

Analysis of Folded Structure and Folding Thermodynamics in Heterogeneous-Backbone Proteomimetics

Jacqueline R. Santhouse, Shilpa R. Rao, W. Seth Horne*

Department of Chemistry, University of Pittsburgh, Pittsburgh, PA, USA

E-mail: horne@pitt.edu

Running title: Folded Structure and Folding Thermodynamics in Proteomimetics

Keywords: Protein mimetics, X-ray crystallography, nuclear magnetic resonance spectroscopy, circular dichroism spectroscopy, isothermal titration calorimetry

Abstract

Recent years have seen a growing number of examples of designed oligomeric molecules with artificial backbone connectivity that are capable of adopting complex folded tertiary structures analogous to those seen in natural proteins. A range of experimental techniques from structural biology and biophysics have been brought to bear in the study of these proteomimetic agents. Here, we discuss some considerations encountered in the characterization of high-resolution folded structure as well as folding thermodynamics of protein-like artificial backbones. We provide an overview of the use of X-ray crystallography and NMR spectroscopy in such systems and review example applications of these methods in the primary literature. Further, we provide detailed protocols for two experiments that have proved useful in our prior and ongoing efforts to compare folding thermodynamics between natural protein domains and heterogeneous-backbone counterparts.

1. Introduction

Chemists have long sought to create artificial molecules with structural characteristics similar to proteins. In proteins, the sequence of amino acid side chains displayed on a polypeptide backbone specifies a corresponding three-dimensional folded conformation, and this folded shape gives rise to function. Mimicry of a particular protein fold provides a means for recreating the corresponding biological function. In the hierarchy of protein structure, primary sequence encodes for formation of local secondary structure motifs, which arrange into a diverse array of complex unimolecular tertiary folding patterns and multimolecular quaternary assemblies. Artificial molecules that mimic proteins share this same structural hierarchy. Most examples of synthetic scaffolds created to mimic peptides and proteins in prior work have involved reproduction of extended chain conformation or secondary structure with peptidomimetic agents (Pelay-Gimeno, Glas, Koch, & Grossmann, 2015). Seeking to broaden the scope of function possible in protein mimetics, research has moved beyond secondary structure as the target for mimicry toward artificial proteomimetic scaffolds that reproduce intricate tertiary folding patterns (Horne & Grossmann, 2020).

Achieving a defined tertiary fold with a synthetic scaffold is comparatively more difficult than secondary structure. The tertiary fold of a protein results from numerous long-range non-covalent interactions in the chain that are individually weak. Recreating the array of interactions typical in an evolutionarily optimized sequence with an artificial scaffold poses a considerable design challenge; however, meeting this challenge is important, as tertiary fold is the foundation for function in most proteins in nature. Among approaches developed for the creation of proteomimetic structures, one that has proved versatile is engineering backbone connectivity in natural domains, reviewed recently (Cabalteja & Horne, 2021). This method, sometimes termed

backbone engineering, entails the replacement of one or more α -amino acid residues within a sequence that specifies a particular fold with some counterpart that differs from nature in backbone covalent connectivity. As the molecules in question are typically prepared by chemical synthesis, the backbone compositions possible in such protein analogues are remarkably diverse. Whatever the artificial monomer type(s) and density of modification involved, the result is a heterogeneous-backbone analogue of a native protein, in which natural α -residues exist alongside artificial backbone units within the chain. In most reported examples of protein backbone engineering, modification has been isolated to a single site; in some cases, artificial connectivity has been interspersed throughout a chain. Regardless of the degree of unnatural backbone content, careful choice of monomer type and placement can yield molecules that reproduce a variety of complex tertiary folds encoded by an array of prototype sequences.

As the size and complexity of folded structure in artificial backbones has evolved from short chains that form isolated secondary structures to protein-sized scaffolds with complex tertiary folds, methods applied for the characterization of these entities have evolved in stride. In general, characterization of folded structure and folded stability in protein-sized artificial chains is more difficult than shorter peptide mimetics. A wealth of techniques from biophysics and structural biology historically applied to natural proteins have been successfully adapted to determine folded structure and folded stability in heterogeneous-backbone proteomimetics. Most of these methods translate well from biomacromolecules to artificial counterparts; however, some technical issues arise that are unique to the proteomimetic context. In the present work, we discuss methods applied for the characterization of folded structure and folded stability in heterogeneous-backbone tertiary structures. We discuss some considerations that are unique to proteomimetic analytes and review primary literature on the application of these methods.

Finally, we provide step-by-step protocols for two experiments we have found useful in comparing folding thermodynamics between natural domains and heterogeneous-backbone counterparts.

2. Characterization of folded structure in heterogeneous-backbone proteomimetics

A fundamental hypothesis in guiding work toward the development of heterogeneous-backbone proteomimetics is that the complex tertiary folds of diverse prototype natural sequences can be recreated by appropriately designed synthetic variants that differ from natural L- α -polypeptides in backbone covalent connectivity. Vital to obtaining data that can be used to address this hypothesis is the characterization of high-resolution folded structure in artificial, protein-sized entities. In the context of the broader field of protein backbone engineering, studies involving high-resolution structural characterization represent a small subset; however, these efforts have provided crucial insights into the interplay between backbone composition and folded structure. The standard methods most often applied for the characterization of folded structure in natural proteins—X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy—have also been shown amenable to proteomimetics. In a few cases, atomistic molecular dynamics (MD) simulation has been a useful complementary technique. Below we discuss some general considerations in the use of X-ray, NMR, and MD in the context of heterogeneous-backbone proteomimetics and review published examples describing the application of these methods.

2.1 X-ray crystallography

2.1.1 Overview of the method

Among biomacromolecular structures deposited in the Protein Data Bank (Berman et al., 2000), about nine out of ten have been determined by X-ray crystallography. Reflecting its central role in the characterization of natural proteins, X-ray crystallography is also the most common method applied in prior work on the characterization of tertiary structure in protein-like chains with artificial backbone composition. Crystallography provides an important benefit in such efforts, as data of sufficient quality can yield electron density maps that provide unambiguous information about both the overall fold of a proteomimetic domain as well as details on the behavior of the artificial monomer(s) it contains. An in-depth discussion of macromolecular X-ray crystallographic methods is outside the scope of the present chapter, which will focus instead on specific issues encountered in its application to heterogeneous backbones.

Two important challenges in the use of X-ray crystallography with proteomimetic analytes are also encountered with natural proteins: (1) the requirement for diffraction quality single crystals and (2) the need for a way to solve the so-called “phase problem” in order to obtain an electron density map from a measured diffraction pattern. To the first point, crystallinity of a prototype protein is often a good predictor of the ability to obtain crystals of heterogeneous-backbone mimetics. However, the demands of material quantity for a typical crystallization screening campaign can be more difficult to meet for protein-sized chains prepared by total chemical synthesis compared to natural proteins produced by heterologous expression. In terms of the phase problem, any method used in the characterization of natural biomacromolecules is potentially applicable to artificial backbones. Both molecular replacement and experimental phasing have been used successfully, as detailed below. With molecular replacement, a search model derived from the natural domain that is the prototype for mimicry

often provides a convenient basis for generating an initial solution; however, steps must be taken to minimize phase bias from the input model around the artificial residue(s) of interest in the resulting electron density map. In the realm of experimental phasing, the fact that most heterogeneous-backbone proteomimetics are prepared by chemical synthesis opens a range of possibilities (e.g., covalent incorporation of heavy atom labels using modified amino acids).

A separate set of issues impacting the application of X-ray crystallography to heterogeneous backbones that is unique to the artificial chain relates to software used for structure determination. Most packages developed for macromolecular X-ray structure refinement, such as Refmac (Murshudov et al., 2011) and Phenix (Liebschner et al., 2019), can readily handle proteins with side chains beyond the 20 canonical amino acids; however, they are not designed with the artificial backbones encountered in many proteomimetics in mind. Refinement of modified backbones with protein-like composition requires parameters for both artificial residues of interest as well as linkages that define their connectivity in the chain. Simple geometric terms (i.e., bond lengths, angles) for artificial units can be obtained from related natural monomers already parameterized in the software and/or from X-ray structures of related small molecules. In contrast, accurate potentials for torsional angles and non-bonded interactions are not as straightforward to derive. For higher resolution X-ray structures, the electron density map often provides unambiguous information about artificial residue conformation, and the accuracy of parameters beyond basic geometry is not essential to the end result of refinement. However, the quality of these parameters can have a more pronounced effect for modest resolution maps (i.e., below ~ 3 Å). A final class of software impacted by altered backbone composition are algorithms used for validation of refined models. Consultation of the metrics output by these programs, such as Molprobity (Williams et al., 2018), is important to both inform

an ongoing refinement and assess the quality of a final structure. As with software for refinement, these algorithms are typically not designed to treat many backbone types found in proteomimetics, and the incorporation of artificial connectivity can lead to false identification of severe problems. Statistics for the natural α -residues present in a heterogeneous backbone can be a useful overall gauge of model quality; however, the utility of this information varies based on the density of modification in a given chain.

2.1.2 Example applications in the literature

The above experimental challenges notwithstanding, X-ray crystallography has proved invaluable in the high-resolution structural characterization of a number of heterogeneous-backbone tertiary folds. Most work in this area has involved chains with altered backbone composition limited to a single site. Such modification can be used to engineer protein properties as well as shed light on fundamental aspects of folding. One of the earliest types of isolated backbone engineering explored in the context of a tertiary fold was replacement of the amide nitrogen atom – first with sulfur (Schnolzer & Kent, 1992) and later with oxygen (Lu, Qasim, Laskowski, & Kent, 1997). In the case of substitution of an α -amino acid with the corresponding α -hydroxy acid analogue, the resulting amide $\rightarrow$ ester replacement has proved a useful means to explore the role of backbone hydrogen bonding in folding (Yang, Wang, & Fitzgerald, 2004). A key assumption underlying such work is that the chemical change has minimal impact on folded structure. In several cases, this hypothesis has been tested through X-ray crystallographic characterization of an ester-containing protein. Reported examples include ester variants of the bacterial potassium channel KcsA (Valiyaveetil, Sekedat, MacKinnon, & Muir, 2006), the enzyme HIV-1 protease (Torbeev et al., 2011), and a coiled coil derived from the yeast transcription factor GCN4 (Dadon, Samiappan, Shahar, Zarivach, & Ashkenasy, 2013).

Another common class of point substitution to the protein backbone is that where the change is made to alter folding propensity. For example, the incorporation of D-amino acid residues results in heterochiral chains with locally altered conformational preferences, and the D-residue can be used to rigidify the backbone when replacing a glycine that adopts a conformation disallowed for L- α -residues. A related approach to backbone rigidification is the incorporation of the C $_{\alpha}$ dimethylated monomer 2-aminoisobutyric acid (Aib), which has a restricted accessible conformational space. X-ray crystallography has been used to characterize D-residues as well as Aib in the context of folded protein domains, including ubiquitin (Figure 1A,B) (Bang et al., 2006; Bang, Makhatadze, Tereshko, Kossiakoff, & Kent, 2005), HIV-1 protease (Torbeev et al., 2011), and the KcsA potassium channel (Valiyaveetil, Leonetti, Muir, & MacKinnon, 2006). A final example of an isolated rigidifying protein backbone substitution characterized by X-ray is an artificial dipeptide surrogate used to reinforce a β -turn found in the trimeric fibrin foldon domain (Eckhardt, Grosse, Essen, & Geyer, 2010).

In addition to those above, the elongation of the backbone in folded protein domains utilizing β -amino acid analogues of natural α -residues represents another common class of modifications. In an example of the exploration of the structural consequences of β -residue incorporation in a tertiary fold, point $\alpha \rightarrow \beta$ residue substitutions were made in the villin headpiece (VHP), a compact helix-rich domain (Kreitler, Mortenson, Forest, & Gellman, 2016). Notable from a methodological standpoint was the use of quasiracemic crystallography in the structure determination. Racemic crystallography involves the crystallization of a 1:1 mixture of natural L-protein and its D-enantiomer; such racemic samples tend to be more crystalline and the resulting diffraction data sets easier to phase compared to enantiopure counterparts (Yeates & Kent, 2012). Quasiracemic crystallography is a variation of the above method where there is

some subtle structural difference between the two protein enantiomers. In the study of VHP, the mixture crystallized consisted of the D-enantiomer of the native domain and the L-enantiomer of a variant bearing a single $\alpha \rightarrow \beta$ residue substitution. Besides facilitating crystallization and phasing, an added benefit of quasiracemic crystallography in the context of heterogeneous backbones is the built-in control structure for comparison through inversion of the native-backbone coordinates from the co-crystal followed by overlay. Additional work on the crystallographic characterization of proteins with isolated $\alpha \rightarrow \beta$ residue substitution has been done in the Pin1 WW domain (Mortenson et al., 2018).

In all the examples discussed above, the modification to the protein backbone was isolated to a single site in the sequence. X-ray crystallography has also proved useful in the study of protein-like chains with backbones deviating more substantially from nature in chemical composition. In the realm of completely artificial backbones, structures of protein-sized helix-bundle assemblies have been documented for sequences composed of β -residues (Daniels, Petersson, Qiu, & Schepartz, 2007), urea-based monomers (see Chapter 5 of this volume) (Collie et al., 2017; Collie et al., 2015), and aromatic monomers (De et al., 2018). A number of high resolution X-ray structures have also been reported for a series of helix bundles formed by heterogeneous α/β -peptide backbones displaying side-chain sequences derived from the dimerization domain of the yeast transcription factor GCN4 (Giuliano, Horne, & Gellman, 2009; Horne, Price, & Gellman, 2008; Horne, Price, Keck, & Gellman, 2007; Price, Horne, & Gellman, 2010). A related quaternary structure resulting from peptide self-assembly that has been subjected to backbone engineering is the collagen triple helix, where X-ray crystallography was applied to variants in which the C_α atom in a glycine residue was replaced by nitrogen (Kasznel, Zhang, Hai, & Chenoweth, 2017).

The above precedent on high resolution structural characterization in proteomimetics involves two complementary areas: isolated modification in the context of protein-sized tertiary folds and extensive backbone modification in assemblies of smaller helical peptidomimetics. An important challenge exists at the interface of these areas—creating protein-sized chains with extensively modified backbones and complex tertiary folding patterns. We have reported an approach to this problem based on the application of diverse backbone alteration types in a single chain to produce heterogeneous-backbone mimics of tertiary folding patterns encoded by natural protein sequences (George & Horne, 2018). In the development of this approach, we made extensive use of X-ray crystallography to compare the structural impact of different backbone alteration strategies in the protein GB1 (Figure 1C,D) (Reinert & Horne, 2014; Reinert, Lengyel, & Horne, 2013; Tavenor, Reinert, Lengyel, Griffith, & Horne, 2016). All these structures were solved by molecular replacement, using either the native domain or an already characterized variant as search model. In an example of the structural characterization of a proteomimetic with an extensively modified backbone in the context of a protein-protein interface, a backbone-modified analogue of a helix-turn-helix domain was crystallized in complex with its receptor VEGF (Checco et al., 2015).

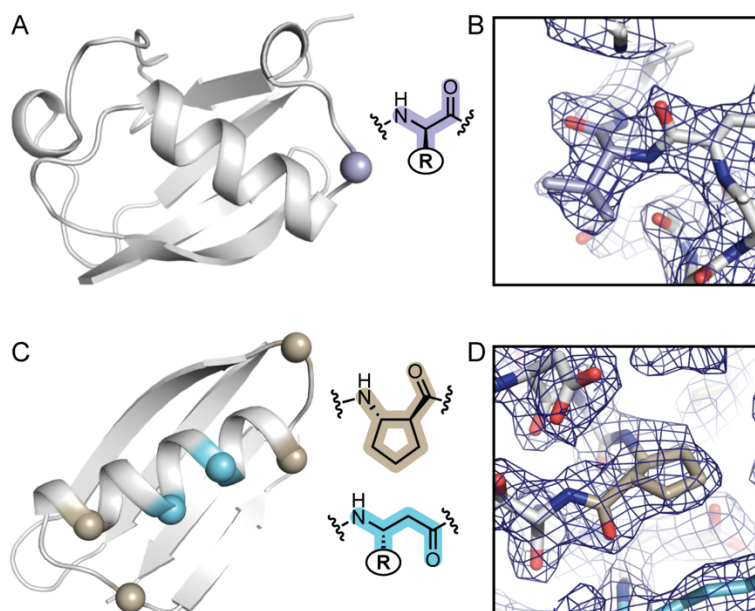


Figure 1. Example structures of heterogeneous-backbone proteomimetics determined by X-ray crystallography. (A) Cartoon representation of the crystal structure of a variant of ubiquitin (PDB 2FCN) bearing a single D-residue (sphere) alongside the chemical structure of the artificial monomer. (B) View of the electron density map ($2F_o - F_c$ contoured at 1σ) surrounding the artificial residue in the refined structure. (C) Cartoon representation of the X-ray crystal structure of a variant of GB1 (PDB 4OZC) bearing cyclic and acyclic β -residues (spheres) alongside chemical structures of the artificial monomers. (D) View of the electron density map ($2F_o - F_c$ contoured at 1σ) surrounding one of the artificial residues in the refined structure.

Another protein class where multi-site backbone engineering has yielded important structural insights are hairpin domains derived from amyloid-forming sequences (Kreutzer & Nowick, 2018). X-ray crystallographic characterization of these scaffolds has provided valuable structural information about soluble assemblies formed by amyloidogenic sequences relevant to neurodegenerative disease (Spencer, Li, & Nowick, 2014) as well as potential inhibitors of

amyloid formation (Cheng, Liu, Zhao, Eisenberg, & Nowick, 2012) (see Chapter 8 of this volume). Of note from a methods perspective, experimental phasing facilitated by heavy atom incorporation during chemical synthesis proved instrumental in solving structures for the large, multi-chain unit cells typical in these systems.

2.2 Nuclear magnetic resonance spectroscopy

2.2.1 Overview of the method

Alongside X-ray crystallography, NMR spectroscopy has found widespread use in the characterization of folded structure in proteomimetics. Below, we discuss some considerations in the application of this technique to artificial, protein-sized chains. As is the case for natural proteins, a key advantage of NMR over X-ray for characterization of proteomimetics is that NMR does not require a domain of interest be crystalline. Further, NMR can provide information about dynamics and folding pathways as well as high resolution structure. While there are technical barriers to application of NMR to very large folds, the limits in chain length accessible by chemical synthesis means that proteomimetic systems of interest are well below the molecular weight thresholds where these issues become relevant. Material quantity demands for NMR characterization are similar to that needed for crystallization; a few milligrams is typically sufficient for a complete set of experiments to support determination of a high-resolution folded structure.

One essential feature distinguishing NMR of heterogeneous-backbone proteomimetics from natural counterparts relates to the incorporation of NMR-active isotopes of low natural abundance (i.e., ^{15}N , ^{13}C). In proteins produced by heterologous expression, global isotopic labeling is straightforward, cost effective, and standard. In contrast, incorporation of isotopic labels in synthetic proteins, while feasible from a chemical standpoint, is impeded by the

extremely high cost of isotopically labeled protected amino acid monomers. This issue makes routine acquisition of the multidimensional heteronuclear experiments common in NMR characterization of natural proteins difficult or impossible for proteomimetic variants. Homonuclear experiments are typically sufficient for complete assignment of proton chemical shifts and structure determination for small domains; however, resonance overlap can lead to complications in larger systems. As discussed below, judicious incorporation of isotopic labels at specific sites in a synthetic construct has proved beneficial in some instances. In other cases, heteronuclear experiments obtained with natural isotope abundance material have been shown to be a viable alternative. Finally, the application of methods for NMR structure calculation that allow for ambiguity in input peak assignments has proved effective.

As with X-ray crystallographic analysis, software used for calculation and validation of structure ensembles for proteins from NMR experimental restraints must be modified to handle heterogeneous backbones. The quality of the force field parameters for artificial monomers has a comparatively larger impact in the case of NMR than X-ray, as the NMR experimental restraints (predominantly interatomic distances) are inherently more ambiguous than an electron density map. Moreover, limitations in programs used to assess quality of protein models pose more of an issue in the case of NMR, because NMR structures lack an analogous metric to the R/R_{free} values that guide crystal structure refinement. One software-based innovation we have found valuable in the structural characterization of proteomimetics by NMR is to make use of ambiguous distance restraints during the calculation, as implemented in the program ARIA (Rieping et al., 2006). The idea underlying the algorithm is to replace the typical manual generation of interatomic distance restraints from assignment of peaks in the NOESY spectrum. Instead, ARIA generates distance restraints in automated fashion in the course of a structure calculation from an

input set of chemical shift assignments and list of unassigned peaks from the NOESY spectrum. The assignment of the NOESY peaks and resulting NOE distance restraints is refined by the program in an iterative fashion, with models output from each cycle in the overall structure calculation informing later iterations. The use of this automated approach minimizes bias in the obtained structure from potential inaccurate assignments in manually assigned peaks and allows for more restraints to inform the final ensemble through inclusion of NOESY signals with multiple contributing interatom correlations. In the case of natural proteins, ARIA has been shown to yield structural ensembles of improved convergence and quality compared to structures determined from the same source NMR data by other methods (Mareuil, Malliavin, Nilges, & Bardiaux, 2015).

2.2.2 Example applications in the literature

As in application of crystallography, most reported examples describing the use of NMR for structural characterization of heterogeneous-backbone proteomimetics involve modification at a single site in a protein of interest. Some of this work has been conducted in β -turns, where NMR yielded structures of various turn mimetics in the context of a zinc finger domain (Viles et al., 1998), the disulfide-rich domain scyllatoxin (Jean et al., 1998), GB1 (Odaert et al., 1999), and the WW domain of Pin1 (Figure 2A) (Fuller et al., 2009). Notable from a methodology perspective, the above efforts all involved synthetic material with natural isotope abundance and homonuclear experiments (e.g., TOCSY, COSY, NOESY). NMR has also been applied to structure determination for a heterochiral variant of the Trp cage miniprotein TC5b bearing a single D-residue (Rodriguez-Granillo, Annavarapu, Zhang, Koder, & Nanda, 2011). In this work, homonuclear experiments were complemented by a natural-abundance sensitivity-enhanced $^1\text{H}/^{13}\text{C}$ -HSQC. Further, resolution of ambiguous assignments in the single D-glutamine required

analysis of a second peptide bearing uniform $^{15}\text{N}/^{13}\text{C}$ labeling at this site. To avoid the cost associated with the isotopically labeled D-residue, the authors prepared the corresponding enantiomeric D-protein bearing a single isotope-enriched L-glutamine at the site of interest. Another study made use of site-specific isotopic labeling to explore the consequences of backbone alteration in variants of HIV-1 protease (Torbeev et al., 2011). Here, the structure of the variant was determined by crystallography (discussed above), and NMR was used to obtain complementary information on dynamics.

In a few recent cases, NMR has been used for determination of tertiary folded structure in proteomimetics with more extensively modified backbones. One example involved a heterochiral variant of a disulfide-rich knottin domain in which a solvent-exposed loop was replaced with an all-D-residue segment (Mong et al., 2017). In this effort, homonuclear ^1H experiments were complemented by natural abundance $^1\text{H}/^{15}\text{N}$ - and $^1\text{H}/^{13}\text{C}$ -HSQC. NMR has also been applied to the characterization of zinc fingers with highly modified backbones. We have determined structures of heterogeneous-backbone variants of two domains from the DNA binding region of the transcription factor Sp1 (George & Horne, 2017; Rao & Horne, 2020). Between our first and second report involving the Sp1 system, we began to apply iteratively assigned ambiguous restraints as implemented in ARIA during structure calculations. The impact of this change is seen in comparison of the final models obtained for the closely related Sp1 domain 2 mimetic (automated NOE assignment with ARIA) and the Sp1 domain 3 mimetic (manual NOE assignment). While the overall folds found for the proteomimetic domains by the two different methods are very similar, statistics obtained from Molprobitry suggest the model obtained using ARIA is of higher quality (average Molprobitry score for the ensemble 2.2 ± 0.4 with ARIA compared to 3.9 ± 0.3 without). We have also applied ARIA in the characterization of a

heterogeneous-backbone mimic of a disulfide-rich miniprotein (Figure 2B) (Cabalteja, Mihalko, & Horne, 2019), where the approach proved similarly effective. NMR has also been used to determine the structure of a zinc finger domain mimetic in which the entire α -helix was replaced by an artificial oligourea-based backbone (Lombardo et al., 2019). Of methodological note in that study, resolving complexities in resonance assignments in the abiotic helix required homonuclear and heteronuclear experiments as well as data acquisition on a cryoprobe-equipped 950 MHz spectrometer.

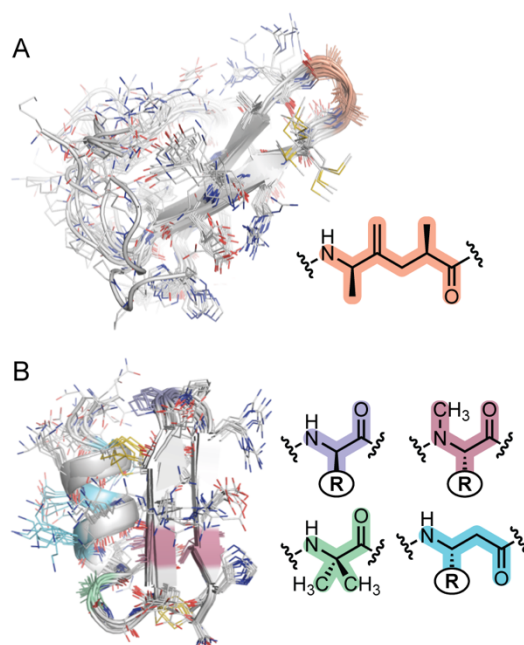


Figure 2. Example structures of heterogeneous-backbone proteomimetics determined by NMR spectroscopy. (A) NMR structure ensemble for a variant of the Pin1 WW domain (PDB 2KBU) with a dipeptide surrogate (chemical structure shown) as an artificial β -turn inducer. (B) NMR structure ensemble for a variant of a disulfide-rich miniprotein (PDB 6E5J) bearing four different classes of artificial monomers (chemical structures shown) in the chain.

2.3 Molecular dynamics simulation

Complementary to the experimental characterization of proteomimetics by NMR and X-ray crystallography, molecular dynamics (MD) simulation has also been applied to study these systems. MD can provide atomistic information about dynamics difficult to obtain by other methods. The quality of data obtained from an MD simulation is crucially dependent on the force field used for the calculations, and the development of force fields for treatment of natural biomacromolecules is a vibrant area of research (Lopes, Guvench, & MacKerell, 2015). Application of MD to artificial backbones requires the determination of suitable force field parameters for all the residue types involved. In the case of heterogeneous backbones, these must be compatible with corresponding parameters for native α -amino acids. Perhaps as a result of this technical hurdle, most prior work involving MD simulations of heterogeneous-backbone proteomimetics has involved chains with isolated D-residue substitution, which can be parameterized based on natural L-residues in an existing force field. MD simulations have been applied to characterize heterochiral variants of a Trp cage miniprotein TC5b (Rodriguez-Granillo et al., 2011), determine optimal D-residue placement in a helix-turn-helix domain (Simon et al., 2016), and elucidate the impact of backbone modification on active site dynamics in the enzyme HIV-1 protease (Torbeev et al., 2011).

Turning to simulations involving heterogeneous backbones containing monomer classes beyond D-residues, modified versions of the CHARMM force field have been developed for the treatment of N-alkylglycine residues (Mirijanian, Mannige, Zuckermann, & Whitelam, 2014) as well as C $_{\alpha}$ - and N-methylated analogues of α -residues (Vanommeslaeghe & MacKerell, 2015). In collaboration with Elcock, we applied MD to investigate the behavior of ethylene glycol oligomers as backbone elements in the loops of a bacterial protein (Reinert, Musselman, Elcock,

& Horne, 2012); in that study, we adapted previously developed force field parameters for ethylene glycol polymers (Fischer, Paschek, Geiger, & Sadowski, 2008). In collaboration with Chong, we have recently worked to broaden the scope of atomistic MD simulation of proteomimetics through development of a force field to simulate heterogeneous backbones bearing multiple artificial monomer classes in a single chain (Bogetti et al., 2020). The initial version, denoted AMBER ff15ipq-m, includes parameters for β^3 -residue analogues of the 20 natural proteinogenic amino acids, two cyclic β -amino acid residues, D- α -residues, and the C $_{\alpha}$ -Me- α -residue Aib. The parameters for these residues were derived using the implicitly polarized charge method (Debiec et al., 2016) and are compatible with the AMBER lineage of force fields. MD simulations conducted using AMBER ff15ipq-m were shown to reproduce experimental observables in tripeptides as well as heterogeneous-backbone tertiary folds (Bogetti et al., 2020).

3. Analysis of folding energetics in proteomimetics

In addition to structure, determining folded stability and folding thermodynamics are important goals in research on heterogeneous-backbone proteomimetics. In terms of methodology, the entire range of biophysical techniques developed and applied to assess folded stability in natural proteins is potentially applicable to artificial chains. Unlike the case for determination of folded structure, most of these experiments can be performed without any modification needed to account for artificial backbone content in a domain of interest. A detailed discussion of methods for determination of protein folded stability is outside the scope of this review; the interested reader is directed to an authoritative text on the subject (Fersht, 1999). As with structural studies, the bulk of research on folding thermodynamics in heterogeneous-backbone proteomimetics has involved modification at a single site in a folded domain. Most of the references involving structure determination by X-ray or NMR cited above include

experiments aimed at measuring folded stability; additional work in this area has been reviewed recently (Cabalteja & Horne, 2021). Below, we provide detailed protocols for two techniques we have found valuable for determining the impact of backbone modification on folding thermodynamics in proteomimetics with highly modified chains. Previously developed and applied to the characterization of folded stability in natural protein domains, both have proved vital in refining design strategies toward heterogeneous-backbone proteomimetics that retain folded stability of a prototype natural sequence.

3.1 Comparison of folding thermodynamics for a bacterial protein and variant by circular dichroism

Circular dichroism (CD) spectroscopy is based on the measurement of differential absorption of opposing circularly polarized light and is a sensitive means to gain insight into peptide and protein folding (Adler, Greenfield, & Fasman, 1973). The shape of a protein CD spectrum in the far-UV region is dominated by electronic transitions of backbone amides and varies in a systematic way with secondary structure content. Interpretation of CD scans for heterogeneous backbones is complicated by the impact of the artificial monomers on typical CD signatures for various secondary structures. Thus, CD is more straightforward in artificial systems as a means to follow a change in folded structure as a function of disruption by temperature or some chaotropic agent. A more in-depth understanding of the thermodynamics of a system can be achieved by monitoring unfolding as a function of both these variables in a single experiment. Thus, measurement of multiple CD thermal melts on samples with varying concentrations of chemical denaturant yields a data set that can be globally fit to determine the change in free energy (ΔG), enthalpy (ΔH), and entropy (ΔS) associated with folding as well as the heat capacity difference between the folded and unfolded states (ΔC_p) and the dependence of

folding free energy on denaturant concentration (m). The latter parameters, ΔC_p and m , have embedded in them insights about the aspects of folding, such as solvation, that are difficult to assess by other methods. This method was first used to determine isotope effects in protein folding (Kuhlman & Raleigh, 1998) and later applied to characterize folding in other systems, including an artificial protein domain with fluorinated side chains (Buer, Levin, & Marsh, 2012). We have found it a powerful means to characterize folding thermodynamics in heterogeneous-backbone proteomimetic variants of the bacterial protein GB1 (Reinert & Horne, 2014; Tavenor et al., 2016) and other systems (Haney, Werner, McKay, & Horne, 2016). In collaboration with Chenoweth and Petersson, we have also used the experiment to determine the thermodynamic effect of backbone amide→thioamide substitution on protein folding (Walters et al., 2017). Below is a step-by-step protocol for comparison of a native protein and backbone-modified variant by the method.

3.1.1 Equipment

1. CD spectrometer (Olis DSM-17) equipped with Peltier temperature controller

Note: If using alternate equipment, the experiment is facilitated by use of a CD instrument equipped with a sample changer, which enables thermal melts to be conducted on several samples in parallel.

2. 1-mm path length quartz cuvettes (Starna Cells)

Note: Sample volumes in the protocol below will need to be adjusted if alternate size cells are used. Due to the high absorbance of chemical denaturants in the far-UV region, narrow path length cuvettes with appropriate holders typically yield the best results.

1. Cuvette washer (Fisher)
2. UV-visible spectrophotometer (Olis HP 8452)

3. pH meter (Fisher Accumet AB15+)

3.1.2 Reagents

1. 8 M guanidinium chloride
2. 2.5 mM solution of each synthetic protein of interest, concentration determined by UV absorbance (Gill & von Hippel, 1989)
3. 0.2 M phosphate buffer, pH 7

Note: Optimal buffer selection may vary with the system under study. Factors to consider include the desired pH, compatibility issues with the domain of interest, and buffer absorbance in the wavelength range for CD measurements. In general, we have found buffer conditions suitable for CD characterization of a given native protein translate well to heterogeneous-backbone variants of that protein.

4. Methanol
5. 18 MΩ H₂O

Note: All stock solutions and subsequent dilutions are prepared using 18 MΩ water.

3.1.3 Preparation of samples for analysis

1. Determine the range of guanidinium chloride concentrations needed to yield a complete chemical denaturation curve for the protein of interest at 2 °C. Divide this concentration range into 10-12 increments (e.g., 0-6 M, sampling every half molar).

Note: When comparing variants with different folded stabilities, it may be necessary to adjust the concentration range of guanidinium and sampling increment. It is best to maintain a similar number of concentrations for each variant and avoid oversampling the unfolded baseline.

2. Calculate dilutions required to yield a series of 300 µL samples for analysis composed of 50 µM protein, 10 mM buffer, and guanidinium at each concentration determined in the

prior step. Calculate dilutions required to yield a corresponding series of blanks identical to the above samples but without protein.

3. Prepare samples and corresponding blanks in individually labeled microcentrifuge tubes.
4. Vortex all samples and blanks to ensure homogeneous solutions.
5. Wash all cuvettes with water and methanol using a cuvette cleaning set up, and dry with nitrogen.

Note: For a more rigorous cleaning of cuvettes prior to the experiment fill each cuvette ~75% of the way with water, add one drop of concentrated nitric acid, and let stand for 10 min.

Thoroughly rinse cuvettes with water, and dispose of resulting waste (HAZARD: aqueous nitric acid waste must be kept separate from organics). Rinse cuvettes again with water, then with methanol. Dry under nitrogen prior to use.

3.1.4 Data collection

1. Determine wavelength to monitor in unfolding experiments from a full CD scan at 20°C.
 - a. Add blank corresponding to the 0 M guanidinium sample into cuvette, clean cuvette with lens paper, and insert into CD sample chamber.
 - b. Collect a wavelength scan in the range 200-260 nm (1 nm increment, 3 sec averaging time).
 - c. Clean out cuvette, add corresponding 0 M guanidinium protein sample, and collect as above.
 - d. Determine the most intense minimum above 210 nm in the resulting baseline-corrected CD spectrum. As guanidinium absorbs strongly below 210 nm, monitoring subsequent melts above this cutoff avoids complicating factors from the denaturant.

2. Set instrument to collect two wavelengths for each subsequent scan: 260 nm and the wavelength determined in step 1 (3 sec averaging time).

Note: Collecting a second wavelength that lacks contributions from the protein (260 nm) provides an internal check for variation of the baseline as a function of time or temperature.

3. Transfer blanks corresponding to the first set of guanidinium concentrations to empty cuvettes. Collect scans for each at 20°C.

Note: The 0 M blank does not need to be measured again. The data can be obtained from the full scan collected in step 1.

4. Wash out cuvettes and add samples containing protein. Match each protein sample to the cuvette used for the corresponding guanidinium concentration blank. To collect thermal melts for these samples, set CD to repeat scans as a function of temperature (2-98°C, 2°C increment, 2 min equilibration at each new temperature). Include a scan at 25°C after the melt is complete.
5. Repeat steps 3-4 with new sets of guanidinium concentrations until all samples have been analyzed.

Note: Clean the outside of each cuvette with lens paper before each measurement. Ensure cuvettes are oriented consistently in the sample chamber for each measurement. Place cuvettes in the same position of sample changer for corresponding blank and sample for a given concentration of denaturant.

3.1.5 Data analysis

1. Export results in ascii formatted files for analysis.
2. For each concentration of guanidinium, correct the raw measured ellipticity for the sample (θ_{obs}) using values for the corresponding blank (θ_{blank}) following the equation:

$$(\theta_{220,\text{obs}} - \theta_{220,\text{blank}}) - (\theta_{260,\text{obs}} - \theta_{260,\text{blank}})$$

Convert resulting raw CD values to molar ellipticity (Adler et al., 1973). Repeat this step for each guanidinium concentration of interest.

3. Generate an ascii formatted text file consisting of an array of values for molar ellipticity as a function of temperature and denaturant concentration (i.e., each row consisting of a set of three values: [guanidinium] (M), T (K), and $[\theta]$ ($\text{deg cm}^2 \text{ dmol}^{-1} \text{ residue}^{-1}$)).
4. Perform global fit of the resulting set of data points to a two-state folding model using equations described previously (Kuhlman & Raleigh, 1998). We employ the *NonlinearModelFit* function in Mathematica for this step, but other programs with similar capabilities are also applicable. A typical fit result is shown in Figure 3A.

Note: If a single denaturant concentration from the experiment shows an obvious mismatched baseline or some other anomaly (e.g., Figure 3B), exclude it from the fit or prepare a fresh sample and blank for that concentration and recollect. During the fitting process, restraints to baseline parameters may be appropriate in some cases. The assumption in the model is a linear dependence of molar ellipticity on temperature as well as on denaturant for both the folded and unfolded states. In some cases of variants with high sensitivity to chemical denaturant, linear dependence on the concentration of denaturant can be difficult to determine reliably (Figure 3C). In these instances, it is best to restrain the corresponding slope to 0 during the fit.

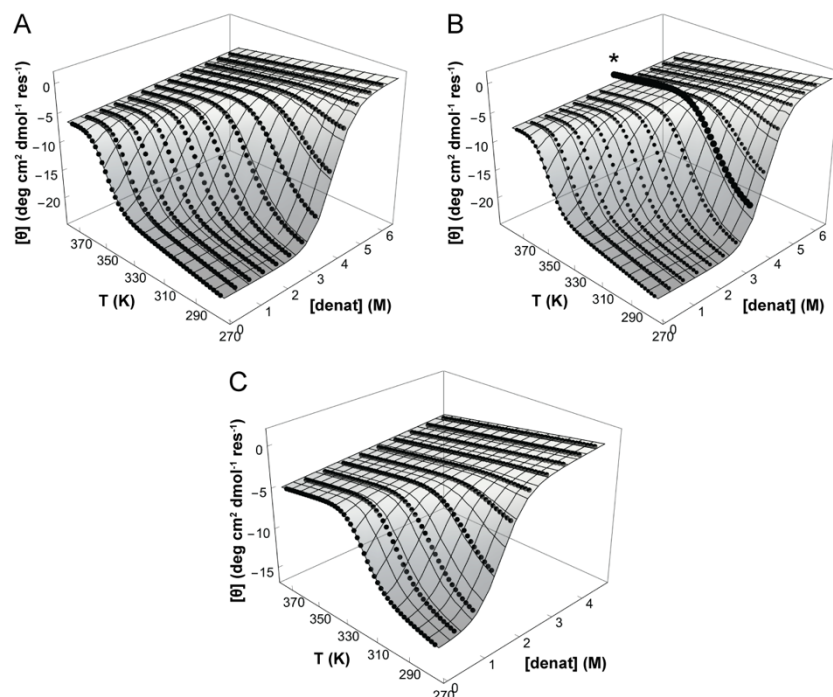


Figure 3. Simulated data points (circles) and fits (surfaces) depicting results from typical experiments utilizing CD to monitor unfolding of a heterogeneous-backbone proteomimetic as a function of temperature and chemical denaturant. (A) Example of a data set without any anomalies. (B) Example of an experiment with an outlier temperature series (marked with an asterisk) corresponding to a single denaturant concentration. (C) Example result where the linear dependence of the folded baseline ellipticity as a function of chemical denaturant is poorly defined.

5. Compile the resulting table of thermodynamic parameters and uncertainties along with graphs of the resulting fits.
6. Repeat entire experiment for series of variants for a given protein of interest to obtain insights into the impact of backbone alteration on folding energetics.

3.2 Comparison of folding thermodynamics for a zinc finger domain and variant by isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is a common biophysical method used to investigate the thermodynamics of binding interactions involving proteins (Grossoehme, Spuches, & Wilcox, 2010). In a typical ITC experiment, some ligand that binds to a protein of interest is titrated into a cell containing a solution of that protein. The sample cell is maintained under isothermal conditions with a reference cell, and the heat change associated with interaction between the ligand and protein measured over the course of the titration. The applied power needed to maintain isothermal conditions between the sample and reference cells over time is plotted in the thermogram, the peaks from which are integrated and plotted against molar ratio to generate a sigmoidal binding curve. Fitting this curve to an appropriate binding model yields the change in enthalpy (ΔH), affinity constant (K), and stoichiometry (n) associated with binding. These values can be used, in turn, to determine the free energy (ΔG) and entropy (ΔS) of the process. While ITC is more directly tied to energetics of binding than of folding, the method can provide useful information about folding thermodynamics in systems where folding and binding are closely coupled. One system where this is the case is the zinc finger domain.

Zinc finger domains are a ubiquitous class of metallopeptides whose tertiary structure formation is driven by tetrahedral coordination of a zinc metal ion by four Cys and/or His residues (Klug, 2010). Because of their coupled folding and metal binding, the thermodynamics measured by ITC upon addition of zinc to a zinc finger sequence has contributions from metal-peptide binding as well as folding. When thermodynamic values are obtained under identical experimental conditions for a series of sequences with identical metal-binding residues, thermodynamics involving metal-peptide and metal-buffer interactions cancel out and differences observed by ITC

are dominated by differences in folding energetics (Berg & Godwin, 1997). As a result of the above considerations, ITC has proved a powerful tool in the characterization of zinc finger folding. In the realm of proteomimetics, we have used ITC to explore the impact of backbone modification on folding thermodynamics in heterogeneous backbone mimics of zinc finger sequences (George & Horne, 2017; Rao & Horne, 2020). Below, we present a step-by-step protocol for comparison of folding energetics between a native zinc finger domain and variant by the method.

3.2.1 Equipment

1. Isothermal titration calorimeter (Malvern MicroCal iTC200)

Note: The Malvern MicroCal iTC200 system has a 200 μL sample cell and 40 μL syringe. Some details of the protocol below will require modification for alternate instrumentation.

2. UV-visible spectrophotometer (Olis HP 8452)
3. pH meter (Fisher Accumet AB15+)

3.2.2 Reagents

1. Lyophilized synthetic peptide
2. Zinc chloride (ZnCl_2)
3. 18 M Ω water
4. 200 mM HEPES, pH 7.4

Note: Alternate buffers may be used, but those that can chelate Zn^{+2} ions or have large protonation enthalpies should be avoided. The coordination environment of zinc and number of protons released upon metal binding are highly dependent on pH. Both buffer identity and pH should be kept consistent across all experiments in a series of variants to allow comparison of folding energetics from ITC results.

5. 200 mM sodium chloride

6. Contrad 70 detergent
7. Ellman's reagent – 2,2'-dithio-bis-(2-nitrobenzoic acid) (DTNB)
8. Zincon reagent – 2-carboxy-2'-hydroxy-5'-sulfoformazyl benzene

3.2.3 Preparation of samples for analysis

1. Dissolve lyophilized zinc finger peptide or variant of interest in water, targeting a concentration of ~1 mM.

Note: Water used to prepare this and all solutions below for the ITC experiment should be degassed by argon bubbling or freeze-pump-thaw cycles.

2. Determine the concentration of this stock solution using Ellman's assay to quantify free sulfhydryl groups (Ellman, 1959).
3. Prepare a 10 mM zinc stock solution in water. Measure the concentration of this stock solution by Zincon assay (Säbel, Neureuther, & Siemann, 2010).
4. Prepare a 200 mM stock solution of HEPES at pH 7.4 and 200 mM stock solution of NaCl.
5. Calculate dilutions required to yield the following solutions from stocks above.
 - a. 500 μ L of a 150 μ M peptide in 50 mM NaCl and 50 mM HEPES at pH 7.4
 - b. 5 mL of a 1 mM zinc in 50 mM NaCl and 50 mM HEPES at pH 7.4
 - c. 30 mL of 50 mM NaCl and 50 mM HEPES at pH 7.4

Note: In general terms, the peptide needs to be in sufficient concentration such that interaction with zinc provides a heat change detectable by the instrument. Further, the concentration should be guided by the binding affinity for zinc. A useful metric is Wiseman's parameter (c); a value of c in the range of 1 and 1000 gives a sigmoidal binding curve from which the thermodynamic parameters can be reliably calculated (Wiseman, Williston, Brandts, & Lin, 1989). The concentration of the zinc stock should be 10-20 times that of the peptide.

6. Prepare samples corresponding to solutions from step 4. Adjust the pH of each sample to the desired value after all components are present.

Tip: To avoid additional heat effects from the buffer mismatches during the titration, exercise great care while preparing these solutions. Avoid pH differences of greater than ± 0.05 units.

Note: Errors in concentration determination for peptide and/or zinc directly affect the thermodynamic values obtained from the fit.

3.2.4 Data collection

1. Before performing the titration experiment, wash the ITC cell and syringe with water and methanol. To ensure that no methanol is left in the syringe, thoroughly dry it using the automated option on the instrument. After the wash steps, rinse the sample cell and syringe with the buffer solution.

Note: For a more rigorous wash of the system prior to the above step, soak the sample cell with detergent solution and then rinse thoroughly with water. The cell can also be washed with 0.1 M EDTA then rinsed to remove residual zinc (Reddi & Gibney, 2007).

2. Fill the sample cell with the peptide solution, fill the reference cell with 18 M Ω water, and fill the syringe with the zinc solution. See section 3.4 for details on solution compositions. Care must be taken while filling the sample and reference cells to avoid air bubbles.

Note: If the sequence under study is highly susceptible to oxidation, additional experimental modifications may be employed to exclude oxygen during the titration. We employ a plastic adapter that blankets parts of the ITC instrument with inert gas (Figure 4A). Other options include sealing the gap between the sample cell and syringe with parafilm or situating the entire ITC in a glovebox. If the peptide is highly susceptible to oxidation, reducing agents such as tris(2-

carboxyethyl)phosphine or β -mercaptoethanol may be employed; it should be present at equal concentration in the syringe and sample cell.

3. Set following titration parameters in the instrument software:

Total number of injections – 18

Cell temperature – 25 °C

Note: A lower temperature may be used if the zinc finger peptide is unstable in the absence of zinc at room temperature.

Initial delay – 60 s

Reference power – 10 μ cal/sec. *Note*, this assumes an exothermic reaction upon zinc binding; for an endothermic process, set reference power to 1-2 μ cal/sec.

Syringe concentration – 1 mM; cell concentration – 150 μ M

Stirring speed – 750 RPM

Feedback mode/gain – high

4. Under the injection parameters, set the injection volume to 2 μ L and the duration of each injection to 4 s (or to a duration in seconds twice the value of injection volume in μ L). The spacing parameter should be set such that the heat signal returns to baseline between each injection. A setting in the range 150-180 seconds is usually appropriate.

Note: Depending on the concentrations of peptide and zinc in the cell and syringe, respectively, the injection volume and number of injections may need to be adjusted.

Tip: As the first injection is excluded from data analysis, set the injection volume to a small value (0.2-0.4 μ L).

Tip: Before performing the actual titration experiment, it is recommended to perform a water/water and buffer/buffer titration using the above settings. This ensures that the baseline is not noisy/drifts and there is no residual heat from methanol contamination in the syringe or cell.

5. After the above zinc into peptide titration, perform a zinc into buffer control titration. Wash the sample cell with detergent, then with water, and then with buffer solution. Fill the sample cell with the buffer solution and refill the syringe with zinc solution. Perform the titration using the same method parameters that were used for the zinc into peptide titration above.

Note: Poor data can be caused by improper cleaning of cell between injections, buffer mismatch, methanol contamination from washes, and/or air bubbles in the cells

3.2.5 Data analysis

1. After opening the relevant files in the analysis software provided with the instrument (Origin), make sure that the concentration values for each experiment are entered correctly. While the software automatically creates a baseline and integrates the peaks, the user may choose to manually adjust integration settings per peak.
2. Remove the data point corresponding to the first injection.
3. Subtract the data for the control zinc into buffer titration from the zinc into peptide titration using the point-by-point method.

Note: If one or more data points in the zinc into buffer titration deviate significantly from the overall trend, the mean value of control heat injections may be used for the subtraction.

4. Fit the resulting curve to a 1:1 binding model using one set of sites option (Figure 4B). After the curve fitting, n , K , ΔH and ΔS values will be displayed.

Tip: Multiple iterations of the fitting procedure will help in reducing the χ^2 value and in obtaining a good fit. During the fitting procedure, certain parameters like n , K , ΔH can be kept constant or varied.

Note: For a 1:1 binding reaction, the n value should be 1. For an ideal value of $n = 1$, it is assumed that all the peptide present in the sample cell is available for binding. In the case of zinc fingers, the active concentration of peptide may be less than the measured concentration due to oxidation during the titration, resulting in a value of n less than 1. Typical n values obtained in the above protocol are ~ 0.6 - 0.8 .

5. Compile the resulting table of thermodynamic parameters and uncertainties along with graphs of the resulting fits.
6. Repeat entire experiment for series of variants for a given protein of interest to obtain insights into the impact of backbone alteration on folding energetics.

Note: Although ITC experiments provide valuable information, the data should be interpreted with care (Kluska, Adamczyk, & Krężel, 2018). In particular, K values may be underestimated when the interaction is very high affinity ($> 10^8 \text{ M}^{-1}$), as is the case for many zinc fingers binding to zinc. In such instances, differences in measured ΔH values are still reliable and thus differences in folding enthalpy as a function of backbone composition informative.

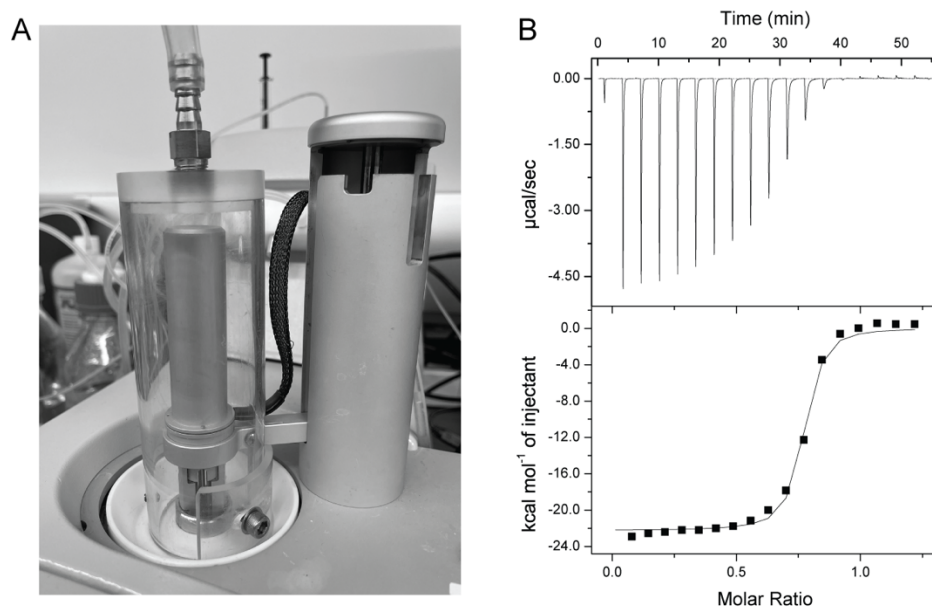


Figure 4. (A) Photograph showing a custom-built plastic adapter that allows the area surrounding the ITC syringe and sample cell to be blanketed with inert gas during the titration. (B) Fitting results from a typical ITC experiment monitoring the binding of a heterogeneous-backbone zinc finger mimetic to Zn^{2+} ; source data from (Rao & Horne, 2020).

4. Summary

In summary, continuing increases in the size and complexity of protein-inspired synthetic chains with artificial backbone composition pose challenges to determination of their folded structure as well as folding thermodynamics. Bringing a combination of classical and cutting-edge methods from structural biology and biophysics to bear on these systems has yielded important insights into the effects of altered backbone on folding. While some experimental challenges arise that are unique to the context of synthetic proteomimetic analytes, a range of prior and ongoing efforts have shown that many techniques applied in the characterization of natural proteins can be successfully adapted to synthetic analogues.

Acknowledgement

Funding for our work involving the development and application of methods for characterizing folded structure and folding thermodynamics in heterogeneous-backbone proteomimetics has been provided by the National Institutes of Health (GM107161) and the National Science Foundation (CHE-1807301).

References

- Adler, A. J., Greenfield, N. J., & Fasman, G. D. (1973). Circular dichroism and optical rotatory dispersion of proteins and polypeptides. *Methods in Enzymology*, 27, 675-735.
- Bang, D., Gribenko, A. V., Tereshko, V., Kossiakoff, A. A., Kent, S. B., & Makhatadze, G. I. (2006). Dissecting the energetics of protein α -helix C-cap termination through chemical protein synthesis. *Nature Chemical Biology*, 2, 139-143.
- Bang, D., Makhatadze, G. I., Tereshko, V., Kossiakoff, A. A., & Kent, S. B. (2005). Total chemical synthesis and X-ray crystal structure of a protein diastereomer: [D-Gln 35]ubiquitin. *Angewandte Chemie International Edition*, 44, 3852-3856.
- Berg, J. M., & Godwin, H. A. (1997). Lessons from zinc-binding peptides. *Annual Review of Biophysics and Biomolecular Structure*, 26, 357-371.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., . . . Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Research*, 28, 235-242.
- Bogetti, A. T., Piston, H. E., Leung, J. M. G., Cabaltea, C. C., Yang, D. T., DeGrave, A. J., . . . Chong, L. T. (2020). A twist in the road less traveled: The AMBER ff15ipq-m force field for protein mimetics. *Journal of Chemical Physics*, 153, 064101.

- Buer, B. C., Levin, B. J., & Marsh, E. N. G. (2012). Influence of fluorination on the thermodynamics of protein folding. *Journal of the American Chemical Society*, *134*, 13027-13034.
- Cabalteja, C. C., & Horne, W. S. (2021). Application of chemical synthesis to engineer protein backbone connectivity. In A. Brik, P. E. Dawson, & L. Liu (Eds.), *Total Chemical Synthesis of Proteins* (pp. 515-532). Weinheim: Wiley-VCH.
- Cabalteja, C. C., Mihalko, D. S., & Horne, W. S. (2019). Heterogeneous-backbone foldamer mimics of a computationally designed, disulfide-rich miniprotein. *ChemBioChem*, *20*, 103-110.
- Checco, J. W., Kreitler, D. F., Thomas, N. C., Belair, D. G., Rettko, N. J., Murphy, W. L., . . . Gellman, S. H. (2015). Targeting diverse protein–protein interaction interfaces with α/β -peptides derived from the Z-domain scaffold. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 4552-4557.
- Cheng, P.-N., Liu, C., Zhao, M., Eisenberg, D., & Nowick, J. S. (2012). Amyloid β -sheet mimics that antagonize protein aggregation and reduce amyloid toxicity. *Nature Chemistry*, *4*, 927-933.
- Collie, G. W., Bailly, R., Pulka-Ziach, K., Lombardo, C. M., Mauran, L., Taib-Maamar, N., . . . Guichard, G. (2017). Molecular recognition within the cavity of a foldamer helix bundle: Encapsulation of primary alcohols in aqueous conditions. *Journal of the American Chemical Society*, *139*, 6128-6137.
- Collie, G. W., Pulka-Ziach, K., Lombardo, C. M., Fremaux, J., Rosu, F., Decossas, M., . . . Guichard, G. (2015). Shaping quaternary assemblies of water-soluble non-peptide helical foldamers by sequence manipulation. *Nature Chemistry*, *7*, 871.

- Dadon, Z., Samiappan, M., Shahar, A., Zarivach, R., & Ashkenasy, G. (2013). A high-resolution structure that provides insight into coiled-coil thiodepsipeptide dynamic chemistry. *Angewandte Chemie International Edition*, 52, 9944-9947.
- Daniels, D. S., Petersson, E. J., Qiu, J. X., & Schepartz, A. (2007). High-resolution structure of a β -peptide bundle. *Journal of the American Chemical Society*, 129, 1532-1533.
- De, S., Chi, B., Granier, T., Qi, T., Maurizot, V., & Huc, I. (2018). Designing cooperatively folded abiotic uni- and multimolecular helix bundles. *Nature Chemistry*, 10, 51-57.
- Debiec, K. T., Cerutti, D. S., Baker, L. R., Gronenborn, A. M., Case, D. A., & Chong, L. T. (2016). Further along the road less traveled: AMBER ff15ipq, an original protein force field built on a self-consistent physical model. *Journal of Chemical Theory and Computation*, 12, 3926-3947.
- Eckhardt, B., Grosse, W., Essen, L.-O., & Geyer, A. (2010). Structural characterization of a β -turn mimic within a protein–protein interface. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 18336-18341.
- Ellman, G. L. (1959). Tissue sulfhydryl groups. *Archives of Biochemistry and Biophysics*, 82, 70-77.
- Fersht, A. (1999). *Structure and Mechanism in Protein Science*. New York: W. H. Freeman and Company.
- Fischer, J., Paschek, D., Geiger, A., & Sadowski, G. (2008). Modeling of aqueous poly(oxyethylene) solutions: 1. Atomistic simulations. *Journal of Physical Chemistry B*, 112, 2388-2398.

- Fuller, A. A., Du, D., Liu, F., Davoren, J. E., Bhabha, G., Kroon, G., . . . Kelly, J. W. (2009). Evaluating β -turn mimics as β -sheet folding nucleators. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 11067-11072.
- George, K. L., & Horne, W. S. (2017). Heterogeneous-backbone foldamer mimics of zinc finger tertiary structure. *Journal of the American Chemical Society*, *139*, 7931-7938.
- George, K. L., & Horne, W. S. (2018). Foldamer tertiary structure through sequence-guided protein backbone alteration. *Accounts of Chemical Research*, *51*, 1220-1228.
- Gill, S. C., & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Analytical Biochemistry*, *182*, 319-326.
- Giuliano, M. W., Horne, W. S., & Gellman, S. H. (2009). An α/β -Peptide helix bundle with a pure β^3 -amino acid core and a distinctive quaternary structure. *Journal of the American Chemical Society*, *131*, 9860-9861.
- Grossoehme, N. E., Spuches, A. M., & Wilcox, D. E. (2010). Application of isothermal titration calorimetry in bioinorganic chemistry. *Journal of Biological Inorganic Chemistry*, *15*, 1183-1191.
- Haney, C. M., Werner, H. M., McKay, J. J., & Horne, W. S. (2016). Thermodynamic origin of α -helix stabilization by side-chain cross-links in a small protein. *Organic & Biomolecular Chemistry*, *14*, 5768-5773.
- Horne, W. S., & Grossmann, T. N. (2020). Proteomimetics as protein-inspired scaffolds with defined tertiary folding patterns. *Nature Chemistry*, *12*, 331-337.
- Horne, W. S., Price, J. L., & Gellman, S. H. (2008). Interplay among side chain sequence, backbone composition, and residue rigidification in polypeptide folding and assembly.

- Proceedings of the National Academy of Sciences of the United States of America*, 105, 9151-9156.
- Horne, W. S., Price, J. L., Keck, J. L., & Gellman, S. H. (2007). Helix bundle quaternary structure from α/β -peptide foldamers. *Journal of the American Chemical Society*, 129, 4178-4180.
- Jean, F., Buisine, E., Melnyk, O., Drobecq, H., Odaert, B., Hugues, M., . . . Tartar, A. (1998). Synthesis and structural and functional evaluation of a protein modified with a β -turn mimic. *Journal of the American Chemical Society*, 120, 6076-6083.
- Kasznel, A. J., Zhang, Y., Hai, Y., & Chenoweth, D. M. (2017). Structural basis for aza-glycine stabilization of collagen. *Journal of the American Chemical Society*, 139, 9427-9430.
- Klug, A. (2010). The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annual Review of Biochemistry*, 79, 213-231.
- Kluska, K., Adamczyk, J., & Krężel, A. (2018). Metal binding properties, stability and reactivity of zinc fingers. *Coordination Chemistry Reviews*, 367, 18-64.
- Kreitler, D. F., Mortenson, D. E., Forest, K. T., & Gellman, S. H. (2016). Effects of single α -to- β residue replacements on structure and stability in a small protein: Insights from quasiracemic crystallization. *Journal of the American Chemical Society*, 138, 6498-6505.
- Kreutzer, A. G., & Nowick, J. S. (2018). Elucidating the structures of amyloid oligomers with macrocyclic β -hairpin peptides: Insights into alzheimer's disease and other amyloid diseases. *Accounts of Chemical Research*, 51, 706-718.
- Kuhlman, B., & Raleigh, D. P. (1998). Global analysis of the thermal and chemical denaturation of the N-terminal domain of the ribosomal protein L9 in H₂O and D₂O. Determination of

- the thermodynamic parameters, ΔH° , ΔS° , and ΔC°_p , and evaluation of solvent isotope effects. *Protein Science*, 7, 2405-2412.
- Liebschner, D., Afonine, P. V., Baker, M. L., Bunkoczi, G., Chen, V. B., Croll, T. I., . . . Adams, P. D. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallographica Section D*, 75, 861-877.
- Lombardo, C. M., Kumar M. V., V., Douat, C., Rosu, F., Mergny, J.-L., Salgado, G. F., & Guichard, G. (2019). Design and structure determination of a composite zinc finger containing a nonpeptide foldamer helical domain. *Journal of the American Chemical Society*, 141, 2516-2525.
- Lopes, P. E. M., Guvench, O., & MacKerell, A. D. (2015). Current status of protein force fields for molecular dynamics simulations. In A. Kukol (Ed.), *Molecular Modeling of Proteins* (pp. 47-71). New York, NY: Springer New York.
- Lu, W., Qasim, M. A., Laskowski, M., & Kent, S. B. H. (1997). Probing intermolecular main chain hydrogen bonding in serine proteinase–protein inhibitor complexes: Chemical synthesis of backbone-engineered turkey ovomucoid third domain. *Biochemistry*, 36, 673-679.
- Mareuil, F., Malliavin, T. E., Nilges, M., & Bardiaux, B. (2015). Improved reliability, accuracy and quality in automated NMR structure calculation with ARIA. *Journal of Biomolecular NMR*, 62, 425-438.
- Mirijanian, D. T., Mannige, R. V., Zuckermann, R. N., & Whitlam, S. (2014). Development and use of an atomistic CHARMM-based forcefield for peptoid simulation. *Journal of Computational Chemistry*, 35, 360-370.

- Mong, S. K., Cochran, F. V., Yu, H., Graziano, Z., Lin, Y.-S., Cochran, J. R., & Pentelute, B. L. (2017). Heterochiral knottin protein: Folding and solution structure. *Biochemistry*, *56*, 5720-5725.
- Mortenson, D. E., Kreitler, D. F., Thomas, N. C., Guzei, I. A., Gellman, S. H., & Forest, K. T. (2018). Evaluation of β -amino acid replacements in protein loops: Effects on conformational stability and structure. *ChemBioChem*, *19*, 604-612.
- Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., . . . Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D*, *67*, 355-367.
- Odaert, B., Jean, F., Melnyk, O., Tartar, A., Lippens, G., Boutillon, C., & Buisine, E. (1999). Synthesis, folding, and structure of the β -turn mimic modified B1 domain of streptococcal protein G. *Protein Science*, *8*, 2773-2783.
- Pelay-Gimeno, M., Glas, A., Koch, O., & Grossmann, T. N. (2015). Structure-based design of inhibitors of protein-protein interactions: Mimicking peptide binding epitopes. *Angewandte Chemie-International Edition*, *54*, 8896-8927.
- Price, J. L., Horne, W. S., & Gellman, S. H. (2010). Structural consequences of β -amino acid preorganization in a self-assembling α/β -peptide: Fundamental studies of foldameric helix bundles. *Journal of the American Chemical Society*, *132*, 12378-12387.
- Rao, S. R., & Horne, W. S. (2020). Proteomimetic zinc finger domains with modified metal-binding β -turns. *Peptide Science*, *112*, e24177.
- Reddi, A. R., & Gibney, B. R. (2007). Role of protons in the thermodynamic contribution of a Zn(II)-Cys₄ site toward metalloprotein stability. *Biochemistry*, *46*, 3745-3758.

- Reinert, Z. E., & Horne, W. S. (2014). Folding thermodynamics of protein-like oligomers with heterogeneous backbones. *Chemical Science*, 5, 3325-3330.
- Reinert, Z. E., Lengyel, G. A., & Horne, W. S. (2013). Protein-like tertiary folding behavior from heterogeneous backbones. *Journal of the American Chemical Society*, 135, 12528-12531.
- Reinert, Z. E., Musselman, E. D., Elcock, A. H., & Horne, W. S. (2012). A PEG-based oligomer as a backbone replacement for surface-exposed loops in a protein tertiary structure. *ChemBioChem*, 13, 1107-1111.
- Rieping, W., Habeck, M., Bardiaux, B., Bernard, A., Malliavin, T. E., & Nilges, M. (2006). ARIA2: Automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics*, 23, 381-382.
- Rodriguez-Granillo, A., Annavarapu, S., Zhang, L., Koder, R. L., & Nanda, V. (2011). Computational design of thermostabilizing D-amino acid substitutions. *Journal of the American Chemical Society*, 133, 18750-18759.
- Säbel, C. E., Neureuther, J. M., & Siemann, S. (2010). A spectrophotometric method for the determination of zinc, copper, and cobalt ions in metalloproteins using Zincon. *Analytical Biochemistry*, 397, 218-226.
- Schnolzer, M., & Kent, S. (1992). Constructing proteins by dovetailing unprotected synthetic peptides: Backbone-engineered HIV protease. *Science*, 256, 221-225.
- Simon, M. D., Maki, Y., Vinogradov, A. A., Zhang, C., Yu, H., Lin, Y.-S., . . . Pentelute, B. L. (2016). D-Amino acid scan of two small proteins. *Journal of the American Chemical Society*, 138, 12099-12111.

- Spencer, R. K., Li, H., & Nowick, J. S. (2014). X-Ray crystallographic structures of trimers and higher-order oligomeric assemblies of a peptide derived from A β ₁₇₋₃₆. *Journal of the American Chemical Society*, 136, 5595-5598.
- Tavenor, N. A., Reinert, Z. E., Lengyel, G. A., Griffith, B. D., & Horne, W. S. (2016). Comparison of design strategies for α -helix backbone modification in a protein tertiary fold. *Chemical Communications*, 52, 3789-3792.
- Torbeev, V. Y., Raghuraman, H., Hamelberg, D., Tonelli, M., Westler, W. M., Perozo, E., & Kent, S. B. H. (2011). Protein conformational dynamics in the mechanism of HIV-1 protease catalysis. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 20982-20987.
- Valiyaveetil, F. I., Leonetti, M., Muir, T. W., & MacKinnon, R. (2006). Ion selectivity in a semisynthetic K⁺ channel locked in the conductive conformation. *Science*, 314, 1004-1007.
- Valiyaveetil, F. I., Sekedat, M., MacKinnon, R., & Muir, T. W. (2006). Structural and functional consequences of an amide-to-ester substitution in the selectivity filter of a potassium channel. *Journal of the American Chemical Society*, 128, 11591-11599.
- Vanommeslaeghe, K., & MacKerell, A. D. (2015). *A CHARMM compatible force field for N- and alpha-methylated peptides with (phi,psi) energy correction grid*. Paper presented at the Biophysical Society 59th Annual Meeting.
- Viles, J. H., Patel, S. U., Mitchell, J. B. O., Moody, C. M., Justice, D. E., Uppenbrink, J., . . . Thornton, J. M. (1998). Design, synthesis and structure of a zinc finger with an artificial β -turn. *Journal of Molecular Biology*, 279, 973-986.

- Walters, C. R., Szantai-Kis, D. M., Zhang, Y., Reinert, Z. E., Horne, W. S., Chenoweth, D. M., & Petersson, E. J. (2017). The effects of thioamide backbone substitution on protein stability: a study in α -helical, β -sheet, and polyproline II helical contexts. *Chemical Science*, 8, 2868-2877.
- Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., . . . Richardson, D. C. (2018). MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science*, 27, 293-315.
- Wiseman, T., Williston, S., Brandts, J. F., & Lin, L.-N. (1989). Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Analytical Biochemistry*, 179, 131-137.
- Yang, X., Wang, M., & Fitzgerald, M. C. (2004). Analysis of protein folding and function using backbone modified proteins. *Bioorganic Chemistry*, 32, 438-449.
- Yeates, T. O., & Kent, S. B. H. (2012). Racemic protein crystallography. *Annual Review of Biophysics*, 41, 41-61.