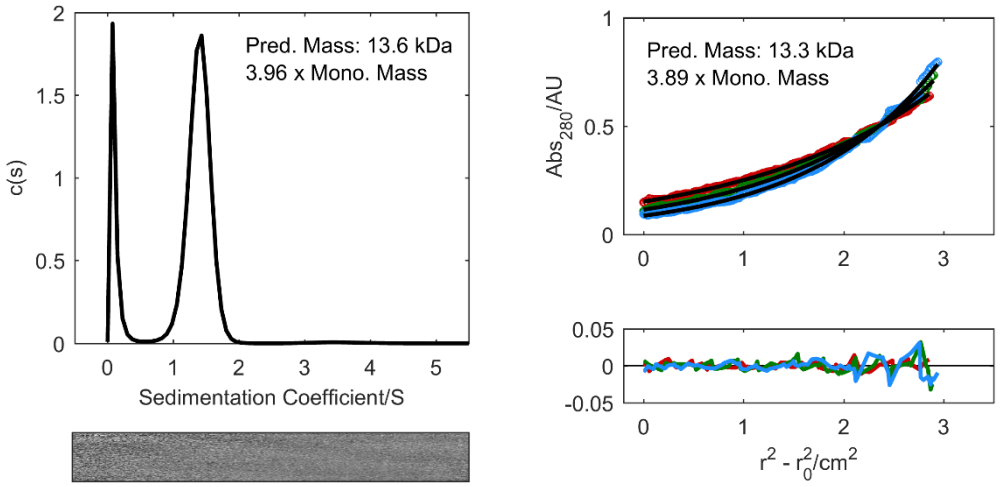


Figure 8-118 AUC data for 2-LV-AB ($\bar{v}$ = 0.7580 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.39, $s_{20,w}$ = 1.42, f/f_0 = 1.28 and Mw = 13552 Da (3.96 x mean monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 27 krpm (red), 30 krpm (green) and 33 krpm (blue), returning Mw = 13290 Da (3.89 x mean monomer mass, 99 % confidence limits: 13183-13396 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

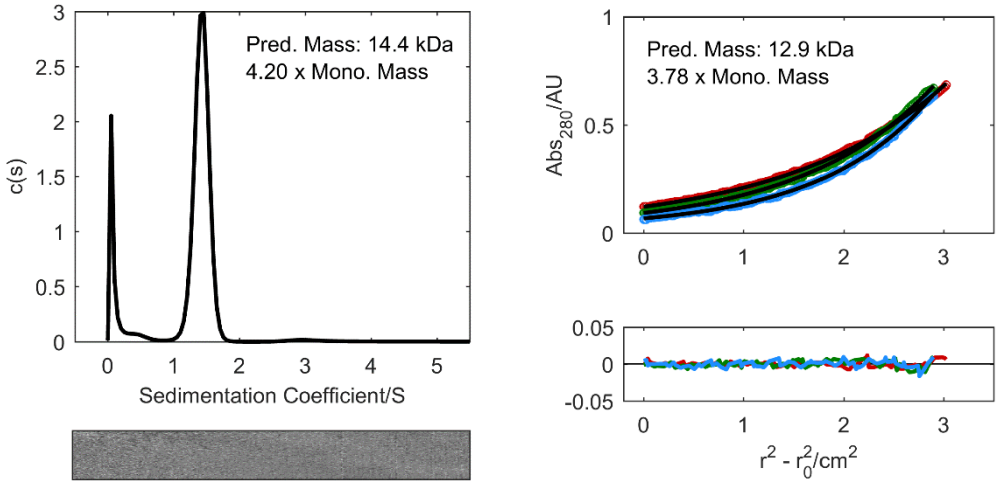


Figure 8-119 AUC data for 3-LV-AB ($\bar{v}$ = 0.7580 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.43, $s_{20,w}$ = 1.46, f/f_0 = 1.29 and Mw = 14363 Da (4.20 x mean monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 33 krpm (green) and 36 krpm (blue), returning Mw = 12910 Da (3.78 x mean monomer mass, 99 % confidence limits: 12840-12982 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

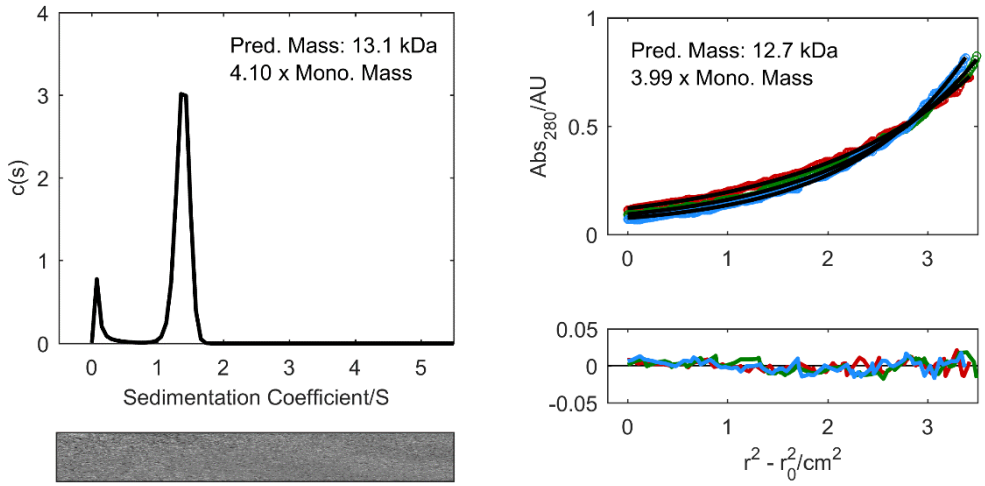


Figure 8-120 AUC data for 1-LV-AB in 10 mM sodium phosphate (pH 7.4) in the presence of 0.0 M NaCl ($\bar{v}$ = 0.7708 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.38, $s_{20,w}$ = 1.39, f/f_0 = 1.20 and Mw = 13080 Da (4.10 x mean monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 33 krpm (green) and 36 krpm (blue), returning Mw = 12730 Da (3.99 mean monomer mass, 99 % confidence limits: 12664-12788 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

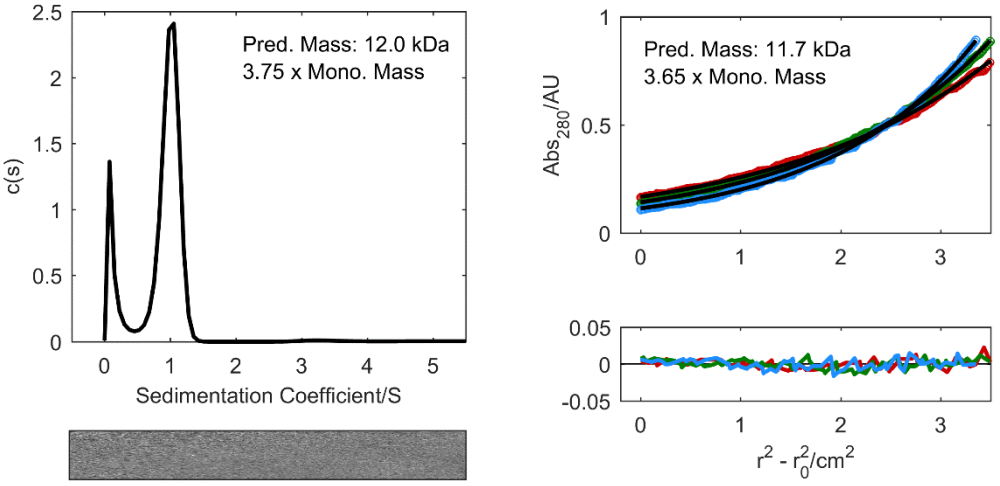


Figure 8-121 AUC data for 1-LV-AB in 10 mM sodium phosphate (pH 7.4) in the presence of 1.0 M NaCl ($\bar{v}$ = 0.7708 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 0.992, $s_{20,w}$ = 1.27, f/f_0 = 1.24 and Mw = 11979 Da (3.75 x mean monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 33 krpm (green) and 36 krpm (blue), returning Mw = 11650 Da (3.65 mean monomer mass, 99 % confidence limits: 11592-11715 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

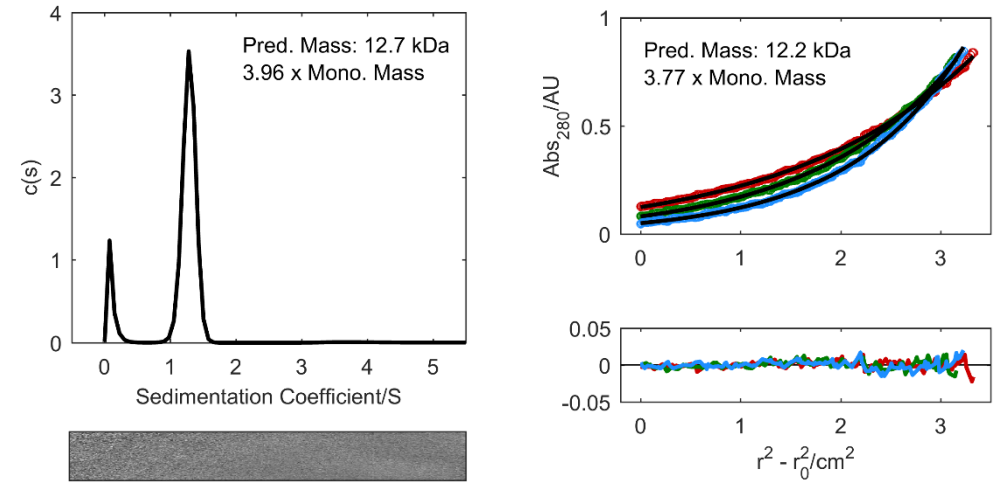


Figure 8-122 AUC data for 1-LI-A/1-LV-B ($\bar{v}$ = 0.7745 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.29, $s_{20,w}$ = 1.32, f/f_0 = 1.22 and Mw = 12737 Da (3.96 x mean monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 34 krpm (green) and 38 krpm (blue), returning Mw = 12150 Da (3.77 x mean monomer mass, 99 % confidence limits: 12091-12213 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

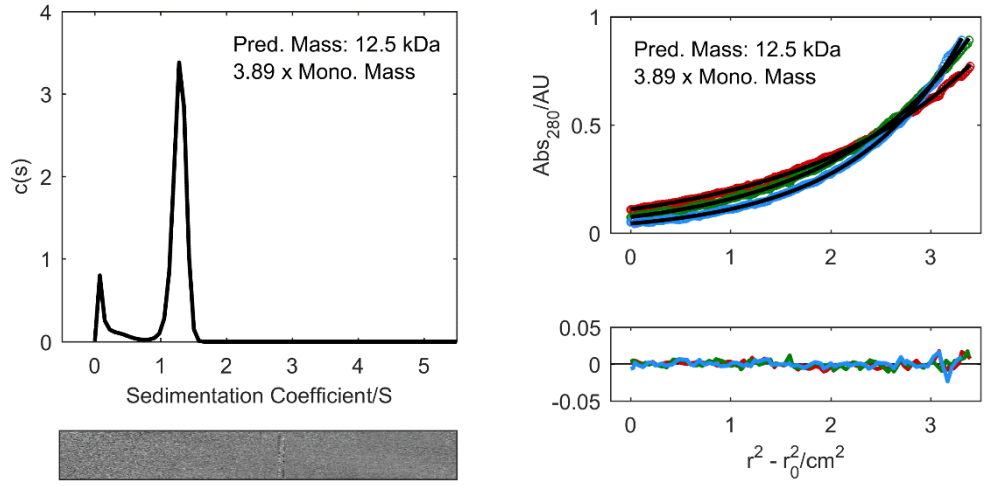


Figure 8-123 AUC data for 1-LV-A/1-LI-B ($\bar{v}$ = 0.7745 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.28, $s_{20,w}$ = 1.31, f/f_0 = 1.21 and Mw = 12530 Da (3.89 x mean monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 34 krpm (green) and 38 krpm (blue), returning Mw = 12530 Da (3.89 x mean monomer mass, 99 % confidence limits: 12492-12592 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

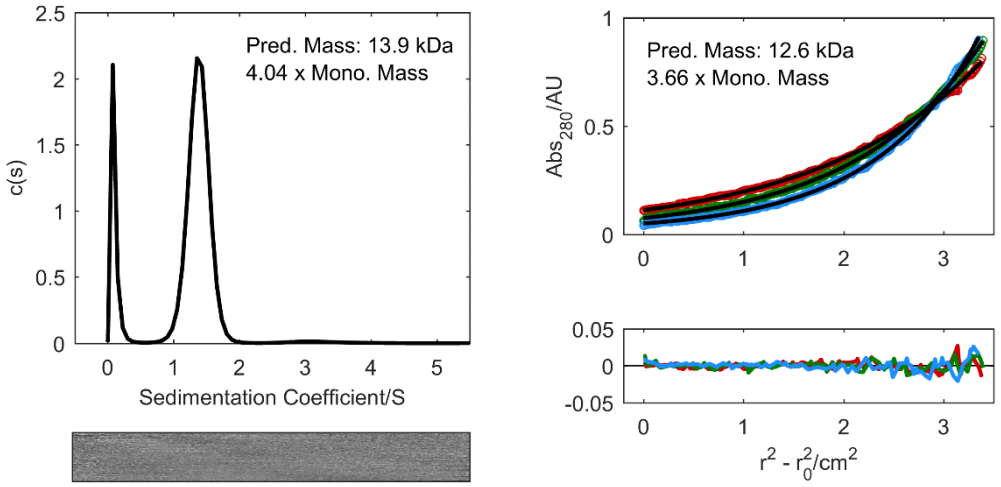


Figure 8-124 AUC data for 2-LI-A/2-LV-B ($\bar{v}$ = 0.7615 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.38, $s_{20,w}$ = 1.41, f/f_0 = 1.29 and Mw = 13931 Da (4.04 x mean monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 34 krpm (green) and 38 krpm (blue), returning Mw = 12600 Da (3.66 x mean monomer mass, 99 % confidence limits: 12529-12681 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

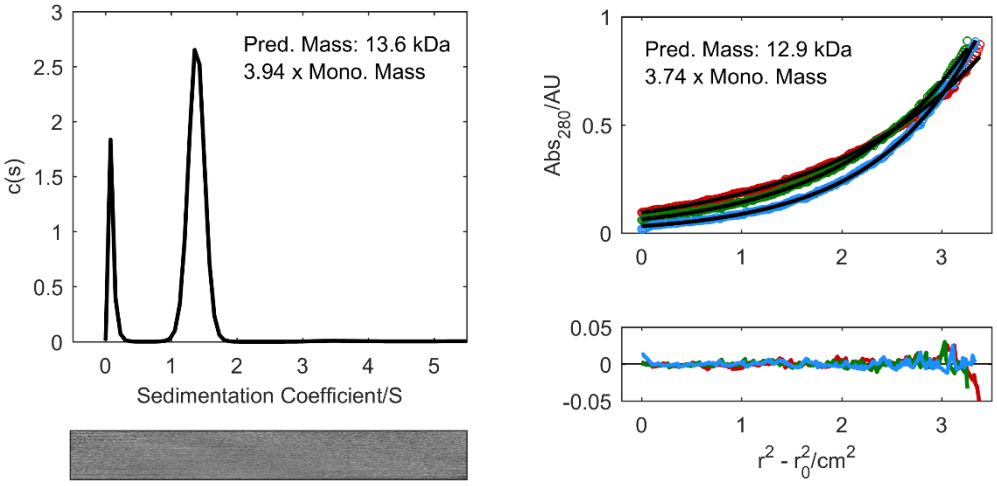


Figure 8-125 AUC data for 2-LV-A/2-LI-B ($\bar{v}$ = 0.7615 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.38, $s_{20,w}$ = 1.42, f/f_0 = 1.26 and Mw = 13594 Da (3.94 x mean monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 34 krpm (green) and 38 krpm (blue), returning Mw = 12890 Da (3.74 x mean monomer mass, 99 % confidence limits: 12827-12948 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

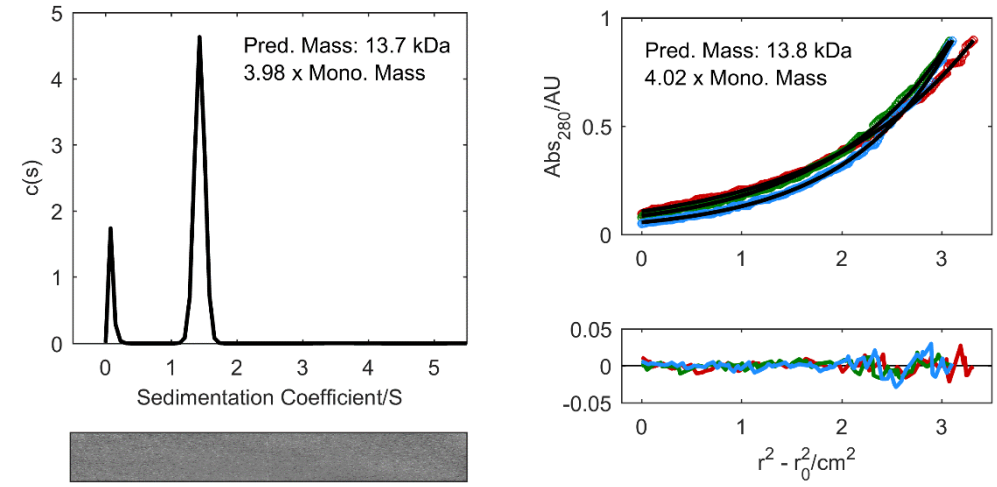


Figure 8-126 AUC data for 3-LI-A/3-LV-B ($\bar{v}$ = 0.7615 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.43, $s_{20,w}$ = 1.47, f/f_0 = 1.23 and Mw = 13733 Da (3.98 x mean monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 33 krpm (green) and 36 krpm (blue), returning Mw = 13840 Da (4.02 x mean monomer mass, 99 % confidence limits: 13779-13899 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

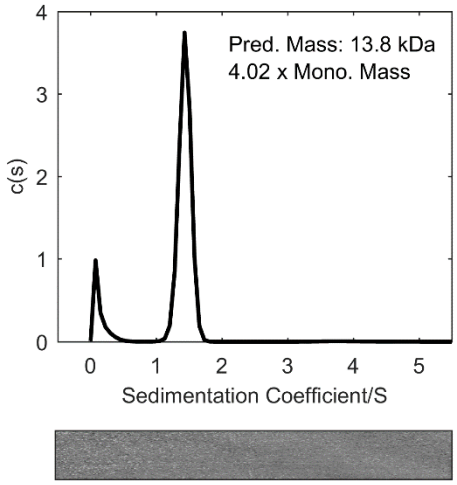


Figure 8-127 AUC data for 3-LV-A/3-LI-B ($\bar{v}$ = 0.7615 $cm^3.g^{-1}$). SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.44, $s_{20,w}$ = 1.47, f/f_0 = 1.23 and Mw = 13843 Da (4.02 x mean monomer mass). Predicted molecular weight is also indicated in figure inset. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

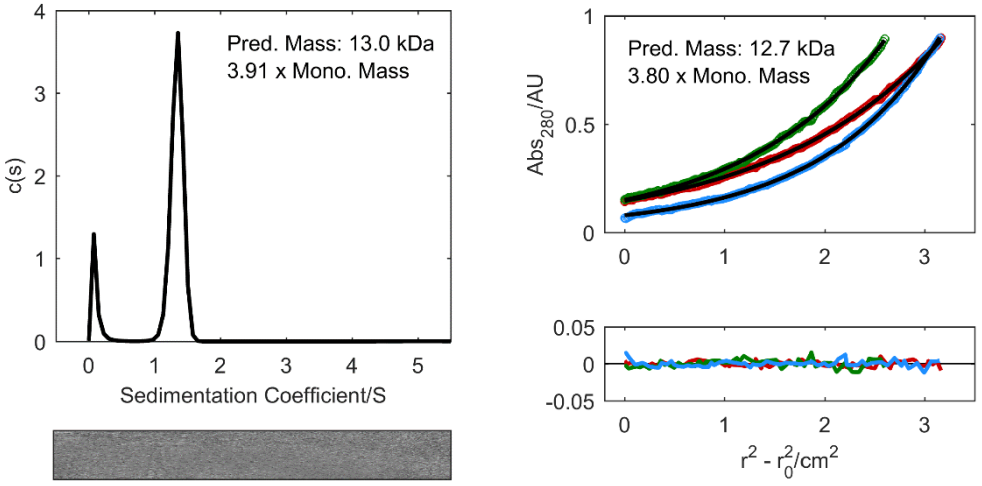


Figure 8-128 AUC data for 1-LI-A/2-LV-B ($\bar{v}$ = 0.7677 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.34, $s_{20,w}$ = 1.38, f/f_0 = 1.23 and Mw = 13025 Da (3.91 x mean monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 33 krpm (green) and 36 krpm (blue), returning Mw = 12650 Da (3.80 x mean monomer mass, 99 % confidence limits: 12589-12717 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

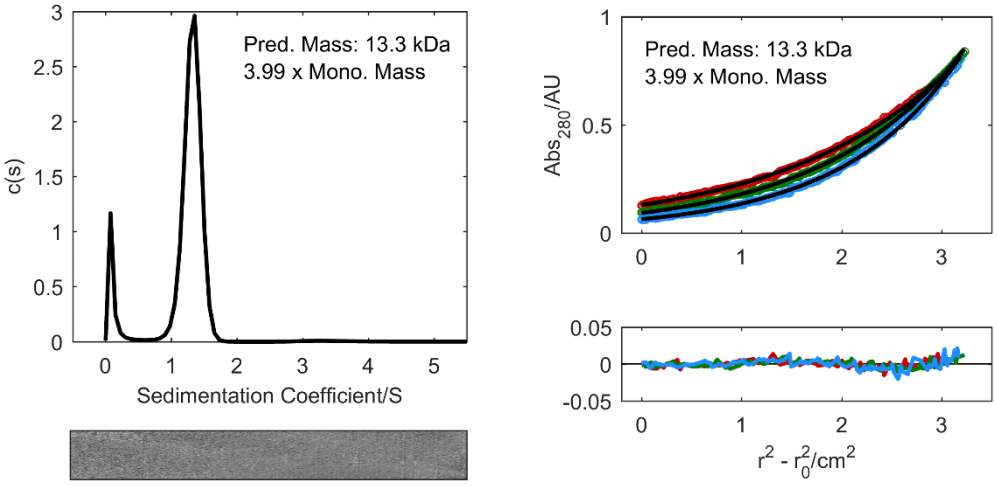


Figure 8-129 AUC data for 1-LI-A/3-LV-B ($\bar{v}$ = 0.7677 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.32, $s_{20,w}$ = 1.35, f/f_0 = 1.27and Mw = 13295 Da (3.99 x mean monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 33 krpm (green) and 36 krpm (blue), returning Mw = 13300 Da (3.99 x mean monomer mass, 99 % confidence limits: 13253-13356 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

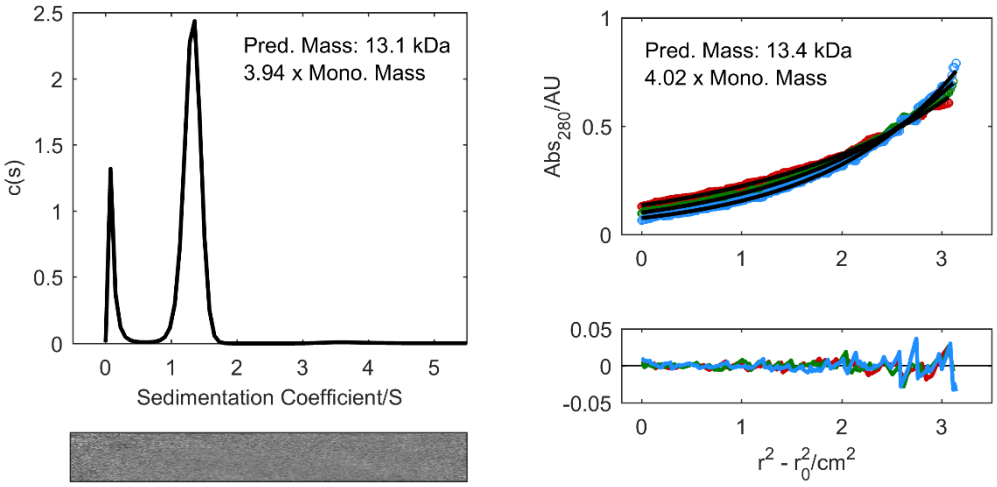


Figure 8-130 AUC data for 2-LV-A/1-LI-B ($\bar{v}$ = 0.7615 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.32, $s_{20,w}$ = 1.35, f/f_0 = 1.26 and Mw = 13128 Da (3.94 x mean monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 27 krpm (red), 30 krpm (green) and 33 krpm (blue), returning Mw = 13400 Da (4.02 x mean monomer mass, 99 % confidence limits: 13280-13517 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

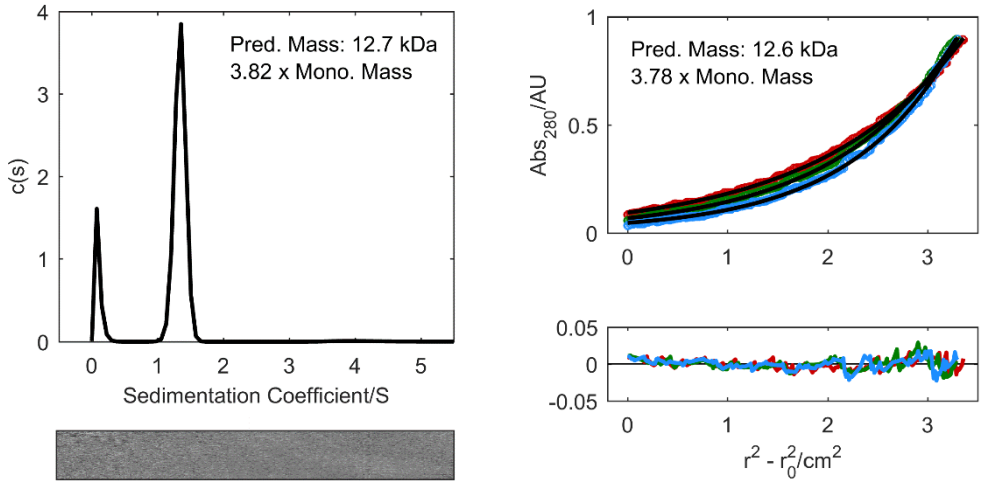


Figure 8-131 AUC data for 3-LV-A/1-LI-B ($\bar{v}$ = 0.7677 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.34, $s_{20,w}$ = 1.37 f/f_0 = 1.21 and Mw = 12723 Da (3.82 x mean monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 33 krpm (green) and 36 krpm (blue), returning Mw = 12610 Da (3.78 x mean monomer mass, 99 % confidence limits: 12558-12659 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

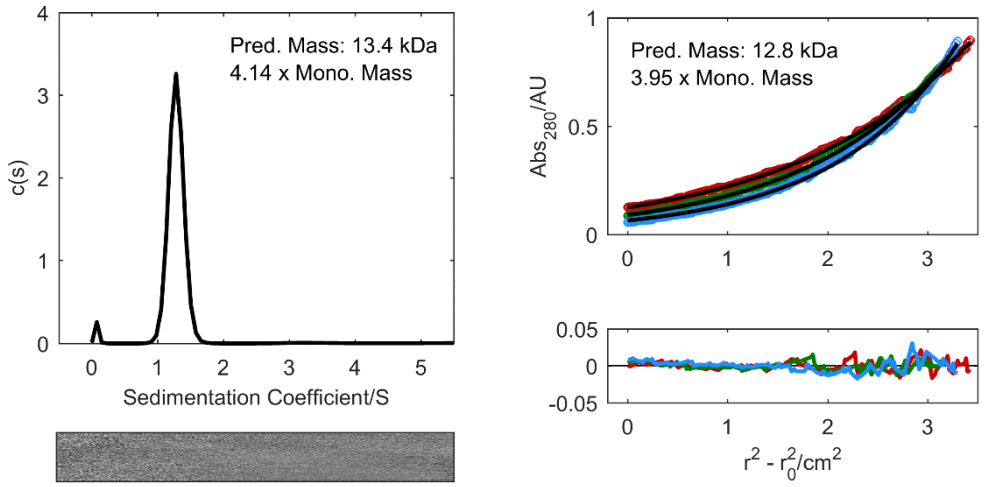


Figure 8-132 AUC data for 2-LIA-AB ($\bar{v}$ = 0.7780 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.28, $s_{20,w}$ = 1.32, f/f_0 = 1.25 and Mw = 13445 Da (4.14 x mean monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 33 krpm (green) and 36 krpm (blue), returning Mw = 12810 Da (3.95 x mean monomer mass, 99 % confidence limits: 12757-12872 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

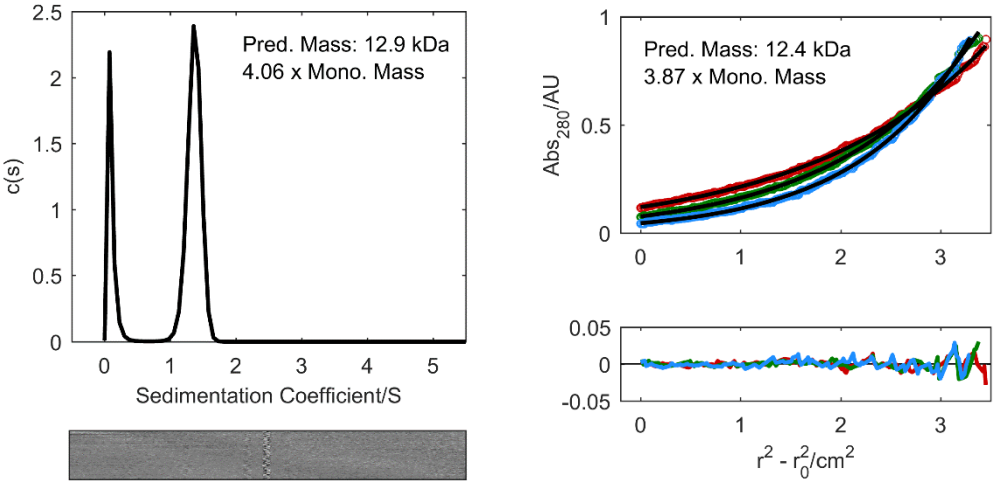


Figure 8-133 AUC data for 2-LVA-AB ($\bar{v}$ = 0.7708 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.37, $s_{20,w}$ = 1.40, f/f_0 = 1.19 and Mw = 12945 Da (4.06 x mean monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 34 krpm (green) and 38 krpm (blue), returning Mw = 12360 Da (3.87 x mean monomer mass, 99 % confidence limits: 12300-12418 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

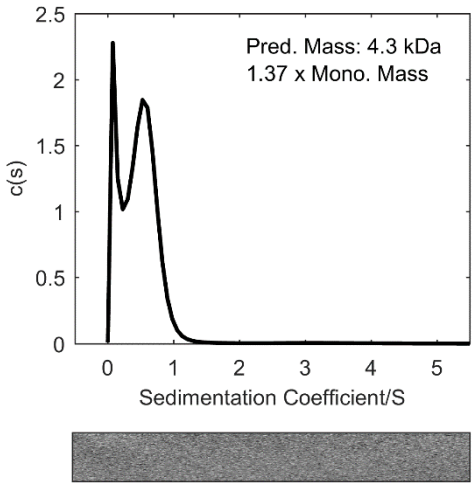


Figure 8-134 SV data for 1-LI-AB* ($\bar{v}$ = 0.7675 $cm^3.g^{-1}$). Continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 0.56, $s_{20,w}$ = 0.58, f/f_0 = 1.41 and Mw = 4335 Da (1.37 x mean monomer mass). Predicted molecular weight is also indicated in figure inset. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

8.5.4 *ABCD heteromers*

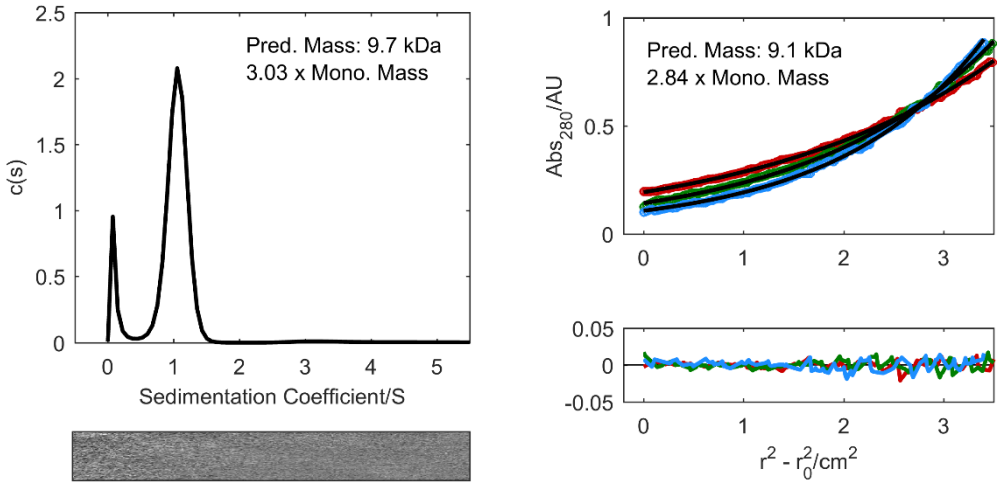


Figure 8-135 AUC data for 1234 ($\bar{v}$ = 0.7665 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.05, $s_{20,w}$ = 1.08, f/f_0 = 1.30 and Mw = 9743 Da (3.03 x mean monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 34 krpm (green) and 38 krpm (blue), returning Mw = 9127 Da (2.84 x mean monomer mass, 99 % confidence limits: 9078-9177 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

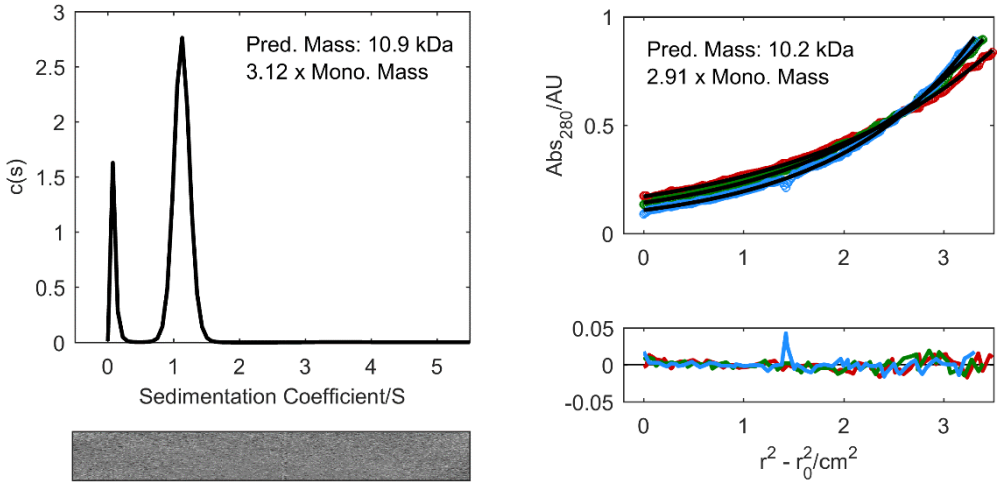


Figure 8-136 AUC data for 5678 ($\bar{v}$ = 0.7555 $cm^3.g^{-1}$). Left: SV continuous c(s) distribution fits (top) and residuals (bottom) at 50 krpm returning s = 1.13, $s_{20,w}$ = 1.15, f/f_0 = 1.38 and Mw = 10884 Da (3.12 x mean monomer mass). Right, top: SE data (circles) fitted to single-ideal species model curves (black lines) at 30 krpm (red), 33 krpm (green) and 36 krpm (blue), returning Mw = 10170 Da (2.91 x mean monomer mass, 99 % confidence limits: 10122-10222 Da). Right, bottom: residuals for the above fit, same colours. Predicted molecular weights are also indicated in figure insets. Pred. mass, predicted mass; mono. mass, (mean) monomer mass.

8.6 X-ray crystallography conditions, data collection statistics and refinement statistics

Peptide	Buffer	pH	Salt	Precipitant
2-LI-EK	0.1 M PCTP	7.0	-	25 % (w/v) PEG 1500
3-LI-EK	0.1 M Sodium HEPES	7.5	0.2 M Sodium citrate tribasic dihydrate	20 % (v/v) 2-propanol
ELAEIK	0.08 M sodium cacodylate	6.5	0.16 M calcium acetate	14.4 % w/v PEG 8000, 20 % v/v glycerol

Table 8-2 Crystallisation buffer conditions for peptide X-ray crystal structures. PEG, polyethylene glycol; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid. PCTP: sodium propionate, sodium cacodylate trihydrate, bis-tris propane.

Collection		
Wavelength (Å)		0.9763
Beamline (Diamond, UK)		i03
Space Group		C222
Cell Dimensions	a, b, c (Å)	47.56, 50.84, 43.94
	α, β, γ (°)	90, 90, 90
Resolution (Å)		43.94–1.70 (1.76–1.70)
R_{merge}		0.080 (0.303)
R_{meas}		0.091 (0.347)
I/σI		12.8 (1.9)
CC ½ (%)		99.0 (97.8)
Completeness (%)		99.6 (100)
Redundancy		7.9 (7.7)
Wilson B Factor (Å^2)		19.72
Refinement		
Resolution (Å)		1.70
Reflections/Unique		48148/6093
R_{work}/R_{free}		0.1889/0.1949
No. atoms	Protein (all)	489
	Ligand/ion	10
	Water	32
B-factors (Å^2)	Main chain/protein	21.698/ 27.883
	Water/ligand	41.816/53.858
R.m.s.d	Bond lengths (Å)	0.005
	Bond angles (°)	0.740
Ramachandran (%)	Favoured	100.0
	Allowed	0.0
	Outliers	0.0

Table 8-3 Data collection and refinement statistics for peptide 2-LI-EK. Overall data are shown with outer shell data given in parentheses.

Collection		
Wavelength (Å)		0.9795
Beamline (Diamond, UK)		i04
Space Group		$P12_11$
Cell Dimensions	a, b, c (Å)	29.46, 48.55, 36.57
	α, β, γ (°)	90.00, 96.19, 90.00
Resolution (Å)		48.55–1.11 (1.12–1.11)
R_{merge}		0.051 (0.908)
R_{meas}		0.062 (1.127)
I/σI		12.7 (1.9)
CC ½ (%)		99.8 (62.9)
Completeness (%)		94.5 (90.6)
Redundancy		6.1 (5.3)
Wilson B Factor (Å^2)		12.88
Refinement		
Resolution (Å)		1.1
Reflections/Unique		236063/38599
R_{work}/R_{free}		0.1606/0.1967
No. atoms	Protein (all)	1015
	Ligand/ion	0
	Water	125
B-factors (Å^2)	Main chain/protein	15.982/18.938
	Water	36.699
R.m.s.d	Bond lengths (Å)	0.005
	Bond angles (°)	0.600
Ramachandran (%)	Favoured	100.0
	Allowed	0.0
	Outliers	0.0

Table 8-4 Data collection and refinement statistics for peptide 3-LI-EK. Overall data are shown with outer shell data given in parentheses.

Collection		
Wavelength (Å)		0.9282
Beamline (Diamond, UK)		i04-1
Space Group		$P2_12_12$
Cell Dimensions	a, b, c (Å)	35.110, 40.040, 44.590
	α, β, γ (°)	90, 90, 90
Resolution (Å)		29.79–1.44 (1.49–1.44)
R_{merge}		0.068 (0.862)
R_{meas}		0.080 (1.021)
I/σI		11.4 (1.4)
CC ½ (%)		0.999 (0.99)
Completeness (%)		96.0.7 (96.4)
Redundancy		6.0 (6.3)
Wilson B Factor (Å^2)		16.79
Refinement		
Resolution (Å)		1.44
Reflections/Unique		68076/11409
R_{work}/R_{free}		0.1822/0.2006
No. atoms	Protein (all)	1019
	Ligand/ion	22
	Water	61
B-factors (Å^2)	Main chain/protein	20.6015/34.8105
	Water	40.515/53.512
R.m.s.d	Bond lengths (Å)	0.009
	Bond angles (°)	0.939
Ramachandran (%)	Favoured	100.0
	Allowed	0.0
	Outliers	0.0

Table 8-5 Data collection and refinement statistics for peptide ELAEIK. Overall data are shown with outer shell data given in parentheses.

8.7 Plasmid maps

All plasmids included in this work are listed in Table 8-6. Selected plasmid maps are shown in Figures 8-137 to 8-155. Plasmid maps were generated using SnapGene Viewer.

No.	Name	Size (bp)	Resistance gene
1	pVRb-LacUV5-SFGFP	4318	*Kan*^{R}
2	pVRb-O1O1-LacUV5-SFGFP	4404	*Kan*^{R}
3	pBAD-LacI-WT	5152	*Amp*^{R}
4	pBAD-LacI-L251	5092	*Amp*^{R}
5	pBAD-LacI-L251A-WTtet	5152	*Amp*^{R}
6	pVRc-LacI-L251A	5943	*Cm*^{R}
7	pBAD-LacI-L251A	5092	*Amp*^{R}
8	pBAD-LacI-ΔHTH	4951	*Amp*^{R}
9	pBAD-LacI-ΔHTH-L251A	4951	*Amp*^{R}
10	pBAD-6H-SUMO	4324	*Amp*^{R}
11	pBAD-6H-T7-Xpress	4099	*Amp*^{R}
12	pBAD-6H	4030	*Amp*^{R}
13	pBAD-LacI-L251A-ccDi	5178	*Amp*^{R}
14	pVRc-LacI-L251A-A1	6029	*Cm*^{R}
15	pVRc-LacI-L251A-A2	6029	*Cm*^{R}
16	pVRc-LacI-L251A-A3	6029	*Cm*^{R}
17	pVRc-LacI-L251A-A1*	6029	*Cm*^{R}
18	pBAD-LacI-L251A-A1	5178	*Amp*^{R}
19	pVRc-LacI-L251A-B1	6029	*Cm*^{R}
20	pBAD-LacI-L251A-B1	5178	*Amp*^{R}
21	pBAD-LacI-L251A-B2	5178	*Amp*^{R}
22	pBAD-LacI-L251A-B3	5178	*Amp*^{R}
23	pBAD-LacI-L251A-B1*	5178	*Amp*^{R}
24	pBAD-LacI-ΔHTH-B1	5037	*Amp*^{R}
25	pBAD-LacI-ΔHTH-L251A-B1	5037	*Amp*^{R}
26	pBAD-6H-SUMO-B1	4410	*Amp*^{R}
27	pBAD-6H-SUMO-B2	4410	*Amp*^{R}
28	pBAD-6H-SUMO-B3	4410	*Amp*^{R}
29	pBAD-6H-T7-Xpress-B1	4185	*Amp*^{R}
30	pBAD-6H-B1	4116	*Amp*^{R}
31	pBAD-6H-B2	4116	*Amp*^{R}
32	pBAD-6H-B3	4116	*Amp*^{R}

33	pBAD-LacI-L251A-Y282A-Wttet	5152	*Amp*^{R}
34	pBAD-LacI-L251A-Y282D-Wttet	5152	*Amp*^{R}
35	pBAD-LacI-L251A-Y282A	5092	*Amp*^{R}
36	pBAD-LacI-L251A-Y282D	5092	*Amp*^{R}
37	pVRc-LacI-L251A-Y282A	5943	*Cm*^{R}
38	pBAD-LacI-L251A-Y282A-ccDi	5178	*Amp*^{R}
39	pBAD-LacI-L251A-Y282D-ccDi	5178	*Amp*^{R}
40	pVRc-LacI-L251A-Y282A-A1	6029	*Cm*^{R}
41	pBAD-LacI-L251A-Y282A-B1	5178	*Amp*^{R}
42	pBAD-pro1-LacI-L251A	4046	*Amp*^{R}
43	pBAD-pro1-LacI-L251A-ccDi	4132	*Amp*^{R}
44	pVRc-pro1-LacI-L251A	4897	*Cm*^{R}
45	pVRc-pro1-LacI-L251A-A1	4983	*Cm*^{R}
46	pVRc-pro1-LacI-L251A-Y282A	4897	*Cm*^{R}
47	pVRc-pro1-LacI-L251A-Y282A-A1	4983	*Cm*^{R}
48	pBAD-pro5-LacI-L251A	4046	*Amp*^{R}
49	pBAD-pro5-LacI-L251A-ccDi	4132	*Amp*^{R}
50	pE-SUMOpro-Kan-LbCpf1	9285	*Kan*^{R}

Table 8-6 Names, sizes and antibiotic resistance markers of all plasmids discussed in this thesis. Plasmid sizes are in bp. *Kan*^{R} encodes an aminoglycoside phosphotransferase for resistance to kanamycin (an inhibitor of ribosomal synthesis in gram negative bacteria, binds 30S ribosomal subunit). *Cm*^{R} encodes chloramphenicol acetyltransferase for resistance to chloramphenicol (an inhibitor of ribosomal synthesis in gram negative bacteria, binds 50S ribosomal subunit). *Amp*^{R} encodes β-lactamase for resistance to ampicillin (an inhibitor of cell wall biogenesis in gram negative bacteria).

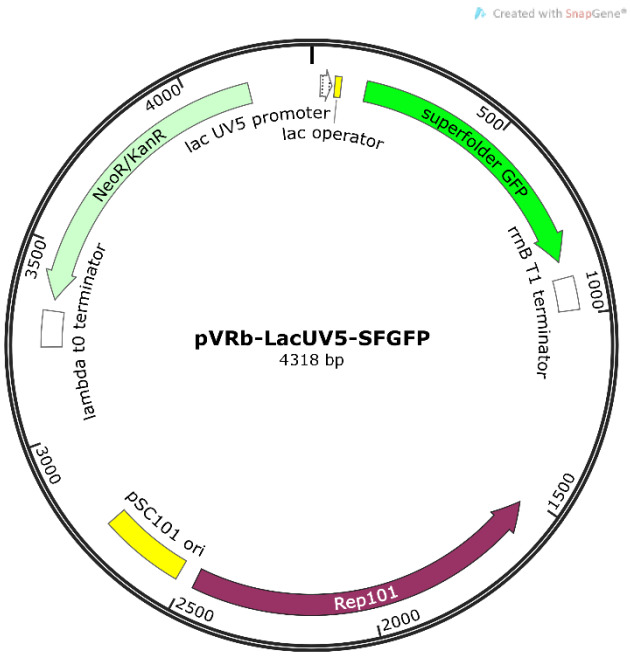


Figure 8-137 Map of plasmid pVRb-LacUV5-SFGFP.

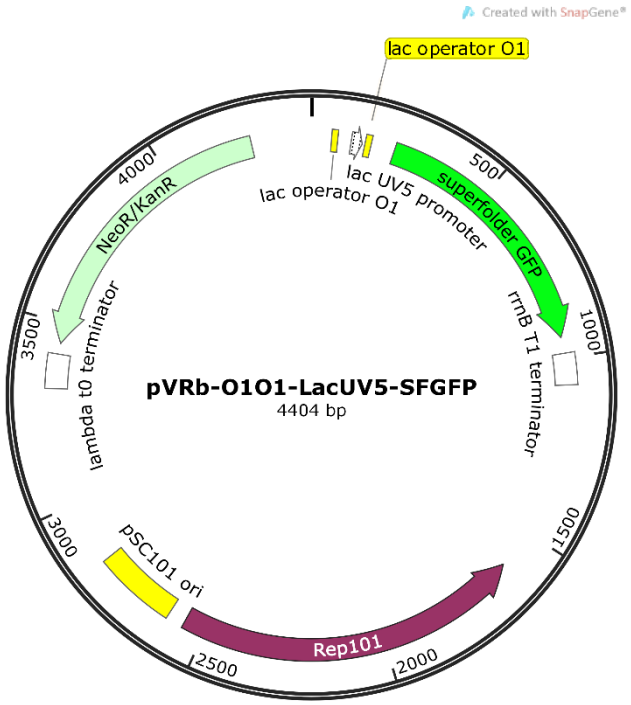


Figure 8-138 Map of plasmid pVRb-O1O1-LacUV5-SFGFP.

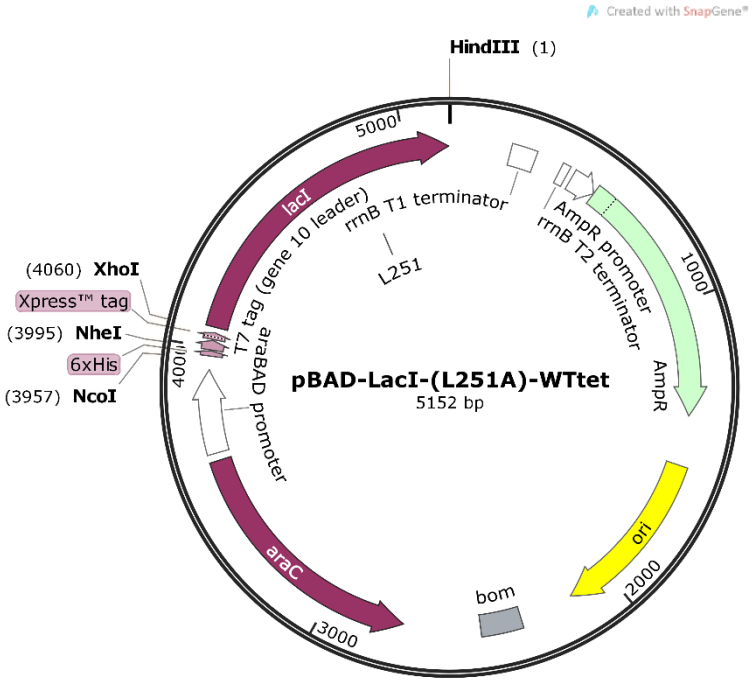


Figure 8-139 Map for plasmids pBAD-LacI-WT (carries wild type lac repressor gene) and pBAD-LacI-L251A-WTtet (carries an L251A substitution in the Lac repressor protein). Residue 251 is labelled.

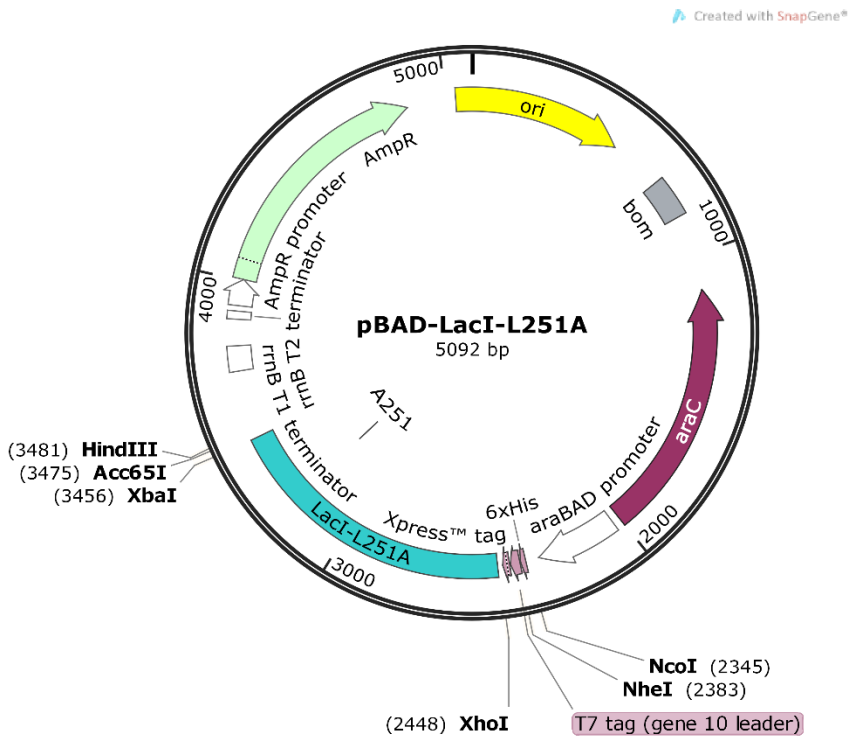


Figure 8-140 Map of plasmid pBAD-LacI-L251A. Residue 251 is labelled.

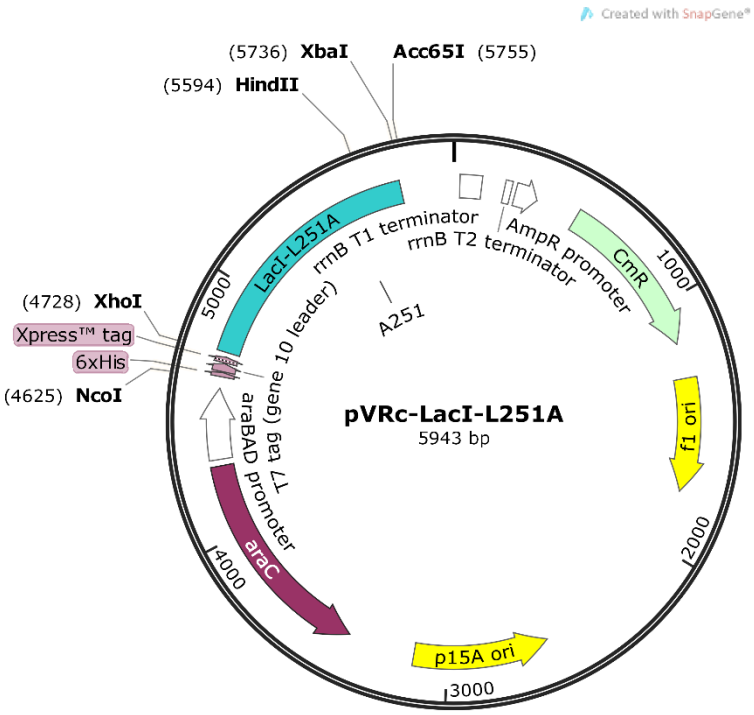


Figure 8-141 Map of plasmid pVRc-LacI-L251A. Residue 251 is labelled.

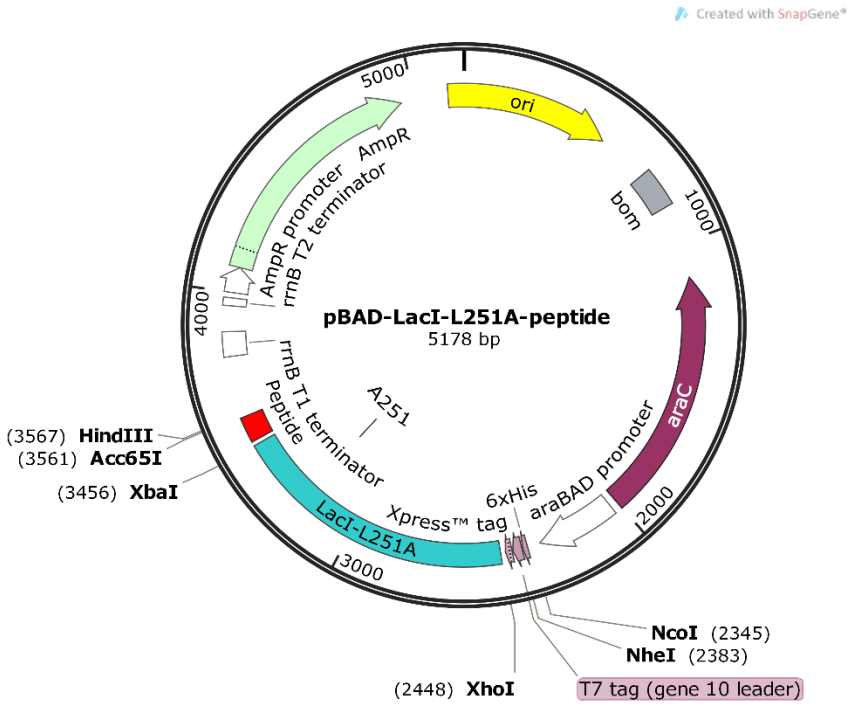


Figure 8-142 Map of plasmids pBAD-LacI-L251A-A1, pBAD-LacI-L251A-B1, pBAD-LacI-L251A-B2, pBAD-LacI-L251A-B3, pBAD-LacI-L251A-B1* and pBAD-LacI-L251A-ccDi where “peptide” represents the position of the genes for A1, B1, B2, B3, B1* and ccDi, respectively. Residue 251 is labelled.

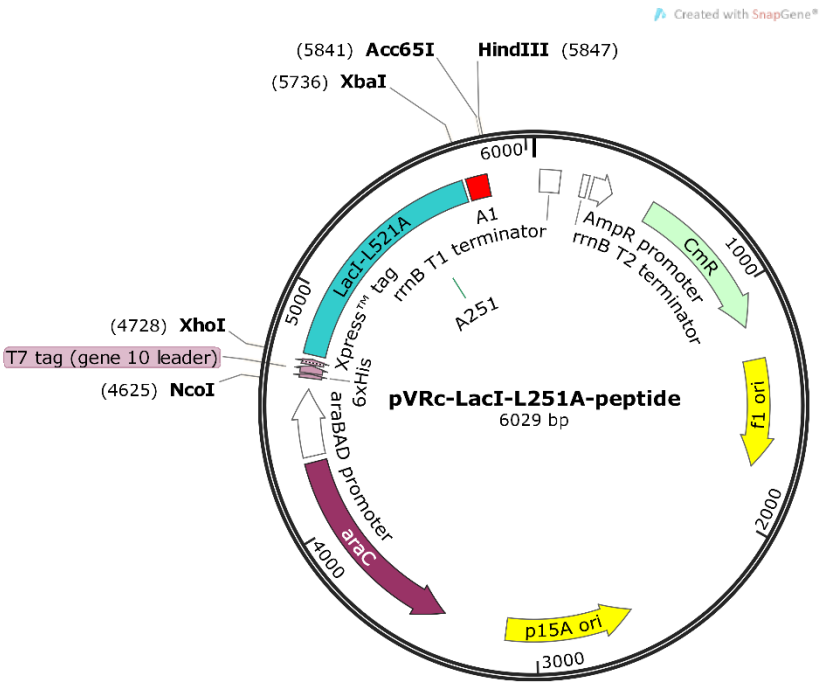


Figure 8-143 Map of plasmids pVRc-LacI-L251A-A1, pVRc-LacI-L251A-A2, pVRc-LacI-L251A-A3, pVRc-LacI-L251A-A1* and pVRc-LacI-L251A-B1 were "peptide" represents the position of the genes for A1, A2, A3, A1* and B1, respectively. Residue 251 is labelled.

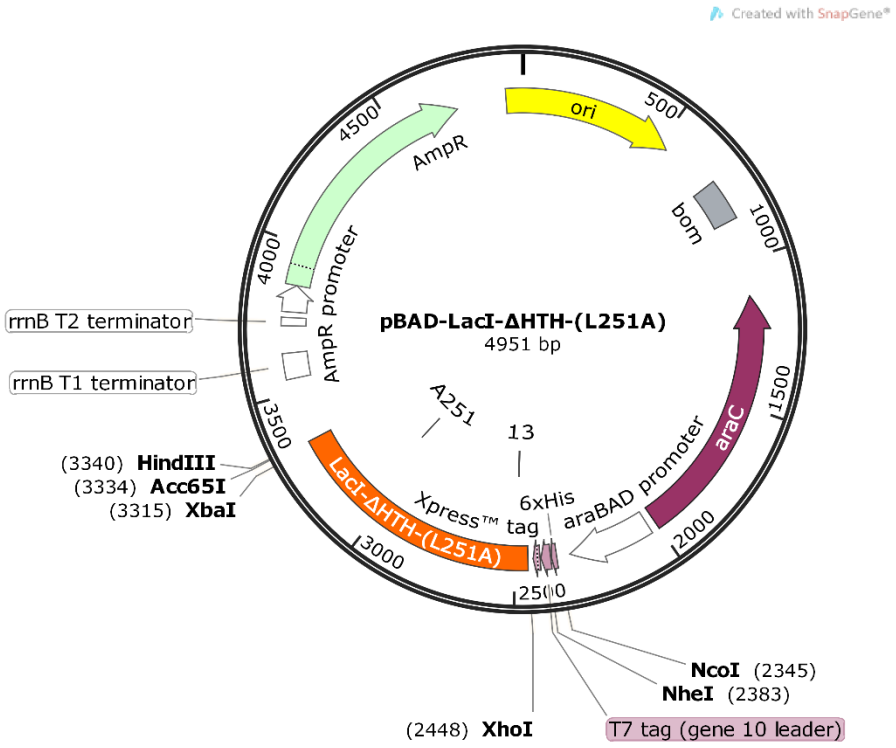


Figure 8-144 Map of plasmids pBAD-LacI-ΔHTH and pBAD-LacI-ΔHTH-L251A. Residue 251 is labelled.

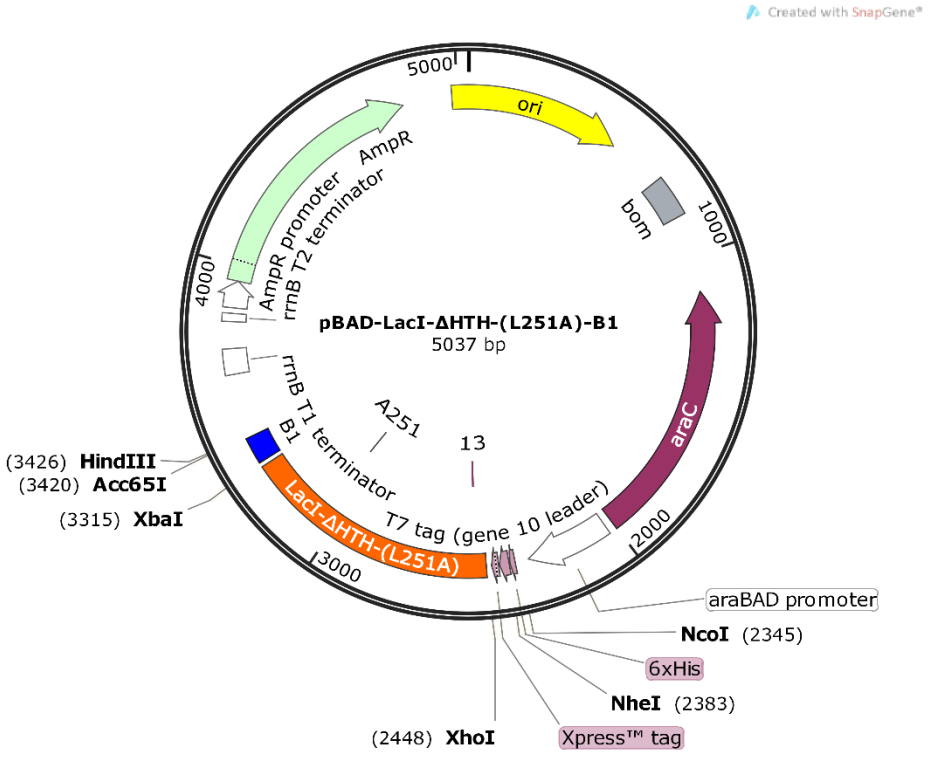


Figure 8-145 Map of plasmids pBAD-LacI-ΔHTH-B1 and pBAD-LacI-ΔHTH-L251A-B1. Residue 251 is labelled.

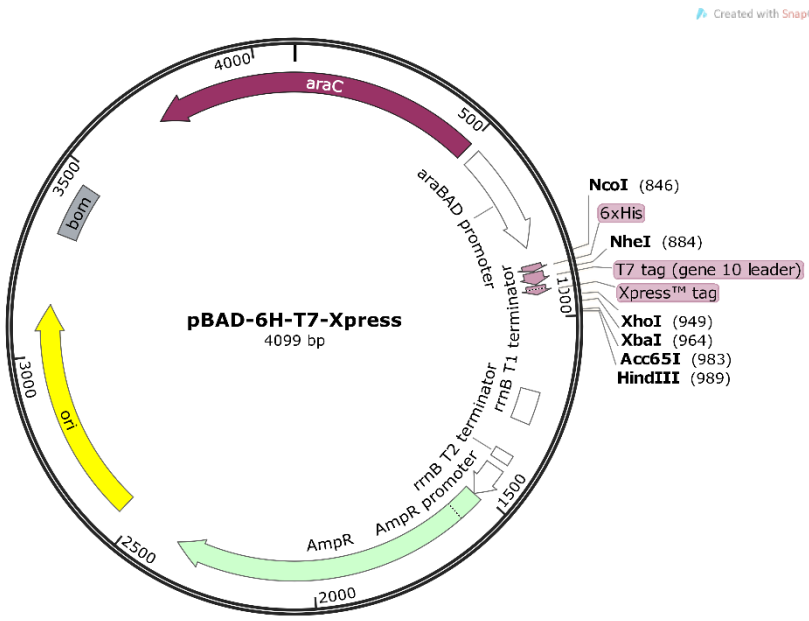


Figure 8-146 Map of plasmid pBAD-6H-T7-Xpress.

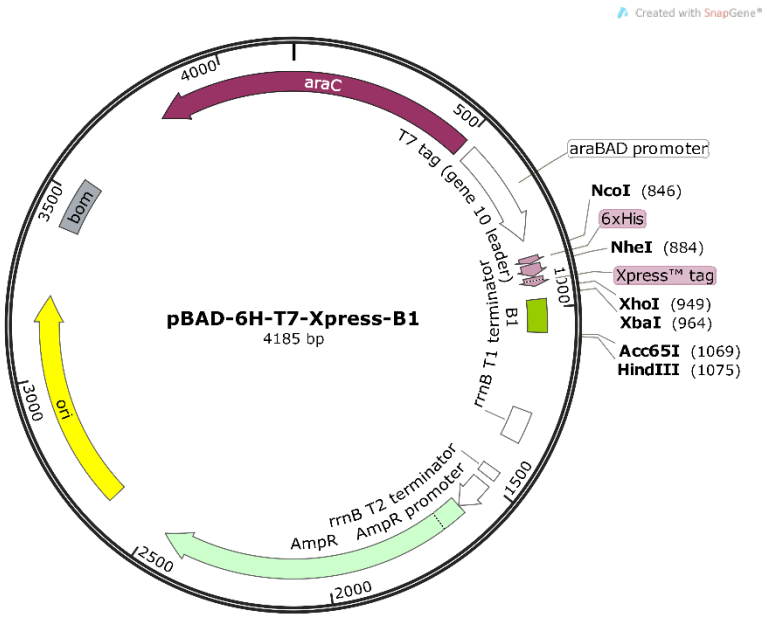


Figure 8-147 Map of plasmid pBAD-6H-T7-Xpress-B1.

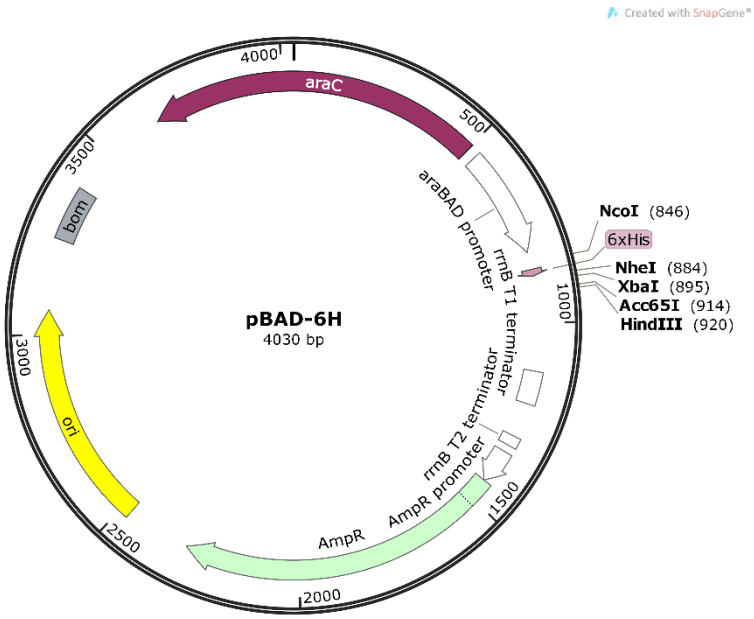


Figure 8-148 Map of plasmid pBAD-6H.

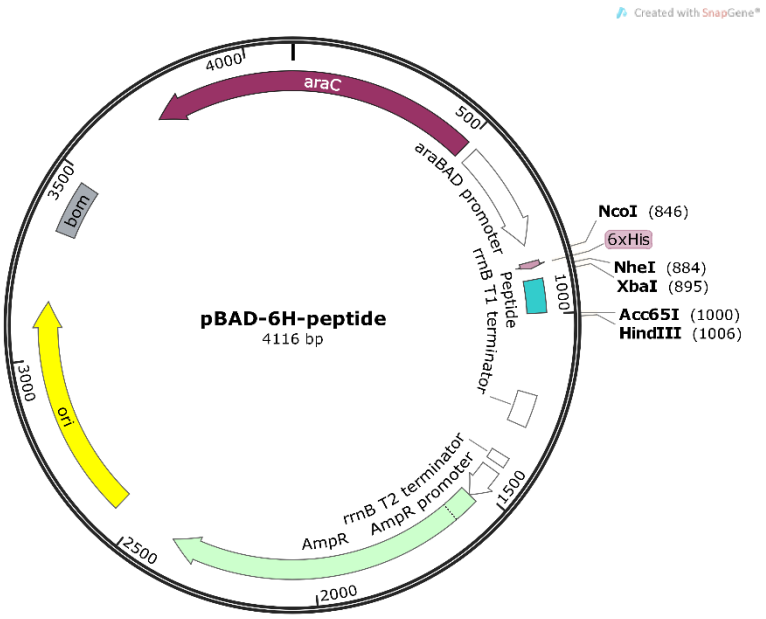


Figure 8-149 Map of plasmids pBAD-6H-B1, pBAD-6H-B2 and pBAD-6H-B3 where “peptide” represents the position of the gene B1, B2 and B3, respectively.

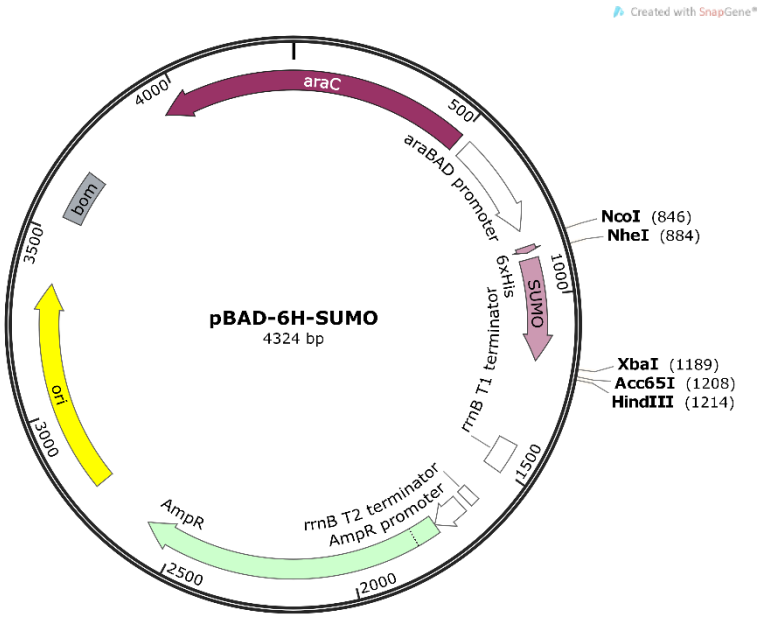


Figure 8-150 Map of plasmid pBAD-6H-SUMO.

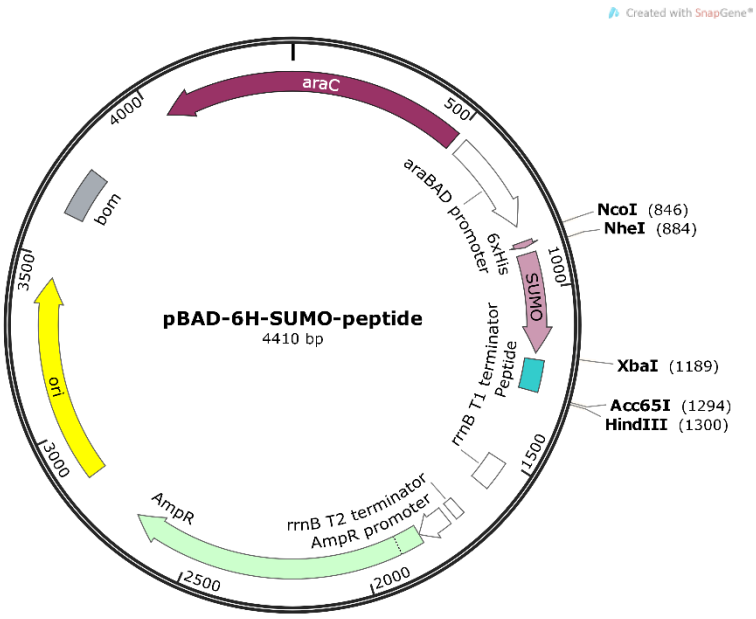


Figure 8-151 Map of plasmids pBAD-6H-SUMO-B1, pBAD-6H-SUMO-B2 and pBAD-6H-SUMO-B3 where “peptide” represents the position of the gene B1, B2 and B3, respectively.

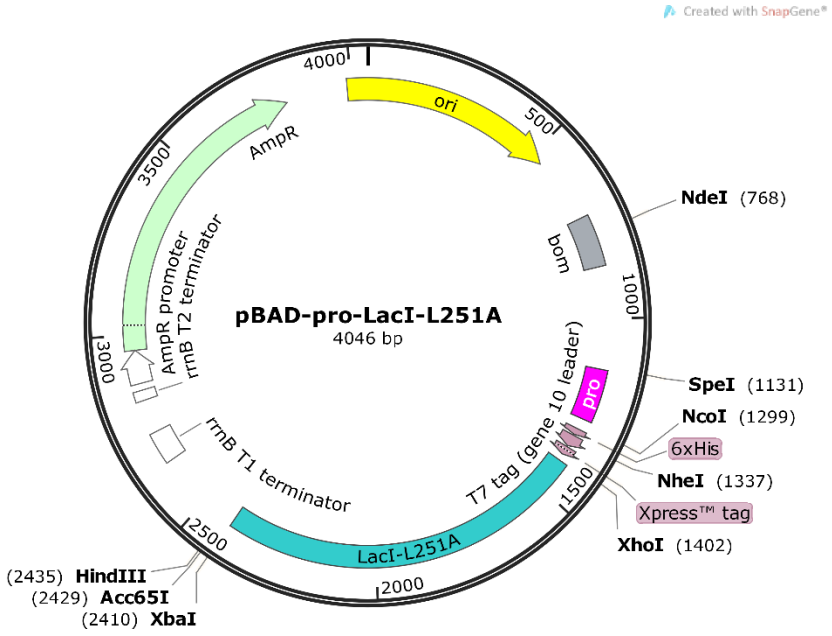


Figure 8-152 Map of plasmids pBAD-pro1-LacI-L251A and pBAD-pro5-LacI-L251A where “pro” represents the position of the constitutive promoters pro1 and pro5, respectively.

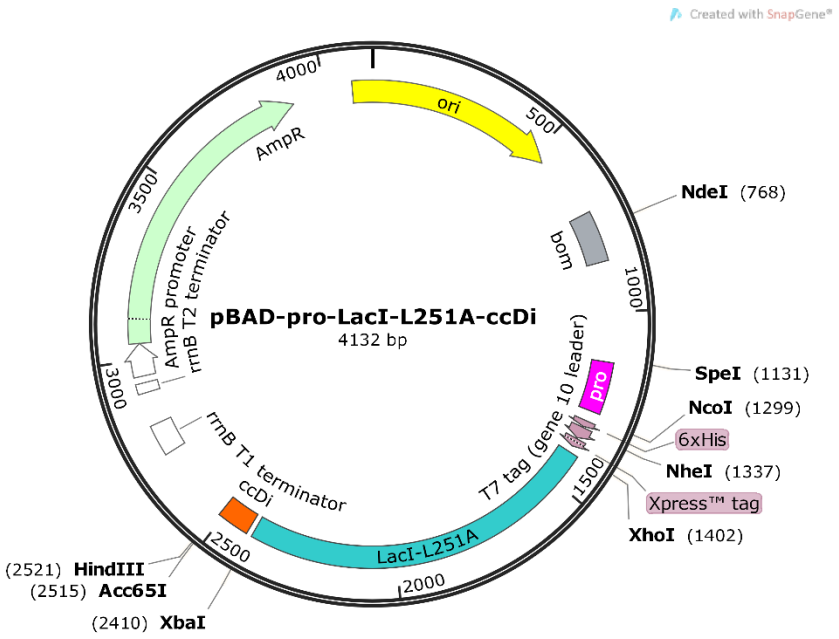


Figure 8-153 Map of plasmids pBAD-pro1-LacI-L251A-ccDi and pBAD-pro5-LacI-L251A-ccDi where "pro" represents the position of the constitutive promoters pro1 and pro5, respectively.

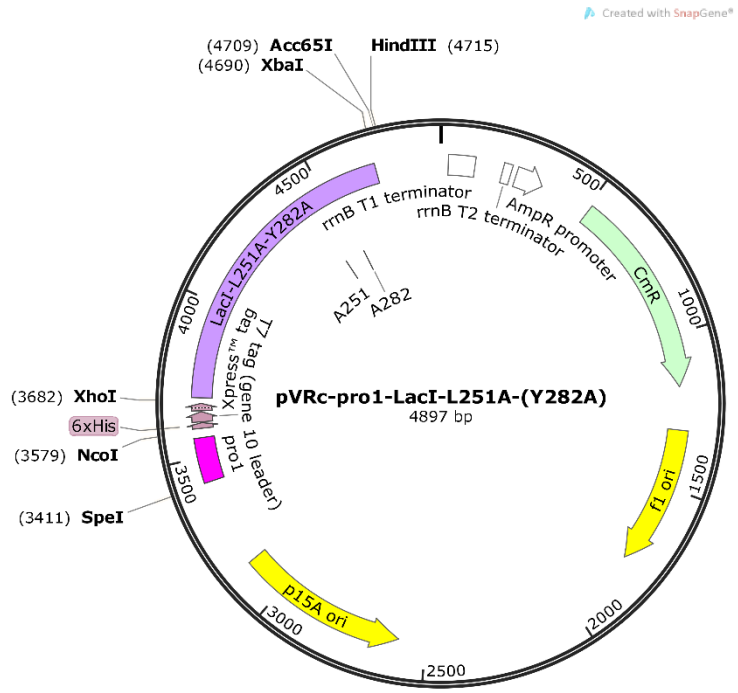


Figure 8-154 Map of plasmids pVRc-pro1-LacI-L251A and pVRc-pro1-LacI-L251A-Y282A. Residues 251 and 282 are labelled.

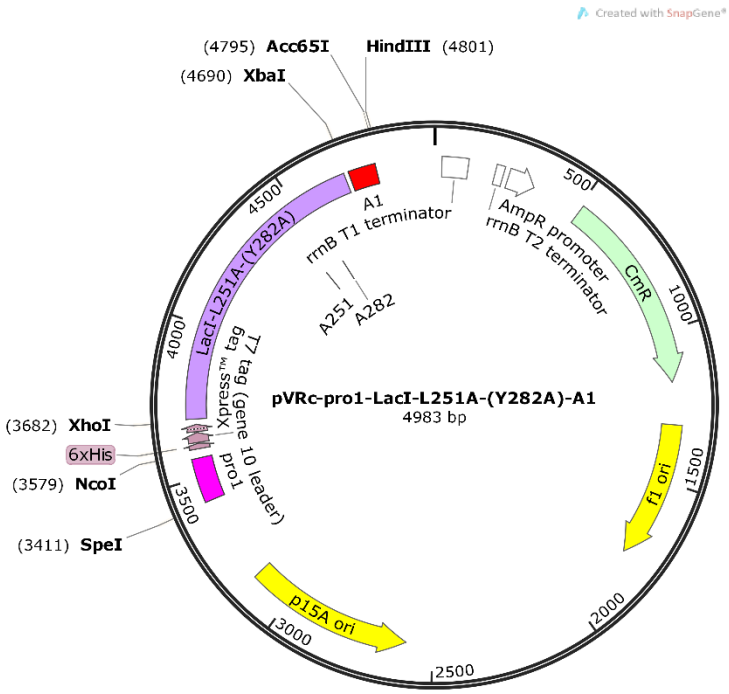


Figure 8-155 Map of plasmids pVRc-pro1-LacI-L251A-A1 and pVRc-pro1-LacI-L251A-Y282A-A1. Residues 251 and 282 are labelled.

8.8 Proteins and linkers

Protein	Size (kDa)
LacI (WT)	42.3
LacI-Trunc	39.8
LacI-L251A (LacI*-WTtet)	42.2
LacI-L251A-Trunc (LacI*)	39.7
LacI-L251A-A1 (LacI*-A1)	42.9
LacI-L251A-A2 (LacI*-A2)	43.0
LacI-L251A-A3 (LacI*-A3)	43.0
LacI-L251A-A1* (LacI*-A1*)	42.8
LacI-L251A-B1 (LacI*-B1)	42.9
LacI-L251A-B2 (LacI*-B2)	43.0
LacI-L251A-B3 (LacI*-B3)	43.0
LacI-L251A-B1* (LacI*-B1*)	42.8
LacI-L251A-ccDi (LacI*-ccDi)	42.9
LacI-ΔHTH-Trunc (LacI-ΔHTH)	34.7
LacI-ΔHTH-L251A-Trunc (LacI*-ΔHTH)	34.6
LacI-ΔHTH-B1 (LacI-ΔHTH-B1)	37.8
LacI-ΔHTH-L251A-B1 (LacI*-ΔHTH-B1)	37.7
6H-SUMO (SUMO)	13.1
6H-SUMO-B1 (SUMO-B1)	16.2
6H-SUMO-B2 (SUMO-B2)	16.4
6H-SUMO-B3 (SUMO-B3)	16.4
6H	1.8
6H-B1	4.9
6H-B2	5.1
6H-B3	5.1
6H-T7-Xpress (6H-Tags)	4.3
6H-T7-Xpress-B1 (6H-Tags-B1)	7.5
LacI-L251A-Y282A (LacI^{AA}-WTtet)	42.1
LacI-L251A-Y282D (LacI^{AD}-WTtet)	42.2
LacI-L251A-Y282A-Trunc (LacI^{AA})	39.6
LacI-L251A-Y282D-Trunc (LacI^{AD})	39.7
LacI-L251A-Y282A-ccDi (LacI^{AA}-ccDi)	42.8
LacI-L251A-Y282D-ccDi (LacI^{AD}-ccDi)	42.8
LacI-L251A-Y282A-A1 (LacI^{AA}-A1)	42.8
LacI-L251A-Y282A-B1 (LacI^{AA}-B1)	42.8

Table 8-7 Names and masses (kDa) of all proteins discussed in this thesis.

Figure 8-156 Representative C-terminal sequences demonstrating protein linker sequences. (a) Wild type Lac repressor (where linker and coiled coil regions were assigned by inspection of the crystal structure, PDB ID: 1TLF [315]). (b) C-terminally truncated Lac repressor. (c) C-terminally truncated Lac repressor fused to a *de novo* peptide. (d) SUMO. (e) SUMO fused to a *de novo* peptide. (f) 6His tag. (g) 6His tag fused to a *de novo* peptide. The sequence for *de novo* peptide B1 (aka 1-LI-B) is shown as an example. Six residues of each fusion partner (LacI, SUMO or 6H) are shown. Ellipses denote that a portion of the protein's N-terminal sequence has been omitted.

References

1 Grosskopf, T. & Soyer, O. S. Synthetic microbial communities. *Curr Opin Microbiol* **18**, 72-77, doi:10.1016/j.mib.2014.02.002 (2014).

2 Nielsen, J. & Keasling, J. D. Engineering Cellular Metabolism. *Cell* **164**, 1185-1197, doi:10.1016/j.cell.2016.02.004 (2016).

3 Huang, P. S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537**, 320-327, doi:10.1038/nature19946 (2016).

4 Channon, K., Bromley, E. H. C. & Woolfson, D. N. Synthetic biology through biomolecular design and engineering. *Curr Opin Struc Biol* **18**, 491-498, doi:10.1016/j.sbi.2008.06.006 (2008).

5 Schwille, P. Jump-starting life? Fundamental aspects of synthetic biology. *J Cell Biol* **210**, 687-690, doi:10.1083/jcb.201506125 (2015).

6 Anosova, I. *et al.* The structural diversity of artificial genetic polymers. *Nucleic Acids Res* **44**, 1007-1021, doi:10.1093/nar/gkv1472 (2016).

7 Gopfrich, K., Platzman, I. & Spatz, J. P. Mastering Complexity: Towards Bottom-up Construction of Multifunctional Eukaryotic Synthetic Cells. *Trends Biotechnol* **36**, 938-951, doi:10.1016/j.tibtech.2018.03.008 (2018).

8 Gibson, D. G. *et al.* Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science* **329**, 52-56, doi:10.1126/science.1190719 (2010).

9 Hutchison, C. A. *et al.* Design and synthesis of a minimal bacterial genome. *Science* **351**, doi:10.1126/science.aad6253 (2016).

10 Isaacs, F. J. *et al.* Precise Manipulation of Chromosomes in Vivo Enables Genome-Wide Codon Replacement. *Science* **333**, 348-353, doi:10.1126/science.1205822 (2011).

11 Ostrov, N. *et al.* Design, synthesis, and testing toward a 57-codon genome. *Science* **353**, 819-822, doi:10.1126/science.aaf3639 (2016).

12 Annaluru, N. *et al.* Total Synthesis of a Functional Designer Eukaryotic Chromosome. *Science* **344**, 55-58, doi:10.1126/science.1249252 (2014).

13 Richardson, S. M. *et al.* Design of a synthetic yeast genome. *Science* **355**, doi:10.1126/science.aaf4557 (2017).

14 Mitchell, L. A. *et al.* Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. *Science* **355**, doi:10.1126/science.aaf4831 (2017).

15 Brophy, J. A. N. & Voigt, C. A. Principles of genetic circuit design. *Nat Methods* **11**, 508-520, doi:10.1038/Nmeth.2926 (2014).

16 Den Haan, R., Rose, S. H., Lynd, L. R. & van Zyl, W. H. Hydrolysis and fermentation of amorphous cellulose by recombinant Saccharomyces cerevisiae. *Metab Eng* **9**, 87-94, doi:10.1016/j.ymben.2006.08.005 (2007).

17 Wisselink, H. W. *et al.* Engineering of Saccharomyces cerevisiae for efficient anaerobic alcoholic fermentation of L-arabinose. *Appl Environ Microb* **73**, 4881-4891, doi:10.1128/Aem.00177-07 (2007).

18 Demeke, M. M. *et al.* Development of a D-xylose fermenting and inhibitor tolerant industrial Saccharomyces cerevisiae strain with high performance in lignocellulose hydrolysates using metabolic and evolutionary engineering. *Biotechnol Biofuels* **6**, doi:10.1186/1754-6834-6-89 (2013).

19 Enquist-Newman, M. *et al.* Efficient ethanol production from brown macroalgae sugars by a synthetic yeast platform. *Nature* **505**, 239-+, doi:10.1038/nature12771 (2014).

20 Weissman, K. J. Genetic engineering of modular PKSs: from combinatorial biosynthesis to synthetic biology. *Nat Prod Rep* **33**, 203-230, doi:10.1039/c5np00109a (2016).

21 Qiao, K. *et al.* Engineering lipid overproduction in the oleaginous yeast Yarrowia lipolyticale. *Metab Eng* **29**, 56-65, doi:10.1016/j.ymben.2015.02.005 (2015).

22 Martin, V. J. J., Pitera, D. J., Withers, S. T., Newman, J. D. & Keasling, J. D. Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nat Biotechnol* **21**, 796-802, doi:10.1038/nbt833 (2003).

23 Atsumi, S. *et al.* Metabolic engineering of Escherichia coli for 1-butanol production. *Metab Eng* **10**, 305-311, doi:10.1016/j.ymben.2007.08.003 (2008).

24 Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403**, 339-342, doi:Doi 10.1038/35002131 (2000).

25 Lienert, F. *et al.* Two- and three-input TALE-based AND logic computation in embryonic stem cells. *Nucleic Acids Res* **41**, 9967-9975, doi:10.1093/nar/gkt758 (2013).

26 Friedland, A. E. *et al.* Synthetic Gene Networks That Count. *Science* **324**, 1199-1202, doi:10.1126/science.1172005 (2009).

27 Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335-338, doi:Doi 10.1038/35002125 (2000).

28 Levskaya, A. *et al.* Engineering Escherichia coli to see light - These smart bacteria 'photograph' a light pattern as a high-definition chemical image. *Nature* **438**, 441-442, doi:10.1038/nature04405 (2005).

29 Taylor, N. D. *et al.* Engineering an allosteric transcription factor to respond to new ligands. *Nat Methods* **13**, 177-+, doi:10.1038/Nmeth.3696 (2016).

30 Loakes, D., Gallego, J., Pinheiro, V. B., Kool, E. T. & Holliger, P. Evolving a Polymerase for Hydrophobic Base Analogues. *J Am Chem Soc* **131**, 14827-14837, doi:10.1021/ja9039696 (2009).

31 Korkegian, A., Black, M. E., Baker, D. & Stoddard, B. L. Computational thermostabilization of an enzyme. *Science* **308**, 857-860, doi:10.1126/science.1107387 (2005).

32 Sharma, P. K. *et al.* Engineering of a metagenome derived lipase toward thermal tolerance: Effect of asparagine to lysine mutation on the protein surface. *Gene* **491**, 264-271, doi:10.1016/j.gene.2011.09.028 (2012).

33 Walter, K. U., Vamvaca, K. & Hilvert, D. An active enzyme constructed from a 9-amino acid alphabet. *J Biol Chem* **280**, 37742-37746, doi:10.1074/jbc.M507210200 (2005).

34 Yurke, B., Turberfield, A. J., Mills, A. P., Simmel, F. C. & Neumann, J. L. A DNA-fuelled molecular machine made of DNA. *Nature* **406**, 605-608, doi:Doi 10.1038/35020524 (2000).

35 Goodman, R. P., Berry, R. M. & Turberfield, A. J. The single-step synthesis of a DNA tetrahedron. *Chem Commun*, 1372-1373, doi:DOI 10.1039/b402293a (2004).

36 Rothemund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297-302, doi:10.1038/nature04586 (2006).

37 McKee, M. L. *et al.* Programmable One-Pot Multistep Organic Synthesis Using DNA Junctions. *J Am Chem Soc* **134**, 1446-1449, doi:10.1021/ja2101196 (2012).

38 Pinheiro, V. B. *et al.* Synthetic Genetic Polymers Capable of Heredity and Evolution. *Science* **336**, 341-344, doi:10.1126/science.1217622 (2012).

39 Taylor, A. I. *et al.* Catalysts from synthetic genetic polymers. *Nature* **518**, 427-430, doi:10.1038/nature13982 (2015).

40 Taylor, A. I. *et al.* Nanostructures from Synthetic Genetic Polymers. *Chembiochem* **17**, 1107-1110, doi:10.1002/cbic.201600136 (2016).

41 Kuhlman, B. *et al.* Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364-1368, doi:DOI 10.1126/science.1089427 (2003).

42 Huang, P. S. *et al.* High thermodynamic stability of parametrically designed helical bundles. *Science* **346**, 481-485, doi:10.1126/science.1257481 (2014).

43 Thomson, A. R. *et al.* Computational design of water-soluble alpha-helical barrels. *Science* **346**, 485-488, doi:10.1126/science.1257452 (2014).

44 Boyken, S. E. *et al.* De novo design of protein homo-oligomers with modular hydrogen bond network-mediated specificity. *Protein Sci* **25**, 52-53 (2016).

45 Summerer, D. *et al.* A genetically encoded fluorescent amino acid. *P Natl Acad Sci USA* **103**, 9785-9789, doi:DOI 10.1073/pnas.0603965103 (2006).

46 Wang, J. Y., Xie, J. M. & Schultz, P. G. A genetically encoded fluorescent amino acid. *J Am Chem Soc* **128**, 8738-8739, doi:10.1021/ja062666k (2006).

47 Liu, D. R., Magliery, T. J., Pasternak, M. & Schultz, P. G. Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. *P Natl Acad Sci USA* **94**, 10092-10097, doi:DOI 10.1073/pnas.94.19.10092 (1997).

48 Wang, L., Magliery, T. J., Liu, D. R. & Schultz, P. G. A new functional suppressor tRNA/aminoacyl-tRNA synthetase pair for the in vivo incorporation of unnatural amino acids into proteins. *J Am Chem Soc* **122**, 5010-5011, doi:DOI 10.1021/ja000595y (2000).

49 Cropp, T. A., Anderson, J. C. & Chin, J. W. Reprogramming the amino-acid substrate specificity of orthogonal aminoacyl-tRNA synthetases to expand the genetic code of eukaryotic cells. *Nat Protoc* **2**, 2590-2600, doi:10.1038/nprot.2007.378 (2007).

50 Sachdeva, A., Wang, K. H., Elliott, T. & Chin, J. W. Concerted, Rapid, Quantitative, and Site-Specific Dual Labeling of Proteins. *J Am Chem Soc* **136**, 7785-7788, doi:10.1021/ja4129789 (2014).

51 De Poli, M. *et al.* Conformational photoswitching of a synthetic peptide foldamer bound within a phospholipid bilayer. *Science* **352**, 575-580, doi:10.1126/science.aad8352 (2016).

52 Stanton, B. C. *et al.* Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nat Chem Biol* **10**, 99-105 (2014).

53 Vassylyev, D. G. *et al.* Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 angstrom resolution. *Nature* **417**, 712-719, doi:DOI 10.1038/nature752 (2002).

54 Murakami, K. S. X-ray Crystal Structure of Escherichia coli RNA Polymerase sigma(70) Holoenzyme. *J Biol Chem* **288**, 9126-9134, doi:10.1074/jbc.M112.430900 (2013).

55 Busby, S. & Ebright, R. H. Promoter Structure, Promoter Recognition, and Transcription Activation in Prokaryotes. *Cell* **79**, 743-746, doi:Doi 10.1016/0092-8674(94)90063-9 (1994).

56 Zhang, G. Y. *et al.* Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 angstrom resolution. *Cell* **98**, 811-824, doi:Doi 10.1016/S0092-8674(00)81515-9 (1999).

57 Zaychikov, E. *et al.* Mapping of catalytic residues in the RNA polymerase active center. *Science* **273**, 107-109, doi:DOI 10.1126/science.273.5271.107 (1996).

58 Ebright, R. H. & Busby, S. The Escherichia coli RNA polymerase alpha subunit: structure and function. *Curr Opin Genet Dev* **5**, 197-203 (1995).

59 Kimura, M. & Ishihama, A. Functional Map of the Alpha-Subunit of Escherichia-Coli Rna-Polymerase - Amino-Acid Substitution within the Amino-Terminal Assembly Domain. *J Mol Biol* **254**, 342-349, doi:DOI 10.1006/jmbi.1995.0621 (1995).

60 Ito, K., Iwakura, Y. & Ishihama, A. Biosynthesis of Rna-Polymerase in Escherichia-Coli .3. Identification of Intermediates in Assembly of Rna-Polymerase. *J Mol Biol* **96**, 257-271, doi:Doi 10.1016/0022-2836(75)90347-2 (1975).

61 Ross, W. *et al.* A 3rd Recognition Element in Bacterial Promoters - DNA-Binding by the Alpha-Subunit of Rna-Polymerase. *Science* **262**, 1407-1413, doi:DOI 10.1126/science.8248780 (1993).

62 Estrem, S. T. *et al.* Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit. *Gene Dev* **13**, 2134-2147, doi:DOI 10.1101/gad.13.16.2134 (1999).

63 Gourse, R. L., Ross, W. & Gaal, T. UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. *Mol Microbiol* **37**, 687-695, doi:DOI 10.1046/j.1365-2958.2000.01972.x (2000).

64 Severinov, K. RNA polymerase structure-function: insights into points of transcriptional regulation. *Curr Opin Microbiol* **3**, 118-125, doi:Doi 10.1016/S1369-5274(00)00062-X (2000).

65 Igarashi, K. & Ishihama, A. Bipartite Functional Map of the Escherichia-Coli Rna Polymerase-Alpha Subunit - Involvement of the C-Terminal Region in Transcription Activation by Camp-Crp. *Cell* **65**, 1015-1022, doi:Doi 10.1016/0092-8674(91)90553-B (1991).

66 Blatter, E. E., Ross, W., Tang, H., Gourse, R. L. & Ebright, R. H. Domain Organization of Rna-Polymerase Alpha-Subunit - C-Terminal-85 Amino-Acids Constitute a Domain Capable of Dimerization and DNA-Binding. *Cell* **78**, 889-896, doi:Doi 10.1016/S0092-8674(94)90682-3 (1994).

67 Jeon, Y. H., Yamazaki, T., Otomo, T., Ishihama, A. & Kyogoku, Y. Flexible linker in the RNA polymerase alpha subunit facilitates the independent motion of the C-terminal activator contact domain. *J Mol Biol* **267**, 953-962, doi:DOI 10.1006/jmbi.1997.0902 (1997).

68 Gentry, D. R. & Burgess, R. R. Cross-Linking of Escherichia-Coli RNA-Polymerase Subunits - Identification of Beta' as the Binding-Site of Omega. *Biochemistry-Us* **32**, 11224-11227, doi:DOI 10.1021/bi00092a036 (1993).

69 Mukherjee, K. & Chatterji, D. Studies on the omega subunit of Escherichia coli RNA polymerase - Its role in the recovery of denatured enzyme activity. *Eur J Biochem* **247**, 884-889, doi:DOI 10.1111/j.1432-1033.1997.00884.x (1997).

70 Ghosh, P., Ishihama, A. & Chatterji, D. Escherichia coli RNA polymerase subunit omega and its N-terminal domain bind full-length beta ' to facilitate

incorporation into the alpha(2)beta subassembly. *Eur J Biochem* **268**, 4621-4627, doi:DOI 10.1046/j.1432-1327.2001.02381.x (2001).

71 Mathew, R. & Chatterji, D. The evolving story of the omega subunit of bacterial RNA polymerase. *Trends Microbiol* **14**, 450-455, doi:10.1016/j.tim.2006.08.002 (2006).

72 Burgess, R. R., Travers, A. A., Dunn, J. J. & Bautz, E. K. F. Factor Stimulating Transcription by Rna Polymerase. *Nature* **221**, 43-&, doi:DOI 10.1038/221043a0 (1969).

73 Sharma, U. K. & Chatterji, D. Transcriptional switching in Escherichia coli during stress and starvation by modulation of Sigma 70 activity. *Fems Microbiol Rev* **34**, 646-657, doi:10.1111/j.1574-6976.2010.00223.x (2010).

74 Paget, M. S. B. & Helmann, J. D. Protein family review - The sigma(70) family of sigma factors. *Genome Biol* **4**, doi:DOI 10.1186/gb-2003-4-1-203 (2003).

75 Wigneshweraraj, S. *et al.* Modus operandi of the bacterial RNA polymerase containing the sigma(54) promoter-specificity factor. *Mol Microbiol* **68**, 538-546, doi:10.1111/j.1365-2958.2008.06181.x (2008).

76 Campbell, E. A. *et al.* Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Mol Cell* **9**, 527-539, doi:Doi 10.1016/S1097-2765(02)00470-7 (2002).

77 Murakami, K. S. & Darst, S. A. Bacterial RNA polymerases: the wholo story. *Curr Opin Struc Biol* **13**, 31-39, doi:10.1016/S0959-440x(02)00005-2 (2003).

78 Lee, J. & Borukhov, S. Bacterial RNA Polymerase-DNA Interaction: The Driving Force of Gene Expression and the Target for Drug Action. *Frontiers in Molecular Bioscience* **3**, doi:10.3389/fmolb.2016.00073 (2016).

79 Saecker, R. M., Record, M. T. & Dehaseth, P. L. Mechanism of Bacterial Transcription Initiation: RNA Polymerase - Promoter Binding, Isomerization to Initiation-Competent Open Complexes, and Initiation of RNA Synthesis. *J Mol Biol* **412**, 754-771, doi:10.1016/j.jmb.2011.01.018 (2011).

80 Bustamante, C., Guthold, M., Zhu, X. S. & Yang, G. L. Facilitated target location on DNA by individual Escherichia coli RNA polymerase molecules observed with the scanning force microscope operating in liquid. *J Biol Chem* **274**, 16665-16668, doi:DOI 10.1074/jbc.274.24.16665 (1999).

81 Feklistov, A. RNA polymerase: in search of promoters. *Ann Ny Acad Sci* **1293**, 25-32, doi:10.1111/nyas.12197 (2013).

82 Murakami, K. S., Masuda, S. & Darst, S. A. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 angstrom resolution. *Science* **296**, 1280-1284, doi:DOI 10.1126/science.1069594 (2002).

83 Zuo, Y. H. & Steitz, T. A. Crystal Structures of the E. coli Transcription Initiation Complexes with a Complete Bubble. *Mol Cell* **58**, 534-540, doi:10.1016/j.molcel.2015.03.010 (2015).

84 Carpousis, A. J. & Gralla, J. D. Cycling of Ribonucleic-Acid Polymerase to Produce Oligonucleotides during Initiation Invitro at the Lac Uv5 Promoter. *Biochemistry-Us* **19**, 3245-3253, doi:DOI 10.1021/bi00555a023 (1980).

85 Landick, R. The regulatory roles and mechanism of transcriptional pausing. *Biochem Soc T* **34**, 1062-1066, doi:Doi 10.1042/Bst0341062 (2006).

86 Washburn, R. S. & Gottesman, M. E. Regulation of Transcription Elongation and Termination. *Biomolecules* **5**, 1063-1078, doi:10.3390/biom5021063 (2015).

87 Lowerygo.C & Richards.Jp. Rna-Dependent Nucleoside Triphosphate Phosphohydrolase (Atpase) Associated with Rho Termination Factor. *P Natl Acad Sci USA* **71**, 2003-2007, doi:DOI 10.1073/pnas.71.5.2003 (1974).

88 Brennan, C. A., Dombroski, A. J. & Platt, T. Transcription Termination Factor-Rho Is an Rna-DNA Helicase. *Cell* **48**, 945-952, doi:Doi 10.1016/0092-8674(87)90703-3 (1987).

89 Farnham, P. J. & Platt, T. Rho-Independent Termination - Dyad Symmetry in DNA Causes Rna-Polymerase to Pause during Transcription Invitro. *Nucleic Acids Res* **9**, 563-577, doi:DOI 10.1093/nar/9.3.563 (1981).

90 Wilson, K. S. & Vonhippel, P. H. Transcription Termination at Intrinsic Terminators - the Role of the Rna Hairpin. *P Natl Acad Sci USA* **92**, 8793-8797, doi:DOI 10.1073/pnas.92.19.8793 (1995).

91 Gruber, T. M. & Gross, C. A. Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu Rev Microbiol* **57**, 441-466, doi:10.1146/annurev.micro.57.030502.090913 (2003).

92 Grossman, A. D., Erickson, J. W. & Gross, C. A. The Htpr Gene-Product of Escherichia-Coli Is a Sigma-Factor for Heat-Shock Promoters. *Cell* **38**, 383-390, doi:Doi 10.1016/0092-8674(84)90493-8 (1984).

93 Grossman, A. D., Straus, D. B., Walter, W. A. & Gross, C. A. Sigma-32 Synthesis Can Regulate the Synthesis of Heat-Shock Proteins in Escherichia-Coli. *Gene Dev* **1**, 179-184, doi:DOI 10.1101/gad.1.2.179 (1987).

94 Straus, D. B., Walter, W. A. & Gross, C. A. The Heat-Shock Response of Escherichia-Coli Is Regulated by Changes in the Concentration of Sigma-32. *Nature* **329**, 348-351, doi:DOI 10.1038/329348a0 (1987).

95 Morita, M. T. *et al.* Translational induction of heat shock transcription factor sigma(32): evidence for a built-in RNA thermosensor. *Gene Dev* **13**, 655-665, doi:DOI 10.1101/gad.13.6.655 (1999).

96 Malik, S., Zalenskaya, K. & Goldfarb, A. Competition between Sigma Factors for Core Rna-Polymerase. *Nucleic Acids Res* **15**, 8521-8530, doi:DOI 10.1093/nar/15.20.8521 (1987).

97 Maeda, H., Fujita, N. & Ishihama, A. Competition among seven Escherichia coli sigma subunits: relative binding affinities to the core RNA polymerase. *Nucleic Acids Res* **28**, 3497-3503, doi:DOI 10.1093/nar/28.18.3497 (2000).

98 Nonaka, G., Blankschien, M., Herman, C., Gross, C. A. & Rhodius, V. A. Regulon and promoter analysis of the E-coli heat-shock factor, sigma(32), reveals a multifaceted cellular response to heat stress. *Gene Dev* **20**, 1776-1789, doi:DOI 10.1101/gad.1428206 (2006).

99 Tripathi, L., Zhang, Y. & Lin, Z. Bacterial Sigma Factors as Targets for Engineered or Synthetic Transcriptional Control. *Front Bioeng Biotechnol* **2**, doi:10.3389/fbioe.2014.00033 (2014).

100 Bervoets, I. *et al.* A sigma factor toolbox for orthogonal gene expression in Escherichia coli. *Nucleic Acids Res* **46**, 2133-2144, doi:10.1093/nar/gky010 (2018).

101 Rhodius, V. A. *et al.* Design of orthogonal genetic switches based on a crosstalk map of sigmas, anti-sigmas, and promoters. *Mol Syst Biol* **9**, doi:10.1038/msb.2013.58 (2013).

102 Perez-Rueda, E. & Collado-Vides, J. The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12. *Nucleic Acids Res* **28**, 1838-1847, doi:10.1093/nar/28.8.1838 (2000).

103 Martinez-Antonio, A. & Collado-Vides, J. Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr Opin Microbiol* **6**, 482-489, doi:10.1016/j.mib.2003.09.002 (2003).

104 Shimada, T., Ogasawara, H. & Ishihama, A. Single-target regulators form a minor group of transcription factors in Escherichia coli K-12. *Nucleic Acids Res* **46**, 3921-3936, doi:10.1093/nar/gky138 (2018).

105 Babu, M. M. & Teichmann, S. A. Evolution of transcription factors and the gene regulatory network in Escherichia coli. *Nucleic Acids Res* **31**, 1234-1244, doi:10.1093/nar/gkg210 (2003).

106 Busby, S. & Ebright, R. H. Transcription activation by catabolite activator protein (CAP). *J Mol Biol* **293**, 199-213, doi:DOI 10.1006/jmbi.1999.3161 (1999).

107 Niu, W., Kim, Y., Tau, G., Heyduk, T. & Ebright, R. H. Transcription activation at class II CAP-dependent promoters: Two interactions between CAP and RNA polymerase. *Cell* **87**, 1123-1134, doi:Doi 10.1016/S0092-8674(00)81806-1 (1996).

108 Dove, S. L., Darst, S. A. & Hochschild, A. Region 4 of sigma as a target for transcription regulation. *Mol Microbiol* **48**, 863-874, doi:DOI 10.1046/j.1365-2958.2003.03467.x (2003).

109 Dove, S. L., Joung, J. K. & Hochschild, A. Activation of prokaryotic transcription through arbitrary protein-protein contacts. *Nature* **386**, 627-630, doi:DOI 10.1038/386627a0 (1997).

110 Dove, S. L. & Hochschild, A. Conversion of the omega subunit of Escherichia coli RNA polymerase into a transcriptional activator or an activation target. *Gene Dev* **12**, 745-754, doi:DOI 10.1101/gad.12.5.745 (1998).

111 Hussey, B. J. & McMillen, D. R. Programmable T7-based synthetic transcription factors. *Nucleic Acids Res* **46**, 9842-9854, doi:10.1093/nar/gky785 (2018).

112 Brown, N. L., Stoyanov, J. V., Kidd, S. P. & Hobman, J. L. The MerR family of transcriptional regulators. *Fems Microbiol Rev* **27**, 145-163, doi:10.1016/S0168-6445(03)00051-2 (2003).

113 Wilson, C. J., Zhan, H., Swint-Kruse, L. & Matthews, K. S. The lactose repressor system: paradigms for regulation, allosteric behavior and protein folding. *Cell Mol Life Sci* **64**, 3-16, doi:10.1007/s00018-006-6296-z (2007).

114 Oehler, S., Eismann, E. R., Kramer, H. & Mullerhill, B. The 3 Operators of the Lac Operon Cooperate in Repression. *Embo J* **9**, 973-979 (1990).

115 Kramer, H. *et al.* Lac Repressor Forms Loops with Linear DNA Carrying 2 Suitably Spaced Lac Operators. *Embo J* **6**, 1481-1491 (1987).

116 Becker, N. A., Peters, J. P., Lionberger, T. A. & Maher, L. J. Mechanism of promoter repression by Lac repressor-DNA loops. *Nucleic Acids Res* **41**, 156-166, doi:10.1093/nar/gks1011 (2013).

117 Riggs, A. D., Newby, R. F. & Bourgeois, S. Lac Repressor-Operator Interaction .2. Effect of Galactosides and Other Ligands. *J Mol Biol* **51**, 303-+, doi:Doi 10.1016/0022-2836(70)90144-0 (1970).

118 Lewis, M. *et al.* Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* **271**, 1247-1254, doi:DOI 10.1126/science.271.5253.1247 (1996).

119 Bertrandburggraf, E., Hurstel, S., Daune, M. & Schnarr, M. Promoter Properties and Negative Regulation of the Uvra Gene by the Lexa Repressor and Its Amino-Terminal DNA-Binding Domain. *J Mol Biol* **193**, 293-302, doi:Doi 10.1016/0022-2836(87)90220-8 (1987).

120 Williams, D. R., Motallebiveshareh, M. & Thomas, C. M. Multifunctional Repressor Korb Can Block Transcription by Preventing Isomerization of Rna-Polymerase Promoter Complexes. *Nucleic Acids Res* **21**, 1141-1148, doi:DOI 10.1093/nar/21.5.1141 (1993).

121 Monsalve, M., Mencia, M., Rojo, F. & Salas, M. Activation and repression of transcription at two different phage Phi 29 promoters are mediated by interaction of the same residues of regulatory protein p4 with RNA polymerase. *Embo J* **15**, 383-391, doi:DOI 10.1002/j.1460-2075.1996.tb00368.x (1996).

122 Politz, M. C., Copeland, M. F. & Pfleger, B. F. Artificial repressors for controlling gene expression in bacteria. *Chem Commun* **49**, 4325-4327, doi:10.1039/c2cc37107c (2013).

123 Feng, J. *et al.* A general strategy to construct small molecule biosensors in eukaryotes. *Elife* **4**, doi:10.7554/eLife.10606 (2015).

124 Garg, A., Lohmueller, J. J., Silver, P. A. & Armel, T. Z. Engineering synthetic TAL effectors with orthogonal target sites. *Nucleic Acids Res* **40**, 7584-7595, doi:10.1093/nar/gks404 (2012).

125 Park, K. S. *et al.* Phenotypic alteration of eukaryotic cells using randomized libraries of artificial transcription factors. *Nat Biotechnol* **21**, 1208-1214, doi:10.1038/nbt868 (2003).

126 Lee, J. Y. *et al.* Phenotypic engineering by reprogramming gene transcription using novel artificial transcription factors in Escherichia coli. *Nucleic Acids Res* **36**, doi:10.1093/nar/gkn449 (2008).

127 Falke, D., Fisher, M., Ye, D. & Juliano, R. L. Design of artificial transcription factors to selectively regulate the pro-apoptotic bax gene. *Nucleic Acids Res* **31**, doi:10.1093/nar/gng010 (2003).

128 Beltran, A. S., Sun, X. G., Lizard, P. M. & Blancafort, P. Reprogramming epigenetic silencing: artificial transcription factors synergize with chromatin remodeling drugs to reactivate the tumor suppressor mammary serine protease inhibitor. *Mol Cancer Ther* **7**, 1080-1090, doi:10.1158/1535-7163.Mct-07-0526 (2008).

129 Mino, T. *et al.* Inhibition of DNA replication of human papillomavirus by artificial zinc finger proteins. *J Virol* **80**, 5405-5412, doi:10.1128/Jvi.01795-05 (2006).

130 Chang, H. J. *et al.* A Modular Receptor Platform To Expand the Sensing Repertoire of Bacteria. *Acs Synth Biol* **7**, 166-175, doi:10.1021/acssynbio.7b00266 (2018).

131 Salinas, R. K. *et al.* Altered specificity in DNA binding by the lac repressor: A mutant lac headpiece that mimics the gal repressor. *Chembiochem* **6**, 1628-1637, doi:10.1002/cbic.200500049 (2005).

132 Krueger, M., Scholz, O., Wisshak, S. & Hillen, W. Engineered Tet repressors with recognition specificity for the tetO-4C5G operator variant. *Gene* **404**, 93-100, doi:10.1016/j.gene.2007.09.002 (2007).

133 Pavletich, N. P. & Pabo, C. O. Zinc Finger DNA Recognition - Crystal-Structure of a Zif268-DNA Complex at 2.1-A. *Science* **252**, 809-817, doi:DOI 10.1126/science.2028256 (1991).

134 Klug, A. The Discovery of Zinc Fingers and Their Applications in Gene Regulation and Genome Manipulation. *Annual Review of Biochemistry, Vol 79* **79**, 213-231, doi:10.1146/annurev-biochem-010909-095056 (2010).

135 Rebar, E. J. & Pabo, C. O. Zinc-Finger Phage - Affinity Selection of Fingers with New DNA-Binding Specificities. *Science* **263**, 671-673, doi:DOI 10.1126/science.8303274 (1994).

136 Choo, Y. & Klug, A. Toward a Code for the Interactions of Zinc Fingers with DNA - Selection of Randomized Fingers Displayed on Phage. *P Natl Acad Sci USA* **91**, 11163-11167, doi:DOI 10.1073/pnas.91.23.11163 (1994).

137 Jamieson, A. C., Kim, S. H. & Wells, J. A. In-Vitro Selection of Zinc Fingers with Altered DNA-Binding Specificity. *Biochemistry-Us* **33**, 5689-5695, doi:DOI 10.1021/bi00185a004 (1994).

138 Choo, Y. & Klug, A. Selection of DNA-Binding Sites for Zinc Fingers Using Rationally Randomized DNA Reveals Coded Interactions. *P Natl Acad Sci USA* **91**, 11168-11172, doi:DOI 10.1073/pnas.91.23.11168 (1994).

139 Desjarlais, J. R. & Berg, J. M. Use of a Zinc-Finger Consensus Sequence Framework and Specificity Rules to Design Specific DNA-Binding Proteins. *P Natl Acad Sci USA* **90**, 2256-2260, doi:DOI 10.1073/pnas.90.6.2256 (1993).

140 Choo, Y. & Klug, A. Physical basis of a protein-DNA recognition code. *Curr Opin Struc Biol* **7**, 117-125, doi:Doi 10.1016/S0959-440x(97)80015-2 (1997).

141 Sera, T. & Uranga, C. Rational design of artificial zinc-finger proteins using a nondegenerate recognition code table. *Biochemistry-Us* **41**, 7074-7081, doi:10.1021/bi020095c (2002).

142 Ramirez, C. L. *et al.* Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat Methods* **5**, 374-375, doi:10.1038/nmeth0508-374 (2008).

143 Sera, T. Zinc-finger-based artificial transcription factors and their applications. *Adv Drug Deliver Rev* **61**, 513-526, doi:10.1016/j.addr.2009.03.012 (2009).

144 Boch, J. & Bonas, U. Xanthomonas AvrBs3 Family-Type III Effectors: Discovery and Function. *Annu Rev Phytopathol* **48**, 419-436, doi:10.1146/annurev-phyto-080508-081936 (2010).

145 Mak, A. N. S., Bradley, P., Bogdanove, A. J. & Stoddard, B. L. TAL effectors: function, structure, engineering and applications. *Curr Opin Struc Biol* **23**, 93-99, doi:10.1016/j.sbi.2012.11.001 (2013).

146 Mak, A. N. S., Bradley, P., Cernadas, R. A., Bogdanove, A. J. & Stoddard, B. L. The Crystal Structure of TAL Effector PthXo1 Bound to Its DNA Target. *Science* **335**, 716-719, doi:10.1126/science.1216211 (2012).

147 Gao, H. S., Wu, X. J., Chai, J. J. & Han, Z. F. Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region. *Cell Res* **22**, 1716-1720, doi:10.1038/cr.2012.156 (2012).

148 Moscou, M. J. & Bogdanove, A. J. A Simple Cipher Governs DNA Recognition by TAL Effectors. *Science* **326**, 1501-1501, doi:10.1126/science.1178817 (2009).

149 Boch, J. *et al.* Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science* **326**, 1509-1512, doi:10.1126/science.1178811 (2009).

150 Rinaldi, F. C., Doyle, L. A., Stoddard, B. L. & Bogdanove, A. J. The effect of increasing numbers of repeats on TAL effector DNA binding specificity. *Nucleic Acids Res* **45**, 6960-6970, doi:10.1093/nar/gkx342 (2017).

151 Li, Y., Moore, R., Guinn, M. & Bleris, L. Transcription activator-like effector hybrids for conditional control and rewiring of chromosomal transgene expression. *Sci Rep-Uk* **2**, doi:10.1038/srep00897 (2012).

152 Konermann, S. *et al.* Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **500**, 472-+, doi:10.1038/nature12466 (2013).

153 Mercer, A. C., Gaj, T., Sirk, S. J., Lamb, B. M. & Barbas, C. F. Regulation of Endogenous Human Gene Expression by Ligand-Inducible TALE Transcription Factors. *Acs Synth Biol* **3**, 723-730, doi:10.1021/sb400114p (2014).

154 Copeland, M. F., Politz, M. C., Johnson, C. B., Markley, A. L. & Pfleger, B. F. A transcription activator-like effector (TALE) induction system mediated by proteolysis. *Nat Chem Biol* **12**, 254-+, doi:10.1038/Nchembio.2021 (2016).

155 Kabadi, A. M. & Gersbach, C. A. Engineering synthetic TALE and CRISPR/Cas9 transcription factors for regulating gene expression. *Methods* **69**, 188-197, doi:10.1016/j.ymeth.2014.06.014 (2014).

156 Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).

157 Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res* **41**, 7429-7437, doi:10.1093/nar/gkt520 (2013).

158 ElrodErickson, M., Rould, M. A., Nekludova, L. & Pabo, C. O. Zif268 protein-DNA complex refined at 1.6 angstrom: A model system for understanding zinc finger-DNA interactions. *Structure* **4**, 1171-1180, doi:Doi 10.1016/S0969-2126(96)00125-6 (1996).

159 Younger, A. K. D., Dalvie, N. C., Rottinghaus, A. G. & Leonard, J. N. Engineering Modular Biosensors to Confer Metabolite-Responsive Regulation of Transcription. *Acs Synth Biol* **6**, 311-325, doi:10.1021/acssynbio.6b00184 (2017).

160 Schreiber, S. L. Chemistry and Biology of the Immunophilins and Their Immunosuppressive Ligands. *Science* **251**, 283-287, doi:DOI 10.1126/science.1702904 (1991).

161 Clackson, T. *et al.* Redesigning an FKBP-ligand interface to generate chemical dimerizers with novel specificity. *P Natl Acad Sci USA* **95**, 10437-10442, doi:DOI 10.1073/pnas.95.18.10437 (1998).

162 Spencer, D. M., Wandless, T. J., Schreiber, S. L. & Crabtree, G. R. Controlling Signal-Transduction with Synthetic Ligands. *Science* **262**, 1019-1024, doi:DOI 10.1126/science.7694365 (1993).

163 Rollins, C. T. *et al.* A ligand-reversible dimerization system for controlling protein-protein interactions. *P Natl Acad Sci USA* **97**, 7096-7101, doi:DOI 10.1073/pnas.100101997 (2000).

164 Becker, N. A., Schwab, T. L., Clark, K. J. & Maher, L. J. Bacterial gene control by DNA looping using engineered dimeric transcription activator like effector (TALE) proteins. *Nucleic Acids Res* **46**, 2690-2696, doi:10.1093/nar/gky047 (2018).

165 Sonneson, G. J. & Horn, J. R. Hapten-Induced Dimerization of a Single-Domain VHH Camelid Antibody. *Biochemistry-Us* **48**, 6693-6695, doi:10.1021/bi900862r (2009).

166 Himmelfarb, H. J., Pearlberg, J., Last, D. H. & Ptashne, M. Gal11p - a Yeast Mutation That Potentiates the Effect of Weak Gal4-Derived Activators. *Cell* **63**, 1299-1309, doi:Doi 10.1016/0092-8674(90)90425-E (1990).

167 Joung, J. K., Ramm, E. I. & Pabo, C. O. A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. *P Natl Acad Sci USA* **97**, 7382-7387, doi:DOI 10.1073/pnas.110149297 (2000).

168 Pomerantz, J. L., Wolfe, S. A. & Pabo, C. O. Structure-based design of a dimeric zinc finger protein. *Biochemistry-Us* **37**, 965-970, doi:DOI 10.1021/bi972464o (1998).

169 Wang, B. S. & Pabo, C. O. Dimerization of zinc fingers mediated by peptides evolved in vitro from random sequences. *P Natl Acad Sci USA* **96**, 9568-9573, doi:DOI 10.1073/pnas.96.17.9568 (1999).

170 Hou, J. R. *et al.* Engineering the Ultrasensitive Transcription Factors by Fusing a Modular Oligomerization Domain. *Acs Synth Biol* **7**, 1188-1194, doi:10.1021/acssynbio.7b00414 (2018).

171 Wolfe, S. A., Ramm, E. I. & Pabo, C. O. Combining structure-based design with phage display to create new Cys(2)His(2) zinc finger dimers. *Struct Fold Des* **8**, 739-750, doi:Doi 10.1016/S0969-2126(00)00161-1 (2000).

172 Wolfe, S. A., Grant, R. A. & Paboll, C. O. Structure of a designed dimeric zinc finger protein bound to DNA. *Biochemistry-Us* **42**, 13401-13409, doi:10.1021/bi034830b (2003).

173 Smith, A. J. *Tuning transcription regulation with de novo designed protein-protein interactions* (University of Bristol, 2018).

174 Thomas, F., Boyle, A. L., Burton, A. J. & Woolfson, D. N. A Set of de Novo Designed Parallel Heterodimeric Coiled Coils with Quantified Dissociation Constants in the Micromolar to Sub-nanomolar Regime. *J Am Chem Soc* **135**, 5161-5166, doi:10.1021/ja312310g (2013).

175 Thompson, K. E., Bashor, C. J., Lim, W. A. & Keating, A. E. SYNZIP Protein Interaction Toolbox: in Vitro and in Vivo Specifications of Heterospecific

Coiled-Coil Interaction Domains. *Acs Synth Biol* **1**, 118-129, doi:10.1021/sb200015u (2012).

176 Kostelny, S. A., Cole, M. S. & Tso, J. Y. Formation of a Bispecific Antibody by the Use of Leucine Zippers. *J Immunol* **148**, 1547-1553 (1992).

177 Pack, P., Muller, K., Zahn, R. & Pluckthun, A. Tetravalent Miniantibodies with High Avidity Assembling in Escherichia-Coli. *J Mol Biol* **246**, 28-34, doi:DOI 10.1006/jmbi.1994.0062 (1995).

178 Swainsbury, D. J. K. *et al.* Directed assembly of defined oligomeric photosynthetic reaction centres through adaptation with programmable extra-membrane coiled-coil interfaces. *Bba-Bioenergetics* **1857**, 1829-1839, doi:10.1016/j.bbabio.2016.09.002 (2016).

179 Liu, J. F. & Rost, B. Comparing function and structure between entire proteomes. *Protein Sci* **10**, 1970-1979, doi:DOI 10.1110/ps.10101 (2001).

180 Rose, A., Schraegle, S. J., Stahlberg, E. A. & Meier, I. Coiled-coil protein composition of 22 proteomes - differences and common themes in subcellular infrastructure and traffic control. *Bmc Evol Biol* **5**, doi:10.1186/1471-2148-5-66 (2005).

181 Dill, K. A. & MacCallum, J. L. The Protein-Folding Problem, 50 Years On. *Science* **338**, 1042-1046, doi:10.1126/science.1219021 (2012).

182 Gruber, M., Soding, J. & Lupas, A. N. Comparative analysis of coiled-coil prediction methods. *J Struct Biol* **155**, 140-145, doi:10.1016/j.jsb.2006.03.009 (2006).

183 Vincent, T. L., Green, P. J. & Woolfson, D. N. LOGICOIL-multi-state prediction of coiled-coil oligomeric state. *Bioinformatics* **29**, 69-76, doi:10.1093/bioinformatics/bts648 (2013).

184 Hicks, M. R., Holberton, D. V., Kowalczyk, C. & Woolfson, D. N. Coiled-coil assembly by peptides with non-heptad sequence motifs. *Fold Des* **2**, 149-158, doi:Doi 10.1016/S1359-0278(97)00021-7 (1997).

185 Hicks, M. R., Walshaw, J. & Woolfson, D. N. Investigating the tolerance of coiled-coil peptides to nonheptad sequence inserts. *J Struct Biol* **137**, 73-81, doi:10.1006/jsbi.2002.4462 (2002).

186 Offer, G., Hicks, M. R. & Woolfson, D. N. Generalized crick equations for modeling noncanonical coiled coils. *J Struct Biol* **137**, 41-53, doi:10.1006/jsbi.2002.4448 (2002).

187 Harbury, P. B., Zhang, T., Kim, P. S. & Alber, T. A Switch between 2-Stranded, 3-Stranded and 4-Stranded Coiled Coils in Gcn4 Leucine-Zipper Mutants. *Science* **262**, 1401-1407, doi:DOI 10.1126/science.8248779 (1993).

188 Oshea, E. K., Lumb, K. J. & Kim, P. S. Peptide Velcro - Design of a Heterodimeric Coiled-Coil. *Curr Biol* **3**, 658-667, doi:Doi 10.1016/0960-9822(93)90063-T (1993).

189 Monera, O. D., Zhou, N. E., Lavigne, P., Kay, C. M. & Hodges, R. S. Formation of parallel and antiparallel coiled-coils controlled by the relative positions of alanine residues in the hydrophobic core. *J Biol Chem* **271**, 3995-4001 (1996).

190 Solan, A., Ratia, K. & Fairman, R. Exploring the role of alanine in the structure of the Lac repressor tetramerization domain, a ferritin-like Alacoil. *J Mol Biol* **317**, 601-612, doi:10.1006/jmbi.2001.5427 (2002).

191 Litowski, J. R. & Hodges, R. S. Designing heterodimeric two-stranded alpha-helical coiled-coils: the effect of chain length on protein folding, stability and specificity. *J Pept Res* **58**, 477-492, doi:DOI 10.1034/j.1399-3011.2001.10972.x (2001).

192 Vu, C., Robblee, J., Werner, K. M. & Fairman, R. Effects of charged amino acids at b and c heptad positions on specificity and stability of four-chain coiled coils. *Protein Sci* **10**, 631-637, doi:DOI 10.1110/ps.41101 (2001).

193 Pauling, L., Corey, R. B. & Branson, H. R. The Structure of Proteins - 2 Hydrogen-Bonded Helical Configurations of the Polypeptide Chain. *P Natl Acad Sci USA* **37**, 205-211, doi:DOI 10.1073/pnas.37.4.205 (1951).

194 Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of Polypeptide Chain Configurations. *J Mol Biol* **7**, 95-&, doi:Doi 10.1016/S0022-2836(63)80023-6 (1963).

195 Hollingsworth, S. A. & Karplus, P. A. A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. *Biomol Concepts* **1**, 271-283, doi:10.1515/BMC.2010.022 (2010).

196 Hollingsworth, S. A., Berkholz, D. S. & Karplus, P. A. On the occurrence of linear groups in proteins. *Protein Sci* **18**, 1321-1325, doi:10.1002/pro.133 (2009).

197 Crick, F. H. C. The Fourier Transform of a Coiled-Coil. *Acta Crystallogr* **6**, 685-689, doi:Doi 10.1107/S0365110x53001952 (1953).

198 Lupas, A. N. & Gruber, M. The structure of alpha-helical coiled coils. *Adv Protein Chem* **70**, 37-+, doi:10.1016/S0065-3233(04)70003-0 (2005).

199 Truebestein, L. & Leonard, T. A. Coiled-coils: The long and short of it. *Bioessays* **38**, 903-916, doi:10.1002/bies.201600062 (2016).

200 Crick, F. H. C. The Packing of Alpha-Helices - Simple Coiled-Coils. *Acta Crystallogr* **6**, 689-697, doi:Doi 10.1107/S0365110x53001964 (1953).

201 Walshaw, J. & Woolfson, D. N. SOCKET: A program for identifying and analysing coiled-coil motifs within protein structures. *J Mol Biol* **307**, 1427-1450, doi:DOI 10.1006/jmbi.2001.4545 (2001).

202 Walshaw, J. & Woolfson, D. N. Open-and-shut cases in coiled-coil assembly: alpha-sheets and alpha-cylinders. *Protein Sci* **10**, 668-673, doi:DOI 10.1110/ps.36901 (2001).

203 Walshaw, J. & Woolfson, D. N. Extended knobs-into-holes packing in classical and complex coiled-coil assemblies. *J Struct Biol* **144**, 349-361, doi:10.1016/j.jsb.2003.10.014 (2003).

204 Woolfson, D. N., Bartlett, G. J., Bruning, M. & Thomson, A. R. New currency for old rope: from coiled-coil assemblies to alpha-helical barrels. *Curr Opin Struc Biol* **22**, 432-441, doi:10.1016/j.sbi.2012.03.002 (2012).

205 Deng, Y. Q. *et al.* Antiparallel four-stranded coiled coil specified by a 3-3-1 hydrophobic heptad repeat. *Structure* **14**, 247-255, doi:10.1016/j.str.2005.10.010 (2006).

206 Liu, J. *et al.* A parallel coiled-coil tetramer with offset helices. *Biochemistry-Us* **45**, 15224-15231, doi:10.1021/bi061914m (2006).

207 Woolfson, D. N. The design of coiled-coil structures and assemblies. *Adv Protein Chem* **70**, 79-+, doi:10.1016/S0065-3233(04)70004-2 (2005).

208 Fletcher, J. M. *et al.* A Basis Set of de Novo Coiled-Coil Peptide Oligomers for Rational Protein Design and Synthetic Biology. *Acs Synth Biol* **1**, 240-250, doi:10.1021/sb300028q (2012).

209 Pelletier, J. N., Arndt, K. M., Pluckthun, A. & Michnick, S. W. An in vivo library-versus-library selection of optimized protein-protein interactions. *Nat Biotechnol* **17**, 683-690 (1999).

210 Arndt, K. M. *et al.* A heterodimeric coiled-coil peptide pair selected in vivo from a designed library-versus-library ensemble. *J Mol Biol* **295**, 627-639, doi:DOI 10.1006/jmbi.1999.3352 (2000).

211 Das, A., Wei, Y. A., Pelczer, I. & Hecht, M. H. Binding of small molecules to cavity forming mutants of a de novo designed protein. *Protein Sci* **20**, 702-711, doi:10.1002/pro.601 (2011).

212 Cherny, I., Korolev, M., Koehler, A. N. & Hecht, M. H. Proteins from an Unevolved Library of de novo Designed Sequences Bind a Range of Small Molecules. *Acs Synth Biol* **1**, 130-138, doi:10.1021/sb200018e (2012).

213 Zaccai, N. R. *et al.* A de novo peptide hexamer with a mutable channel. *Nat Chem Biol* **7**, 935-941, doi:10.1038/Nchembio.692 (2011).

214 Woolfson, D. N. in *Fibrous Proteins: Structures and Mechanisms* Vol. 82 *Subcellular Biochemistry* (ed D. Parry, Squire, J.) Ch. 2, 35-61 (Springer International Publishing AG, 2017).

215 Aronsson, C. *et al.* Self-sorting heterodimeric coiled coil peptides with defined and tuneable self-assembly properties. *Sci Rep-Uk* **5**, doi:10.1038/srep14063 (2015).

216 Koshland, D. E., Goldbeter, A. & Stock, J. B. Amplification and Adaptation in Regulatory and Sensory Systems. *Science* **217**, 220-225, doi:DOI 10.1126/science.7089556 (1982).

217 Cherry, J. L. & Adler, F. R. How to make a Biological Switch. *J Theor Biol* **203**, 117-133, doi:DOI 10.1006/jtbi.2000.1068 (2000).

218 Bradley, R. W., Buck, M. & Wang, B. J. Recognizing and engineering digital-like logic gates and switches in gene regulatory networks. *Curr Opin Microbiol* **33**, 74-82, doi:10.1016/j.mib.2016.07.004 (2016).

219 Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183-1186, doi:DOI 10.1126/science.1070919 (2002).

220 Swain, P. S., Elowitz, M. B. & Siggia, E. D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *P Natl Acad Sci USA* **99**, 12795-12800, doi:10.1073/pnas.162041399 (2002).

221 Maienschein-Cline, M., Warmflash, A. & Dinner, A. R. Defining cooperativity in gene regulation locally through intrinsic noise. *Iet Syst Biol* **4**, 379-392, doi:10.1049/iet-syb.2009.0070 (2010).

222 Deng, Y. Q. *et al.* Self-assembly of coiled-coil tetramers in the 1.40 angstrom structure of a leucine-zipper mutant. *Protein Sci* **16**, 323-328, doi:10.1110/ps.062590807 (2007).

223 Liu, J. *et al.* Conformational specificity of the Lac repressor coiled-coil tetramerization domain. *Biochemistry-Us* **46**, 14951-14959, doi:10.1021/bi701930d (2007).

224 Root, B. C., Pellegrino, L. D., Crawford, E. D., Kokona, B. & Fairman, R. Design of a heterotetrameric coiled coil. *Protein Sci* **18**, 329-336, doi:10.1002/pro.30 (2009).

225 Fairman, R. *et al.* Characterization of a New 4-Chain Coiled-Coil - Influence of Chain-Length on Stability. *Protein Sci* **4**, 1457-1469, doi:DOI 10.1002/pro.5560040803 (1995).

226 Fairman, R. *et al.* Design of heterotetrameric coiled coils: Evidence for increased stabilization by Glu(-)-Lys(+) ion pair interactions. *Biochemistry-Us* **35**, 2824-2829, doi:DOI 10.1021/bi952784c (1996).

227 Monera, O. D., Sonnichsen, F. D., Hicks, L., Kay, C. M. & Hodges, R. S. The relative positions of alanine residues in the hydrophobic core control the formation of two-stranded or four-stranded alpha-helical coiled-coils. *Protein Eng* **9**, 353-363, doi:DOI 10.1093/protein/9.4.353 (1996).

228 Harbury, P. B., Plecs, J. J., Tidor, B., Alber, T. & Kim, P. S. High-resolution protein design with backbone freedom. *Science* **282**, 1462-1467, doi:DOI 10.1126/science.282.5393.1462 (1998).

229 Lumb, K. J. & Kim, P. S. A Buried Polar Interaction Imparts Structural Uniqueness in a Designed Heterodimeric Coiled-Coil. *Biochemistry-Us* **34**, 8642-8648, doi:DOI 10.1021/bi00027a013 (1995).

230 Kamtekar, S. & Hecht, M. H. Protein Motifs .7. The 4-Helix Bundle - What Determines a Fold. *Faseb J* **9**, 1013-1022 (1995).

231 Cohen, C. & Parry, D. A. D. Alpha-Helical Coiled Coils and Bundles - How to Design an Alpha-Helical Protein. *Proteins* **7**, 1-15, doi:DOI 10.1002/prot.340070102 (1990).

232 Eisenberg, D. *et al.* The design, synthesis, and crystallization of an alpha-helical peptide. *Proteins* **1**, 16-22, doi:10.1002/prot.340010105 (1986).

233 Osterhout, J. J. *et al.* Characterization of the Structural-Properties of Alpha-1b, a Peptide Designed to Form a 4-Helix Bundle. *J Am Chem Soc* **114**, 331-337, doi:DOI 10.1021/ja00027a043 (1992).

234 Ho, S. P. & Degrado, W. F. Design of a 4-Helix Bundle Protein - Synthesis of Peptides Which Self-Associate into a Helical Protein. *J Am Chem Soc* **109**, 6751-6758, doi:DOI 10.1021/ja00256a032 (1987).

235 Regan, L. & Degrado, W. F. Characterization of a Helical Protein Designed from 1st Principles. *Science* **241**, 976-978, doi:DOI 10.1126/science.3043666 (1988).

236 Raleigh, D. P. & Degrado, W. F. A De novo Designed Protein Shows a Thermally Induced Transition from a Native to a Molten Globule-Like State. *J Am Chem Soc* **114**, 10079-10081, doi:DOI 10.1021/ja00051a061 (1992).

237 Handel, T. & Degrado, W. F. De novo Design of a Zn2+-Binding Protein. *J Am Chem Soc* **112**, 6710-6711, doi:DOI 10.1021/ja00174a039 (1990).

238 Choma, C. T. *et al.* Design of a Heme-Binding 4-Helix Bundle. *J Am Chem Soc* **116**, 856-865, doi:DOI 10.1021/ja00082a005 (1994).

239 Kamtekar, S., Schiffer, J. M., Xiong, H. Y., Babik, J. M. & Hecht, M. H. Protein Design by Binary Patterning of Polar and Nonpolar Amino-Acids. *Science* **262**, 1680-1685, doi:DOI 10.1126/science.8259512 (1993).

240 Wei, Y. N. *et al.* Stably folded de novo proteins from a designed combinatorial library. *Protein Sci* **12**, 92-102, doi:10.1110/ps.0228003 (2003).

241 Bradley, L. H., Kleiner, R. E., Wang, A. F., Hecht, M. H. & Wood, D. W. An intein-based genetic selection allows the construction of a high-quality library of binary patterned de novo protein sequences. *Protein Eng Des Sel* **18**, 201-207, doi:10.1093/protein/gzi020 (2005).

242 Roy, S. & Hecht, M. H. Cooperative thermal denaturation of proteins designed by binary patterning of polar and nonpolar amino acids. *Biochemistry-Us* **39**, 4603-4607, doi:DOI 10.1021/bi992328e (2000).

243 Go, A., Kim, S., Baum, J. & Hecht, M. H. Structure and dynamics of de novo proteins from a designed superfamily of 4-helix bundles. *Protein Sci* **17**, 821-832, doi:10.1110/ps.073377908 (2008).

244 Wei, Y. N., Kim, S., Fela, D., Baum, J. & Hecht, M. H. Solution structure of a de novo protein from a designed combinatorial library. *P Natl Acad Sci USA* **100**, 13270-13273, doi:10.1073/pnas.1835644100 (2003).

245 Davidson, A. R., Lumb, K. J. & Sauer, R. T. Cooperatively Folded Proteins in Random Sequence Libraries. *Nat Struct Biol* **2**, 856-864, doi:DOI 10.1038/nsb1095-856 (1995).

246 Davidson, A. R. & Sauer, R. T. Folded Proteins Occur Frequently in Libraries of Random Amino-Acid-Sequences. *P Natl Acad Sci USA* **91**, 2146-2150, doi:DOI 10.1073/pnas.91.6.2146 (1994).

247 Rojas, N. R. L. *et al.* De novo heme proteins from designed combinatorial libraries. *Protein Sci* **6**, 2512-2524 (1997).

248 Patel, S. C., Bradley, L. H., Jinadasa, S. P. & Hecht, M. H. Cofactor binding and enzymatic activity in an unevolved superfamily of de novo designed 4-helix bundle proteins. *Protein Sci* **18**, 1388-1400, doi:10.1002/pro.147 (2009).

249 Wei, Y. N. & Hecht, M. H. Enzyme-like proteins from an unselected library of designed amino acid sequences. *Protein Eng Des Sel* **17**, 67-75, doi:10.1093/protein/gzh007 (2004).

250 Burton, A. J., Thomson, A. R., Dawson, W. M., Brady, R. L. & Woolfson, D. N. Installing hydrolytic activity into a completely de novo protein framework. *Nat Chem* **8**, 837-844, doi:10.1038/Nchem.2555 (2016).

251 Koder, R. L. *et al.* Design and engineering of an O-2 transport protein. *Nature* **458**, 305-U364, doi:10.1038/nature07841 (2009).

252 Wang, W., Knovich, M. A., Coffman, L. G., Torti, F. M. & Torti, S. V. Serum ferritin: Past, present and future. *Bba-Gen Subjects* **1800**, 760-769, doi:10.1016/j.bbagen.2010.03.011 (2010).

253 Holt, C. Unfolded phosphopolypeptides enable soft and hard tissues to coexist in the same organism with relative ease. *Curr Opin Struc Biol* **23**, 420-425, doi:10.1016/j.sbi.2013.02.010 (2013).

254 Joh, N. H. *et al.* De novo design of a transmembrane Zn2+-transporting four-helix bundle. *Science* **346**, 1520-1524, doi:10.1126/science.1261172 (2014).

255 Daber, R., Stayrook, S., Rosenberg, A. & Lewis, M. Structural analysis of lac repressor bound to allosteric effectors. *J Mol Biol* **370**, 609-619, doi:10.1016/j.jmb.2007.04.028 (2007).

256 Lin, Q., Barbas, C. F. & Schultz, P. G. Small-molecule switches for zinc finger transcription factors. *J Am Chem Soc* **125**, 612-613, doi:10.1021/ja0248408e (2003).

257 Efimov, V. P., Lustig, A. & Engel, J. The Thrombospondin-Like Chains of Cartilage Oligomeric Matrix Protein Are Assembled by a 5-Stranded Alpha-Helical Bundle between Residue-20 and Residue-83. *Febs Lett* **341**, 54-58, doi:Doi 10.1016/0014-5793(94)80239-4 (1994).

258 Malashkevich, V. N., Kammerer, R. A., Efimov, V. P., Schulthess, T. & Engel, J. The crystal structure of a five-stranded coiled coil in COMP: A prototype ion channel? *Science* **274**, 761-765, doi:DOI 10.1126/science.274.5288.761 (1996).

259 Guo, Y. *et al.* All-trans retinol, vitamin D and other hydrophobic compounds bind in the axial pore of the five-stranded coiled-coil domain of cartilage

oligomeric matrix protein. *Embo J* **17**, 5265-5272, doi:DOI 10.1093/emboj/17.18.5265 (1998).

260 MacFarlane, A. A. *et al.* The Pentameric Channel of COMPcc in Complex with Different Fatty Acids. *Plos One* **7**, doi:10.1371/journal.pone.0048130 (2012).

261 Liu, J., Yong, W., Deng, Y. Q., Kallenbach, N. R. & Lu, M. Atomic structure of a tryptophan-zipper pentamer. *P Natl Acad Sci USA* **101**, 16156-16161, doi:10.1073/pnas.0405319101 (2004).

262 Xu, C. F. *et al.* Rational Design of Helical Nanotubes from Self-Assembly of Coiled-Coil Lock Washers. *J Am Chem Soc* **135**, 15565-15578, doi:10.1021/ja4074529 (2013).

263 Burgess, N. C. *et al.* Modular Design of Self-Assembling Peptide-Based Nanotubes. *J Am Chem Soc* **137**, 10554-10562, doi:10.1021/jacs.5b03973 (2015).

264 Thomas, F. *et al.* De Novo-Designed alpha-Helical Barrels as Receptors for Small Molecules. *Acs Synth Biol* **7**, 1808-1816, doi:10.1021/acssynbio.8b00225 (2018).

265 Stetefeld, J. *et al.* Crystal structure of a naturally occurring parallel right-handed coiled coil tetramer. *Nat Struct Biol* **7**, 772-776, doi:Doi 10.1038/79006 (2000).

266 Eriksson, M. *et al.* Utilization of a Right-handed Coiled-coil Protein from Archaebacterium Staphylothermus marinus as a Carrier for Cisplatin. *Anticancer Res* **29**, 11-18 (2009).

267 Holsinger, L. J. & Lamb, R. A. Influenza Virus-M2 Integral Membrane-Protein Is a Homotetramer Stabilized by Formation of Disulfide Bonds. *Virology* **183**, 32-43, doi:Doi 10.1016/0042-6822(91)90115-R (1991).

268 Sugrue, R. J. & Hay, A. J. Structural Characteristics of the M2 Protein of Influenza-a Viruses - Evidence That It Forms a Tetrameric Channel. *Virology* **180**, 617-624, doi:Doi 10.1016/0042-6822(91)90075-M (1991).

269 Manzoor, R., Igarashi, M. & Takada, A. Influenza A Virus M2 Protein: Roles from Ingress to Egress. *Int J Mol Sci* **18**, doi:10.3390/ijms18122649 (2017).

270 Jing, X. H. *et al.* Functional studies indicate amantadine binds to the pore of the influenza A virus M2 proton-selective ion channel. *P Natl Acad Sci USA* **105**, 10967-10972, doi:10.1073/pnas.0804958105 (2008).

271 Stouffer, A. L. *et al.* Structural basis for the function and inhibition of an influenza virus proton channel. *Nature* **451**, 596-U513, doi:10.1038/nature06528 (2008).

272 Cady, S. D. *et al.* Structure of the amantadine binding site of influenza M2 proton channels in lipid bilayers. *Nature* **463**, 689-U127, doi:10.1038/nature08722 (2010).

273 Balgi, A. D. *et al.* Inhibitors of the Influenza A Virus M2 Proton Channel Discovered Using a High-Throughput Yeast Growth Restoration Assay. *Plos One* **8**, doi:10.1371/journal.pone.0055271 (2013).

274 Oshea, E. K., Rutkowski, R. & Kim, P. S. Evidence That the Leucine Zipper Is a Coiled Coil. *Science* **243**, 538-542, doi:DOI 10.1126/science.2911757 (1989).

275 Gonzalez, L., Plecs, J. J. & Alber, T. An engineered allosteric switch in leucine-zipper oligomerization. *Nat Struct Biol* **3**, 510-515, doi:DOI 10.1038/nsb0696-510 (1996).

276 Yadav, M. K. *et al.* Structure-based engineering of internal cavities in coiled-coil peptides. *Biochemistry-Us* **44**, 9723-9732, doi:10.1021/bi050742a (2005).

277 Mizuno, T. *et al.* Organic Ligand Binding by a Hydrophobic Cavity in a Designed Tetrameric Coiled-Coil Protein. *Chem-Eur J* **15**, 1491-1498, doi:10.1002/chem.200800855 (2009).

278 Obataya, I., Sakamoto, S., Ueno, A. & Mihara, H. Design and synthesis of 3 alpha-helix peptides forming a cavity for a fluorescent ligand. *Biopolymers* **59**, 65-71, doi:Doi 10.1002/1097-0282(200108)59:2<65::Aid-Bip1006>3.0.Co;2-V (2001).

279 Johansson, J. S., Scharf, D., Davies, L. A., Reddy, K. S. & Eckenhoff, R. G. A designed four-alpha-helix bundle that binds the volatile general anesthetic halothane with high affinity. *Biophys J* **78**, 982-993, doi:Doi 10.1016/S0006-3495(00)76656-2 (2000).

280 Johansson, J. S., Rabanal, F. & Dutton, P. L. Binding of the volatile anesthetic halothane to the hydrophobic core of a tetra-alpha-helix-bundle protein. *J Pharmacol Exp Ther* **279**, 56-61 (1996).

281 Robertson, D. E. *et al.* Design and Synthesis of Multi-Heme Proteins. *Nature* **368**, 425-431, doi:DOI 10.1038/368425a0 (1994).

282 Ghirlanda, G. *et al.* De novo design of a D-2-symmetrical protein that reproduces the diheme four-helix bundle in cytochrome bc(1). *J Am Chem Soc* **126**, 8141-8147, doi:10.1021/ja039935g (2004).

283 Cochran, F. V. *et al.* Computational de novo design and characterization of a four-helix bundle protein that selectively binds a nonbiological cofactor. *J Am Chem Soc* **127**, 1346-1347, doi:10.1021/ja044129a (2005).

284 Polizzi, N. F. *et al.* De novo design of a hyperstable non-natural protein-ligand complex with sub-angstrom accuracy. *Nat Chem* **9**, 1157-1164, doi:10.1038/Nchem.2846 (2017).

285 Watkins, D. W. *et al.* Construction and in vivo assembly of a catalytically proficient and hyperthermostable de novo enzyme. *Nat Commun* **8**, doi:10.1038/s41467-017-00541-4 (2017).

286 Kodali, G. *et al.* Design and engineering of water-soluble light-harvesting protein maquettes. *Chem Sci* **8**, 316-324, doi:10.1039/c6sc02417c (2017).

287 Peacock, A. F. A. Recent advances in designed coiled coils and helical bundles with inorganic prosthetic groups - from structural to functional applications. *Curr Opin Chem Biol* **31**, 160-165, doi:10.1016/j.cbpa.2016.03:009 (2016).

288 Suzuki, K., Hiroaki, H., Kohda, D., Nakamura, H. & Tanaka, T. Metal ion induced self-assembly of a designed peptide into a triple-stranded alpha-helical bundle: A novel metal binding site in the hydrophobic core. *J Am Chem Soc* **120**, 13008-13015, doi:DOI 10.1021/ja982768d (1998).

289 Cerasoli, E., Sharpe, B. K. & Woolfson, D. N. ZiCo: A peptide designed to switch folded state upon binding zinc. *J Am Chem Soc* **127**, 15008-15009, doi:10.1021/ja0543604 (2005).

290 Ambroggio, X. I. & Kuhlman, B. Computational design of a single amino acid sequence that can switch between two distinct protein folds. *J Am Chem Soc* **128**, 1154-1161, doi:10.1021/ja054718w (2006).

291 Oheix, E. & Peacock, A. F. A. Metal-Ion-Regulated Miniature DNA-Binding Proteins Based on GCN4 and Non-native Regulation Sites. *Chem-Eur J* **20**, 2829-2839, doi:10.1002/chem.201303747 (2014).

292 Fong, J. H., Keating, A. E. & Singh, M. Predicting specificity in bZIP coiled-coil protein interactions. *Genome Biol* **5**, doi:DOI 10.1186/gb-2004-5-2-r11 (2004).

293 Wood, C. W. *et al.* ISAMBARD: an open-source computational environment for biomolecular analysis, modelling and design. *Bioinformatics* **33**, 3043-3050, doi:10.1093/bioinformatics/btx352 (2017).

294 McIntosh-Smith, S., Wilson, T., Ibarra, A. A., Crisp, J. & Sessions, R. B. Benchmarking Energy Efficiency, Power Costs and Carbon Emissions on Heterogeneous Systems. *Comput J* **55**, 192-205, doi:10.1093/comjnl/bxr091 (2012).

295 Myers, J. K., Pace, C. N. & Scholtz, J. M. A direct comparison of helix propensity in proteins and peptides. *P Natl Acad Sci USA* **94**, 2833-2837, doi:DOI 10.1073/pnas.94.7.2833 (1997).

296 Brown, P. H. & Schuck, P. Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation. *Biophys J* **90**, 4651-4661, doi:DOI 10.1529/biophysj.106.081372 (2006).

297 Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D* **67**, 271-281, doi:10.1107/S0907444910048675 (2011).

298 Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr D* **69**, 1204-1214, doi:10.1107/S0907444913000061 (2013).

299 Mccoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674, doi:10.1107/S0021889807021206 (2007).

300 Kantardjieff, K. A. & Rupp, B. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci* **12**, 1865-1871, doi:10.1110/ps.0350503 (2003).

301 Wood, C. W. & Woolfson, D. N. CCBuilder 2.0: Powerful and accessible coiled-coil modeling. *Protein Sci* **27**, 103-111, doi:10.1002/pro.3279 (2018).

302 Rodriguez, D. D. *et al.* Crystallographic ab initio protein structure solution below atomic resolution. *Nat Methods* **6**, 651-U639, doi:10.1038/Nmeth.1365 (2009).

303 Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr D* **67**, 235-242, doi:10.1107/S0907444910045749 (2011).

304 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).

305 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).

306 Davis, J. H., Rubin, A. J. & Sauer, R. T. Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res* **39**, 1131-1141, doi:10.1093/nar/gkq810 (2011).

307 Bernhardt, T. G. & de Boer, P. A. J. The Escherichia coli amidase AmiC is a periplasmic septal ring component exported via the twin-arginine transport pathway. *Mol Microbiol* **48**, 1171-1182, doi:10.1046/j.1365-2958.2003.03511.x (2003).

308 Pedelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol* **24**, 79-88, doi:10.1038/nbt1172 (2006).

309 Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671-675, doi:10.1038/nmeth.2089 (2012).

310 Chao, H. M., Bautista, D. L., Litowski, J., Irvin, R. T. & Hodges, R. S. Use of a heterodimeric coiled-coil system for biosensor application and affinity purification. *J Chromatogr B* **715**, 307-329, doi:Doi 10.1016/S0378-4347(98)00172-8 (1998).

311 Yano, Y. *et al.* Coiled-coil tag - Probe system for quick labeling of membrane receptors in living cells. *Acs Chem Biol* **3**, 341-345, doi:10.1021/cb8000556 (2008).

312 Wang, X. W., Zhong, P. Y., Luo, P. P. & Wang, K. C. Antibody Engineering Using Phage Display with a Coiled-Coil Heterodimeric Fv Antibody Fragment. *Plos One* **6**, doi:10.1371/journal.pone.0019023 (2011).

313 Reinhardt, U. *et al.* Peptide-Templated Acyl Transfer: A Chemical Method for the Labeling of Membrane Proteins on Live Cells. *Angew Chem Int Edit* **53**, 10237-10241, doi:10.1002/anie.201403214 (2014).

314 Friedman, P. N., Chen, X. B., Bargonetti, J. & Prives, C. The P53 Protein Is an Unusually Shaped Tetramer That Binds Directly to DNA. *P Natl Acad Sci USA* **90**, 3319-3323, doi:DOI 10.1073/pnas.90.8.3319 (1993).

315 Friedman, A. M., Fischmann, T. O. & Steitz, T. A. Crystal-Structure of Lac Repressor Core Tetramer and Its Implications for DNA Looping. *Science* **268**, 1721-1727, doi:DOI 10.1126/science.7792597 (1995).

316 Munson, M., Obrien, R., Sturtevant, J. M. & Regan, L. Redesigning the Hydrophobic Core of a 4-Helix-Bundle Protein. *Protein Sci* **3**, 2015-2022, doi:DOI 10.1002/pro.5560031114 (1994).

317 Hill, R. B., Raleigh, D. P., Lombardi, A. & Degrado, W. F. De novo design of helical bundles as models for understanding protein folding and function. *Accounts Chem Res* **33**, 745-754, doi:10.1021/ar970004h (2000).

318 Woolfson, D. N. Core-directed protein design. *Curr Opin Struc Biol* **11**, 464-471, doi:Doi 10.1016/S0959-440x(00)00234-7 (2001).

319 Betz, S. F., Liebman, P. A. & DeGrado, W. F. De novo design of native proteins: Characterization of proteins intended to fold into antiparallel, rop-like, four-helix bundles. *Biochemistry-Us* **36**, 2450-2458, doi:DOI 10.1021/bi961704h (1997).

320 Boyken, S. E. *et al.* De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science* **352**, 680-687, doi:10.1126/science.aad8865 (2016).

321 Bromley, E. H. C., Sessions, R. B., Thomson, A. R. & Woolfson, D. N. Designed alpha-Helical Tectons for Constructing Multicomponent Synthetic Biological Systems. *J Am Chem Soc* **131**, 928-+, doi:10.1021/ja804231a (2009).

322 Small, L. S. R. *et al.* Construction of a Chassis for a Tripartite Protein-Based Molecular Motor. *Acs Synth Biol* **6**, 1096-1102, doi:10.1021/acssynbio.7b00037 (2017).

323 Ljubetic, A. *et al.* Design of coiled-coil protein-origami cages that self-assemble in vitro and in vivo. *Nat Biotechnol* **35**, 1094-+, doi:10.1038/nbt.3994 (2017).

324 Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc* **1**, 2876-2890, doi:10.1038/nprot.2006.202 (2006).

325 Greenfield, N. J. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat Protoc* **1**, 2527-2535, doi:10.1038/nprot.2006.204 (2006).

326 Greenfield, N. J. Determination of the folding of proteins as a function of denaturants, osmolytes or ligands using circular dichroism. *Nat Protoc* **1**, 2733-2741, doi:10.1038/nprot.2006.229 (2006).

327 Holzwarth, G. & Doty, P. Ultraviolet Circular Dichroism of Polypeptides. *J Am Chem Soc* **87**, 218-+, doi:DOI 10.1021/ja01080a015 (1965).

328 Cole, J. L., Lary, J. W., Moody, T. P. & Laue, T. M. Analytical ultracentrifugation: Sedimentation velocity and sedimentation equilibrium. *Biophysical Tools for Biologists: Vol 1 in Vitro Techniques* **84**, 143-179, doi:10.1016/S0091-679x(07)84006-4 (2008).

329 Gonzalez, L., Woolfson, D. N. & Alber, T. Buried polar residues and structural specificity in the GCN4 leucine zipper. *Nat Struct Biol* **3**, 1011-1018, doi:DOI 10.1038/nsb1296-1011 (1996).

330 Akey, D. L., Malashkevich, V. N. & Kim, P. S. Buried polar residues in coiled-coil interfaces. *Biochemistry-Us* **40**, 6352-6360, doi:10.1021/bi002829w (2001).

331 Fletcher, J. M. *et al.* N@a and N@d: Oligomer and Partner Specification by Asparagine in Coiled-Coil Interfaces. *Acs Chem Biol* **12**, 528-538, doi:10.1021/acschembio.6b00935 (2017).

332 Oshea, E. K., Klemm, J. D., Kim, P. S. & Alber, T. X-Ray Structure of the Gcn4 Leucine Zipper, a 2-Stranded, Parallel Coiled Coil. *Science* **254**, 539-544, doi:DOI 10.1126/science.1948029 (1991).

333 Zhu, H., Celinski, S. A., Scholtz, J. M. & Hu, J. C. The contribution of buried polar groups to the conformational stability of the GCN4 coiled coil. *J Mol Biol* **300**, 1377-1387, doi:10.1006/jmbi.2000.3936 (2000).

334 Thomas, F., Niitsu, A., Oregioni, A., Bartlett, G. J. & Woolfson, D. N. Conformational Dynamics of Asparagine at Coiled-Coil Interfaces. *Biochemistry-Us* **56**, 6544-6554, doi:10.1021/acs.biochem.7b00848 (2017).

335 Schreiber, G. & Fersht, A. R. Rapid, electrostatically assisted association of proteins. *Nat Struct Biol* **3**, 427-431, doi:DOI 10.1038/nsb0596-427 (1996).

336 Selzer, T., Albeck, S. & Schreiber, G. Rational design of faster associating and tighter binding protein complexes. *Nat Struct Biol* **7**, 537-541 (2000).

337 Deng, Y. Q. *et al.* A Heterospecific Leucine Zipper Tetramer. *Chem Biol* **15**, 908-919, doi:10.1016/j.chembiol.2008.07.008 (2008).

338 Lau, S. Y. M., Taneja, A. K. & Hodges, R. S. Synthesis of a Model Protein of Defined Secondary and Quaternary Structure - Effect of Chain-Length on the Stabilization and Formation of 2-Stranded Alpha-Helical Coiled-Coils. *J Biol Chem* **259**, 3253-3261 (1984).

339 Zhu, B. Y., Zhou, N. E., Kay, C. M. & Hodges, R. S. Packing and Hydrophobicity Effects on Protein Folding and Stability - Effects of Beta-Branched Amino-Acids, Valine and Isoleucine, on the Formation and

Stability of 2-Stranded Alpha-Helical Coiled Coils Leucine Zippers. *Protein Sci* **2**, 383-394 (1993).

340 Litowski, J. R. & Hodges, R. S. Designing heterodimeric two-stranded alpha-helical coiled-coils - Effects of hydrophobicity and alpha-helical propensity on protein folding, stability, and specificity. *J Biol Chem* **277**, 37272-37279, doi:10.1074/jbc.M204257200 (2002).

341 Tsai, C. J., Maizel, J. V. & Nussinov, R. The hydrophobic effect: A new insight from cold denaturation and a two-state water structure. *Crit Rev Biochem Mol* **37**, 55-69, doi:Doi 10.1080/10409230290771456 (2002).

342 Dias, C. L. *et al.* The hydrophobic effect and its role in cold denaturation. *Cryobiology* **60**, 91-99, doi:10.1016/j.cryobiol.2009.07.005 (2010).

343 Testa, O. D., Moutevelis, E. & Woolfson, D. N. CC plus : a relational database of coiled-coil structures. *Nucleic Acids Res* **37**, D315-D322, doi:10.1093/nar/gkn675 (2009).

344 Walshaw, J., Shipway, J. M. & Woolfson, D. N. Guidelines for the assembly of novel coiled-coil structures: alpha-sheets and alpha-cylinders. *Biochem Soc Symp*, 111-123 (2001).

345 Liu, J. *et al.* A seven-helix coiled coil. *P Natl Acad Sci USA* **103**, 15457-15462, doi:10.1073/pnas.0604871103 (2006).

346 Zeng, X. G., Zhu, H., Lashuel, H. A. & Hu, J. C. Oligomerization properties of GCN4 leucine zipper e and g position mutants. *Protein Sci* **6**, 2218-2226 (1997).

347 Bairoch, A. & Boeckmann, B. The Swiss-Prot Protein-Sequence Data-Bank. *Nucleic Acids Res* **19**, 2247-2248, doi:DOI 10.1093/nar/19.suppl.2247 (1991).

348 Gernert, K. M., Surles, M. C., Labean, T. H., Richardson, J. S. & Richardson, D. C. The Alacoil - a Very Tight, Antiparallel Coiled-Coil of Helices. *Protein Sci* **4**, 2252-2260, doi:DOI 10.1002/pro.5560041102 (1995).

349 Banner, D. W., Kokkinidis, M. & Tsernoglou, D. Structure of the Co1e1 Rop Protein at 1.7 a Resolution. *J Mol Biol* **196**, 657-675, doi:Doi 10.1016/0022-2836(87)90039-8 (1987).

350 Betz, S. F., Raleigh, D. P. & Degrado, W. F. De-Novo Protein Design - from Molten Globules to Native-Like States. *Curr Opin Struc Biol* **3**, 601-610, doi:Doi 10.1016/0959-440x(93)90090-8 (1993).

351 Hawe, A., Sutter, M. & Jiskoot, W. Extrinsic fluorescent dyes as tools for protein characterization. *Pharm Res-Dordr* **25**, 1487-1499, doi:10.1007/s11095-007-9516-9 (2008).

352 Zhou, N. E., Kay, C. M. & Hodges, R. S. The Role of Interhelical Ionic Interactions in Controlling Protein-Folding and Stability - De-Novo Designed Synthetic 2-Stranded Alpha-Helical Coiled-Coils. *J Mol Biol* **237**, 500-512, doi:DOI 10.1006/jmbi.1994.1250 (1994).

353 Lewis, M. The lac repressor. *Cr Biol* **328**, 521-548, doi:10.1016/j.crvi.2005.04.004 (2005).

354 Reznikoff, W. S., Winter, R. B. & Hurley, C. K. Location of Repressor Binding-Sites in Lac Operon. *P Natl Acad Sci USA* **71**, 2314-2318, doi:DOI 10.1073/pnas.71.6.2314 (1974).

355 Eismann, E., Vonwilckenbergmann, B. & Mullerhill, B. Specific Destruction of the 2nd Lac Operator Decreases Repression of the Lac Operon in Escherichia-Coli Fivefold. *J Mol Biol* **195**, 949-952, doi:Doi 10.1016/0022-2836(87)90499-2 (1987).

356 Matthews, K. S. DNA Looping. *Microbiol Rev* **56**, 123-136 (1992).

357 Dong, F. M. *et al.* Dimerisation mutants of Lac repressor. I. A monomeric mutant, L251A, that binds lac operator DNA as a dimer. *J Mol Biol* **290**, 653-666, doi:DOI 10.1006/jmbi.1999.2902 (1999).

358 Johnsrud, L. Contacts between Escherichia-Coli Rna-Polymerase and a Lac Operon Promoter. *P Natl Acad Sci USA* **75**, 5314-5318, doi:DOI 10.1073/pnas.75.11.5314 (1978).

359 Wanner, B. L., Kodaira, R. & Neidhardt, F. C. Physiological Regulation of a Decontrolled Lac Operon. *J Bacteriol* **130**, 212-222 (1977).

360 Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight Regulation, Modulation, and High-Level Expression by Vectors Containing the Arabinose P-Bad Promoter. *J Bacteriol* **177**, 4121-4130, doi:DOI 10.1128/jb.177.14.4121-4130.1995 (1995).

361 Schleif, R. AraC protein: a love-hate relationship. *Bioessays* **25**, 274-282, doi:10.1002/bies.10237 (2003).

362 Lobell, R. B. & Schleif, R. F. DNA Looping and Unlooping by Arac Protein. *Science* **250**, 528-532, doi:DOI 10.1126/science.2237403 (1990).

363 Schleif, R. AraC protein, regulation of the l-arabinose operon in Escherichia coli, and the light switch mechanism of AraC action. *Fems Microbiol Rev* **34**, 779-796, doi:10.1111/j.1574-6976.2010.00226.x (2010).

364 Johnson, C. M. & Schleif, R. F. Cooperative action of the catabolite activator protein and AraC in vitro at the araFGH promoter. *J Bacteriol* **182**, 1995-2000, doi:Doi 10.1128/Jb.182.7.1995-2000.2000 (2000).

365 Maloy, S., Stewart, V. & Taylor, R. *Genetic analysis of pathogenic bacteria: a laboratory manual*. (Cold Spring Harbor Laboratory Press, 1996).

366 Haldimann, A., Daniels, L. L. & Wanner, B. L. Use of new methods for construction of tightly regulated arabinose and rhamnose promoter fusions in studies of the Escherichia coli phosphate regulon. *J Bacteriol* **180**, 1277-1286 (1998).

367 Alberti, S., Oehler, S., Vonwilckenbergmann, B. & Mullerhill, B. Genetic-Analysis of the Leucine Heptad Repeats of Lac Repressor - Evidence for a 4-Helical Bundle. *Embo J* **12**, 3227-3236 (1993).

368 Geiger-Schuller, K. *et al.* Extreme stability in de novo-designed repeat arrays is determined by unusually stable short-range interactions. *P Natl Acad Sci USA* **115**, 7539-7544, doi:10.1073/pnas.1800283115 (2018).

369 Griffith, J., Hochschild, A. & Ptashne, M. DNA Loops Induced by Cooperative Binding of Lambda-Repressor. *Nature* **322**, 750-752, doi:DOI 10.1038/322750a0 (1986).

370 Daly, T. J. & Matthews, K. S. Characterization and Modification of a Monomeric Mutant of the Lactose Repressor Protein. *Biochemistry-Us* **25**, 5474-5478, doi:DOI 10.1021/bi00367a019 (1986).

371 Chakerian, A. E. & Matthews, K. S. Characterization of Mutations in Oligomerization Domain of Lac Repressor Protein. *J Biol Chem* **266**, 22206-22214 (1991).

372 Nichols, J. C. & Matthews, K. S. Combinatorial mutations of lac repressor - Stability of monomer-monomer interface is increased by apolar substitution at position 84. *J Biol Chem* **272**, 18550-18557, doi:DOI 10.1074/jbc.272.30.18550 (1997).

373 Bell, C. E. & Lewis, M. A closer view of the conformation of the Lac repressor bound to operator. *Nat Struct Biol* **7**, 209-214 (2000).

374 Marblestone, J. G. *et al.* Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO. *Protein Sci* **15**, 182-189, doi:10.1110/ps.051812706 (2006).

375 Wood, C. *Unpublished undergraduate dissertation* (University of Bristol, 2018).

376 Parsell, D. A. & Sauer, R. T. Induction of a Heat Shock-Like Response by Unfolded Protein in Escherichia Coli: Dependence on Protein Level Not Protein-Degradation. *Gene Dev* **3**, 1226-1232, doi:DOI 10.1101/gad.3.8.1226 (1989).

377 Parsell, D. A. & Sauer, R. T. The Structural Stability of a Protein Is an Important Determinant of Its Proteolytic Susceptibility in Escherichia-Coli. *J Biol Chem* **264**, 7590-7595 (1989).

378 Baneyx, F. & Mujacic, M. Recombinant protein folding and misfolding in Escherichia coli. *Nat Biotechnol* **22**, 1399-1408, doi:10.1038/nbt1029 (2004).

379 Truscott, K. N., Bezawork-Geleta, A. & Dougan, D. A. Unfolded Protein Responses in Bacteria and Mitochondria: A Central Role for the ClpXP Machine. *Iubmb Life* **63**, 955-963, doi:10.1002/iub.526 (2011).

380 Tompa, P., Dosztanyi, Z. & Simon, I. Prevalent structural disorder in E. coli and S. cerevisiae proteomes. *J Proteome Res* **5**, 1996-2000, doi:10.1021/pr0600881 (2006).

381 Xue, B., Dunker, A. K. & Uversky, V. N. Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three

domains of life. *J Biomol Struct Dyn* **30**, 137-149, doi:10.1080/07391102.2012.675145 (2012).

382 Gsponer, J., Futschik, M. E., Teichmann, S. A. & Babu, M. M. Tight Regulation of Unstructured Proteins: From Transcript Synthesis to Protein Degradation. *Science* **322**, 1365-1368, doi:10.1126/science.1163581 (2008).

383 Vavouri, T., Semple, J. I., Garcia-Verdugo, R. & Lehner, B. Intrinsic Protein Disorder and Interaction Promiscuity Are Widely Associated with Dosage Sensitivity. *Cell* **138**, 198-208, doi:10.1016/j.cell.2009.04.029 (2009).

384 Zilberstein, D., Agmon, V., Schuldiner, S. & Padan, E. Escherichia-Coli Intracellular Ph, Membrane-Potential, and Cell-Growth. *J Bacteriol* **158**, 246-252 (1984).

385 Wilks, J. C. & Slonczewski, J. L. pH of the cytoplasm and periplasm of Escherichia coli: Rapid measurement by green fluorescent protein fluorimetry. *J Bacteriol* **189**, 5601-5607, doi:10.1128/Jb.00615-07 (2007).

386 Schavemaker, P. E., Smigiel, W. M. & Poolman, B. Ribosome surface properties may impose limits on the nature of the cytoplasmic proteome. *Elife* **6**, doi:10.7554/elife.30084 (2017).

387 Nautiyal, S., Woolfson, D. N., King, D. S. & Alber, T. A Designed Heterotrimeric Coiled-Coil. *Biochemistry-Us* **34**, 11645-11651, doi:DOI 10.1021/bi00037a001 (1995).

388 Nautiyal, S. & Alber, T. Crystal structure of a designed, thermostable; heterotrimeric coiled coil. *Protein Sci* **8**, 84-90 (1999).

389 Hinrichs, W. *et al.* Structure of the Tet Repressor Tetracycline Complex and Regulation of Antibiotic-Resistance. *Science* **264**, 418-420, doi:DOI 10.1126/science.8153629 (1994).

390 Mahr, R. & Frunzke, J. Transcription factor-based biosensors in biotechnology: current state and future prospects. *Appl Microbiol Biot* **100**, 79-90, doi:10.1007/s00253-015-7090-3 (2016).

391 Peet, D. J., Doyle, D. F., Corey, D. R. & Mangelsdorf, D. J. Engineering novel specificities for ligand-activated transcription in the nuclear hormone receptor RXR. *Chem Biol* **5**, 13-21, doi:Doi 10.1016/S1074-5521(98)90083-7 (1998).

392 Beerli, R. R., Schopfer, U., Dreier, B. & Barbas, C. F. Chemically regulated zinc finger transcription factors. *J Biol Chem* **275**, 32617-32627, doi:DOI 10.1074/jbc.M005108200 (2000).

393 Schnell, J. R. & Chou, J. J. Structure and mechanism of the M2 proton channel of influenza A virus. *Nature* **451**, 591-U512, doi:10.1038/nature06531 (2008).

394 Pielak, R. M., Schnell, J. R. & Chou, J. J. Mechanism of drug inhibition and drug resistance of influenza A M2 channel. *P Natl Acad Sci USA* **106**, 7379-7384, doi:10.1073/pnas.0902548106 (2009).

395 Cady, S. D., Mishanina, T. V. & Hong, M. Structure of Amantadine-Bound M2 Transmembrane Peptide of Influenza A in Lipid Bilayers from Magic-Angle-Spinning Solid-State NMR: The Role of Ser31 in Amantadine Binding. *J Mol Biol* **385**, 1127-1141, doi:10.1016/j.jmb.2008.11.022 (2009).

396 Kochendoerfer, G. G. *et al.* Total chemical synthesis of the integral membrane protein influenza A virus M2: Role of its C-terminal domain in tetramer assembly. *Biochemistry-Us* **38**, 11905-11913, doi:DOI 10.1021/bi990720m (1999).

397 Salom, D., Hill, B. R., Lear, J. D. & DeGrado, W. F. pH-dependent tetramerization and amantadine binding of the transmembrane helix of M2 from the influenza A virus. *Biochemistry-Us* **39**, 14160-14170, doi:10.1021/bi001799u (2000).

398 Tian, W., Chen, C., Lei, X., Zhao, J. L. & Liang, J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res* **46**, W363-W367, doi:10.1093/nar/gky473 (2018).

399 Williams, D. M., Wang, D. X. & Cole, P. A. Chemical rescue of a mutant protein-tyrosine kinase. *J Biol Chem* **275**, 38127-38130, doi:DOI 10.1074/jbc.C000606200 (2000).

400 Qiao, Y. F., Molina, H., Pandey, A., Zhang, J. & Cole, P. A. Chemical rescue of a mutant enzyme in living cells. *Science* **311**, 1293-1297, doi:10.1126/science.1122224 (2006).

401 Goedl, C. & Nidetzky, B. The phosphate site of trehalose phosphorylase from Schizophyllum commune probed by site-directed mutagenesis and chemical rescue studies. *Febs J* **275**, 903-913, doi:10.1111/j.1742-4658.2008.06254.x (2008).

402 Deckert, K., Budiardjo, S. J., Brunner, L. C., Lovell, S. & Karanicolas, J. Designing Allosteric Control into Enzymes by Chemical Rescue of Structure. *J Am Chem Soc* **134**, 10055-10060, doi:10.1021/ja301409g (2012).

403 Xia, Y. *et al.* The Designability of Protein Switches by Chemical Rescue of Structure: Mechanisms of Inactivation and Reactivation. *J Am Chem Soc* **135**, 18840-18849, doi:10.1021/ja407644b (2013).

404 Guo, Z. H., Zhou, D. M. & Schultz, P. G. Designing small-molecule switches for protein-protein interactions. *Science* **288**, 2042-+, doi:DOI 10.1126/science.288.5473.2042 (2000).

405 Hassan, A. Q. & Koh, J. T. Selective chemical rescue of a thyroid-hormone-receptor mutant, TR beta(H435Y), identified in pituitary carcinoma and resistance to thyroid hormone. *Angew Chem Int Edit* **47**, 7280-7283, doi:10.1002/anie.200801742 (2008).

406 London, E. & Feigenson, G. W. Convenient and Sensitive Fluorescence Assay for Phospholipid Vesicles Using Diphenylhexatriene. *Anal Biochem* **88**, 203-211, doi:Doi 10.1016/0003-2697(78)90412-8 (1978).

407 Fulford, A. J. C. & Peel, W. E. Lateral Pressures in Biomembranes Estimated from the Dynamics of Fluorescent-Probes. *Biochim Biophys Acta* **598**, 237-246, doi:Doi 10.1016/0005-2736(80)90002-4 (1980).

408 Lentz, B. R. Membrane Fluidity as Detected by Diphenylhexatriene Probes. *Chem Phys Lipids* **50**, 171-190, doi:Doi 10.1016/0009-3084(89)90049-2 (1989).

409 Dyson, H. J. & Wright, P. E. Coupling of folding and binding for unstructured proteins. *Curr Opin Struc Biol* **12**, 54-60, doi:Doi 10.1016/S0959-440x(02)00289-0 (2002).

410 Mollica, L. *et al.* Binding Mechanisms of Intrinsically Disordered Proteins: Theory, Simulation, and Experiment. *Frontiers in Molecular Bioscience* **3**, doi:10.3389/fmolb.2016.00052 (2016).

411 Tucker, C. L. & Fields, S. A yeast sensor of ligand binding. *Nat Biotechnol* **19**, 1042-1046, doi:DOI 10.1038/nbt1101-1042 (2001).

412 Iwamoto, M., Bjorklund, T., Lundberg, C., Kirik, D. & Wandless, T. J. A General Chemical Method to Regulate Protein Stability in the Mammalian Central Nervous System. *Chem Biol* **17**, 981-988, doi:10.1016/j.chembiol.2010.07.009 (2010).

413 Banaszynski, L. A., Chen, L. C., Maynard-Smith, L. A., Ooi, A. G. L. & Wandless, T. J. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* **126**, 995-1004, doi:10.1016/j.cell.2006.07.025 (2006).

414 Miyazaki, Y., Imoto, H., Chen, L. C. & Wandless, T. J. Destabilizing Domains Derived from the Human Estrogen Receptor. *J Am Chem Soc* **134**, 3942-3945, doi:10.1021/ja209933r (2012).

415 Navarro, R., Chen, L. C., Rakhit, R. & Wandless, T. J. A Novel Destabilizing Domain Based on a Small-Molecule Dependent Fluorophore. *Acs Chem Biol* **11**, 2101-2104, doi:10.1021/acschembio.6b00234 (2016).

416 Morell, M., Ventura, S. & Aviles, F. X. Protein complementation assays: Approaches for the in vivo analysis of protein interactions. *Febs Lett* **583**, 1684-1691, doi:10.1016/j.febslet.2009.03.002 (2009).

417 Ghosh, I., Hamilton, A. D. & Regan, L. Antiparallel leucine zipper-directed protein reassembly: Application to the green fluorescent protein. *J Am Chem Soc* **122**, 5658-5659, doi:DOI 10.1021/ja994421w (2000).

418 Hu, C. D., Chinenov, Y. & Kerppola, T. K. Visualization of interactions among bZip and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol Cell* **9**, 789-798, doi:Doi 10.1016/S1097-2765(02)00496-3 (2002).

419 Rossi, F., Charlton, C. A. & Blau, H. M. Monitoring protein-protein interactions in intact eukaryotic cells by beta-galactosidase complementation. *P Natl Acad Sci USA* **94**, 8405-8410, doi:DOI 10.1073/pnas.94.16.8405 (1997).

420 Karimova, G., Pidoux, J., Ullmann, A. & Ladant, D. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *P Natl Acad Sci USA* **95**, 5752-5756, doi:DOI 10.1073/pnas.95.10.5752 (1998).

421 Pelletier, J. N., Campbell-Valois, F. X. & Michnick, S. W. Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. *P Natl Acad Sci USA* **95**, 12141-12146, doi:DOI 10.1073/pnas.95.21.12141 (1998).

422 Galarneau, A., Primeau, M., Trudeau, L. E. & Michnick, S. W. beta-Lactamase protein fragment complementation assays as in vivo and in vitro sensors of protein-protein interactions. *Nat Biotechnol* **20**, 619-622, doi:10.1038/nbt0602-619 (2002).

423 Pu, J. Y., Zinkus-Boltz, J. & Dickinson, B. C. Evolution of a split RNA polymerase as a versatile biosensor platform. *Nat Chem Biol* **13**, 432-+ (2017).

424 Finucane, M. D., Tuna, M., Lees, J. H. & Woolfson, D. N. Core-directed protein design. I. An experimental method for selecting stable proteins from combinatorial libraries. *Biochemistry-Us* **38**, 11604-11612, doi:DOI 10.1021/bi990765n (1999).

425 Bai, Y. W. & Feng, H. Q. Selection of stably folded proteins by phage-display with proteolysis. *Eur J Biochem* **271**, 1609-1614, doi:10.1111/j.1432-1033.2004.04074.x (2004).

426 Finucane, M. D. & Woolfson, D. N. Core-directed protein design. II. Rescue of a multiply mutated and destabilized variant of ubiquitin. *Biochemistry-Us* **38**, 11613-11623, doi:DOI 10.1021/bi990766f (1999).

427 Rocklin, G. J. *et al.* Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science* **357**, 168-174, doi:10.1126/science.aan0693 (2017).

428 Tinberg, C. E. *et al.* Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* **501**, 212-+, doi:10.1038/nature12443 (2013).

429 Malisi, C. *et al.* Binding Pocket Optimization by Computational Protein Design. *Plos One* **7**, doi:10.1371/journal.pone.0052505 (2012).

430 Feldmeier, K. & Hocker, B. Computational protein design of ligand binding and catalysis. *Curr Opin Chem Biol* **17**, 929-933, doi:10.1016/j.cbpa.2013.10.002 (2013).

431 Bick, M. J. *et al.* Computational design of environmental sensors for the potent opioid fentanyl. *Elife* **6**, doi:10.7554/eLife.28909 (2017).

432 Stratton, M. M. & Loh, S. N. Converting a protein into a switch for biosensing and functional regulation. *Protein Sci* **20**, 19-29, doi:10.1002/pro.541 (2011).

433 Kumita, J. R., Smart, O. S. & Woolley, G. A. Photo-control of helix content in a short peptide. *P Natl Acad Sci USA* **97**, 3803-3808, doi:DOI 10.1073/pnas.97.8.3803 (2000).

434 Kumita, J. R., Flint, D. G., Woolley, G. A. & Smart, O. S. Achieving photo-control of protein conformation and activity: producing a photo-controlled leucine zipper. *Faraday Discuss* **122**, 89-103, doi:10.1039/b200897a (2003).

435 Renner, C. & Moroder, L. Azobenzene as conformational switch in model peptides. *Chembiochem* **7**, 869-878, doi:10.1002/cbic.200500531 (2006).

436 Woolley, G. A. *et al.* Reversible photocontrol of DNA binding by a designed GCN4-bZIP protein. *Biochemistry-Us* **45**, 6075-6084, doi:10.1021/bi060142r (2006).

437 Blanco-Lomas, M., Samanta, S., Campos, P. J., Woolley, G. A. & Sampedro, D. Reversible Photocontrol of Peptide Conformation with a Rhodopsin-like Photoswitch. *J Am Chem Soc* **134**, 6960-6963, doi:10.1021/ja301868p (2012).

438 Zhang, F. Z., Timm, K. A., Arndt, K. M. & Woolley, G. A. Photocontrol of Coiled-Coil Proteins in Living Cells. *Angew Chem Int Edit* **49**, 3943-3946, doi:10.1002/anie.201000909 (2010).

439 Pandya, M. J. *et al.* Sequence and structural duality: Designing peptides to adopt two stable conformations. *J Am Chem Soc* **126**, 17016-17024, doi:10.1021/ja045568c (2004).

440 GarciaEcheverria, C. Disruption of coiled coil formation by methionine oxidation. *Bioorg Med Chem Lett* **6**, 229-232, doi:Doi 10.1016/0960-894x(96)00009-1 (1996).

441 Dado, G. P. & Gellman, S. H. Redox Control of Secondary Structure in a Designed Peptide. *J Am Chem Soc* **115**, 12609-12610, doi:DOI 10.1021/ja00079a060 (1993).

442 Andrew, C. D., Warwicker, J., Jones, G. R. & Doig, A. J. Effect of phosphorylation on alpha-helix stability as a function of position. *Biochemistry-Us* **41**, 1897-1905, doi:10.1021/bi0113216 (2002).

443 Szilak, L., Moitra, J., Krylov, D. & Vinson, C. Phosphorylation destabilizes alpha-helices. *Nat Struct Biol* **4**, 112-114, doi:DOI 10.1038/nsb0297-112 (1997).

444 Szilak, L., Moitra, J. & Vinson, C. Design of a leucine zipper coiled coil stabilized 1.4 kcal mol(-1) by phosphorylation of a serine in the e position. *Protein Sci* **6**, 1273-1283, doi:DOI 10.1002/pro.5560060615 (1997).

445 Signarvic, R. S. & DeGrado, W. F. De novo design of a molecular switch: Phosphorylation-dependent association of designed peptides. *J Mol Biol* **334**, 1-12, doi:10.1016/j.jmb.2003.09.041 (2003).

446 Errington, N. & Doig, A. J. A phosphoserine-lysine salt bridge within an alpha-helical peptide, the strongest alpha-helix side-chain interaction

measured to date. *Biochemistry-Us* **44**, 7553-7558, doi:10.1021/bi050297j (2005).

447 Groger, K., Gavins, G. & Seitz, O. Strand Displacement in Coiled-Coil Structures: Controlled Induction and Reversal of Proximity. *Angew Chem Int Edit* **56**, 14217-14221, doi:10.1002/anie.201705339 (2017).

448 Groth, M. C., Rink, W. M., Meyer, N. F. & Thomas, F. Kinetic studies on strand displacement in de novo designed parallel heterodimeric coiled coils. *Chem Sci* **9**, 4308-4316, doi:10.1039/c7sc05342h (2018).

449 Dueber, J. E. *et al.* Synthetic protein scaffolds provide modular control over metabolic flux. *Nat Biotechnol* **27**, 753-U107, doi:10.1038/nbt.1557 (2009).