



2018-07-01

Measuring the Interaction and Cooperativity Between Ionic, Aromatic, and Nonpolar Amino Acids in Protein Structure

Mason Scott Smith
Brigham Young University

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>

 Part of the [Physical Sciences and Mathematics Commons](#)

BYU ScholarsArchive Citation

Smith, Mason Scott, "Measuring the Interaction and Cooperativity Between Ionic, Aromatic, and Nonpolar Amino Acids in Protein Structure" (2018). *All Theses and Dissertations*. 7443.
<https://scholarsarchive.byu.edu/etd/7443>

This Dissertation is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Measuring the Interactions and Cooperativity Between Ionic, Aromatic, and Nonpolar Amino
Acids in Protein Structure

Mason Scott Smith

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Joshua L. Price, Chair
Barry M. Willardson
Steven W. Graves
Joshua L. Anderson
David J. Michealis

Department of Chemistry and Biochemistry
Brigham Young University

Copyright © 2018 Mason Scott Smith

All Rights Reserved

ABSTRACT

Measuring the Interactions and Cooperativity Between Ionic, Aromatic, and Nonpolar Amino Acids in Protein Structure

Mason Scott Smith
Department of Chemistry and Biochemistry, BYU
Doctor of Philosophy

Protein folding studies have provided important insights about the key role of non-covalent interactions in protein structure and conformational stability. Some of these interactions include salt bridges, cation- π , and anion- π interactions. Understanding these interactions is crucial to developing methods for predicting protein secondary, tertiary, quaternary structure from primary sequence and understanding protein-protein interactions and protein-ligand interactions. Several studies have described how the interaction between two amino acid side chains have a substantial effect on protein structure and conformational stability. This is under the assumption that the interaction between the two amino acids is independent of surrounding interactions. We are interested in understanding how salt bridges, cation- π , and anion- π interactions affect each other when they are in close proximity.

Chapter 1 is a brief introduction on noncovalent interactions and noncovalent interaction cooperativity. Chapter 2 describes the progress we have made measuring the cooperativity between noncovalent interactions involving cations, anions and aromatic amino acids in a coiled-coil alpha helix model protein. Chapter 3 describes cooperativity between cation, anion, and nonaromatic hydrophobic amino acids in the context of a coiled-coil alpha helix. In chapter 4 we describe a strong anion- π interaction in a reverse turn that stabilizes a beta sheet model protein. In chapter 5 we measure the interaction between a cysteine linked maleimide and two lysines in a helix and show that it is a general strategy to stabilize helical structure.

Key words: cation- π , anion- π , noncovalent interaction, coiled coil, protein structure, conformational stability

ACKNOWLEDGEMENTS

First, I would like to give the greatest thanks to my advisor and friend Joshua Price. I consider it an honor to have been trained in his lab. I appreciate his advice, ideas, and mentorship without which the work presented in this thesis would not be possible.

Thanks to my wife and son, Trisha and Elliott for their support. My wife has spent many hours keeping me company while I run protein melts and mount protein crystals. Thanks to Paul Lawrence and Brijesh Pandey, who, among other things, were good friends and mentors in the beginning of graduate school. Thanks for Steven Draper and Qiang Xiao for being good friends and patient with me at the end graduate school. Thanks to Wendy, Kimberlee, Katie, Natalie, and Mackenzie who have helped me with synthesis, purification, and characterization and the analysis of the many peptides used in my research. This dissertation would also not be possible without them. Thanks to Frank Whitby for his mentorship in protein crystallography and his friendship.

TABLE OF CONTENTS

ABSTRACT.....	ii
1 Introduction.....	1
1.1 Noncovalent interaction cooperativity.....	1
1.1.1 Cation- π , anion- π , and π - π interactions	2
1.1.2 Cooperativity between arenes and ions.....	7
1.2 Protein Folding Cooperativity.....	10
1.2.1 Measuring non-covalent interactions in proteins.....	11
1.2.2 Measuring ionic, aromatic, nonpolar cooperativity in protein structure.....	14
1.3 References.....	14
2 Measuring A helical Salt bridge enhanced by aromatic amino acids.	22
2.1 Introduction.....	22
2.2 Results and discussion	23
2.2.1 Cooperativity between a long-range salt bridges mediated by Tryptophan.	23
2.2.2 The structural influence of the arene on the long-range salt bridge.	27
2.2.3 Crystal structures of 1CW variants.....	29
2.2.4 NMR model system	35
2.2.5 Exploring the Long-Range Salt-bridge within a dimeric helix bundle.....	38
2.3 Conclusions.....	40
2.4 Supporting Information.....	41
2.4.1 Protein Synthesis, Purification, and Characterization.....	41
2.4.2 ESI-TOF data	45
2.4.3 Analytical HPLC data.....	74
2.4.4 Size exclusion chromatography	108
2.4.5 Folded Free energy measurement, Circular Dichroism spectropolarimetry.....	113
2.4.6 Single crystal X-ray Diffraction Data.	166
2.4.7 Justification for using Ser at position 14 instead of Ala as a control for evaluating Trp-based enhancement of the Glu10-Lys18 Interaction.	170
2.4.8 Impact of Salt on Phe based enhancement of Glu10:Lys18 salt bridge	172
2.4.9 NMR Tocsy and Roesy data	173
2.5 References.....	185

3	Measuring A helical long-range Salt bridge enhanced by nonaromatic hydrophobic amino acids.	188
3.1	Introduction.....	188
3.2	Results and discussion	190
3.2.1	Long range salt-bridge enhanced by nonpolar amino acids.....	190
3.2.2	Crystal structures	192
3.2.3	Conclusions.....	195
3.3	Supporting Information.....	196
3.3.1	Protein synthesis, characterization, and Purification.	196
3.3.2	ESI-TOF mass spec data.....	197
3.3.3	Analytical HPLC data.....	207
3.3.4	Size Exclusion Chromatography.....	220
3.3.5	Folded Free energy measurement, Circular Dichroism spectropolarimetry.	223
3.3.6	Single crystal diffraction data.	234
3.4	References.....	239
4	An Anion- π Interaction Strongly Stabilizes the β -Sheet Protein WW	241
4.1	Introduction.....	241
4.2	Results and discussion	242
4.3	Conclusions.....	246
4.4	Supporting information.....	246
4.4.1	Protein Synthesis.....	246
4.4.2	Protein Purification and Characterization.....	246
4.4.3	Circular Dichroism Spectropolarimetry.....	255
4.5	References.....	276
5	The Cys _i -Lys _{i+3} -Lys _{i+4} Triad: A General Approach for PEG-based Stabilization of α -helical Proteins.	279
5.1	Introduction.....	279
5.2	Results and Discussion	280
5.3	Conclusions.....	288
5.4	Contributions.....	289
5.5	Supporting Information.....	289

5.5.1	Protein Synthesis.....	289
5.5.2	Protein Purification and Characterization.....	290
5.5.3	Variable Temperature CD experiments	310
5.5.4	Sedimentation Equilibrium experiments.....	317
5.5.5	Triple Mutant Cycle Analysis.....	335
5.6	References.....	337

1 INTRODUCTION

1.1 Noncovalent interaction cooperativity

Covalent bonds are strong interactions between atoms where the electrons are shared between atoms ($10-100 \text{ kcal mol}^{-1}$).¹ When two atoms form a covalent bond but one atom is more electronegative than the other the two atoms do not share the electrons equally, consequently leaving one atom electron rich and the other electron poor. This results in the tendency for molecules to have weaker noncovalent interactions with other molecules. Noncovalent interactions do not share electrons but are governed by a dispersive electronic interactions between atoms or molecules where electron rich atoms or molecules interact with electron poor atoms or molecules.²⁻³ Generally these interactions are weaker than covalent bonds ($0-5 \text{ kcal mol}^{-1}$), but are important for the formation of supramolecular structure found in biological macromolecules such as DNA, RNA, proteins, carbohydrates and lipids.⁴ Noncovalent interactions include hydrogen bonding, cation- π , anion- π , London dispersion, π - π , CH- π , hydrophobic along with several others.

Hydrogen bonding is an interaction between an electron poor hydrogen, bound to an electronegative atom, interacting with a lone pair of an adjacent atom. Now consider a water-water hydrogen bond ($3-8 \text{ kcal mol}^{-1}$) which when in solution with other waters form dimers, trimer, or even a dodecamer hydrogen bonding networks.⁵ As more molecules associate with the hydrogen bonding network each individual hydrogen bond becomes stronger energetically almost

doubling once ten waters are in the bonding network.⁶⁻⁸ This shows that the energy found in a hydrogen bonding network is not additive, or when 1 interaction + 1 interaction \neq sum of the interactions.

When multiple noncovalent interactions are non-additive this is due to cooperativity or non-cooperativity. When two noncovalent interactions strengthen each other in a non-additive way we define that energy as cooperative, and when two interactions weaken each other it is anti-cooperative. Cooperativity has historically been studied in hydrogen bonding systems but there are fewer experimental measurements of cooperativity in ionic, aromatic and nonpolar systems. Most commonly found are quantum computational methods that predict cooperativity between these interactions.

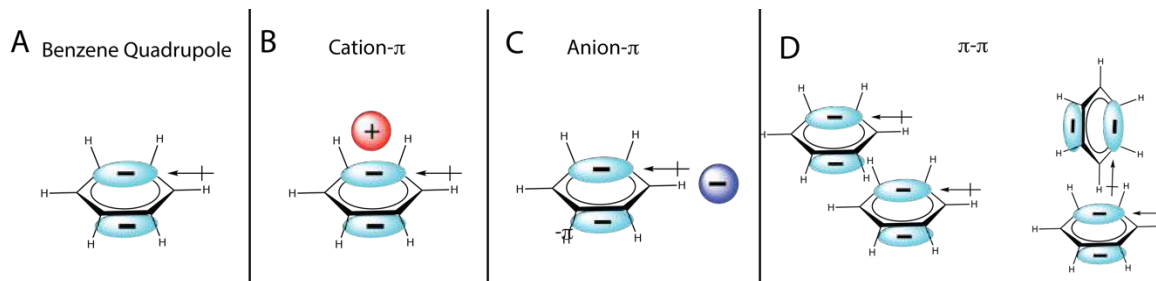


Figure 1-1. schematic of arene ion and arene-arene interactions. (A) electronic configuration of benzene with electron density above and below the ring and positively dipoled edges of the ring. (B) cation- π interaction. (C) anion- π interaction. (D) π - π interactions with both the parallel stacking and “t-shaped” configurations.

1.1.1 Cation- π , anion- π , and π - π interactions

Electron-rich arenes like benzene have large amounts of electron density in the center of the ring due to the electronegative sp^2 hybridized carbons, leaving the less electronegative hydrogens on the ring edges electron poor, this electronic configuration is referred to as a quadrupole (Figure 1-1A). The large electron density in the center of the ring is able to form a

strong interaction with cations, where the cation is positioned above or below the electron density in the ring (Figure 1-1B). This is called a cation- π interaction which has been extensively reviewed elsewhere.⁹⁻¹⁴ The edges of electron rich rings are electron poor and able to interact with anions (Figure 1-1C).¹⁵⁻¹⁸ π - π interactions form when two arene quadrupoles interact with each other forming either “offset stacked” or “edge to face” orientations. Both orientations position the electron rich portion of the arene over the electron poor portion (Figure 1-1D).¹⁹⁻²¹

Almost forty years ago, gas phase studies demonstrated the binding of a potassium cation with benzene is stronger than the binding of a potassium cation with a water molecule.²² This result suggested that cation- π interactions could be energetically important in aqueous solution. Computational analysis predict that the interaction is stronger when the cation is positioned perpendicular to the plane of the ring, and it has a distance dependence relative to R^{-n} where R is the distance between the cation and the center of the arene and n is less than 2. This indicates that a cation can be moved from its optimal position without significant cost to the strength of the interaction, with a distance limit being 6-7 angstroms.^{14, 23} A probe of 600 nonhomologous proteins show that there are over 3,000 energetically significant cation- π interactions.⁹ The study indicates that there is on average 1 cation- π for every 77 amino acids in the PDB. Figure 1-2 highlights some of the cation- π examples from this PDB screen. First, an aromatic box in glucoamylase where a Lys108 is surrounded by four aromatic amino acids. Second, the cation- π ladder in the human growth hormone where there is an alternating pattern of aromatic and positively charged amino acids along two strands of a beta sheet (Figure 1-2A, 1-2B)^{9, 24-28}

Subsequent biological studies show that several protein-protein, protein-DNA, and protein-ligand interactions are mediated through cation- π interactions.²⁹⁻³³ Figure 1-3C Highlights the VWF protein domain binding to its target gene which has a conserved cation π

interaction between Arg 632 and an adenine, misregulation with VWF-DNA protein complex has been associated with a blood clotting disorder called Von Willebrand disease.³⁴ Histone effector proteins, which help regulate gene activation, bind histone methylated lysines. The PHD domain, a common trimethyl lysine effector protein, has two conserved aromatics at positions 215 and 238 which are critical to the protein-histone interactions and specificity (Figure 1-2D). Acetylcholine receptors, important for proper neural function, binds its positively charged ligands with aromatic amino acids (Figure 1-3E).^{14, 35-39} Figure 1-2F shows Trp123 of lysozyme binding to a sodium cation, an example of a protein side chain binding to a metal cation.⁴⁰⁻⁴³

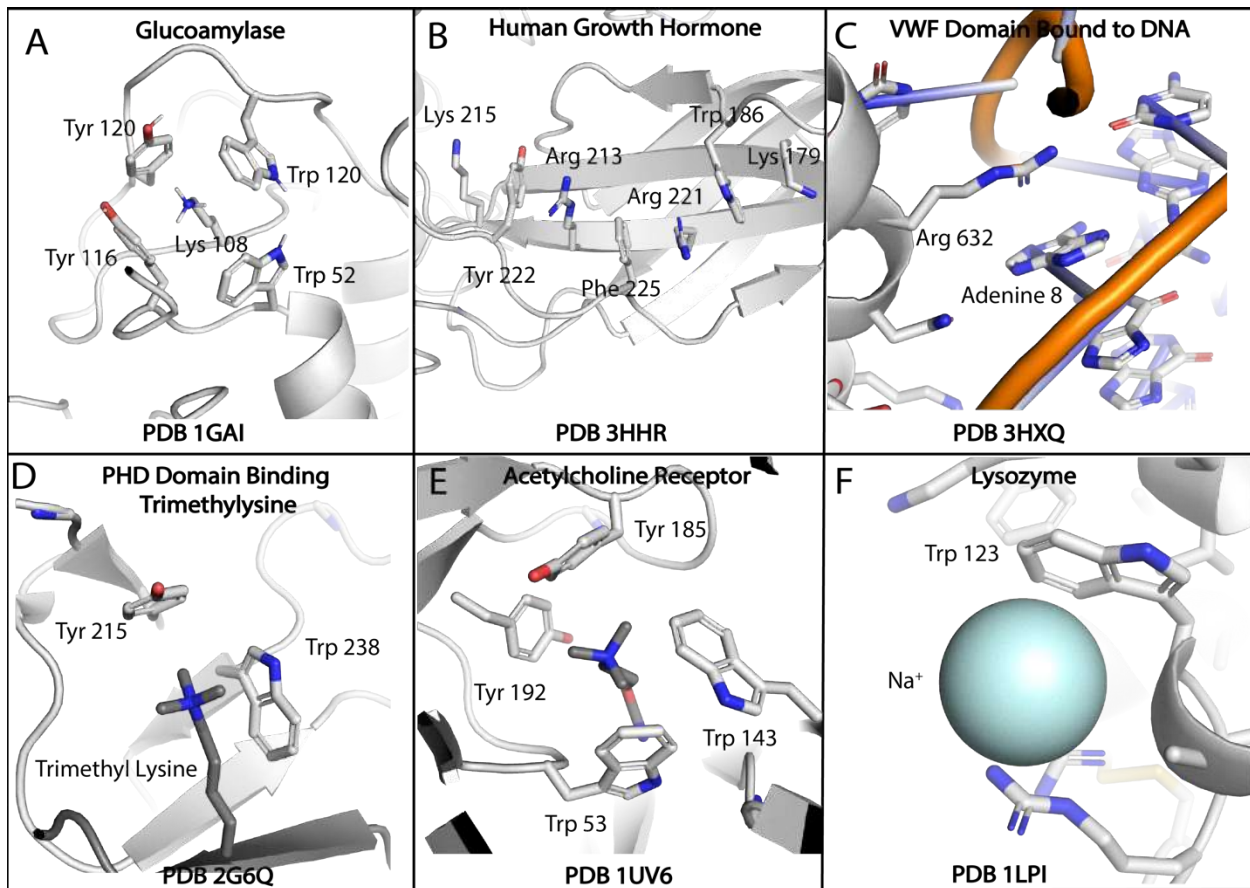


Figure 1-2. Examples of cation- π interactions found the PBD (A) The aromatic box of glucoamylase with a cation π interaction found in between lys108 and Tyr 120, Tyr116, Trp52, and Trp120. (B) the cation- π ladder in human growth hormone. (C) Cation- π at the interface of the VWF domain and it's DNA binding partner. (D) PHD domain interacting with trimethyllysine on Histone 3. (E) Acetylcholine

receptor mediation by Tyr185, Tyr192, Trp54, and Trp143 all participating in a cation π with the trimethylammonium of acetylcholine. (F) Cation π between a sodium atom and Trp 123.

Unlike cation- π interactions, anion- π interactions are believed to be unfavorable when positioned at the face of an electron rich arene.^{18, 44-45} Anions are positioned close to the positively dipoled edges of electron rich arenes.⁴⁶ In biology, probes for anion- π interactions in the PDB show that Glu and Asp side chains tend to be oriented to the edges of Phe, Tyr, Trp.⁴⁶⁻⁴⁷ An example can be found in the enzyme keto steroid isomerase where a conserved Glu 38 forms a edgewise anion- π interaction with both Phe 116, and Phe 54 (Figure1-4A).⁴⁷⁻⁴⁸ Unsurprisingly, changing the electronic character of benzene to hexafluorobenzene shows that the electron poor face can tolerate a facewise interaction with an anion.^{44, 49-50} Other ion-arene interaction studies indicate that modifications in the electronic character of an arene can change its facewise preference from a cation to an anion.⁵¹⁻⁵⁴ Interestingly Chakravarty et. al. searched the PDB for face-wise anion- π interactions both in protein structure and at protein-DNA interfaces.⁵⁵ They probed 1500 non-homologous structures and searched for anions that were less 4.5 angstroms from the center of the aromatic ring and were within 15° perpendicular to the plane of the ring. Although less common than edgewise anion- π interactions, they unambiguously found face-wise interactions in the pdb and showed that the interactions were conserved within protein families suggesting importance to protein folding or function. A facewise interaction found in the enzyme franketolase, which catalyzes the formation of glyceraldehyde-3-phosphate, shows Glu102 directly above Tyr105 (Figure 1-3B). Additionally, anion- π interactions have been shown useful for engineering active sites for enzymes.⁵⁶ Stabilizing the conformation of proteins and beta sheet peptides,⁵⁷⁻⁵⁹ and significant anion- π interactions at the interface of proteins and DNA suggesting an importance for DNA regulation and transcription.^{17, 60-61}

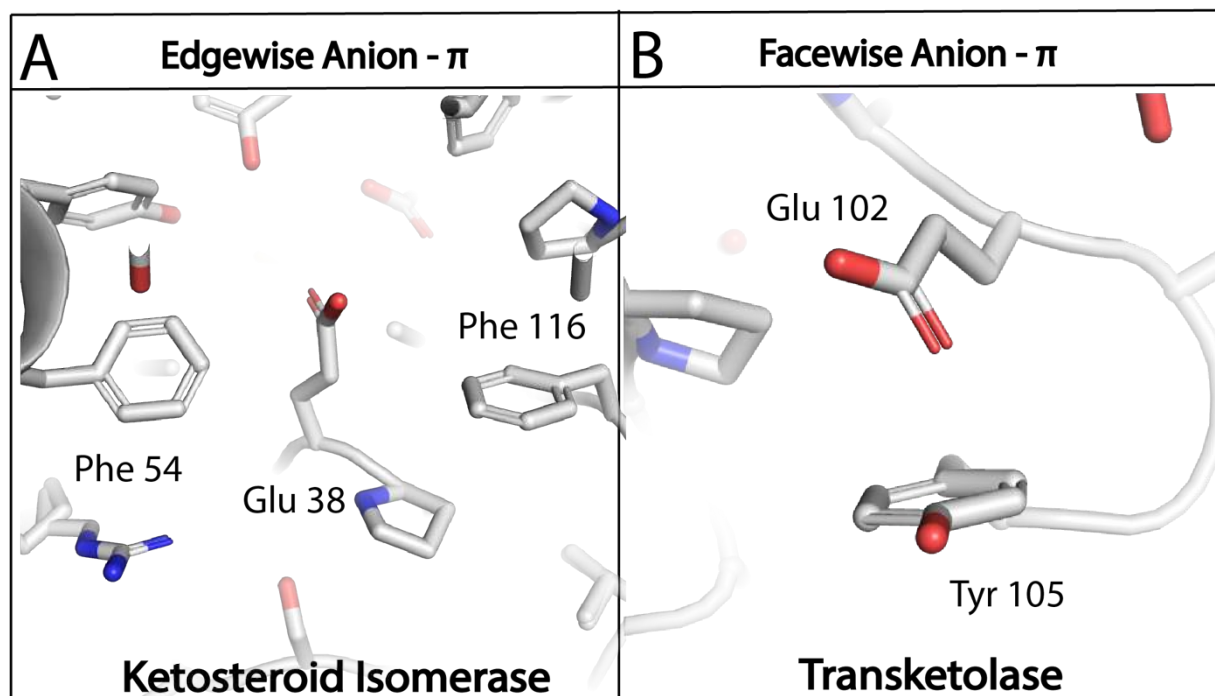


Figure 1-3. Anion π interactions. (A) Anion π interactions found in Ketosteroid Isomerase, Glu 38 forms an edgewise anion π with Phe 116, and Phe 54. (B) Facewise anion π interactions in Transketolase between Glu 102 and Tyr 105.

π - π interactions are usually weaker than ion-arene interactions but have still been used to stabilize coiled-coil peptides, specify protein oligermization, and stabilize protein conformation (Figure 1-4).⁶²⁻⁶⁶ The rings can form the “offset stacked” orientation where the planes of the rings are parallel and the center of one ring is positioned over the edge of another ring. Two arenes can also form a “edge to face” orientation where the positive edge of one ring forms a favorable interaction with the center of the second ring. π - π stacking has been used to engineer an interesting pentameric coiled-coil where the interior of the coil is a tight packing of Trp residues (Figure 1-4A).⁶⁵ Other studies show that this pentamer can tolerate other aromatic amino acids and has shed light on how tetrameric, pentameric, and hexameric coiled coils form.⁶⁷⁻⁶⁸ A de novo helix-loop-helix dimer (alpha 2 delta) designed by Degrado et al. has a π π interaction between Phe29, and

Phe 10 that contributes substantially to the dimer association state, mutation of either Phe causes the dimer to not fold (Figure 1-4B).⁶⁹⁻⁷⁰ In DNA and RNA structure π - π stacking contributes to the supramolecular structure mainly when the aromatic components of the nucleic acids stack on top of each other (Figure 1-4C).⁷¹⁻⁷² π - π interactions are also critical to protein tertiary structure, and protein-ligand interactions.^{66, 69, 73-75}

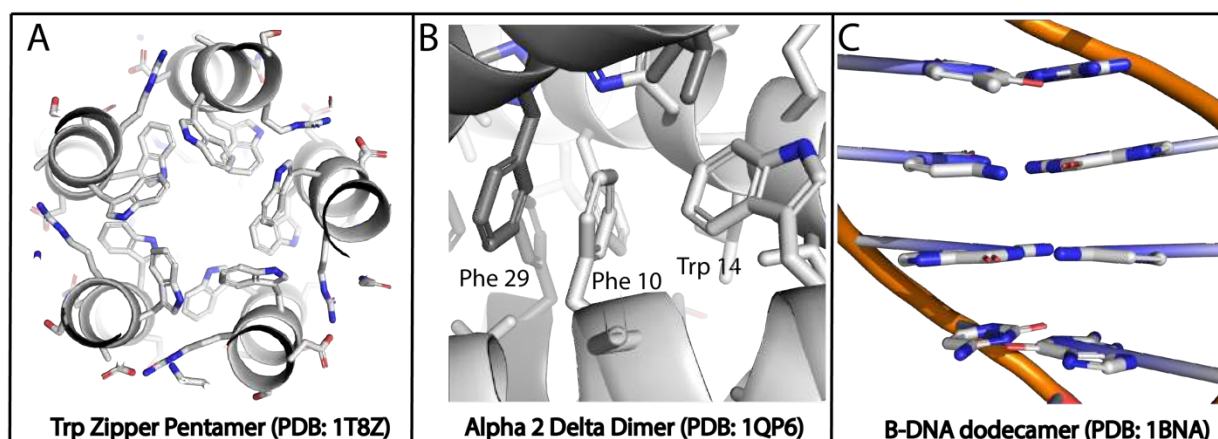


Figure 1-4. (A) Trp π - π interactions in the core of a pentameric coiled coil. (B) Helical homodimer mediated by a Phe29, Phe10, Trp 14 π - π interaction. (C)

1.1.2 Cooperativity between arenes and ions.

Cation- π interactions have been computationally shown to affect π - π , anion- π interactions. Cation- π interactions have been shown to enhance π - π stacking interactions by 5-fold.⁷⁶⁻⁷⁷ Included in these studies was a PDB search of protein structures showing that metal- π - π interactions occur just as frequently as do metal- π interactions. They indicated that a T-shaped π - π orientation was preferred where the metal bound the face of one ring with another ring edge binding the opposite face. Anion- π interactions also can have cooperative interactions with π - π interactions where the arenes are electron deficient.^{49, 78} Interestingly there appears to be a four-way cooperative effect when a cation was positioned on one face of a stacked π - π interaction and

an anion positioned on the other face.⁵⁰ These studies indicate that ions have extensive cooperative effects with arenes.

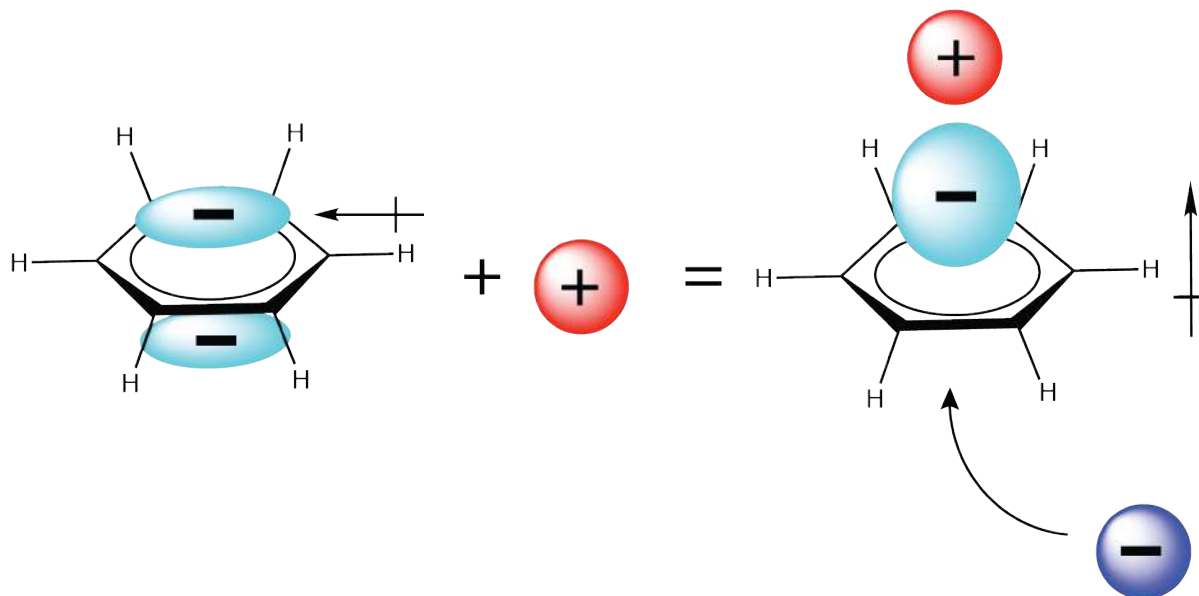


Figure 1-5. Schematic of the cooperativity predicted for a cation- π -anion interaction

Cation- π interactions can have an interesting effect on anion- π interactions. It is believed that face-wise anion- π interactions with electron rich arenes should be destabilizing but computational and supramolecular work show that facewise anion- π interactions are possible if the electronics of the ring are changed. Interesting studies show if a cation is positioned on the opposing face of the ring then the ring electron density is polarized towards the cation leaving the opposing face electron poor (figure 1-5).⁷⁹⁻⁸¹ Dougherty et al. also show that synthetic aromatic host systems that bind cations are more effective if anionic carboxylates are engineered on the opposing face of the aromatic rings. They confirmed these results by computationally modelling a chlorine atom on one face of a benzene and showed that the benzene indeed was more predisposed to interact with a sodium cation on the opposite face.⁸¹ Binding of anionic guests

within the internal arene-lined cavities of calixarenes, cryptophanes, and cyclotrimeratrylenes increases substantially when the exterior surfaces of these hosts are bound by transition metal cations.⁸²⁻⁸⁶ A more recent bioinformatics probe by of the PDB protein structures showed that about 3% of anion- π interactions show that the aromatic side chains is also participating in a cation- π interaction.⁴⁸ The same paper describes how the anion- π -cation interaction is prevalent at protein-protein interfaces, this seems reasonable due to polar amino acids being very prevalent at the surface of proteins instead of the interior. Figure 1-6 highlights six examples of cation- π -anion motifs found in protein structure and between protein-protein interfaces.

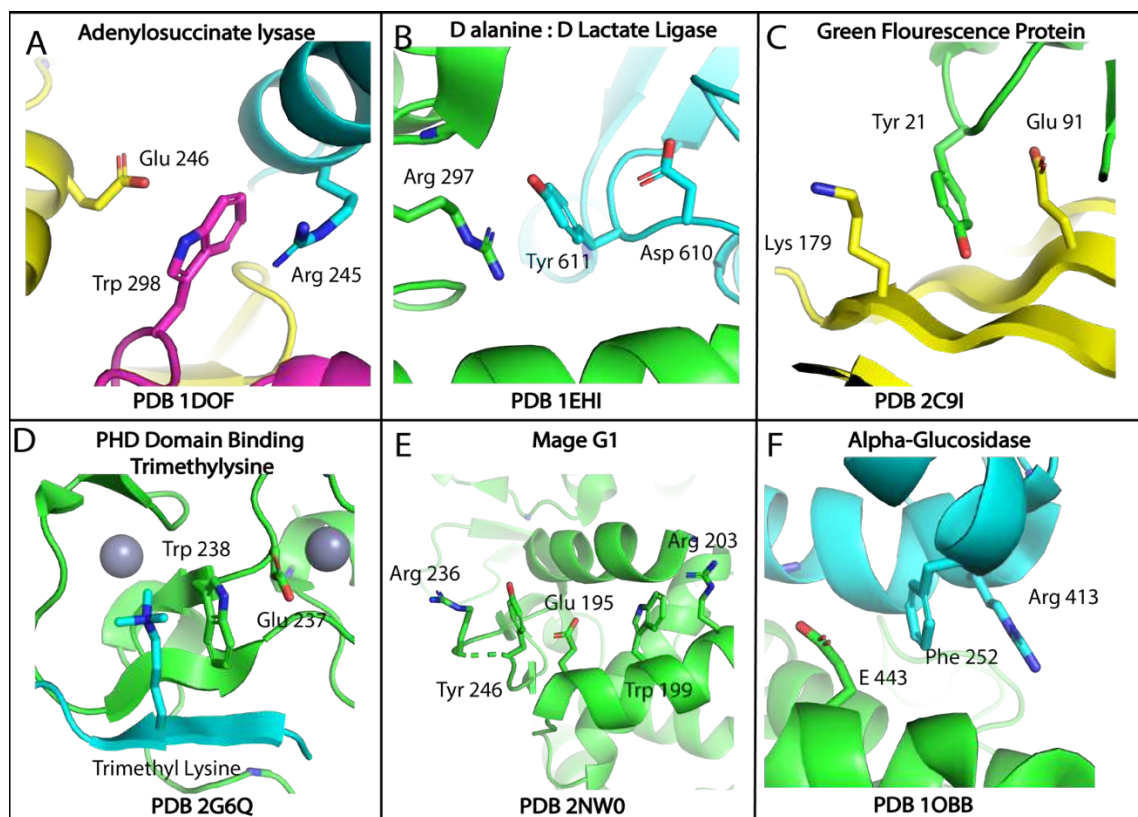


Figure 1-6. Examples described in a bioinformatics probe for cation π anion interactions in protein structure.⁴⁸ Differing colors represent different protein chains within the crystal structure. The six examples include (A) adenylosuccinate lysase, (B) D:alanine :Dlactate ligase, (C) Green fluorescence protein tetramer, (D) PHD domain binding to H3K4 histone peptide, (E) MAGE G1 protein, and (F) Alpha-Glucosidase.

1.2 Protein Folding Cooperativity

Proteins are polymers of amino acids that form complex structural conformations. The structure of the protein governs the protein function and biological purpose, so much so that several disease states are associated with protein structural misfolding.⁸⁷⁻⁹² Protein structure is mostly governed by noncovalent interactions with the exception of covalently linked disulfide bonds, because of this some proteins are only marginally stable at room temperature. Due to the low energy barrier between a folded protein and its unfolded state no type of molecular interaction within the protein is unimportant.

Two main determinants in protein folding are hydrophobic amino acids packing in the interior of the protein isolated from solvent and hydrogen bonding.⁹³⁻⁹⁴ Hydrophobic packing stabilizes proteins in two ways: Hydrophobic amino acids form London dispersion forces, and isolating the hydrophobic amino acids from water solvent allows water to form a hydrogen bonding network where water has more entropy than when bound to nonpolar amino acids. The two main secondary structures of proteins are alpha helices and beta sheets, these two structures form extensive hydrogen bonding networks between the main chain carbonyls and amides giving the secondary structure stability and direction.⁹⁴⁻⁹⁶

Some labs have proposed that protein folding is mostly cooperative in nature.⁹⁷⁻¹⁰³ Dill et al. propose a hydrophobic “zipper” hypothesis where initial hydrophobic contacts bring additional contacts in close proximity further contributing to the hydrophobic cooperativity of a protein core and allowing water to form a cooperative network in the solvent.^{93, 101, 103} Baker et al. suggest that the observed cooperativity on protein folding is a result of natural selection and supports this conclusion by creating de novo protein folds, or folds that are not observed in nature and therefore

cannot be subject to natural selection. They suggest that biology over the years has selected pathways for proteins that fold cooperatively.^{102, 104} A recent experimental study showed that cooperativity between already folded domains contribute to the folded tertiary structure of a protein.⁹⁹

Although cooperativity is established in protein folding, there are limited examples of experimental measurements of cooperativity between specific amino acids side chains. As described in section 1.1.2 several labs predict extensive cooperativity between cations, anions, and arenes. To our knowledge experimental measurements of the cooperativity between anions, cations, and aromatics have not been measured in the context of protein structure. This dissertation describes the first experimental measurements of the cooperative interaction between negatively and positively charged amino acids with aromatic amino acids.

1.2.1 Measuring non-covalent interactions in proteins

Specific noncovalent interactions, in addition to hydrogen bonds and hydrophobic interactions, are critical to protein function and structure. These interactions include cation- π , anion- π , π - π , and n- π^* , along with other interactions. Understanding these interactions is crucial to developing methods for predicting protein secondary, tertiary, quaternary structure from primary sequence, especially for proteins with no known function and with little homology with known proteins.¹⁰⁵⁻¹⁰⁸

To measure the contributions of noncovalent interactions to protein conformational stability, labs have developed double mutant cycle, and triple mutant cycle analysis.^{2, 109-112} This approach has led to insights into the impact of non-covalent interactions on protein folding and stability. A great example of a double mutant cycle was done by Ferst et. al where they measured a salt bridge interaction between an arginine at position 69 and an aspartate at position

93 found in the protein barnase (Figure 1-7A).¹¹³⁻¹¹⁴ To measure the interaction Ferst et. al. made four barnase variants, wild-type, R69S, D93N, and R69S D93N double mutant. This group folded and unfolded each barnase variant using chemical denaturants and monitored the folding process using a stop-flow fluorimeter. They extrapolated folded free energies for each variant and calculated the interaction between R69 and D93 (Figure 1-7B). The partially buried salt bridge was estimated to contribute around 3 kcal mol⁻¹ of energy to the folded conformation of the wildtype barnase.

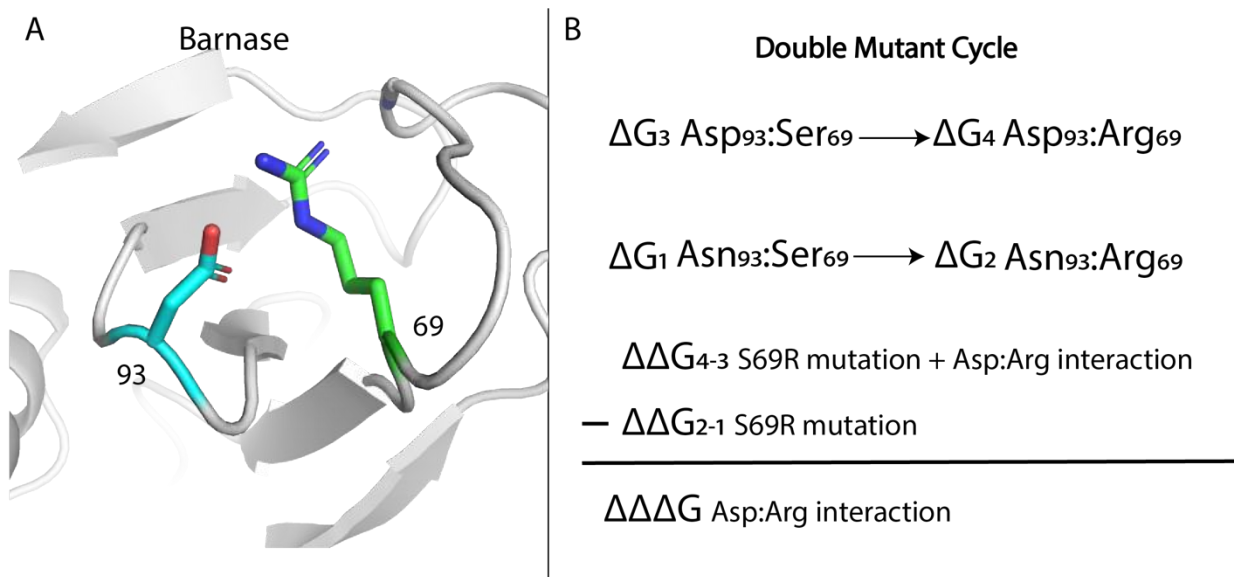


Figure 1-7. Structure of a partially buried salt bridge between Asp93 and Arg 69 (Left). Double mutant cycle measuring the interaction energy between Asp 93 and Arg 69 (Right).

To measure the impact of a third side chain on the binary non-covalent interaction two double mutant cycles can be combined to form a triple mutant cycle cube. Ferst et. al. were interested in salt bridge cooperativity and continued using the protein barnase as a model system.¹¹⁵ In the protein structure they noticed two salt bridges formed by Asp8, Asp12, and Arg110 (Figure 1-8A). They made eight barnase variants and measured each variant's folded

free energy then calculated the triple mutant cycle analysis (figure 1-8B). Each side of the triple mutant cycle cube is a double mutant cycle which measures the interaction between two amino acids in the presence of a control side chain (usually alanine) and again in the presence of a side chain of interest. When you compare opposite cube faces, you measure the effect of a third amino acid position on a double mutant cycle binary interaction. They showed the salt bridge between Asp 8 and Arg110 was worth -0.22 kcal/mol while position 12 was an alanine (red double mutant cycle in figure 1-8B). Then they showed the same salt bridge gained -0.77 kcal mol⁻¹ in interaction energy when position 12 was the wild-type aspartate (green double mutant cycle in figure 1-8B). The Asp8:Arg110 salt bridge was enhanced by -0.77 kcal mol⁻¹ if the Arg110:Glu12 salt bridge is present in the structure (Figure 3C). This demonstrates a cooperativity between the two salt bridges where one salt bridge enhances the second.

When doing double or triple mutant cycle analysis two main assumptions are made: 1. All the protein variants must not have significant structural rearrangements 2. The control side chains should not interact with each other. In this example, the group confirmed structural integrity of each variant by NMR and protein crystallography. The group chose alanine as the mutant side chain and assumed that the Ala8, Ala12, Ala110 would not have substantial interaction. This example illustrates the power of double and triple mutant cycle analysis, showing the strength of binary non-covalent interactions can have can be substantially influenced by a third amino acid.

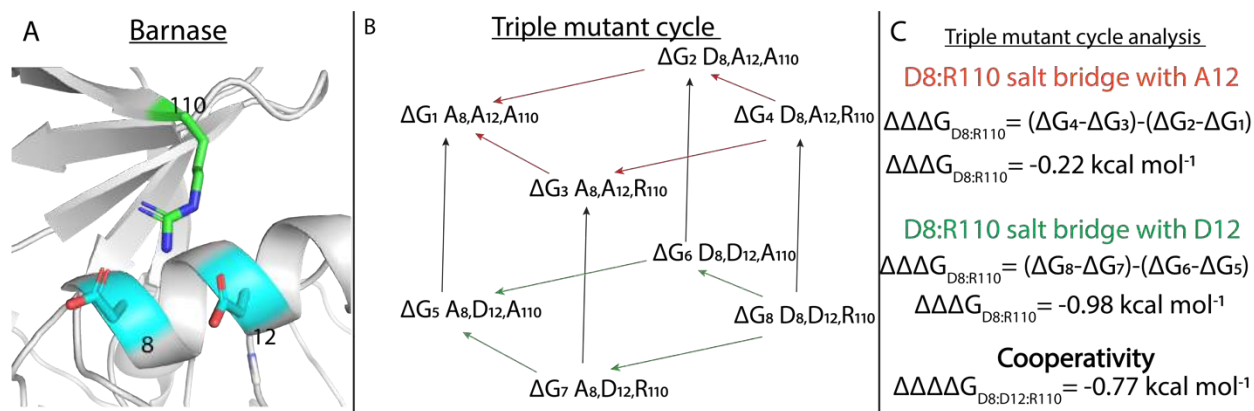


Figure 1-8. Example of a triple mutant cycle analysis between Asp8, Asp12, and Arg110 in the protein barnase (A) structural data representing positions 8, 12, and 110. (B) The eight barnase variants required for the triple mutant cube analysis. (C) The two double mutant cycles that measure the cooperativity between Asp8, Asp12, and Arg110.

1.2.2 Measuring ionic, aromatic, nonpolar cooperativity in protein structure.

Our lab is interested in measuring the cooperativity of interactions involving anion, cation, and aromatic side chains found in protein structure. Chapter 2 describes the progress we have made measuring the cooperativity of cations, anions and aromatic amino acids in a coiled-coil alpha helix model protein. Chapter 3 describes cooperativity between cation, anion, and nonaromatic hydrophobic amino acids in the context of a coiled-coil alpha helix. In Chapter 4 we were tested the strength of an anion- π interaction to a reverse turn found in a beta sheet model system.

1.3 References

1. Blanksby, S. J.; Ellison, G. B., Bond dissociation energies of organic molecules. *Acc Chem Res* **2003**, *36* (4), 255-63.
2. Mahadevi, A. S.; Sastry, G. N., Cooperativity in Noncovalent Interactions. *Chem Rev* **2016**, *116* (5), 2775-825.

3. Schneider, H. J., Dispersive interactions in solution complexes. *Acc Chem Res* **2015**, *48* (7), 1815-22.
4. Antony, J.; Bruske, B.; Grimme, S., Cooperativity in noncovalent interactions of biologically relevant molecules. *Phys Chem Chem Phys* **2009**, *11* (38), 8440-7.
5. Hus, M.; Urbic, T., Strength of hydrogen bonds of water depends on local environment. *J Chem Phys* **2012**, *136* (14), 144305.
6. Stokely, K.; Mazza, M. G.; Stanley, H. E.; Franzese, G., Effect of hydrogen bond cooperativity on the behavior of water. *Proc Natl Acad Sci U S A* **2010**, *107* (4), 1301-6.
7. Perez, C.; Zaleski, D. P.; Seifert, N. A.; Temelso, B.; Shields, G. C.; Kisiel, Z.; Pate, B. H., Hydrogen bond cooperativity and the three-dimensional structures of water nonamers and decamers. *Angew Chem Int Ed Engl* **2014**, *53* (52), 14368-72.
8. Keutsch, F. N.; Cruzan, J. D.; Saykally, R. J., The water trimer. *Chem Rev* **2003**, *103* (7), 2533-77.
9. Gallivan, J. P.; Dougherty, D. A., Cation-pi interactions in structural biology. *Proc Natl Acad Sci U S A* **1999**, *96* (17), 9459-64.
10. Oh, K. S.; Lee, C. W.; Choi, H. S.; Lee, S. J.; Kim, K. S., Origin of the high affinity and selectivity of novel receptors for NH₄⁺ over K⁺: charged hydrogen bonds vs cation-pi interaction. *Org Lett* **2000**, *2* (17), 2679-81.
11. Pellequer, J. L.; Zhao, B.; Kao, H. I.; Bell, C. W.; Li, K.; Li, Q. X.; Karu, A. E.; Roberts, V. A., Stabilization of bound polycyclic aromatic hydrocarbons by a pi-cation interaction. *J Mol Biol* **2000**, *302* (3), 691-9.
12. Tsou, L. K.; Tatko, C. D.; Waters, M. L., Simple cation-pi interaction between a phenyl ring and a protonated amine stabilizes an alpha-helix in water. *J Am Chem Soc* **2002**, *124* (50), 14917-21.
13. Shi, Z.; Olson, C. A.; Kallenbach, N. R., Cation-pi interaction in model alpha-helical peptides. *J Am Chem Soc* **2002**, *124* (13), 3284-91.
14. Dougherty, D. A., Cation-pi interactions in chemistry and biology: A new view of benzene, Phe, Tyr, and Trp. *Science* **1996**, *271* (5246), 163-168.
15. Frontera, A.; Quinonero, D.; Deya, P. M., Cation-pi and anion-pi interactions. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2011**, *1* (3), 440-459.
16. Philip, V.; Harris, J.; Adams, R.; Nguyen, D.; Spiers, J.; Baudry, J.; Howell, E. E.; Hinde, R. J., A Survey of Aspartate-Phenylalanine and Glutamate-Phenylalanine Interactions in the Protein Data Bank: Searching for Anion-pi Pairs. *Biochemistry* **2011**, *50* (14), 2939-2950.
17. Chifotides, H. T.; Dunbar, K. R., Anion-pi Interactions in Supramolecular Architectures. *Accounts of Chemical Research* **2013**, *46* (4), 894-906.
18. Quinonero, D.; Garau, C.; Rotger, C.; Frontera, A.; Ballester, P.; Costa, A.; Deya, P. M., Anion-pi interactions: Do they exist? *Angewandte Chemie-International Edition* **2002**, *41* (18), 3389-3392.
19. Hunter, C. A.; Sanders, J. K. M., THE NATURE OF PI-PI INTERACTIONS. *J. Am. Chem. Soc.* **1990**, *112* (14), 5525-5534.
20. Li, J.; Du, X.; Hashim, S.; Shy, A.; Xu, B., Aromatic-Aromatic Interactions Enable alpha-Helix to beta-Sheet Transition of Peptides to Form Supramolecular Hydrogels. *J Am Chem Soc* **2017**, *139* (1), 71-74.
21. Bhattacharyya, R.; Samanta, U.; Chakrabarti, P., Aromatic-aromatic interactions in and around alpha-helices. *Protein Eng* **2002**, *15* (2), 91-100.

22. Sunner, J.; Nishizawa, K.; Kebarle, P., ION-SOLVENT MOLECULE INTERACTIONS IN THE GAS-PHASE - THE POTASSIUM-ION AND BENZENE. *Journal of Physical Chemistry* **1981**, *85* (13), 1814-1820.
23. Marshall, M. S.; Steele, R. P.; Thanthiriwatte, K. S.; Sherrill, C. D., Potential energy curves for cation-pi interactions: off-axis configurations are also attractive. *J Phys Chem A* **2009**, *113* (48), 13628-32.
24. Tsou, L. K.; Tatko, C. D.; Waters, M. L., Simple cation-pi interaction between a phenyl ring and a protonated amine stabilizes an alpha-helix in water. *J. Am. Chem. Soc.* **2002**, *124* (50), 14917-14921.
25. Spector, S.; Young, P.; Raleigh, D. P., Nativelike structure and stability in a truncation mutant of a protein minidomain: The peripheral subunit-binding domain. *Biochemistry* **1999**, *38* (13), 4128-4136.
26. Ackermann, E. J.; Ang, E. T.; Kanter, J. R.; Tsigelny, I.; Taylor, P., Identification of pairwise interactions in the alpha-neurotoxin-nicotinic acetylcholine receptor complex through double mutant cycles. *J Biol Chem* **1998**, *273* (18), 10958-64.
27. Anderson, M. A.; Ogbay, B.; Arimoto, R.; Sha, W.; Kisselev, O. G.; Cistola, D. P.; Marshall, G. R., Relative strength of cation-pi vs salt-bridge interactions: the Galpha(340-350) peptide/rhodopsin system. *J Am Chem Soc* **2006**, *128* (23), 7531-41.
28. Tatko, C. D.; Waters, M. L., The geometry and efficacy of cation-pi interactions in a diagonal position of a designed beta-hairpin. *Protein Science* **2003**, *12* (11), 2443-2452.
29. Leavens, F. M.; Churchill, C. D.; Wang, S.; Wetmore, S. D., Evaluating how discrete water molecules affect protein-DNA pi-pi and pi(+)-pi stacking and T-shaped interactions: the case of histidine-adenine dimers. *J Phys Chem B* **2011**, *115* (37), 10990-1003.
30. Wells, R. A.; Kellie, J. L.; Wetmore, S. D., Significant strength of charged DNA-protein pi-pi interactions: a preliminary study of cytosine. *J Phys Chem B* **2013**, *117* (36), 10462-74.
31. Stewart, M.; Dunlap, T.; Dourlain, E.; Grant, B.; McFail-Isom, L., Cations form sequence selective motifs within DNA grooves via a combination of cation-pi and ion-dipole/hydrogen bond interactions. *PLoS One* **2013**, *8* (8), e71420.
32. Heindl, H.; Greenwell, P.; Weingarten, N.; Kiss, T.; Terstyanszky, G.; Weinzierl, R. O., Cation-pi interactions induce kinking of a molecular hinge in the RNA polymerase bridge-helix domain. *Biochem Soc Trans* **2011**, *39* (1), 31-5.
33. Rutledge, L. R.; Churchill, C. D.; Wetmore, S. D., A preliminary investigation of the additivity of pi-pi or pi+-pi stacking and T-shaped interactions between natural or damaged DNA nucleobases and histidine. *J Phys Chem B* **2010**, *114* (9), 3355-67.
34. Huang, R. H.; Fremont, D. H.; Diener, J. L.; Schaub, R. G.; Sadler, J. E., A structural explanation for the antithrombotic activity of ARC1172, a DNA aptamer that binds von Willebrand factor domain A1. *Structure* **2009**, *17* (11), 1476-84.
35. Post, M. R.; Tender, G. S.; Lester, H. A.; Dougherty, D. A., Secondary Ammonium Agonists Make Dual Cation-pi Interactions in alpha4beta2 Nicotinic Receptors. *eNeuro* **2017**, *4* (2).
36. Pathak, A. K.; Bandyopadhyay, T., Ortho-7 bound to the active-site gorge of free and OP-conjugated acetylcholinesterase: cation-pi interactions. *Biopolymers* **2016**, *105* (1), 10-20.
37. Dougherty, D. A., Cation-pi interactions involving aromatic amino acids. *J Nutr* **2007**, *137* (6 Suppl 1), 1504S-1508S; discussion 1516S-1517S.

38. Bartoli, S.; Roelens, S., Binding of acetylcholine and tetramethylammonium to a cyclophane receptor: Anion's contribution to the cation-pi interaction. *J. Am. Chem. Soc.* **2002**, *124* (28), 8307-8315.
39. Dougherty, D. A.; Stauffer, D. A., ACETYLCHOLINE BINDING BY A SYNTHETIC RECEPTOR - IMPLICATIONS FOR BIOLOGICAL RECOGNITION. *Science* **1990**, *250* (4987), 1558-1560.
40. Anitha, P.; Sivasakthi, V.; Lavanya, P.; Bag, S.; Kumar, K. M.; Anbarasu, A.; Ramaiah, S., Arginine and Lysine interactions with pi residues in metalloproteins. *Bioinformation* **2012**, *8* (17), 820-6.
41. Yang, C. M.; Zhang, J., Insights into intramolecular Trp and His side-chain orientation and stereospecific pi interactions surrounding metal centers: an investigation using protein metal-site mimicry in solution. *Chemistry* **2010**, *16* (35), 10854-65.
42. Lombardi, A.; Summa, C. M.; Geremia, S.; Randaccio, L.; Pavone, V.; DeGrado, W. F., Retrostructural analysis of metalloproteins: Application to the design of a minimal model for diiron proteins. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97* (12), 6298-6305.
43. Wouters, J., Cation-pi (Na⁺-Trp) interactions in the crystal structure of tetragonal lysozyme. *Protein Sci* **1998**, *7* (11), 2472-5.
44. Garau, C.; Frontera, A.; Quinonero, D.; Ballester, P.; Costa, A.; Deya, P. M., Cation-pi versus anion-pi interactions: Energetic, charge transfer, and aromatic aspects. *Journal of Physical Chemistry A* **2004**, *108* (43), 9423-9427.
45. Garau, C.; Frontera, A.; Quinonero, D.; Ballester, P.; Costa, A.; Deya, P. M., Cation-pi vs anion-pi interactions: a complete pi-orbital analysis. *Chemical Physics Letters* **2004**, *399* (1-3), 220-225.
46. Jackson, M. R.; Beahm, R.; Duvvuru, S.; Narasimhan, C.; Wu, J.; Wang, H. N.; Philip, V. M.; Hinde, R. J.; Howell, E. E., A preference for edgewise interactions between aromatic rings and carboxylate anions: the biological relevance of anion-quadrupole interactions. *J Phys Chem B* **2007**, *111* (28), 8242-9.
47. Philip, V.; Harris, J.; Adams, R.; Nguyen, D.; Spiers, J.; Baudry, J.; Howell, E. E.; Hinde, R. J., A survey of aspartate-phenylalanine and glutamate-phenylalanine interactions in the protein data bank: searching for anion-pi pairs. *Biochemistry* **2011**, *50* (14), 2939-50.
48. Lucas, X.; Bauza, A.; Frontera, A.; Quinonero, D., A thorough anion-pi interaction study in biomolecules: on the importance of cooperativity effects. *Chemical Science* **2016**, *7* (2), 1038-1050.
49. Zaccheddu, M.; Filippi, C.; Buda, F., Anion-pi and pi-pi cooperative interactions regulating the self-assembly of nitrate-triazine-triazine complexes. *J Phys Chem A* **2008**, *112* (7), 1627-32.
50. Wheeler, S. E.; Houk, K. N., Are Anion/pi Interactions Actually a Case of Simple Charge-Dipole Interaction? *Journal of Physical Chemistry A* **2010**, *114* (33), 8658-8664.
51. Franconetti, A.; Nunez-Franco, R.; de Gonzalo, G.; Iglesias-Siguenza, J.; Alvarez, E.; Cabrera-Escribano, F., Fingerprinting the Nature of Anions in Pyrylium Complexes: Dual Binding Mode for Anion-pi Interactions. *Chemphyschem* **2018**, *19* (3), 327-334.
52. Ribic, V. R.; Stojanovic, S. D.; Zlatovic, M. V., Anion-pi interactions in active centers of superoxide dismutases. *Int J Biol Macromol* **2018**, *106*, 559-568.
53. Fujisawa, K.; Humbert-Droz, M.; Letrun, R.; Vauthey, E.; Wesolowski, T. A.; Sakai, N.; Matile, S., Ion Pair-pi Interactions. *J Am Chem Soc* **2015**, *137* (34), 11047-56.

54. Giese, M.; Albrecht, M.; Rissanen, K., Anion- π Interactions with Fluoroarenes. *Chem Rev* **2015**, *115* (16), 8867-95.
55. Chakravarty, S.; Sheng, Z. Z.; Iverson, B.; Moore, B., " η^6 "-Type anion- π in biomolecular recognition. *FEBS Lett* **2012**, *586* (23), 4180-5.
56. Cotelle, Y.; Lebrun, V.; Sakai, N.; Ward, T. R.; Matile, S., Anion- π Enzymes. *ACS Cent Sci* **2016**, *2* (6), 388-93.
57. Yurenko, Y. P.; Bazzi, S.; Marek, R.; Kozelka, J., Anion- π Interactions in Flavoproteins Involve a Substantial Charge-Transfer Component. *Chemistry* **2017**, *23* (14), 3246-3250.
58. Estarellas, C.; Frontera, A.; Quinonero, D.; Deya, P. M., Anion- π interactions in flavoproteins. *Chem Asian J* **2011**, *6* (9), 2316-8.
59. Smith, M. S.; Lawrence, E. E. K.; Billings, W. M.; Larsen, K. S.; Becar, N. A.; Price, J. L., An Anion- π Interaction Strongly Stabilizes the beta-Sheet Protein WW. *ACS Chem Biol* **2017**, *12* (10), 2535-2537.
60. Chifotides, H. T.; Dunbar, K. R., Anion- π interactions in supramolecular architectures. *Acc Chem Res* **2013**, *46* (4), 894-906.
61. Chakravarty, S.; Ung, A. R.; Moore, B.; Shore, J.; Alshamrani, M., A Comprehensive Analysis of Anion-Quadrupole Interactions in Protein Structures. *Biochemistry* **2018**, *57* (12), 1852-1867.
62. McGaughey, G. B.; Gagne, M.; Rappe, A. K., π -stacking interactions - Alive and well in proteins. *J. Biol. Chem.* **1998**, *273* (25), 15458-15463.
63. Zhao, R.; Zhang, R. Q., A new insight into π - π stacking involving remarkable orbital interactions. *Phys Chem Chem Phys* **2016**, *18* (36), 25452-25457.
64. Liu, Y.; Liu, B. Y.; Hao, P.; Li, X.; Li, Y. X.; Wang, J. F., π - π Stacking mediated drug-drug interactions in human CYP2E1. *Proteins* **2013**, *81* (6), 945-54.
65. Liu, J.; Yong, W.; Deng, Y.; Kallenbach, N. R.; Lu, M., Atomic structure of a tryptophan-zipper pentamer. *Proc Natl Acad Sci U S A* **2004**, *101* (46), 16156-61.
66. Frank, B. S.; Vardar, D.; Buckley, D. A.; McKnight, C. J., The role of aromatic residues in the hydrophobic core of the villin headpiece subdomain. *Protein Sci* **2002**, *11* (3), 680-7.
67. Liu, J.; Zheng, Q.; Deng, Y.; Kallenbach, N. R.; Lu, M., Conformational transition between four and five-stranded phenylalanine zippers determined by a local packing interaction. *J Mol Biol* **2006**, *361* (1), 168-79.
68. 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 Struct Biol* **2012**, *22* (4), 432-41.
69. Pace, C. J.; Gao, J., Exploring and exploiting polar- π interactions with fluorinated aromatic amino acids. *Acc Chem Res* **2013**, *46* (4), 907-15.
70. Hill, R. B.; DeGrado, W. F., Solutions structure of alpha D-2, a natively like de novo designed protein. *J. Am. Chem. Soc.* **1998**, *120* (6), 1138-1145.
71. Matta, C. F.; Castillo, N.; Boyd, R. J., Extended weak bonding interactions in DNA: π -stacking (base-base), base-backbone, and backbone-backbone interactions. *J Phys Chem B* **2006**, *110* (1), 563-78.
72. Sponer, J.; Riley, K. E.; Hobza, P., Nature and magnitude of aromatic stacking of nucleic acid bases. *Phys Chem Chem Phys* **2008**, *10* (19), 2595-610.
73. Reddy, S. M.; Shanmugam, G., Role of Intramolecular Aromatic π - π Interactions in the Self-Assembly of Di-l-Phenylalanine Dipeptide Driven by Intermolecular Interactions: Effect of Alanine Substitution. *Chemphyschem* **2016**, *17* (18), 2897-907.

74. An, Y.; Bloom, J. W.; Wheeler, S. E., Quantifying the pi-Stacking Interactions in Nitroarene Binding Sites of Proteins. *J Phys Chem B* **2015**, *119* (45), 14441-50.
75. He, T.; Gershenson, A.; Eyles, S. J.; Lee, Y. J.; Liu, W. R.; Wang, J.; Gao, J.; Roberts, M. F., Fluorinated Aromatic Amino Acids Distinguish Cation-pi Interactions from Membrane Insertion. *J Biol Chem* **2015**, *290* (31), 19334-42.
76. Chelli, R.; Procacci, P., Comment on "from subtle to substantial: role of metal ions on pi-pi interactions". *J Phys Chem B* **2006**, *110* (20), 10204-5; discussion 10206-7.
77. Reddy, A. S.; Vijay, D.; Sastry, G. M.; Sastry, G. N., From subtle to substantial: role of metal ions on pi-pi interactions. *J Phys Chem B* **2006**, *110* (6), 2479-81.
78. Quinonero, D.; Frontera, A.; Garau, C.; Ballester, P.; Costa, A.; Deya, P. M., Interplay between cation-pi, anion-pi and pi-pi interactions. *Chemphyschem* **2006**, *7* (12), 2487-91.
79. Frontera, A.; Quinonero, D.; Costa, A.; Ballester, P.; Deya, P. M., MP2 study of cooperative effects between cation-pi, anion-pi and pi-pi interactions. *New Journal of Chemistry* **2007**, *31* (4), 556-560.
80. Garau, C.; Quinonero, D.; Frontera, A.; Ballester, P.; Costa, A.; Deya, P. M., Anion-pi interactions: must the aromatic ring be electron deficient? *New Journal of Chemistry* **2003**, *27* (2), 211-214.
81. Ngola, S. M.; Kearney, P. C.; Mecozzi, S.; Russell, K.; Dougherty, D. A., A selective receptor for arginine derivatives in aqueous media. Energetic consequences of salt bridges that are highly exposed to water. *J. Am. Chem. Soc.* **1999**, *121* (6), 1192-1201.
82. Fairchild, R. M.; Holman, K. T., Selective anion encapsulation by a metalated cryptophane with a pi-acidic interior. *J. Am. Chem. Soc.* **2005**, *127* (47), 16364-16365.
83. Staffilani, M.; Bonvicini, G.; Steed, J. W.; Holman, K. T.; Atwood, J. L.; Elsegood, M. R. J., Bowl vs saddle conformations in cyclononatriene-based anion binding hosts. *Organometallics* **1998**, *17* (9), 1732-1740.
84. Staffilani, M.; Hancock, K. S. B.; Steed, J. W.; Holman, K. T.; Atwood, J. L.; Juneja, R. K.; Burkhalter, R. S., Anion binding within the cavity of pi-metalated calixarenes. *J. Am. Chem. Soc.* **1997**, *119* (27), 6324-6335.
85. Atwood, J. L.; Holman, K. T.; Steed, J. W., Laying traps for elusive prey: Recent advances in the non-covalent binding of anions. *Chemical Communications* **1996**, (12), 1401-1407.
86. Holman, K. T.; Halihan, M. M.; Steed, J. W.; Jurisson, S. S.; Atwood, J. L., Hosting a Radioactive Guest - Binding of (Tco4-)-Tc-99 by a Metalated Cyclotrimeratrylene. *J. Am. Chem. Soc.* **1995**, *117* (29), 7848-7849.
87. Chiti, F.; Dobson, C. M., Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu Rev Biochem* **2017**, *86*, 27-68.
88. Schmitz, M.; Cramm, M.; Llorens, F.; Muller-Cramm, D.; Collins, S.; Atarashi, R.; Satoh, K.; Orru, C. D.; Groveman, B. R.; Zafar, S.; Schulz-Schaeffer, W. J.; Caughey, B.; Zerr, I., The real-time quaking-induced conversion assay for detection of human prion disease and study of other protein misfolding diseases. *Nat Protoc* **2016**, *11* (11), 2233-2242.
89. Wang, M.; Kaufman, R. J., Protein misfolding in the endoplasmic reticulum as a conduit to human disease. *Nature* **2016**, *529* (7586), 326-35.
90. Michalczyk, K.; Ziman, M., Current concepts in human prion protein (Prp) misfolding, Prnp gene polymorphisms and their contribution to Creutzfeldt-Jakob Disease (CJD). *Histol Histopathol* **2007**, *22* (10), 1149-59.

91. Chiti, F.; Dobson, C. M., Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* **2006**, *75*, 333-66.
92. Gregersen, N.; Bross, P.; Vang, S.; Christensen, J. H., Protein misfolding and human disease. *Annu Rev Genomics Hum Genet* **2006**, *7*, 103-24.
93. Dill, K. A., Dominant forces in protein folding. *Biochemistry* **1990**, *29* (31), 7133-55.
94. Baker, E. N.; Hubbard, R. E., Hydrogen bonding in globular proteins. *Prog Biophys Mol Biol* **1984**, *44* (2), 97-179.
95. Myers, J. K.; Pace, C. N., Hydrogen bonding stabilizes globular proteins. *Biophys J* **1996**, *71* (4), 2033-9.
96. Stickle, D. F.; Presta, L. G.; Dill, K. A.; Rose, G. D., Hydrogen bonding in globular proteins. *J Mol Biol* **1992**, *226* (4), 1143-59.
97. Kellis, J. T., Jr.; Nyberg, K.; Sali, D.; Fersht, A. R., Contribution of hydrophobic interactions to protein stability. *Nature* **1988**, *333* (6175), 784-6.
98. Cheung, M. S.; Garcia, A. E.; Onuchic, J. N., Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc Natl Acad Sci U S A* **2002**, *99* (2), 685-90.
99. Shank, E. A.; Cecconi, C.; Dill, J. W.; Marqusee, S.; Bustamante, C., The folding cooperativity of a protein is controlled by its chain topology. *Nature* **2010**, *465* (7298), 637-40.
100. Ghosh, K.; Dill, K. A., Theory for protein folding cooperativity: helix bundles. *J Am Chem Soc* **2009**, *131* (6), 2306-12.
101. Weikl, T. R.; Palassini, M.; Dill, K. A., Cooperativity in two-state protein folding kinetics. *Protein Sci* **2004**, *13* (3), 822-9.
102. Scalley-Kim, M.; Baker, D., Characterization of the folding energy landscapes of computer generated proteins suggests high folding free energy barriers and cooperativity may be consequences of natural selection. *J Mol Biol* **2004**, *338* (3), 573-83.
103. Dill, K. A.; Fiebig, K. M.; Chan, H. S., Cooperativity in protein-folding kinetics. *Proc Natl Acad Sci U S A* **1993**, *90* (5), 1942-6.
104. Watters, A. L.; Deka, P.; Corrent, C.; Callender, D.; Varani, G.; Sosnick, T.; Baker, D., The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection. *Cell* **2007**, *128* (3), 613-24.
105. Brinda, K. V.; Vishveshwara, S., A network representation of protein structures: implications for protein stability. *Biophys J* **2005**, *89* (6), 4159-70.
106. de la Cruz, X.; Reverter, J.; Fita, I., Representation of noncovalent interactions in protein structures. *J Mol Graph* **1992**, *10* (2), 96-100, 110.
107. DeBartolo, J.; Colubri, A.; Jha, A. K.; Fitzgerald, J. E.; Freed, K. F.; Sosnick, T. R., Mimicking the folding pathway to improve homology-free protein structure prediction. *Proc Natl Acad Sci U S A* **2009**, *106* (10), 3734-9.
108. Verschuere, E.; Vanhee, P.; Rousseau, F.; Schymkowitz, J.; Serrano, L., Protein-Peptide Complex Prediction through Fragment Interaction Patterns. *Structure* **2013**, *21* (5), 789-797.
109. Horovitz, A., Double-mutant cycles: a powerful tool for analyzing protein structure and function. *Fold Des* **1996**, *1* (6), R121-6.
110. Serrano, L.; Horovitz, A.; Avron, B.; Bycroft, M.; Fersht, A. R., Estimating the contribution of engineered surface electrostatic interactions to protein stability by using double-mutant cycles. *Biochemistry* **1990**, *29* (40), 9343-52.

111. Camara-Campos, A.; Musumeci, D.; Hunter, C. A.; Turega, S., Chemical double mutant cycles for the quantification of cooperativity in H-bonded complexes. *J Am Chem Soc* **2009**, *131* (51), 18518-24.
112. Hunter, C. A.; Anderson, H. L., What is cooperativity? *Angew Chem Int Ed Engl* **2009**, *48* (41), 7488-99.
113. Vaughan, C. K.; Harryson, P.; Buckle, A. M.; Fersht, A. R., A structural double-mutant cycle: estimating the strength of a buried salt bridge in barnase. *Acta Crystallogr D Biol Crystallogr* **2002**, *58* (Pt 4), 591-600.
114. Oliveberg, M.; Fersht, A. R., New approach to the study of transient protein conformations: the formation of a semiburied salt link in the folding pathway of barnase. *Biochemistry* **1996**, *35* (21), 6795-805.
115. Horovitz, A.; Serrano, L.; Avron, B.; Bycroft, M.; Fersht, A. R., Strength and cooperativity of contributions of surface salt bridges to protein stability. *J Mol Biol* **1990**, *216* (4), 1031-44.

2 MEASURING A HELICAL SALT BRIDGE ENHANCED BY AROMATIC AMINO ACIDS.

2.1 Introduction

Salt bridges, or the interaction between negative and positive amino acids contribute significant conformational stability to protein structure, especially when two ionic groups are in close proximity (i.e., within ~ 4 Å of each other).¹⁻⁹ For example, placing complementarily charged amino acid side chains three or four residues apart can substantially stabilize α -helical peptides because they are on the same face of the helix and in close proximity.¹⁰ Analogous approaches also promote helical conformations within β -peptides and other foldamers.¹¹⁻¹³ Increasing the distance between the positively and negatively charged residues causes the residues to be solvated by water and ion molecules from the solvent, thereby attenuating the salt-bridge interaction.¹⁴ Several lines of evidence show that longer range salt bridges are possible when an arene is positioned between the cation, and anion. (see Section 1.1.2).¹⁵⁻¹⁸

We set out to measure the impact of an arene on a Glu-Lys salt bridge in the protein **1CW**. **1CW** is a trimer α -helix coiled coil whose structure has been extensively characterized.^{9, 19-20} **1CW** is a two state folder and is therefore amenable to detailed thermodynamic analysis.^{9, 21} The small size of **1CW** (33 amino acids) facilitates the chemical synthesis of the protein with complete control of site specific mutations necessary for the measurement of noncovalent interactions. Alanine 10, serine 14, and alanine 18 occupy the *i*, the *i*+4, and the *i*+8 positions along the solvent-

exposed faces of each α -helix in 1CW (Figure 2-1). Several studies have shown that the i to $i+3$ or i to $i+4$ spacing in an α -helix facilitates favorable noncovalent interactions due to the close proximity of these positions.^{5, 22-23}

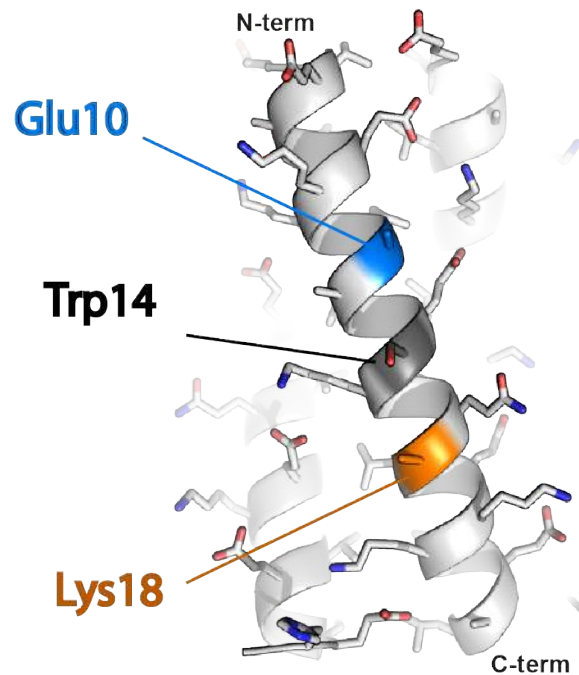


Figure 2-1. Coiled coil trimeric protein. The protein has three solvent exposed sites where we can measure the interaction between anionic, aromatic, and cationic amino acids.

2.1 Results and discussion

Cooperativity between a long-range salt bridges mediated by Tryptophan.

We wondered to what extent Trp14 might enable a favorable long-range salt-bridge interaction between Glu10 and Lys18. Assessing the contribution of an interaction between two amino acid residues to peptide or protein conformational stability requires double mutant cycle

analysis, in which each residue is mutated to a non-interacting surrogate, individually and then in combination. For example, to determine whether a Glu10–Lys18 interaction contributes substantially to the stability of the three-helix bundle formed by **1** (Figure 2-2), we mutated Glu10 to Ala to give peptide **2**; Lys18 to Ala to give peptide **3**; and both Glu10 and Lys18 to Ala to give peptide **4**. Like peptide **1**, each of these variants forms a cooperatively folded helix-bundle homotrimer (See Supporting information section 2.4.4). The difference in the folding free energies of **2** vs. **1** is a measure of the energetic contribution of the putative Glu10–Lys18 interaction plus the impact of the Glu10 to Ala mutation (i.e., $\Delta\Delta G_{21} = \Delta\Delta G_{E\cdot K} + \Delta\Delta G_{E10A}$). In contrast, no Glu10–Lys18 interaction is possible within peptides **3** and **4**, in which Ala occupies position 18. Therefore, the difference in the folding free energies of **4** vs. **3** must only reflect the impact of the Glu to Ala mutation (i.e., $\Delta\Delta G_{43} = \Delta\Delta G_{E10A}$). Subtracting $\Delta\Delta G_{43}$ from $\Delta\Delta G_{21}$ therefore provides a measure of the energetic contribution of the Glu10–Lys18 interaction to helix bundle stability: $\Delta\Delta G_{E\cdot K} = \Delta\Delta G_{21} - \Delta\Delta G_{43} = -2.18 \pm 0.09 \text{ kcal mol}^{-1}$ for the three Glu10–Lys18 interactions present in the helix-bundle formed by **1**; $-0.73 \pm 0.02 \text{ kcal mol}^{-1}$ per individual Glu10–Lys18 interaction. Despite the large distance between Glu10 and Lys18 ($\sim 13 \text{ \AA}$), the energetic contribution of the Glu10–Lys18 interaction is similar in magnitude to that of many previously characterized short-range salt bridges within proteins.^{1-2, 4-5, 8, 24}

Peptide	Sequence	ΔG_f (kcal/mol of trimer)
1	Ac-EVEALEKKV E ALE W KV Q K LEKKVEALEHGWDGR	-17.23 ± 0.03
2	Ac-EVEALEKKV A ALE W KV Q K LEKKVEALEHGWDGR	-16.84 ± 0.02
3	Ac-EVEALEKKV E ALE W KV Q A LEKKVEALEHGWDGR	-15.24 ± 0.03
4	Ac-EVEALEKKV A ALE W KV Q A LEKKVEALEHGWDGR	-17.03 ± 0.04
5	Ac-EVEALEKKV E ALE S KV Q K LEKKVEALEHGWDGR	-14.09 ± 0.02
6	Ac-EVEALEKKV A ALE S KV Q K LEKKVEALEHGWDGR	-15.87 ± 0.03
7	Ac-EVEALEKKV E ALE S KV Q A LEKKVEALEHGWDGR	-15.48 ± 0.02
8	Ac-EVEALEKKV A ALE S KV Q A LEKKVEALEHGWDGR	-16.39 ± 0.04

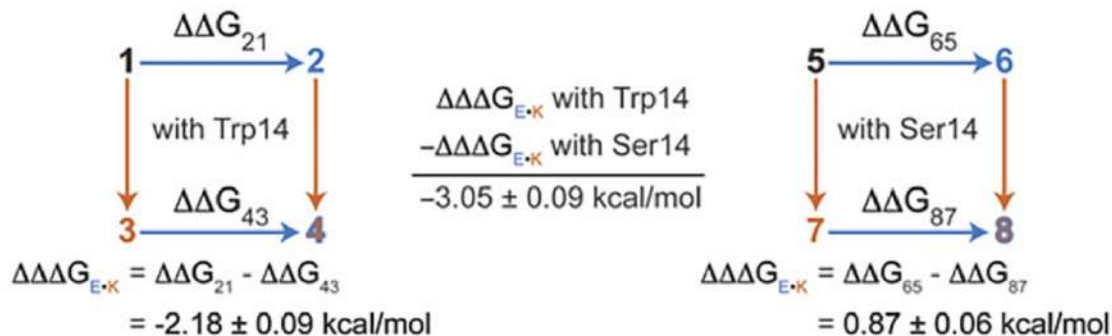


Figure 2-2. Peptides 1-8 used to measure the cooperative interaction between a Glu10, Trp14, and Lys18. The double mutant cycle that measure the interaction in the presence of Trp14 (bottom left) and the double mutant cycle in the presence of Ser14 (bottom right) can be compared to quantify the impact of Trp14 on the Glu10:Lys18 salt bridge. CD experiments were done in 20 mM phosphate buffer pH 7.0 at 30 μ M peptide concentration.

To assess the contribution of Trp14 to the long-range Glu10–Lys18 interaction in peptide 1, we prepared peptide 5, a derivative of 1 in which Trp14 has been replaced by Ser, which occupies position 14 in the parent peptide from which 1 was derived; subsequent experiments demonstrated that Ser is a reasonable mimic of Ala at this position (supporting information section 2.4.7). We then repeated the double mutant cycle analysis to assess the strength of the Glu10–Lys18 interaction in the absence of Trp14 by comparing the stability of 5 with that of peptides 6, 7, and 8 (the Ser14 analogues of peptides 2, 3, and 4 respectively). Interestingly, when Ser occupies position 14 the salt bridge interaction is unfavorable ($\Delta\Delta\Delta G_{E\cdot K} = \Delta\Delta G_{65} - \Delta\Delta G_{87} = 0.89 \pm 0.06$ kcal mol⁻¹ for the three E·K interactions in the peptide 5). Comparing the contribution to

conformational stability of the Glu10-Lys18 salt bridge in the presence of Trp14 vs. Ser14 shows that the Trp14 enhances each salt bridge in **5** by -1.02 ± 0.09 kcal mol⁻¹ per Glu10-Trp14-Lys18 interaction (Figure 2-2).

Table 2-1. Structure activity relationship of Glu10, Lys 18 salt bridge enhanced by Trp 14.^a

Peptide	Sequence	Salt-bridge with Arene $\Delta\Delta\Delta G_f$ (kcal/mol)	Salt-bridge with Ser14 $\Delta\Delta\Delta G_f$ (kcal/mol)	Impact of arene on trimer $\Delta\Delta\Delta G_f$ (kcal/mol)
1	•••KVEALEW KVQK LE•••	-2.25 ± 0.08	0.20 ± 0.06	-2.45 ± 0.10
9	•••KVDALEW KVQK LE•••	-0.74 ± 0.08	0.46 ± 0.07	-1.21 ± 0.11
10	•••KVEALEW KVQO LE•••	-0.23 ± 0.06	-0.25 ± 0.10	0.02 ± 0.08
11	•••KVALEW KVQK LE•••	0.35 ± 0.11	-0.86 ± 0.08	1.21 ± 0.14
12	•••KV K ALEW KVQ ELE•••	1.01 ± 0.12	0.25 ± 0.10	0.76 ± 0.15
13	•••KVEALEW MeKVQK LE•••	-0.79 ± 0.09	0.20 ± 0.06	-1.00 ± 0.11

^a Partial sequences given for 1, 9-13 from residue 8-20. Variable temperature CD experiments done in 20 mM sodium phosphate pH 7.0, 1M Urea, 343.15 K. Data are given as \pm standard error for 30 μ M peptide.

We next explored whether the large Trp-based enhancement of the Glu10–Lys18 interaction could tolerate changes in the position of the Glu10 carboxylate relative to the Lys18 ammonium group (Table 2-1). To that end, we performed triple mutant cycle analyses on peptides **9–12** and their sequence variants, in which we varied the identity and relative location of the positive and negatively charged components of the salt bridge. Some of these variants were so stable in 20 mM sodium phosphate buffer (pH 7) as to preclude direct assessment of their folding free energies. To facilitate direct comparison among peptides **1, 9–12**, we performed variable temperature CD experiments on these compounds in 20 mM sodium phosphate buffer (pH 7) with 1 M urea, which adjusted helix-bundle stability such that we could reliably extract thermodynamic parameters from the variable temperature CD data for each compound. Replacing Glu10 with Asp

(peptide **9**, Table 2-1) decreases the Trp-based enhancement in the salt-bridge interaction by 50%. Replacing Lys18 with ornithine (peptide **10**) abolishes the Trp-based enhancement completely. This is an interesting contrast with earlier work by Waters and coworkers, who showed that Phe interacts more favorably with an i+4 ornithine than an i+4 Lys.²² Moving Glu10 to position 11 (i.e., one residue closer to Trp14; peptide **11**) interestingly causes the salt bridge to be favorable in the presence of Ser14 and non-favorable with Trp14. Peptide **12** shows that Trp14 makes an unfavorable Lys10-Glu18 salt bridge worse. We speculate that the failure of Trp to enhance a long-range salt-bridge between Lys10 and Glu18 (peptide **12**) could reflect an unfavorable interaction between the long-range salt bridge and the helical macrodipole, but the precise mechanism remains unclear. These results suggest that the ability for Trp14 to enhance the interaction between Glu and Lys is highly sensitive to the relative positions of the Glu carboxylate and the Lys ammonium. We also show that the Trp-based enhancement of the Glu10-Lys18 salt bridge appears to have a significant hydrogen bonding component because when methylating the indole nitrogen on Trp 14 side chain (W_{me}) decreases the Trp based stabilization by about 60% (Figure 2-3, peptide **13**).

2.2.2 The structural influence of the arene on the long-range salt bridge.

We then tested the ability of other aromatic side chains to enhance the Glu10-Lys18 salt bridge (Figure 2-3). Tyr (peptide **14**) enhances the stability of the Glu10-Lys18 salt bridge to a similar degree as Trp. However, as with Trp, methylation of the Tyr phenolic oxygen (peptide **15**) decreases Tyr-based stabilization by ~65%, again suggesting a substantial hydrogen-bonding component for Trp or Tyr based enhancement of the Glu10-Lys18 interaction. In agreement with this conclusion, Phe (peptide **16**) is substantially worse at enhancing the Glu10-Lys19 interaction

than both Tyr and Trp. Interestingly, addition of 0.25M NaCl to the buffer increases the Phe-based enhancement of the Glu10–Lys18 interaction (supporting information section 2.4.8), suggesting that the enhanced long-range salt-bridge is resistant to screening by salt, possibly because the non-polar residue excludes ions from the space immediately between Glu10 and Lys18; alternatively, it is possible the the Phe-based enhancement of the Glu10-Lys18 interaction also has a hydrophobic component.

We wondered whether the still sizeable ability of Phe to enhance the Glu10–Lys18 interaction (-1.25 ± 0.09 for the three Phe residues in 16; -0.42 ± 0.04 kcal mol⁻¹ enhancement per individual Phe) was related to a cation– π interaction between the positively charged Lys18 and the electron-rich face of the Phe arene. To test this hypothesis, we replaced Phe with pentafluorophenylalanine (f₅Phe), in which the face of the aromatic ring is electron-poor instead of electron-rich (peptide **17**). Interestingly F5Phe enhances the stability of the Glu10–Lys18 interaction to a similar degree as does Phe. This observation agrees qualitatively with the computational work of Deyá and his coworkers, who predicted that the two charges positioned across hexafluorobenzene would still interact cooperatively.¹⁵ The similar impact f₅Phe and Phe on the Glu10-Lys18 salt bridge suggests that the electrostatic potential of the ring may not be a major determinant of the observed salt-bridge stabilization. To further test this hypothesis, we made peptides **18–21** along with their sequence variants and showed that p-fluorophenylalanine (p-fPhe, Peptide **18**), 3,4-difluorophenylalanine (3,4-f₂Phe, peptide **19**), and p-nitrophenylalanine (p-NO₂Phe, peptide **20**) are each similar or superior to Phe14 in their ability to stabilize Glu10-Lys18 salt bridge (Figure 2-3). An exception to this trend, for reasons that are still unclear, was that 4-pyridinylalanine (PyrA, peptide **21**) had a substantially smaller impact than that of Phe14.

This may reflect perturbation of the pyridinium/pyridine acid-base equilibrium by Glu10 and/or Lys18.


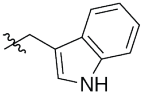
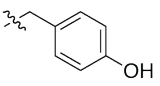
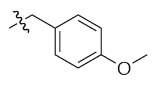
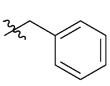
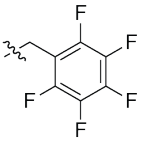
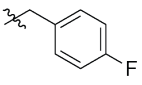
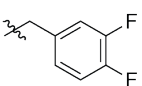
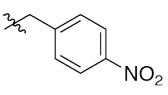
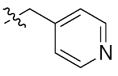
Peptide	Sequence variant	Residue 14	Salt-bridge Interaction $\Delta\Delta\Delta G_f$ (kcal/mol)	Influence of Residue 14 $\Delta\Delta\Delta\Delta G_f$ (kcal/mol)
8	ESK		0.87 ± 0.06	---
1	EWK		-2.18 ± 0.09	-3.05 ± 0.09
14	EYK		-1.93 ± 0.07	-2.80 ± 0.09
15	EY _{me} K		-0.01 ± 0.06	-0.88 ± 0.06
16	EFK		-0.38 ± 0.07	-1.25 ± 0.09
17	EF ₅ PheK		-0.21 ± 0.06	-1.08 ± 0.09
18	E <i>p</i> -fFK		-0.33 ± 0.05	-1.20 ± 0.07
19	E3,4-f ₂ FK		-0.79 ± 0.04	-1.66 ± 0.07
20	E <i>p</i> -NO ₂ FK		-0.61 ± 0.06	-1.48 ± 0.08
21	EPyrAK		0.21 ± 0.05	-0.66 ± 0.07

Figure 2-3. The effect of changing the arene side chain at position 14 on the strength of the Glu10:Lys18 salt bridge

2.2.3 Crystal structures of 1CW variants

Like small molecules, proteins can form crystals when they are supersaturated in solution.²⁵ Under saturated conditions individual protein molecules can pack in ordered arrays

through noncovalent interactions between protein molecules. High-quality crystals are formed when well-ordered packing arrangements are maintained throughout the crystal, the crystal has little contamination, and is large enough to collect diffraction data.²⁶⁻²⁹ Initial crystallization screens of **1 (EWK)** indicated two crystal forms, or two differing packing arrangements occurred depending on solution condition (crystal forms EWK₁ and EWK₂ are shown in Figure 2-4A, details of each crystal packing lattice can be found in the supporting information 2.4.6). Each crystal form folds into a homotrimeric helix bundle. As is typical for helix bundles, the primary sequence of **EWK** is characterized by a repeating seven-residue sequence (i.e., a heptad), in which hydrophobic residues occupy the first and fourth positions (positions a and d), respectively, of an abcdefg heptad.^{19, 30-31} Each heptad spans two turns of the α -helix (at ~ 3.5 residues per turn). Consequently, nonpolar a and d-position residues occupy the same face of the helix; burial of these residues at the inter-helical interface generally provides the major driving force for helix-bundle self-association.³⁰ Similarly, polar residues at b, c, and f-positions in peptide **1** occupy the solvent-exposed surface of the helix bundle; b-position Glu10 is two helical turns away from b-position Lys18, with f-position Trp14 in between them. Figure 2-4B shows a single helix from EWK₁, and EWK₂ superimposed, both crystal forms show the same rotamer for tryptophan. We then measured the distances between the ammonium group on Lys18 and the C δ Glu10 on both EWK crystal forms (Figure 2-4C). The long Glu10:Lys18 distances (13.9 Å,

15.9 Å) indicate the potential for a long-range salt bridge being enhanced by Trp14.

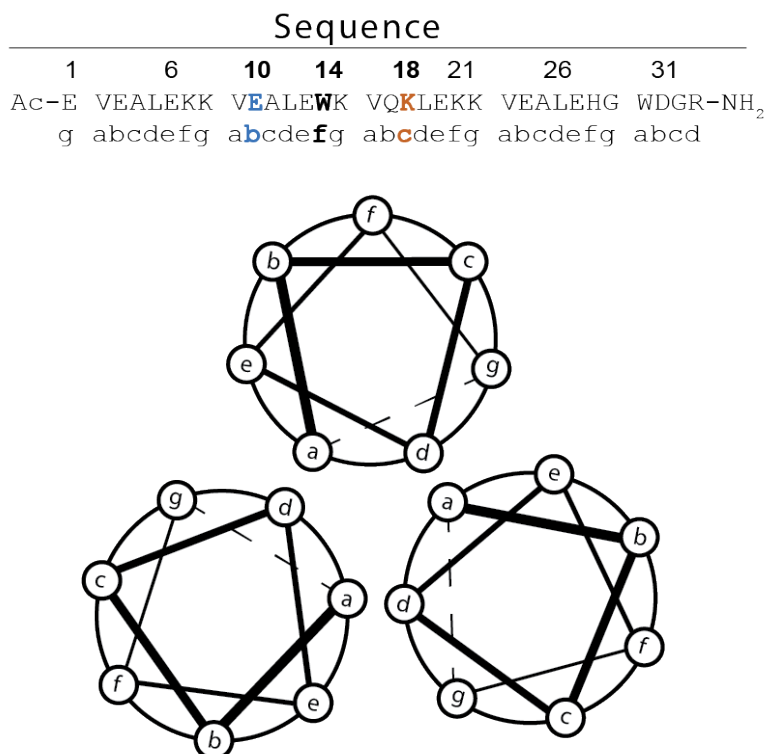


Figure 2-4 Representation of EWK trimer helical wheel where Glu10, Trp14, and Lys18 occupy the solvent exposed b, f, and c positions within the heptad repeat.

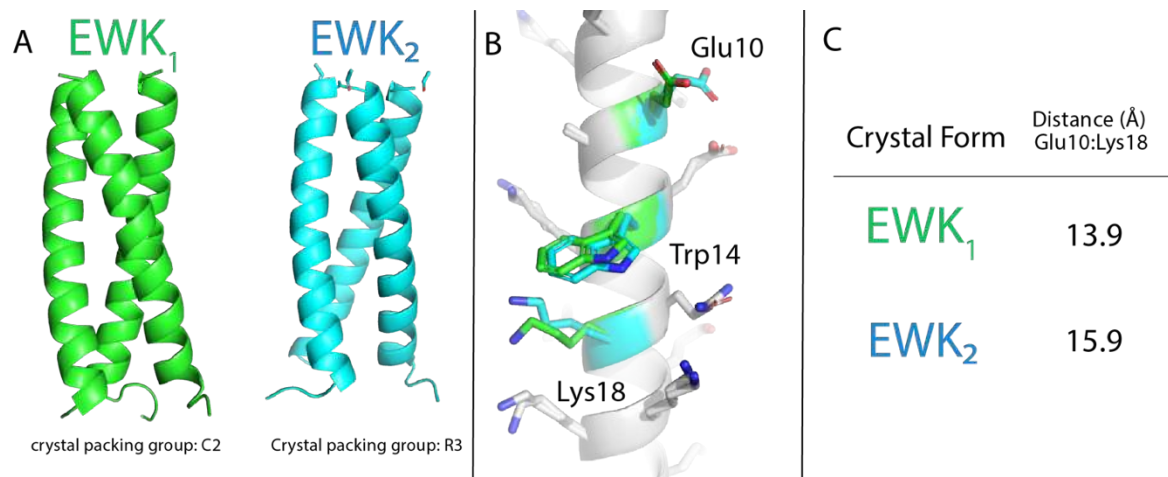


Figure 2-5. (A) Two crystal forms for peptide 1 (EWK). (B) Glu10, Trp14, and Lys 18 overlaid from each crystal form. (C) Distances between the lysine N_{zeta} center of the Trp ring and the distance between the Glu 10 C_{epsilon} and the Trp14 Ring.

The structure of peptide EWK₂ (peptide 1) was overlaid with individual α -helices from crystal structures of five derivatives of EWK, in which Tyr (EYK), Phe (EFK), p-FPhe (Ep-fFK), 3,4-F₂Phe (E3,4-f₂FK), p-NO₂Phe (Ep-NO₂FK), or PyrA (EPyrAK) occupy position 14 instead of Trp (Figure 2-5). We wondered whether the side-chain conformations (i.e. rotamers) adopted by Glu10, Lys18, or the i-position arenes would correlate with the ability of each arene to enhance the Glu10-Lys18 salt-bridge. To explore this possibility, we measured dihedral angles χ_1 (defined by the atoms N α , C α , C β , and C γ) and χ_2 (C α , C β , C γ , and C δ) for each side-chain, and described each χ_1 and χ_2 in one of three possible ways: g⁺ (positive gauche) for angles between 0° and 120°; g⁻ (negative gauche) for angles between 0° and -120°; and t (trans) for angles between 120° and 180° or between -120° and -180° (the symmetry of the Phe, Tyr, F5Phe, p-fPhe, p-NO₂Phe, and PyrA side chains make it impossible to distinguish between g⁺ and g⁻ for χ_2 ; consequently, χ_2 values between 0° and 120° or between 0° and -120° for these side-chains are simply described as g.). We observed no particular trend in the side-chain conformations of the Glu10 or Lys18 from variant to variant. However, PyrA14 in EPyrAK adopted an unusual g⁻,t conformation, in contrast to the more conventional t,g conformation adopted by EWK, EYK, EFK, Ep-fFK, E3,4-f₂FK, and Ep-NO₂FK. PyrA14 had a much smaller impact on the Glu10-Lys18 salt bridge than did the other arenes investigated here, and it seems possible that this effect could be related to the unusual g⁻,t rotamer adopted by PyrA14.

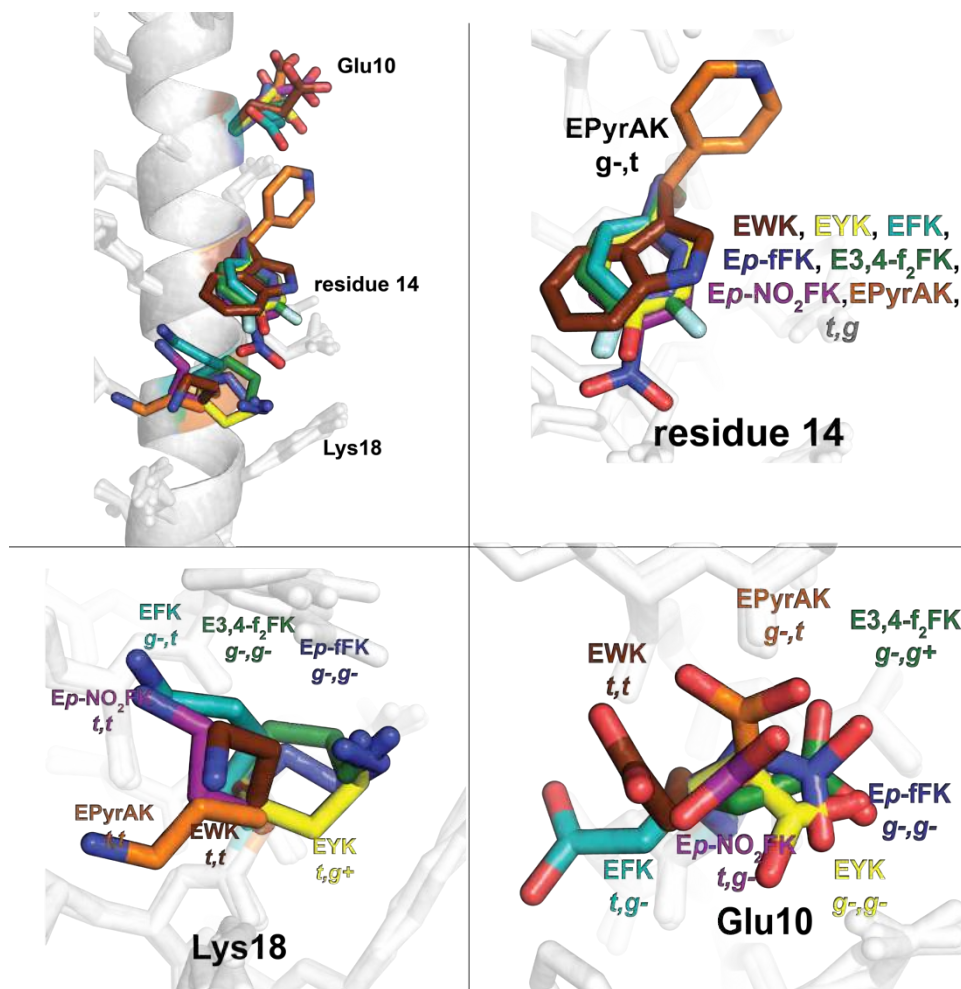


Figure 2-6. EWK, EYK, EFK, Ep-fFK, E3,4-f₂FK, Ep-NO₂FK, and EPyrAK structures overlaid with the comparison of aromatic, Lys18, and Glu 14 side chain rotamers.

Each of these helix-bundle trimers crystallized in a dense lattice with extensive inter-bundle contacts between surface residues (Figure 2-6). These contacts are relevant to the crystalline state, but not to solution, where each helix bundle trimer is surrounded by water and does not interact closely with other trimers. Fig. 2-7 shows some of these inter-bundle contacts involving Glu10, Lys18, and the i-position arenes. In most of these variants, Glu10 from one helix is intimately interdigitated with the i-position arene from another helix within an adjacent bundle. Moreover, the unusual rotamer adopted by PyrA14 in EPyrAK appears to facilitate an inter-bundle

salt bridge with Glu10 from another helix. These intimate crystal contacts suggest that we should be cautious in using these structural data from the crystalline state to draw conclusions about the behavior of the molecule in solution.

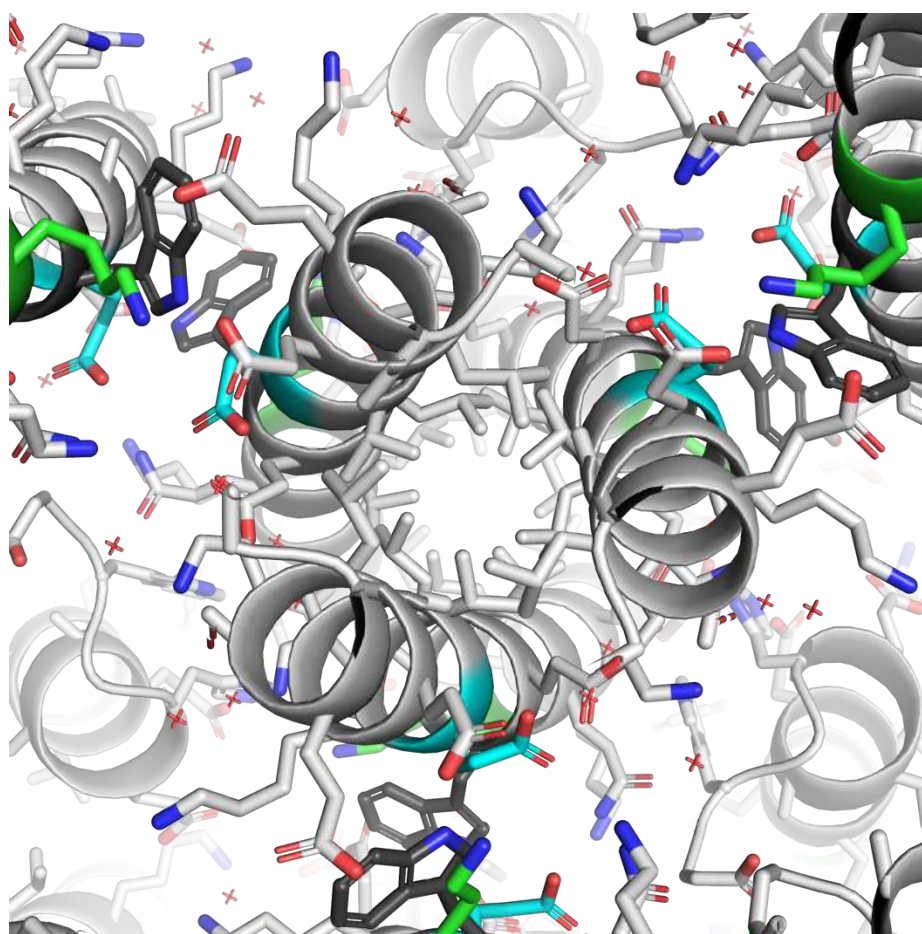


Figure 2-7. Peptide 1 (EWK) crystal packing lattice. Red X symbolized detectable water molecules in the electron density map. Lys10 is colored in green, Trp14 in black, and Glu10 in cyan.

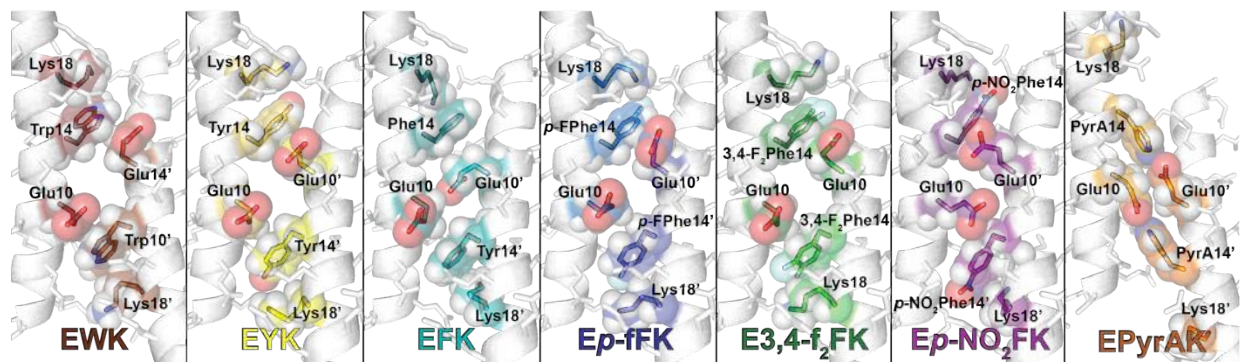


Figure 2-8. Crystal packing interface for EWK, EYK, EFK, Ep-fFK, E3,4-f₂FK, Ep-NO₂FK, and EPyrAK.

2.2.4 NMR model system

We next used NMR spectroscopy to explore the solution behavior of the *i*-4-position Glu, an *i*-position Trp, and an *i*+4-position Lys within an α -helix. Peptide **pA-EWK** (Figure 2-8) is derived from the parent α -helical model peptide **pA** (sequence: Ac-OO-AAAAA-AAAAA-AAAAA-AAAA-OO-Y-NH₂), which contains a 19-residue polyalanine segment (for high helix propensity) flanked by two ornithine residues on either end (to increase solubility in aqueous solution); a C-terminal tyrosine provides a spectroscopic handle for accurate concentration determination.³²⁻³³ Ala12 is centrally located within the 19-residue polyalanine segment; Ala8 and Ala16 occupy the *i*-4 and *i*+4 positions, respectively, relative to Ala12. We replaced Ala16 with Lys, Ala12 with Trp, and Ala8 with Glu to give peptide **pA-EWK**. We obtained two-dimensional TOCSY and ROESY spectra of a 5 mM solution of **pA-EWK** in 20 mM sodium phosphate (pH 7) and assigned the backbone and side-chain resonances for Glu8, Trp12, and Lys16 (see supporting information section 2.4.9 for details). The ROESY spectrum of **pA-EWK** reveals several non-sequential NOEs between backbone and side-chain protons of Glu8 and Trp12 and between backbone and side-chain protons of Trp12 and Lys16 (Figure 2); normalized volumes of these NOE peaks are given in the supporting information. H δ 1 and H ϵ 1 on the indole ring of Trp12

are each involved in multiple NOEs with protons on both Glu8 and Lys16: Trp12H δ 1 has NOEs with Glu8H β , Glu8H γ , Lys16H α , Lys16H γ , Lys16H δ , and Lys16H ϵ , whereas Trp12H ϵ has NOEs with Glu8H β , Glu8H γ , Lys16H γ , Lys16H δ , and Lys16H ϵ , consistent with close contacts between Glu8, Trp12, and Lys16 in **pA-EWK**, centered on the right-hand side of the Trp12 indole ring as drawn in Figure 2-8. Moreover, chemical shifts of the Glu8 H γ protons and of the Lys16 H γ , H δ , and H ϵ protons appear upfield ($\Delta\delta = -0.15$ to -0.30) relative to corresponding protons in random coil reference compounds,³⁴ suggesting that the Glu8 H γ protons and the Lys16 H γ , H δ , and H ϵ protons are within the shielding region of the Trp12 indole ring. These observations indicate that in **pA-EWK** the Glu8 and Lys16 side chains are both on opposing sides of and in close proximity to the Trp12 indole.

We further explored the structural impact of the interaction by preparing peptides **pA-AWK** and **pA-EWA** (Figure 2-8), in which Ala replaces Glu8 or Lys16, respectively. As with **pA-EWK**, the ROESY spectrum for **pA-AWK** shows many NOEs between Trp12 and Lys16 backbone and side-chain protons. However, several important differences between the ROESY spectra of **pA-EWK** vs. **pA-AWK** highlight the impact of Glu8 vs. Ala8 on Trp-Lys contacts (Figure 2-8). For example, NOE peaks corresponding to Trp12H ζ 3/Lys16H α , Trp12H ζ 3/Lys16H δ , and Trp12H η 2/Lys16H δ are present in the spectrum of **pA-AWK** but not in that of **pA-EWK**. In contrast, NOE peaks corresponding to Trp12H ϵ 1/Lys16H γ , Trp12H ϵ 1/Lys16H δ , and Trp12H ϵ 1/Lys16H ϵ are present in the spectrum of **pA-EWK** but not in that of **pA-AWK**. Moreover, normalized peak volumes for the Trp12H ϵ 3/Lys16H δ , Trp12H α /Lys16NH, and Trp12 H δ 1/Lys16H δ NOEs are larger for **pA-EWK** than for **pA-AWK**, whereas normalized peak volumes for the Trp12H δ 1/Lys16H ϵ and Trp12H ζ 2/Lys16H ϵ NOEs are smaller for **pA-EWK** than for **pA-AWK**. These changes suggest that replacing Ala8 with Glu8

allows Lys16H δ to move closer to H δ 1, H ϵ 1, and H ϵ 3 on the right-hand side of the Trp12 indole ring (as drawn in Figure 2-8), whereas Lys16NH moves closer to Trp12H α , possibly due to localized overwinding of the α -helical backbone between Trp12 and Lys16 to allow for optimized interaction geometry between Glu8, Trp12, and Lys16.

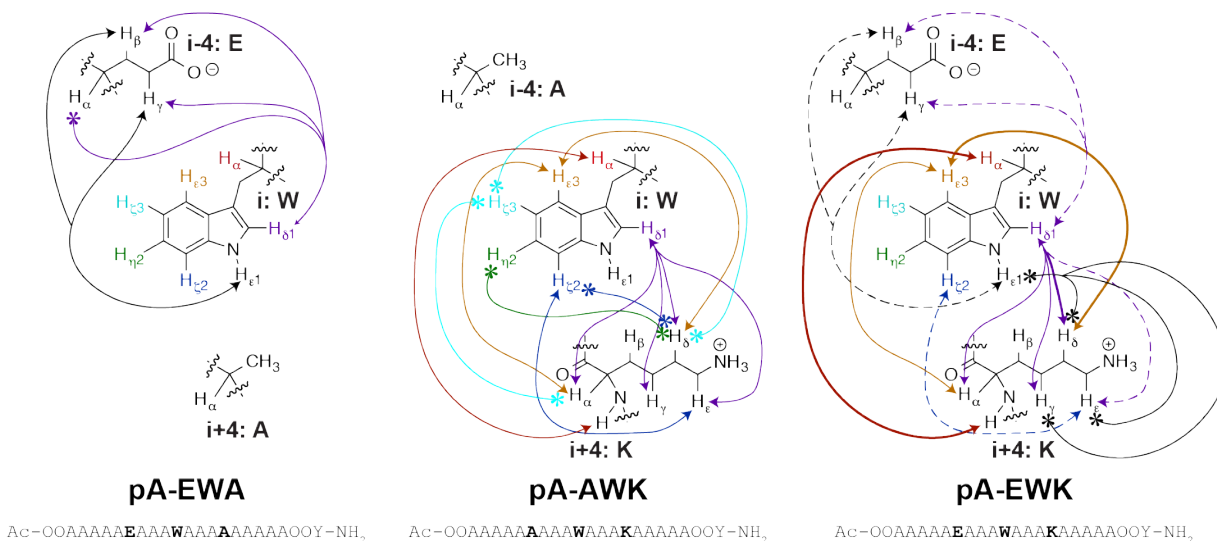


Figure 2-9. Observed NOEs for monomeric α -helical peptides pA-EWK, pA-AWK, and pA-EWA in 20 mM sodium phosphate buffer (pH 7). Lines tipped with asterisks indicate NOEs that are only present in the indicated peptide. Dotted lines indicate NOEs that are weaker in pA-EWK than in pA-EWA or pA-AWK. Bold lines indicate NOEs that are stronger in pA-EWK than in pA-EWA or pA-AWK.

Similarly, differences between the ROESY spectra of pA-EWK vs. pA-EWA highlight impact of Lys16 vs. Ala16 on Glu8-Trp12 contacts. The Trp12H δ 1/Glu8H α NOE is present in the spectrum of pA-EWA, but not that of pA-EWK. Moreover, normalized peak volumes for the Trp12H δ 1/Glu8H β , Trp12H δ 1/Glu8H γ , Trp12H ϵ 1/Glu8H β , and Trp12H ϵ 1/Glu8H γ NOEs are smaller in pA-EWK than in pA-AWK. These observations indicate that the Glu8 side-chain protons are close to H δ 1 and H ϵ 1 of Trp12 (as is also the case for H δ and H ϵ of Lys16, see above), but that replacing Ala16 with Lys16 increases the distance between the Glu8 side chain and the

right side of the Trp12 indole ring (as drawn in Figure 2-8), possibly to compensate for the localized overwinding of the α -helical backbone between Trp12 and Lys16.

2.2.5 Exploring the Long-Range Salt-bridge within a dimeric helix bundle.

We wondered whether the ability of an $i+4$ aromatic amino acid to enhance the stability of a long-range salt bridge between an i -position Glu and an $i+8$ position Lys might also apply in other α -helices in addition to the homotrimeric model system described above. We decided to explore this possibility in the context of the homodimeric α -helical coiled coil GCN4-p1. In the native GCN4-p1 sequence, Lys18 already occupies the $i+8$ -position relative to Glu10, with Ser14 at $i+4$; all three of these residues lie along the solvent-exposed surface of the coiled-coil homodimer, providing an ideal context for assessing the impact of a non-polar residue at position 14 on the long-range interaction between Glu10 and Lys18. However, we worried that Glu11 (adjacent to Glu10) might interfere with our ability to characterize the Glu10- Lys18 interaction. Consequently, in preparing peptide **22** (in which Glu, Phe, and Lys occupy positions 10, 14, and 18, respectively), we also mutated Glu11 to Ala. Circular dichroism (CD) experiments and size-exclusion chromatography experiments indicate that these mutations do not substantially disrupt the homodimeric α -helical coiled-coil quaternary structure of **20** relative to GCN4-p1 (see supporting info 2.4.4). We assessed the impact of Phe14 on the Glu10-Lys18 interaction by replacing Glu10 with Ala; Phe14 with Ser; and/or Lys18 with Ala, in all possible combinations. Comparing the folding free energies of **22-ESK**, **22-ASK**, **22-ESA**, and **22-ASA** reveals that Glu10 and Lys18 do not interact favorably ($\Delta\Delta\Delta G_f = 0.59 \pm 0.09$ kcal/mol) in the presence of Ser14. In contrast with our previous observations described in the main text, placing Phe at

position 14 in the GCN4 system does not substantially change the already unfavorable interaction between Glu10 and Lys18 ($\Delta\Delta\Delta G_f = 0.65 \pm 0.07$ kcal/mol), suggesting that Phe14 does not enhance a long-range salt-bridge between Glu10 and Lys18 (Table 2-2).

Table 2-1 Enhancement of Glu:Lys long range salt bridge in the GCN4-p1 Dimeric helix.

Peptide	Sequence	Salt-bridge with Ser14 $\Delta\Delta\Delta G_f$ (kcal/mol)	Salt-bridge with Phe14 $\Delta\Delta\Delta G_f$ (kcal/mol)	Influence of Phe14 $\Delta\Delta\Delta\Delta G_f$ (kcal/mol)
22	Ac-•••V E A L E F K N Y K L•••-CONH ₂	0.59 ± 0.09	0.65 ± 0.07	0.06 ± 0.12
23	Ac-•••V A E L E F K N Y K L•••-CONH ₂	1.20 ± 0.07	-0.38 ± 0.07	-1.58 ± 0.10

^aData are given ± standard error at 30 μM protein concentration in 20 mM sodium phosphate buffer (pH 7) at 333.15 K.

It is possible that subtle geometric differences between the homodimeric system and the homotrimeric system described above might move Glu10 far enough away from Phe14 and Lys18 as to prevent any Phe-based enhancement in of a long-range Glu10-Lys18 interaction in **22**. Position 11 occupies the same solvent-exposed face of the helix as Glu10, Phe14, and Lys18 but should be closer to Phe14 and Lys18 than position 10. We wondered whether placing Glu at position 11 instead of position 10, might facilitate a favorable long-range i to i+7 salt bridge between Glu11 and Lys18 in the presence of Phe14. To test this hypothesis, we prepared peptide **23**, in which Glu, Phe, and Lys occupy positions 11, 14, and 18, respectively, with Ala at position 10 instead of Glu (again, to avoid interference of Glu10 with any possible interaction among Glu11, Phe14, and Lys18). We also prepared variants of 21 in which we replaced Glu11 with Ala; Phe14 with Ser; and/or Lys18 with Ala, in all possible combinations (names and sequences of these variants are shown in supporting information section 2.4.1; some of these combinations were already accounted for in peptides **22-AFK**, **22-AFA**, **22 ASK**, and **22-ASA**). Comparing the folding free energies of peptides **23-ESK**, **23-ESA**, **22-ASK**, and **22-ASA**

reveals that Glu11 and Lys18 do not interact favorably in the presence of Ser14 ($\Delta\Delta\Delta G_f = 1.20 \pm 0.07$ kcal/mol). However, in the presence of Phe14 (compare peptides **21-EFK**, **21-EFA**, **21-AFK**, and **21-AFA**), the Glu11-Lys18 interaction is substantially favorable ($\Delta\Delta\Delta G_f = -0.38 \pm 0.07$ kcal/mol), a dramatic Phe-based shift of -1.58 ± 0.10 kcal/mol per trimer (figure 2-9).

2.3 Conclusions

Salt bridges are critical to protein stability and to protein function. Through double and triple mutant cycle analysis we have shown that a long-range salt bridge can stabilize the conformation of a helix bundle if an aromatic residue is positioned between the ions. The enhanced salt bridge interaction, in the context of a helix, is strongly dependent on the precise location and orientation of the ionic side-chains and involves a significant hydrogen-bonding component. The three-way interaction is also independent of the electronic configuration of the aromatic side chain. Crystal structures of the interaction show that the trimer packs in a dense crystal lattice and solvent exposed positions of the trimer are likely to be involved in crystal contacts with other trimers and therefore the side chain conformations of positions 10, 14, and 18 must be interpreted with caution and might not correlate to solution thermodynamic data. In the crystal structures, it appears that the Lys18 is packing closely to the arene at position 14 while the Glu 10 from one trimer is packing close to the arene14 of an adjacent trimer. Although, not relevant to solution behavior, it is tempting to speculate that this is potentially an enhanced salt bridge interaction between trimers in the crystal lattice. Solution NMR data for a monomeric pA-EWK helix shows several NOE's between a the Glu8 and Trp12, and between the Lys16 and Trp12, indicating that the ionic side chains are close to the Trp side chain although the exact geometry of the side chains is still unclear.

2.4 Supporting Information

2.4.1 Protein Synthesis, Purification, and Characterization

Peptides **pA-EWK**, **pA-AWK**, **pA-EWA**, and **1-27**, and their sequence variants (sequences shown in Supplementary Table 1) were synthesized as C-terminal amides, by microwave-assisted solid-phase peptide synthesis. Fmoc-protected amino acids were activated by 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) and N-hydroxybenzotriazole hydrate (HOBT), all purchased from Advanced ChemTech. NovaSyn TGR resin was purchased from EMD Biosciences. Peptides were synthesized on a 12.5 μmol scale. A general protocol for manual solid-phase peptide synthesis follows: NovaSyn TGR resin (52.1 mg, 12.5 μmol at 0.24 mmol/g resin loading) was aliquotted into a fritted polypropylene syringe and allowed to swell first in CH_2Cl_2 , and then in dimethylformamide (DMF). Solvent was drained from the resin using a vacuum manifold. To remove the Fmoc protecting group on the resin-linked amino acid, 0.625 mL of 20% piperidine in DMF was added to the resin, and the resulting mixture was allowed to sit at room temperature for 1 minute. The deprotection solution was then drained from the resin with a vacuum manifold. An additional 1.25 mL of 20% piperidine in DMF was then added to the resin, and the reaction vessel was placed in the microwave. The temperature was ramped from room temperature to 80°C over the course of 2 minutes and held at 80°C for 2 minutes. The deprotection solution was drained from the resin using a vacuum manifold, and the resin was rinsed five times with DMF. The Fmoc deprotection solution was changed after an aspartate residue was coupled to avoid aspartimide formation. The modified deprotection solution (0.625 mL of 5% m/v piperazine + 0.1 M HOBT in DMF) was added to the resin, the resulting mixture was allowed to sit at room temperature for 1 minute. The deprotection solution was drained from the resin with a vacuum manifold. An additional 0.625

mL of modified deprotection solution was then added to the resin, and the reaction vessel was placed in the microwave. The temperature was ramped from room temperature to 75°C over the course of 2 minutes and held at 75°C for 3 minutes.

For coupling of an activated amino acid, we prepared a stock coupling solution of 100 mL NMP, 3.17 g HBTU (0.01 mol, 0.1 M) and 1.53 g HOBt (0.01 mol, 0.1 M) for a final concentration of 0.1 M HBTU and 0.1 M HOBt. The desired Fmoc-protected amino acid (125 µmol, 5 eq) was dissolved by vortexing in 1.25 mL coupling solution (125 µmol, 5 eq HBTU; 125 µmol, 5 eq HOBt). To the dissolved amino acid solution was added 44 µL DIEA (250 µmol, 10 eq). The resulting mixture was vortexed briefly and allowed to react for at least 1 min. The activated amino acid solution was then added to the resin, and the reaction vessel was placed in the microwave. The temperature was ramped from room temperature to 70°C over 2 minutes and held at 70°C for 4 minutes. Following the coupling reaction, the activated amino acid solution was drained from the resin with a 2 vacuum manifold, and the resin was subsequently rinsed five times with DMF. The cycles of deprotection and coupling were alternately repeated to give the desired full-length protein.

Acid-labile side-chain protecting groups were globally removed and proteins were cleaved from the resin by stirring the resin for ~4h in a solution of phenol (0.0625 g), water (62.5 µL), thioanisole (62.5 µL), ethanedithiol (31 µL) and triisopropylsilane (12.5 µL) in trifluoroacetic acid (TFA, 1 mL). Following the cleavage reaction, the TFA solution was drained from the resin, the resin was rinsed with additional TFA. Proteins were precipitated from the TFA solution by addition of diethyl ether (~40 mL). Following centrifugation, the ether was decanted, and the pellet was dissolved in ~40mL 1:1 H₂O/MeCN, frozen and lyophilized to remove volatile impurities. The resulting powder was stored at -20°C until purification.

Immediately prior to purification, the crude protein was dissolved in 1:1 H₂O/MeCN. Proteins were purified by preparative reverse-phase high performance liquid chromatography (HPLC) on a C18 column using a linear gradient of water in acetonitrile with 0.1% v/v TFA. Fractions containing the desired protein product were pooled, frozen, and lyophilized. Proteins were identified by electrospray ionization time of flight mass spectrometry (ESI-TOF); expected and observed exact masses mass spectra appear in Table 1. Protein purity was assessed by Analytical HPLC.

Table 2-2. Sequences for peptides 1-27, pA-EWK, pA-AWK, pA-EWA and their sequence derivatives.

Peptide	Sequence
1CW	Ac-EVEALEKKV A ALECKVQ A LEKKVEALEHGWDGR-CONH ₂
1	Ac-EVEALEKKV E ALEWKVQ K LEKKVEALEHGWDGR-CONH ₂
2	Ac-EVEALEKKV A ALEWKVQ K LEKKVEALEHGWDGR-CONH ₂
3	Ac-EVEALEKKV E ALEWKVQ A LEKKVEALEHGWDGR-CONH ₂
4	Ac-EVEALEKKV A ALEWKVQ A LEKKVEALEHGWDGR-CONH ₂
5	Ac-EVEALEKKV E ALESKVQ K LEKKVEALEHGWDGR-CONH ₂
6	Ac-EVEALEKKV A ALESKVQ K LEKKVEALEHGWDGR-CONH ₂
7	Ac-EVEALEKKV E ALESKVQ A LEKKVEALEHGWDGR-CONH ₂
8	Ac-EVEALEKKV A ALESKVQ A LEKKVEALEHGWDGR-CONH ₂
9	Ac-EVEALEKKV D ALEWKVQ K LEKKVEALEHGWDGR-CONH ₂
9-DWA	Ac-EVEALEKKV D ALEWKVQ A LEKKVEALEHGWDGR-CONH ₂
9-DSK	Ac-EVEALEKKV D ALESKVQ K LEKKVEALEHGWDGR-CONH ₂
9-DSA	Ac-EVEALEKKV D ALESKVQ A LEKKVEALEHGWDGR-CONH ₂
10	Ac-EVEALEKKV E ALEWKVQ O LEKKVEALEHGWDGR-CONH ₂
10-AWO	Ac-EVEALEKKV A ALEWKVQ O LEKKVEALEHGWDGR-CONH ₂
10-ESO	Ac-EVEALEKKV E ALESKVQ O LEKKVEALEHGWDGR-CONH ₂
10-ASO	Ac-EVEALEKKV A ALESKVQ O LEKKVEALEHGWDGR-CONH ₂
11	Ac-EVEALEKKV A ELEWKVQ K LEKKVEALEHGWDGR-CONH ₂
11-EWA	Ac-EVEALEKKV A ELEWKVQ A LEKKVEALEHGWDGR-CONH ₂
11-ESK	Ac-EVEALEKKV A ELESKVQ K LEKKVEALEHGWDGR-CONH ₂
11-ESA	Ac-EVEALEKKV A ELESKVQ A LEKKVEALEHGWDGR-CONH ₂
12	Ac-EVEALEKKV K ALEWKVQ E LEKKVEALEHGWDGR-CONH ₂
12-KWA	Ac-EVEALEKKV K ALEWKVQ A LEKKVEALEHGWDGR-CONH ₂
12-AWE	Ac-EVEALEKKV A ALEWKVQ E LEKKVEALEHGWDGR-CONH ₂
12-KSE	Ac-EVEALEKKV K ALESKVQ E LEKKVEALEHGWDGR-CONH ₂

12-KSA	Ac-EVEALEKKV K ALESKVQ A LEKKVEALEHGW DGR-CONH ₂
12-ASE	Ac-EVEALEKKV A ALESKVQ E LEKKVEALEHGW DGR-CONH ₂
13	Ac-EVEALEKKV E ALEW _{me} KVQ K LEKKVEALEHGW DGR-CONH ₂
13-AW _{me} K	Ac-EVEALEKKV A ALEW _{me} KVQ K LEKKVEALEHGW DGR-CONH ₂
13-EW _{me} A	Ac-EVEALEKKV E ALEW _{me} KVQ A LEKKVEALEHGW DGR-CONH ₂
13-AW _{me} A	Ac-EVEALEKKV A ALEW _{me} KVQ A LEKKVEALEHGW DGR-CONH ₂
14	Ac-EVEALEKKV E ALEYKVQ K LEKKVEALEHGW DGR-CONH ₂
14-AYK	Ac-EVEALEKKV A ALEYKVQ K LEKKVEALEHGW DGR-CONH ₂
14-EYA	Ac-EVEALEKKV E ALEYKVQ A LEKKVEALEHGW DGR-CONH ₂
14-AYA	Ac-EVEALEKKV A ALEYKVQ A LEKKVEALEHGW DGR-CONH ₂
15	Ac-EVEALEKKV E ALEY _{me} KVQ K LEKKVEALEHGW DGR-CONH ₂
15-AY _{me} K	Ac-EVEALEKKV A ALEY _{me} KVQ K LEKKVEALEHGW DGR-CONH ₂
15-EY _{me} A	Ac-EVEALEKKV E ALEY _{me} KVQ A LEKKVEALEHGW DGR-CONH ₂
15-AY _{me} A	Ac-EVEALEKKV A ALEY _{me} KVQ A LEKKVEALEHGW DGR-CONH ₂
16	Ac-EVEALEKKV E ALEFKVQ K LEKKVEALEHGW DGR-CONH ₂
16-AFK	Ac-EVEALEKKV A ALEFKVQ K LEKKVEALEHGW DGR-CONH ₂
16-EFA	Ac-EVEALEKKV E ALEFKVQ A LEKKVEALEHGW DGR-CONH ₂
16-AFA	Ac-EVEALEKKV A ALEFKVQ A LEKKVEALEHGW DGR-CONH ₂
17	Ac-EVEALEKKV E ALE(f ₅ F)KVQ K LEKKVEALEHGW DGR-CONH ₂
17-A(f ₅ F)K	Ac-EVEALEKKV A ALE(f ₅ F)KVQ K LEKKVEALEHGW DGR-CONH ₂
17-E(f ₅ F)A	Ac-EVEALEKKV E ALE(f ₅ F)KVQ A LEKKVEALEHGW DGR-CONH ₂
17-A(f ₅ F)A	Ac-EVEALEKKV A ALE(f ₅ F)KVQ A LEKKVEALEHGW DGR-CONH ₂
18	Ac-EVEALEKKV E ALE(P-fF)KVQ K LEKKVEALEHGW DGR-CONH ₂
18-A(p-fF)K	Ac-EVEALEKKV A ALE(P-fF)KVQ K LEKKVEALEHGW DGR-CONH ₂
18-E(p-fF)A	Ac-EVEALEKKV E ALE(P-fF)KVQ A LEKKVEALEHGW DGR-CONH ₂
18-A(p-fF)A	Ac-EVEALEKKV A ALE(P-fF)KVQ A LEKKVEALEHGW DGR-CONH ₂
19	Ac-EVEALEKKV E ALE(3,4-f ₂ F)KVQ K LEKKVEALEHGW DGR-CONH ₂
19-A(3,4-f ₂ F)K	Ac-EVEALEKKV A ALE(3,4-f ₂ F)KVQ K LEKKVEALEHGW DGR-CONH ₂
19-E(3,4-f ₂ F)A	Ac-EVEALEKKV E ALE(3,4-f ₂ F)KVQ A LEKKVEALEHGW DGR-CONH ₂
19-A(3,4-f ₂ F)A	Ac-EVEALEKKV A ALE(3,4-f ₂ F)KVQ A LEKKVEALEHGW DGR-CONH ₂
20	Ac-EVEALEKKV E ALE(P-NO ₂ F)KVQ K LEKKVEALEHGW DGR-CONH ₂
20-A(p-NO ₂ F)K	Ac-EVEALEKKV A ALE(P-NO ₂ F)KVQ K LEKKVEALEHGW DGR-CONH ₂
20-E(p-NO ₂ F)A	Ac-EVEALEKKV E ALE(P-NO ₂ F)KVQ A LEKKVEALEHGW DGR-CONH ₂
20-A(p-NO ₂ F)A	Ac-EVEALEKKV A ALE(P-NO ₂ F)KVQ A LEKKVEALEHGW DGR-CONH ₂
20	Ac-EVEALEKKV E ALE(PyrA)KVQ K LEKKVEALEHGW DGR-CONH ₂
20-A(PyrA)K	Ac-EVEALEKKV A ALE(PyrA)KVQ K LEKKVEALEHGW DGR-CONH ₂
20-E(PyrA)A	Ac-EVEALEKKV E ALE(PyrA)KVQ A LEKKVEALEHGW DGR-CONH ₂
20-A(PyrA)A	Ac-EVEALEKKV A ALE(PyrA)KVQ A LEKKVEALEHGW DGR-CONH ₂
22	Ac-RMKQLEDRV E ALEFKNY K LENEVARLKKLVGER-CONH ₂
22-AFK	Ac-RMKQLEDRV A ALEFKNY K LENEVARLKKLVGER-CONH ₂
22-EFA	Ac-RMKQLEDRV E ALEFKNY A LENEVARLKKLVGER-CONH ₂
22-AFA	Ac-RMKQLEDRV A ALEFKNY A LENEVARLKKLVGER-CONH ₂
22-ESK	Ac-RMKQLEDRV E ALESKNY K LENEVARLKKLVGER-CONH ₂
22-ASK	Ac-RMKQLEDRV A ALESKNY K LENEVARLKKLVGER-CONH ₂
22-ESA	Ac-RMKQLEDRV E ALESKNY A LENEVARLKKLVGER-CONH ₂

22-ASA	Ac-RMKQLEDRVA A LESKNY A LENEVARLKKLVGER-CONH ₂
23	Ac-RMKQLEDRVA E LEFKNY K LENEVARLKKLVGER-CONH ₂
23-EFA	Ac-RMKQLEDRVA E LEFKNY A LENEVARLKKLVGER-CONH ₂
23-ESK	Ac-RMKQLEDRVA E LESKNY K LENEVARLKKLVGER-CONH ₂
23-ESA	Ac-RMKQLEDRVA E LESKNY A LENEVARLKKLVGER-CONH ₂
24	Ac-EVEALEKKV E ALEFKV Q RLEKKVEALEHGW DGR-CONH ₂
24-AFR	Ac-EVEALEKKV A ALEFKV Q RLEKKVEALEHGW DGR-CONH ₂
24-ESR	Ac-EVEALEKKV E ALESKV Q RLEKKVEALEHGW DGR-CONH ₂
24-ASR	Ac-EVEALEKKV A ALESKV Q RLEKKVEALEHGW DGR-CONH ₂
25	Ac-EVEALEKKV E ALEAKV Q KLEKKVEALEHGW DGR-CONH ₂
25-AAK	Ac-EVEALEKKV A ALEAKV Q KLEKKVEALEHGW DGR-CONH ₂
25-EAA	Ac-EVEALEKKV E ALEAKV A LEKKVEALEHGW DGR-CONH ₂
25-AAA	Ac-EVEALEKKV A ALEAKV A LEKKVEALEHGW DGR-CONH ₂
26	Ac-EVEALEKKV A ALEEKV Q ALEKKVEALEHGW DGR-CONH ₂
27	Ac-EVEALEKKV A ALEEKV Q KLEKKVEALEHGW DGR-CONH ₂
pA-EWK	Ac-OOAAAA A EA A WAAA K AAAA A OOY-CONH ₂
pA-EWA	Ac-OOAAAA A EA A WAAA A AAAA A OOY-CONH ₂
pA-AWK	Ac-OOAAAA A AA A WAAA K AAAA A OOY-CONH ₂

2.4.2 ESI-TOF data

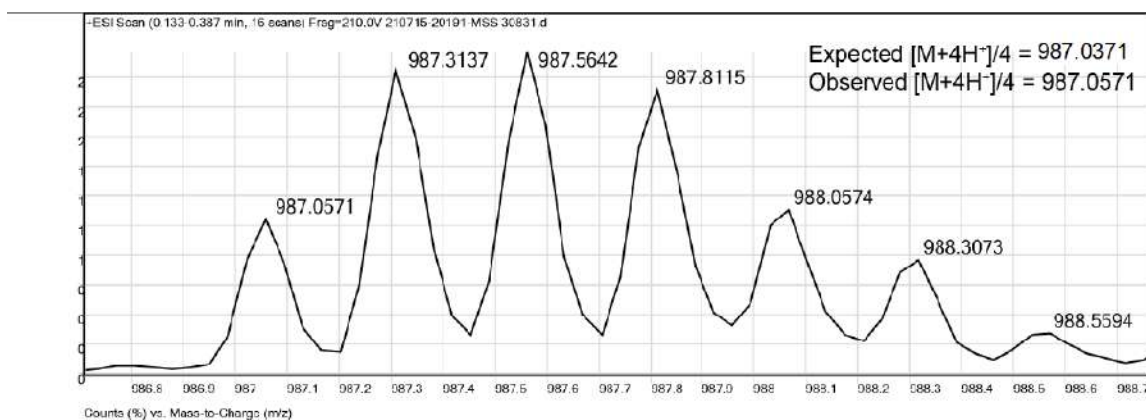


Figure 2-10. ESI TOF spectrum for peptide 1

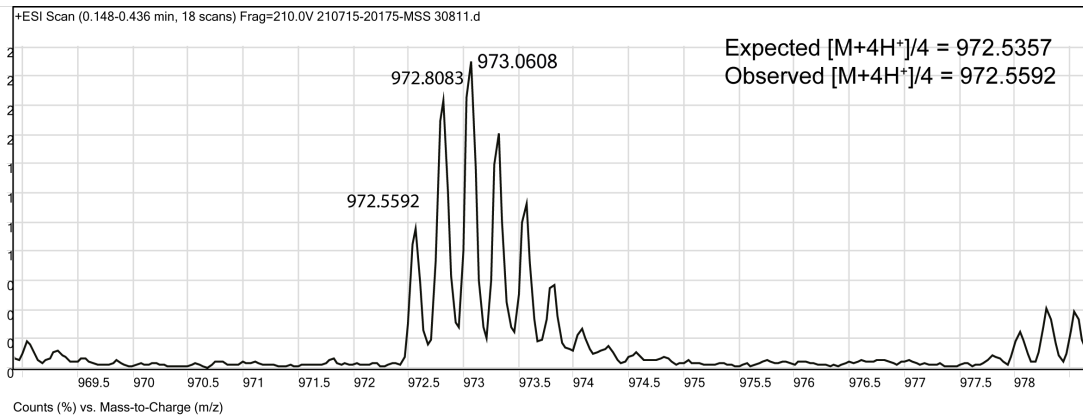


Figure 2-11. ESI TOF spectrum for peptide 2

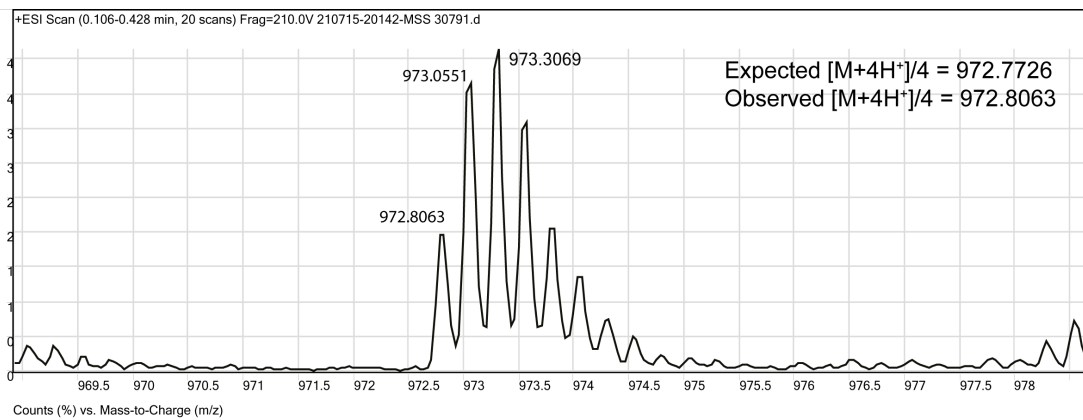


Figure 2-12. ESI TOF spectrum for peptide 3

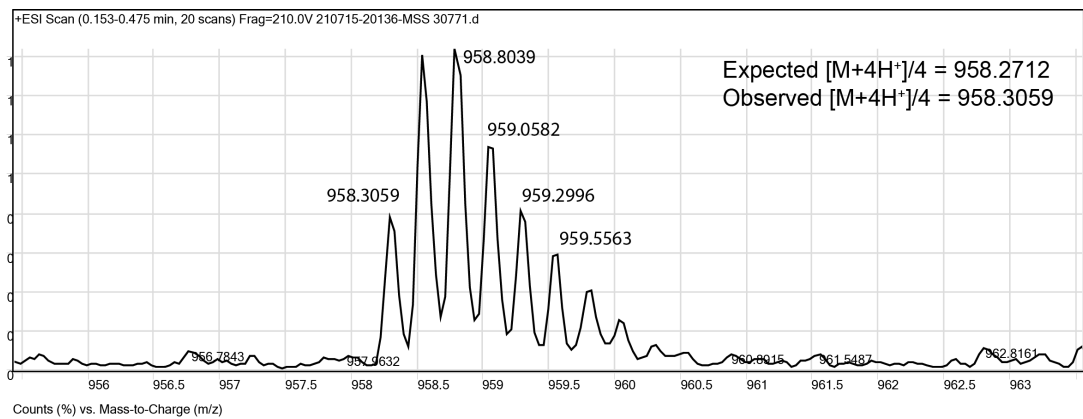


Figure 2-13. ESI TOF spectrum for peptide 4.

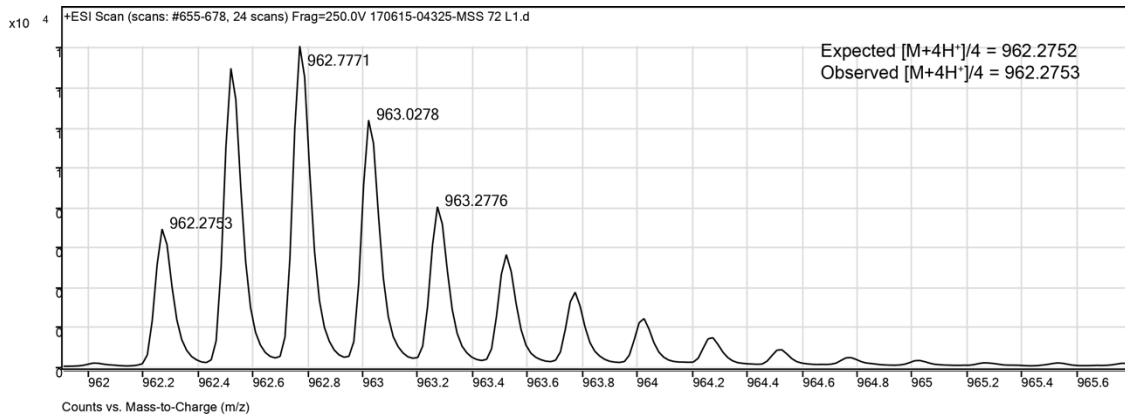


Figure 2-14. ESI TOF spectrum for peptide 5.

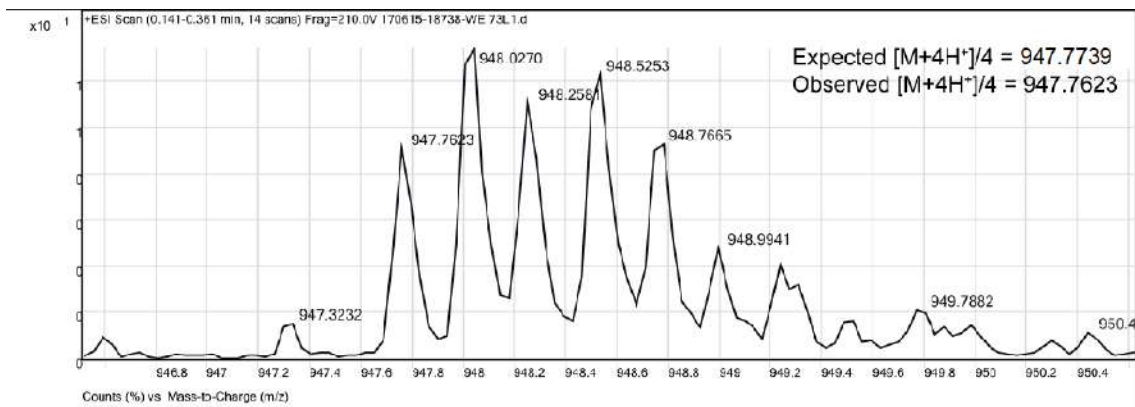


Figure 2-15. ESI TOF spectrum for peptide 6.

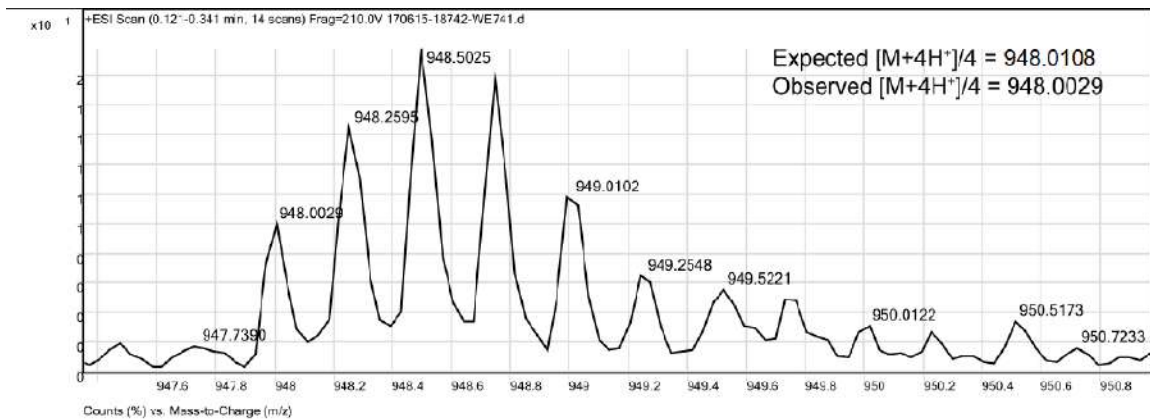


Figure 2-16. ESI TOF spectrum for peptide 7.

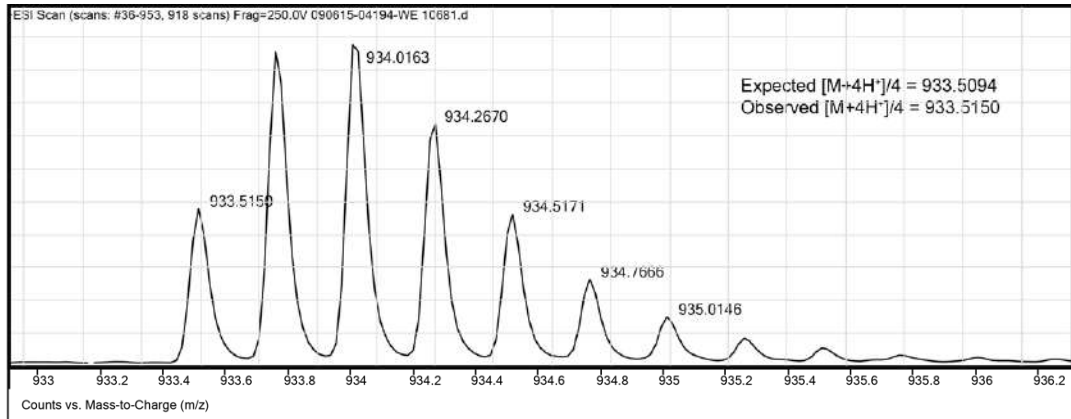


Figure 2-17. ESI TOF spectrum for peptide **8**.

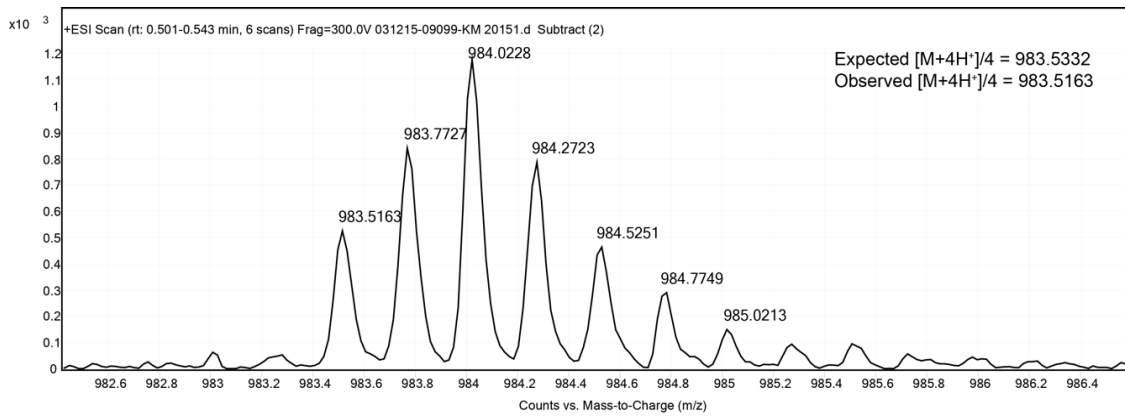


Figure 2-18. ESI TOF spectrum for peptide **9**.

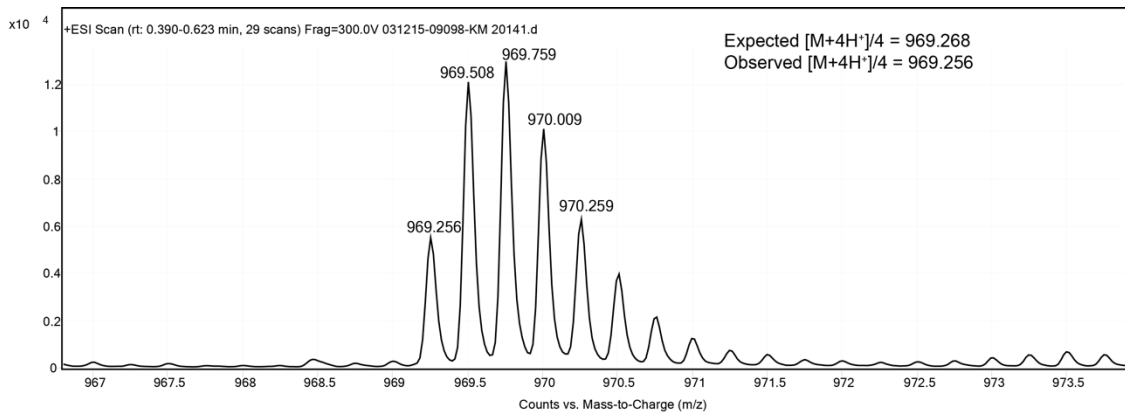


Figure 2-19. ESI TOF spectrum for peptide **9-DWA**

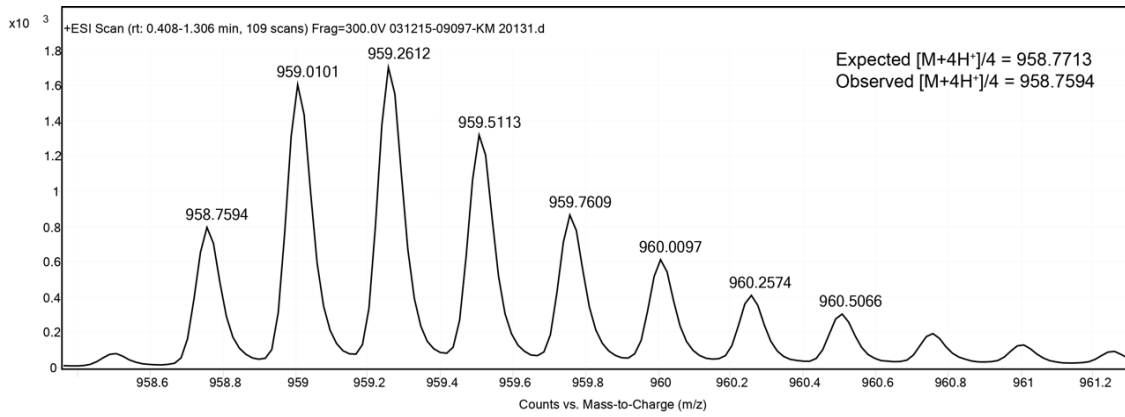


Figure 2-20. ESI TOF spectrum for peptide **9-DSK**.

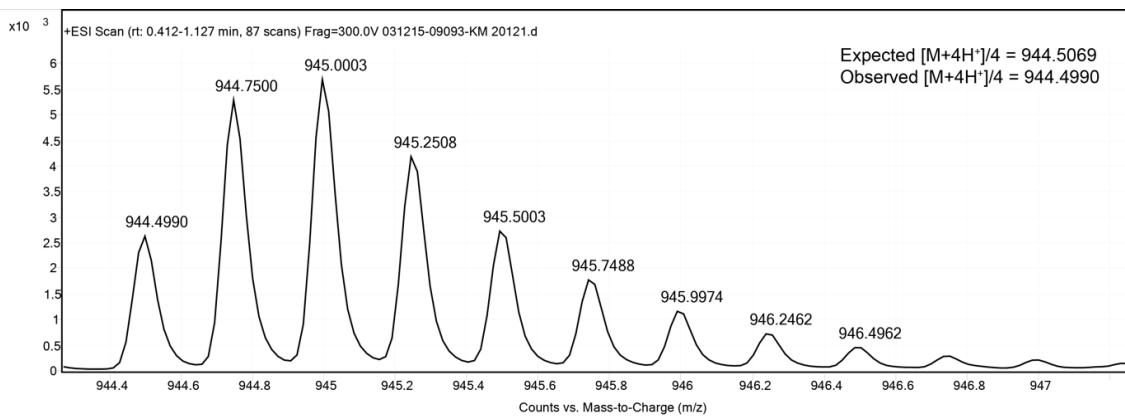


Figure 2-21. ESI TOF spectrum for peptide **9-DSA**.

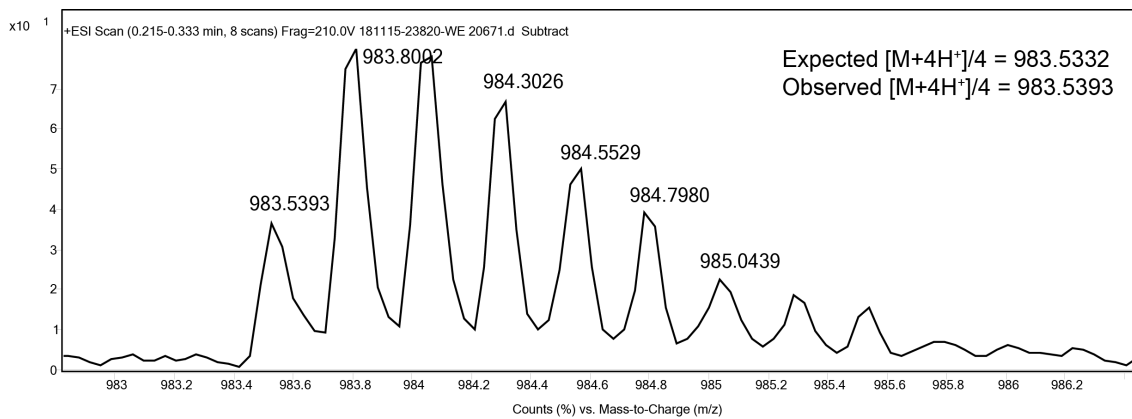


Figure 2-22. ESI TOF spectrum for peptide **10**.

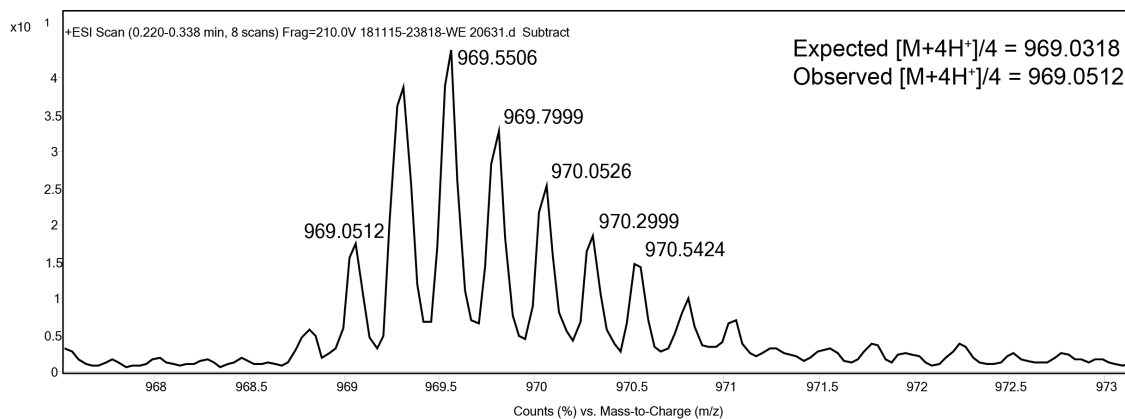


Figure 2-23. ESI TOF spectrum for peptide **10-AWO**.

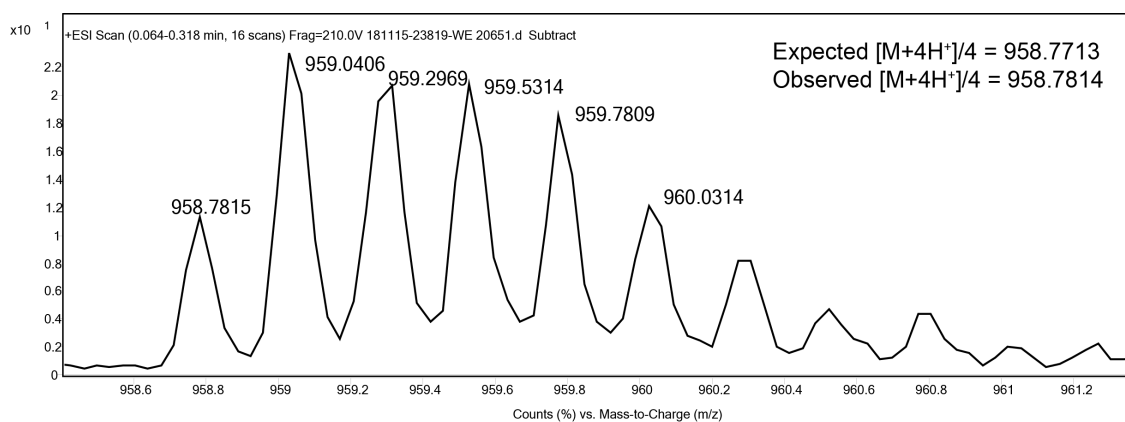


Figure 2-24. ESI TOF spectrum for peptide **10-ESO**.

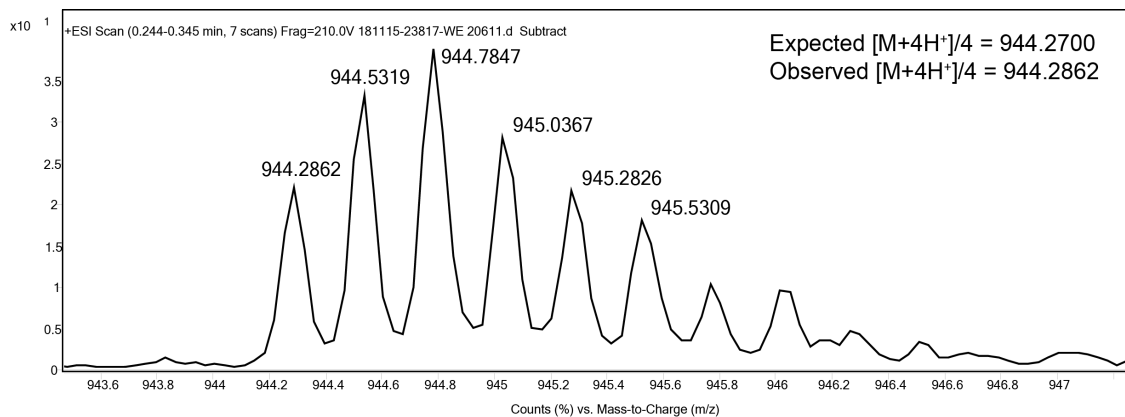


Figure 2-25. ESI TOF spectrum for peptide **10-ASO**.

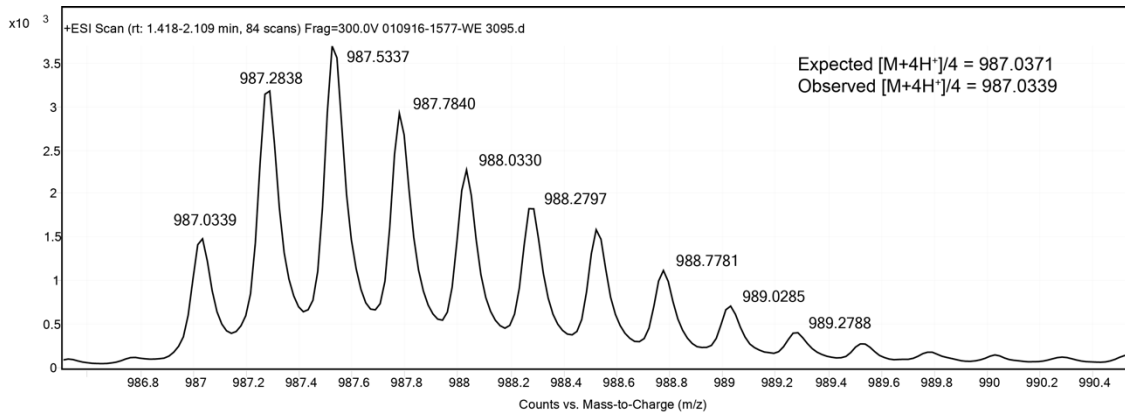


Figure 2-26. ESI TOF spectrum for peptide **11**.

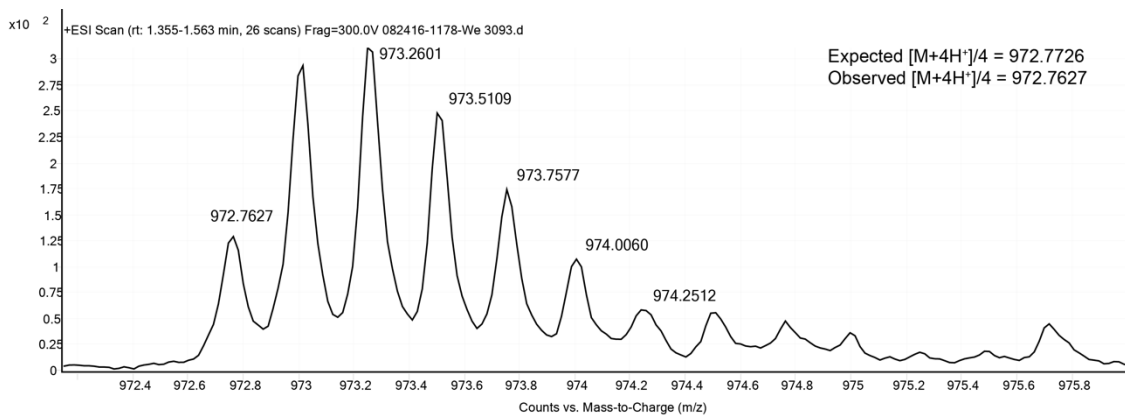


Figure 2-27. ESI TOF spectrum for peptide **11-EWA**.

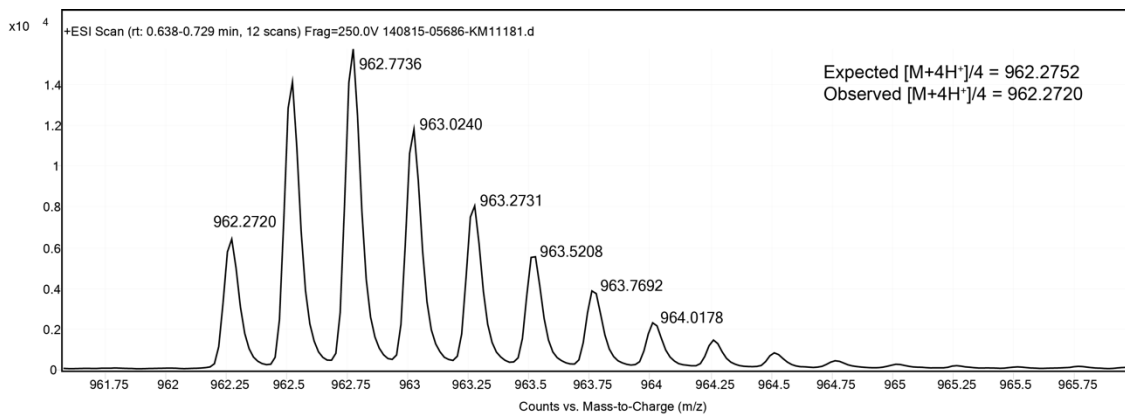


Figure 2-28. ESI TOF spectrum for peptide **11-ESK**.

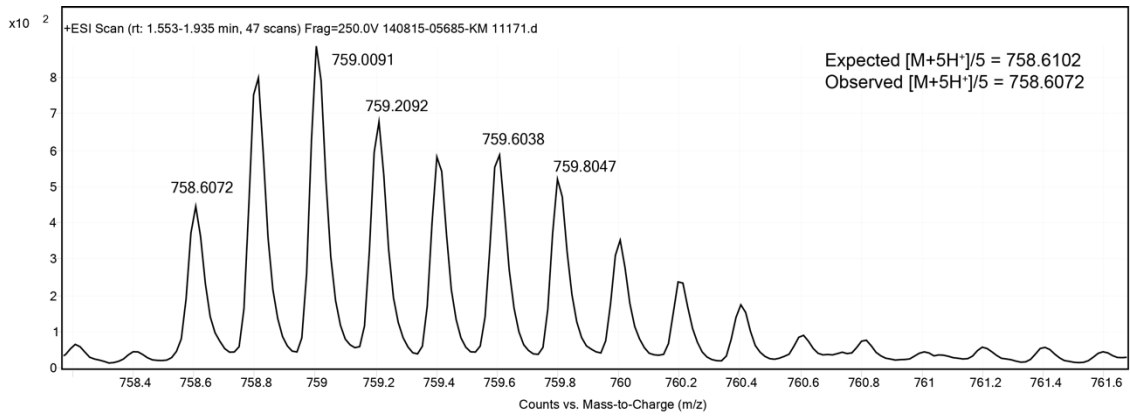


Figure 2-29. ESI TOF spectrum for peptide 11-ESA.

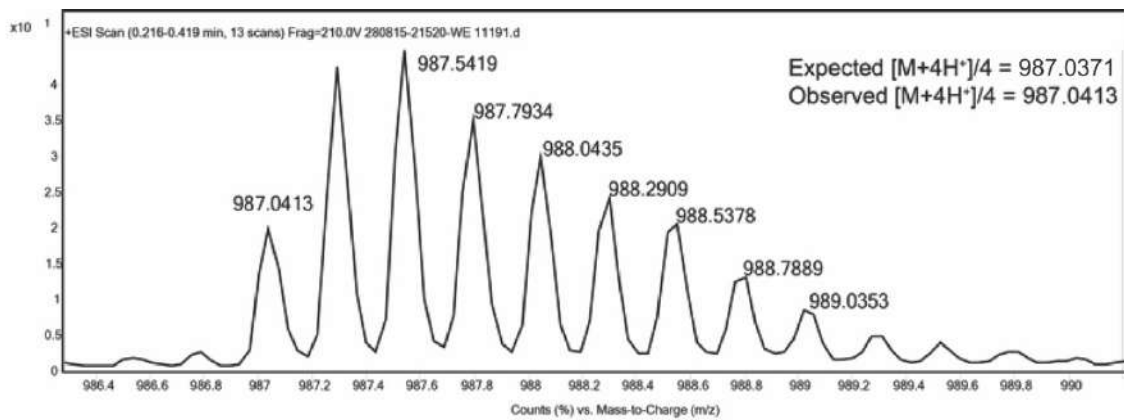


Figure 2-30. ESI TOF spectrum for peptide 12.

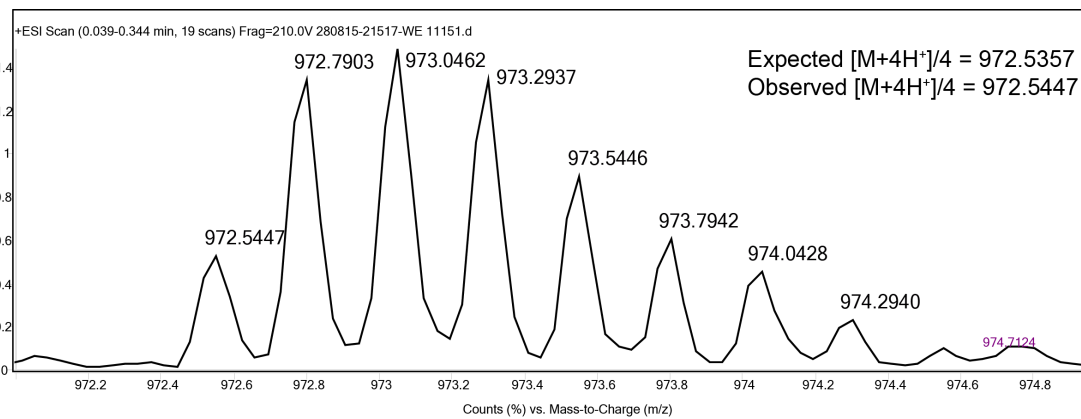


Figure 2-31. ESI TOF spectrum for peptide 12-KWA.

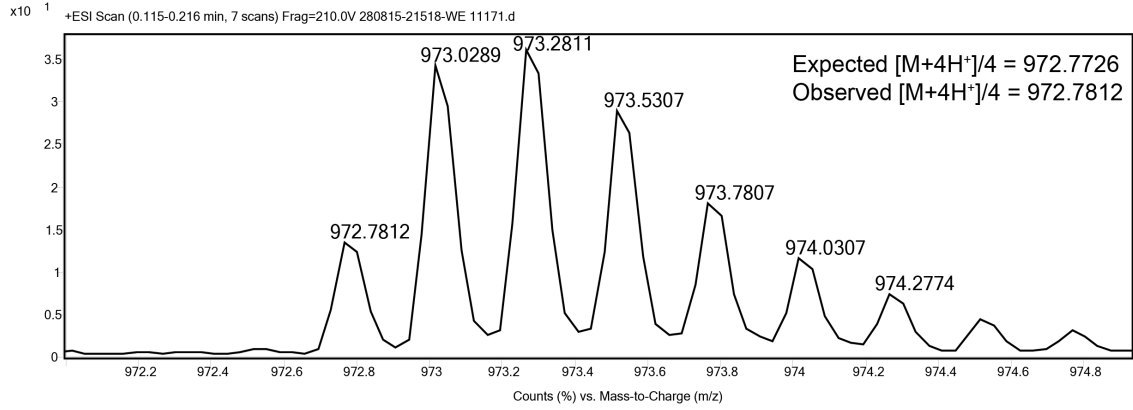


Figure 2-32. ESI TOF spectrum for peptide 12-AWE.

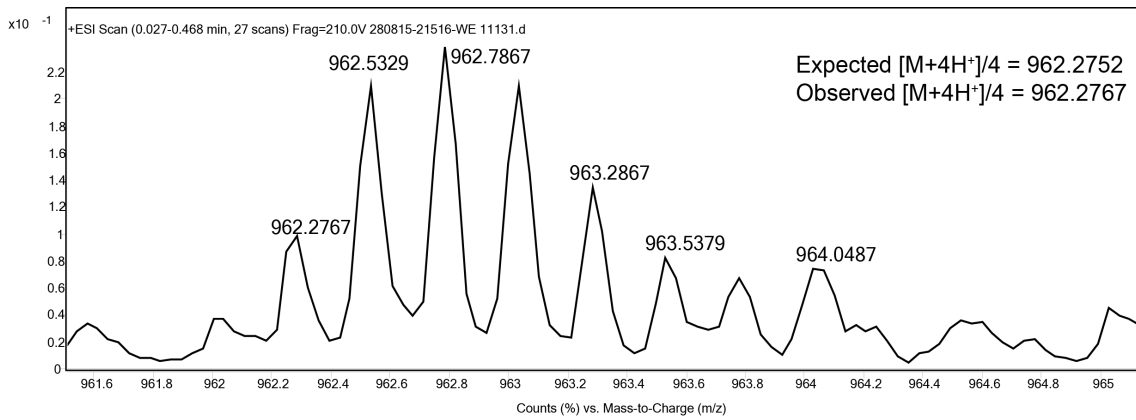


Figure 2-33. ESI TOF spectrum for 12-KSE.

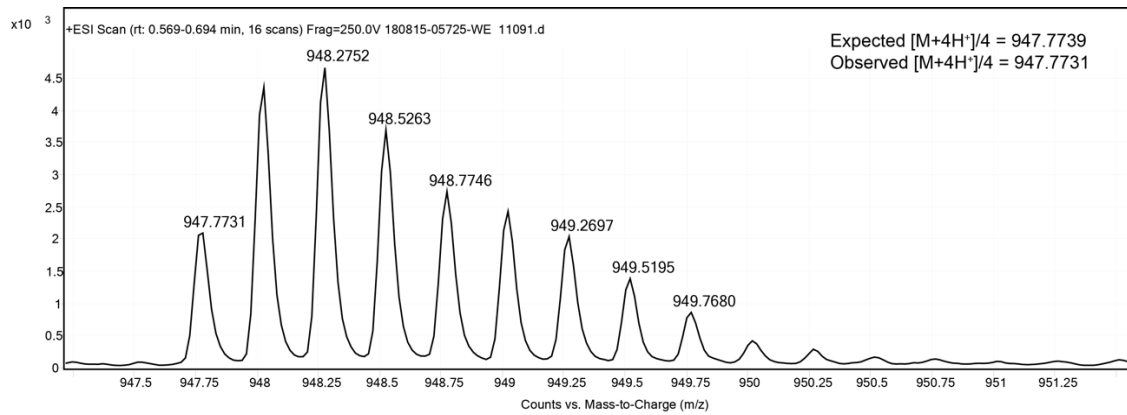


Figure 2-34. ESI TOF spectrum for 12-KSA.

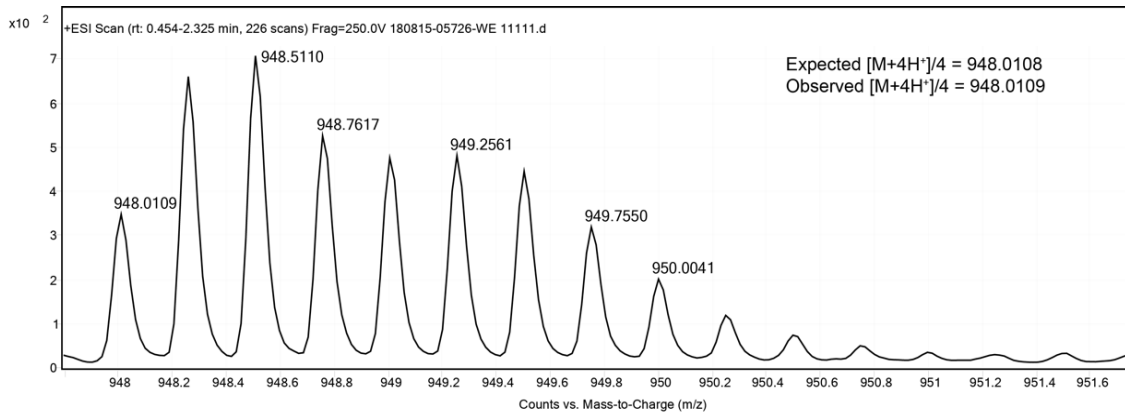


Figure 2-35. ESI TOF spectrum for **12-ASE**.

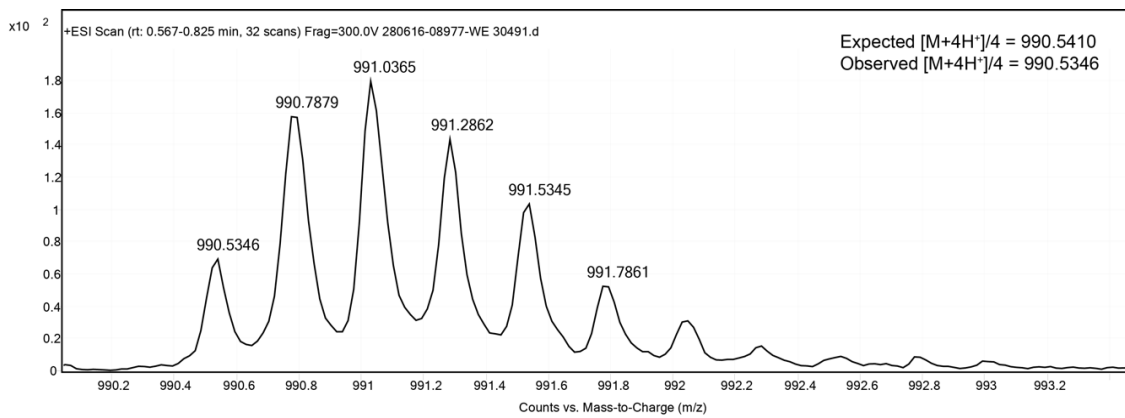


Figure 2-36. ESI TOF spectrum for peptide **13**.

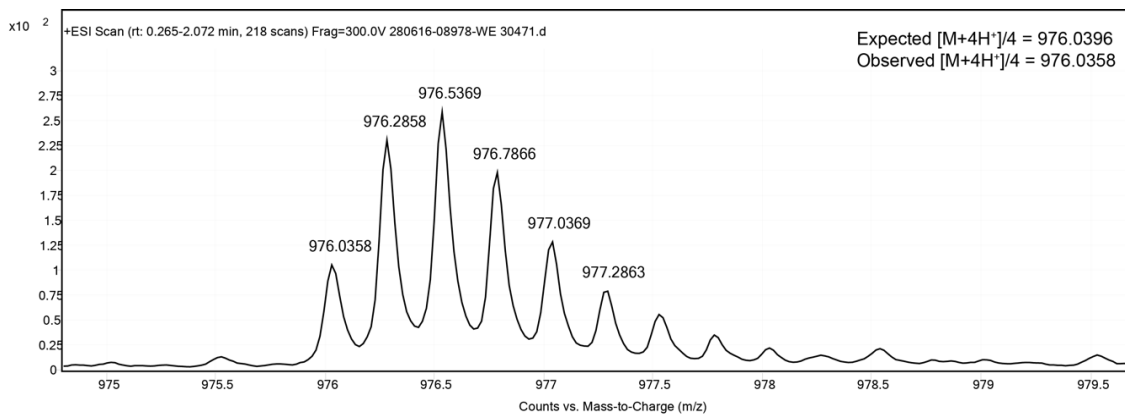


Figure 2-37. ESI TOF spectrum for peptide **13-AW_{me}K**.

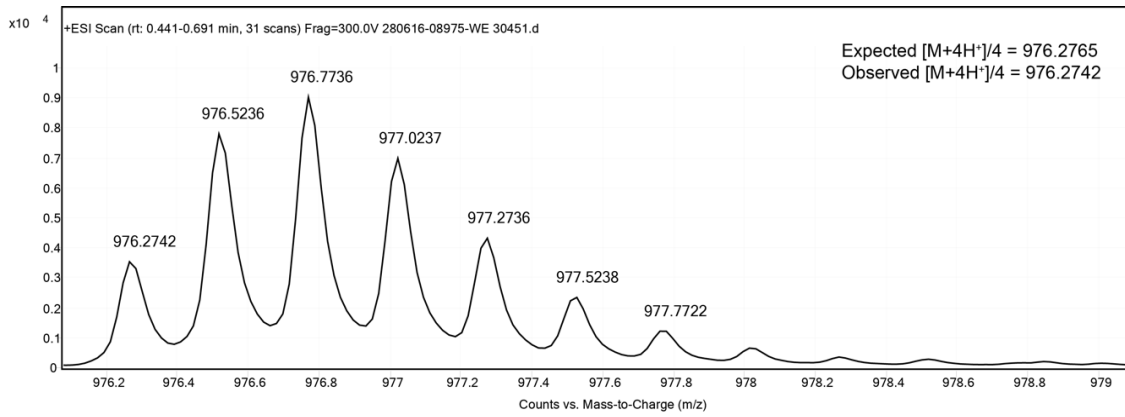


Figure 2-38. ESI TOF spectrum for peptide **13-EW_{meA}**.

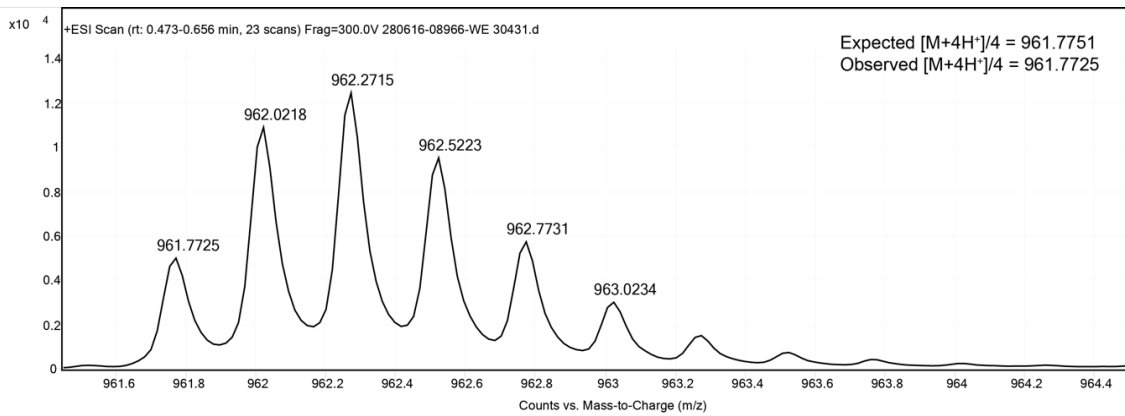


Figure 2-39. ESI TOF spectrum for peptide **13-AW_{meA}**.

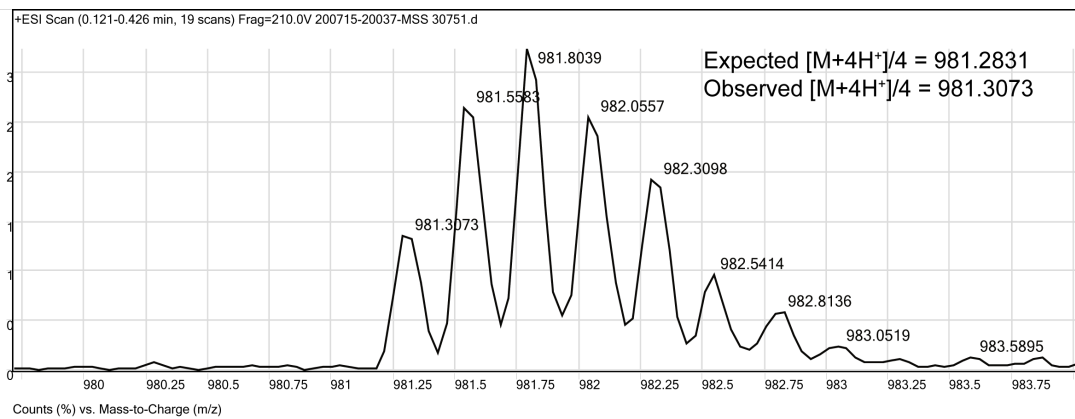


Figure 2-40. ESI TOF spectrum for peptide **14**.

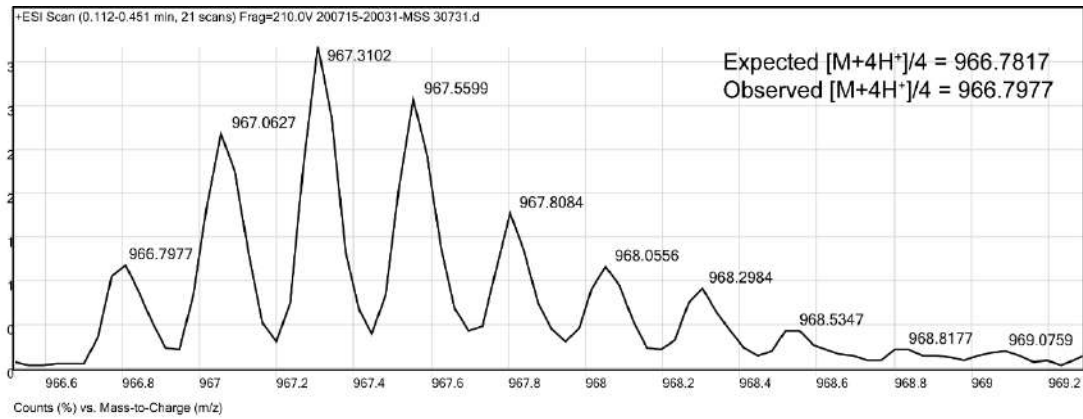


Figure 2-41. ESI TOF spectrum for peptide **14-AYK**.

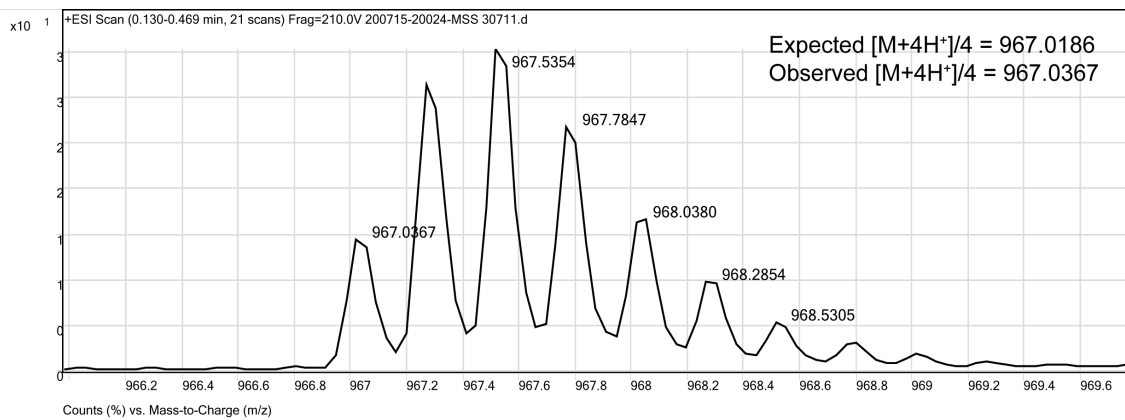


Figure 2-42. ESI TOF spectrum for peptide **14-EYA**.

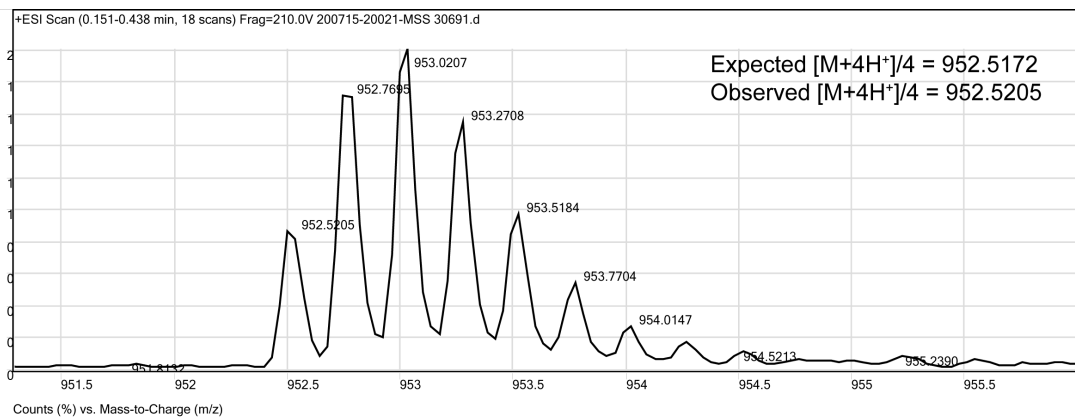


Figure 2-43. ESI TOF spectrum for peptide **14-AYA**.

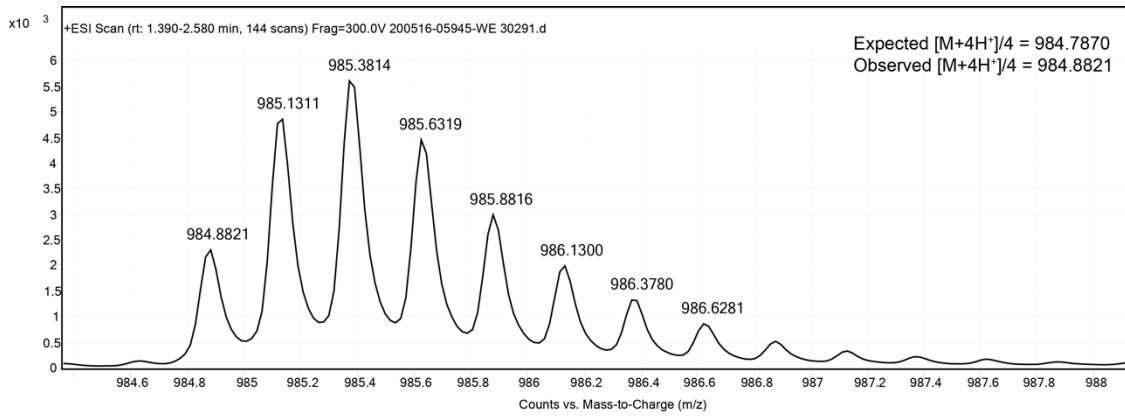


Figure 2-44. ESI TOF spectrum for peptide **15**.

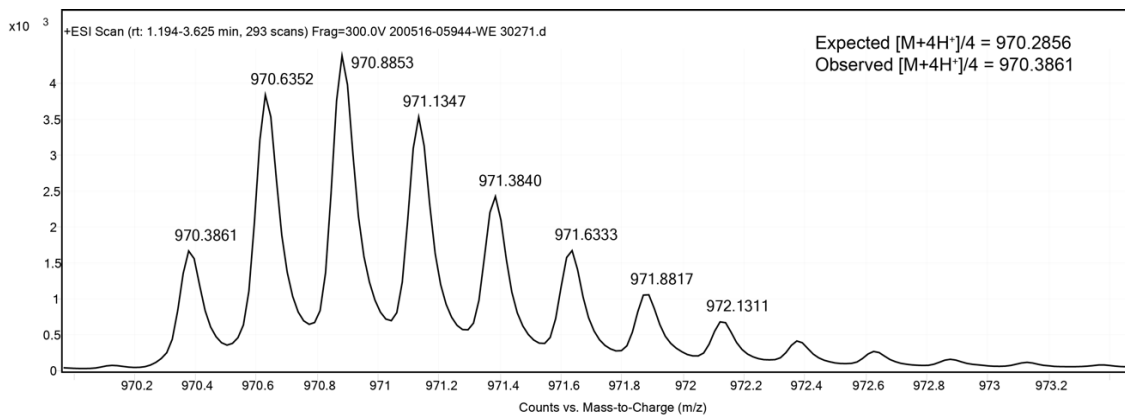


Figure 2-45. ESI TOF spectrum for peptide **15-AY_{me}K**.

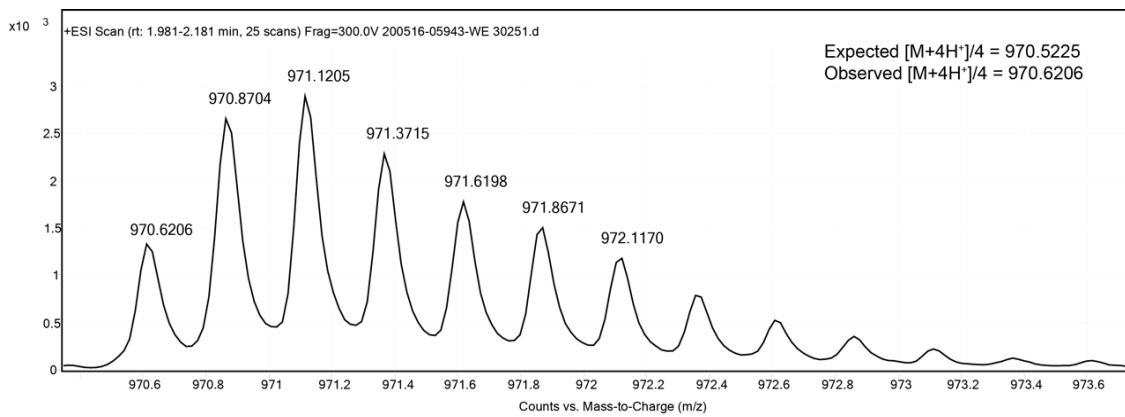


Figure 2-46. ESI TOF spectrum for peptide **15-EY_{me}A**.

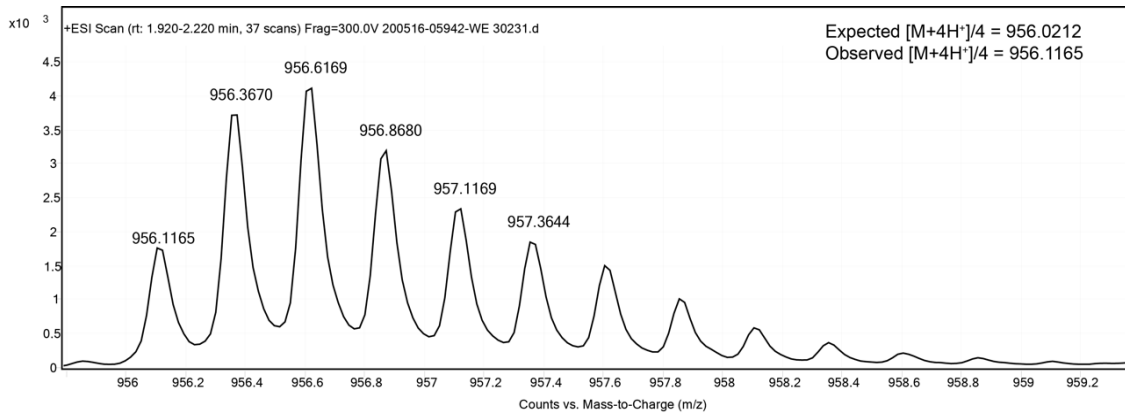


Figure 2-47. ESI TOF spectrum for peptide 15-AY_{mc}A.

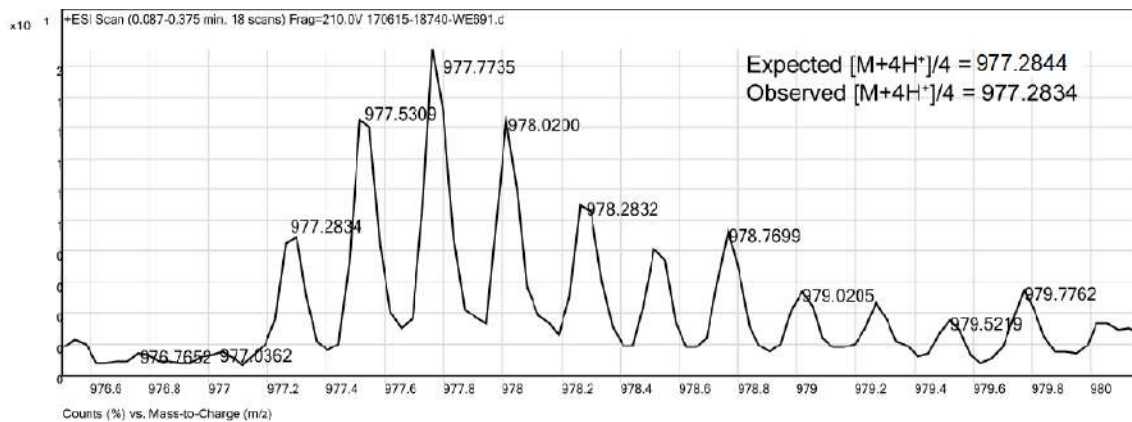


Figure 2-48. ESI TOF spectrum for peptide 16.

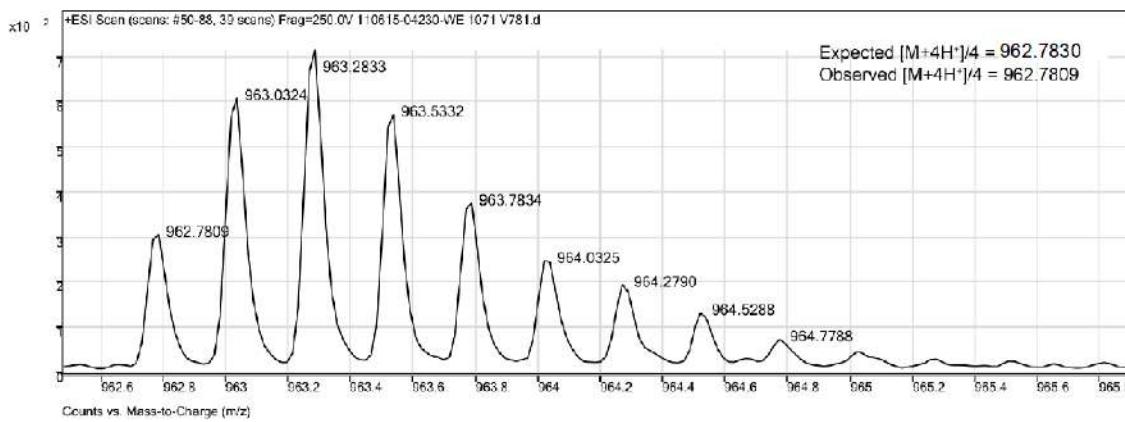


Figure 2-49. ESI TOF spectrum for peptide 16-AFK.

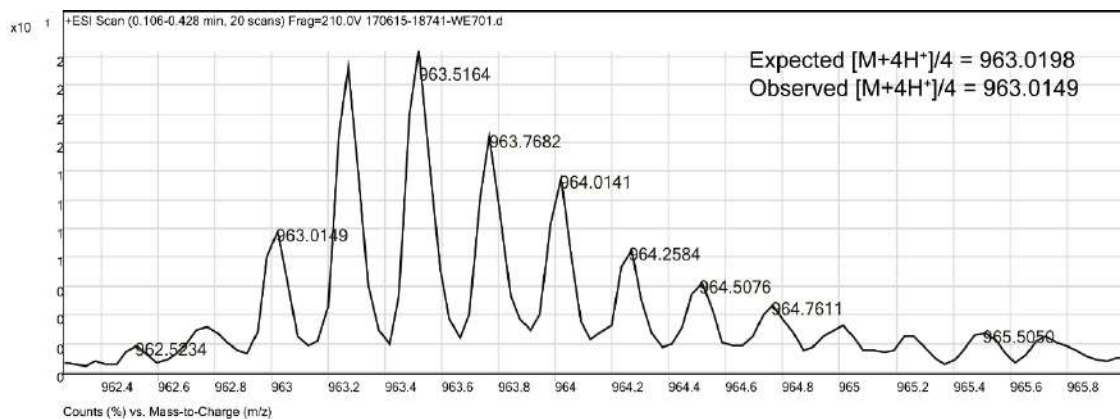


Figure 2-50. ESI TOF spectrum for peptide 16-EFA.

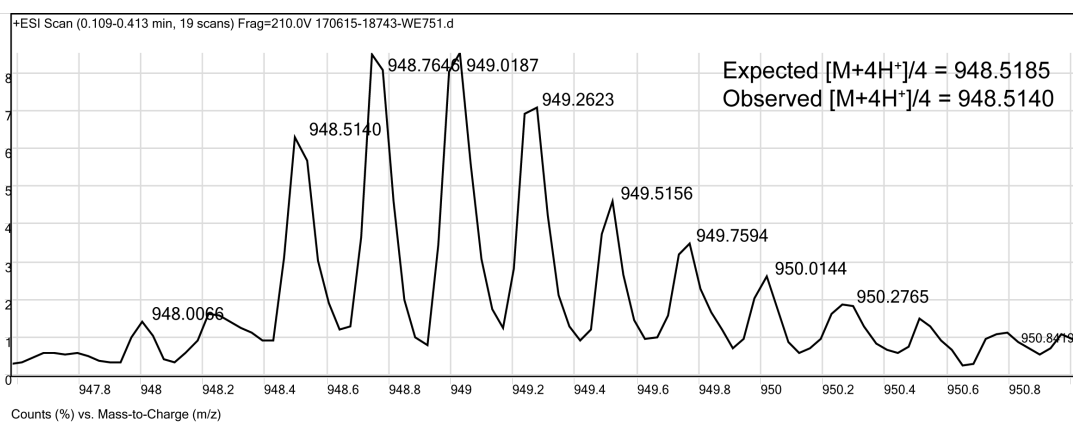


Figure 2-51. ESI TOF spectrum for peptide 16-AFA.

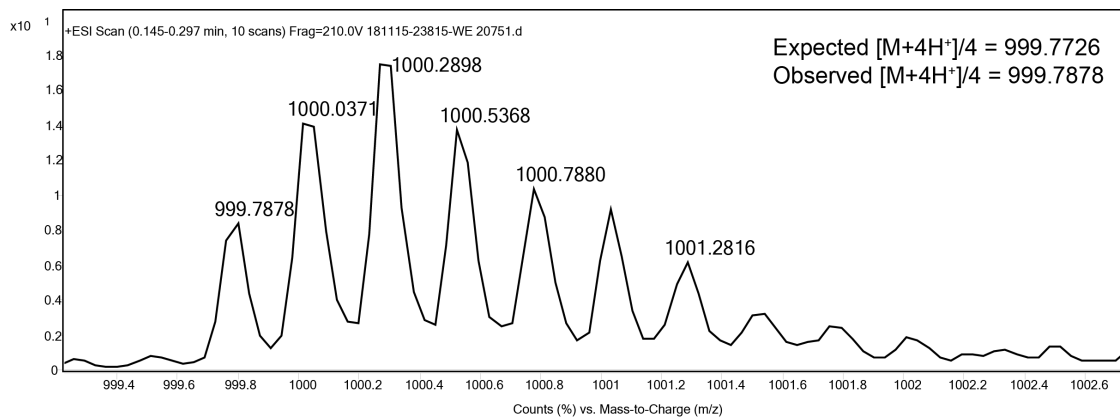


Figure 2-52. ESI TOF spectrum for peptide 17.

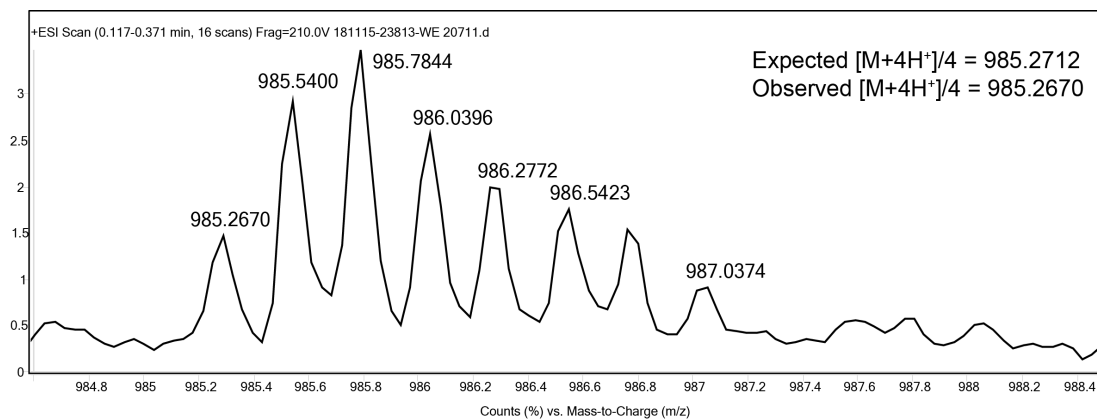


Figure 2-53. ESI TOF spectrum for peptide **17-A(f₅Phe)K**.

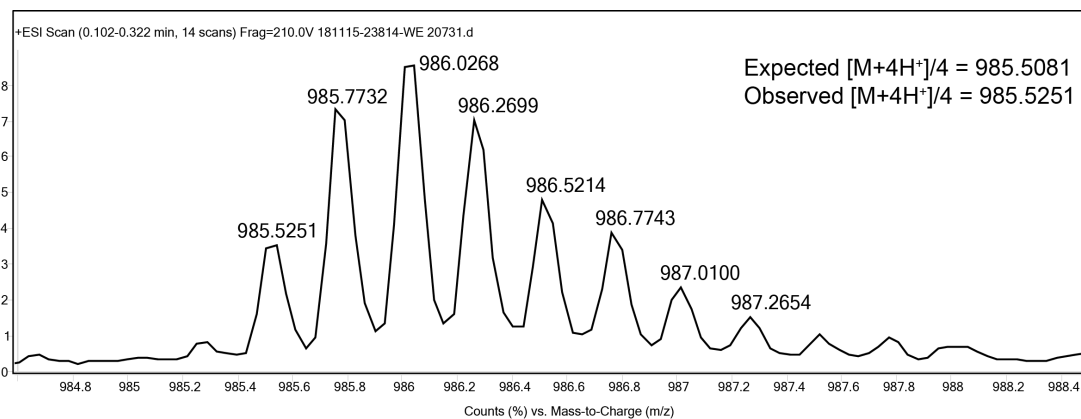


Figure 2-54. ESI TOF spectrum for peptide **17-E(f₅Phe)A**.

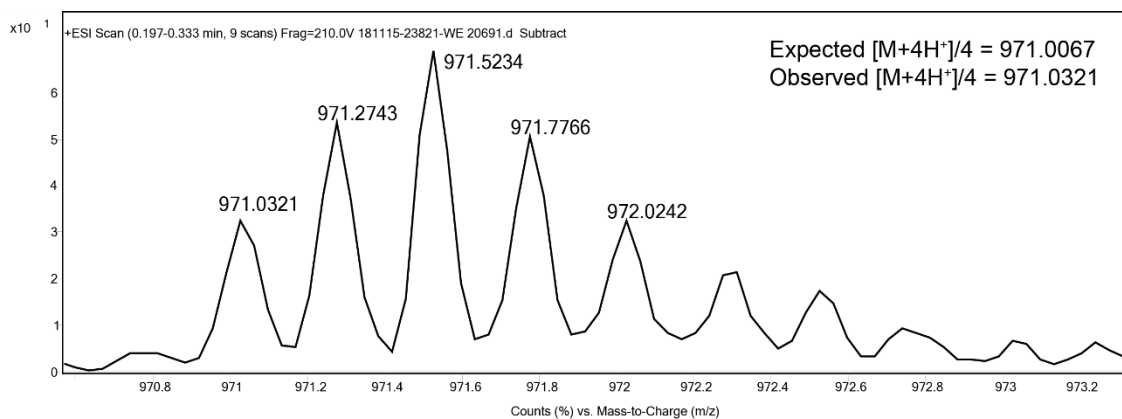


Figure 2-55. ESI TOF spectrum for peptide **17-A(f₅Phe)A**.

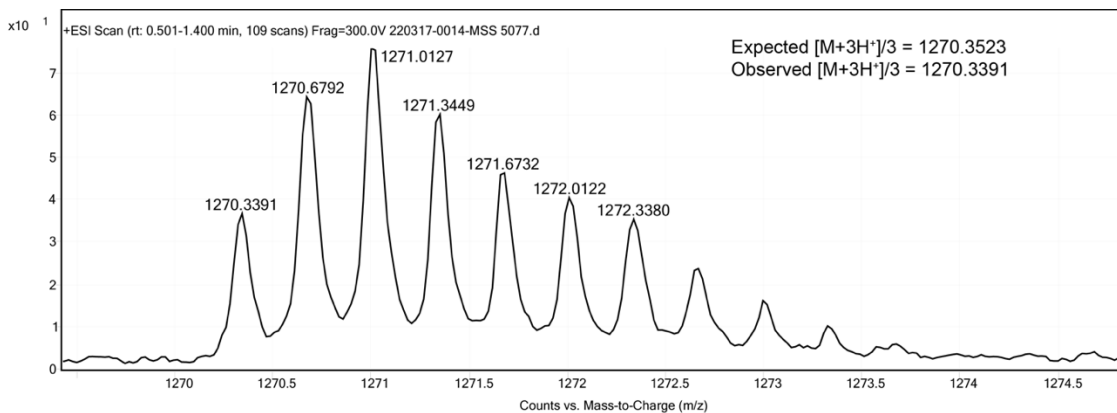


Figure 2-56. ESI TOF spectrum for 18-A(p-fF)A.

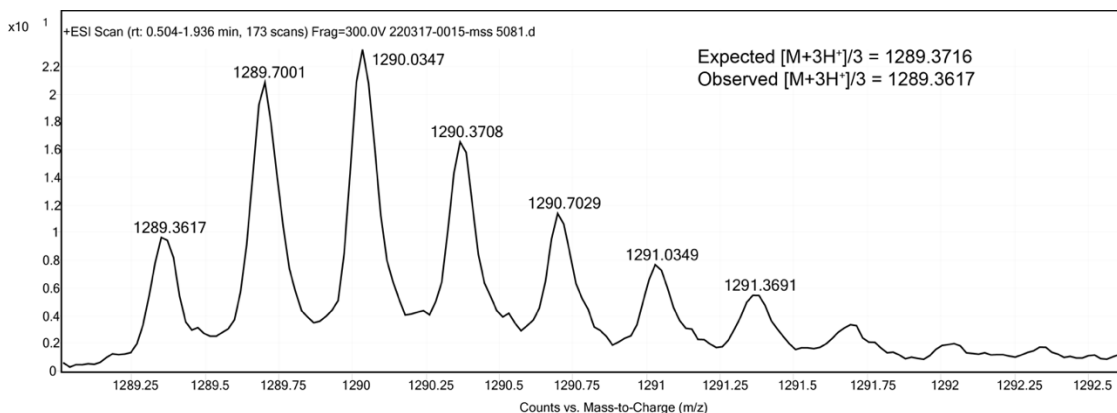


Figure 2-57. ESI TOF spectrum for 18-A(p-fF)K.

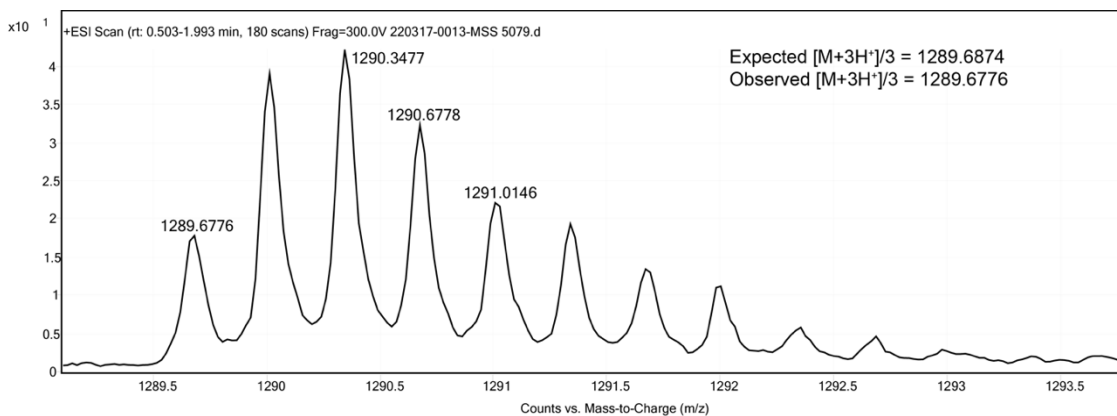


Figure 2-58. ESI TOF spectrum for 18-E(p-fF)A.

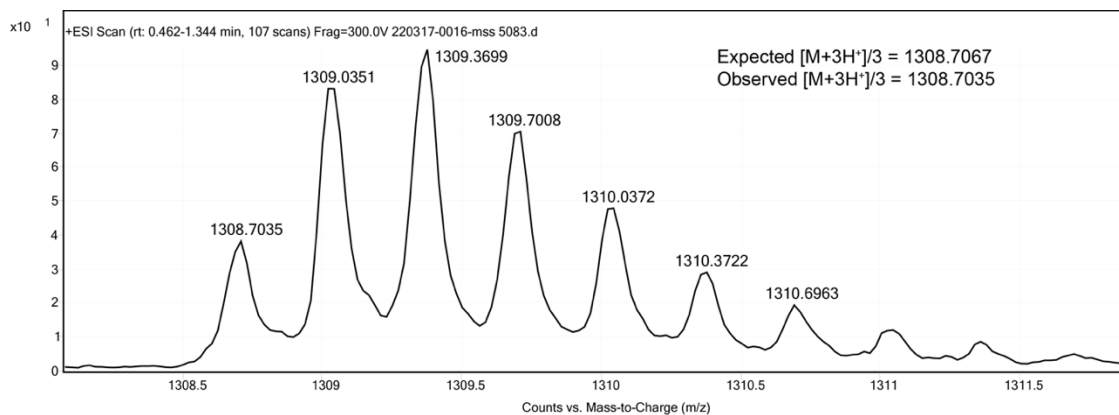


Figure 2-59. ESI TOF spectrum for 18.

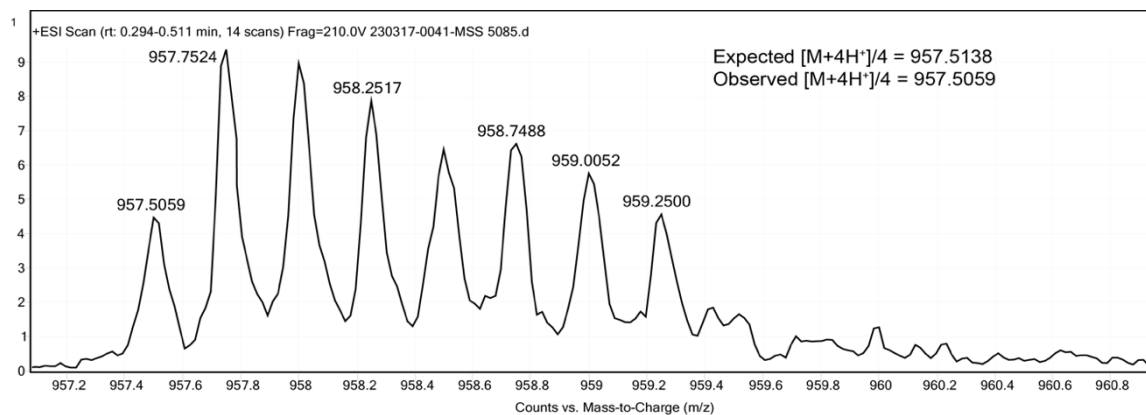


Figure 2-60. ESI TOF spectrum for 19-A(3,4-f₂F)A.

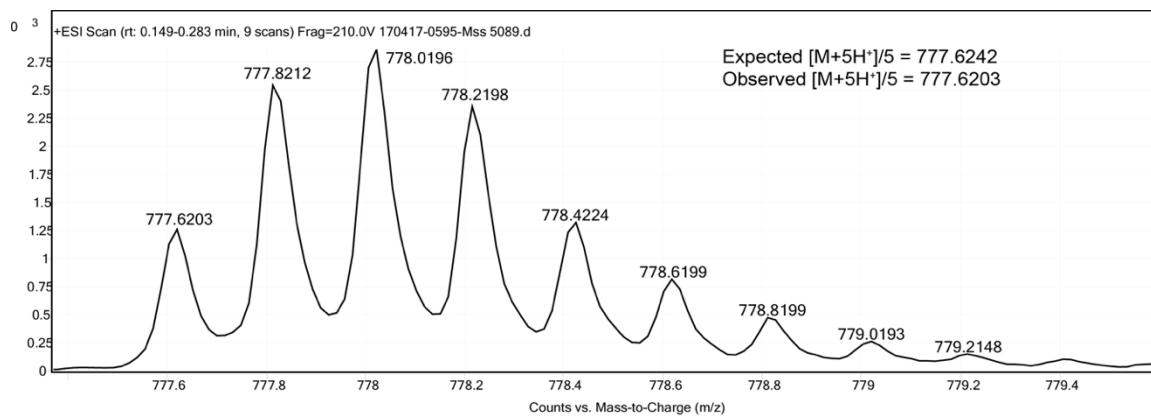


Figure 2-61. ESI TOF spectrum for 19-A(3,4-f₂F)K.

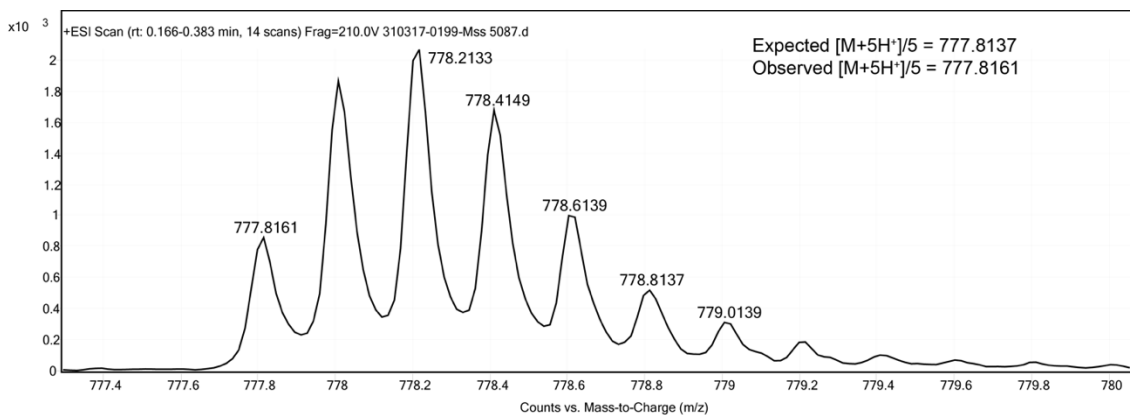


Figure 2-62. ESI TOF spectrum for 19-E(3,4-f₂F)A.

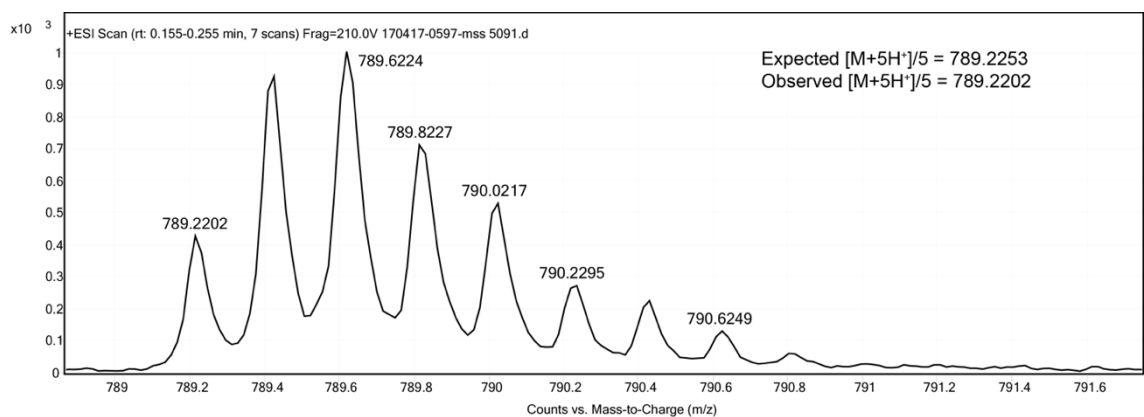


Figure 2-63. ESI TOF spectrum for 19.

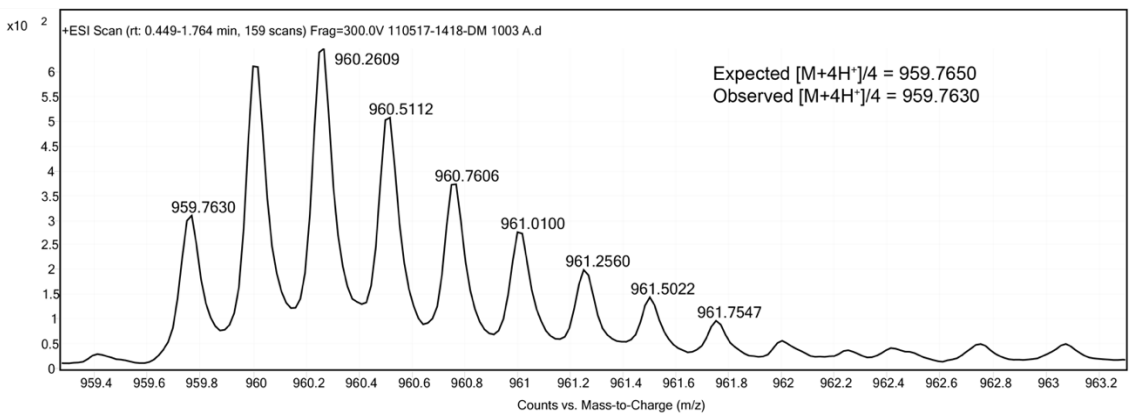


Figure 2-64. ESI TOF spectrum for 20-A(p-NO₂F)A.

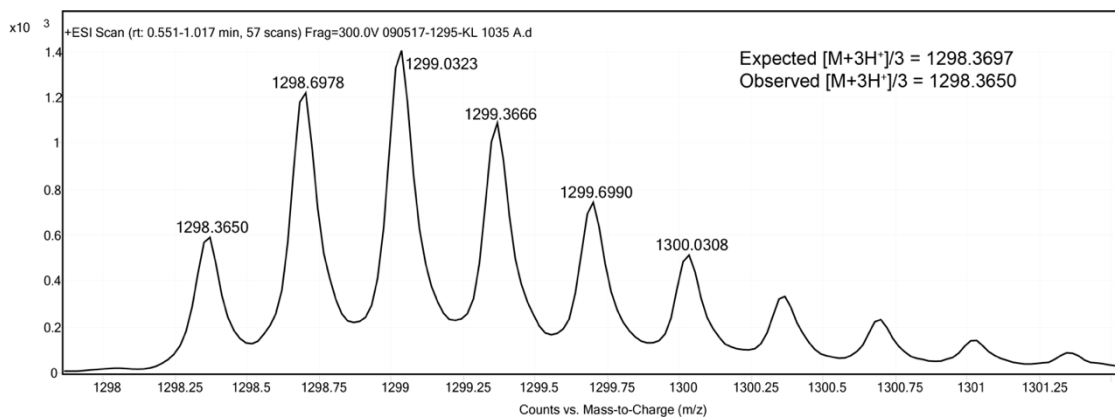


Figure 2-65. ESI TOF spectrum for 20-A(p-NO₂F)K.

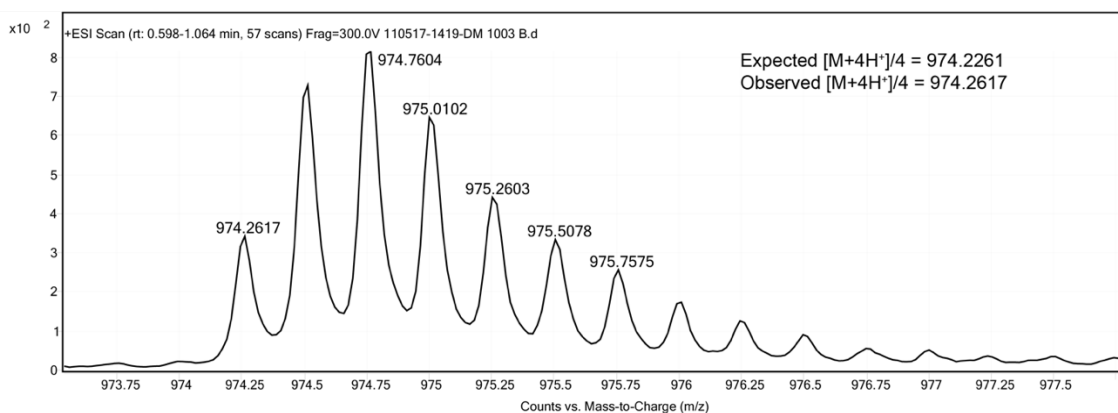


Figure 2-66. ESI TOF spectrum for 20-E(p-NO₂F)A.

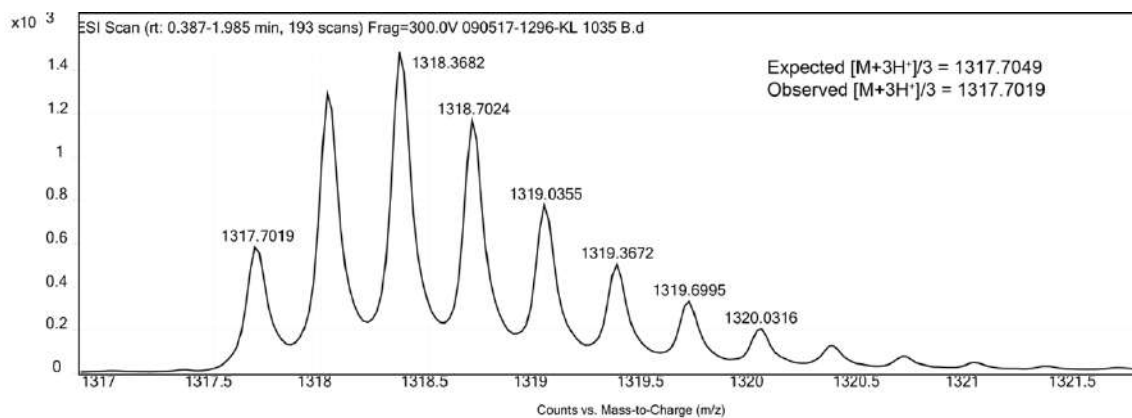


Figure 2-67. ESI TOF spectrum for 20

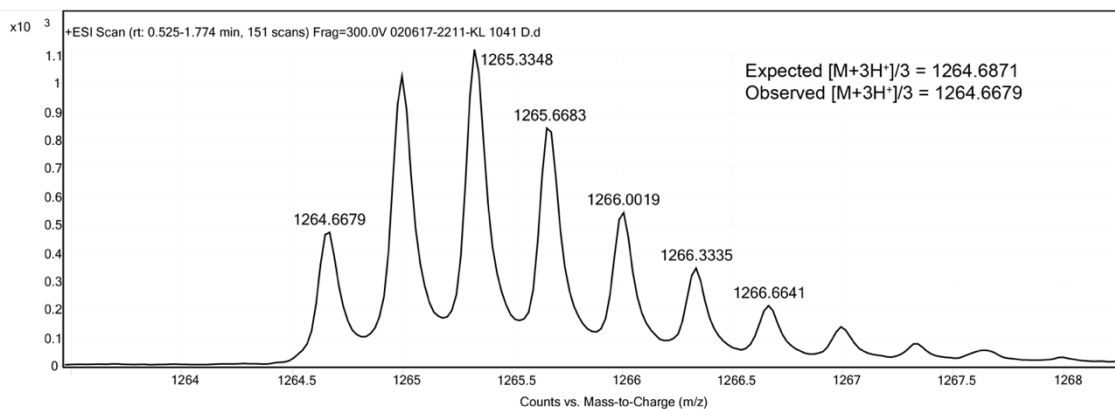


Figure 2-68. ESI TOF spectrum for 21-A(PyrA)A.

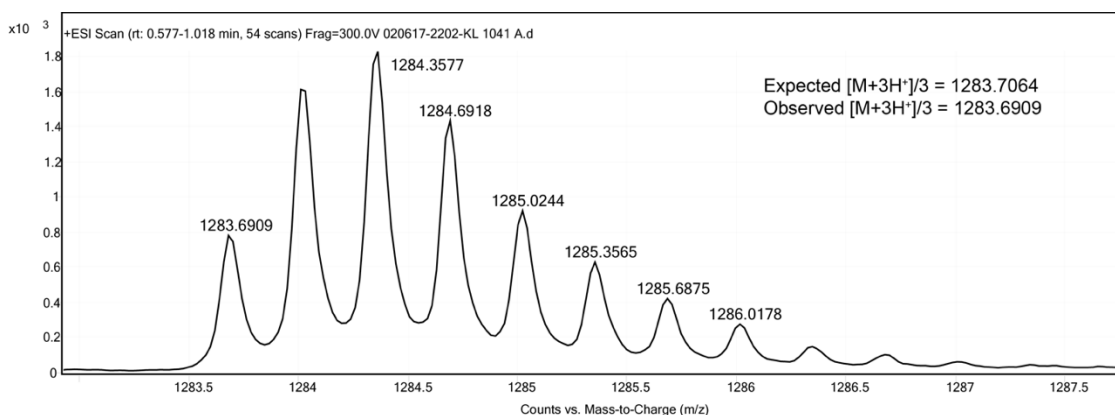


Figure 2-69. ESI TOF spectrum for 21-A(PyrA)K.

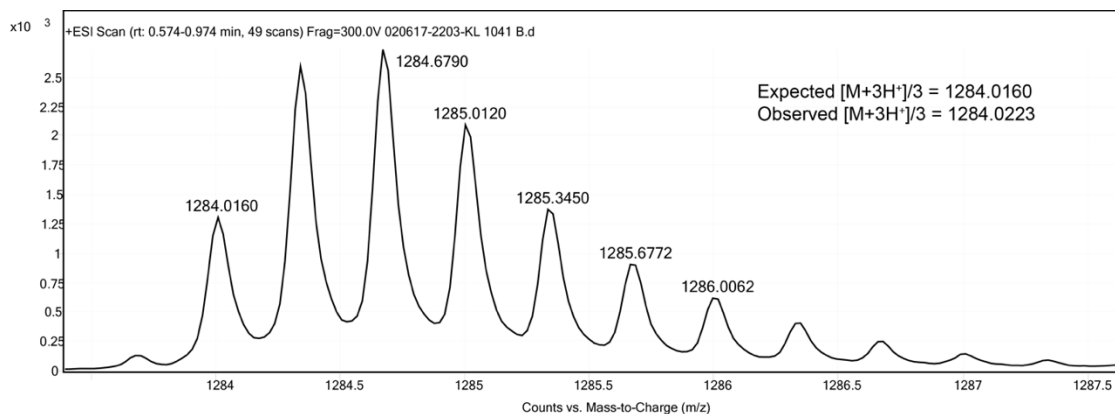


Figure 2-70. ESI TOF spectrum for 21-E(PyrA)A.

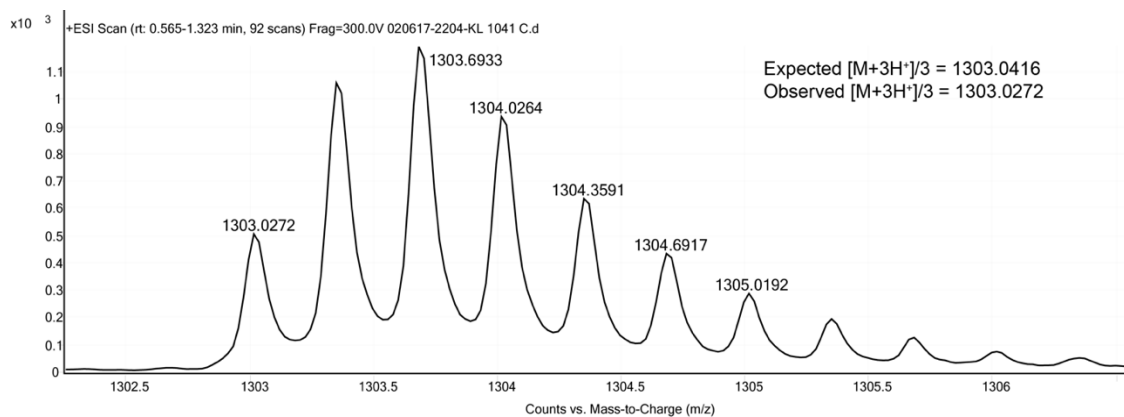


Figure 2-71. ESI TOF spectrum for **21**.

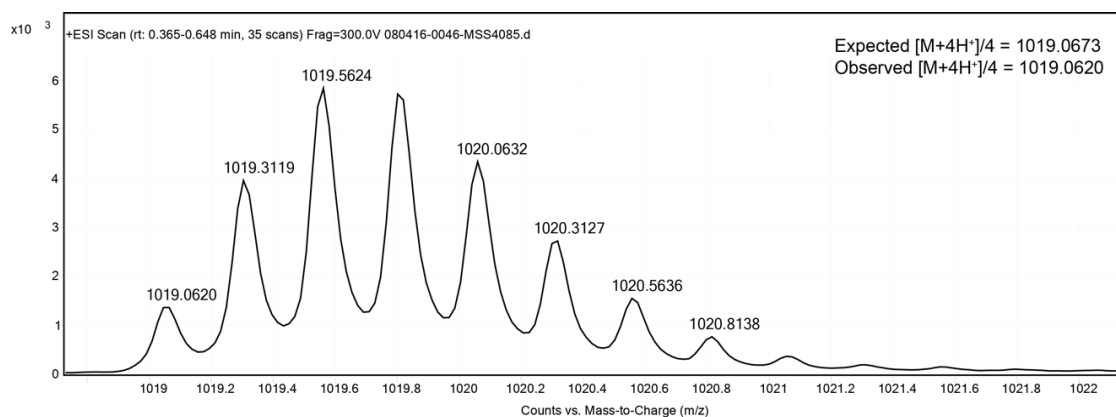


Figure 2-72. ESI TOF spectrum for peptide **22**.

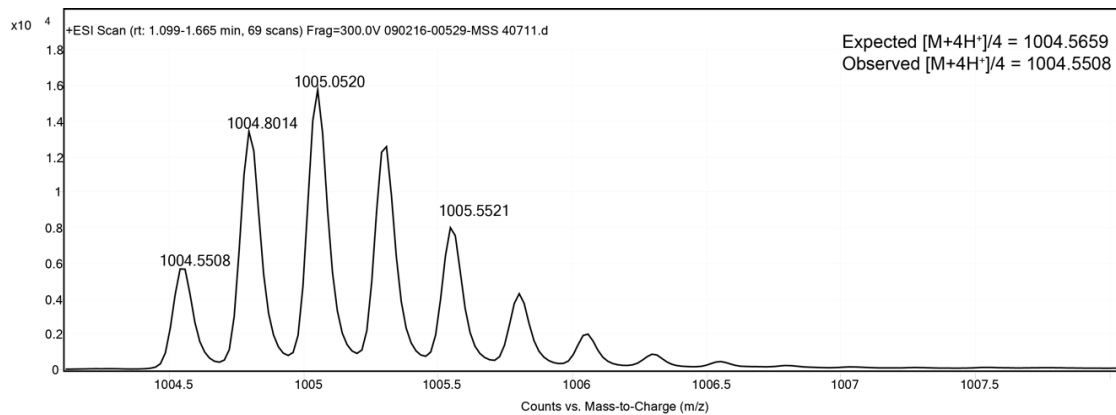


Figure 2-73. ESI TOF spectrum for peptide **22-AFK**.

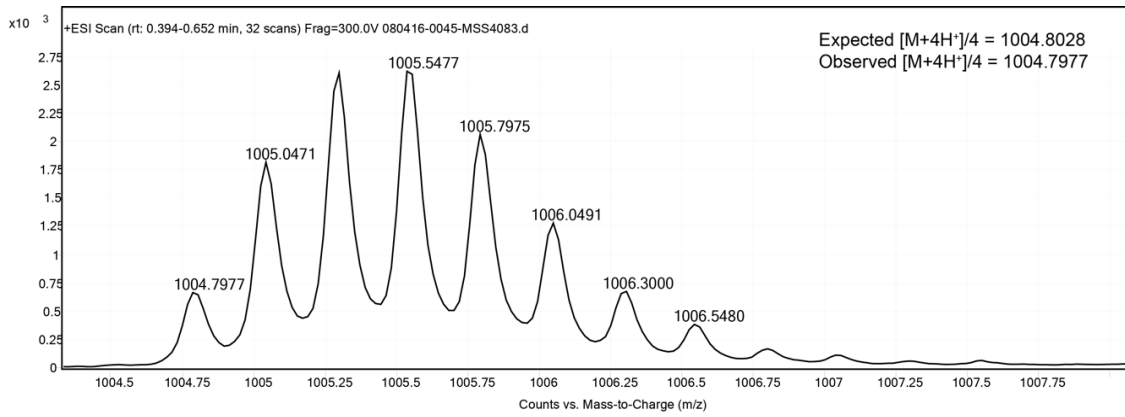


Figure 2-74. ESI TOF spectrum for peptide **22-EFA**.

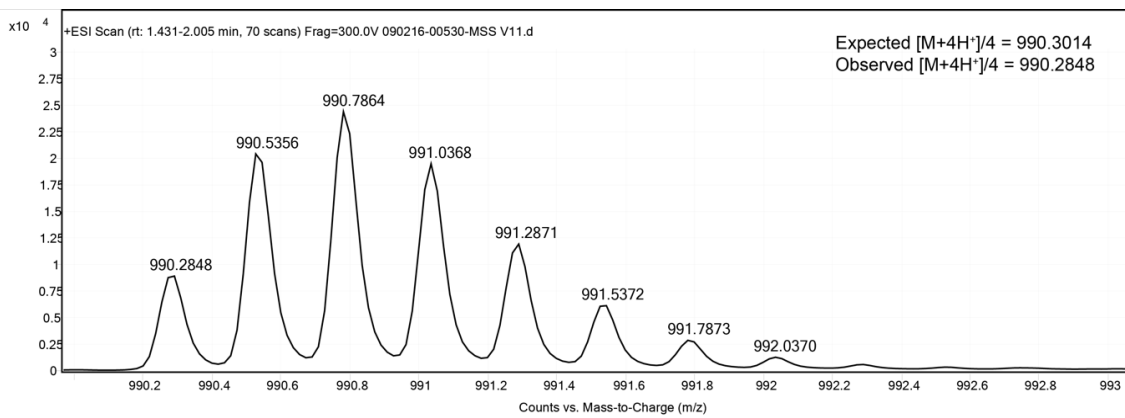


Figure 2-75. ESI TOF spectrum for peptide **22-AFA**.

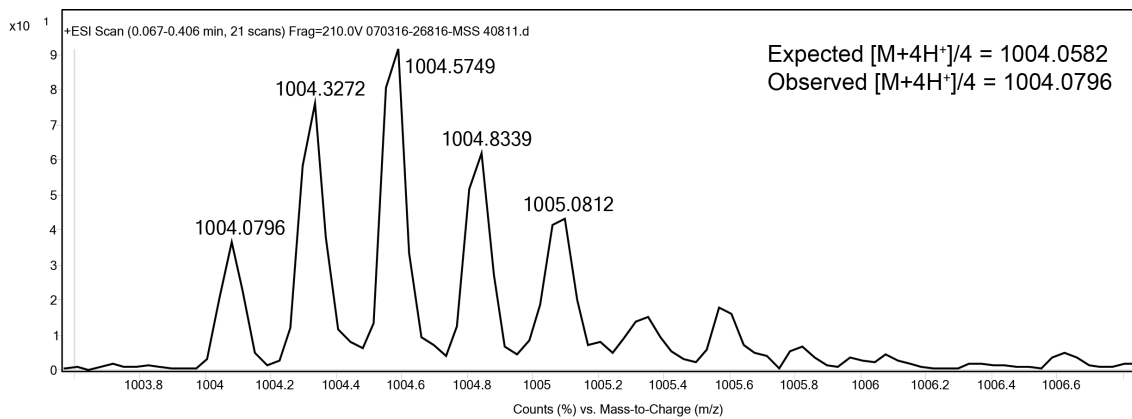


Figure 2-76. ESI TOF spectrum for peptide **22-ESK**.

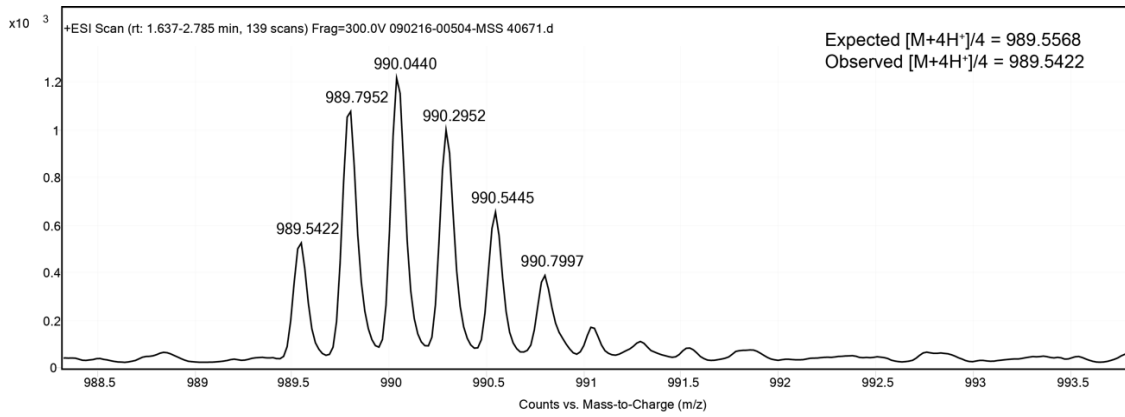


Figure 2-77. ESI TOF spectrum for peptide 22-ASK.

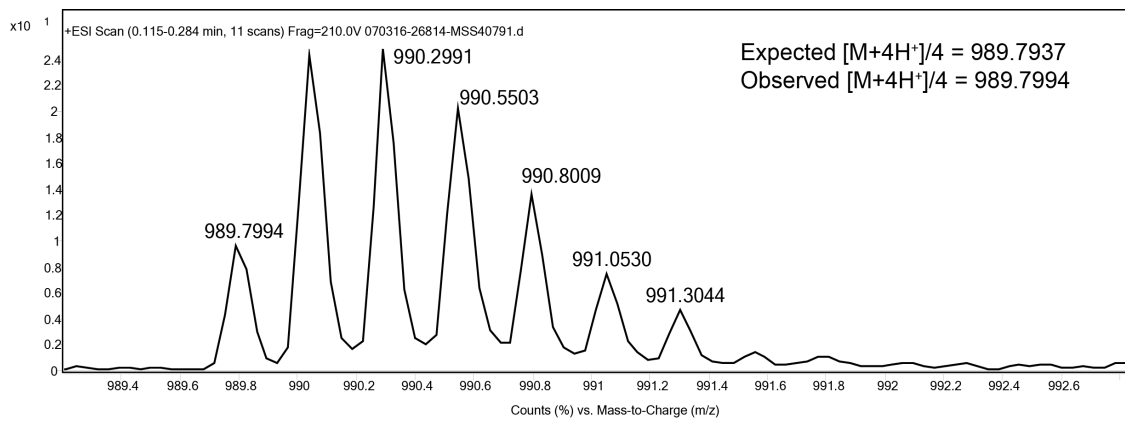


Figure 2-78. ESI TOF spectrum for peptide 22-ESA.

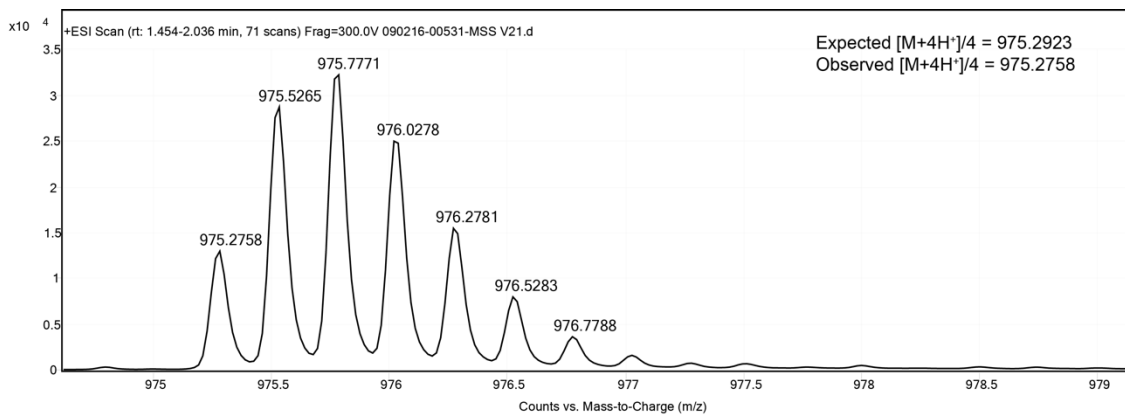


Figure 2-79. ESI TOF spectrum for peptide 22-ASA.

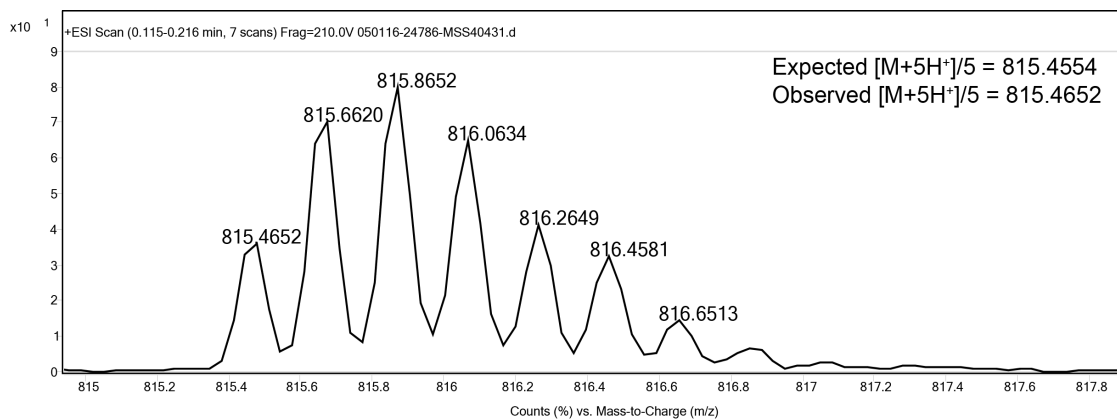


Figure 2-80. ESI TOF spectrum for peptide **23**

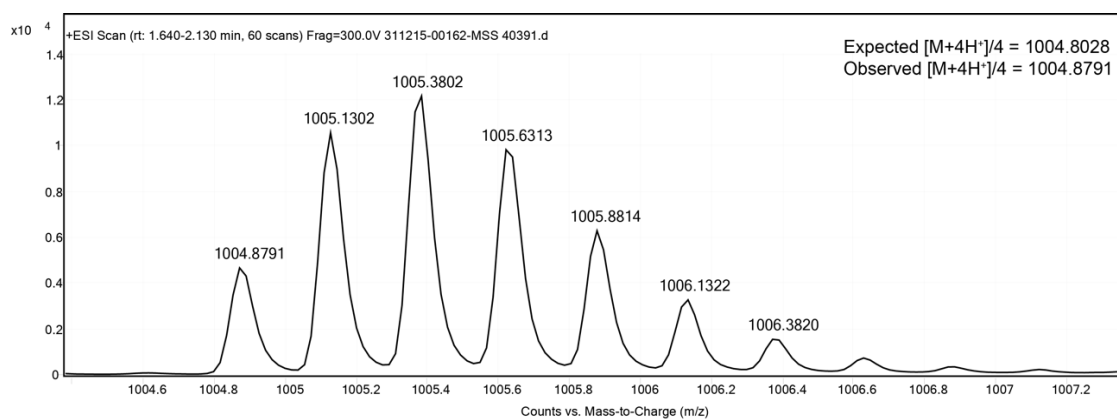


Figure 2-81. ESI TOF spectrum for peptide **23-EFA**.

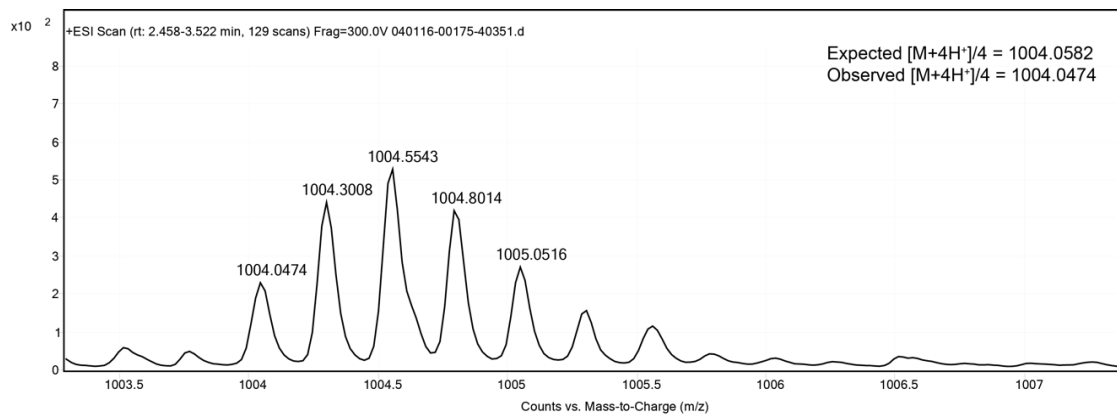


Figure 2-82. ESI TOF spectrum for peptide **23-ESK**.

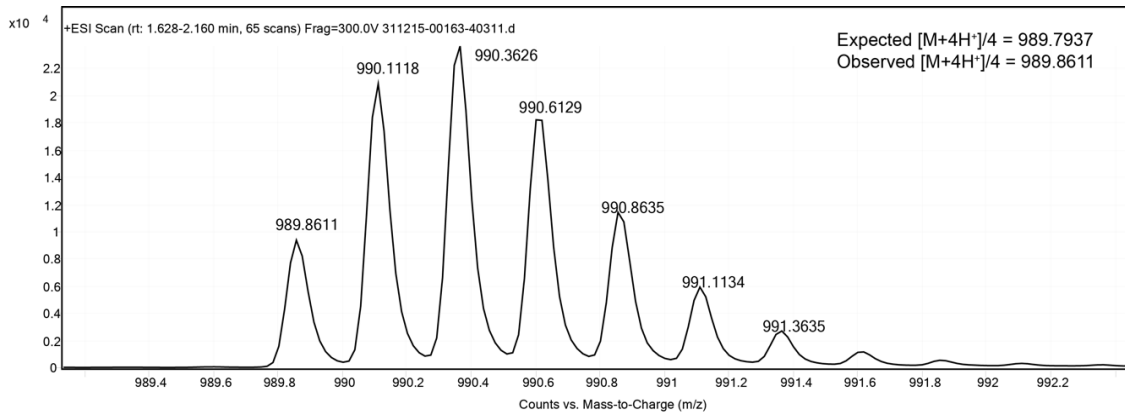


Figure 2-83. ESI TOF spectrum for peptide 23-ESA.

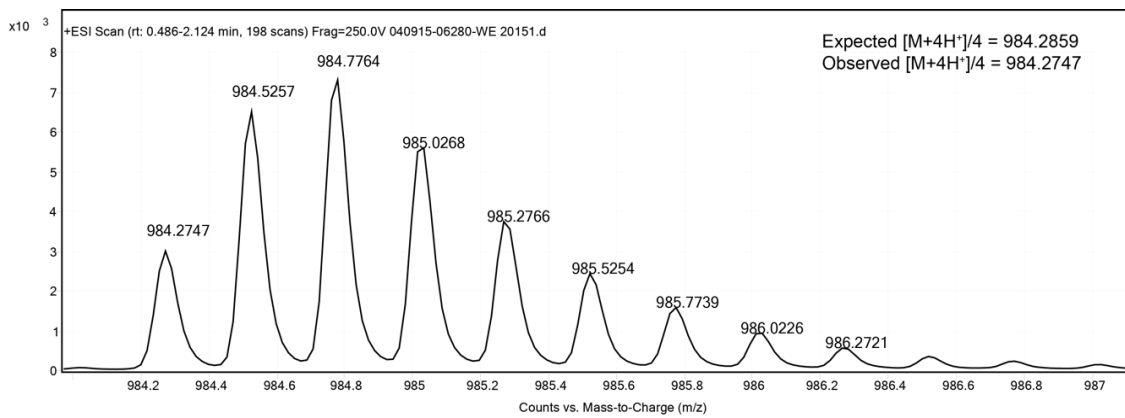


Figure 2-84. ESI TOF spectrum for peptide 24.

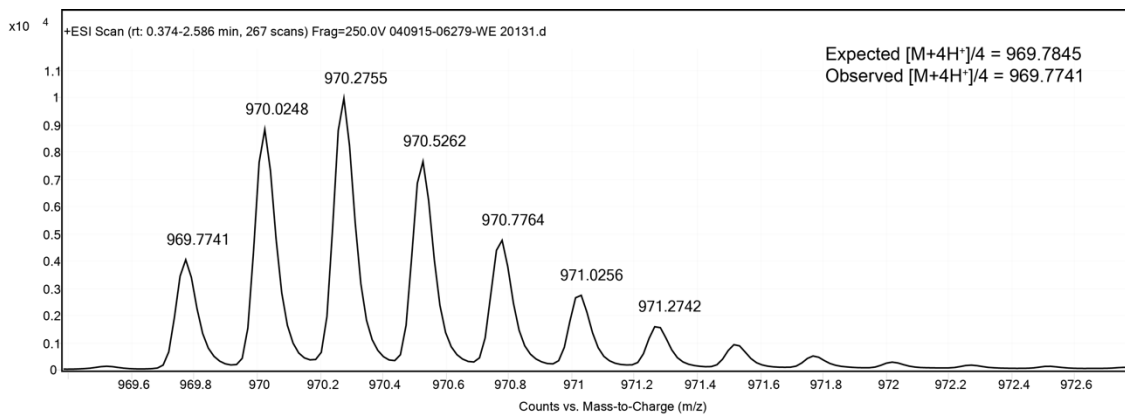


Figure 2-85. ESI TOF spectrum for peptide 24-AFR.

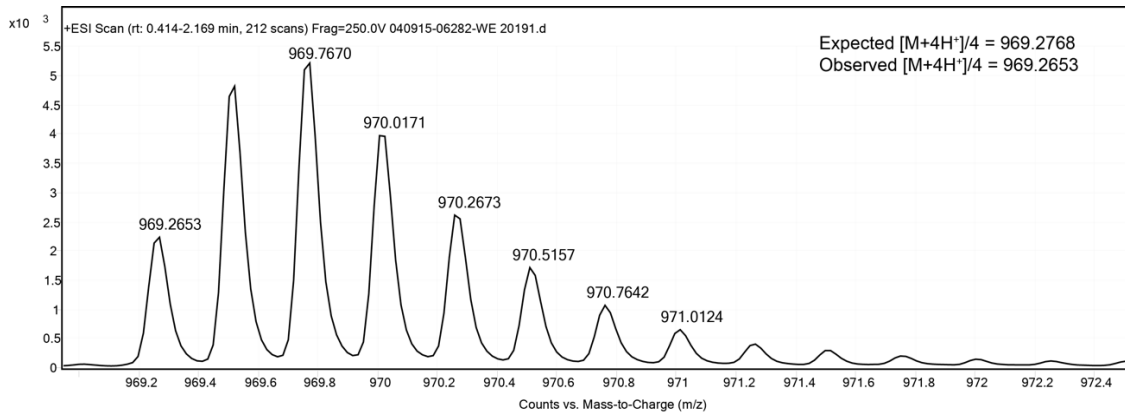


Figure 2-86. ESI TOF spectrum for peptide 24-ESR.

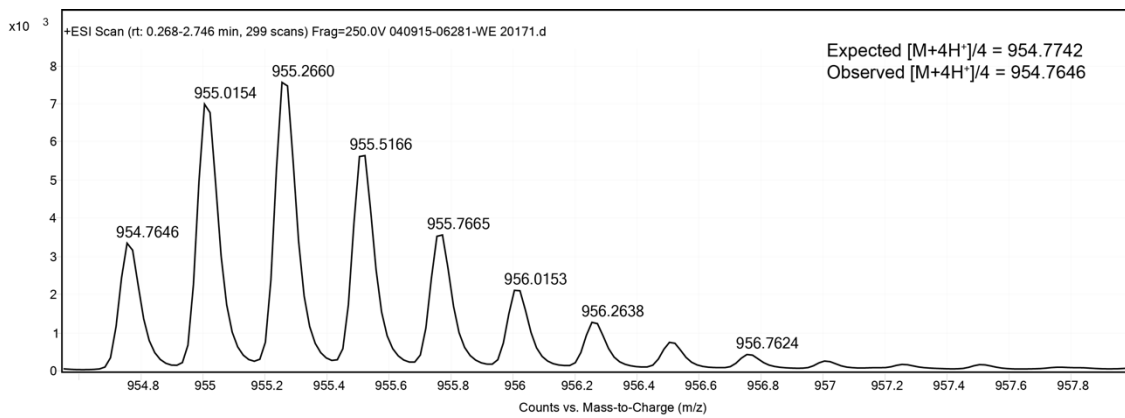


Figure 2-87. ESI TOF spectrum for peptide 24-ASR.

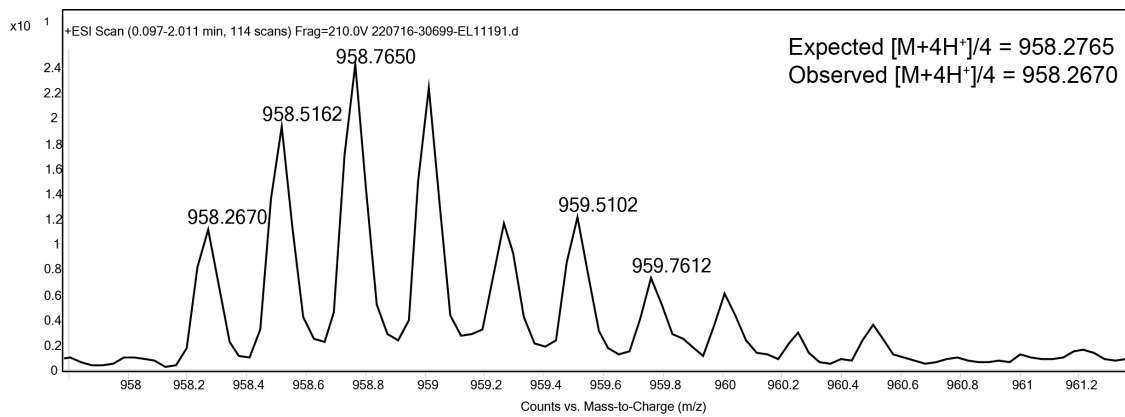


Figure 2-88. ESI TOF spectrum for 25.

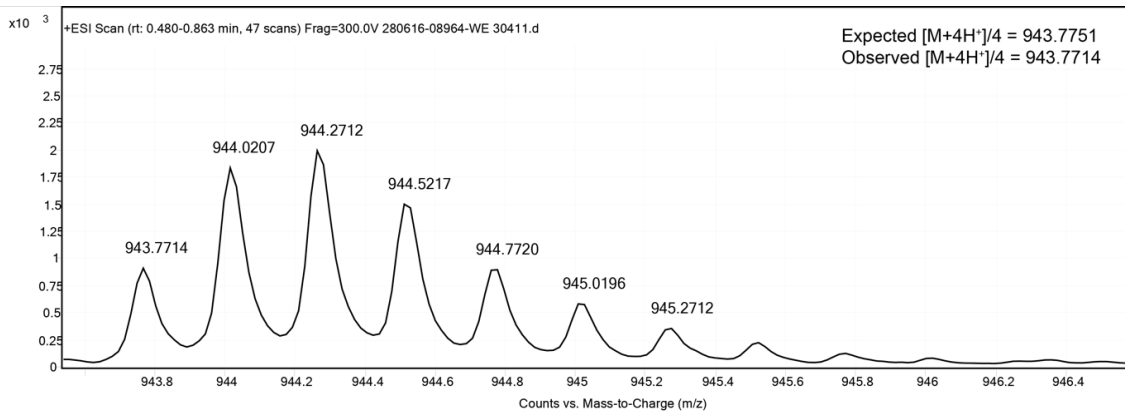


Figure 2-89. ESI TOF spectrum for peptide **25-AAK**.

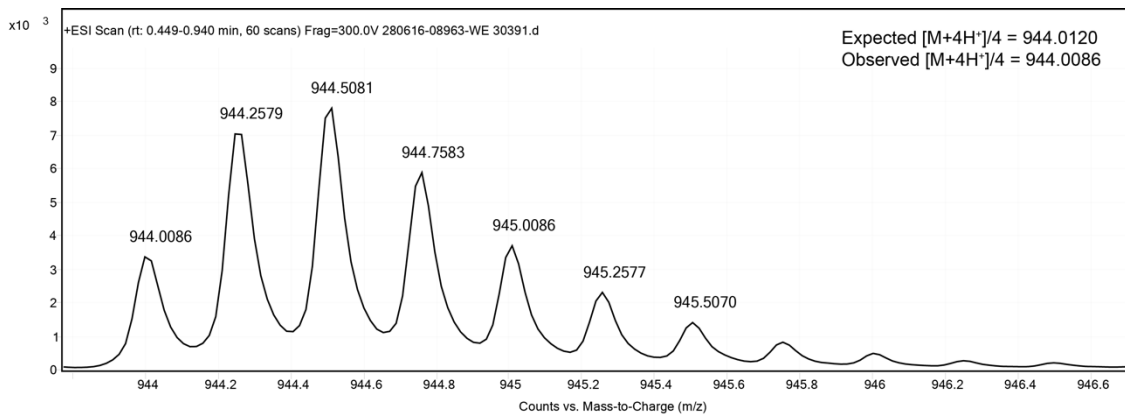


Figure 2-90. ESI TOF spectrum for **25-EAA**.

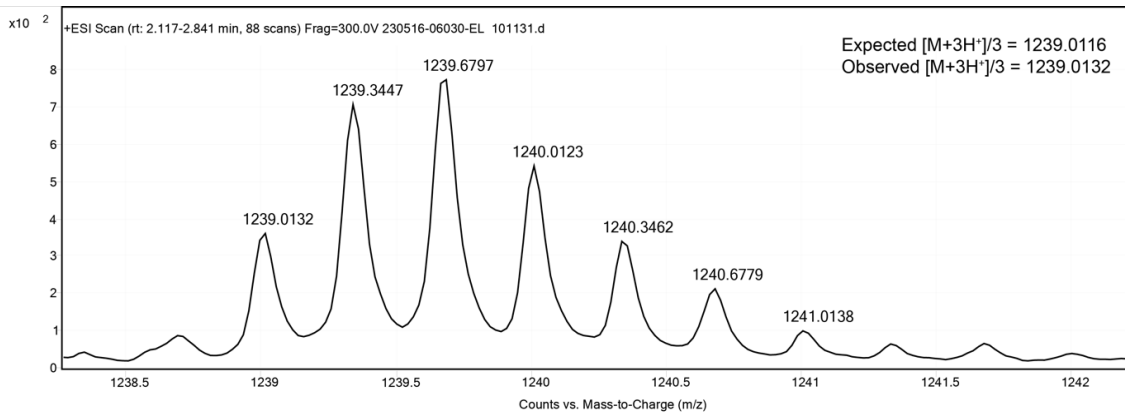


Figure 2-91. ESI TOF spectrum for **25-AAA**.

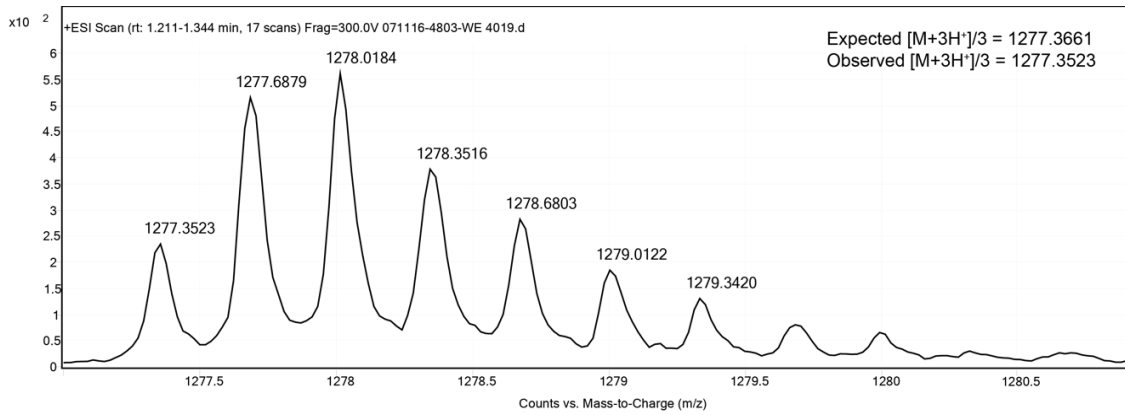


Figure 2-92. ESI TOF spectrum for peptide 26.

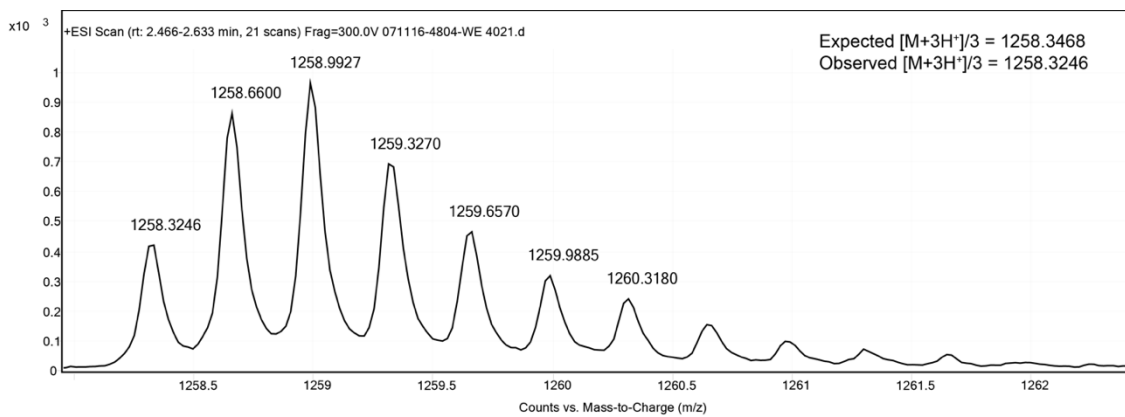


Figure 2-93. ESI TOF spectrum for peptide 27.

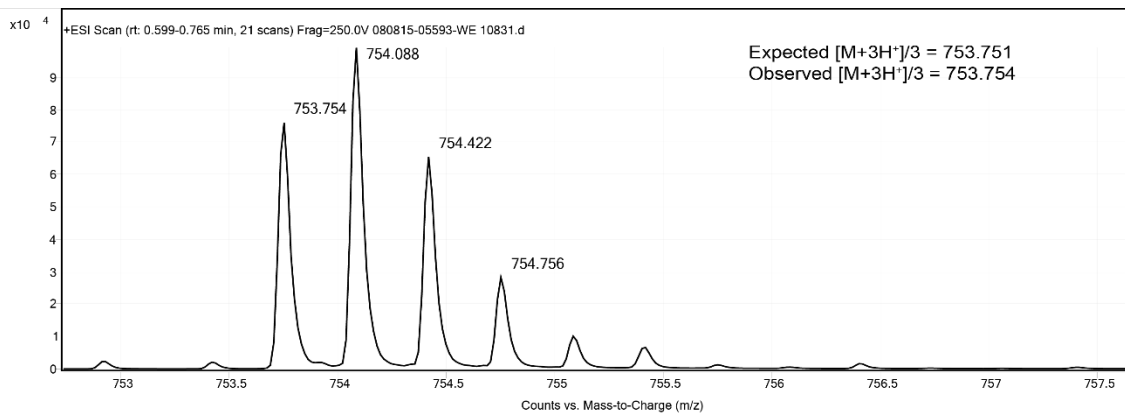


Figure 2-94. ESI TOF spectrum for pA-EWK.

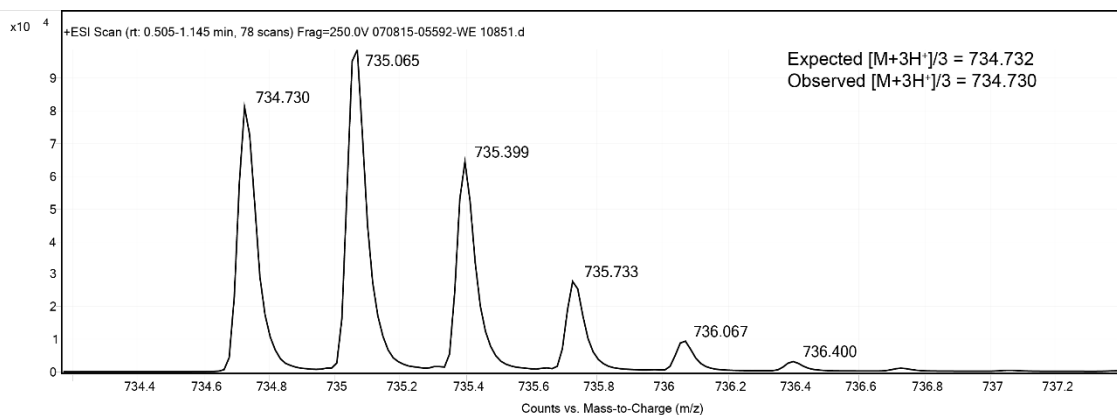


Figure 2-95. ESI TOF spectrum for pA-EWA.

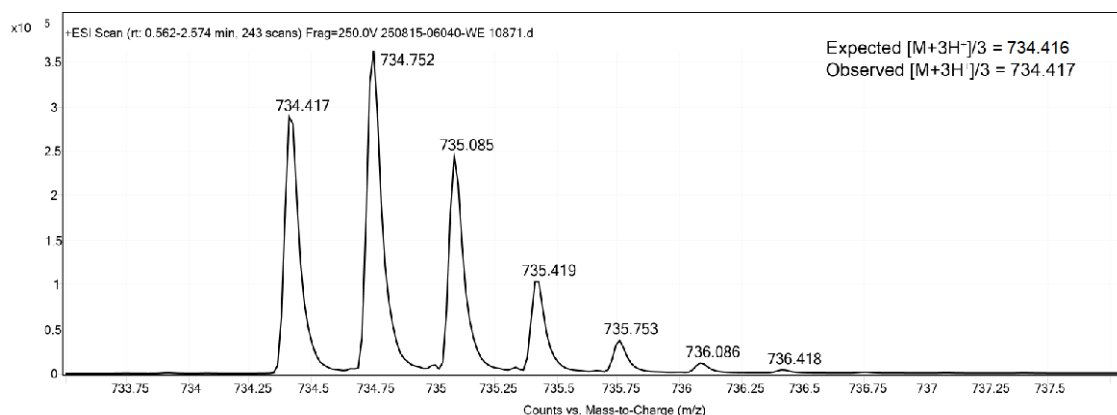


Figure 2-96. ESI TOF spectrum for pA-AWK.

2.4.3 Analytical HPLC data

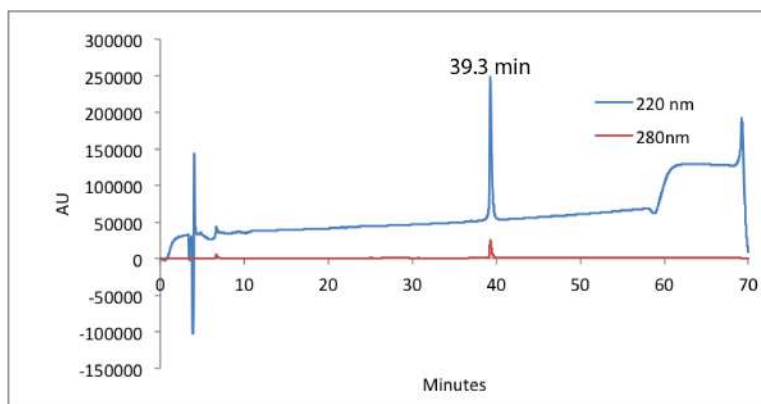


Figure 2-97. Analytical HPLC Data for **1**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

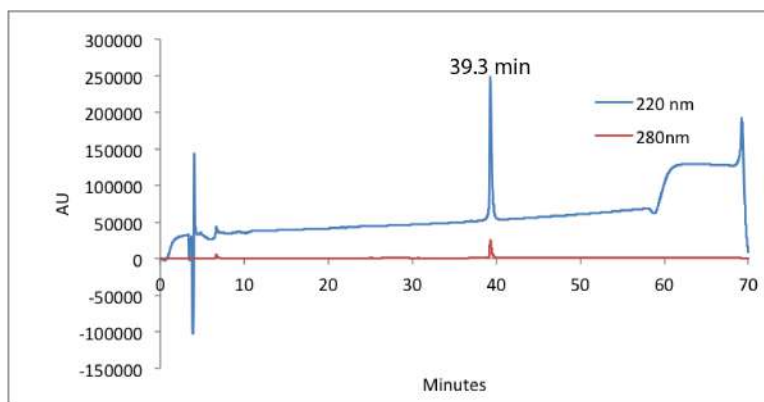


Figure 2-98. Analytical HPLC Data for **2**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

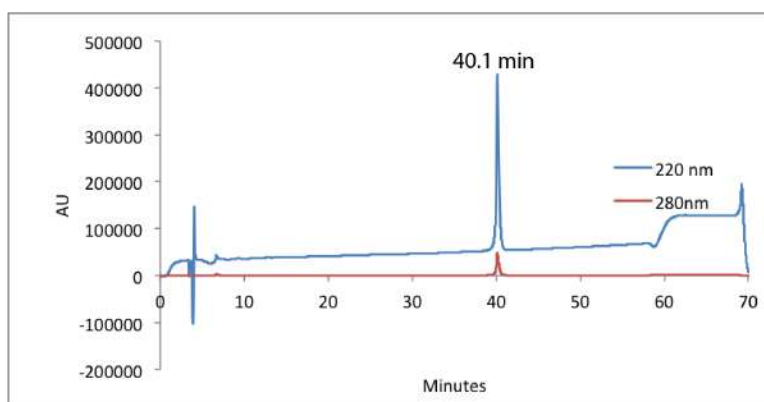


Figure 2-99. Analytical HPLC Data for **3**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

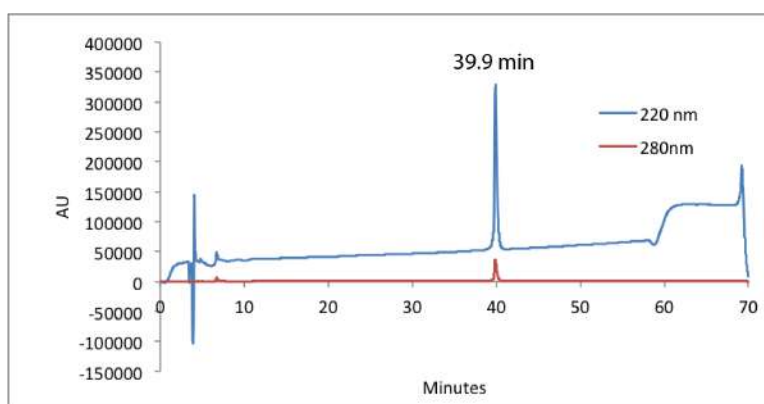


Figure 2-100. Analytical HPLC Data for **4**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

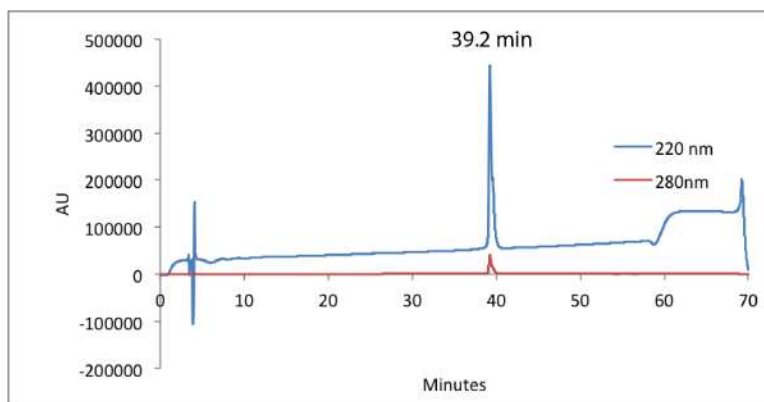


Figure 2-101. Analytical HPLC Data for **5**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

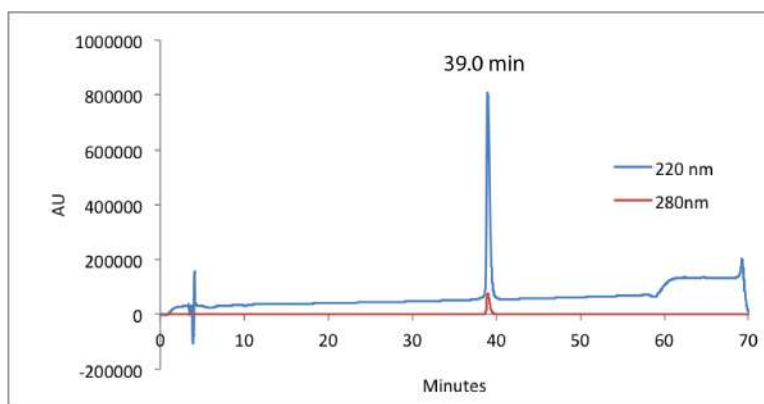


Figure 2-102. Analytical HPLC Data for **6**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

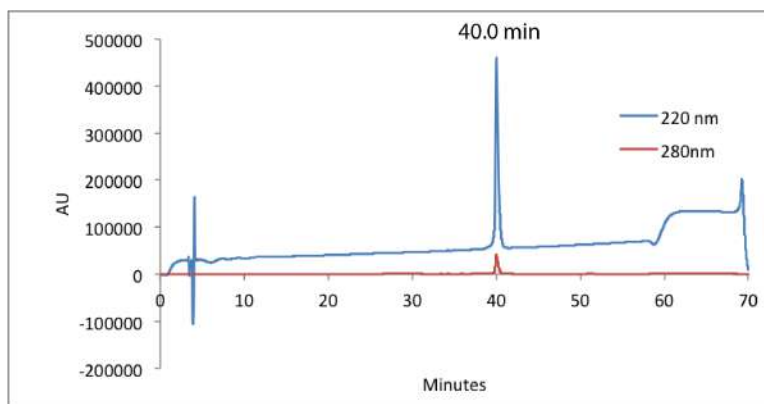


Figure 2-103. Analytical HPLC Data for **7**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50

minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

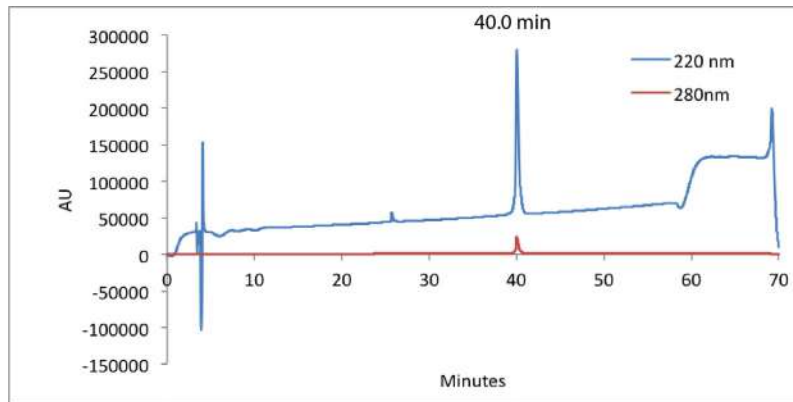


Figure 2-104. Analytical HPLC Data for peptide **8**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

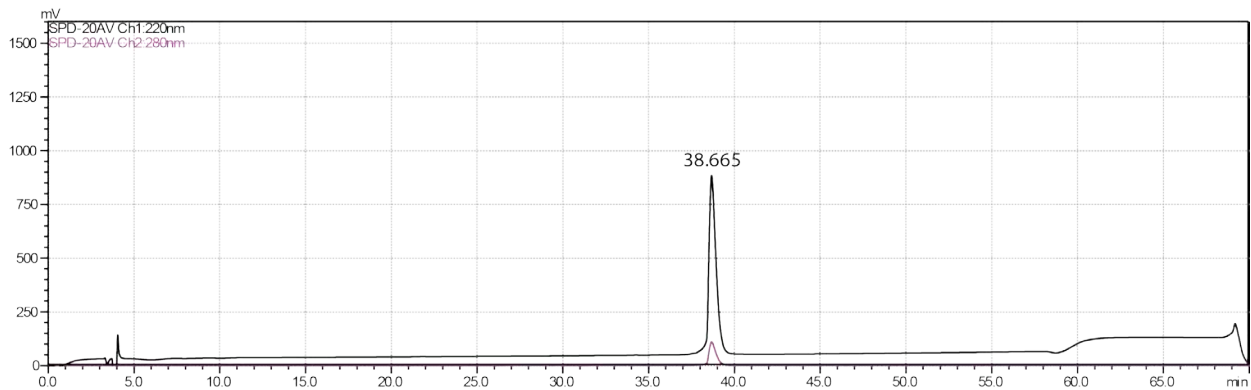


Figure 2-105. Analytical HPLC Data for **9**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

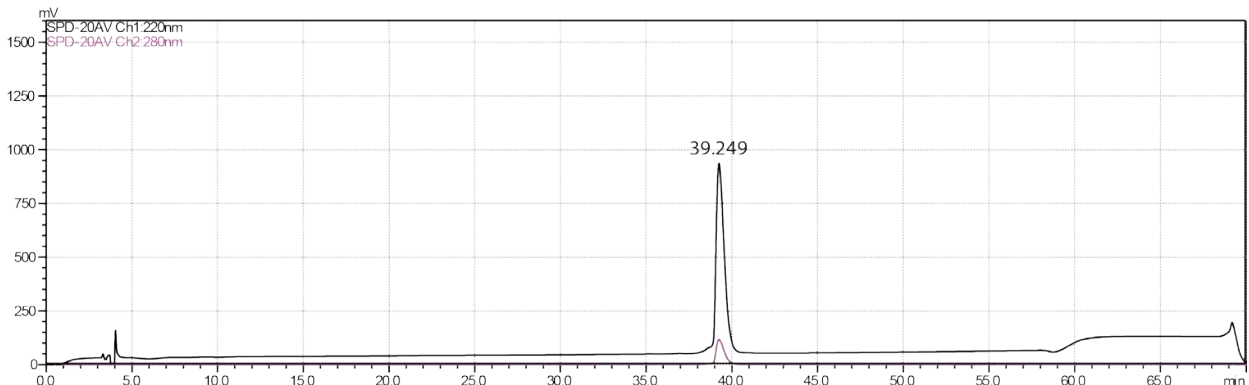


Figure 2-106. Analytical HPLC Data for **9-DWA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over

50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

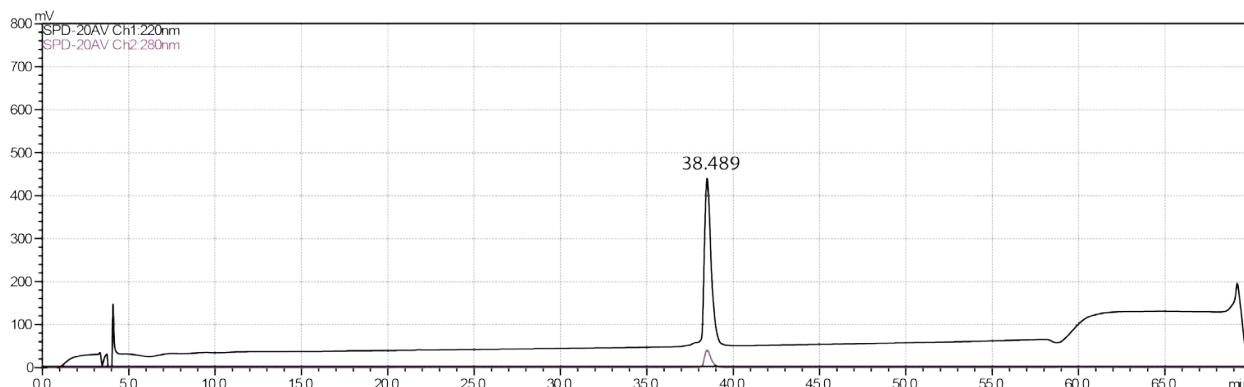


Figure 2-107. Analytical HPLC Data for **9-DSK**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

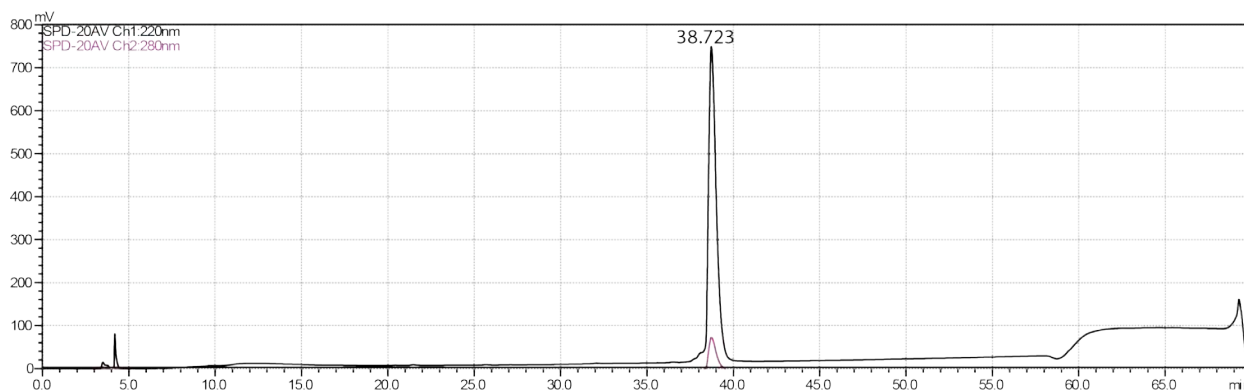


Figure 2-108. Analytical HPLC Data for **9-DSA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

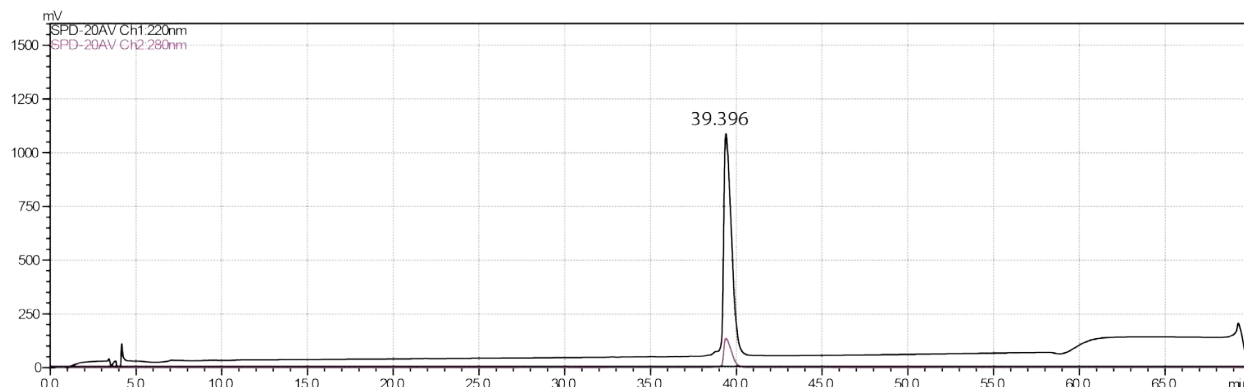


Figure 2-109. Analytical HPLC Data for **10**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50

minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

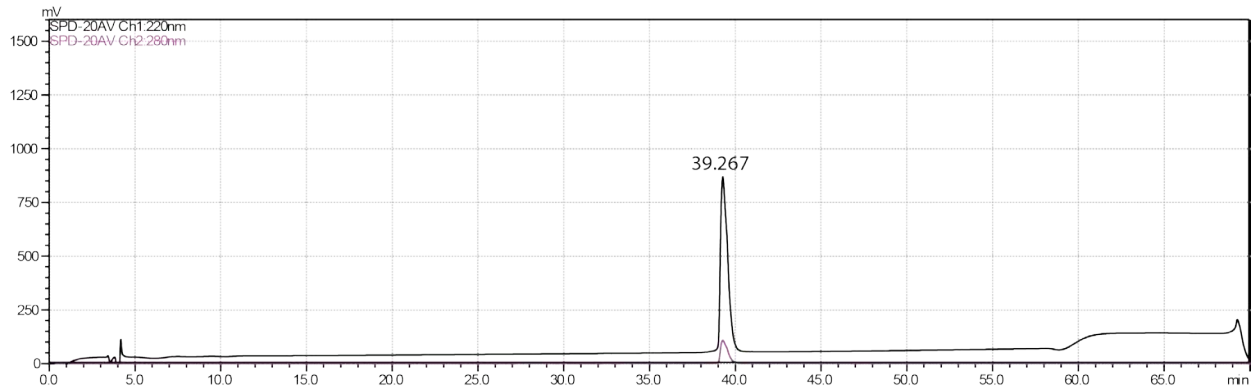


Figure 2-110. Analytical HPLC Data for **10-AWO**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

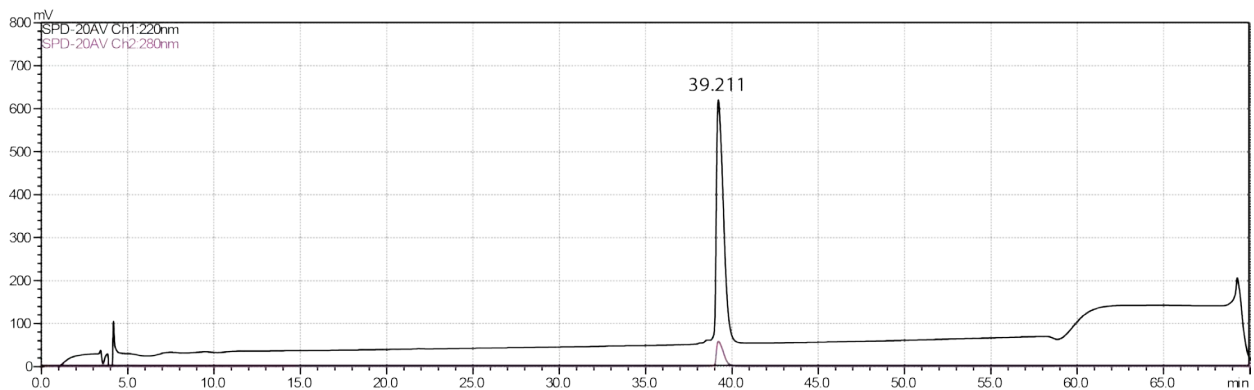


Figure 2-111. Analytical HPLC Data for **10-ESO**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

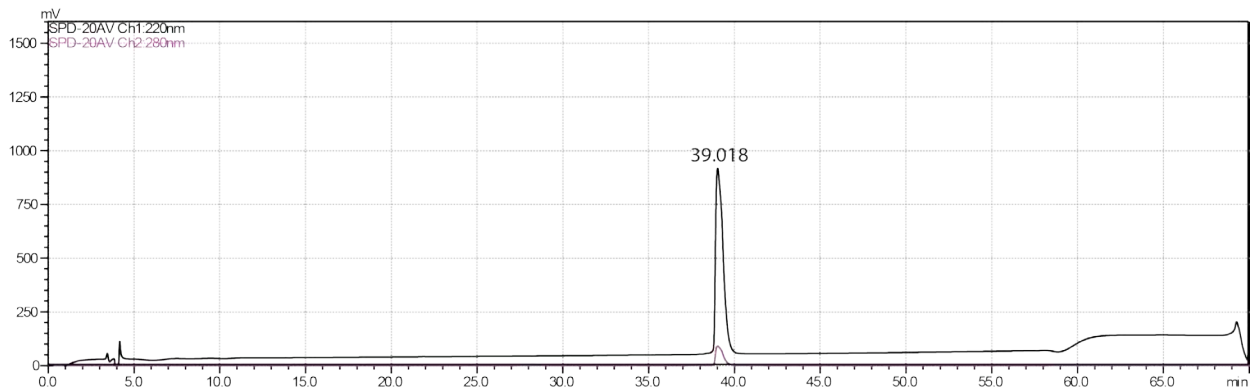


Figure 2-112. Analytical HPLC Data for **10-ASO**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over

50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

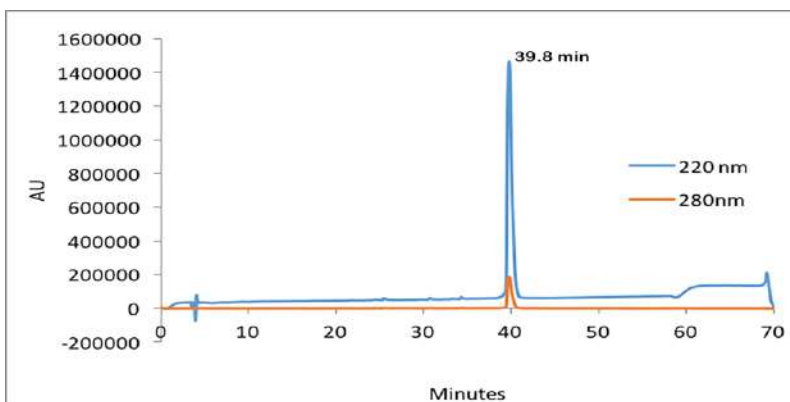


Figure 2-113. Analytical HPLC Data for **11**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

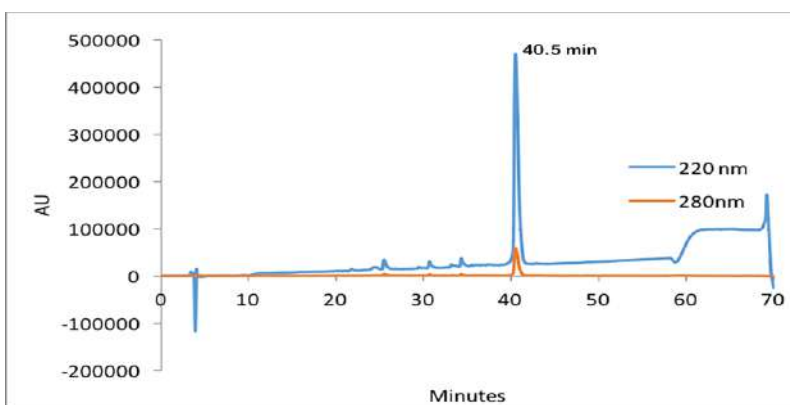


Figure 2-114. Analytical HPLC Data for **11-EWA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

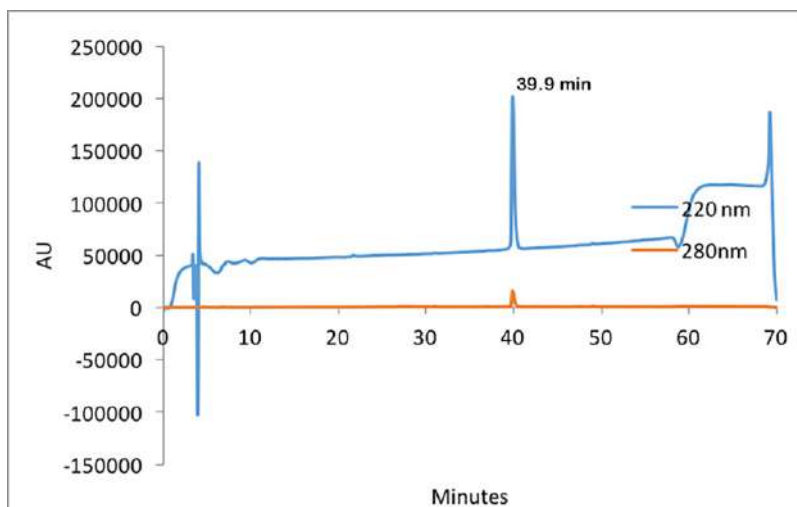


Figure 2-115. Analytical HPLC Data for **11-ESK**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

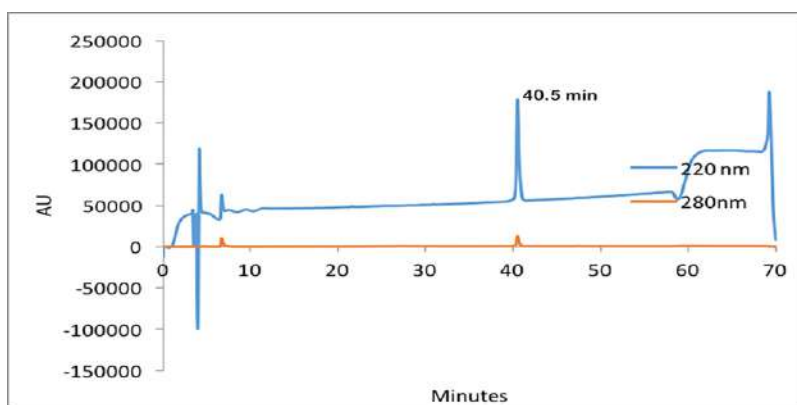


Figure 2-116. Analytical HPLC Data for **11-ESA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

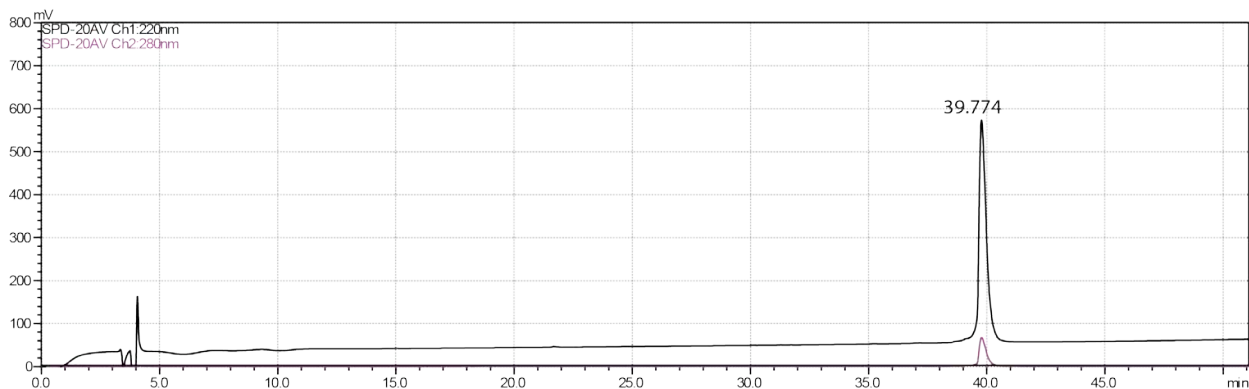


Figure 2-117. Analytical HPLC Data for **12**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min. (this run was ended early)

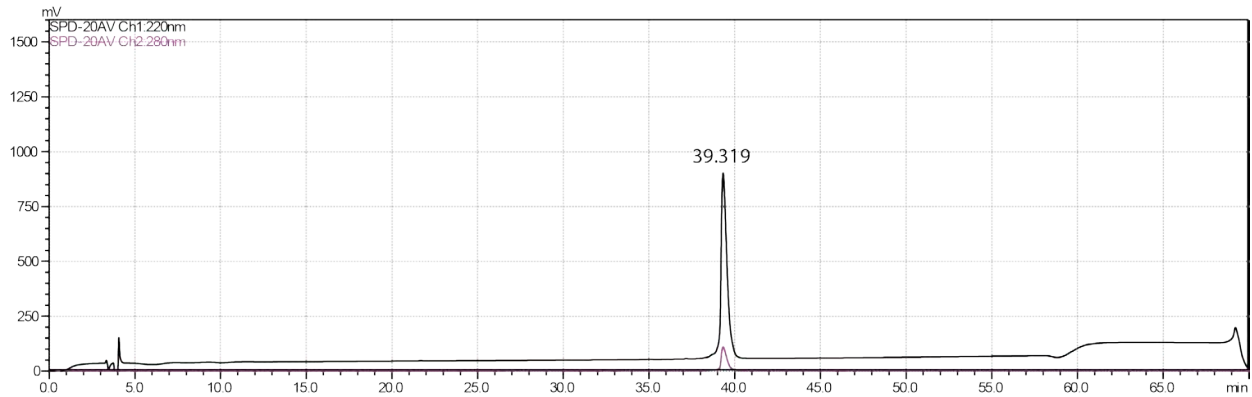


Figure 2-118. Analytical HPLC Data for **12-KWA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

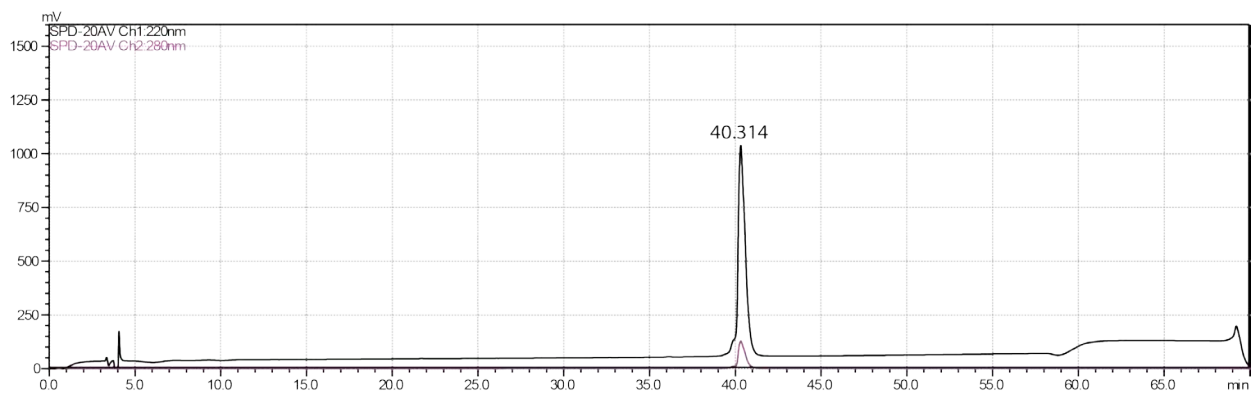


Figure 2-119. Analytical HPLC Data for **12-AWE**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

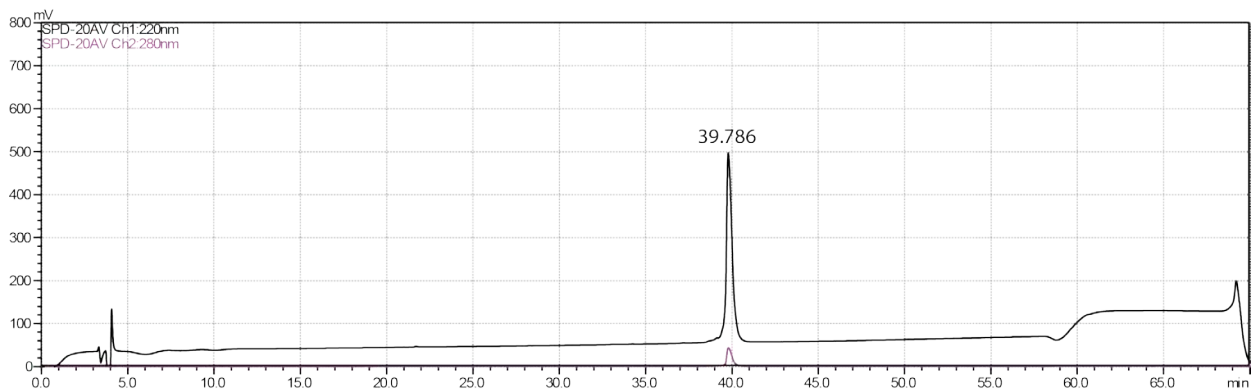


Figure 2-120. Analytical HPLC Data for **12-KSE**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

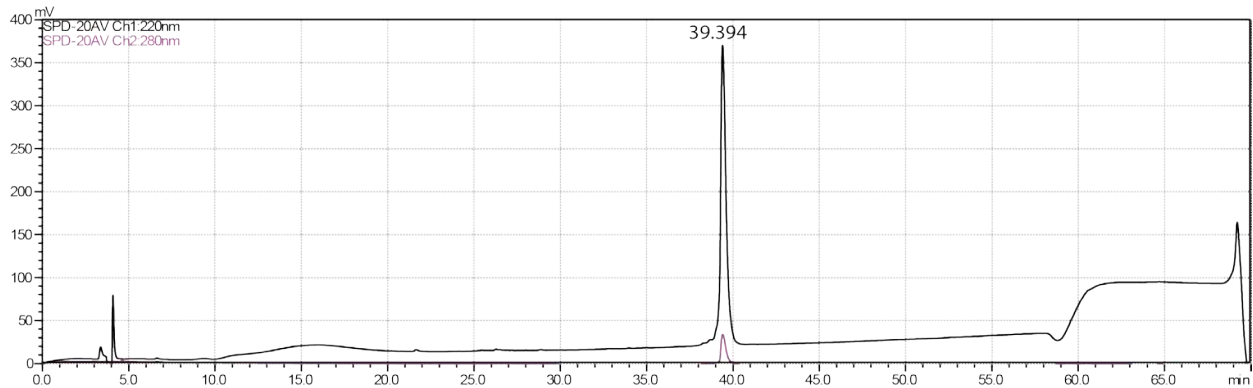


Figure 2-121. Analytical HPLC Data for **12-KSA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

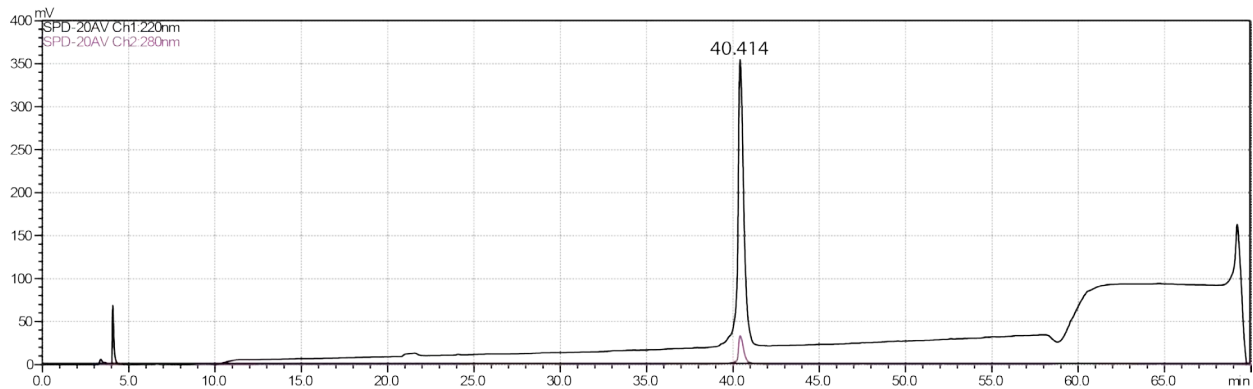


Figure 2-122. Analytical HPLC Data for **12-ASE**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

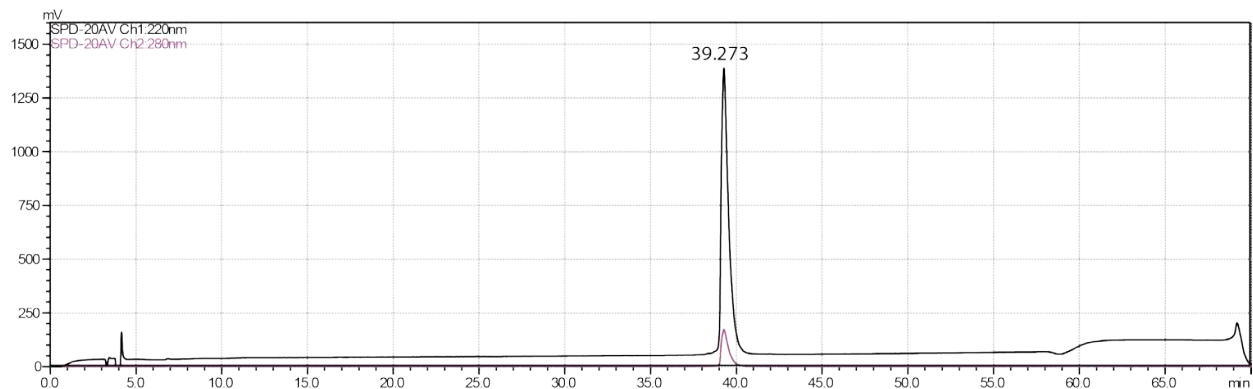


Figure 2-123. Analytical HPLC Data for **13**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

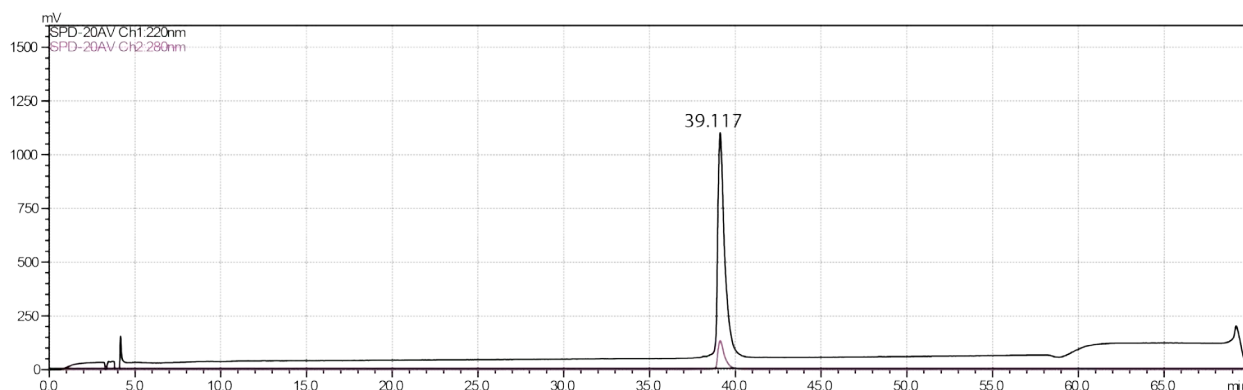


Figure 2-124. Analytical HPLC Data for **13-AW_{me}K**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

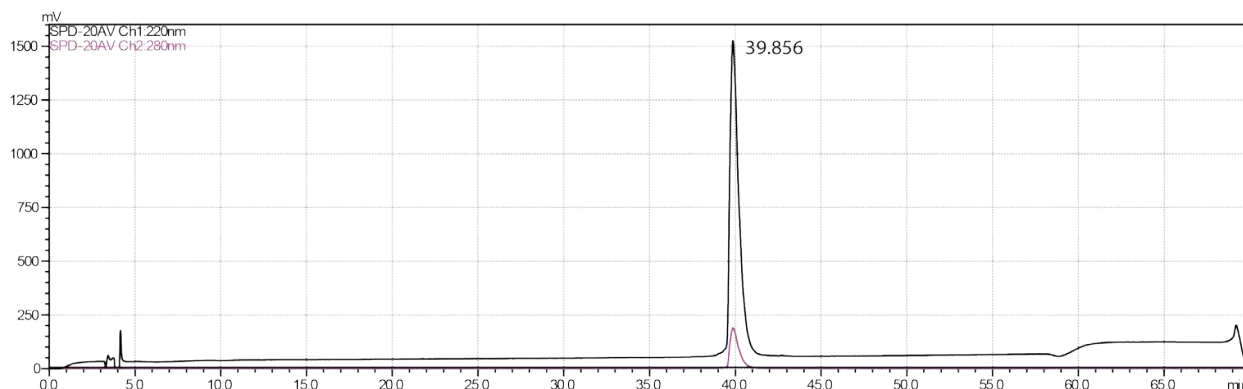


Figure 2-125. Analytical HPLC Data for **13-EW_{me}A**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

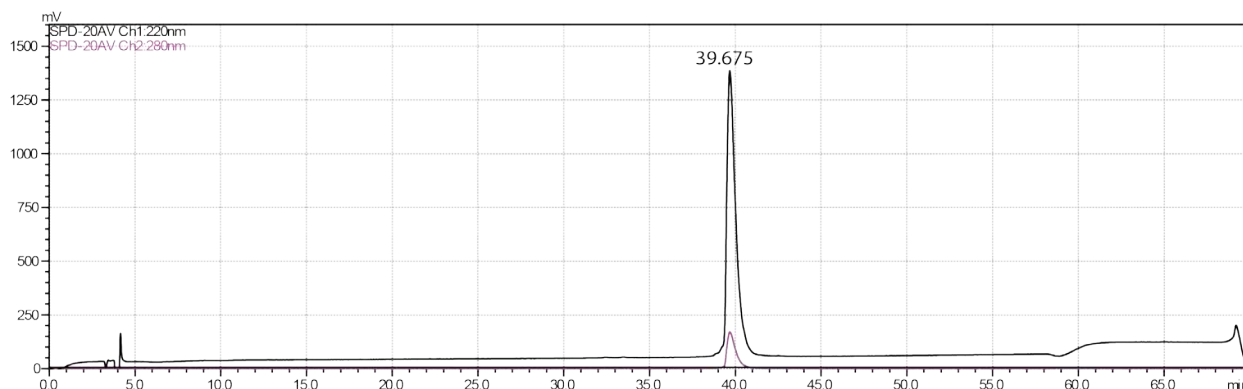


Figure 2-126. Analytical HPLC Data for **13-AW_{me}A**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

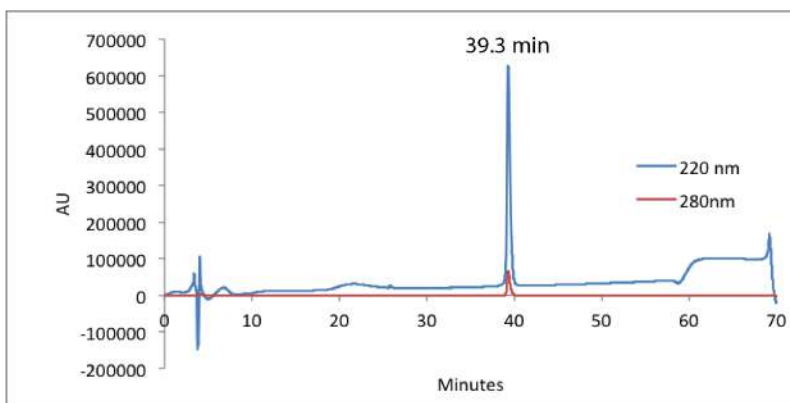


Figure 2-127. Analytical HPLC Data for **14**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

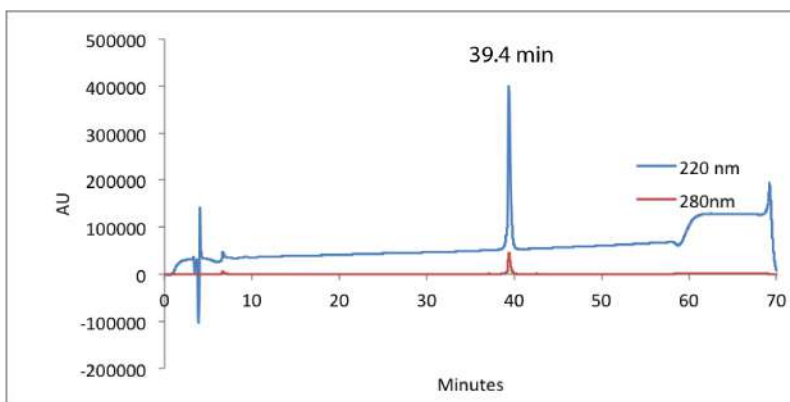


Figure 2-128. Analytical HPLC Data for **14-AYK**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

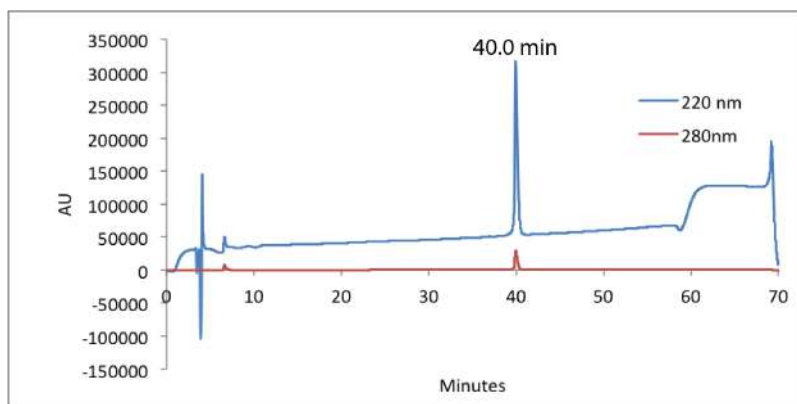


Figure 2-129. Analytical HPLC Data for **14-EYA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

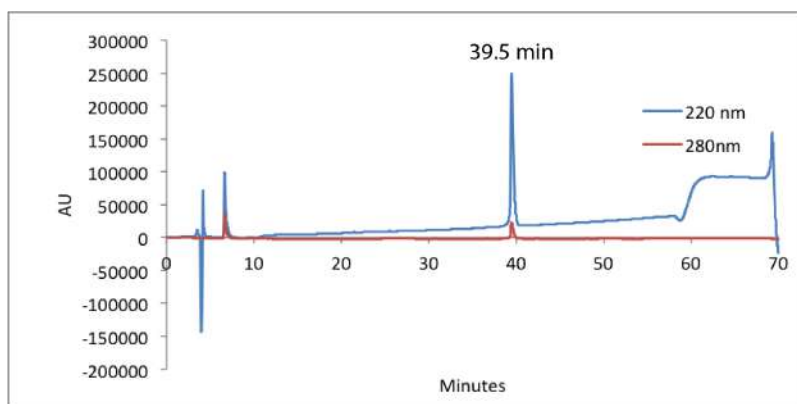


Figure 2-130. Analytical HPLC Data for **14-AYA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

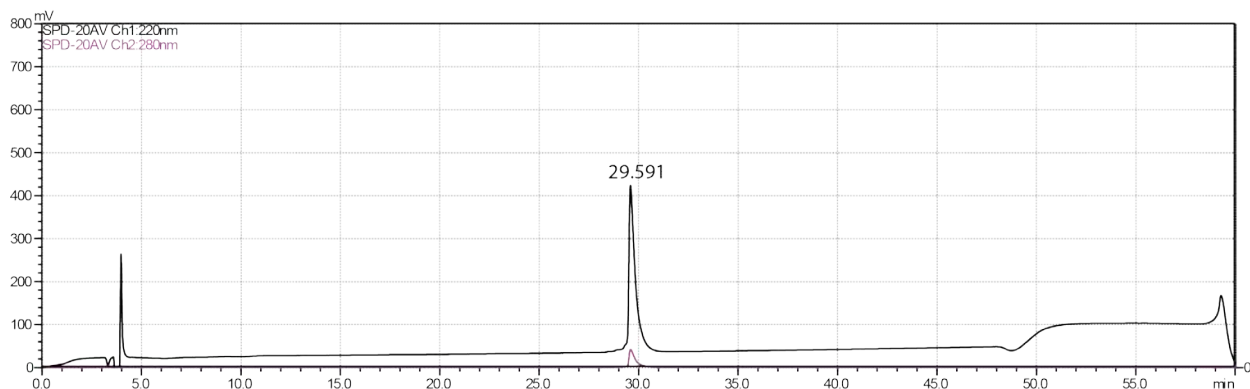


Figure 2-131. Analytical HPLC Data for **15**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 20-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 40 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

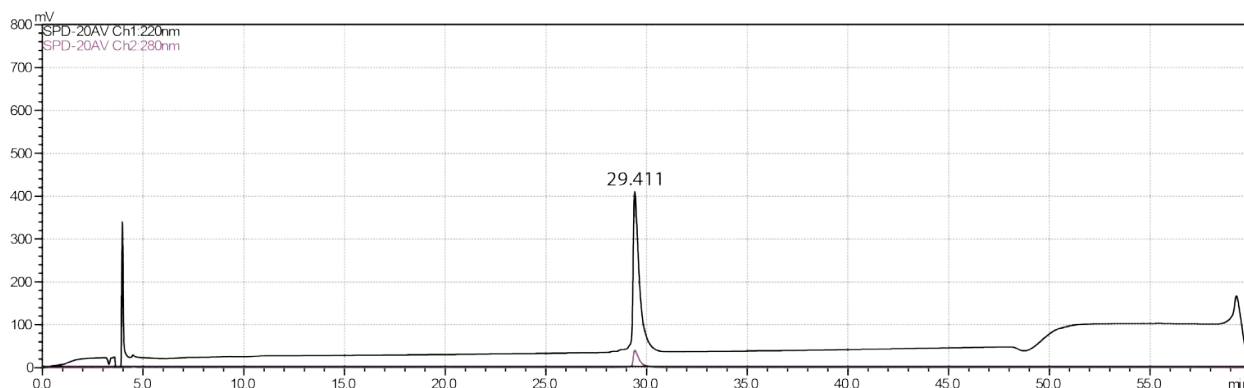


Figure 2-132. Analytical HPLC Data for **15-AY_{me}K**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 20-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 40 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

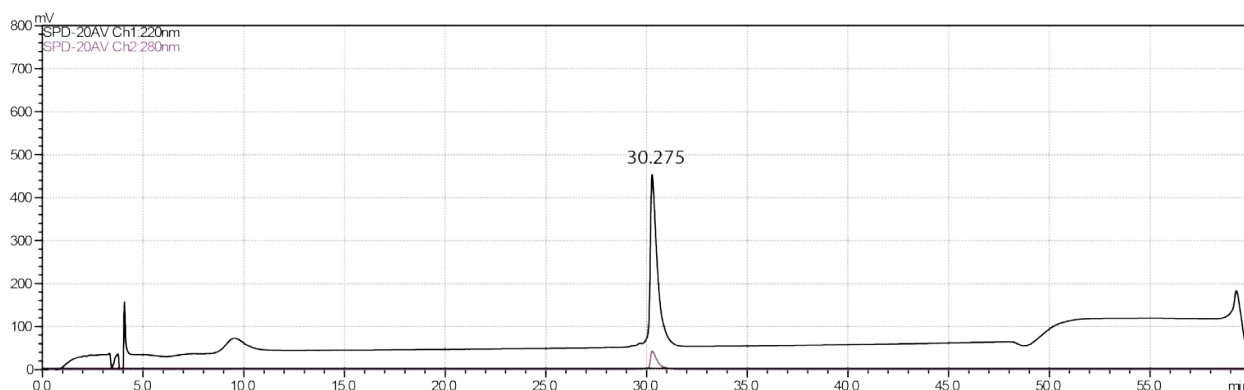


Figure 2-133. Analytical HPLC Data for **15-EY_{me}A**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 20-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 40 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

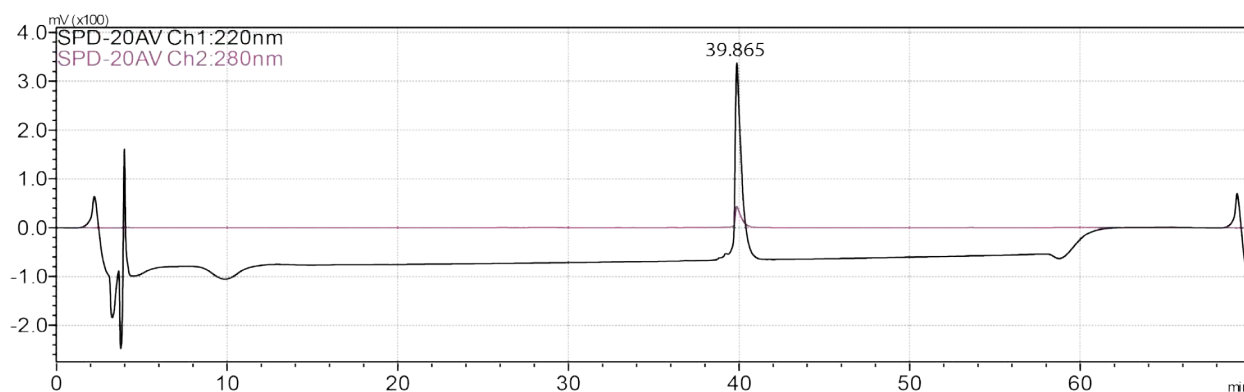


Figure 2-134. Analytical HPLC Data for **15-AY_{me}A**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 20-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 40 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

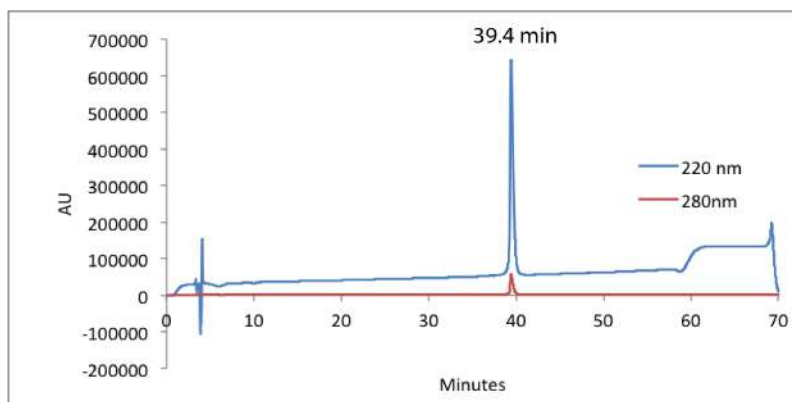


Figure 2-135. Analytical HPLC Data for **16**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

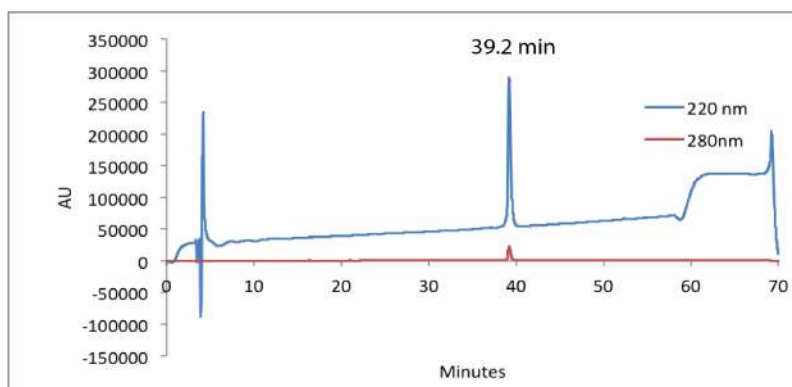


Figure 2-136. Analytical HPLC Data for **16-AFK**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

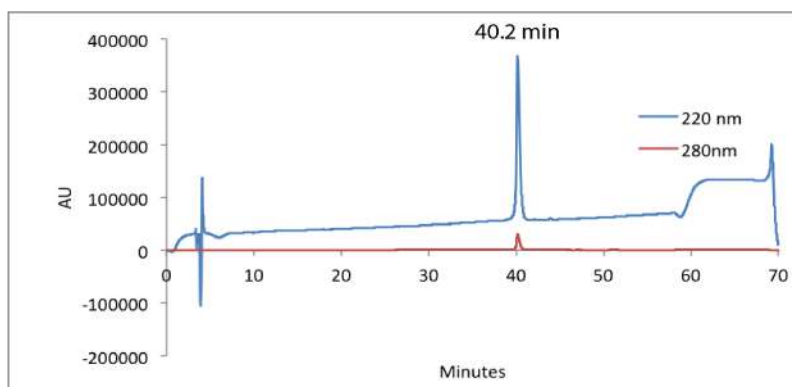


Figure 2-137. Analytical HPLC Data for **16-EFA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

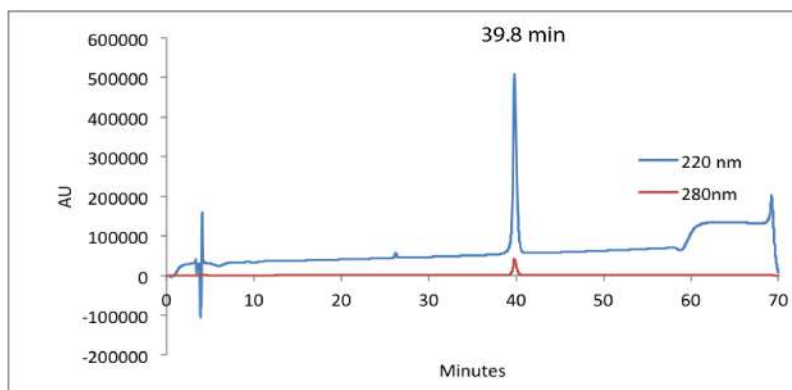


Figure 2-138. Analytical HPLC Data for **16-AFA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

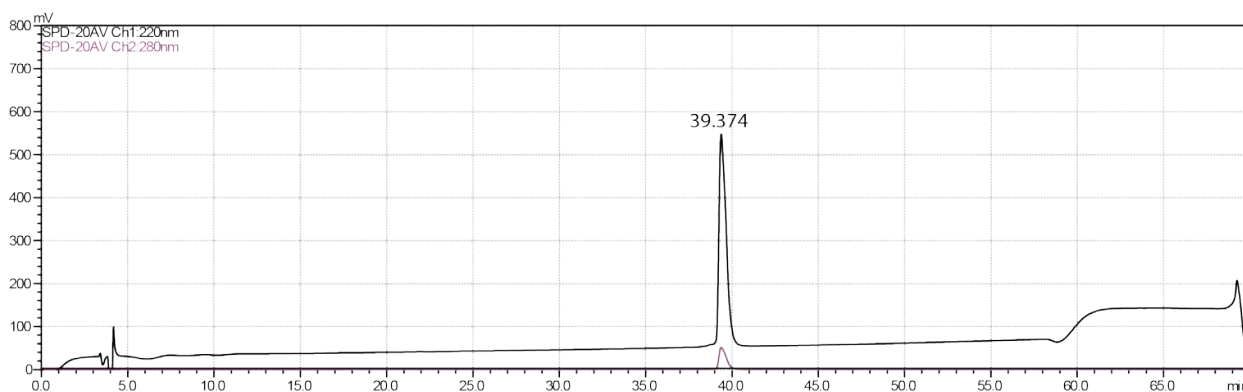


Figure 2-139. Analytical HPLC Data for **17**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

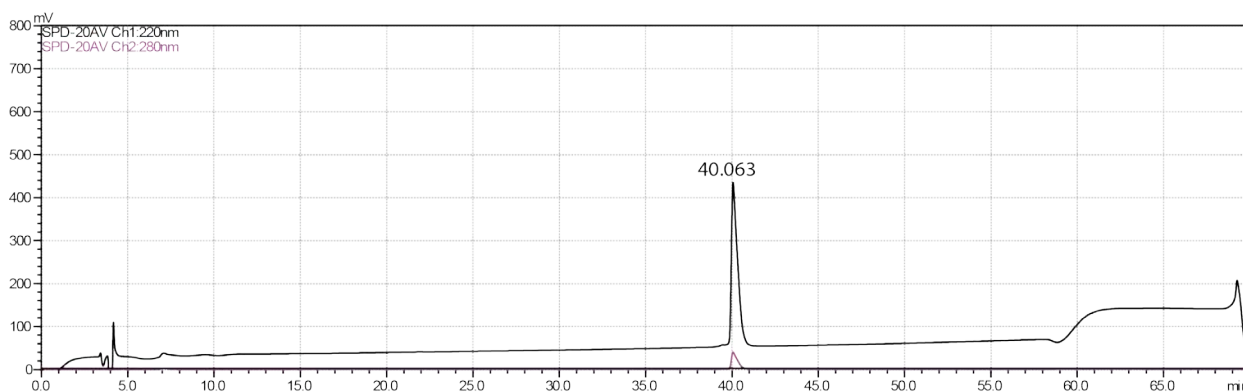


Figure 2-140. Analytical HPLC Data for **17-A(f₅F)K**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

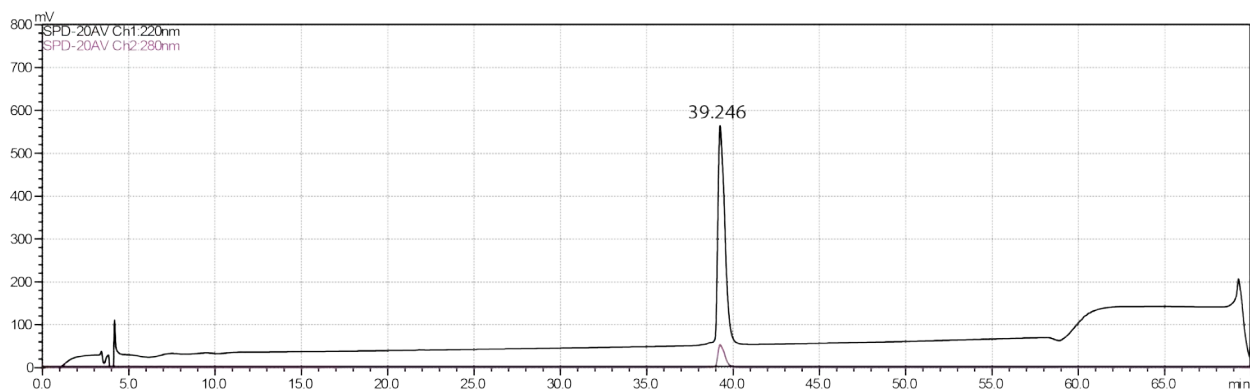


Figure 2-141. Analytical HPLC Data for **17-E(f₅F)A**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

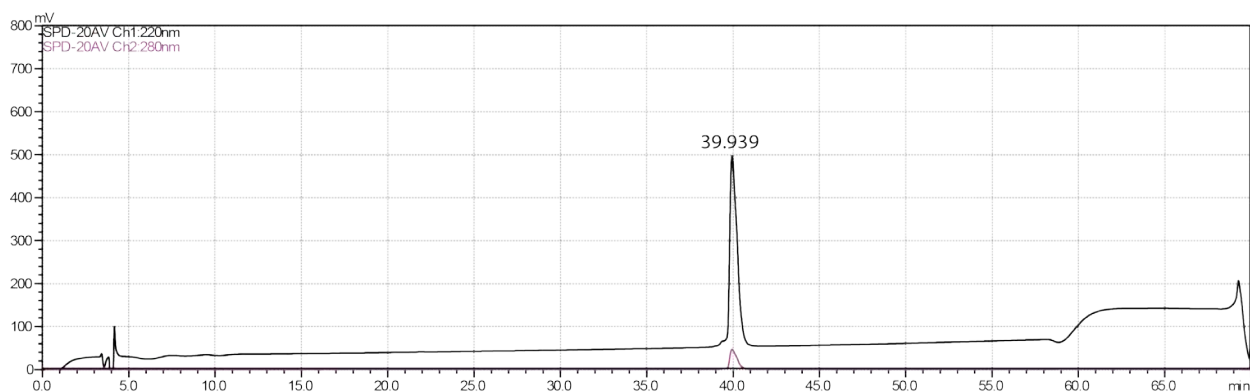


Figure 2-142. Analytical HPLC Data for **17-A(f₅F)A**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

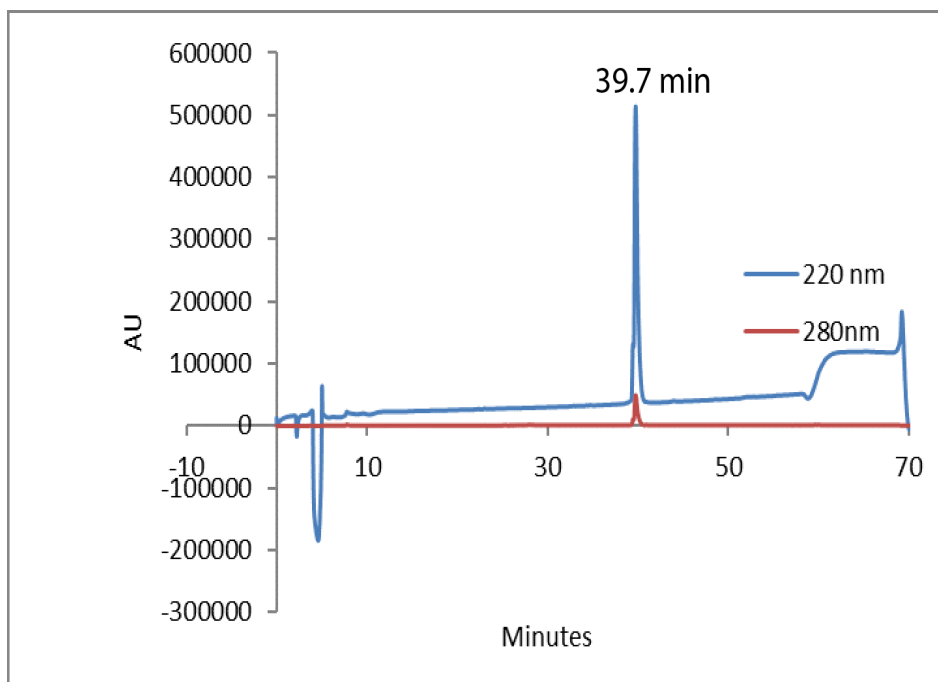


Figure 2-143. Analytical HPLC Data for **18**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min

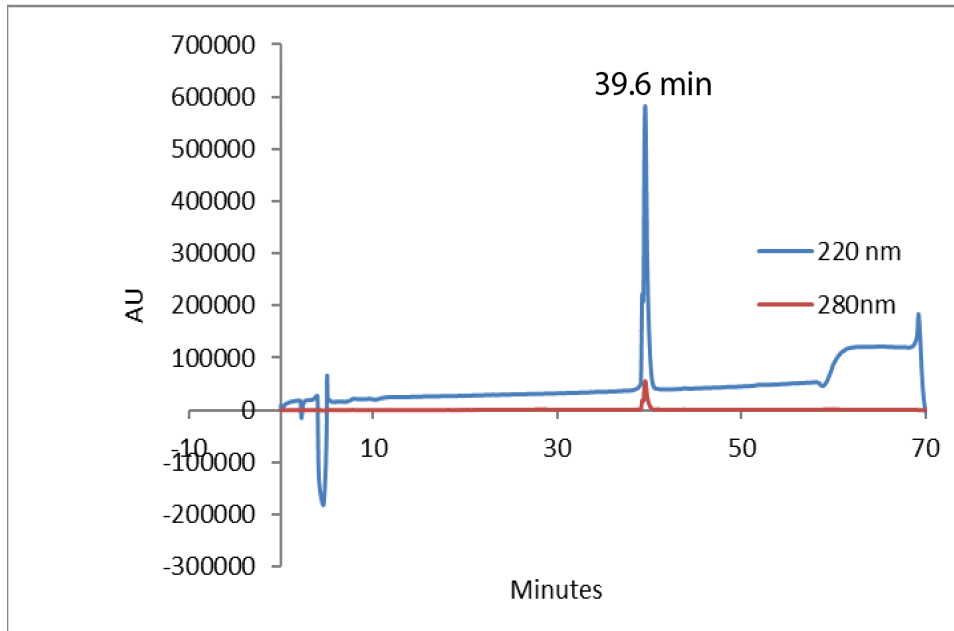


Figure 2-144. Analytical HPLC Data for **18-A(p-fF)K**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min

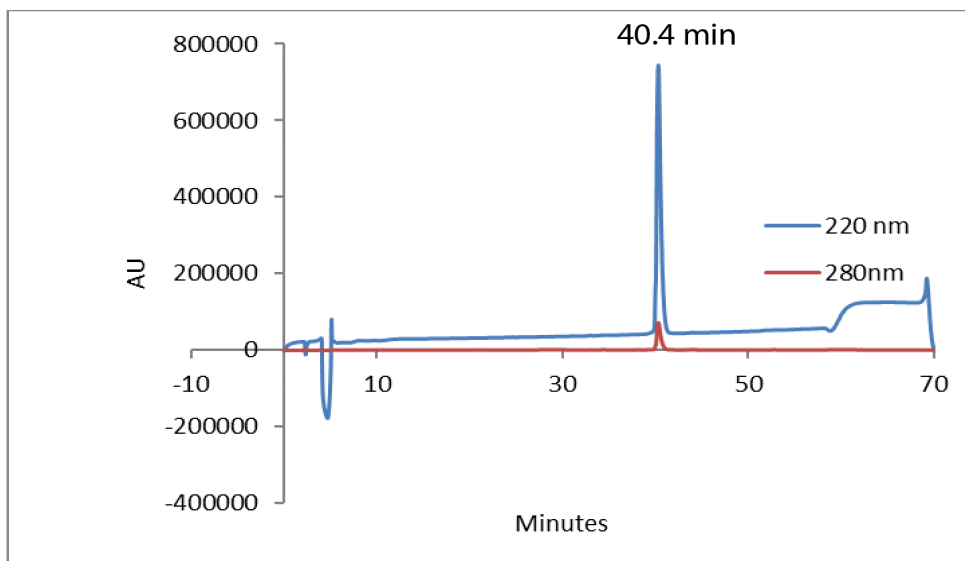


Figure 2-145. Analytical HPLC Data for **18-E(p-fF)A**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min

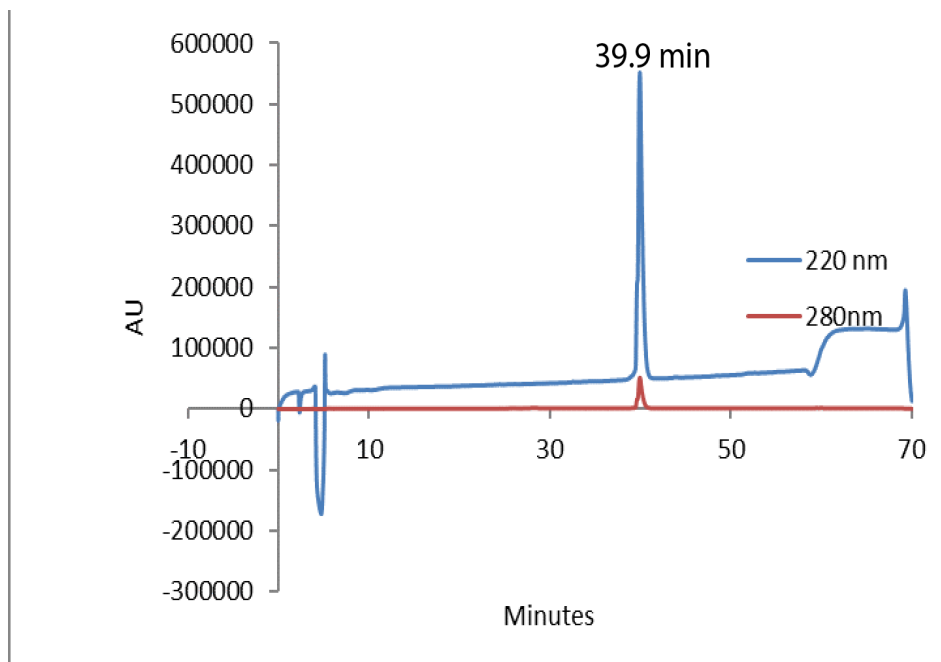


Figure 2-146. Analytical HPLC Data for **18-A(p-fF)A**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min

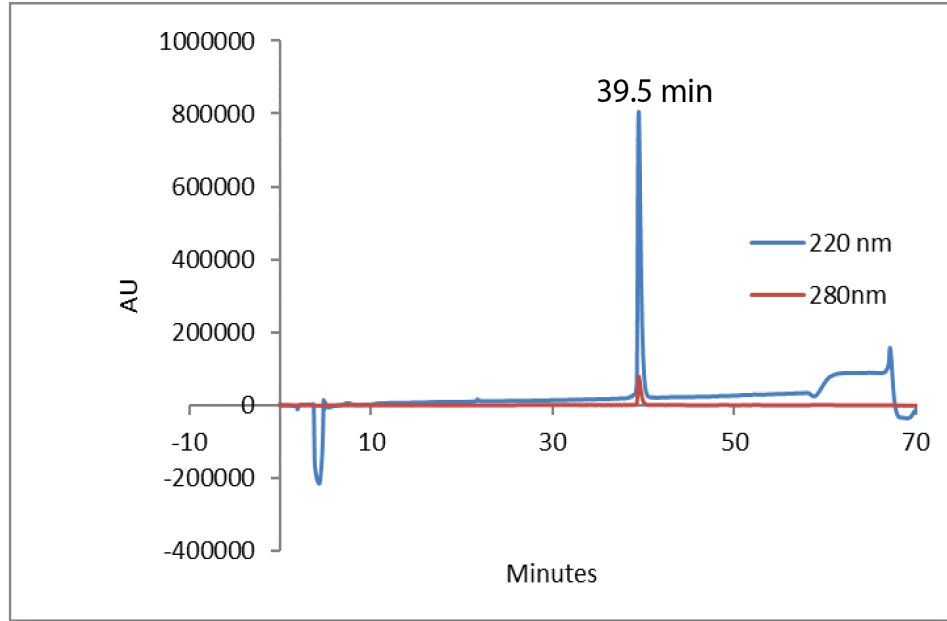


Figure 2-147. Analytical HPLC Data for **19**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

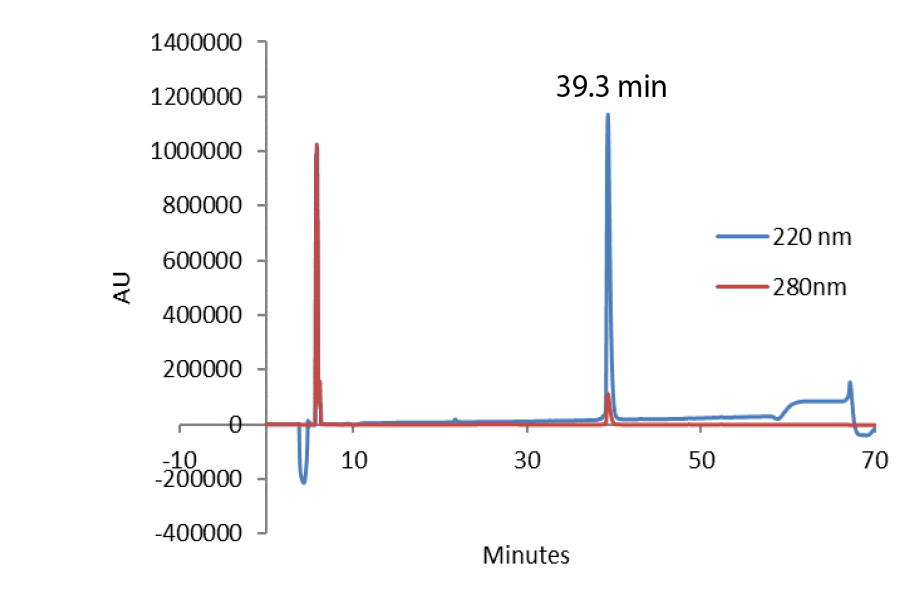


Figure 2-148. Analytical HPLC Data for **19-A(3,4-f₂F)K**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min

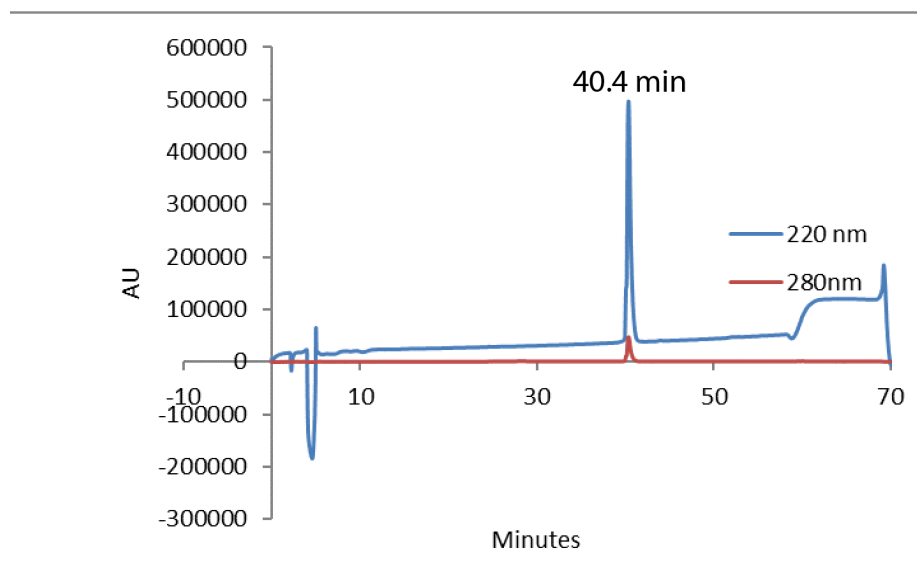


Figure 2-149. Analytical HPLC Data for **19-E(3,4-f₂F)K**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min

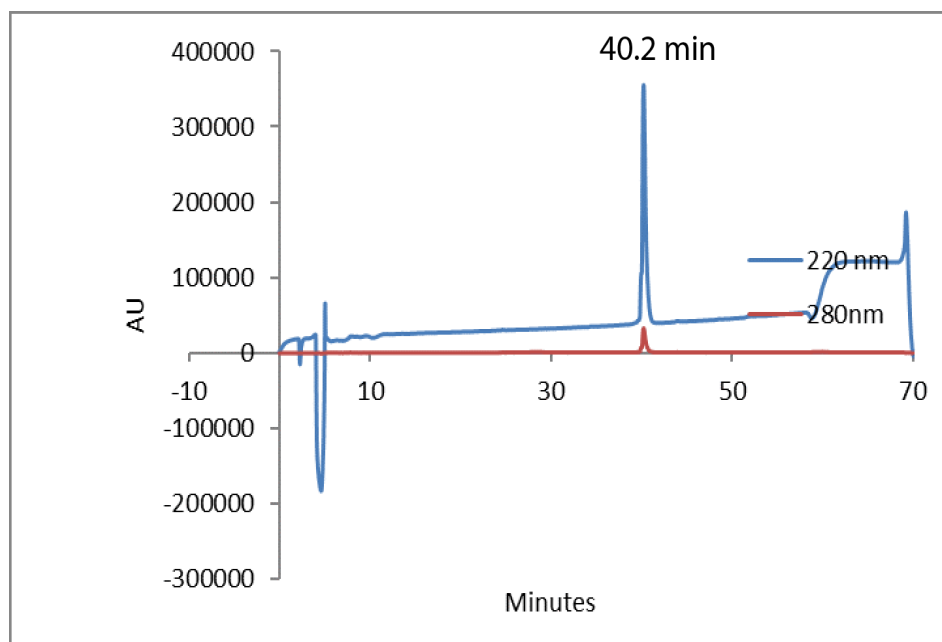


Figure 2-150. Analytical HPLC Data for **19-A(3,4-f₂F)A**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min

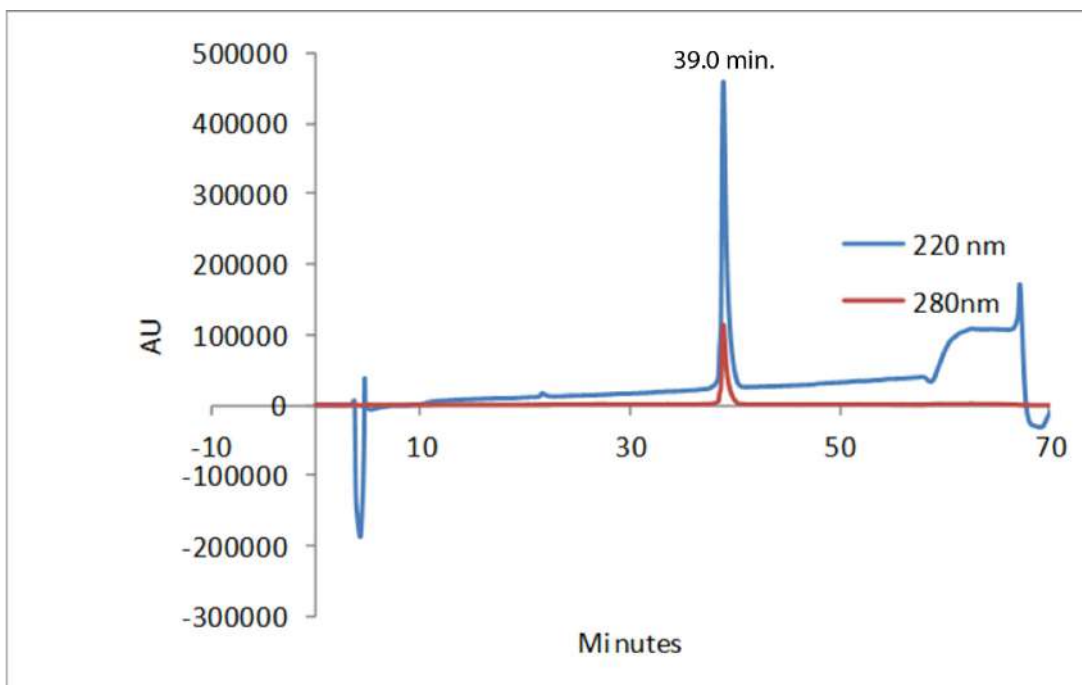


Figure 2-151. Analytical HPLC Data for **20**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

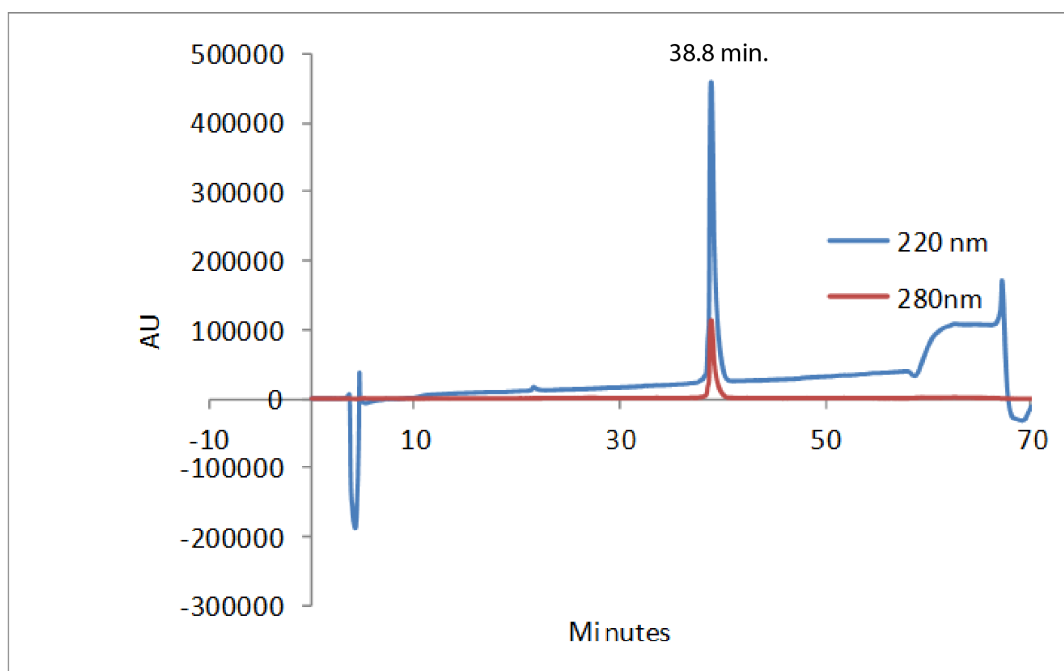


Figure 2-152. Analytical HPLC Data for **20-A(p-NO₂F)K**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min

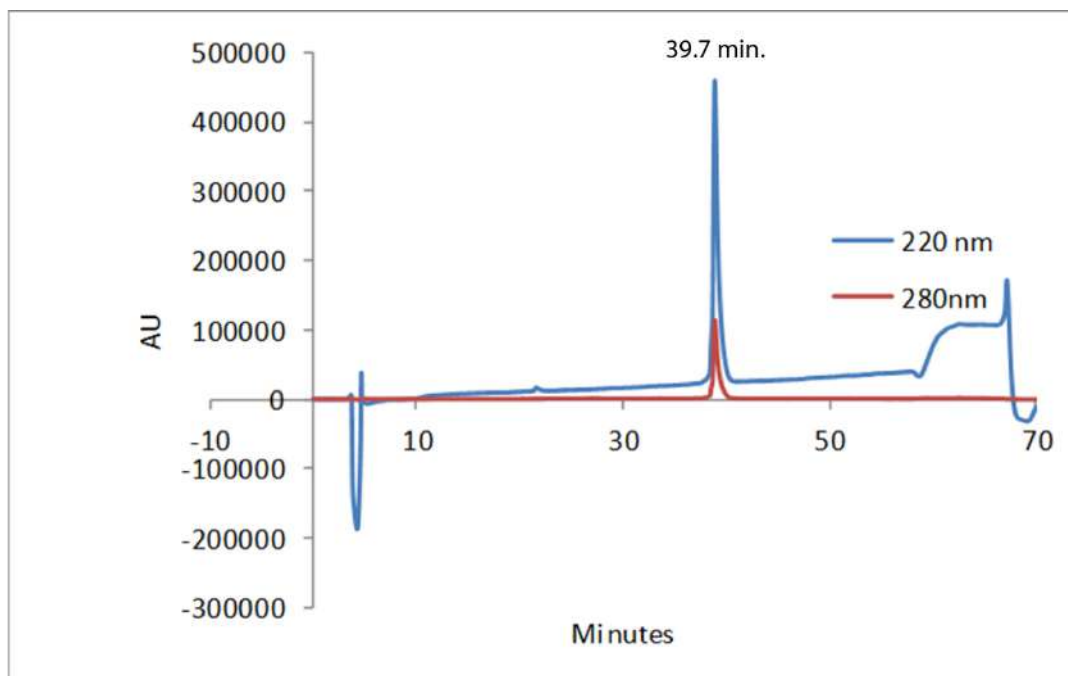


Figure 2-153. Analytical HPLC Data for **20-E(p-NO₂F)A**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min

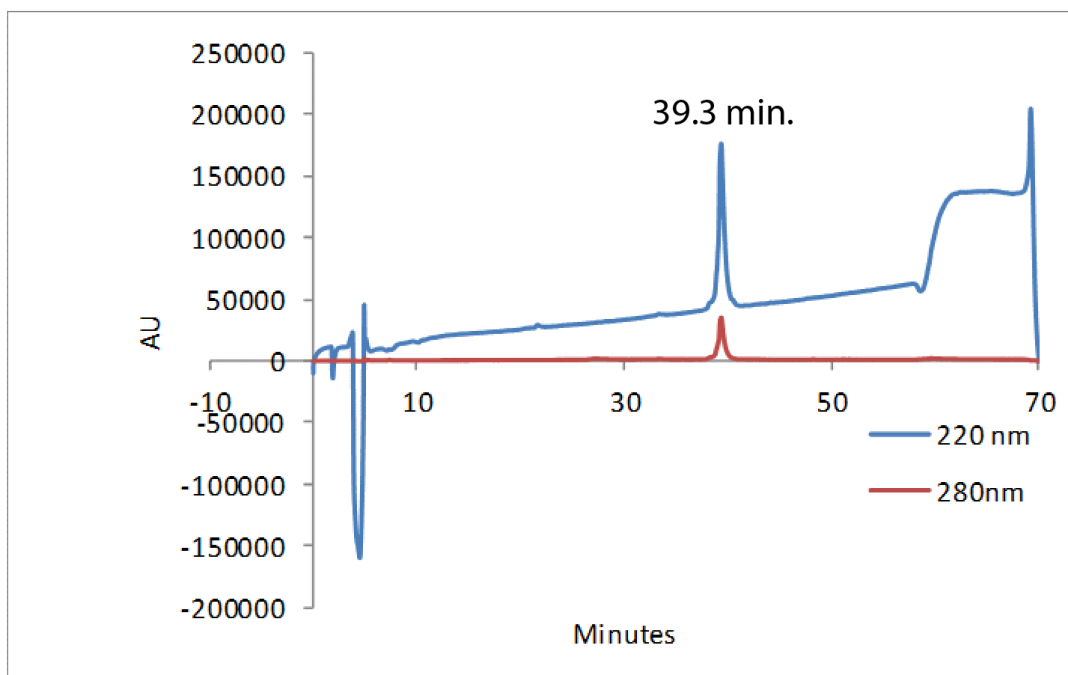


Figure 2-154. Analytical HPLC Data for **20-A(p-NO₂F)A**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

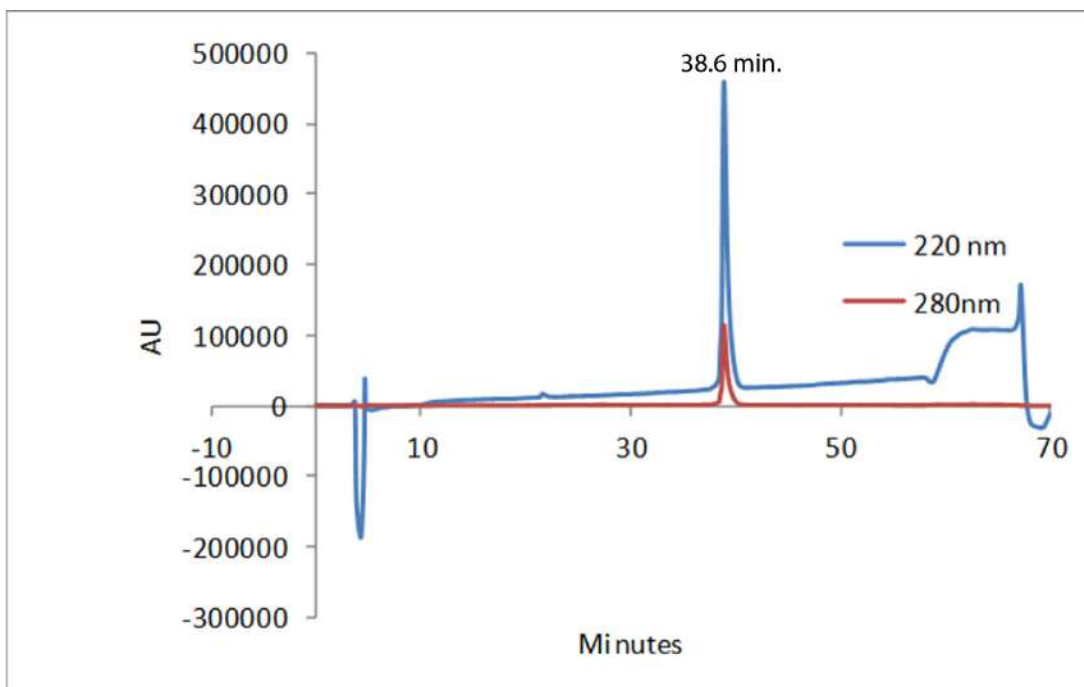


Figure 2-155. Analytical HPLC Data for **21**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

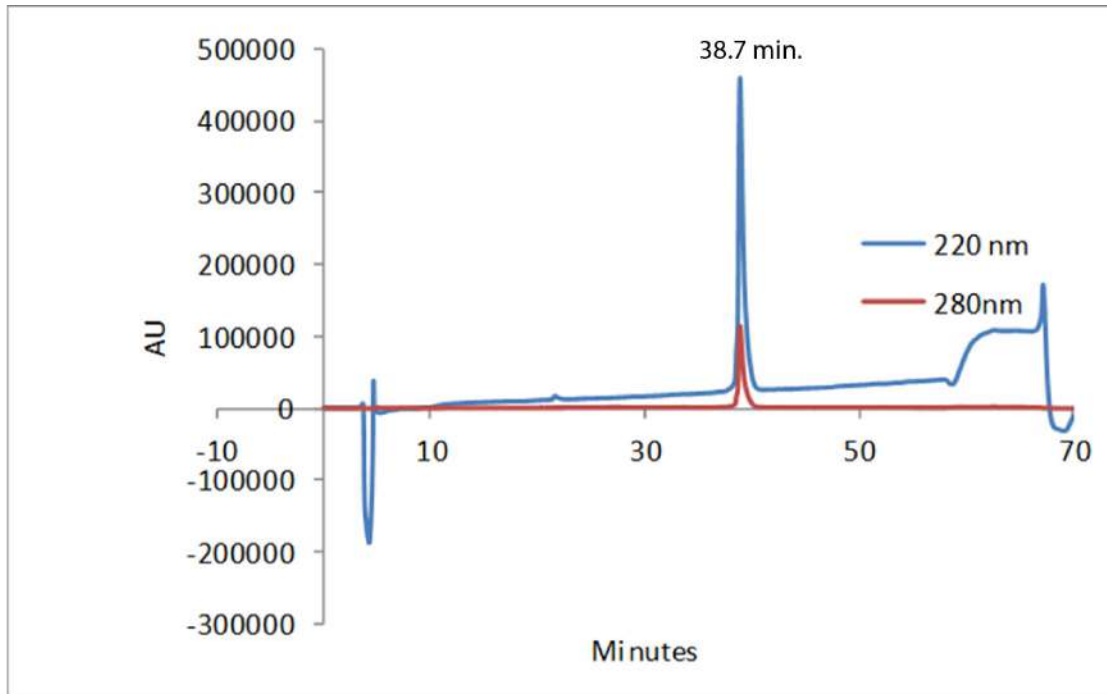


Figure 2-156. Analytical HPLC Data for **21-A(PyrA)K**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

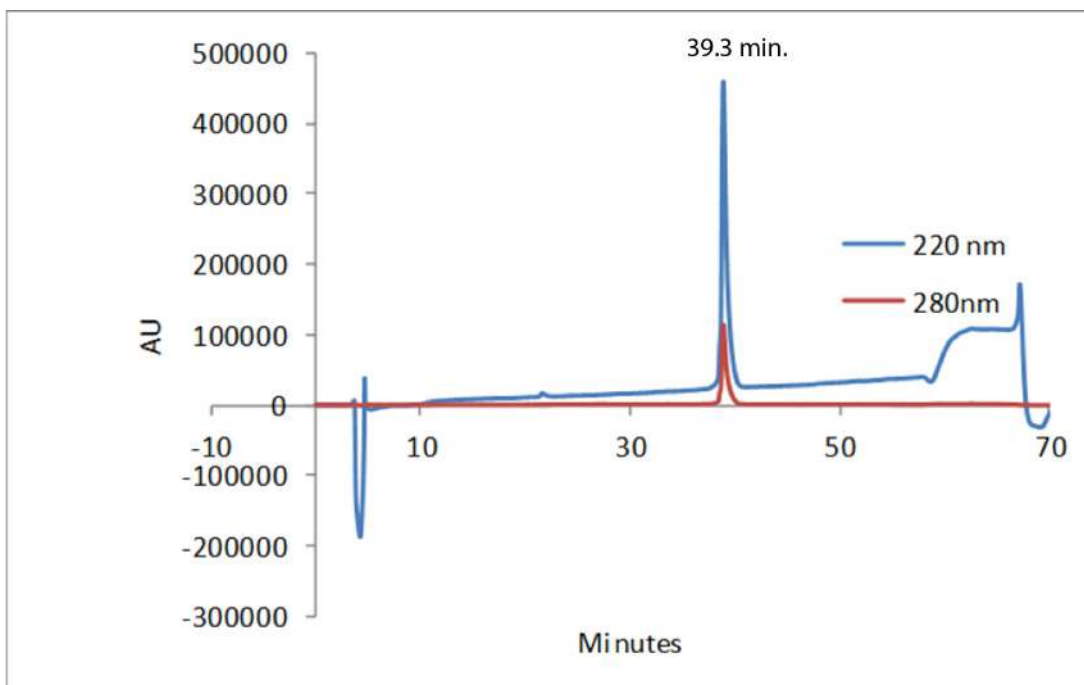


Figure 2-157. Analytical HPLC Data for **21-E(PyrA)A**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

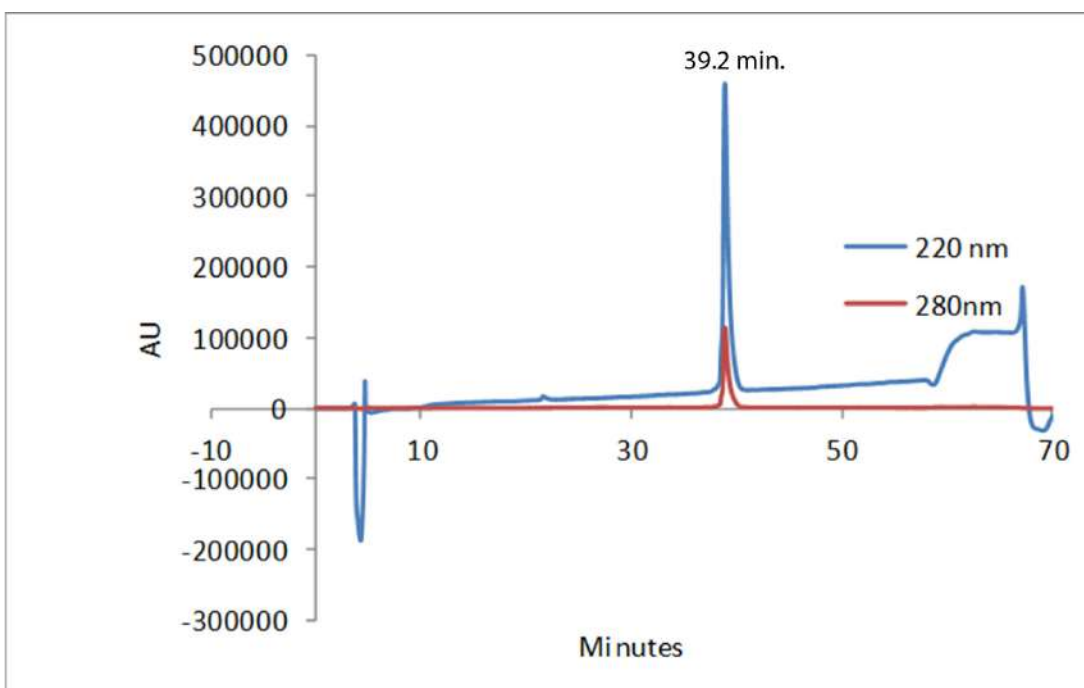


Figure 2-158. Analytical HPLC Data for **21-A(PyrA)A**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

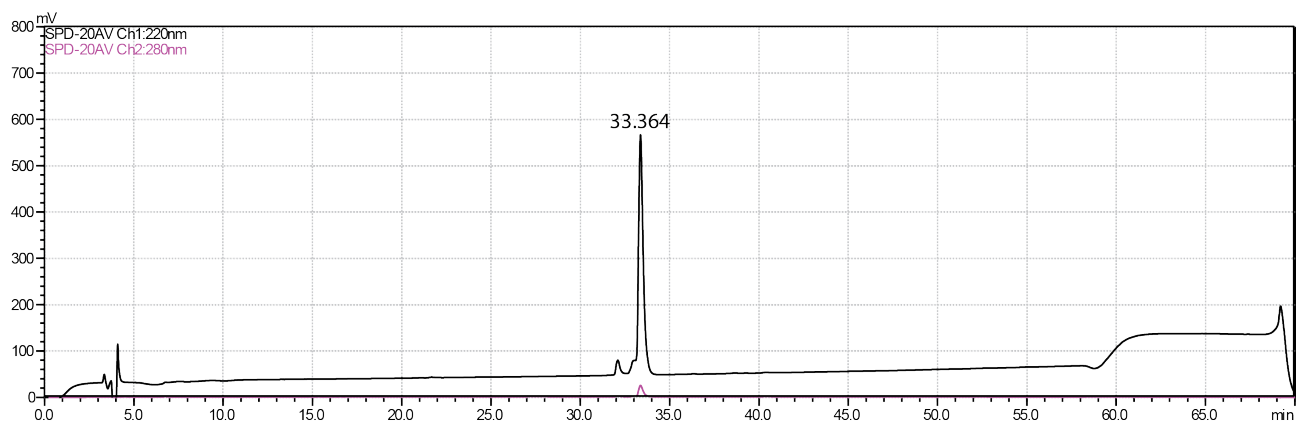


Figure 2-159. Analytical HPLC Data for **22**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

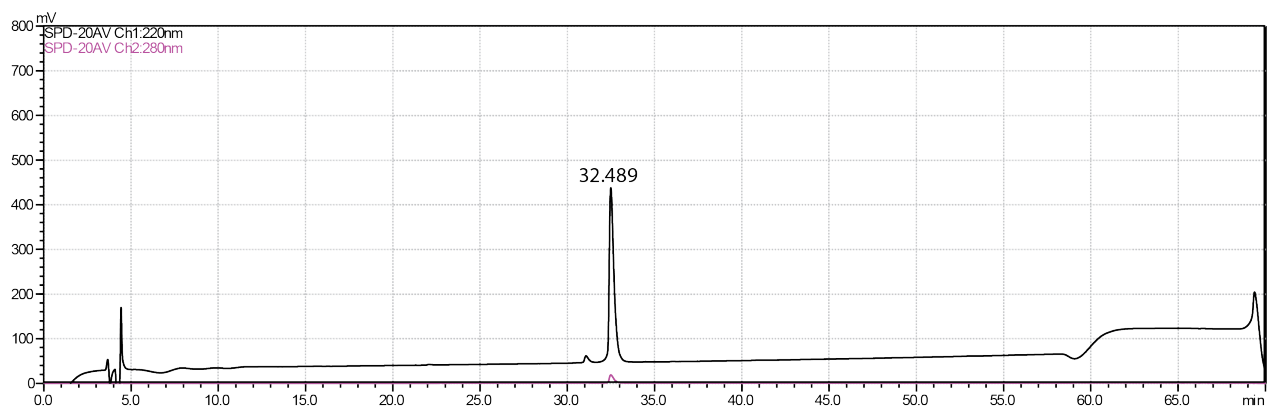


Figure 2-160. Analytical HPLC Data for **22-AFK**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

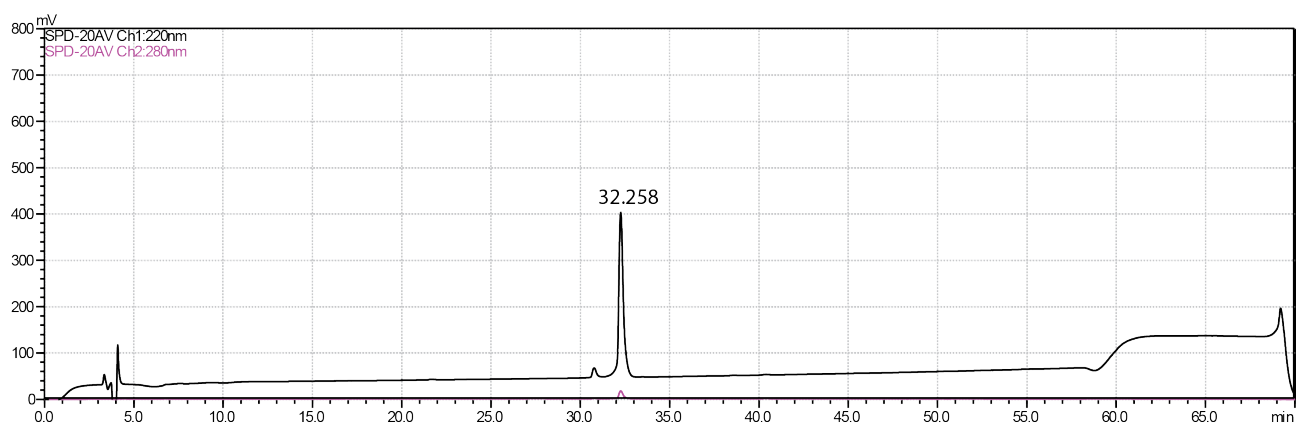


Figure 2-161. Analytical HPLC Data for **22-EFA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over

50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

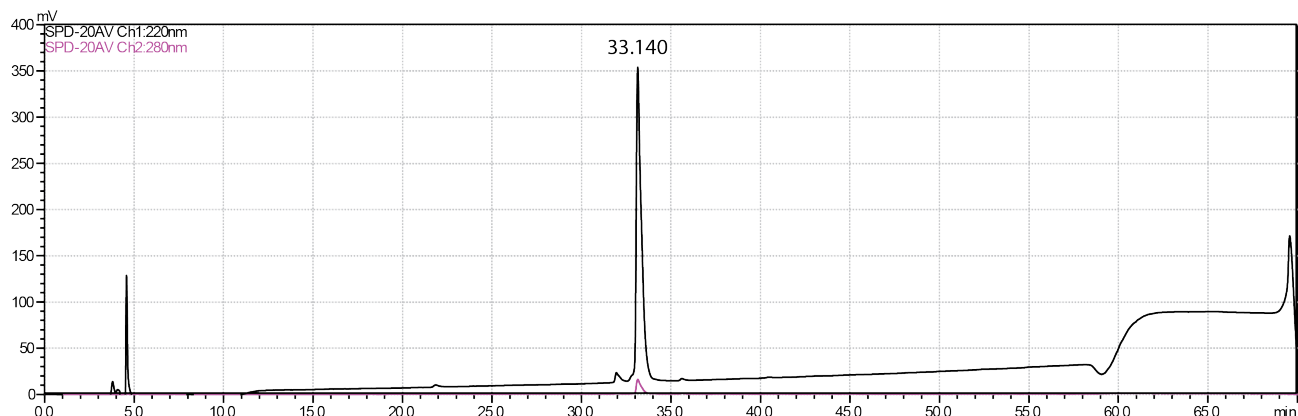


Figure 2-162. Analytical HPLC Data for **22-AFA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

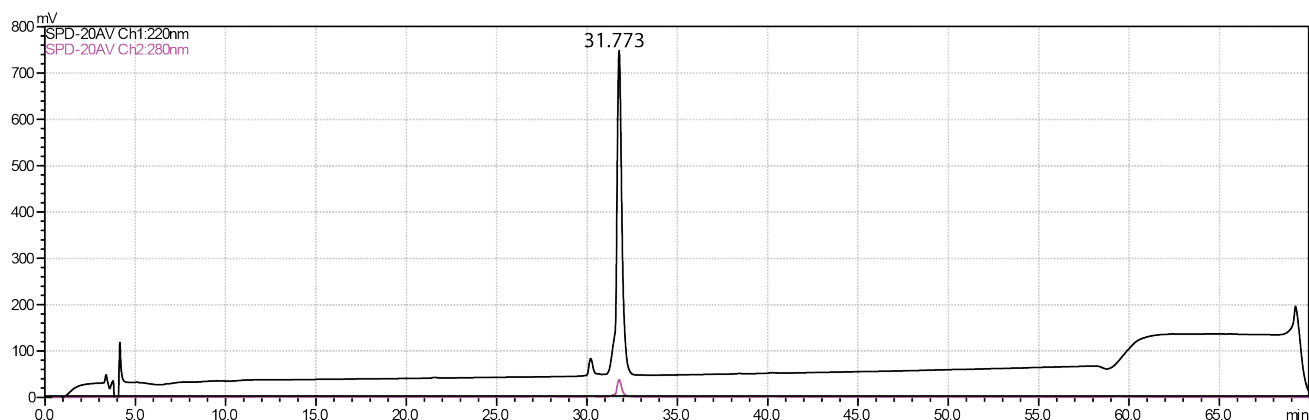


Figure 2-163. Analytical HPLC Data for **22-ESK**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

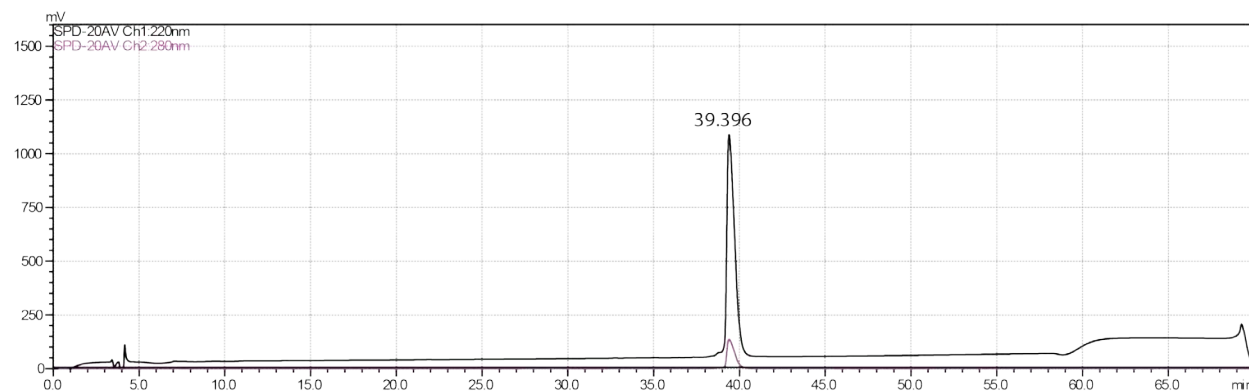


Figure 2-164. Analytical HPLC Data for **22-ASK**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

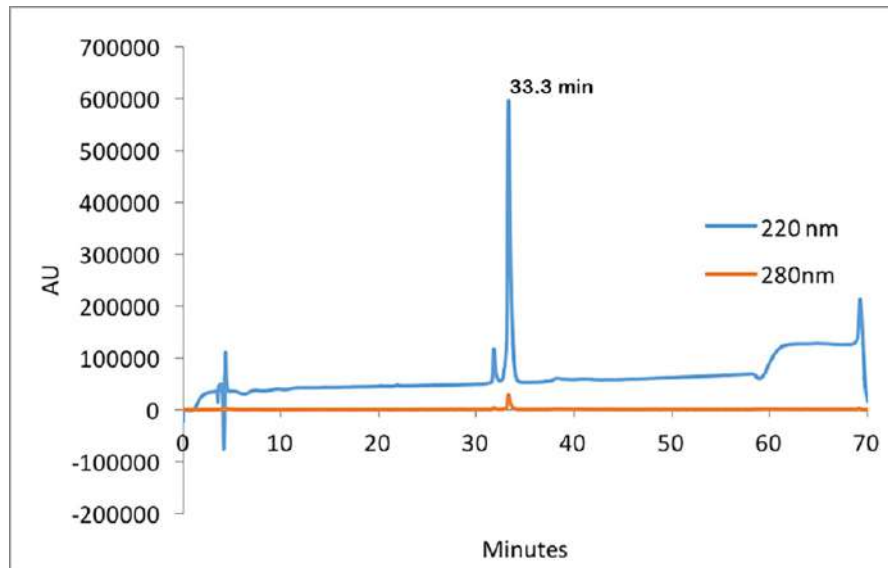


Figure 2-165. Analytical HPLC Data for **22-ESA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

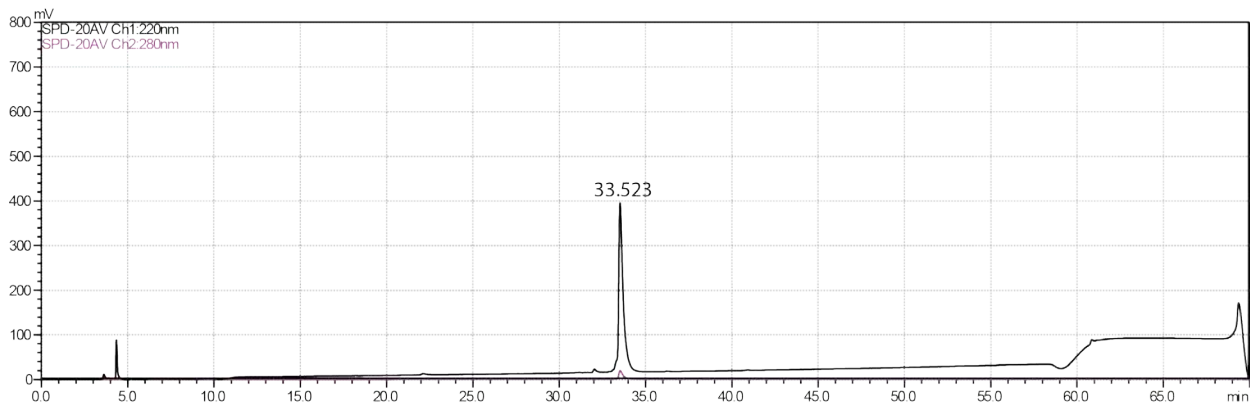


Figure 2-166. Analytical HPLC Data for **22-ASA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

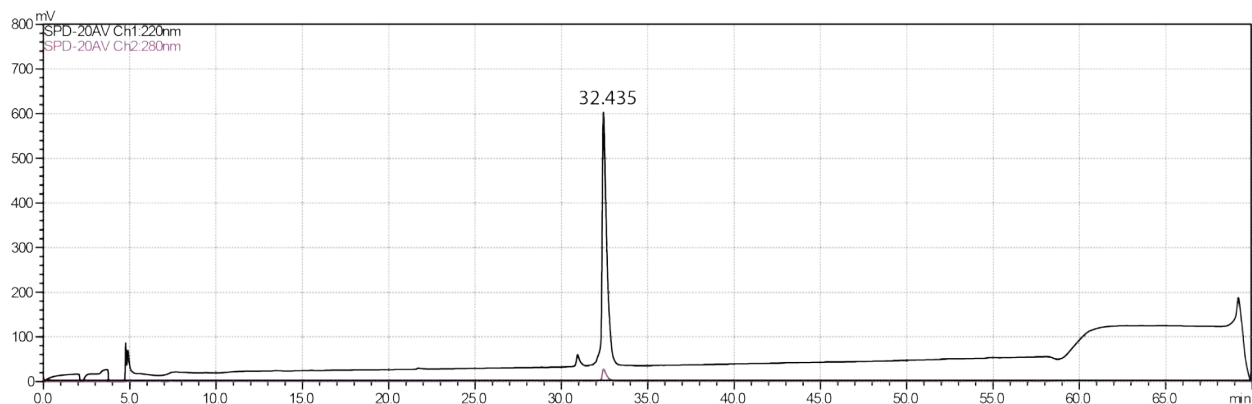


Figure 2-167. Analytical HPLC Data for **23**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min

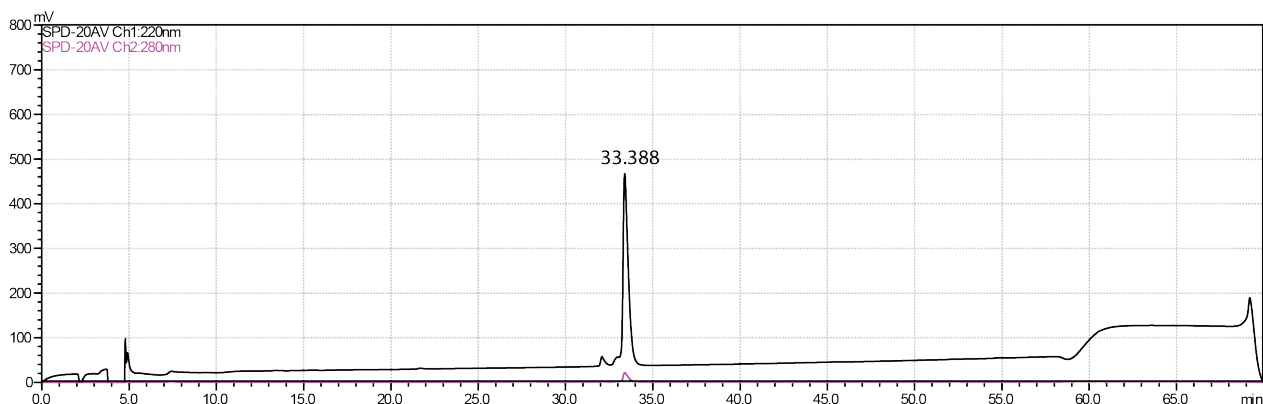


Figure 2-168. Analytical HPLC Data for **23-EFA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

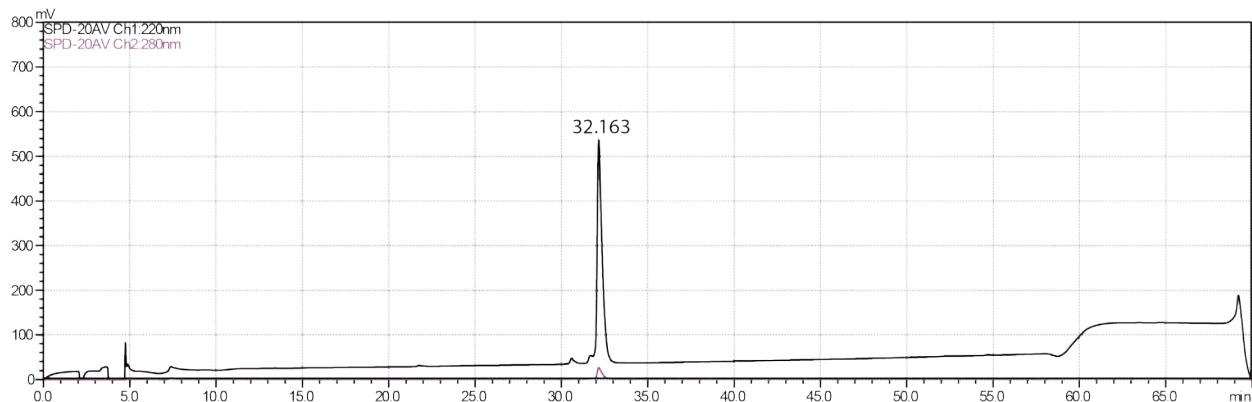


Figure 2-169. Analytical HPLC Data for **23-ESK**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over

50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

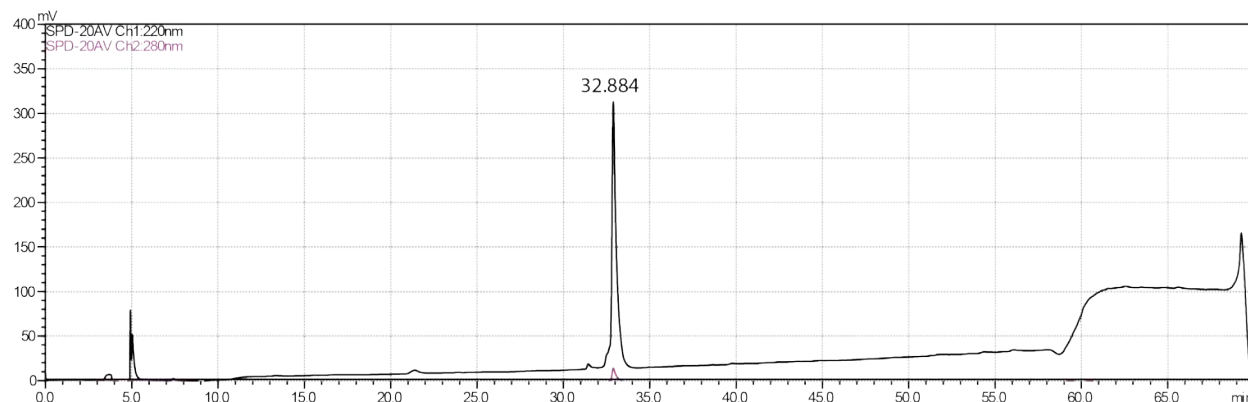


Figure 2-170. Analytical HPLC Data for **23-ESA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

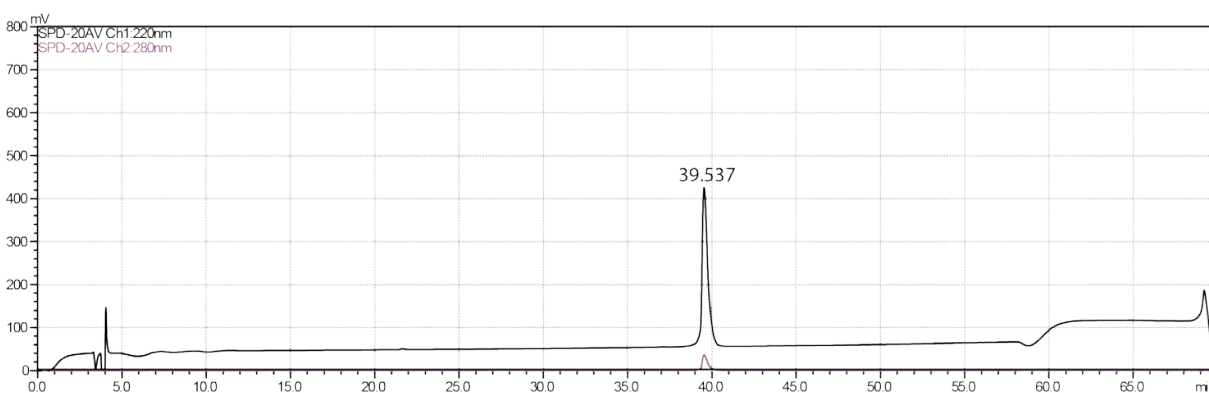


Figure 2-171. Analytical HPLC Data for **24**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

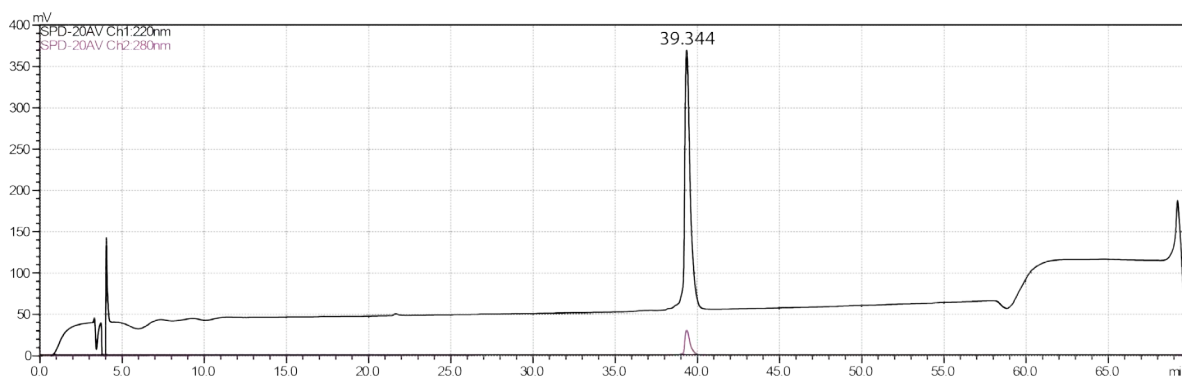


Figure 2-172. Analytical HPLC Data for **24-AFR**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over

50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

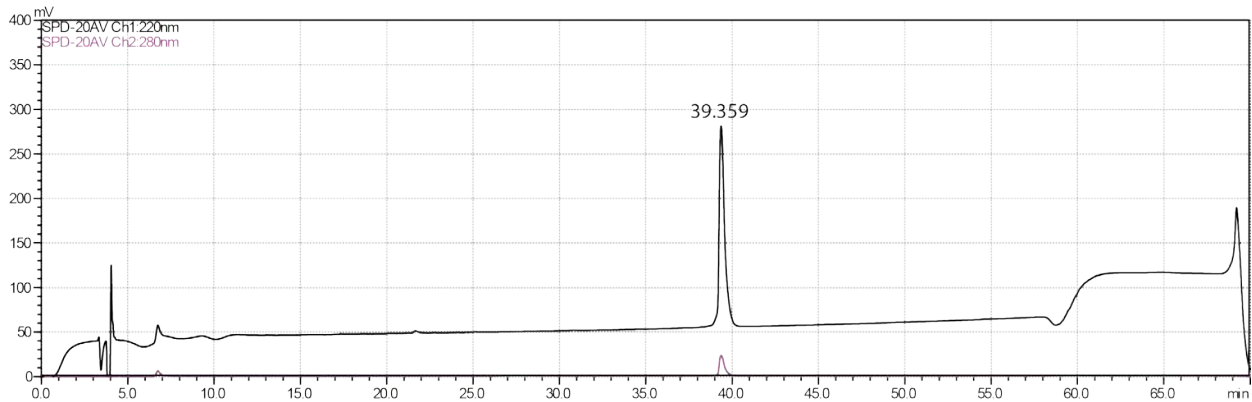


Figure 2-173. Analytical HPLC Data for **24-ESR**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

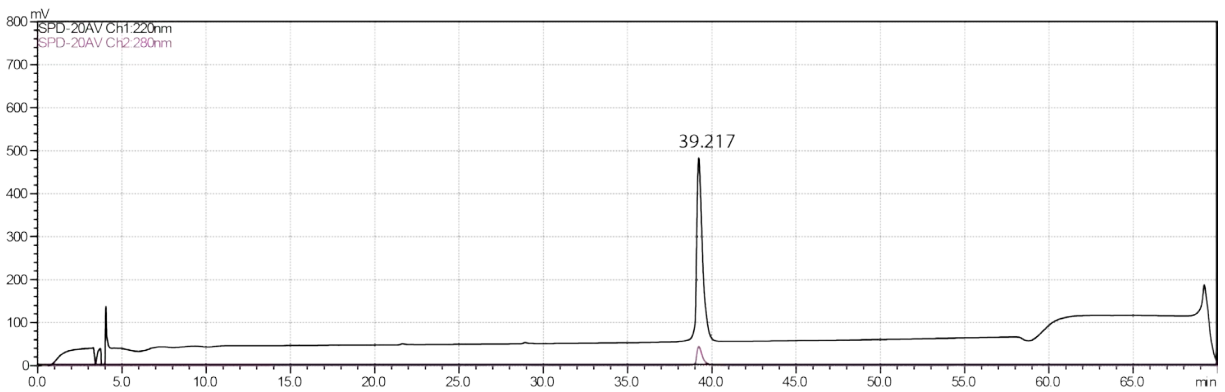


Figure 2-174. Analytical HPLC Data for **24-ASR**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

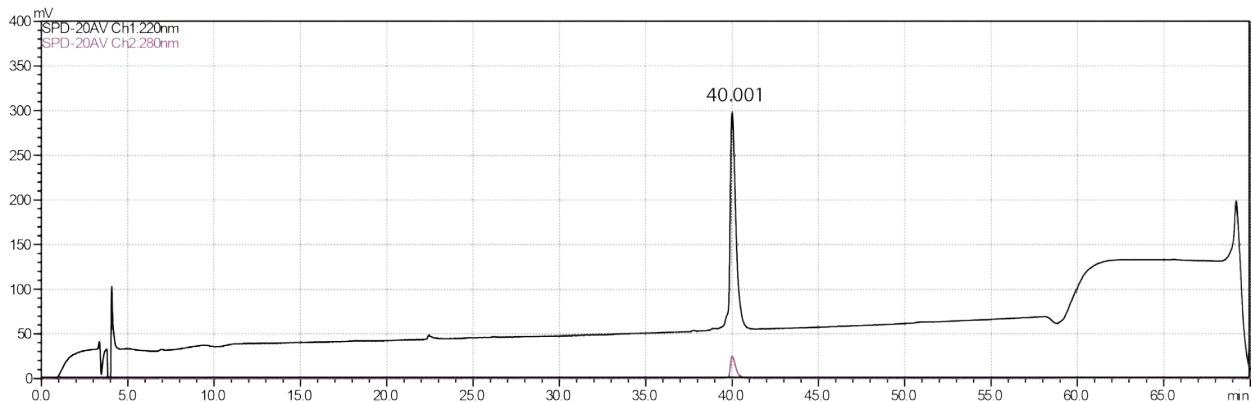


Figure 2-175. Analytical HPLC Data for **25**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

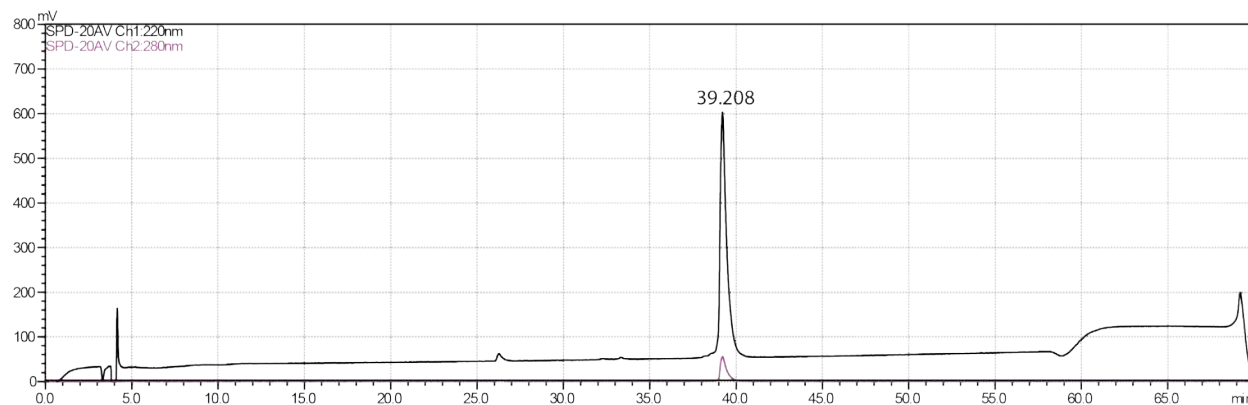


Figure 2-176. Analytical HPLC Data for **25-AAK**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

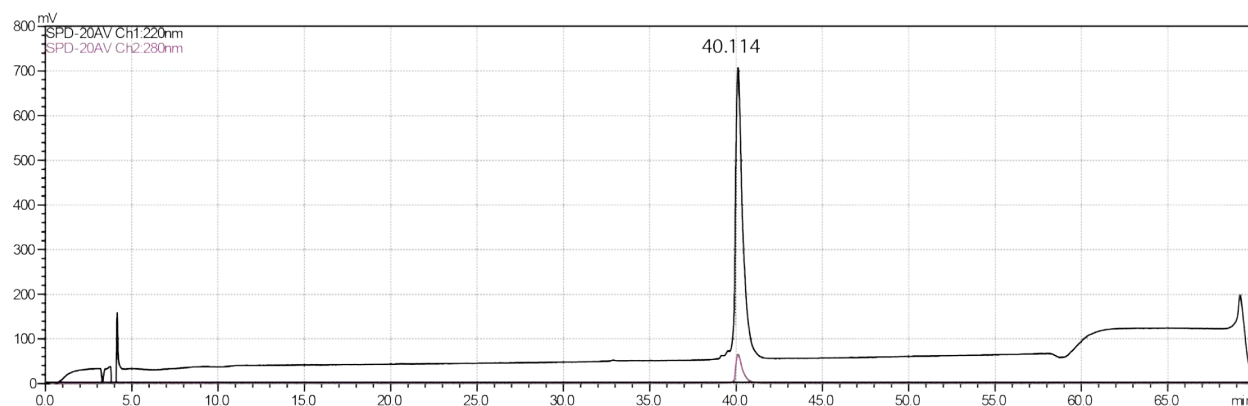


Figure 2-177. Analytical HPLC Data for **25-EAA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

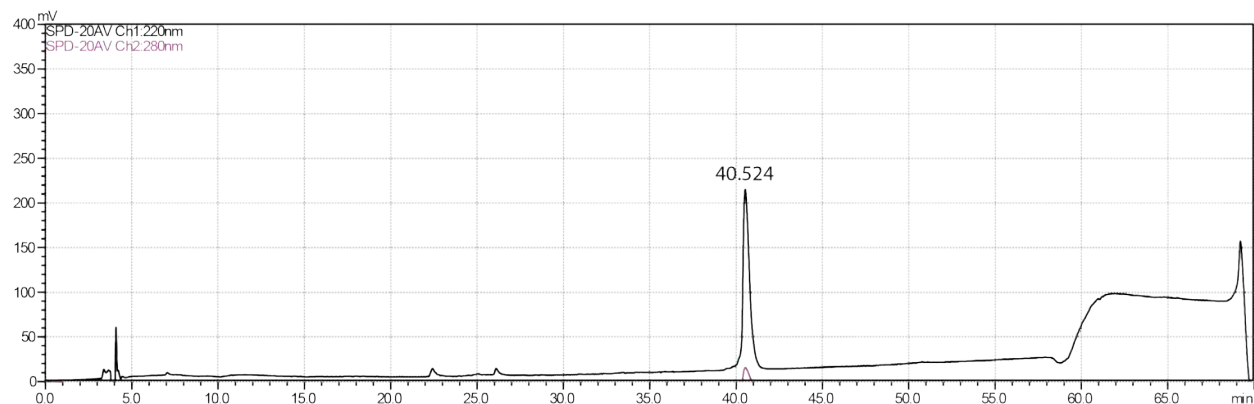


Figure 2-178. Analytical HPLC Data for **25-AAA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

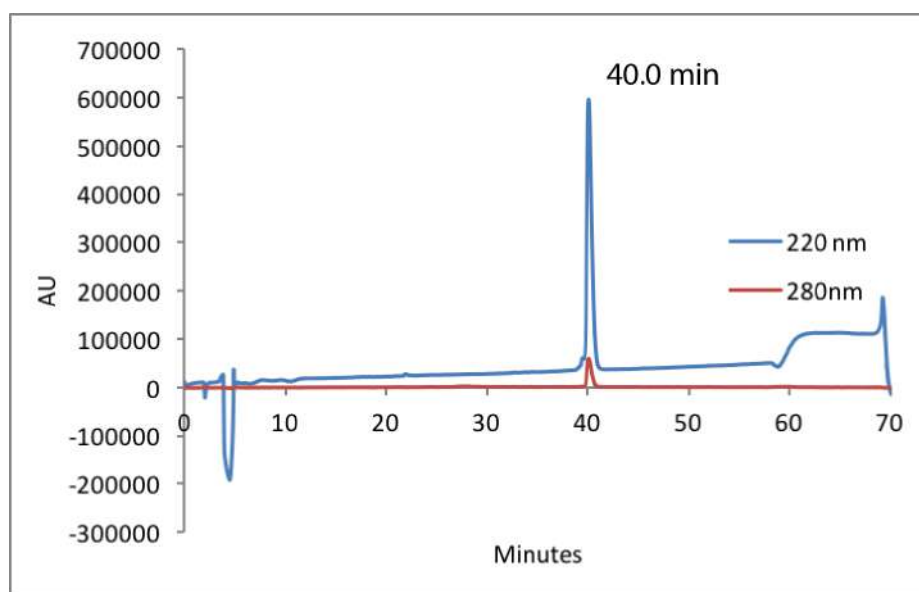


Figure 2-179. Analytical HPLC Data for **26**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

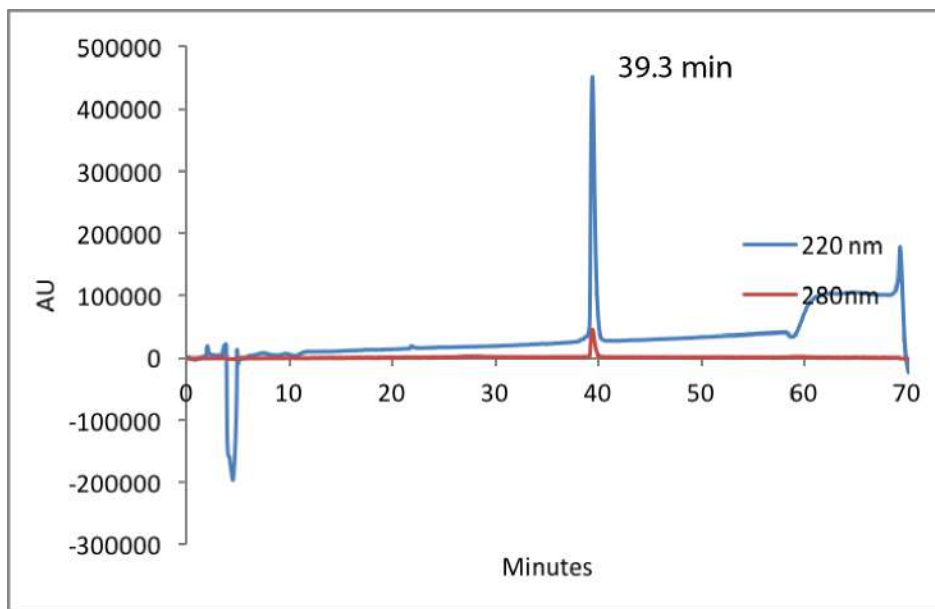


Figure 2-180. Analytical HPLC Data for **27**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

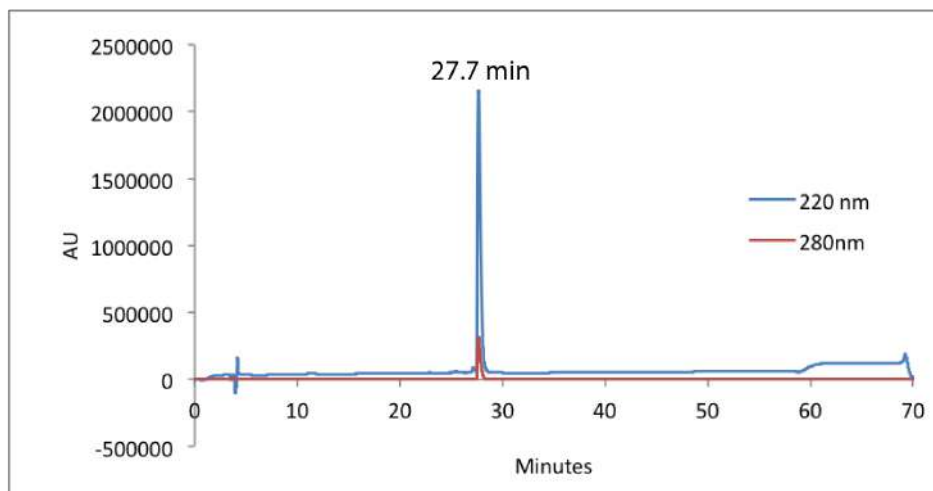


Figure 2-181. Analytical HPLC Data for **pA-EWK**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

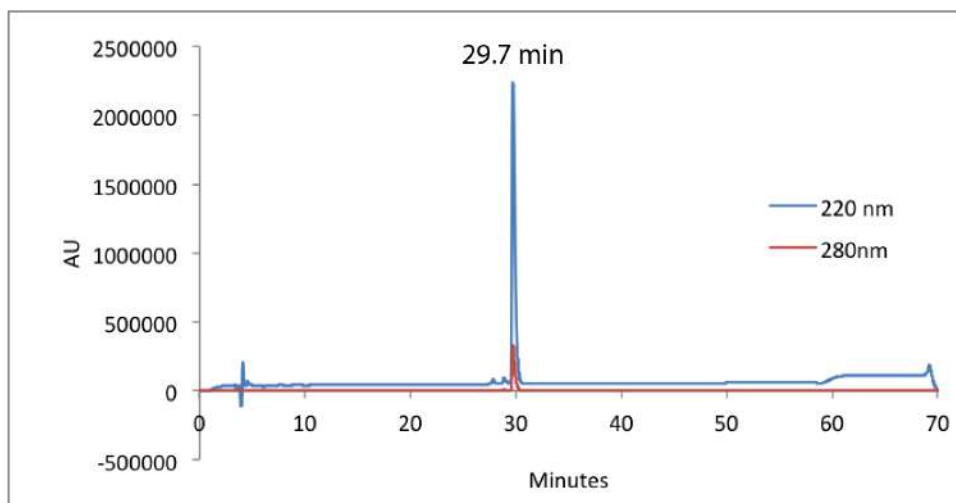


Figure 2-182. Analytical HPLC Data for **pA-EWA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

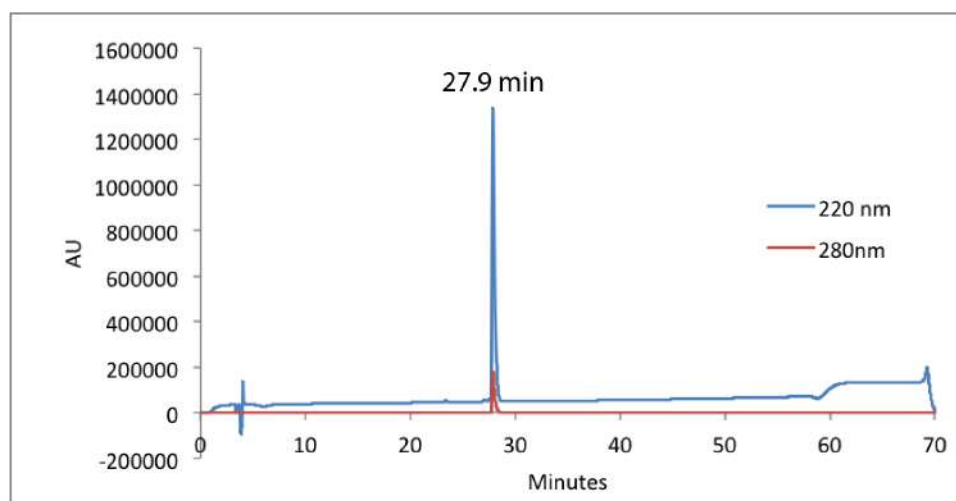


Figure 2-183. Analytical HPLC Data for **pA-AWK**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

2.4.4 Size exclusion chromatography

Previously characterized peptide 1CW adopts a homotrimeric self-association state in solution, whereas GCN4 adopts a homodimeric self-association state in solution. The large number of peptides explored here (78 peptides including 1–27 and their sequence variants shown

in Supplementary Table 1) precluded the use of time and resource-intensive sedimentation equilibrium experiments to characterize their self-association properties. Consequently, we used the higher throughput size exclusion chromatography to characterize the self-association properties of 1–27 (and their sequence variants) by comparing their retention times on a size-exclusion column to the retention times of homotrimeric 1CW, homodimeric GCN4 and monomeric α -helical PSBD36.^{9, 35-36}

Size exclusion chromatography (SEC) was done on a Shimadzu HPLC instrument using a Phenomenex yarra 3u sec-3000 column (batches 1 and 2) or a Zenix-C SEC 100 column (batches 3, 4, and 5). The columns were calibrated with internal 1CW, GCN4, and PSBD36 standards. Previous characterization of 1CW, GCN4, and PSBD36 by sedimentation equilibrium analytical ultracentrifugation under analogous buffer conditions demonstrates that 1CW adopts a trimeric association state; that GCN4 adopts a dimeric state; and that PSBD36 is an α -helical monomer. The retention times derived from SEC experiments on peptides 1-19 and 23-26 are very close to that of trimeric 1CW, suggesting that these variants likewise adopt a trimeric association state. Similarly, the retention times derived from SEC experiments on peptides 20-22 are close to that of dimeric GCN4, suggesting that these variants likewise adopt a dimeric association state. Moreover, published peptide coil-VaLd (which has the same sequence as peptide 5, except that 5 has four additional C-terminal residues) crystallizes as a trimer,¹⁹ as do other coiled-coil peptides in which beta-branched non-polar residues (i.e. Ile and/or Val) occupy a-positions and Leu occupies d-positions in the canonical heptad repeat.³⁷ Therefore, we are reasonably confident that 1-19 and 23-24 adopt trimeric association states and that 20-22 adopt dimeric association states.

Table 2-4 Size exclusion chromatography of variants 1-27 and sequence variants. Batches 1-2 were done on a Phenomenex yarra 3u sec-3000 column. Batches 3-5 were done on a Zenix-C SEC 100 column.

Batch 1			
Peptide	Trimer MW (Da)	Retention Time	Inferred association state
1CW (trimer standard)	11017	10.50	
8	11016	10.50	Trimer
16	11541	10.54	Trimer
16-EFA	11367	10.38	Trimer
16-AFK	11370	10.51	Trimer
5	11361	10.43	Trimer
6	11190	10.51	Trimer
7	11187	10.35	Trimer
16-AFA	11196	10.47	Trimer
14-AYA	11244	10.45	Trimer
14-EYA	11416	10.35	Trimer
14-AYK	11419	10.50	Trimer
14-EYA	11590	10.43	Trimer
4	11313	10.50	Trimer
3	11484	10.37	Trimer
2	11488	10.60	Trimer
1	11659	10.55	Trimer
22-AFR	11452	10.56	Trimer
22	11623	10.43	Trimer
22-ASR	11271	10.52	Trimer
22-ESR	11442	10.41	Trimer
11-EWA	11367	10.43	Trimer

11	11541	10.56	Trimer
11-ESA	11187	10.41	Trimer
11-ESK	11361	10.52	Trimer
12-KSA	11190	10.60	Trimer
12-ASE	11187	10.38	Trimer
12-KSE	11361	10.47	Trimer
12-KWA	11488	10.76	Trimer
12-AWE	11484	10.44	Trimer
12	11659	10.51	Trimer
Batch 2			
1CW	11017	10.19	Trimer
9-DSA	11325	10.08	Trimer
9-DSK	11496	10.26	Trimer
9-DWA	11622	10.21	Trimer
9	11793	10.34	Trimer
10-ASO	11322	10.33	Trimer
10-AWO	11619	10.49	Trimer
10-ESO	11496	10.21	Trimer
10	11793	10.29	Trimer
17-A(f5F)A	11652	10.20	Trimer
17-A(f5F)K	11823	10.28	Trimer
17-E(f5F)A	11826	10.08	Trimer
17	11997	10.22	Trimer
Batch 3			
1CW(monomer standard)	11199	8.83	Trimer
GCN4(dimer standard)	8131	9.43	Dimer

PSBD36(monomer standard)	4001	10.51	Monomer
15-AY_{me}A	11422	8.97	Trimer
15-EY_{me}A	11593	8.83	Trimer
15-AY_{me}K	11596	8.99	Trimer
15	11767	8.88	Trimer
13-AW_{me}A	11491	9.05	Trimer
13-EW_{me}A	11662	8.95	Trimer
13-AW_{me}K	11665	9.10	Trimer
13	11836	8.95	Trimer
25-AAA	11104	8.86	Trimer
25-EAA	11275	8.73	Trimer
25-AAK	11278	8.86	Trimer
25	11449	8.83	Trimer
22-ESA	7909	9.13	Dimer
22-ASA	7794	9.52	Dimer
22-ASK	7911	9.50	Dimer
22-AFA	7915	10.30	Dimer/Monomer
22-AFK	8031	10.23	Dimer/Monomer
22-ESK	8025	9.15	Dimer
22-EFA	8029	9.39	Dimer
22	8145	9.46	Dimer
23-ESA	7909	9.44	Dimer
23-ESK	8025	9.39	Dimer
23-EFA	8029	9.81	Dimer
23-EFK	8145	9.73	Dimer

Batch 4

1CW(trimer standard)	11199	9.01	Trimer
26	11275	8.93	Trimer
27	11449	9.09	Trimer
Batch 5			
Trimer standard	11151	9.15	Trimer
GCN4 (Dimer Standard)	8131	10.31	Dimer
18-A(p-fF)A	11385	9.33	Trimer
18-E(p-fF)A	11556	9.16	Trimer
18-A(p-fF)K	11560	9.36	Trimer
18	11731	9.18	Trimer
19-A(3,4-f₂F)A	11439	9.27	Trimer
19-E(3,4-f₂F)A	11610	9.15	Trimer
19-A(3,4-f₂F)K	11614	9.28	Trimer
19	11785	9.16	Trimer
21-A(PyrA)A	11340	9.28	Trimer
21-E(PyrA)A	11511	9.16	Trimer
21-A(PyrA)K	11515	9.25	Trimer
21	11686	9.19	Trimer
20	11809	9.22	Trimer
20-A(p-NO₂F)K	11638	9.34	Trimer
20-A(p-NO₂F)A	11463	9.28	Trimer
20-E(p-NO₂F)A	11634	9.11	Trimer

2.4.5 Folded Free energy measurement, Circular Dichroism spectropolarimetry

Measurements were made with an Aviv 420 Circular Dichroism Spectropolarimeter, using quartz cuvettes with a path length of 0.1 cm. Protein solutions were prepared in 20 mM sodium phosphate buffer, pH 7, and protein concentrations were determined spectroscopically based on tyrosine and tryptophan absorbance at 280 nm in 6 M guanidine hydrochloride + 20 mM sodium phosphate ($\epsilon_{\text{Trp}} = 5690 \text{ M}^{-1}\text{cm}^{-1}$, $\epsilon_{\text{Tyr}} = 1280 \text{ M}^{-1}\text{cm}^{-1}$).³⁸ Concentrations for variants containing p-nitrophenylalanine (p-NO₂F) were confirmed by analytical HPLC traces monitored at 220nm and integrated and compared to standard 33 amino acid coiled coils ($\epsilon_{\text{p-NO}_2\text{Phe}} = 4439 \text{ M}^{-1}\text{cm}^{-1}$). CD spectra of 30 μM solutions were obtained from 260 to 200 nm at 25°C. Variable temperature CD data were obtained at least in triplicate by monitoring the molar ellipticity at 222 nm of 30 μM solutions each protein variant (30 μM) in 20 mM sodium phosphate (pH 7) from 1 to 95°C at 2 °C intervals, with 120 s equilibration time between data points and 30 s averaging time.

Triplicate variable temperature CD data for **1-20**, and **24-27** and their individual variants were fit globally to a two-state model for thermally-induced unfolding of helix-bundle trimers as shown in equations S1—S9 (see below for sedimentation equilibrium evidence that these 1CW variants are, in fact, trimers).

In this two state model, 3 peptide monomers **M** are in equilibrium with the helix-bundle trimer **T**, where the position of equilibrium is determined by folding equilibrium constant **K**:



(S1)

In turn, **K** is defined by the equation S2:

$$K = \frac{[T]}{[M]^3} \quad (S2)$$

where $[T]$ and $[M]$ are the concentrations of helix-bundle trimer and peptide monomer, respectively. The total concentration of peptide in solution P is defined by the equation S3:

$$P = [M] + 3[T] \quad (S3)$$

By combining equations S2 and S3, we can obtain an expression for P that depends only on $[M]$ and on K_f , as shown in equation S4:

$$P = [M] + 3K[M]^3 \quad (S4)$$

In equation S4, P is a constant, K is a temperature-dependent function (that is constant at a given temperature), and $[M]$ is unknown. Rearranging equation S4 results in the following polynomial equation that is cubic in $[M]$:

$$0 = [M]^3 + \frac{[M]}{3K} - \frac{P}{3K} \quad (S5)$$

Using MATLAB, we found the three roots of this polynomial, two of which are complex, whereas the third is real. The real root of equation (S5) provides an expression for $[M]$ that depends only on P and K_f , as shown in equation (S6):

$$[M] = \left(\frac{P}{6K} + \left(\frac{1}{729K^3} + \frac{P^2}{36K^2} \right)^{\frac{1}{2}} \right)^{\frac{1}{3}} - \frac{1}{9K \left(\frac{P}{6K} + \left(\frac{1}{729K^3} + \frac{P^2}{36K^2} \right)^{\frac{1}{2}} \right)^{\frac{1}{3}}} \quad (S6)$$

As described above, K is a temperature-dependent function that is constant at a given temperature. K is related to the temperature-dependent folding free energy ΔG_f according to equation S7:

$$K = \exp\left(\frac{-\Delta G_f}{RT}\right) \quad (S7)$$

where R is the universal gas constant (0.0019872 kcal/mol/K). In turn, the temperature dependence of ΔG_f can be defined by the following first order polynomial:

$$\Delta G_f = \Delta G_0 + \Delta G_1(T - T_0) \quad (S8)$$

where T is temperature in Kelvin; ΔG_0 and ΔG_1 are parameters to be determined via least-squares regression; and T_0 is an arbitrary reference temperature, ideally chosen to be near the midpoint of the unfolding transition. By combining equations S6–S8, we now have an expression for $[M]$ as a function of temperature that depends only on ΔG_0 , ΔG_1 , P and reference temperature T_0 .

We can use this expression for $[M]$ to fit the variable temperature CD data $[\theta]$ to equation S9, using the actual protein concentration in solution for P ; using a fixed arbitrary value for T_0 (343.15 K); and then varying ΔG_0 and ΔG_1 as parameters of the fit so as to minimize the sum of the squared residuals.

$$[\theta] = (u_0) \left(1 - \frac{3K[M]^3}{P}\right) + (f_0 + f_1 T) \left(\frac{3K[M]^3}{P}\right) \quad (S9)$$

In equation S9, T is the temperature in Kelvin; u_0 defines a horizontal post-transition baseline; and f_0 and f_1 are the intercept and slope of the pre-transition baseline, respectively. We fit the variable trimer variable temperature CD data using distinct ΔG_0 and ΔG_1 values for each peptide; distinct u_0 and f_0 values for each replicate data set of each peptide and a global value for f_1 across all replicate data sets and peptides.

Triplicate variable temperature CD data for each **22-23** and their individual variants were fit globally to a two-state model for thermally-induced unfolding of helix-bundle dimers as shown

in equations S10—S19 (see below for sedimentation equilibrium evidence that these GCN4-p1 variants are dimers).

In this two state model, 2 peptide monomers **M** are in equilibrium with the helix-bundle dimer **D**, where the position of equilibrium is determined by folding equilibrium constant **K**:



(S10)

In turn, **K** is defined by the equation S11:

$$K = \frac{[D]}{[M]^2} \quad (S11)$$

where **[T]** and **[M]** are the concentrations of helix-bundle trimer and peptide monomer, respectively. The total concentration of peptide in solution **P** is defined by the equation S3:

$$P = [M] + 2[D] \quad (S12)$$

By combining equations S12 and S11, we can obtain an expression for **P** that depends only on **[M]** and on **K_f**, as shown in equation S4:

$$P = [M] + 2K[M]^2 \quad (S13)$$

In equation S13, **P** is a constant, **K** is a temperature-dependent function (that is constant at a given temperature), and **[M]** is unknown. Rearranging equation S13 results in the following quadratic:

$$0 = [M]^2 + [M] - P \quad (S14)$$

The solution the the quadratic (S14) provides an expression for $[M]$ that depends only on P and K , as shown in equation (S15):

$$[M] = \frac{\sqrt{1+8KP}+1}{4K} \quad (S15)$$

As described above, K is a temperature-dependent function that is constant at a given temperature. In turn, the temperature dependence of ΔG_f for the dimer can be defined by the following second order polynomial:

$$\Delta G_f = \Delta G_0 + \Delta G_1(T - T_0) + \Delta G_2(T - T_0)^2 \quad (S17)$$

where T is temperature in Kelvin; ΔG_0 , ΔG_1 and ΔG_2 are parameters to be determined via least-squares regression; and T_0 is an arbitrary reference temperature, also chosen near the midpoint of the unfolding transition. By combining equations, we now have an expression for $[M]$ as a function of temperature that depends only on ΔG_0 , ΔG_1 , P and reference temperature T_0 .

We can use this expression for $[M]$ to fit the variable temperature CD data $[\theta]$ to equation S9, using the actual protein concentration in solution for P ; using a fixed arbitrary value for T_0 (343.15 K); and then varying ΔG_0 and ΔG_1 as parameters of the fit so as to minimize the sum of the squared residuals.

$$[\theta] = (u_0)(1 - Ffit) + (f_0 + f_1 T)(Ffit) \quad (S18)$$

In equation S18, T is the temperature in Kelvin; u_0 defines a horizontal post-transition baseline; and f_0 and f_1 are the intercept and slope of the pre-transition baseline, respectively. F_{fit} is the fraction folded as defined by equation (S19) of the protein at temperature T .

$$F_{fit} = 1 + \left(\frac{1}{4KP}\right) - \left(\frac{1}{2KP} + \frac{1}{16K^2P^2}\right)^{\frac{1}{2}} \quad (S19)$$

We fit the dimer variable temperature CD data using distinct ΔG_0 , ΔG_1 and ΔG_2 values for each peptide; distinct u_0 and f_0 values for each replicate data set of each peptide and a global value for f_1 across all replicate data sets and peptides.

CD fit data for compounds **22-23** and their control variants are shown below.

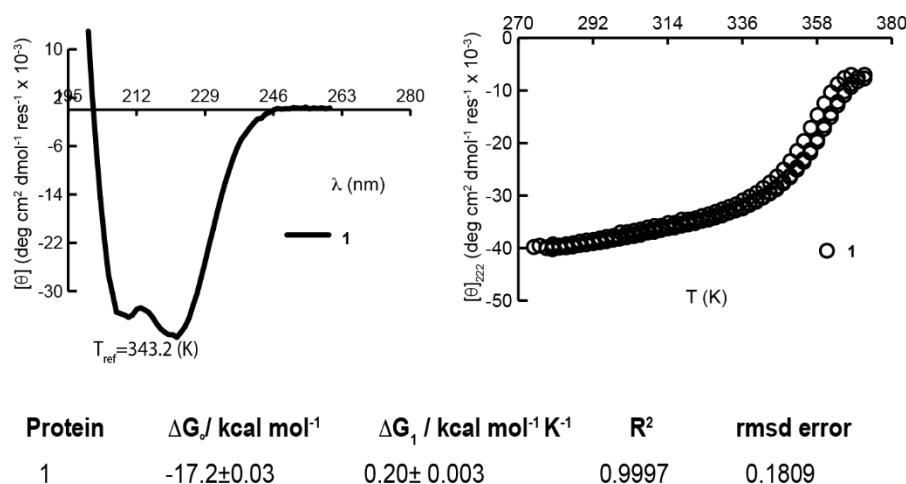
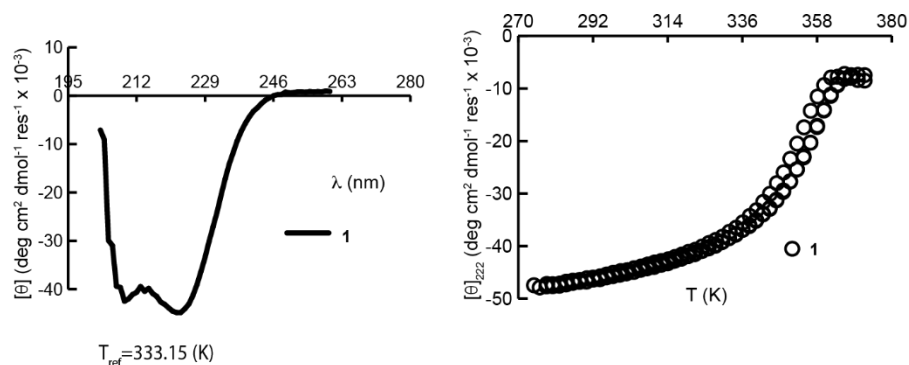
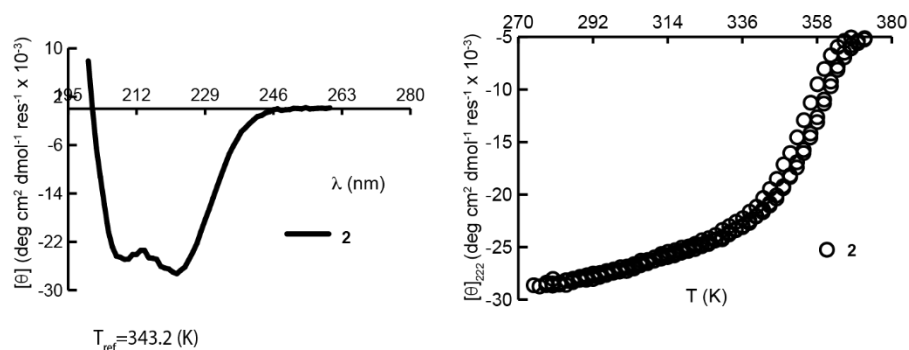


Figure 2-184. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **1** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



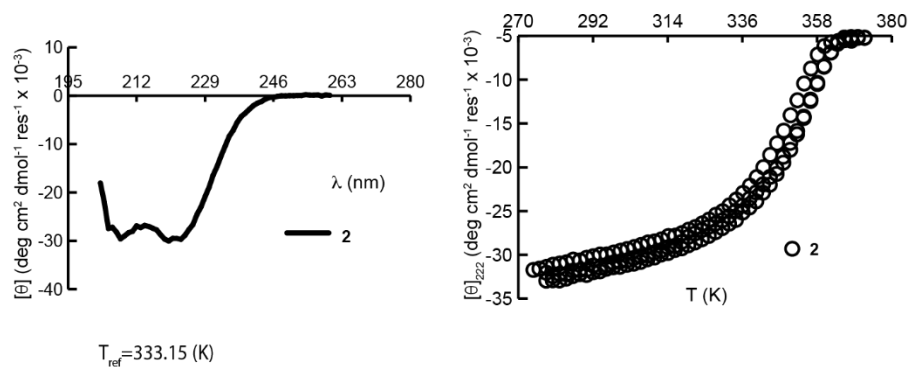
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
1	-16.2 ± 0.05	0.20 ± 0.004	0.9991	0.3976

Figure 2-185. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **1** in 20 mM sodium phosphate, pH 7, **1M urea**. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



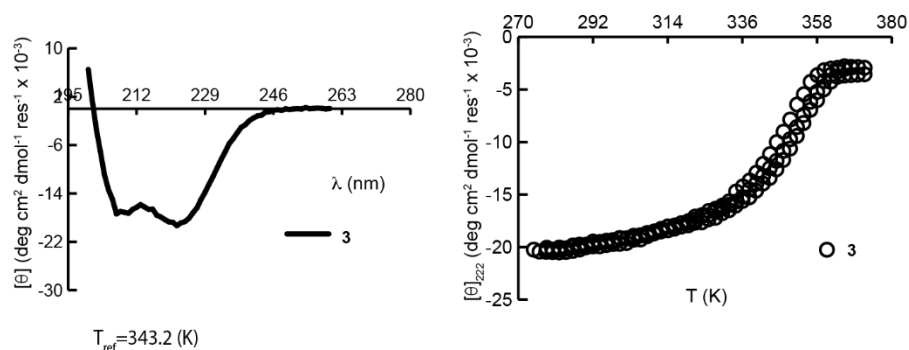
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
2	-16.8 ± 0.03	0.20 ± 0.003	0.9997	0.1351

Figure 2-186. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **2** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



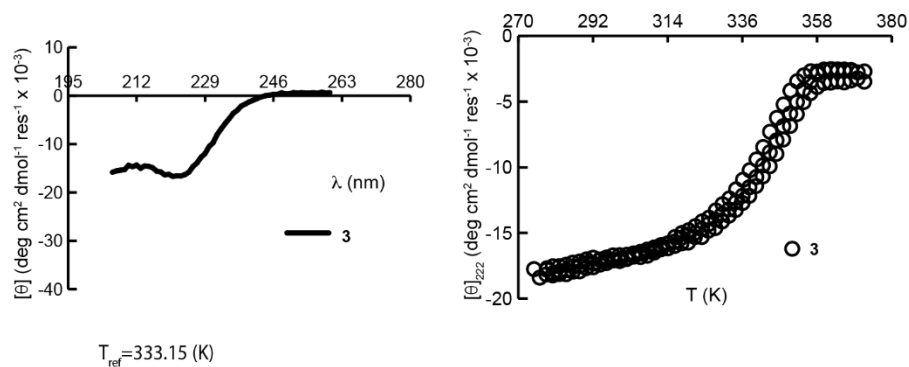
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
2	-15.7 ± 0.04	0.18 ± 0.004	0.9993	0.2509

Figure 2-187. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **2** in 20 mM sodium phosphate, pH 7, **1M urea**. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated



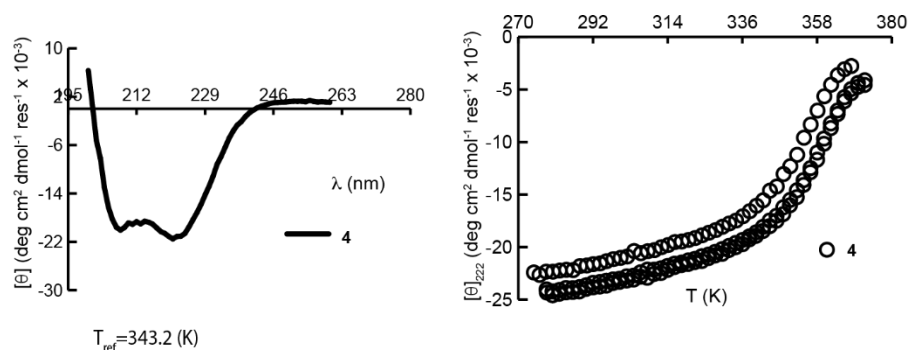
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
3	-15.2 ± 0.03	0.18 ± 0.003	0.9995	0.1447

Figure 2-188. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **3** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



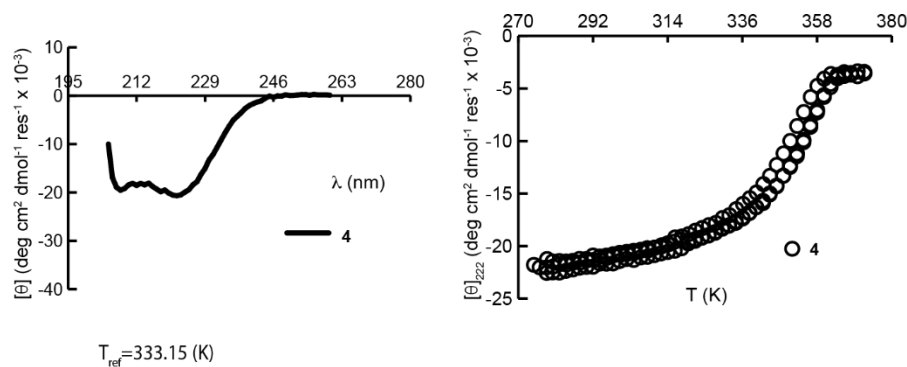
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
3	-14.0 ± 0.03	0.17 ± 0.004	0.9992	0.1631

Figure 2-189. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **3** in 20 mM sodium phosphate, pH 7, 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



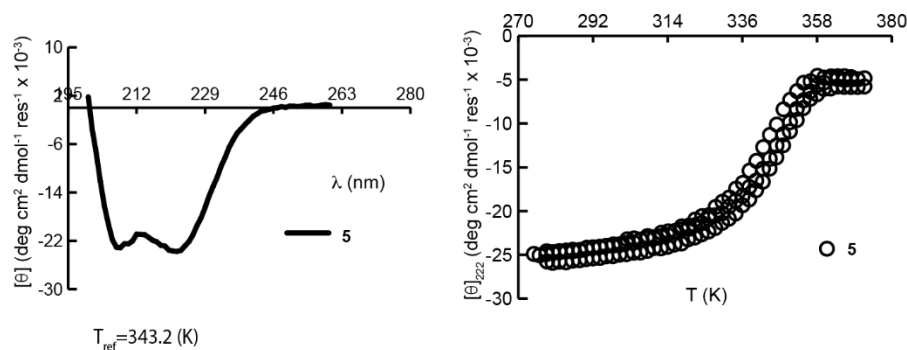
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
4	-17.0 ± 0.04	0.19 ± 0.003	0.9996	0.1297

Figure 2-190. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **4** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
4	-15.8 ± 0.05	0.19 ± 0.005	0.9989	0.2123

Figure 2-191. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **4** in 20 mM sodium phosphate, pH 7, **1M urea**. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
5	-14.1 ± 0.02	0.18 ± 0.003	0.9996	0.1436

Figure 2-192. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **5** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

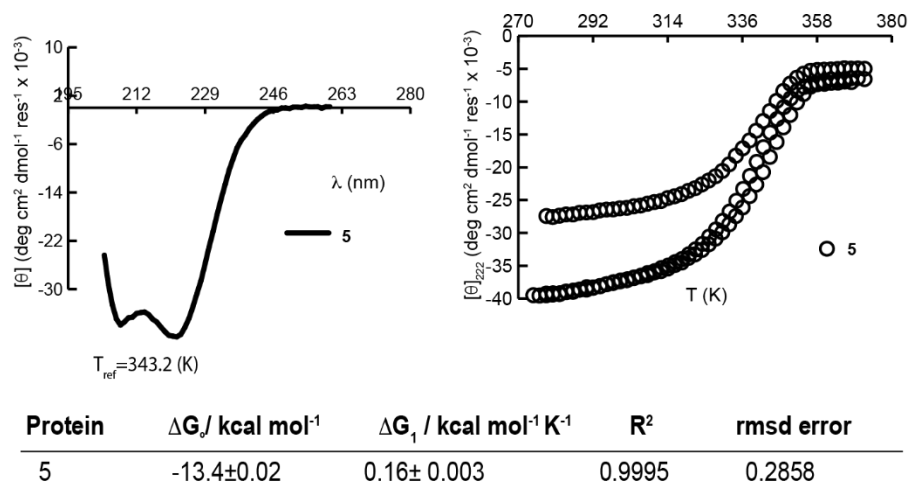


Figure 2-193. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **5** in 20 mM sodium phosphate, pH 7, **1M urea**. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

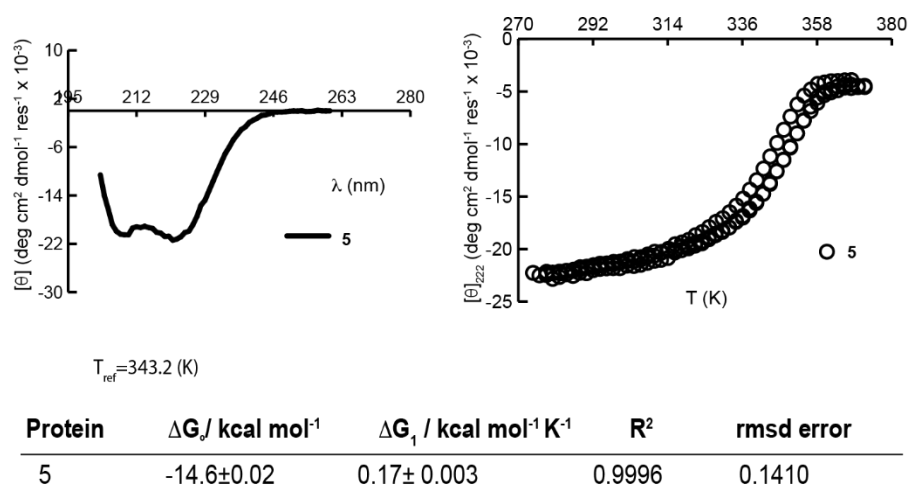


Figure 2-194. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **5** in 20 mM sodium phosphate, pH 7, **0.25M NaCl**. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

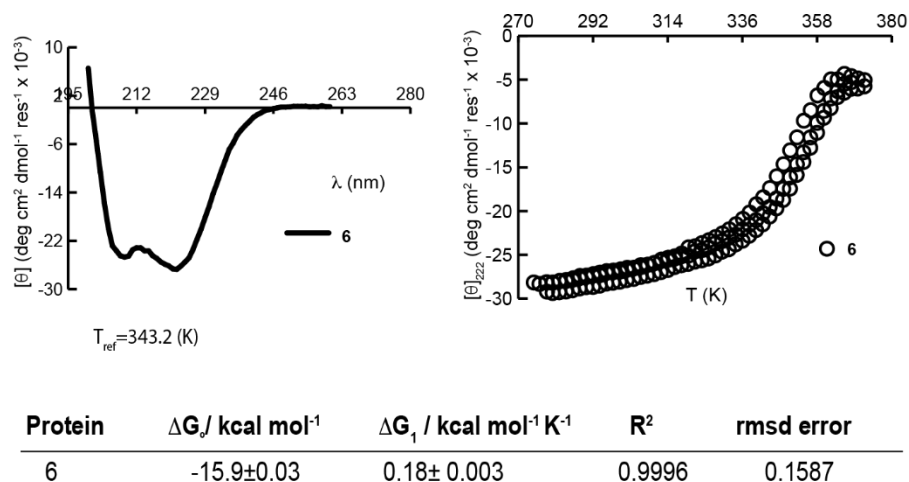


Figure 2-195. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **6** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

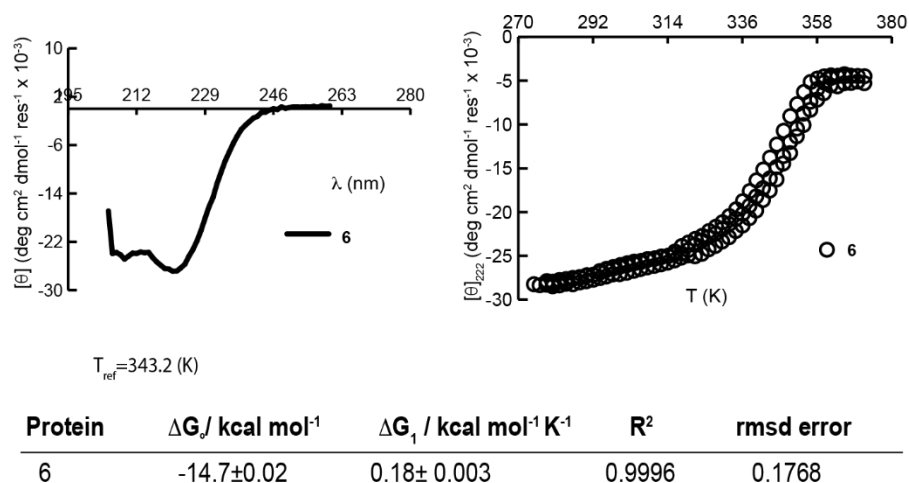


Figure 2-196. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **6** in 20 mM sodium phosphate, pH 7, **1M urea**. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

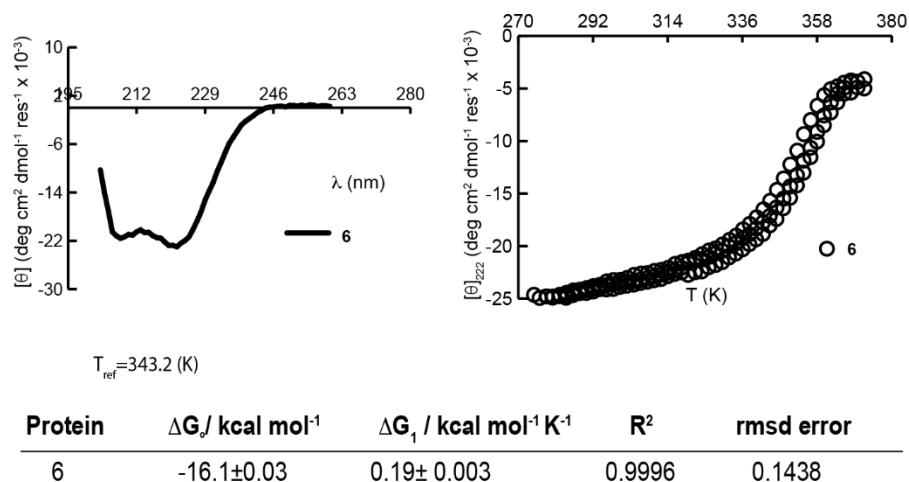


Figure 2-197. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **6** in 20 mM sodium phosphate, pH 7, **0.25M NaCl**. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

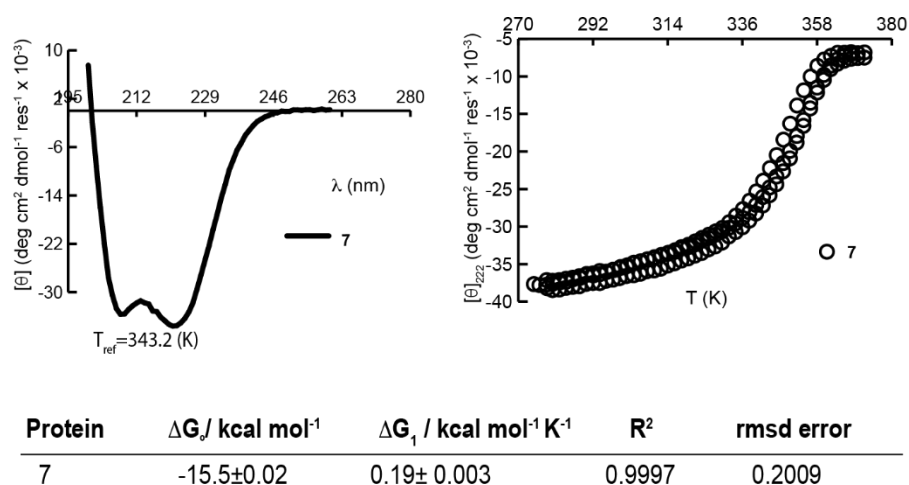


Figure 2-198. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **7** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

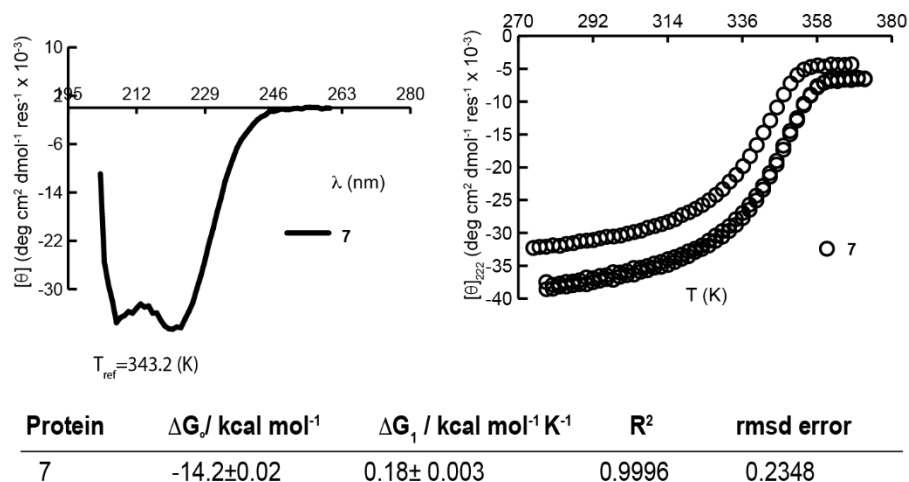


Figure 2-199. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins 7 in 20 mM sodium phosphate, pH 7, 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

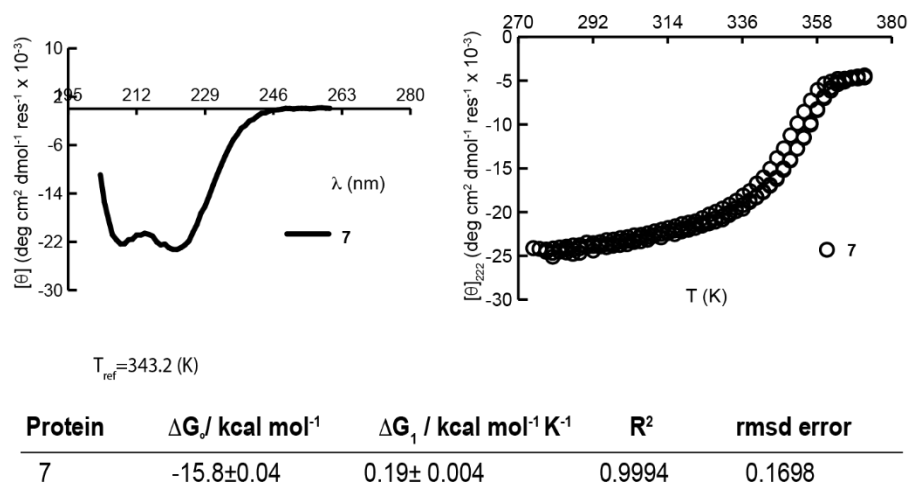
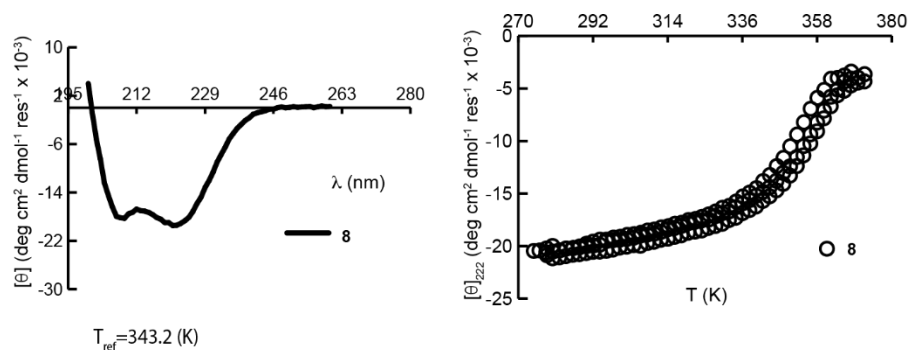
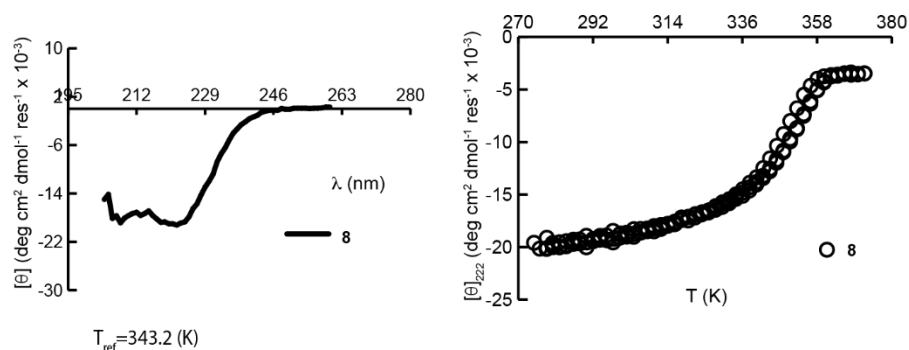


Figure 2-200. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins 7 in 20 mM sodium phosphate, pH 7, 0.25M NaCl. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



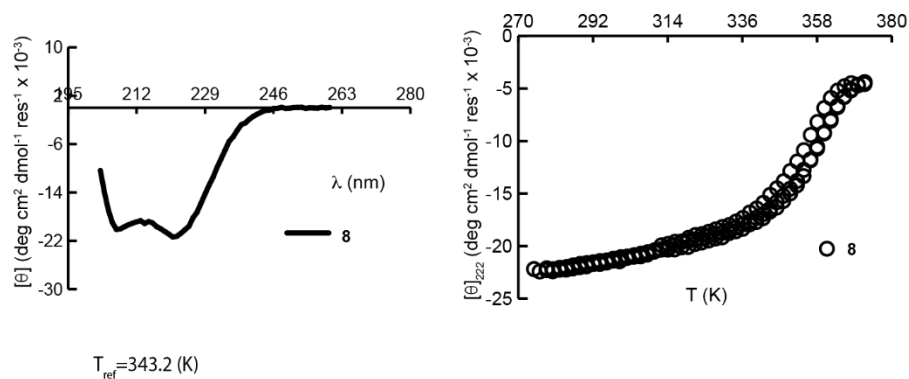
Protein	ΔG_f / kcal mol ⁻¹	ΔG_f / kcal mol ⁻¹ K ⁻¹	R ²	rmsd error
8	-16.4±0.04	0.19± 0.004	0.9994	0.1392

Figure 2-201. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μ M solutions of protein **8** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



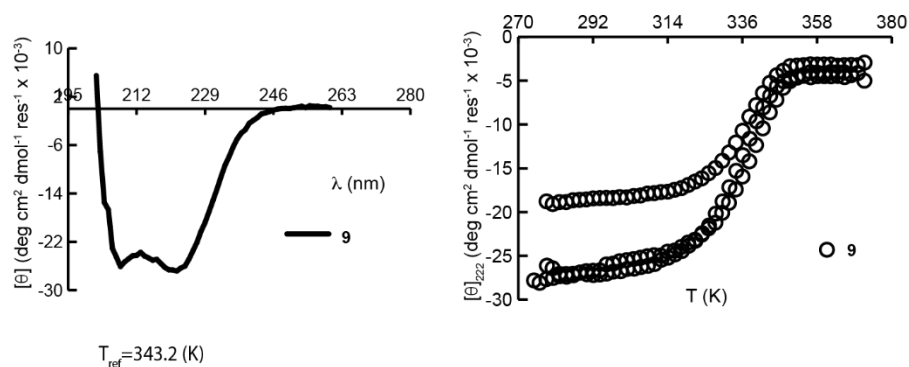
Protein	ΔG_f / kcal mol ⁻¹	ΔG_f / kcal mol ⁻¹ K ⁻¹	R ²	rmsd error
8	-15.3±0.04	0.20± 0.005	0.9989	0.2029

Figure 2-202. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μ M solutions of proteins **8** in 20 mM sodium phosphate, pH 7, **1M urea**. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



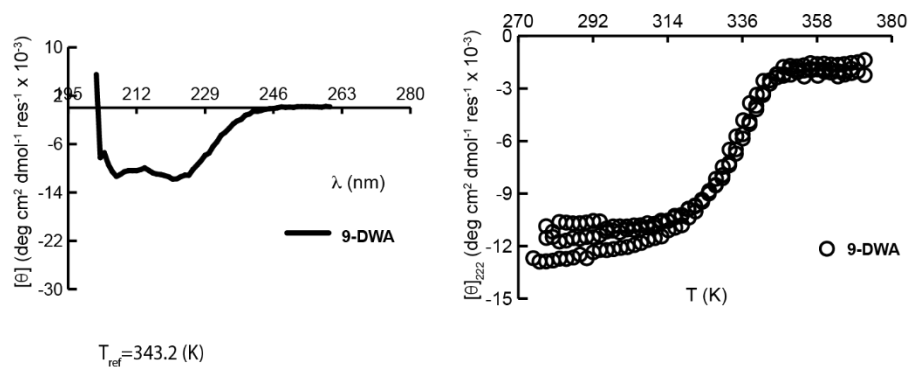
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
8	-17.1 ± 0.05	0.20 ± 0.004	0.9993	0.1430

Figure 2-203. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **8** in 20 mM sodium phosphate, pH 7, **0.25M NaCl**. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



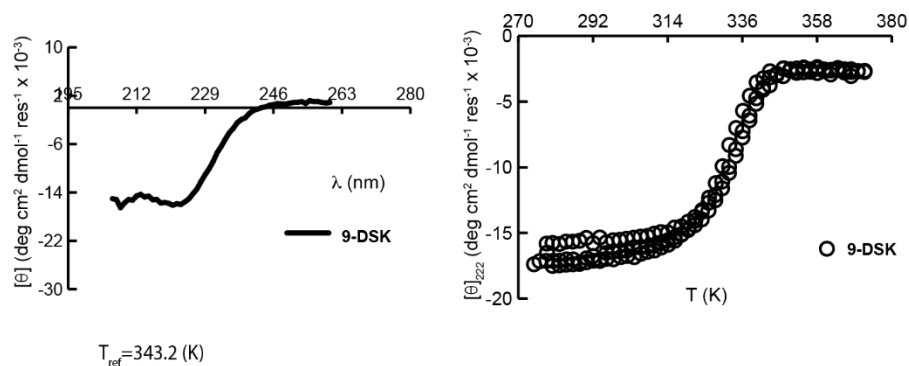
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
9	-12.2 ± 0.03	0.20 ± 0.005	0.9992	0.2675

Figure 2-204. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **9** in 20 mM sodium phosphate, pH 7, 1M Urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



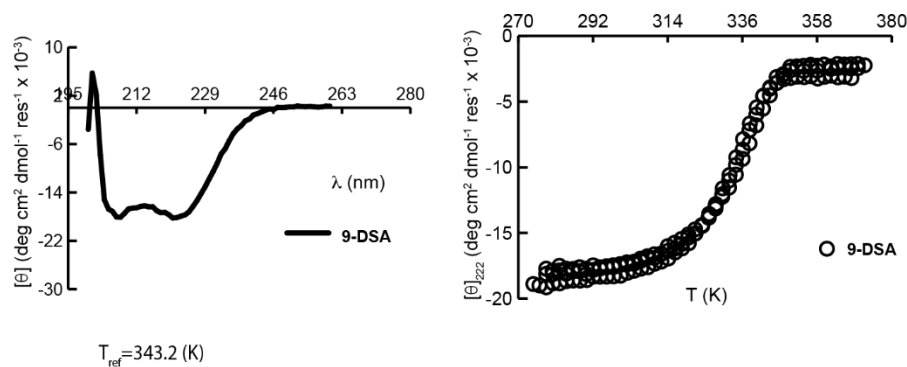
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
9-DWA	-11.5 ± 0.04	0.20 ± 0.004	0.9993	0.1151

Figure 2-205. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **9-DWA** in 20 mM sodium phosphate, pH 7, 1M Urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



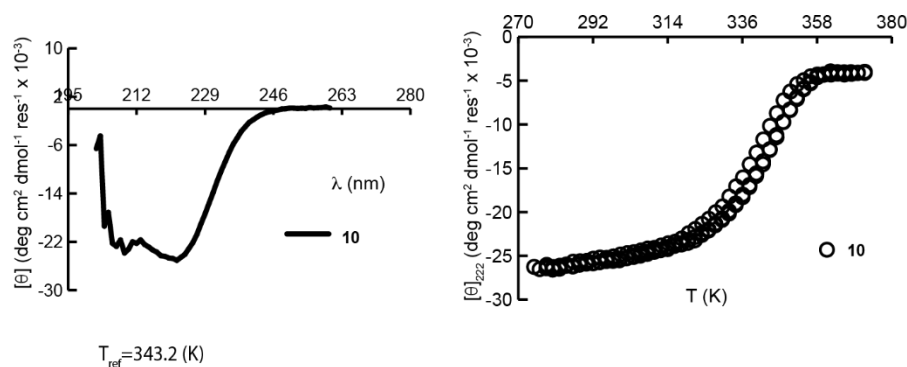
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
9-DSK	-11.0 ± 0.04	0.23 ± 0.004	0.9994	0.1584

Figure 2-206. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **9-DSK** in 20 mM sodium phosphate, pH 7, 1M Urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



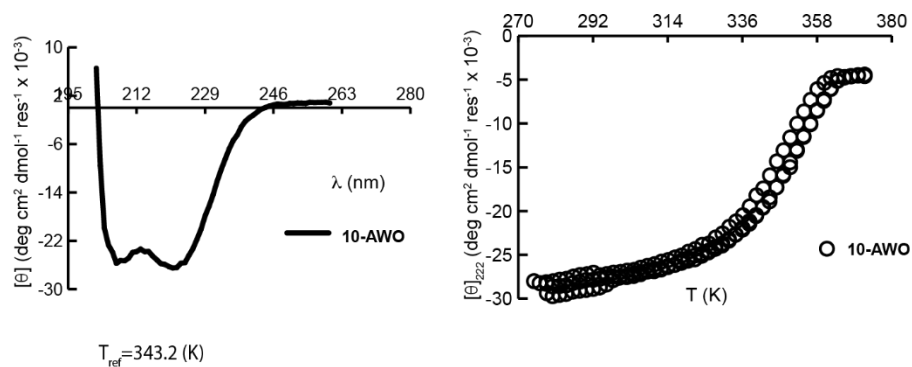
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
9-DSA	-12.0 ± 0.02	0.19 ± 0.003	0.9996	0.1343

Figure 2-207. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **9-DSA** in 20 mM sodium phosphate, pH 7, 1M Urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



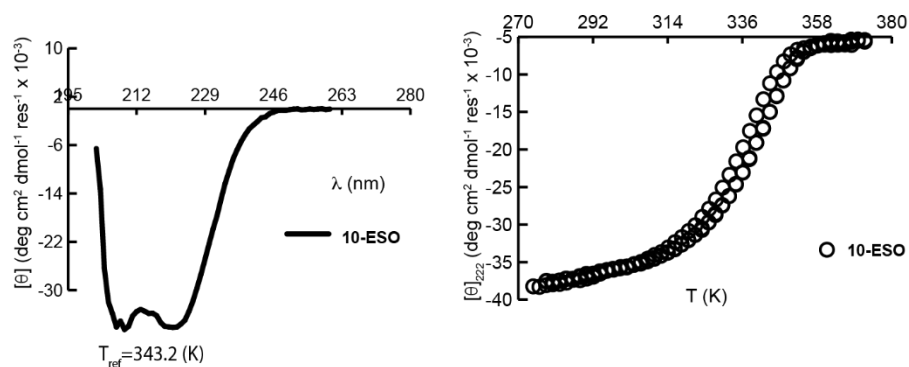
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
10	-13.6 ± 0.01	0.17 ± 0.002	0.9998	0.1146

Figure 2-208. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **10** in 20 mM sodium phosphate, pH 7, 1M Urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



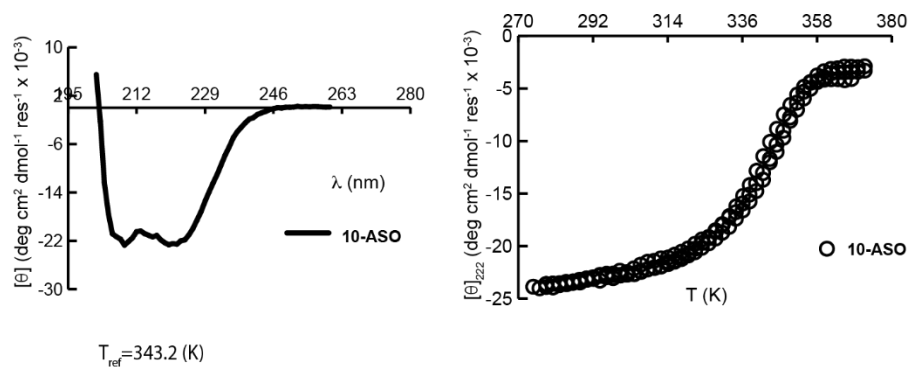
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
10-AWO	-15.1 ± 0.03	0.16 ± 0.003	0.9996	0.1771

Figure 2-209. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **10-AWO** in 20 mM sodium phosphate, pH 7, 1M Urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



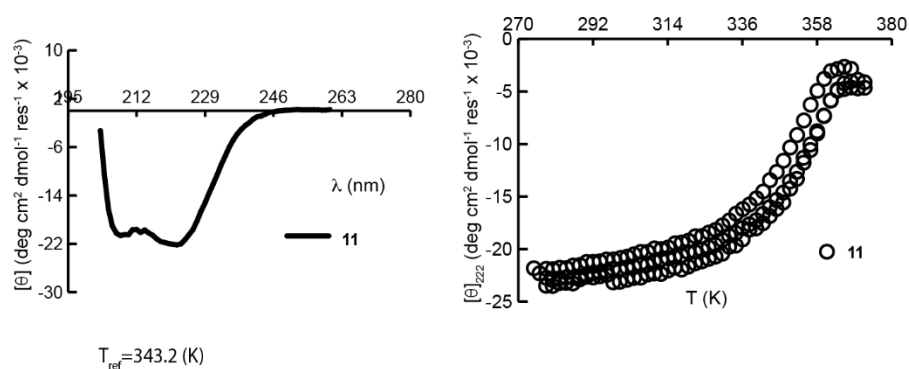
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
10-ESO	-13.0 ± 0.01	0.16 ± 0.001	0.9999	0.1294

Figure 2-210. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **10-ESO** in 20 mM sodium phosphate, pH 7, 1M Urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



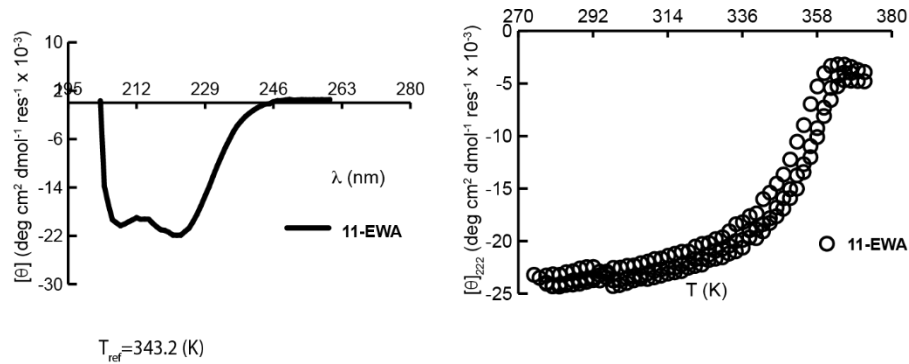
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
10-ASO	-13.9 ± 0.01	0.16 ± 0.002	0.9998	0.1239

Figure 2-211. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **10-ASO** in 20 mM sodium phosphate, pH 7, 1M Urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



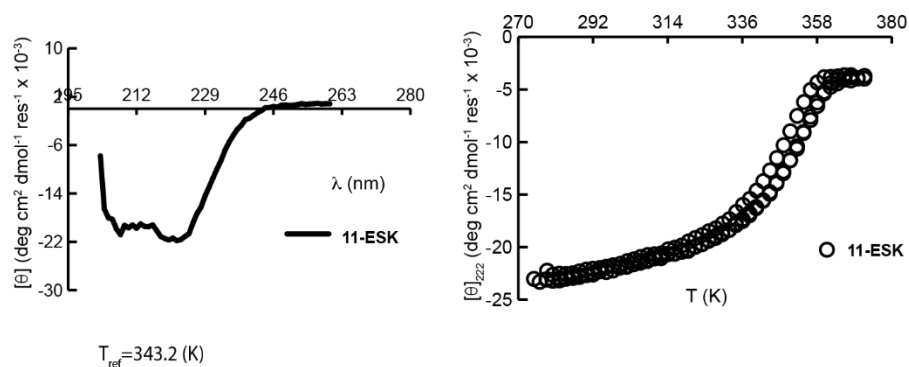
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
11	-16.0 ± 0.06	0.18 ± 0.005	0.9984	0.2666

Figure 2-212. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **11** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



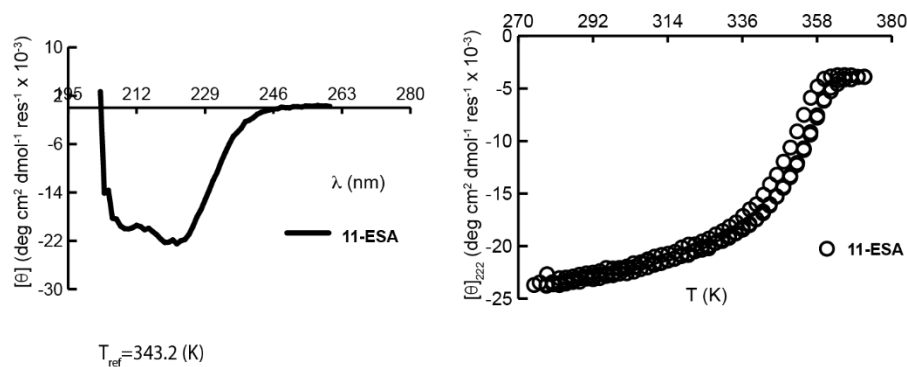
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
11-EWA	-16.3 ± 0.07	0.21 ± 0.007	0.9976	0.3397

Figure 2-213. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **11-EWA** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



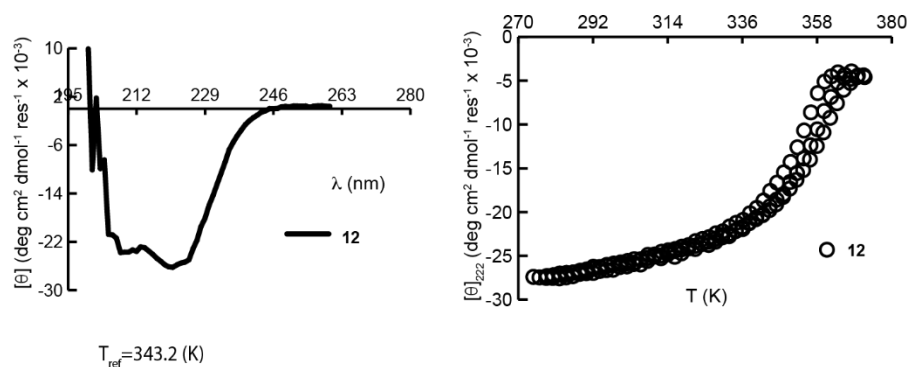
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
11-ESK	-15.3 ± 0.04	0.19 ± 0.004	0.9990	0.2122

Figure 2-214. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **11-ESK** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



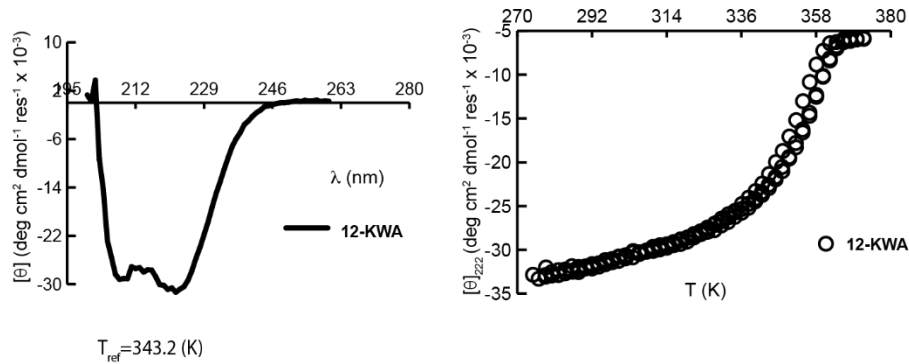
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
11-ESA	-16.0 ± 0.05	0.21 ± 0.005	0.9990	0.2122

Figure 2-215. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **11-ESA** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



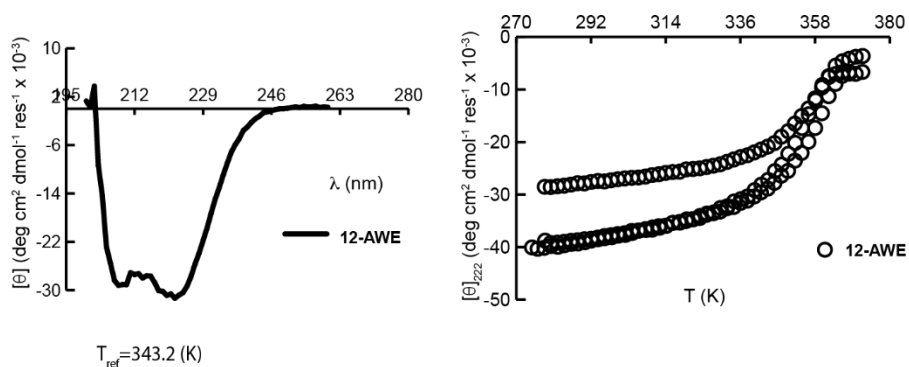
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
12	-16.7 ± 0.07	0.21 ± 0.006	0.9983	0.3091

Figure 2-216. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **12** in 20 mM sodium phosphate, pH 7, 1M Urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



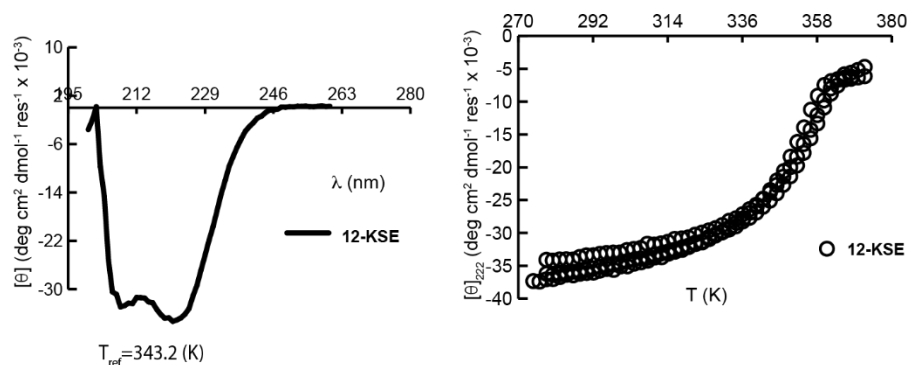
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
12-KWA	-16.5 ± 0.05	0.21 ± 0.005	0.9991	0.2788

Figure 2-217. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **12-KWA** in 20 mM sodium phosphate, pH 7, 1M Urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



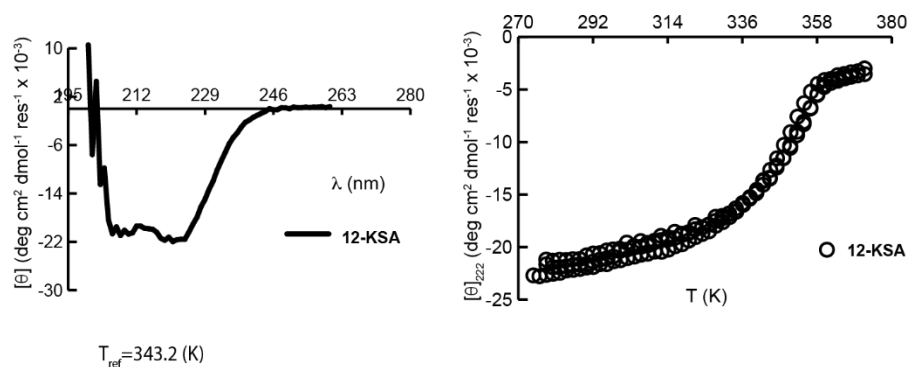
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
12-AWE	-17.1 ± 0.06	0.23 ± 0.005	0.9991	0.3306

Figure 2-218. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **12-AWE** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



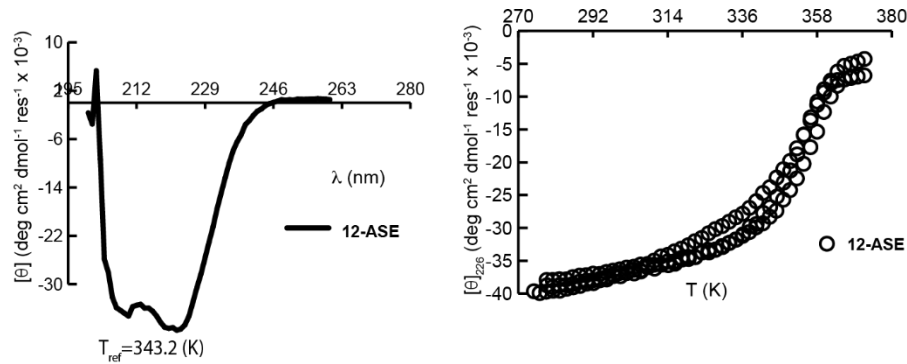
Protein	ΔG_f / kcal mol ⁻¹	ΔG_f / kcal mol ⁻¹ K ⁻¹	R ²	rmsd error
12-KSE	-16.2±0.04	0.20± 0.004	0.9992	0.2882

Figure 2-219. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μ M solutions of proteins **12-KSE** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



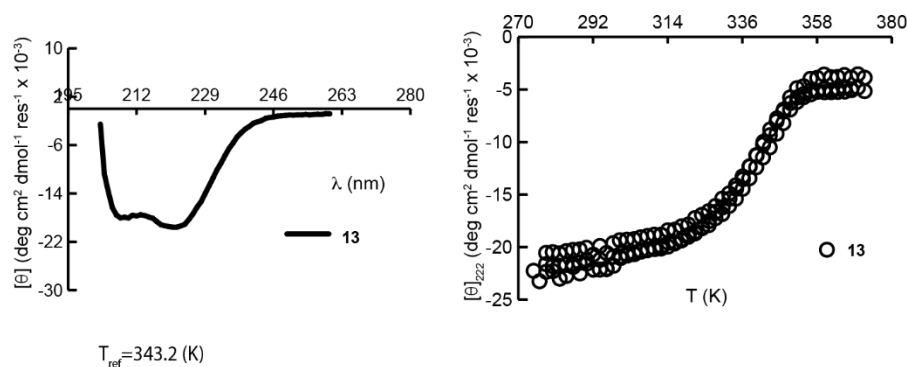
Protein	ΔG_f / kcal mol ⁻¹	ΔG_f / kcal mol ⁻¹ K ⁻¹	R ²	rmsd error
12-KSA	-15.3±0.04	0.19± 0.004	0.9992	0.1773

Figure 2-220. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μ M solutions of proteins **12-KSA** in 20 mM sodium phosphate, pH 7, 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



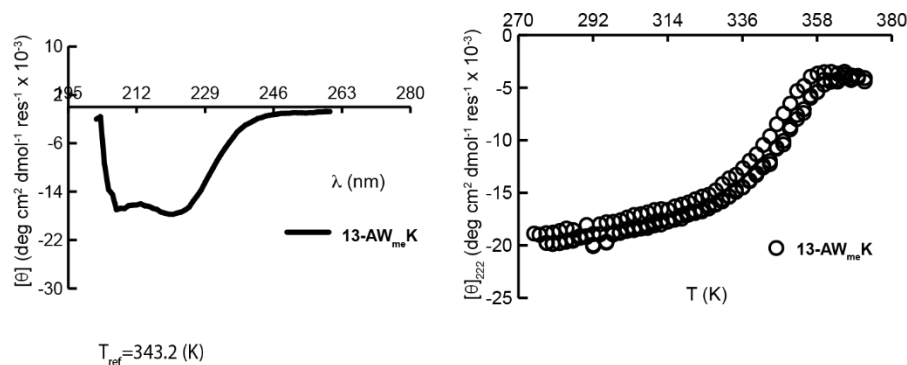
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
12-ASE	-16.4 ± 0.07	0.21 ± 0.006	0.9984	0.4455

Figure 2-221. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **12-ASE** in 20 mM sodium phosphate, pH 7, 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



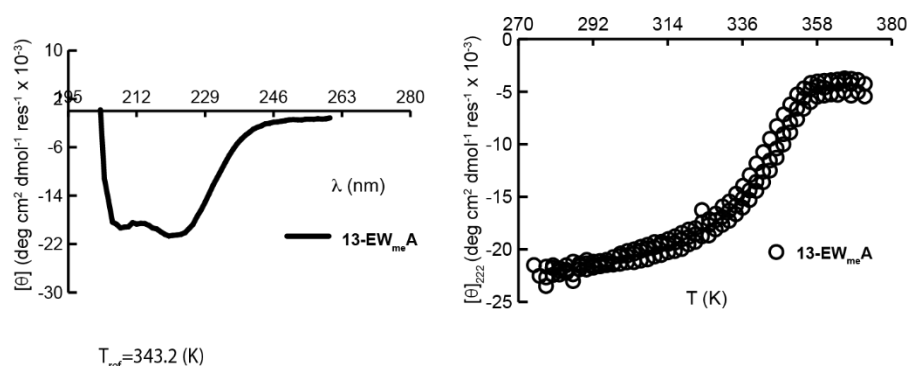
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
13	-13.3 ± 0.04	0.17 ± 0.005	0.9986	0.2572

Figure 2-222. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **13** in 20 mM sodium phosphate, pH 7, 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



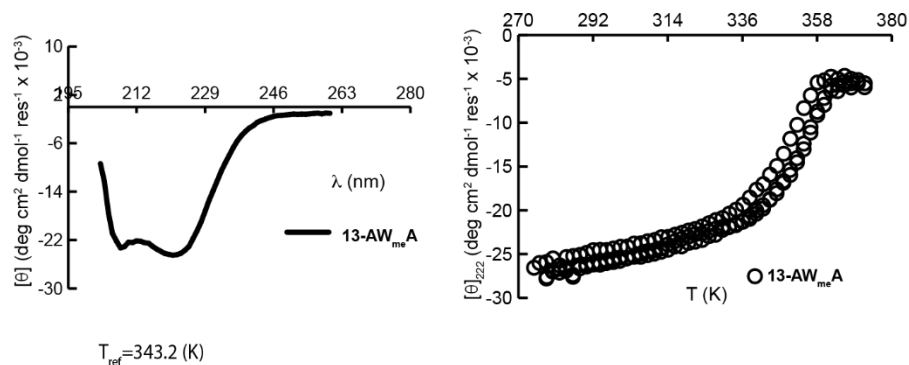
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
13-AW _{me} K	-14.7 ± 0.05	0.18 ± 0.006	0.9984	0.2336

Figure 2-223. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **13-AW_{me}K** in 20 mM sodium phosphate, pH 7, 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



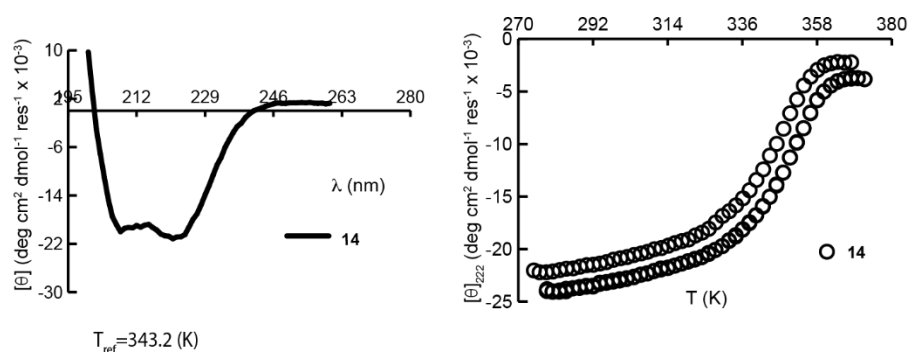
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
13-EW _{me} A	-14.1 ± 0.04	0.19 ± 0.007	0.9982	0.2930

Figure 2-224. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **13-EW_{me}A** in 20 mM sodium phosphate, pH 7, 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



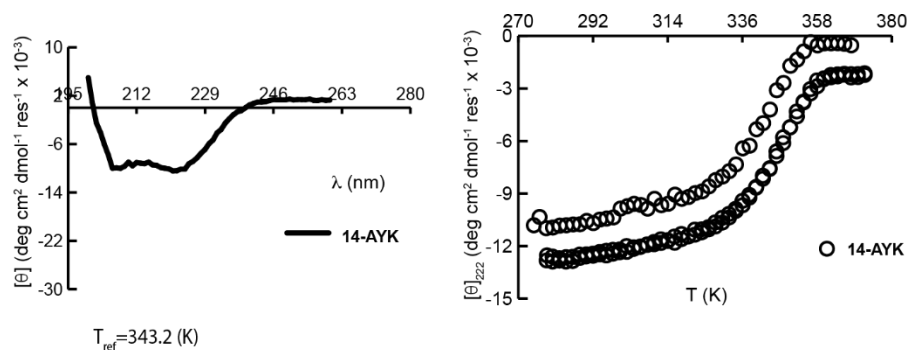
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
13-AW _{me} A	-16.3±0.06	0.25± 0.006	0.9989	0.2523

Figure 2-225. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **13-AW_{me}A** in 20 mM sodium phosphate, pH 7, 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



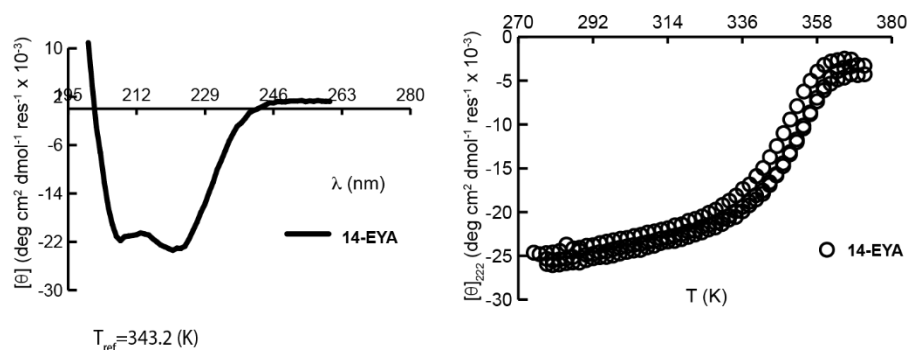
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
14	-15.0±0.02	0.18± 0.002	0.9997	0.1206

Figure 2-226. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **14** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



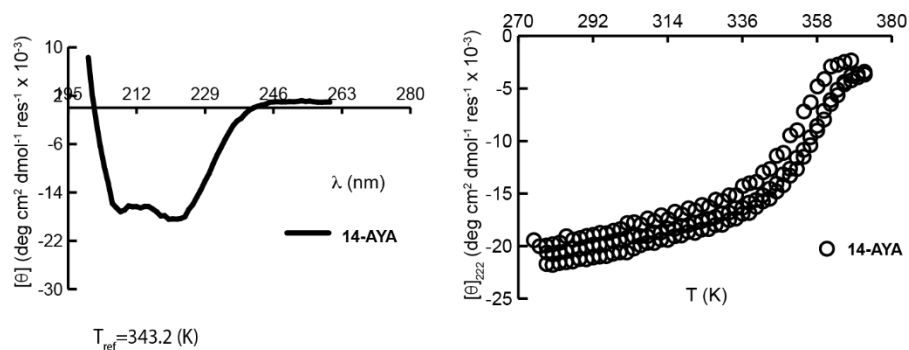
Protein	ΔG_f / kcal mol ⁻¹	ΔG_f / kcal mol ⁻¹ K ⁻¹	R ²	rmsd error
14-AYK	-14.3±0.03	0.19± 0.005	0.9990	0.1395

Figure 2-227. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μ M solutions of proteins **14-AYK** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



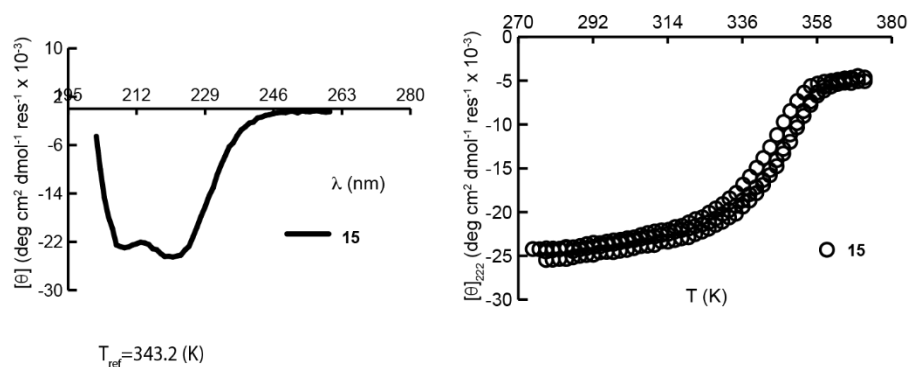
Protein	ΔG_f / kcal mol ⁻¹	ΔG_f / kcal mol ⁻¹ K ⁻¹	R ²	rmsd error
14-EYA	-15.5±0.03	0.18± 0.003	0.9996	0.1537

Figure 2-228. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μ M solutions of proteins **14-EYA** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



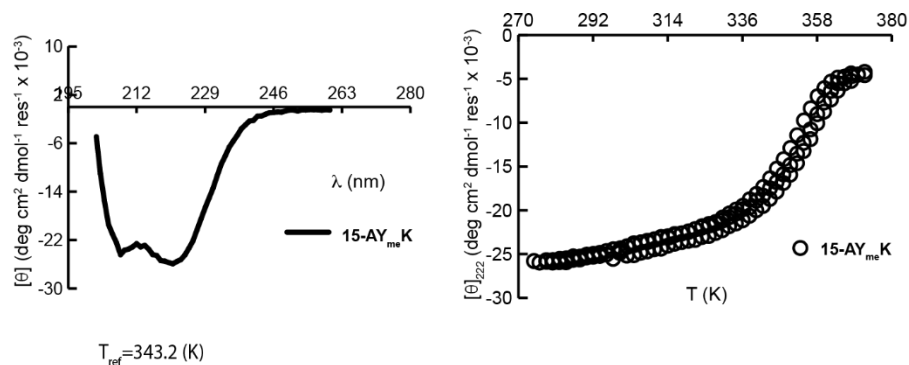
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
14-AYA	-16.8 ± 0.06	0.21 ± 0.005	0.9990	0.1692

Figure 2-229. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **14-AYA** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



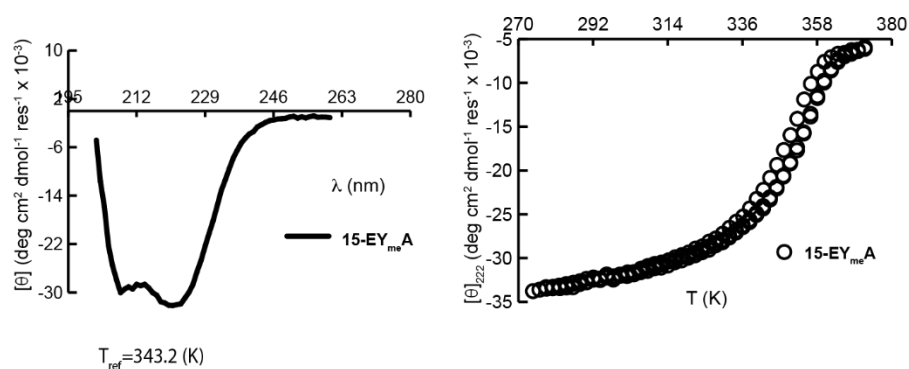
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
15	-14.6 ± 0.02	0.17 ± 0.002	0.9998	0.1098

Figure 2-230. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **15** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



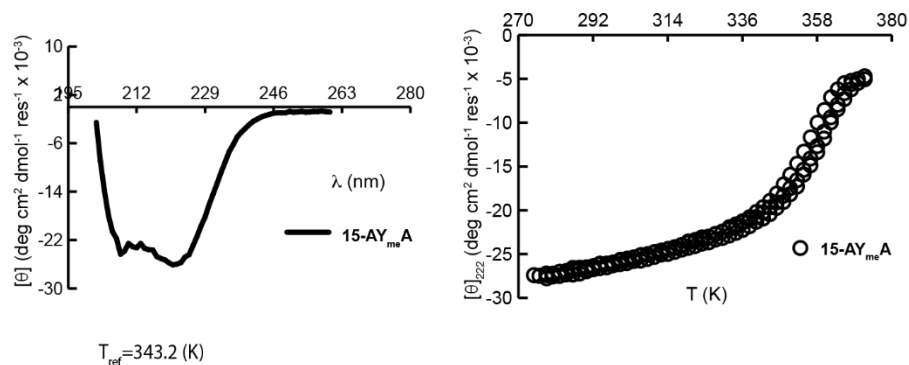
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
15-AY _{me} K	-16.1 ± 0.04	0.19 ± 0.003	0.9994	0.1720

Figure 2-231. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **15-AY_{me}K** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



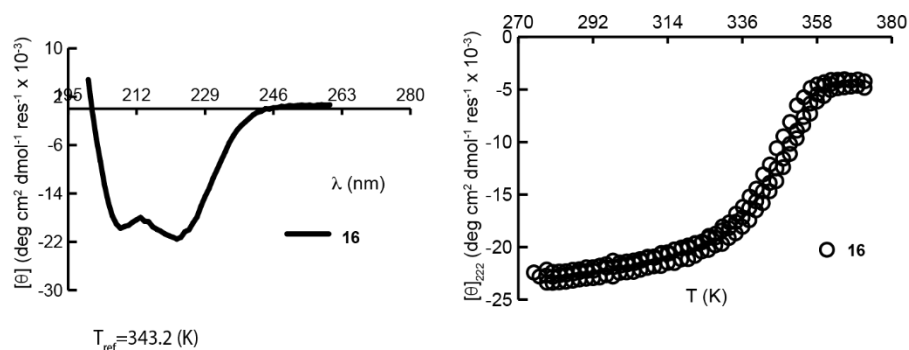
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
15-EY _{me} A	-15.7 ± 0.03	0.18 ± 0.003	0.9996	0.1932

Figure 2-232. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **15-EY_{me}A** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
15-AY _{me} A	-17.2±0.04	0.20± 0.004	0.9995	0.1545

Figure 2-233. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **15-AY_{me}A** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
16	-14.7±0.02	0.18± 0.003	0.9996	0.1417

Figure 2-234. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **16** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

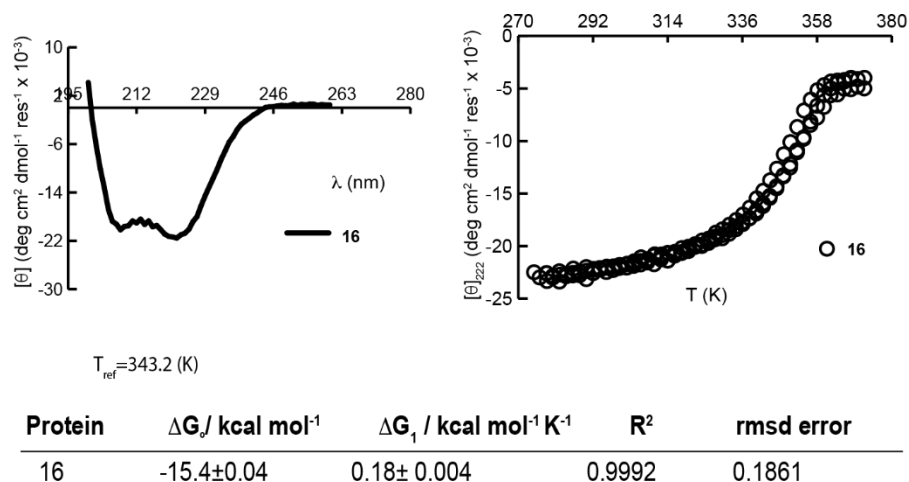


Figure 2-235. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **16** in 20 mM sodium phosphate, pH 7, 0.25M NaCl. Parameters used to fit the variable.

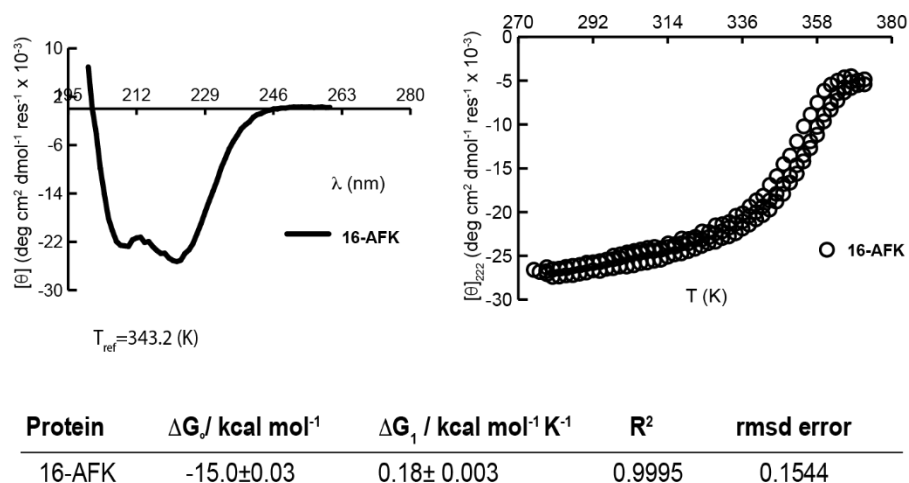
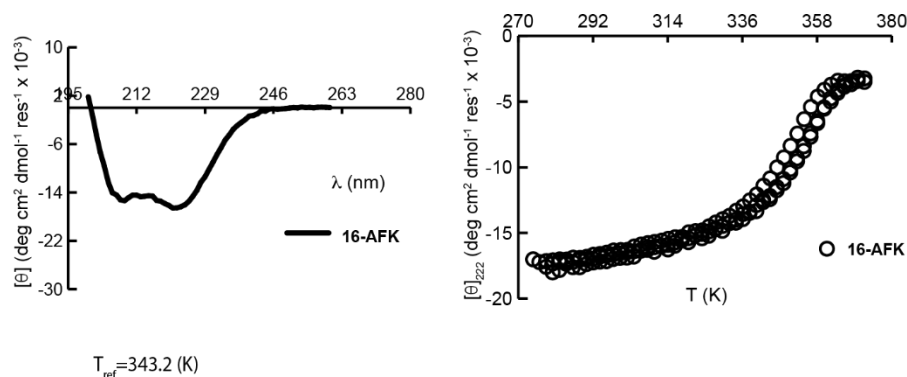
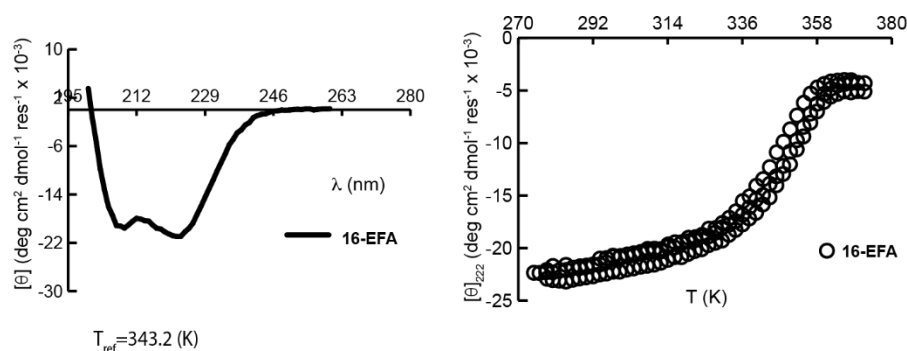


Figure 2-236. ESI TOF spectrum for peptide **3** CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **16-AFK** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
16-AFK	-15.9 ± 0.05	0.18 ± 0.004	0.9990	0.1496

Figure 2-237. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **16-AFK** in 20 mM sodium phosphate, pH 7, **0.25M NaCl**. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
16-EFA	-15.0 ± 0.03	0.18 ± 0.003	0.9995	0.1544

Figure 2-238. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **16-EFA** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

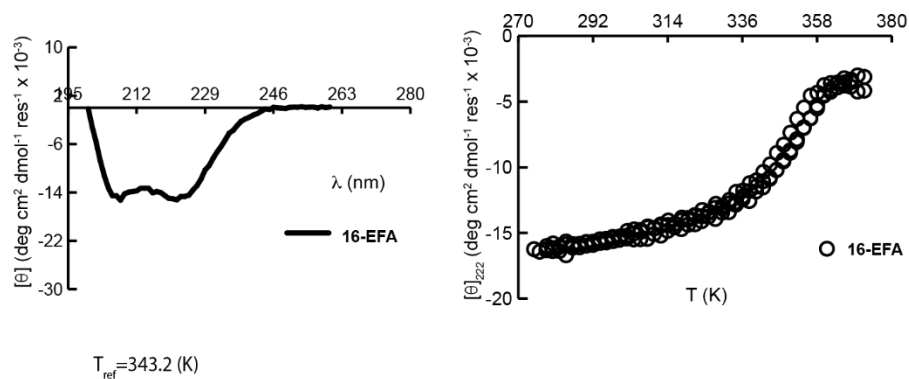


Figure 2-239. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **16-EFA** in 20 mM sodium phosphate, pH 7, **0.25M NaCl**. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

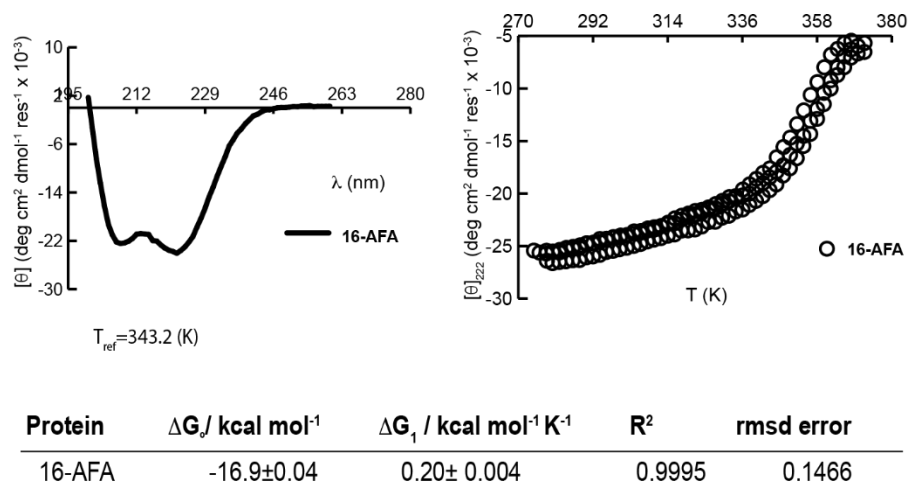


Figure 2-240. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **16-AFA** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

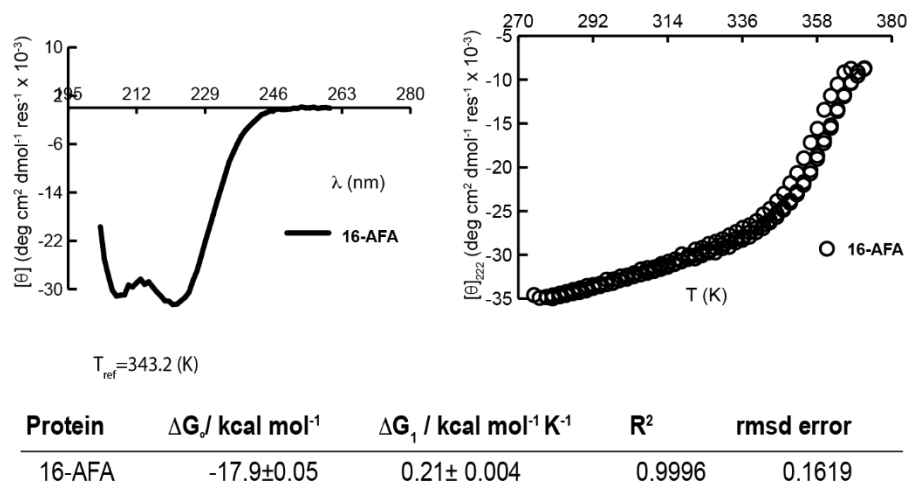


Figure 2-241. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **16-AFA** in 20 mM sodium phosphate, pH 7, **0.25M NaCl**. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

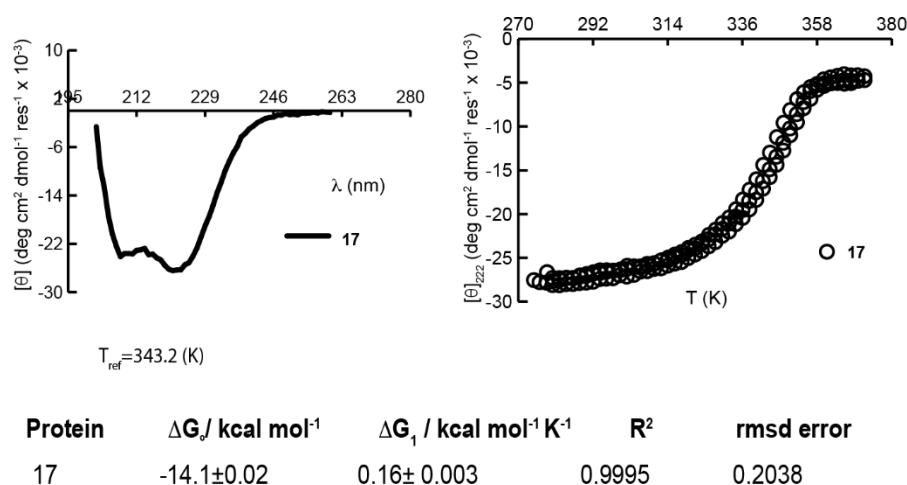
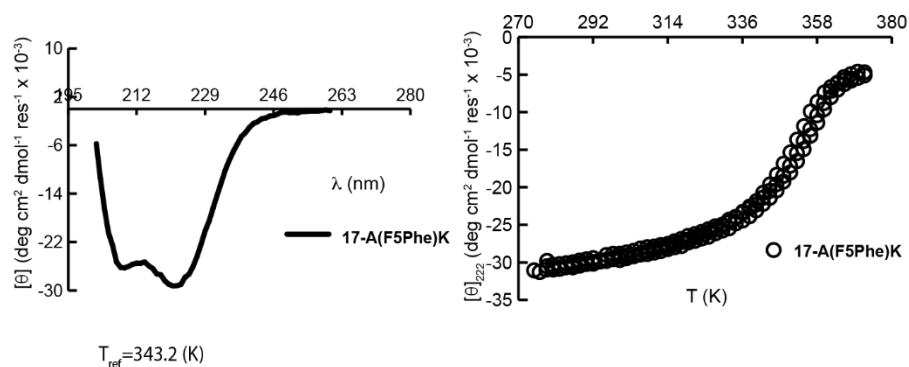
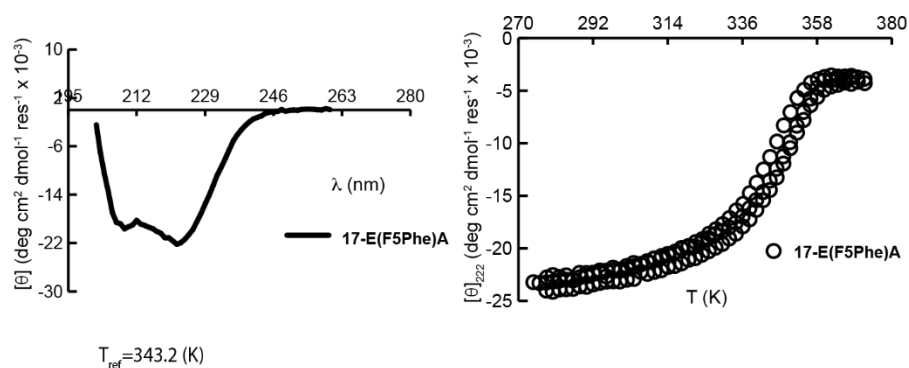


Figure 2-242. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **17** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



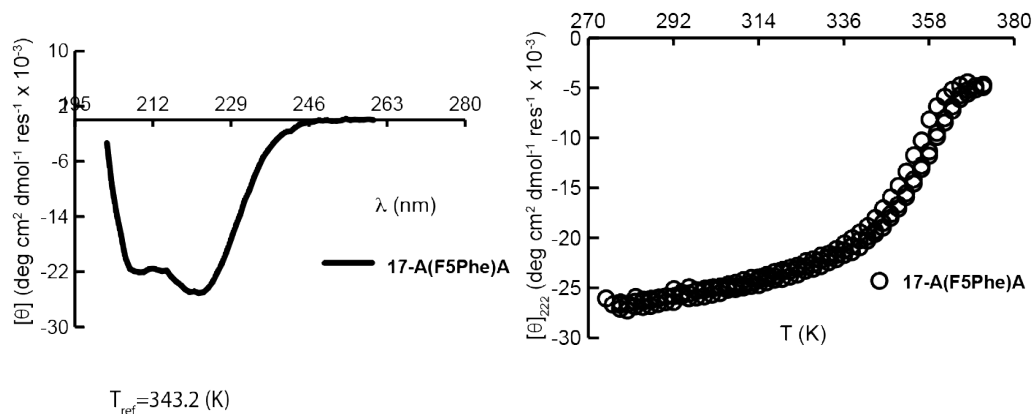
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
17-A(F5Phe)K	-15.9 ± 0.03	0.18 ± 0.003	0.9995	0.1891

Figure 2-243. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **17-A(Ff₅F)K** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
17-E(F5Phe)A	-14.6 ± 0.03	0.19 ± 0.004	0.9994	0.1775

Figure 2-244. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **17-E(f₅F)A** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



Protein	ΔG_f / kcal mol ⁻¹	ΔG_1 / kcal mol ⁻¹ K ⁻¹	R ²	rmsd error
17-A(F5Phe)A	-16.6±0.04	0.19± 0.004	0.9994	0.1788

Figure 2-245. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μ M solutions of proteins **17-A(f5F)A** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

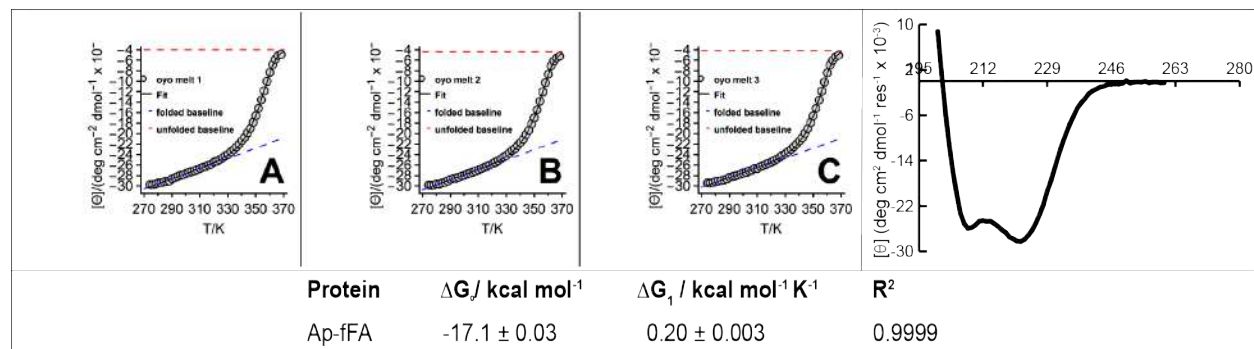


Figure 2-246. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μ M solution of protein **18-A(p-fF)A** in 20 mM sodium phosphate (pH 7). Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

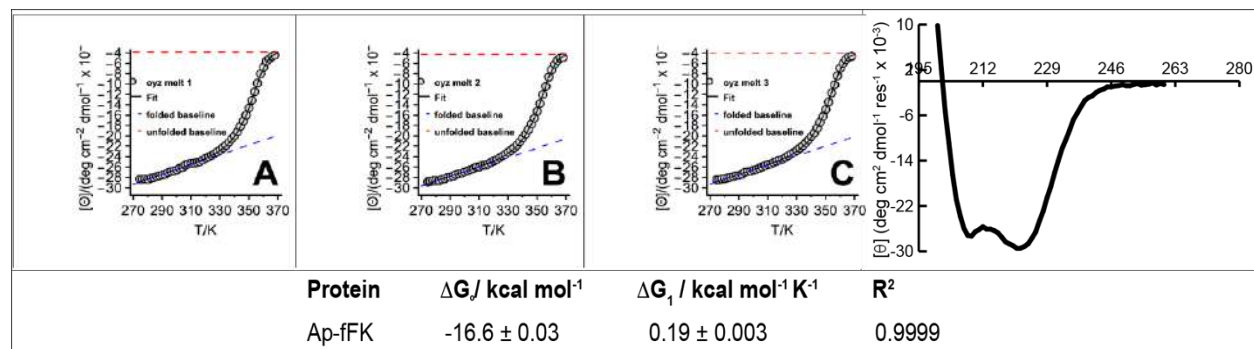


Figure 2-247. ESI TOF spectrum for peptide **3CD** spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μ M solution of protein **18-A(p-fF)K** in 20 mM sodium phosphate (pH 7).

Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

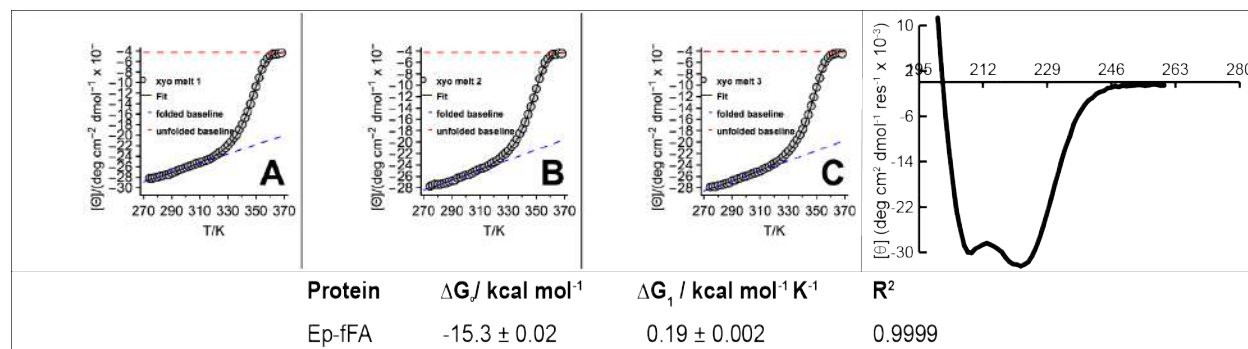


Figure 2-248. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **18-E(p-fF)A** in 20 mM sodium phosphate (pH 7). Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

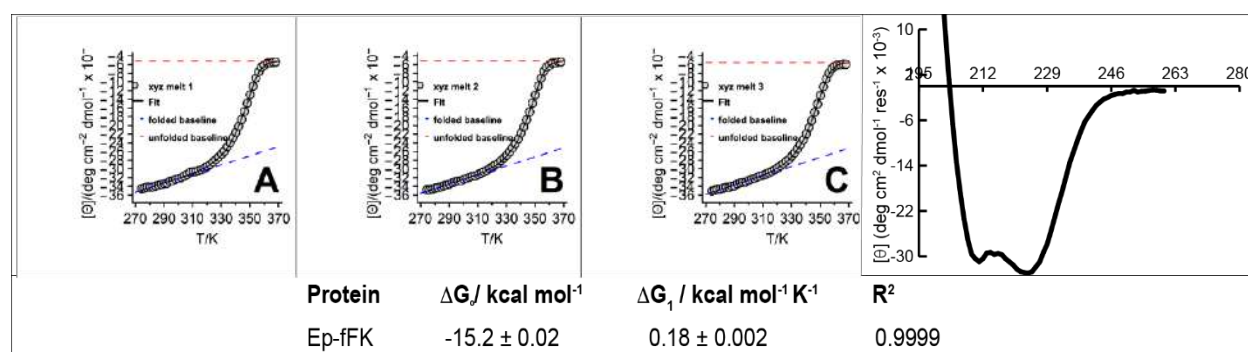


Figure 2-249. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **18** in 20 mM sodium phosphate (pH 7). Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

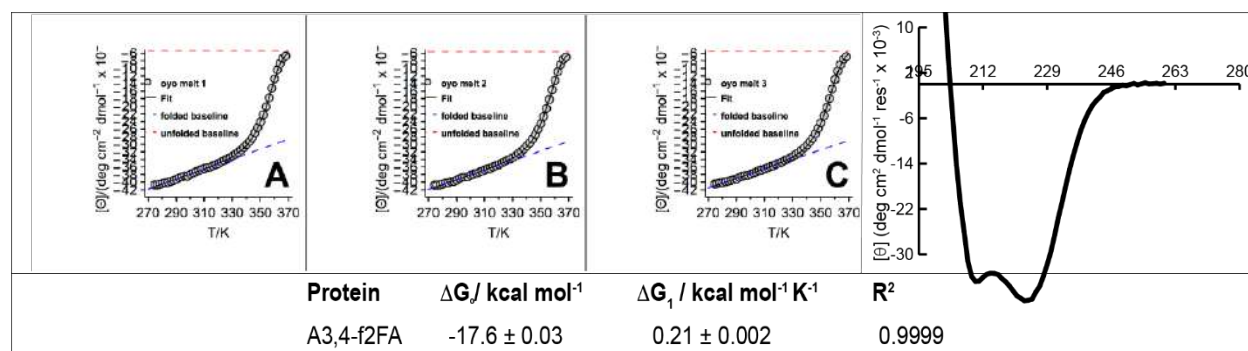


Figure 2-250. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **19-A(3,4-f2F)A** in 20 mM sodium phosphate (pH 7). Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

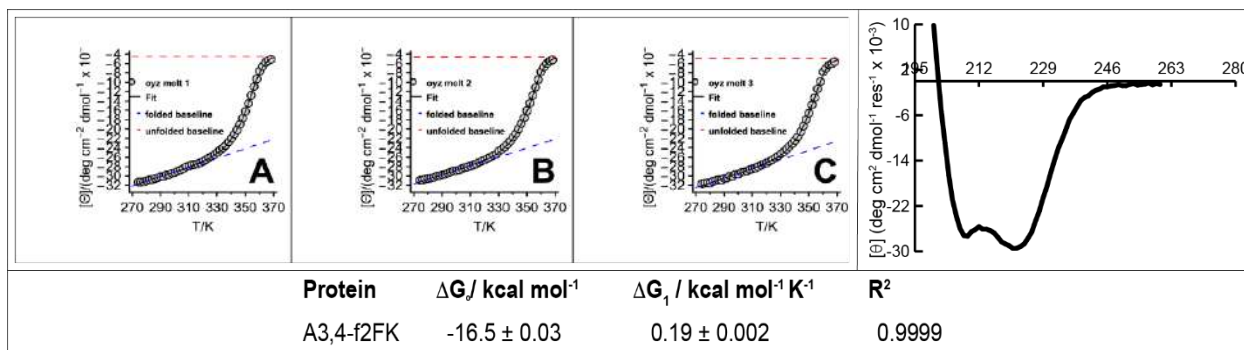


Figure 2-251. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **19-A(3,4-f₂F)K** in 20 mM sodium phosphate (pH 7). Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

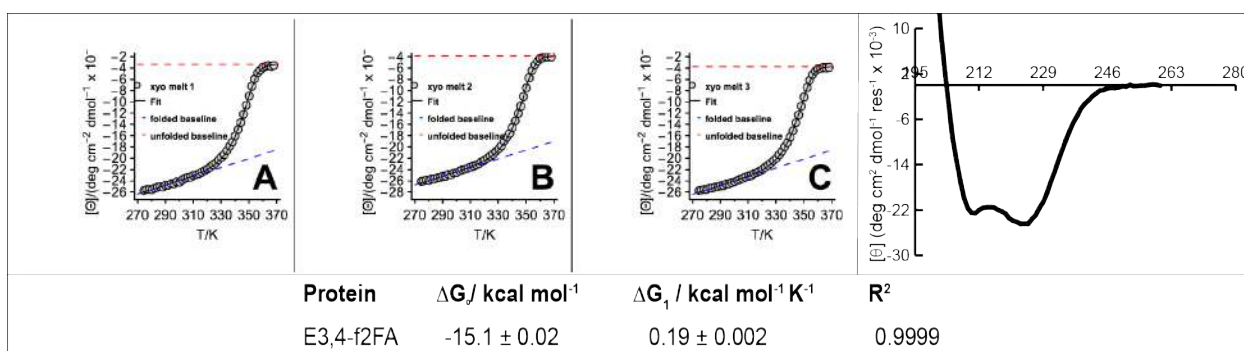


Figure 2-252. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **19-E(3,4-f₂F)A** in 20 mM sodium phosphate (pH 7). Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

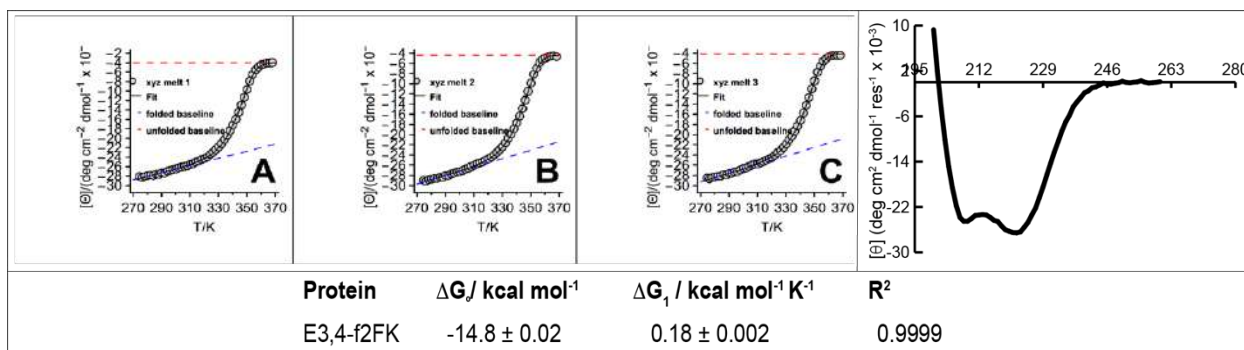


Figure 2-253. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **19** in 20 mM sodium phosphate (pH 7). Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

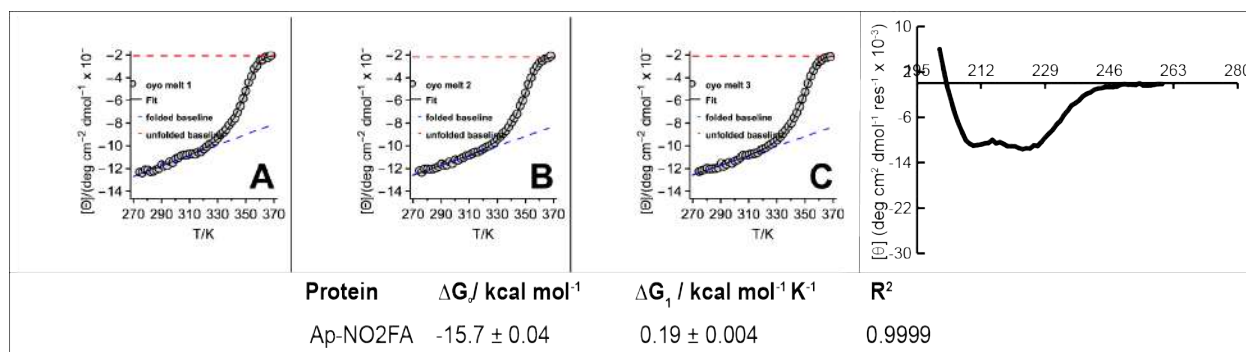


Figure 2-254. ESI TOF spectrum for peptide **3** CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **20-A(p-NO₂F)A** in 20 mM sodium phosphate (pH 7). Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

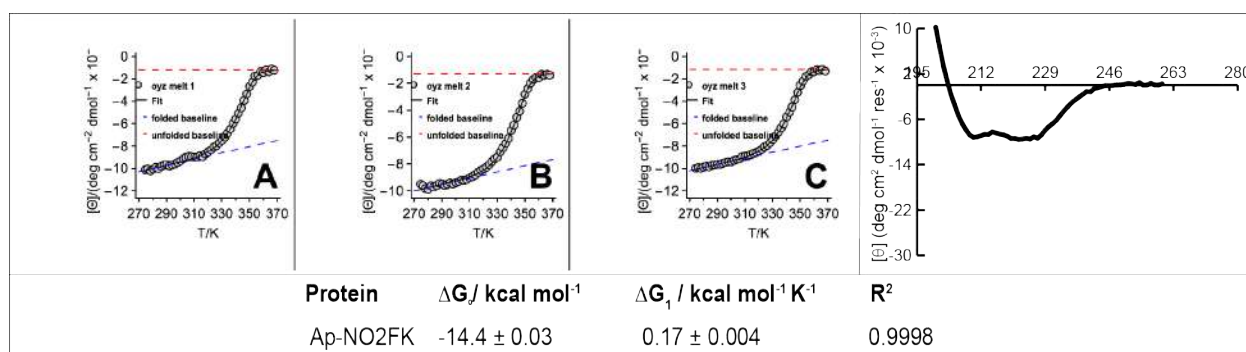


Figure 2-255. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **20-A(p-NO₂F)K** in 20 mM sodium phosphate (pH 7). Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

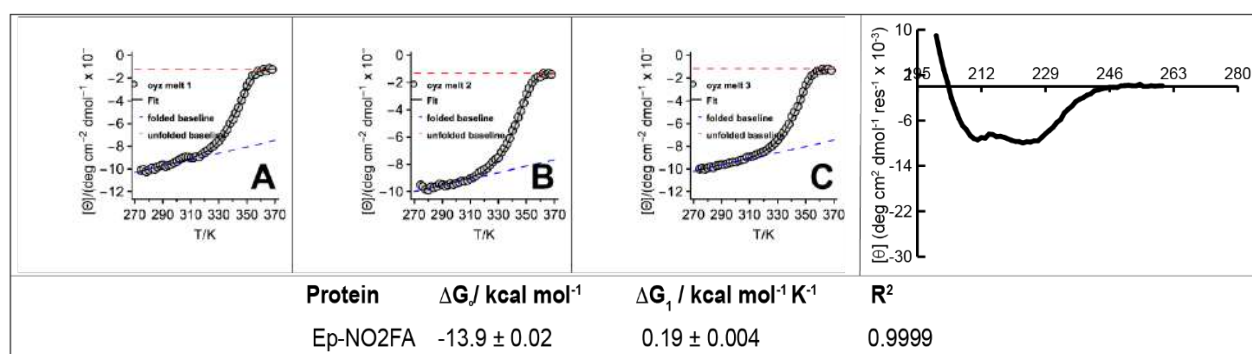


Figure 2-256. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **20-E(p-NO₂F)A** in 20 mM sodium phosphate (pH 7). Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

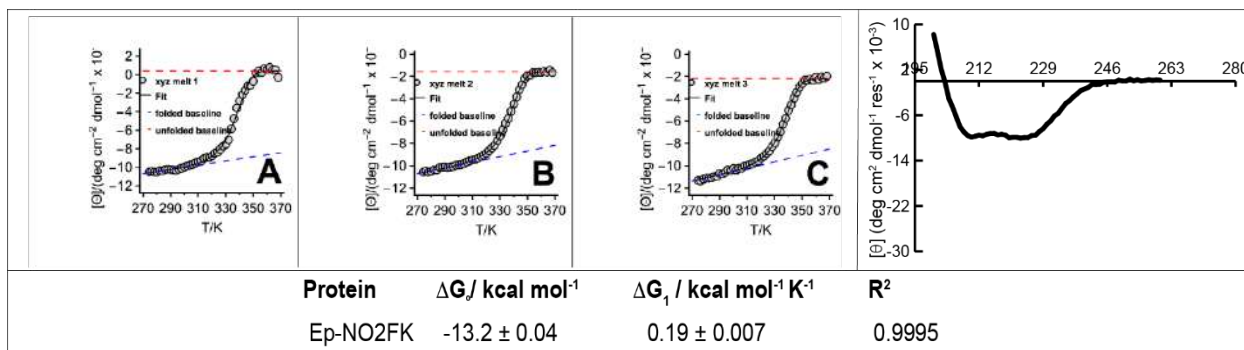


Figure 2-257. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **20** in 20 mM sodium phosphate (pH 7). Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

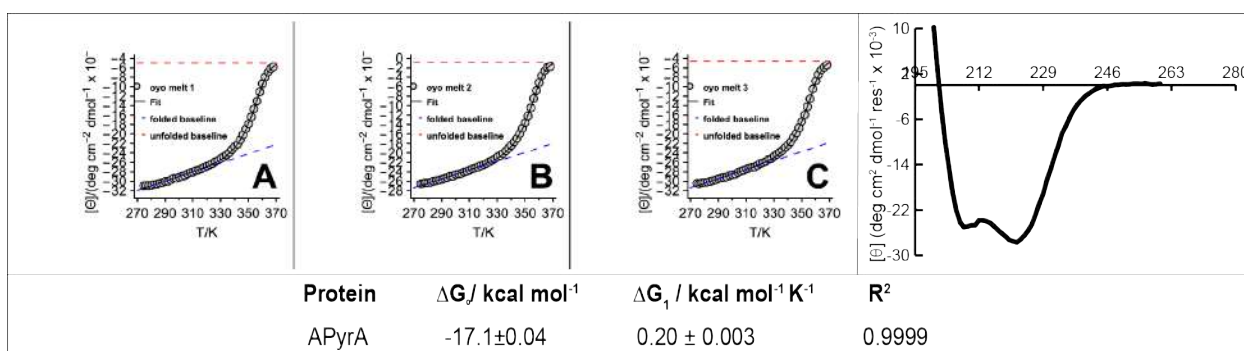


Figure 2-258. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **21-A(Pyra)A** in 20 mM sodium phosphate (pH 7). Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

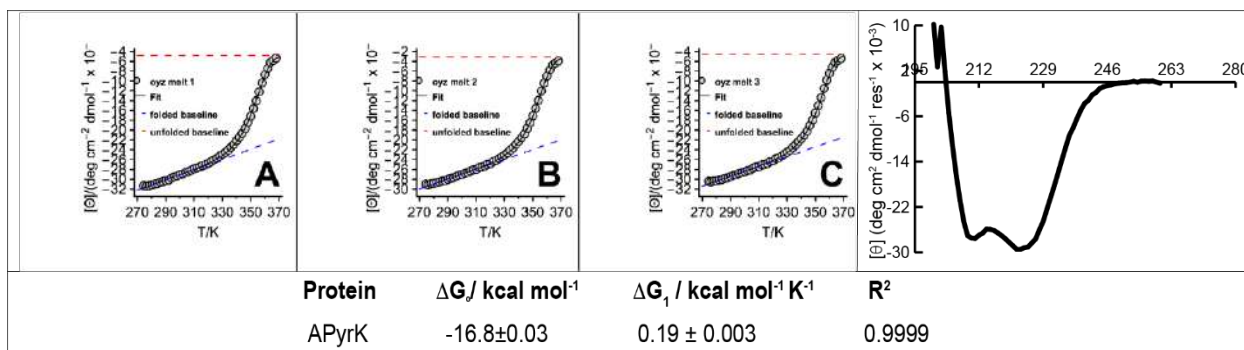


Figure 2-259. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **21-A(Pyra)K** in 20 mM sodium phosphate (pH 7). Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

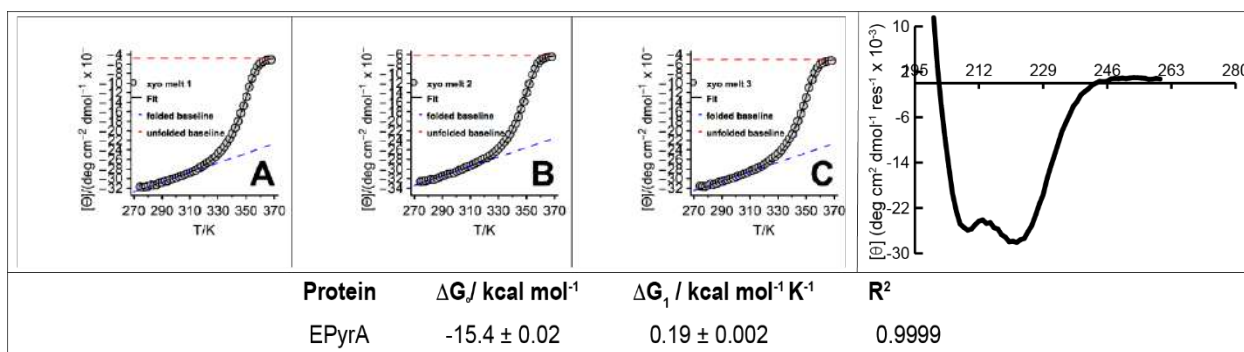


Figure 2-260. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **21-E(PyrA)** in 20 mM sodium phosphate (pH 7). Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

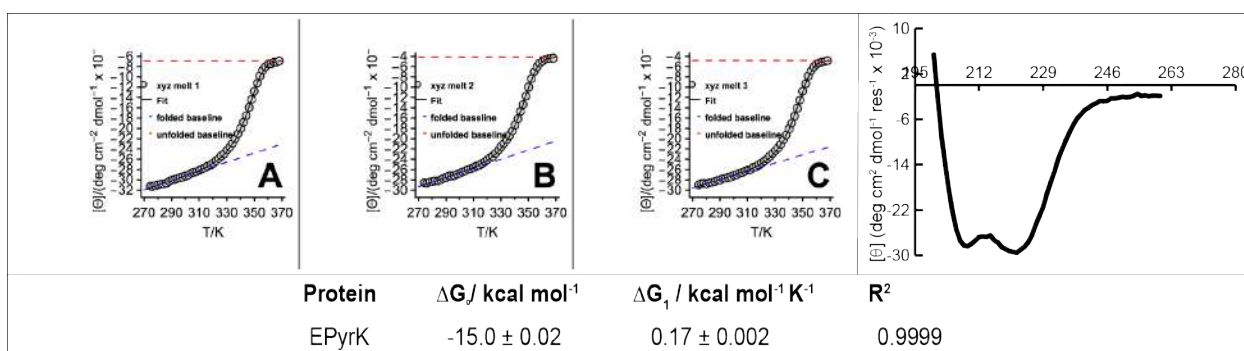
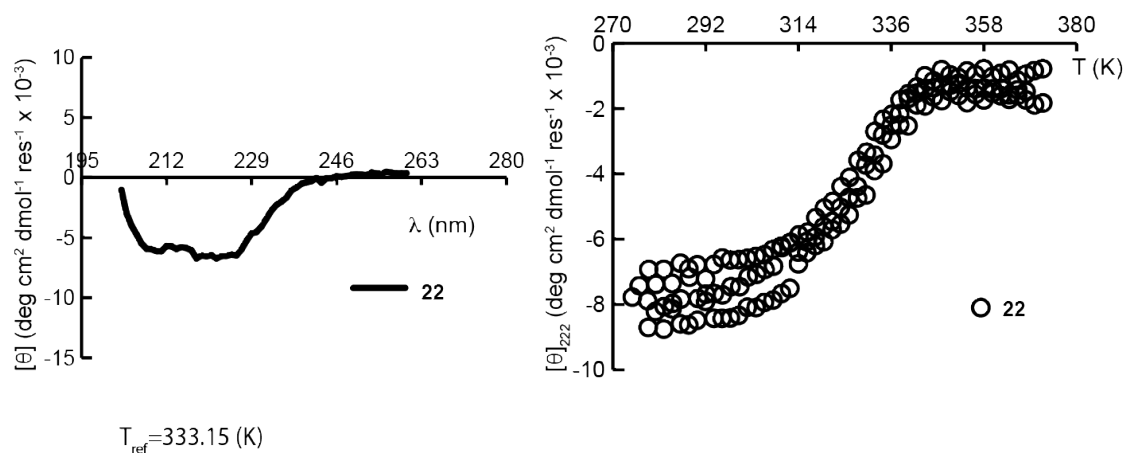
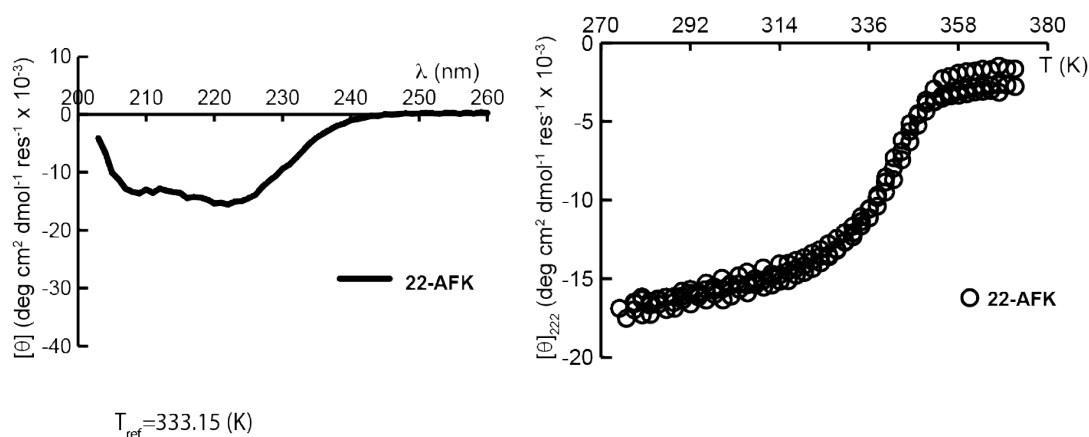


Figure 2-261. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **21** in 20 mM sodium phosphate (pH 7). Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



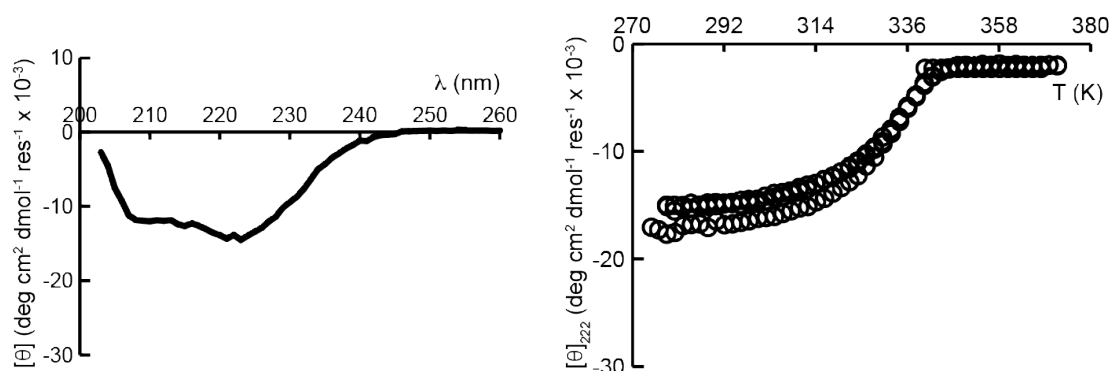
Protein	$\Delta G / \text{kcal mol}^{-1}$	$\Delta G_1 / \text{kcal mol}^{-1} \text{K}^{-1}$	$\Delta G_2 / \text{kcal mol}^{-1} \text{K}^{-2}$	R^2	rmsd error
22	-5.57 ± 0.06	0.116 ± 0.007	0.0012 ± 0.0002	0.9935	0.2217

Figure 2-262. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **22** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



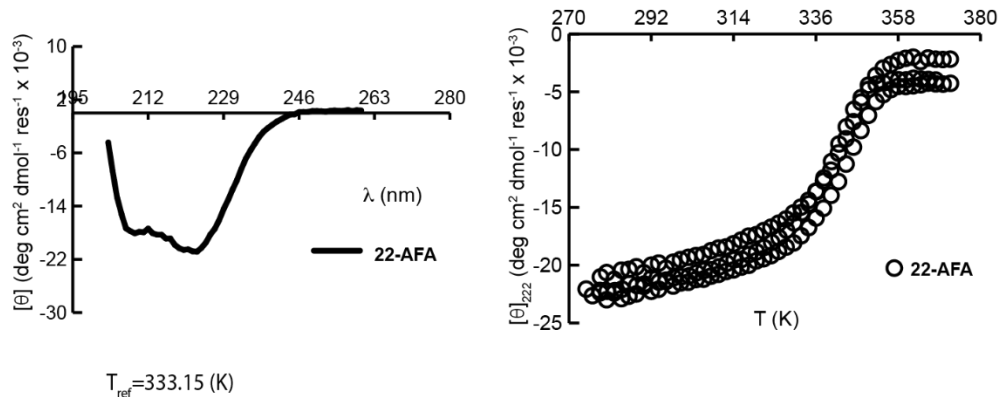
Protein	$\Delta G_0 / \text{kcal mol}^{-1}$	$\Delta G_1 / \text{kcal mol}^{-1} \text{K}^{-1}$	$\Delta G_2 / \text{kcal mol}^{-1} \text{K}^{-2}$	R^2	rmsd error
22-AFK	-7.98 ± 0.04	0.143 ± 0.003	0.0021 ± 0.0001	0.9993	0.1513

Figure 2-263. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **22-AFK** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



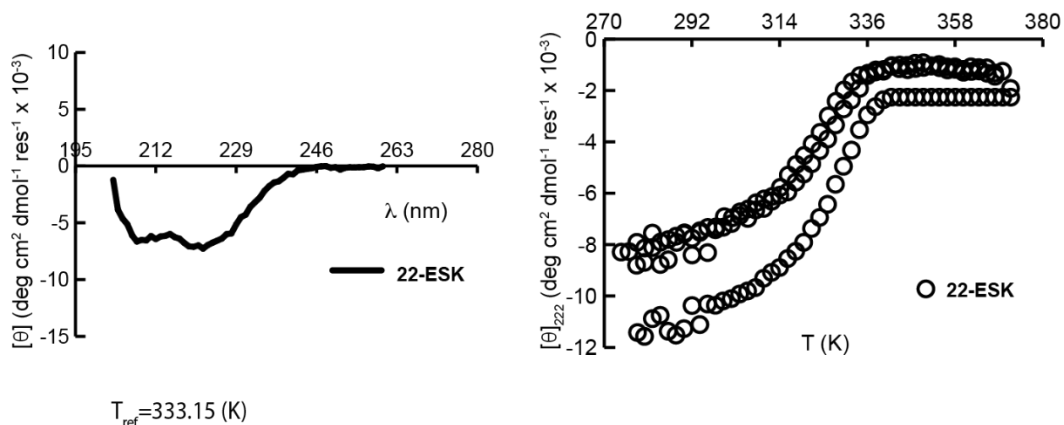
Protein	$\Delta G_0 \text{ (kcal/mol)}$	$\Delta G_1 \text{ (kcal/mol/K)}$	$\Delta G_2 \text{ (kcal/mol/K}^2\text{)}$	$T_{\text{ref}} \text{ (K)}$	R^2	std error
22-EFA	-6.38 ± 0.04	0.179 ± 0.008	0.0027 ± 0.002	333.15	0.9981	0.254

Figure 2-264. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **22-EFA** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



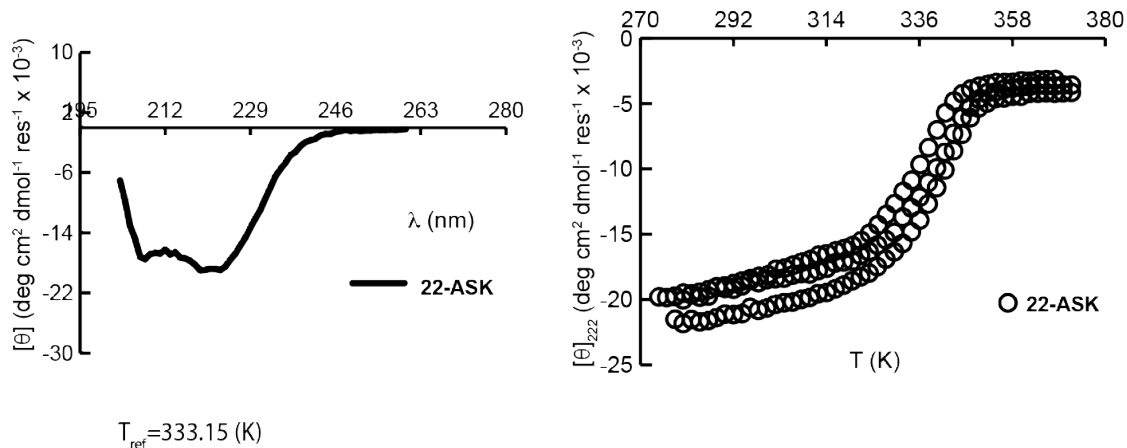
Protein	$\Delta G / \text{kcal mol}^{-1}$	$\Delta G_1 / \text{kcal mol}^{-1} \text{K}^{-1}$	$\Delta G_2 / \text{kcal mol}^{-1} \text{K}^{-2}$	R^2	rmsd error
22-AFA	-8.14 ± 0.03	0.147 ± 0.002	0.002 ± 0.0001	0.9996	0.1469

Figure 2-265. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **22-AFA** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



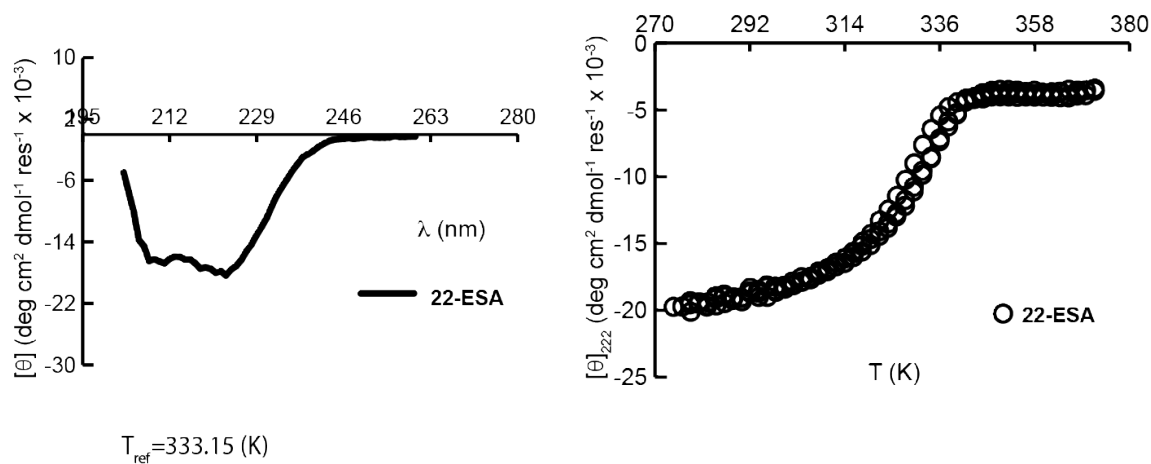
Protein	$\Delta G / \text{kcal mol}^{-1}$	$\Delta G_1 / \text{kcal mol}^{-1} \text{K}^{-1}$	$\Delta G_2 / \text{kcal mol}^{-1} \text{K}^{-2}$	R^2	rmsd error
22-ESK	-4.90 ± 0.08	0.215 ± 0.018	0.002 ± 0.0004	0.9967	0.2014

Figure 2-266. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **22-ESK** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



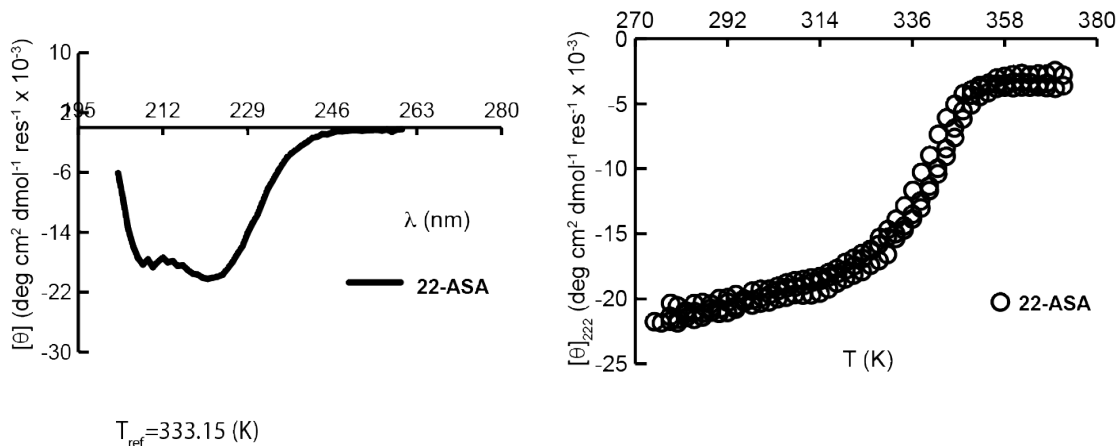
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_1 / \text{kcal mol}^{-1} \text{K}^{-1}$	$\Delta G_2 / \text{kcal mol}^{-1} \text{K}^{-2}$	R^2	rmsd error
22- ASK	-7.43 ± 0.02	0.154 ± 0.003	0.002 ± 0.0002	0.9996	0.1334

Figure 2-267. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **22-ASK** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



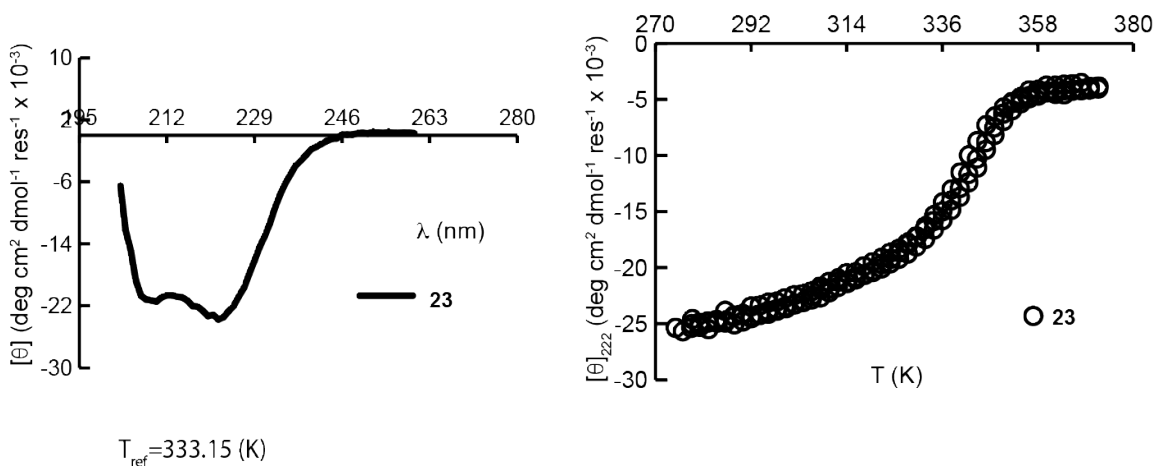
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_1 / \text{kcal mol}^{-1} \text{K}^{-1}$	$\Delta G_2 / \text{kcal mol}^{-1} \text{K}^{-2}$	R^2	rmsd error
22-ESA	-5.85 ± 0.02	0.181 ± 0.004	0.002 ± 0.0001	0.9995	0.1472

Figure 2-268. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **22-ESA** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



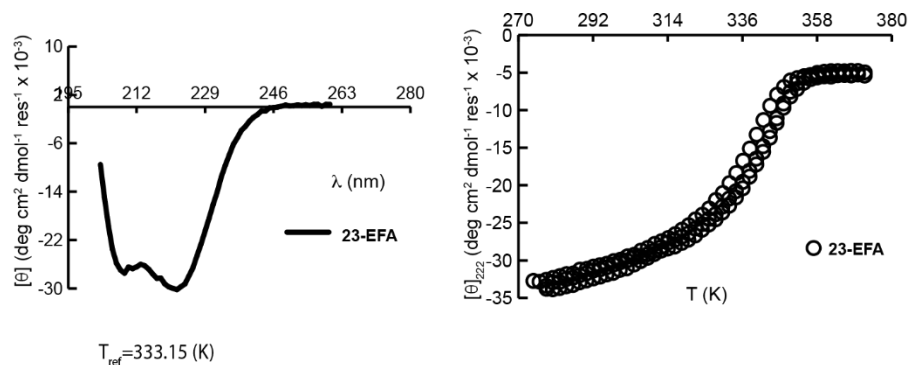
Protein	$\Delta G_0 / \text{kcal mol}^{-1}$	$\Delta G_1 / \text{kcal mol}^{-1} \text{K}^{-1}$	$\Delta G_2 / \text{kcal mol}^{-1} \text{K}^{-2}$	R^2	rmsd error
22-ASA	-7.79 ± 0.04	0.152 ± 0.003	0.002 ± 0.0001	0.9994	0.1819

Figure 2-269. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **22-ASA** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



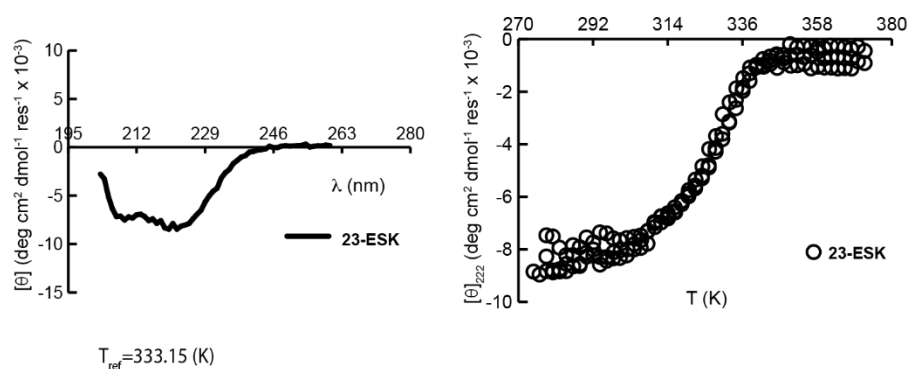
Protein	$\Delta G_0 / \text{kcal mol}^{-1}$	$\Delta G_1 / \text{kcal mol}^{-1} \text{K}^{-1}$	$\Delta G_2 / \text{kcal mol}^{-1} \text{K}^{-2}$	R^2	rmsd error
23	-7.97 ± 0.04	0.129 ± 0.002	0.002 ± 0.0001	0.9996	0.1571

Figure 2-270. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **23** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



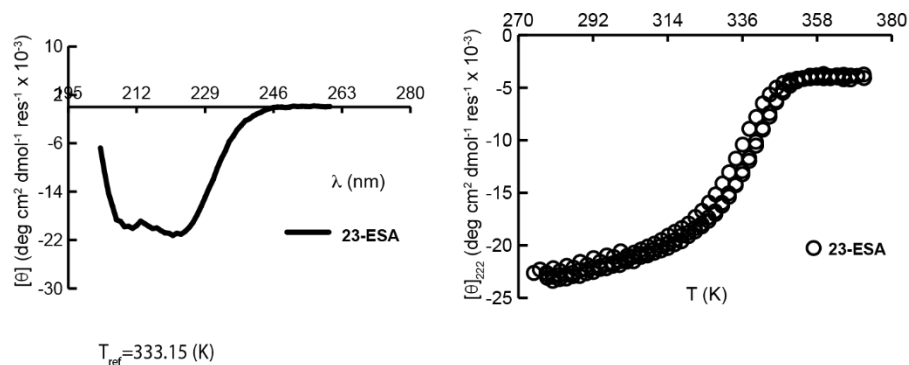
Protein	$\Delta G / \text{kcal mol}^{-1}$	$\Delta G_1 / \text{kcal mol}^{-1} \text{K}^{-1}$	$\Delta G_2 / \text{kcal mol}^{-1} \text{K}^{-2}$	R^2	rmsd error
23-EFA	-7.75 ± 0.02	0.134 ± 0.001	0.002 ± 0.00004	0.9999	0.1114

Figure 2-271. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **23-EFA** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



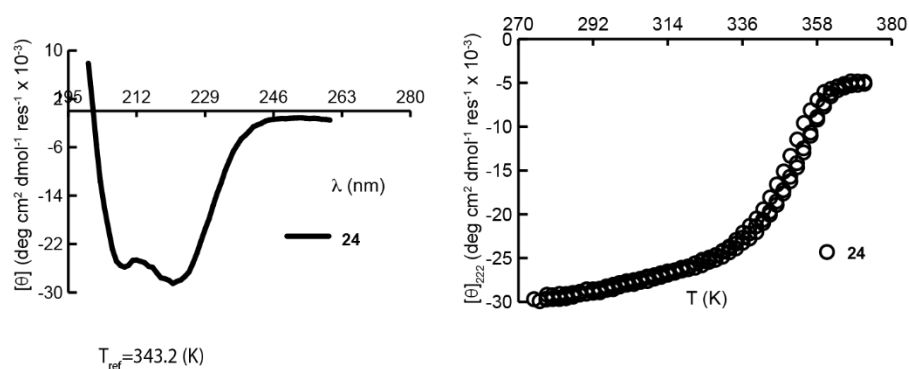
Protein	$\Delta G / \text{kcal mol}^{-1}$	$\Delta G_1 / \text{kcal mol}^{-1} \text{K}^{-1}$	$\Delta G_2 / \text{kcal mol}^{-1} \text{K}^{-2}$	R^2	rmsd error
23-ESK	-5.58 ± 0.05	0.176 ± 0.010	0.003 ± 0.0002	0.9976	0.1675

Figure 2-272. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **23-ESK** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



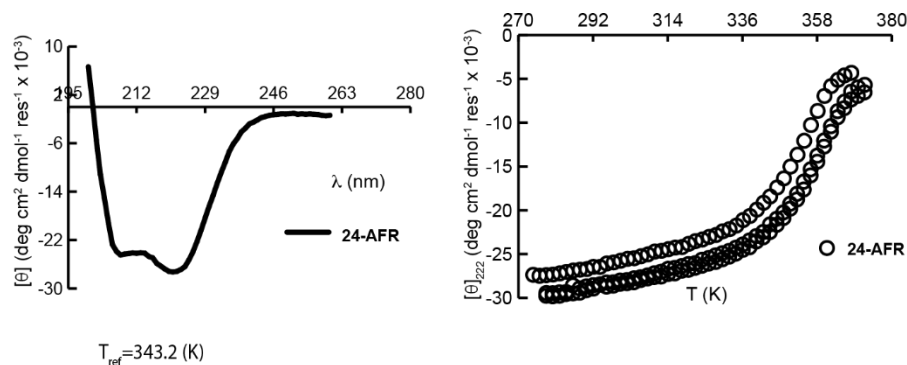
Protein	$\Delta G / \text{kcal mol}^{-1}$	$\Delta G_1 / \text{kcal mol}^{-1} \text{K}^{-1}$	$\Delta G_2 / \text{kcal mol}^{-1} \text{K}^{-2}$	R^2	rmsd error
23-ESA	-7.14 ± 0.02	0.147 ± 0.002	0.002 ± 0.0001	0.9997	0.1370

Figure 2-273. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **23-ESA** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



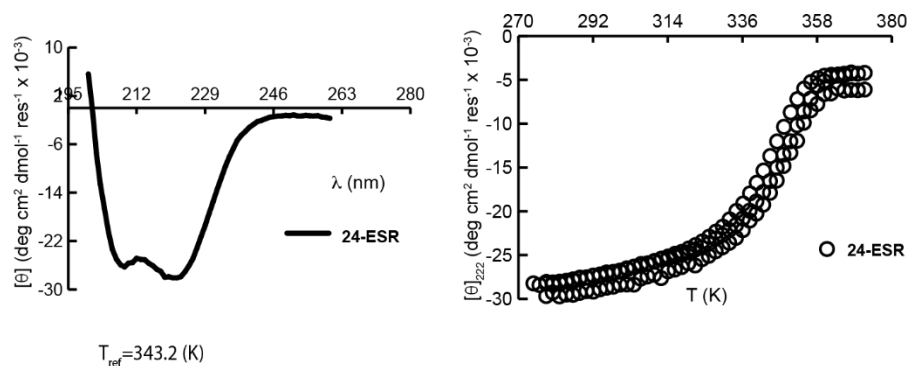
Protein	$\Delta G / \text{kcal mol}^{-1}$	$\Delta G_1 / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
24	-15.6 ± 0.02	0.20 ± 0.002	0.9997	0.1451

Figure 2-274. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **24** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



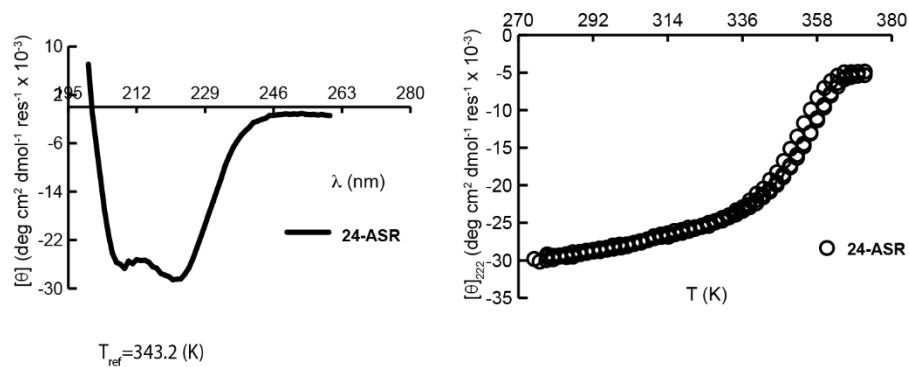
Protein	ΔG_f / kcal mol ⁻¹	ΔG_f / kcal mol ⁻¹ K ⁻¹	R ²	rmsd error
24-AFR	-16.8±0.03	0.19± 0.003	0.9997	0.1337

Figure 2-275. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μ M solutions of proteins **24-AFR** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



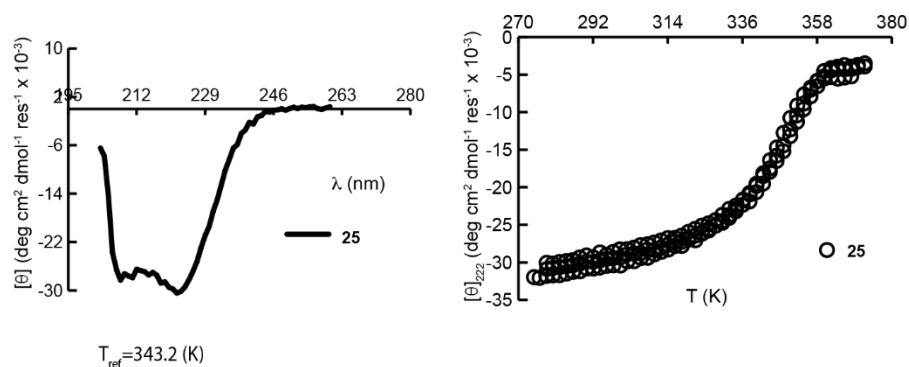
Protein	ΔG_f / kcal mol ⁻¹	ΔG_f / kcal mol ⁻¹ K ⁻¹	R ²	rmsd error
24-ESR	-14.6±0.02	0.18± 0.003	0.9996	0.1732

Figure 2-276. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μ M solutions of proteins **24-ESR** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



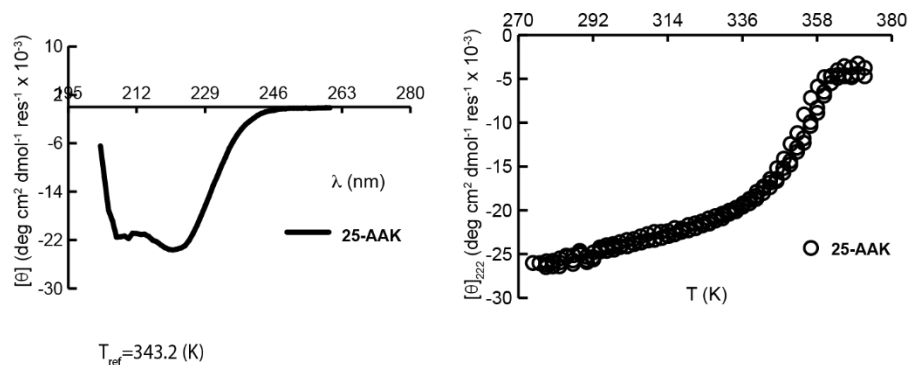
Protein	ΔG_f / kcal mol ⁻¹	ΔG_f / kcal mol ⁻¹ K ⁻¹	R ²	rmsd error
24-ASR	-16.3±0.03	0.20± 0.003	0.9996	0.1608

Figure 2-277. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μ M solutions of proteins **24-ASR** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



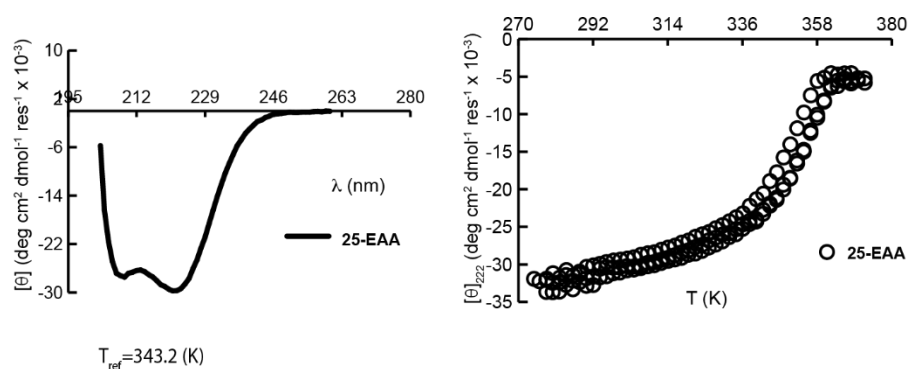
Protein	ΔG_f / kcal mol ⁻¹	ΔG_f / kcal mol ⁻¹ K ⁻¹	R ²	rmsd error
25	-14.8±0.02	0.18± 0.003	0.9996	0.1985

Figure 2-278. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μ M solutions of proteins **25** in 20 mM sodium phosphate, pH 7, 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



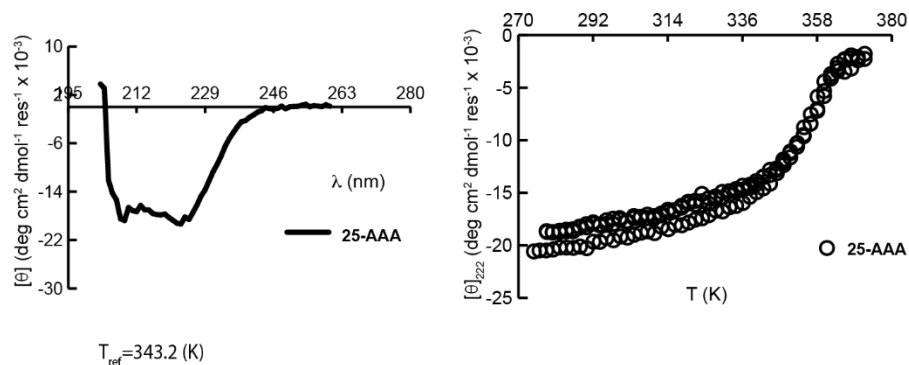
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
25-AAK	-16.7 ± 0.07	0.25 ± 0.007	0.9986	0.2735

Figure 2-279. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **25-AAK** in 20 mM sodium phosphate, pH 7, 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



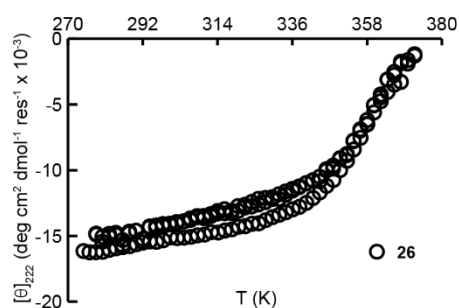
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
25-EAA	-16.3 ± 0.06	0.24 ± 0.006	0.9989	0.3213

Figure 2-280. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **25-EAA** in 20 mM sodium phosphate, pH 7, 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



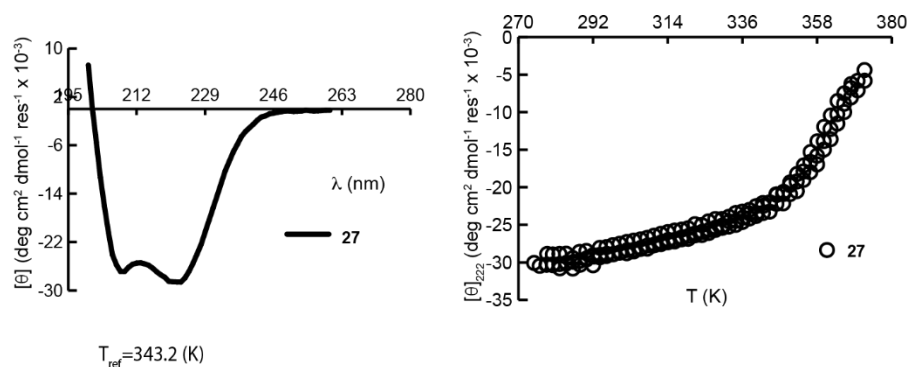
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
25-AAA	-17.3 ± 0.09	0.25 ± 0.007	0.9983	0.2281

Figure 2-281. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **25-AAA** in 20 mM sodium phosphate, pH 7, 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
26	-17.4 ± 0.09	0.21 ± 0.008	0.9977	0.1913

Figure 2-282. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **26** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_f / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
27	-18.2 ± 0.08	0.21 ± 0.006	0.9987	0.2545

Figure 2-283. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **27** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

2.4.6 Single crystal X-ray Diffraction Data.

Freeze dried **EWK** was dissolved in milli Q water at concentration of 5mg/ml and screened against Rigaku Wizard Cryo1 and 2 screens. **EWK₁** and **EWK₂** Crystals were briefly dipped in cryoprotectant (25 % (v/v) glycerol in well solution), then cryo-cooled by plunging into liquid nitrogen prior to data collection. Data were collected at 100 K with a copper rotating anode X-ray source (Rigaku Micromax-007HF), with Varimax-HR confocal optics, and a Rigaku Raxis4++ image plate detector. Each drop contained 2 μl well solution + 2 μl protein. Crystals were cryo-cooled by plunging into liquid nitrogen prior to data collection. Data were collected at 100 K with a copper rotating anode X-ray source (Rigaku Micromax-007HF), with Varimax-HR confocal optics, and a Rigaku Raxis4++ image plate detector.

Freeze dried peptides **14**, **16**, **18**, **19**, **20**, and **21** were dissolved in milli Q water at a concentration of 10 mg/ml and screened against the Rigaku Wizard Cryo 1 and 2 screens. Drops were set with a TPP mosquito liquid handling robot with protein:condition ratios of 1:2, 1:1, and 2:1 at a total drop volume of 300nl. Crystals was harvested in ambient temperatures from sitting

well plates with a nylon loop and immediately frozen in liquid N₂. The crystal was mounted in a stream of cold N₂ and centered in the X-ray beam using a video camera. Low-temperature (100 K) X-ray diffraction data was collected using a MACH3 kappa goniometer coupled to a Bruker Apex II CCD detector with a Bruker-Nonius FR591 rotating anode X-ray source producing Cu K_α radiation ($\lambda = 1.54178 \text{ \AA}$). The Bruker Proteum-3 suite was used to process (integrate and scale) the data

All Structures were determined by molecular replacement with Phaser (CCP4 program suite) using the coordinates of the a coiled-coil trimer structure previously determined (1COI).¹⁹ The initial electron density map indicated two helices each being a helix in a coiled coil trimer; side-chain density was clearly interpretable. Model building was carried out using COOT.³⁹ Refinement was performed with Phenix.⁴⁰ Crystal structure EWK₁ was deposited in the protein databank (PDB ID 5UXT).

Table 2-5 Conditions which produced diffractable crystals for each EXK coiled coil.

Peptide	EXK variant	Precipitant	Buffer	pH	Salt
EWK ₁	EWK	50% PEG 200	0.1 M Na/K Phosphate	6.2	200mM Sodium Chloride
EWK ₂	EWK	30% Jeffamine ED-2001	0.1M HEPES	7.0	
19	E(3,4-f ₂ F)K	40% v/v PEG 300	100 mM Sodium cacodylate/ Hydrochloric acid	6.5	200 mM Calcium acetate
14	EYK	30% v/v PEG 600 (10% glycerol)	100 mM HEPES/ Sodium hydroxide	7.5	50mM lithium sulfate
18	E(p-fPhe)K	40% v/v PEG 300	100 mM Sodium cacodylate/ Hydrochloric acid	6.5	200 mM Calcium acetate
16	EFK	40% v/v PEG 300	100 mM Sodium phosphate dibasic/ Citric acid	4.2	
21	E(PyrA)K	40% v/v PEG 300	100 mM Sodium phosphate dibasic/ Citric acid	4.2	

20	E(p-NO ₂ F)K	40% v/v PEG 300	100 mM Sodium phosphate dibasic/ Citric acid	4.2	
----	-------------------------	-----------------	--	-----	--

Table 2-6 Crystal Data and Refinement Statistics.

	EWK ₁	EWK ₂
Space group	C2	R3
Unit Cell		
a, b, c	84.8, 38.5, 37.0	39.4, 39.4, 98.7
α, β, γ	90, 113.3, 90	90, 90, 120
Data collection		
Resolution	40-2.20	40-1.80
R merge (%)	7.3 (26.5)	2.8 (191.9)
I/ σ (I)	4.0 (1.0)	5(0.5)
completeness	92 (62.7)	99.1(91.4)
Redundancy	3.9	7.3
Refinement		
Resolution	40-2.20	40-1.80
# of reflections	6128	5325
R _{work} /R _{free}	19.4/27.5	23.4/27.8
# of atoms	845	523
Ave B factor	44.6	55
Rmsd		
Bond lengths (Å)	0.007	0.007
Bond angles (°)	0.838	0.950

Table 2-7 Crystal Data and Refinement Statistics.

	E(3,4-f₂F)K	EYK	E(p-fF)K
Space group	R3	R3	R3
Unit Cell			
a, b, c	39.28, 39.28, 98.71	38.04, 38.04, 104.28	39.07, 39.07, 98.62
α, β, γ	90, 90, 120	90, 90, 120	90, 90, 120
Data collection			
Resolution	32.91-2.10	33.28-1.80	32.88-2.15
R merge (%)	15.2 (38.7)	9.9(31.8)	16.3 (35.3)
I/ σ (I)	17.1 (2.0)	12.7(1.4)	7.4 (1.4)
completeness	100 (98.9)	99.8(98.1)	99.7 (100)
Redundancy	4.6 (2.4)	4.7(3.0)	4.1 (2.5)
Refinement			
Resolution	32.91-2.10	33.28-1.80	32.88-2.15
# of reflections	3325	5361	3077
R _{work} /R _{free}	19.7/25.5	14.6/17.3	20.0/27.8
# of atoms	578	664	564
Ave B factor	28.8	18.8	26.7
Rmsd			
Bond lengths (Å)	0.008	0.010	0.007
Bond angles (°)	1.05	1.11	0.98

Table 2-8 Crystal Data and Refinement Statistics.

	EFK	E(PYRA)K	E(p-NO₂F)K
Space group	R3	R3	R3
Unit Cell			
a, b, c	38.89, 38.89, 103.47	38.38, 38.38, 119.43	39.05, 39.05, 101.16
α, β, γ	90, 90, 120	90, 90, 120	90, 90, 120
Data collection			
Resolution	34.49-2.15	39.82-2.2	33.73-2.1
R merge (%)	19.3 (40.6)	17.7 (53.9)	22.0 (62.6)
I/ σ (I)	9.4 (2.1)	7.5 (1.7)	10.7 (1.3)
completeness	99.7 (100)	100 (100)	100 (100)
Redundancy	7.7 (4.4)	4.9 (3.3)	10.9 (6.2)
Refinement			
Resolution	34.49-2.15	39.82-2.2	33.73-2.1
# of reflections	3180	3343	3366
R _{work} /R _{free}	17.6/24.4	18.7/26.8	20.9/27.5
# of atoms	604	595	586
Ave B factor	20.5	25.18	27.52
Rmsd			
Bond lengths (Å)	0.008	0.008	0.007
Bond angles (°)	0.96	0.99	1.00

2.4.7 Justification for using Ser at position 14 instead of Ala as a control for evaluating Trp-based enhancement of the Glu10-Lys18 Interaction.

In the main text, we demonstrated that placing Trp at position 14 enhances the strength of a long-range i to $i+8$ Glu10-Lys18 salt bridge better than does Ser. We chose Ser as a negative control because of its polar character and small size, and because it occupies this position in the sequence of the original 1CW coiled coil from which our variants were derived. However, we wondered whether the use of Ser as a negative control instead of the more conventional Ala might have artificially inflated the observed impact of Trp on the Glu10-Lys18 interaction. To explore this possibility, we prepared peptides **25**, **25-AAK**, **25-EAA**, and **25-AAA**, derivatives of peptides 5-8 in which Ala occupies position 14 instead of Ser. Variable temperature CD experiments revealed that peptides **25**, **25-AAK**, **25-EAA**, and **25-AAA** are too stable in 20 mM sodium phosphate buffer (pH 7) to reliably identify a fully unfolded baseline, and therefore we were unable to extract folding free energy values for these peptides under these conditions. However, we were able to fit variable temperature CD data for peptides **25**, **25-AAK**, **25-EAA**, and **25-AAA** in 20 mM sodium phosphate (pH 7) + 1 M urea; we have folding free energy data for peptides 1-8 under identical conditions. We observed that impact of Trp relative to Ala on the Glu10-Lys18 interaction is substantially larger than the impact of Trp relative to Ser (Supplementary Table 13). These results indicate that using Ser as a negative control leads to a smaller, more conservative estimate for Trp-based enhancement of the Glu10-Lys18 interaction than when Ala is used as negative control, suggesting that our concerns were unfounded.

Table2- 9. Sequences and folding free energies of GCN4-p1 variants **20**, **21**, **22** and their derivatives.^a

Peptide	Sequence	ΔG_f (kcal/mol) in 1M Urea
1	Ac-EVEALEKKV E ALE W KV Q KLEKKVEALEHGWDGR-CONH ₂	-16.23 ± 0.05
2	Ac-EVEALEKKV A ALE W KV Q KLEKKVEALEHGWDGR-CONH ₂	-15.73 ± 0.04

3	Ac-EVEALEKKV E ALE W KV Q ALEKKVEALEHGWDGR-CONH ₂	-14.04 ± 0.03
4	Ac-EVEALEKKV A ALE W KV Q ALEKKVEALEHGWDGR-CONH ₂	-15.78 ± 0.05
5	Ac-EVEALEKKV E ALE S KV Q KLEKKVEALEHGWDGR-CONH ₂	-13.42 ± 0.02
6	Ac-EVEALEKKV A ALE S KV Q KLEKKVEALEHGWDGR-CONH ₂	-14.74 ± 0.02
7	Ac-EVEALEKKV E ALE S KV Q ALEKKVEALEHGWDGR-CONH ₂	-14.18 ± 0.02
8	Ac-EVEALEKKV A ALE S KV Q ALEKKVEALEHGWDGR-CONH ₂	-15.30 ± 0.04
25	Ac-EVEALEKKV E ALE A KV Q KLEKKVEALEHGWDGR-CONH ₂	-14.78 ± 0.02
25-AAK	Ac-EVEALEKKV A ALE A KV Q KLEKKVEALEHGWDGR-CONH ₂	-16.66 ± 0.07
25-EAA	Ac-EVEALEKKV E ALE A KV Q ALEKKVEALEHGWDGR-CONH ₂	-16.31 ± 0.06
25-AAA	Ac-EVEALEKKV A ALE A KV Q ALEKKVEALEHGWDGR-CONH ₂	-17.32 ± 0.09

^aData are given ± standard error at 30 μM protein concentration in 20 mM sodium phosphate buffer (pH 7) + 1 M urea at 343.15 K.

Table 2-10. Impact of Trp14 relative to Ser14 or Ala14 on the long-range interaction between Glu10 and Lys18 in the homotrimeric 1CW coiled coil.^a

	ΔΔΔΔG _r (kcal/mol)
Impact of Trp vs. Ser on the Glu10-Lys18 Interaction	-2.45 ± 0.10
Impact of Trp vs. Ala on the Glu10-Lys18 Interaction	-3.12 ± 0.15

^aData are given ± standard error at 30 μM protein concentration in 20 mM sodium phosphate buffer (pH 7) + 1M urea at 343.15 K. Triple mutant cycle analysis performed for peptides 1–4 in comparison to Ser-containing peptides 5–8, vs. Ala-containing peptides 25-AAK, 25-EAA, 25-AAA.

2.4.8 Impact of Salt on Phe based enhancement of Glu10:Lys18 salt bridge

We wondered whether the ability of Phe to enhance the long-range Glu10-Lys18 interaction depends on the concentration of salt in the buffer. To explore this possibility, we performed triplicate variable temperature CD experiments on peptides 5–8, 16, 16-AFK, 16-EFA, and 16-AFA in 20 mM sodium phosphate (pH 7) and 0.25 M NaCl. The results of this analysis are shown in Supplementary Tables 14 and 15. The impact of Phe on the Glu10-Lys18 interaction gets stronger in the presence of 0.25 M NaCl, indicating that it is resistant to screening by salt.

Table 2-11. Sequences, expected and observed exact masses, and folding free energies peptides 5–8, 16, 16-AFK, 16-EFA, and 16-AFA.^a

Peptide	Sequence	ΔG _r (kcal/mol)
5	Ac-EVEALEKKV E ALE S KV Q KLEKKVEALEHGWDGR-CONH ₂	-14.60 ± 0.02
6	Ac-EVEALEKKV A ALE S KV Q KLEKKVEALEHGWDGR-CONH ₂	-16.12 ± 0.03
7	Ac-EVEALEKKV E ALE S KV Q ALEKKVEALEHGWDGR-CONH ₂	-15.77 ± 0.03
8	Ac-EVEALEKKV A ALE S KV Q ALEKKVEALEHGWDGR-CONH ₂	-17.08 ± 0.05
16	Ac-EVEALEKKV E ALE F KV Q KLEKKVEALEHGWDGR-CONH ₂	-15.37 ± 0.04
16-AFK	Ac-EVEALEKKV A ALE F KV Q KLEKKVEALEHGWDGR-CONH ₂	-15.91 ± 0.05
16-EFA	Ac-EVEALEKKV E ALE F KV Q ALEKKVEALEHGWDGR-CONH ₂	-15.55 ± 0.06

^aData are given ± standard error at 30 μM protein concentration in 20 mM sodium phosphate buffer (pH 7) + 0.25 M NaCl at 343.15 K.

Table 2-12. Impact of Phe14 on the long-range interaction between Glu10 and Lys18 in the homotrimeric 1CW coiled coil in 20 mM sodium phosphate (pH 7), with or without 0.25 M NaCl.^a

	$\Delta\Delta\Delta G_f$ (kcal/mol)
Phe-based stabilization in 20 mM sodium phosphate (pH 7)	-1.25 ± 0.09
Phe-based stabilization in 20 mM sodium phosphate (pH 7) + 0.25 M NaCl	-2.00 ± 0.12

^aData are given ± standard error at 30 μM protein concentration in 20 mM sodium phosphate buffer (pH 7) + 0.25 M NaCl at 343.15 K.

2.4.9 NMR Tocsy and Roesy data

As described in the main text, peptide **pA-EWK** (Table S1) is derived from a previously reported parent α -helical model peptide **pA**⁴¹ (sequence Ac-OOAAAAAAAAAAAAAAAAAAOOY-NH₂), which contains a 19-residue polyalanine segment (for high helix propensity), flanked by two ornithine residues on either end (to increase solubility in aqueous solution), and a C-terminal tyrosine (a spectroscopic handle for accurate concentration determination). In **pA-EWK**, Glu occupies position 8, Trp occupies position 12, and Lys occupies position 16. Peptides **pA-AWK** and **pA-EWA** are derivatives of **pA-EWK** in which Ala replaces Glu8 or Lys16, respectively (Table S1).

NMR samples for **pA-EWA**, **pA-AWK**, and **pA-EWK** were prepared at 5 mM peptide concentration in 20 mM sodium phosphate (pH 7) with 10% D₂O. Samples were transferred to a Varian 500 MHz magnet and data was collect using vnmrJ software. Water suppression was achieved using an excitation sculpting sequence. 2D TOCSY experiments were collected with 8 scans, 400 t1 increments, and 80 ms mixing time. 2D adiabatic ROESY experiments were collected with 64 scans, 256 t1 increments, and 200 ms mixing time. Data was processed using vnmrJ followed by ccpNMR software. Examples of the 2D spectra for **pA-EWA**, **pA-AWK**, and **pA-EWK** are shown in figures S77-S79, summaries of assignments and correlations are shown in tables S3-S8. CD data for **pA-EWA**, **pA-AWK**, and **pA-EWK** (Figure S80) confirm their similar weakly helical secondary structure.

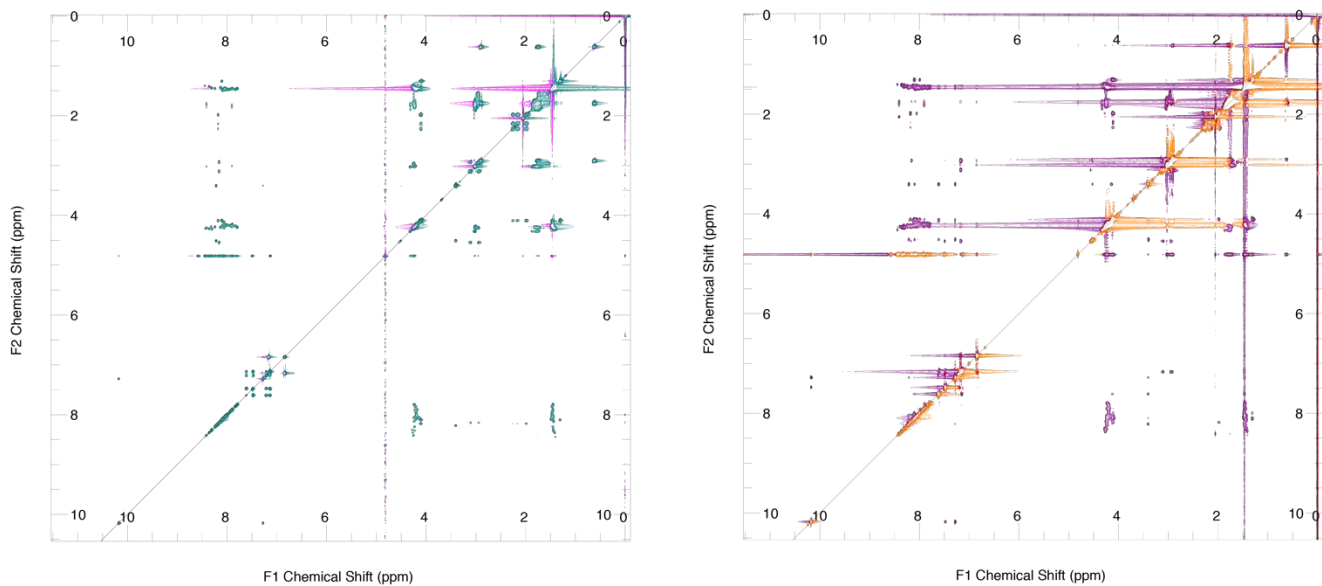


Figure 2-284. NMR Data for pA-EWA. 2D TOCSY data (left), 2D ROESY data (right).

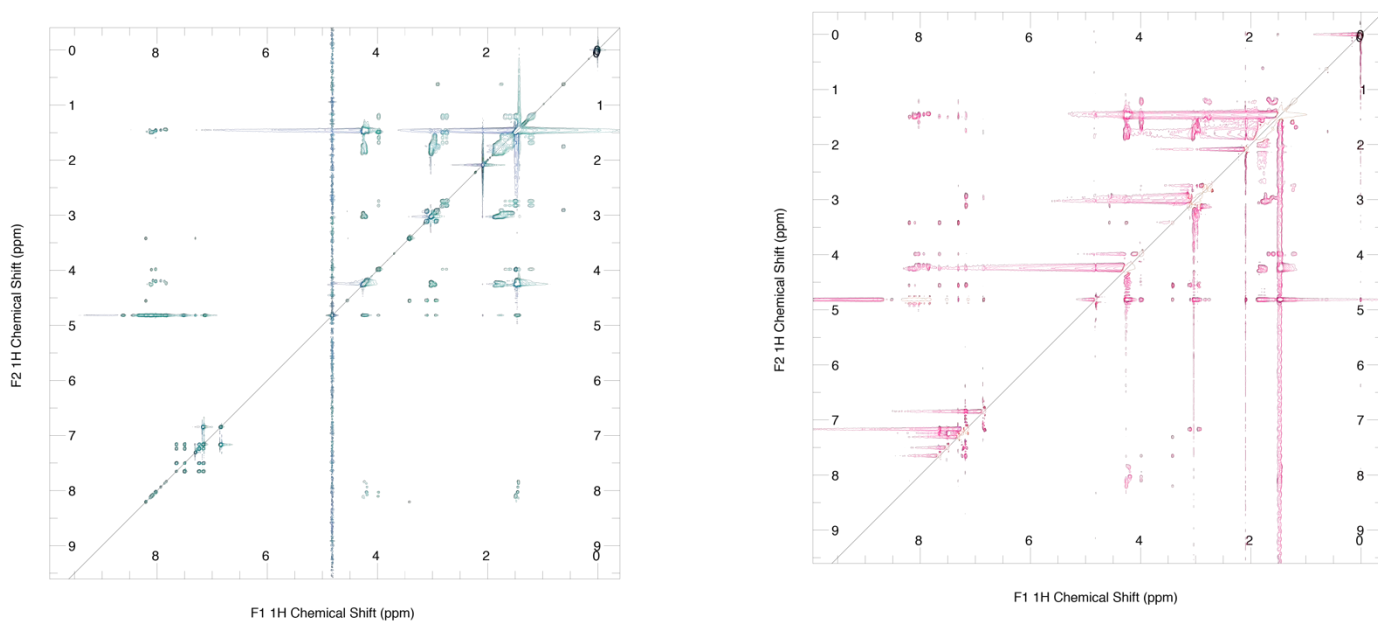


Figure 2-285. NMR Data for pA-AWK. 2D TOCSY data (left), 2D ROESY data (right).

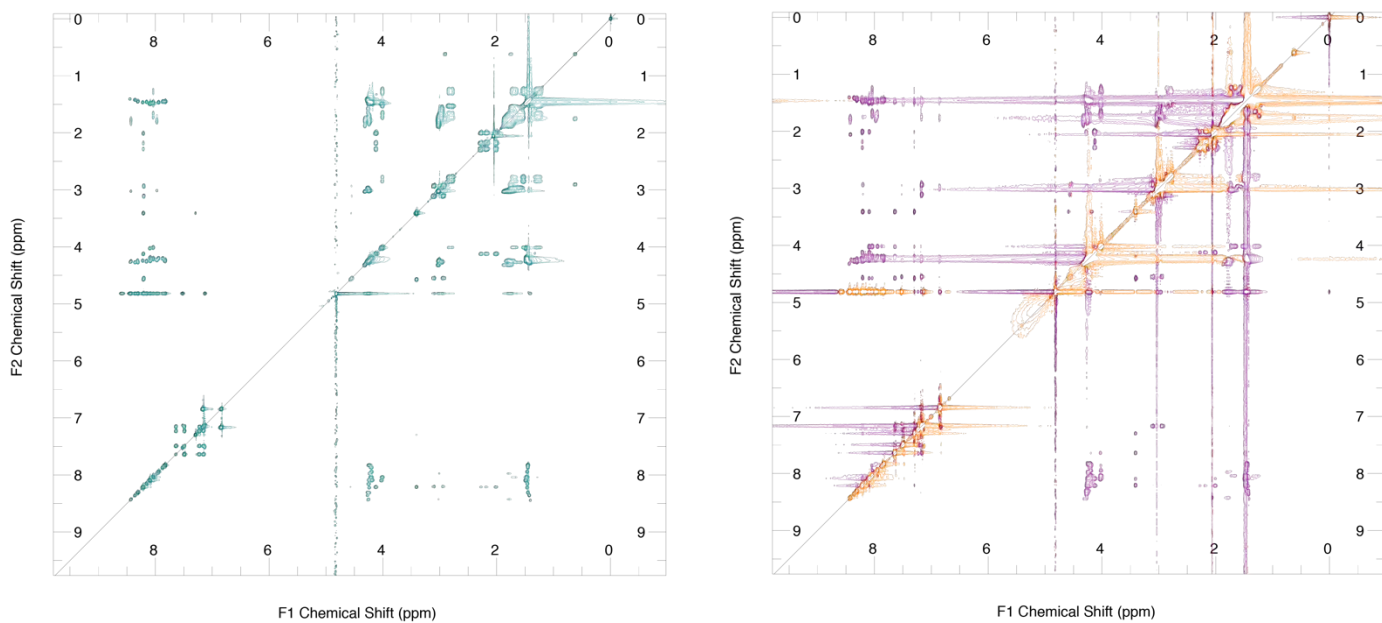


Figure 2-286. NMR Data for **pA-EWK**. 2D TOCSY data (left), 2D ROESY data (right).

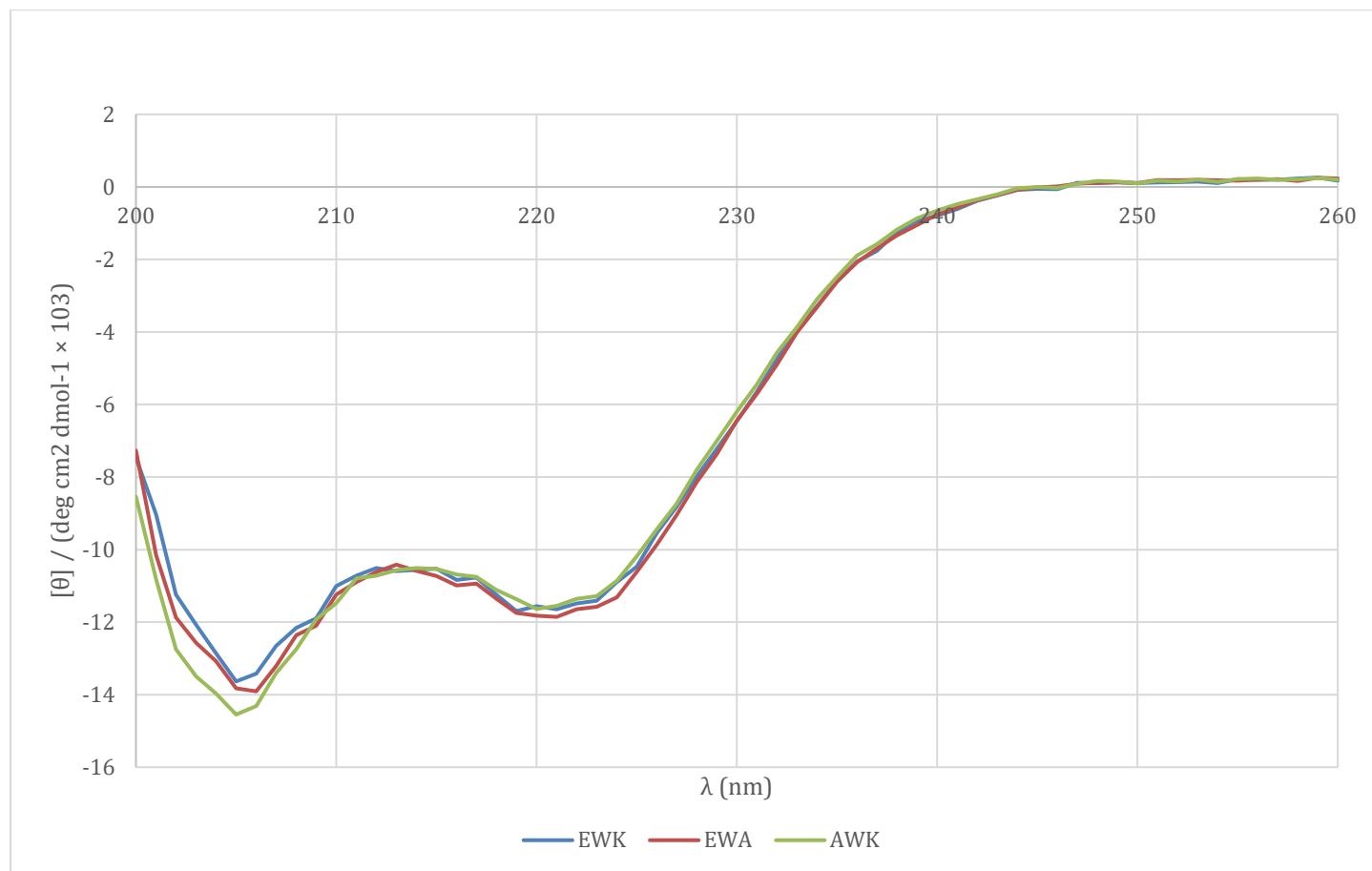


Figure 2-287. CD spectra for **pA-EWK**, **pA-AWK**, **pA-EWA** at 100 μ M concentration in 20 mM sodium phosphate pH 7, 25 $^{\circ}$ C.

A summary of the observed NOEs for **pA-AWK**, **pA-EWA**, and **pA-EWK** appears in Figure S81. As with **pA-EWK**, the ROESY spectrum for **pA-AWK** shows many NOEs between Trp12 and Lys16 backbone and side-chain protons. However, several important differences between the ROESY spectra of **pA-EWK** vs. **pA-AWK** highlight the impact of Glu8 vs. Ala8 on Trp-Lys contacts (Figure S81). For example, NOE peaks corresponding to Trp12H ζ 3/Lys16H α , Trp12H ζ 3/Lys16H δ , and Trp12H η 2/Lys16H δ are present in the spectrum of **pA-AWK** but not in that of **pA-EWK**. In contrast, NOE peaks corresponding to Trp12H ϵ 1/Lys16H γ , Trp12H ϵ 1/Lys16H δ , and Trp12H ϵ 1/Lys16H ϵ are present in the spectrum of **pA-EWK** but not in that of **pA-AWK**. Moreover, normalized peak volumes for the Trp12H ϵ 3/Lys16H δ , Trp12H α /Lys16NH, and Trp12 H δ 1/Lys16H δ NOEs are larger for **pA-EWK** than for **pA-AWK**, whereas normalized peak volumes for the Trp12H δ 1/Lys16H ϵ and Trp12H ζ 2/Lys16H ϵ NOEs are smaller for **pA-EWK** than for **pA-AWK**. These changes suggest that replacing Ala8 with Glu8 allows Lys16H δ to move closer to H δ 1, H ϵ 1, and H ϵ 3 on the right-hand side of the Trp12 indole ring (as drawn in Figures 2), whereas Lys16NH moves closer to Trp12H α , possibly due to localized overwinding of the α -helical backbone between Trp12 and Lys16 to allow for optimized interaction geometry between Glu8, Trp12, and Lys16.

Similarly, differences between the ROESY spectra of **pA-EWK** vs. **pA-EWA** highlight impact of Lys16 vs. Ala16 on Glu8-Trp12 contacts. The Trp12H δ 1/Glu8H α NOE is present in the spectrum of **pA-EWA**, but not that of **pA-EWK**. Moreover, normalized peak volumes for the Trp12H δ 1/Glu8H β , Trp12H δ 1/Glu8H γ , Trp12H ϵ 1/Glu8H β , and Trp12H ϵ 1/Glu8H γ NOEs are smaller in **pA-EWK** than in **pA-AWK**. These observations indicate that the Glu8 side-chain

protons are close to H δ 1 and H ϵ 1 of Trp12 (as is also the case for H δ and H ϵ of Lys16, see above), but that replacing Ala16 with Lys16 increases the distance between the Glu8 side chain and the right side of the Trp12 indole ring (as drawn in Figure S81), possibly to compensate for the localized overwinding of the α -helical backbone between Trp12 and Lys16.

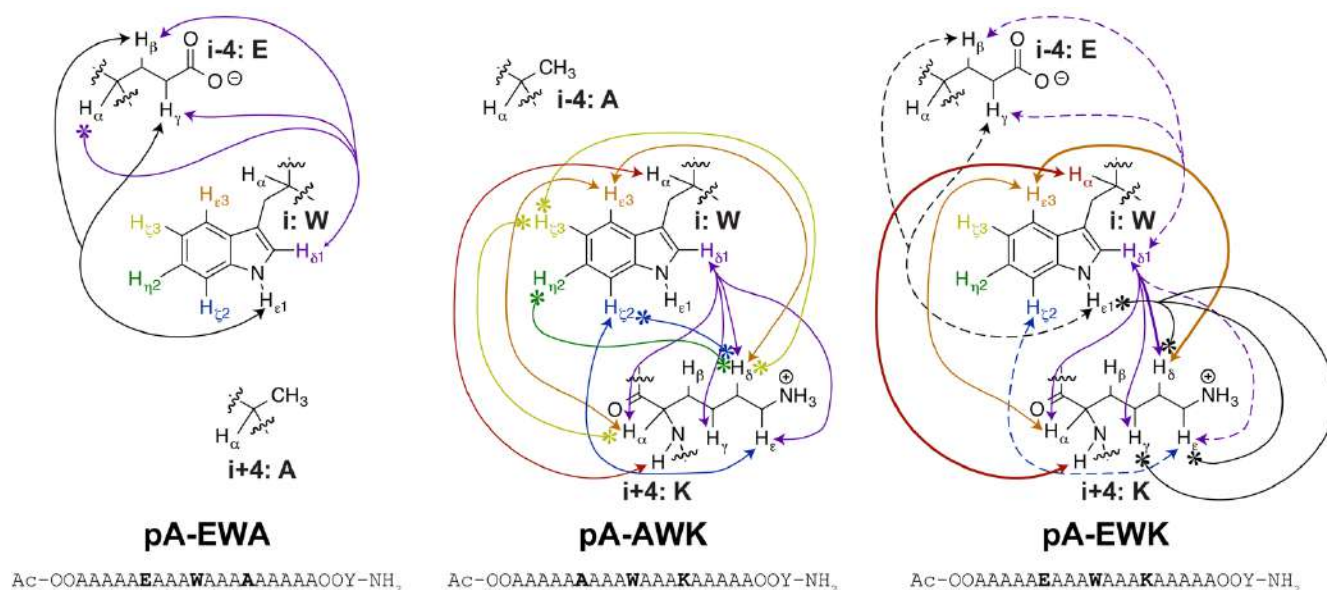


Figure 2-288. Observed NOEs for monomeric α -helical peptides **pA-EWK**, **pA-AWK**, and **pA-EWA** in 20 mM sodium phosphate buffer (pH 7). Lines tipped with asterisks indicate NOEs that are only present in the indicated peptide. Dotted lines indicate NOEs that are weaker in **pA-EWK** than in **pA-EWA** or **pA-AWK**. Bold lines indicate NOEs that are stronger in **pA-EWK** than in **pA-EWA** or **pA-AWK** (see Tables S6-S8).

Table 2-13. TOCSY correlations for pA-AWK.^a

δ (ppm) F1	Assignment F1	δ (ppm) F2	Assignment F2	δ (ppm) F1	Assignment F1	δ (ppm) F2	Assignment F2
8.2	12Trp NH	4.56	12Trp H α	1.67	16Lys H β b	3.98	16Lys H α
8.21	12Trp NH	3.42	12Trp H β *	1.68	16Lys H β b	2.74	16Lys H ϵ a
4.55	12Trp H α	4.55	12Trp H α	1.68	16Lys H β b	2.82	16Lys H ϵ b
3.42	12Trp H β *	4.55	12Trp H α	1.68	16Lys H β b	1.21	16Lys H γ *
7.31	12Trp H δ 1	4.55	12Trp H α	1.49	16Lys H δ *	2.74	16Lys H ϵ a
7.31	12Trp H δ 1	3.42	12Trp H β *	1.49	16Lys H δ *	2.82	16Lys H ϵ b
7.31	12Trp H δ 1	7.31	12Trp H γ 1	1.48	16Lys H δ *	1.21	16Lys H γ *
7.31	12Trp H δ 1	10.2	12Trp H ϵ 1	1.48	16Lys H δ a	3.98	16Lys H α
10.2	12Trp H ϵ 1	7.31	12Trp H δ 1	2.75	16Lys H ϵ a	3.98	16Lys H α
10.2	12Trp H ϵ 1	10.2	12Trp H ϵ 1	2.74	16Lys H ϵ a	3.98	16Lys H α
7.66	12Trp H ϵ 3	7.65	12Trp H ϵ 3	2.75	16Lys H ϵ a	1.6	16Lys H β a
7.66	12Trp H ϵ 3	7.24	12Trp H η 2	2.75	16Lys H ϵ a	1.68	16Lys H β b
7.64	12Trp H ϵ 3	7.24	12Trp H η 2	2.75	16Lys H ϵ a	1.49	16Lys H δ *
7.66	12Trp H ϵ 3	7.5	12Trp H ζ 2	2.75	16Lys H ϵ a	2.75	16Lys H ϵ a
7.64	12Trp H ϵ 3	7.5	12Trp H ζ 2	2.75	16Lys H ϵ a	2.83	16Lys H ϵ b
7.66	12Trp H ϵ 3	7.16	12Trp H ζ 3	2.74	16Lys H ϵ a	2.83	16Lys H ϵ b
7.65	12Trp H ϵ 3	7.16	12Trp H ζ 3	2.75	16Lys H ϵ a	1.21	16Lys H γ *
7.24	12Trp H η 2	7.65	12Trp H ϵ 3	2.82	16Lys H ϵ b	3.98	16Lys H α
7.24	12Trp H η 2	7.24	12Trp H η 2	2.81	16Lys H ϵ b	1.59	16Lys H β a
7.24	12Trp H η 2	7.5	12Trp H ζ 2	2.82	16Lys H ϵ b	1.68	16Lys H β b
7.24	12Trp H η 2	7.16	12Trp H ζ 3	2.82	16Lys H ϵ b	1.49	16Lys H δ *
7.51	12Trp H ζ 2	7.65	12Trp H ϵ 3	2.82	16Lys H ϵ b	2.74	16Lys H ϵ a
7.5	12Trp H ζ 2	7.65	12Trp H ϵ 3	2.82	16Lys H ϵ b	2.82	16Lys H ϵ b
7.49	12Trp H ζ 2	7.24	12Trp H η 2	2.82	16Lys H ϵ b	1.22	16Lys H γ *
7.49	12Trp H ζ 2	7.5	12Trp H ζ 2	1.22	16Lys H γ *	3.98	16Lys H α
7.51	12Trp H ζ 2	7.16	12Trp H ζ 3	1.22	16Lys H γ *	1.59	16Lys H β a
7.49	12Trp H ζ 2	7.16	12Trp H ζ 3	1.21	16Lys H γ *	1.68	16Lys H β b
7.16	12Trp H ζ 3	7.65	12Trp H ϵ 3	1.22	16Lys H γ *	2.74	16Lys H ϵ a
7.16	12Trp H ζ 3	7.24	12Trp H η 2	1.21	16Lys H γ *	2.82	16Lys H ϵ b
7.16	12Trp H ζ 3	7.5	12Trp H ζ 2	1.2	16Lys H γ *	1.2	16Lys H γ *
8.03	16Lys NH	3.98	16Lys H α	2.91	24Tyr H β a	4.55	24Tyr H α
8.03	16Lys NH	1.59	16Lys H β a	2.95	24Tyr H β a	4.55	24Tyr H α
8.03	16Lys NH	1.68	16Lys H β b	2.93	24Tyr H β a	4.55	24Tyr H α
8.03	16Lys NH	1.22	16Lys H γ *	3.13	24Tyr H β b	4.54	24Tyr H α
3.98	16Lys H α	3.98	16Lys H α	3.09	24Tyr H β b	4.55	24Tyr H α
3.98	16Lys H α	1.59	16Lys H β a	7.16	24Tyr H δ *	7.16	24Tyr H δ *
3.98	16Lys H α	1.68	16Lys H β b	7.17	24Tyr H δ *	6.85	24Tyr H ϵ *
3.98	16Lys H α	1.49	16Lys H δ *	7.16	24Tyr H δ *	6.84	24Tyr H ϵ *
3.98	16Lys H α	2.74	16Lys H ϵ a	6.85	24Tyr H ϵ *	7.17	24Tyr H δ *
3.98	16Lys H α	2.82	16Lys H ϵ b	6.84	24Tyr H ϵ *	7.17	24Tyr H δ *
3.98	16Lys H α	1.2	16Lys H γ *	6.85	24Tyr H ϵ *	6.85	24Tyr H ϵ *
1.6	16Lys H β a	3.98	16Lys H α	6.84	24Tyr H ϵ *	6.84	24Tyr H ϵ *
1.6	16Lys H β a	2.74	16Lys H ϵ a				
1.6	16Lys H β a	2.82	16Lys H ϵ b				
1.6	16Lys H β a	1.21	16Lys H γ *				

^aNOE assignments made based on 2D ¹H NMR TOCSY and ROESY experiments on a 5 mM solution of pA-AWK in 20 mM sodium phosphate buffer (pH 7) at room temperature using a 500 MHz Varian INOVA spectrometer.

Table 2-14. TOCSY correlations for pA-EWA.^a

δ (ppm) F1	Assignment F1	δ (ppm) F2	Assignment F2	δ (ppm) F1	Assignment F1	δ (ppm) F2	Assignment F2
8.18	8Glu NH	8.18	8Glu NH	7.16	24Tyr H δ *	6.84	24Tyr H ϵ *
8.18	8Glu NH	4.11	8Glu H α	6.85	24Tyr H ϵ *	7.17	24Tyr H δ *
8.18	8Glu NH	2.17	8Glu H β a	6.84	24Tyr H ϵ *	7.17	24Tyr H δ *
8.18	8Glu NH	2.26	8Glu H β b	6.85	24Tyr H ϵ *	6.84	24Tyr H ϵ *
8.18	8Glu NH	1.98	8Glu H γ *				
4.11	8Glu H α	8.18	8Glu NH				
2.17	8Glu H β a	8.18	8Glu NH				
2.26	8Glu H β b	8.18	8Glu NH				
1.99	8Glu H γ b	8.18	8Glu NH				
8.22	12Trp NH	4.51	12Trp H α				
8.22	12Trp NH	3.41	12Trp H β *				
4.5	12Trp H α	3.41	12Trp H β a				
7.28	12Trp H δ 1	3.41	12Trp H β a				
7.28	12Trp H δ 1	7.28	12Trp H δ 1				
7.28	12Trp H δ 1	10.18	12Trp H ϵ 1				
10.18	12Trp H ϵ 1	7.28	12Trp H δ 1				
10.18	12Trp H ϵ 1	10.18	12Trp H ϵ 1				
7.62	12Trp H ϵ 3	7.61	12Trp H ϵ 3				
7.62	12Trp H ϵ 3	7.21	12Trp H η 2				
7.62	12Trp H ϵ 3	7.48	12Trp H ζ 2				
7.62	12Trp H ϵ 3	7.14	12Trp H ζ 3				
7.21	12Trp H η 2	7.61	12Trp H ϵ 3				
7.21	12Trp H η 2	7.21	12Trp H η 2				
7.21	12Trp H η 2	7.48	12Trp H ζ 2				
7.21	12Trp H η 2	7.14	12Trp H ζ 3				
7.48	12Trp H ζ 2	7.61	12Trp H ϵ 3				
7.47	12Trp H ζ 2	7.61	12Trp H ϵ 3				
7.47	12Trp H ζ 2	7.21	12Trp H η 2				
7.49	12Trp H ζ 2	7.49	12Trp H ζ 2				
7.49	12Trp H ζ 2	7.14	12Trp H ζ 3				
7.14	12Trp H ζ 3	7.61	12Trp H ϵ 3				
7.15	12Trp H ζ 3	7.21	12Trp H η 2				
7.14	12Trp H ζ 3	7.48	12Trp H η 2				
7.14	12Trp H ζ 3	7.14	12Trp H ζ 3				
8.1	16Ala NH	4.11	16Ala H α				
8.1	16Ala NH	1.31	16Ala H β *				
8.08	23Om NH	1.63	23Om H β a				
8.08	23Om NH	1.74	23Om H β b				
8.08	23Om NH	2.96	23Om H δ *				
8.08	23Om NH	1.55	23Om H γ a				
8.08	23Om NH	1.58	23Om H γ b				
8.16	24Tyr NH	4.55	24Tyr H α				
8.16	24Tyr NH	2.94	24Tyr H β a				

8.16	24Tyr NH	3.11	24Tyr H β b
7.16	24Tyr H δ *	7.17	24Tyr H δ *

^aNOE assignments made based on 2D ¹H NMR TOCSY and ROESY experiments on a 5 mM solution of pA-EWA in 20 mM sodium phosphate buffer (pH 7) at room temperature using a 500 MHz Varian INOVA spectrometer.

Table 2-15. TOCSY correlations for pA-EWK.^a

δ (ppm) F1	Assignment F1	δ (ppm) F2	Assignment F2	δ (ppm) F1	Assignment F1	δ (ppm) F2	Assignment F2
8.21	8Glu NH	4.13	8Glu H α	4.02	16Lys H α	4.02	16Lys H α
8.21	8Glu NH	2.19	8Glu H β a	4.01	16Lys H α	1.66	16Lys H β a
8.21	8Glu NH	2.28	8Glu H β b	4.01	16Lys H α	1.72	16Lys H β b
8.21	8Glu NH	2.01	8Glu H γ *	4.02	16Lys H α	1.49	16Lys H δ *
4.12	8Glu H α	2.19	8Glu H β a	4.01	16Lys H α	2.78	16Lys H ϵ a
4.12	8Glu H α	2.28	8Glu H β b	4.01	16Lys H α	2.85	16Lys H ϵ b
4.12	8Glu H α	2.01	8Glu H γ *	4.01	16Lys H α	1.25	16Lys H γ a
8.21	12Trp NH	3.41	12Trp H β *	4.01	16Lys H α	1.29	16Lys H γ b
4.57	12Trp H α	4.57	12Trp H α	1.67	16Lys H β a	2.78	16Lys H ϵ a
4.56	12Trp H α	3.41	12Trp H β *	1.67	16Lys H β a	2.85	16Lys H ϵ b
3.41	12Trp H β *	3.41	12Trp H β *	1.67	16Lys H β a	1.25	16Lys H γ a
7.29	12Trp H δ 1	4.57	12Trp H α	1.67	16Lys H β a	1.3	16Lys H γ b
7.3	12Trp H δ 1	3.41	12Trp H β *	1.73	16Lys H β b	2.78	16Lys H ϵ a
7.29	12Trp H δ 1	7.29	12Trp H δ 1	1.72	16Lys H β b	2.85	16Lys H ϵ b
7.3	12Trp H δ 1	10.19	12Trp H ϵ 1	1.72	16Lys H β b	1.24	16Lys H γ a
10.19	12Trp H ϵ 1	3.41	12Trp H β *	1.72	16Lys H β b	1.3	16Lys H γ b
10.19	12Trp H ϵ 1	7.3	12Trp H δ 1	1.53	16Lys H δ *	2.78	16Lys H ϵ a
10.19	12Trp H ϵ 1	10.19	12Trp H ϵ 1	1.53	16Lys H δ *	2.85	16Lys H ϵ b
10.19	12Trp H ϵ 1	7.64	12Trp H ϵ 3	1.53	16Lys H δ *	1.25	16Lys H γ a
7.64	12Trp H ϵ 3	7.64	12Trp H ϵ 3	1.53	16Lys H δ *	1.3	16Lys H γ b
7.64	12Trp H ϵ 3	7.23	12Trp H η 2	2.78	16Lys H ϵ a	1.53	16Lys H δ *
7.63	12Trp H ϵ 3	7.23	12Trp H η 2	2.79	16Lys H ϵ a	2.79	16Lys H ϵ a
7.64	12Trp H ϵ 3	7.49	12Trp H ζ 2	2.78	16Lys H ϵ a	2.85	16Lys H ϵ b
7.64	12Trp H ϵ 3	7.15	12Trp H ζ 3	2.78	16Lys H ϵ a	1.25	16Lys H γ a
7.63	12Trp H ϵ 3	7.15	12Trp H ζ 3	2.78	16Lys H ϵ a	1.3	16Lys H γ b
7.23	12Trp H η 2	7.64	12Trp H ϵ 3	2.84	16Lys H ϵ b	1.53	16Lys H δ *
7.23	12Trp H η 2	7.23	12Trp H η 2	2.84	16Lys H ϵ b	2.78	16Lys H ϵ a
7.23	12Trp H η 2	7.49	12Trp H ζ 2	2.84	16Lys H ϵ b	2.84	16Lys H ϵ b
7.23	12Trp H η 2	7.15	12Trp H ζ 3	2.84	16Lys H ϵ b	1.25	16Lys H γ a
7.48	12Trp H ζ 2	7.64	12Trp H ϵ 3	2.84	16Lys H ϵ b	1.3	16Lys H γ b
7.5	12Trp H ζ 2	7.64	12Trp H ϵ 3	1.25	16Lys H γ a	2.78	16Lys H ϵ a
7.48	12Trp H ζ 2	7.23	12Trp H η 2	1.3	16Lys H γ b	2.78	16Lys H ϵ a
7.49	12Trp H ζ 2	7.49	12Trp H ζ 2	1.3	16Lys H γ b	2.85	16Lys H ϵ b
7.48	12Trp H ζ 2	7.16	12Trp H ζ 3	8.2	24Tyr NH	2.94	24Tyr H β a
7.15	12Trp H ζ 3	7.64	12Trp H ϵ 3	8.2	24Tyr NH	3.1	24Tyr H β b
7.15	12Trp H ζ 3	7.23	12Trp H η 2	4.55	24Tyr H α	4.55	24Tyr H α
7.15	12Trp H ζ 3	7.49	12Trp H ζ 2	4.55	24Tyr H α	2.94	24Tyr H β a
8.04	16Lys NH	4.01	16Lys H α	4.55	24Tyr H α	3.11	24Tyr H β b
8.04	16Lys NH	1.66	16Lys H β a	7.16	24Tyr H δ *	7.16	24Tyr H δ *
8.04	16Lys NH	1.73	16Lys H β b	7.17	24Tyr H δ *	6.85	24Tyr H ϵ *

8.03	16Lys NH	2.78	16Lys Hεa	7.16	24Tyr Hδ*	6.84	24Tyr Hε*
8.03	16Lys NH	2.85	16Lys Hεb	6.85	24Tyr Hε*	7.17	24Tyr Hδ*
8.04	16Lys NH	1.25	16Lys Hγa	6.84	24Tyr Hε*	7.17	24Tyr Hδ*
8.04	16Lys NH	1.3	16Lys Hγb	6.85	24Tyr Hε*	6.85	24Tyr Hε*
4.01	16Lys Hα	8.04	16Lys H	6.84	24Tyr Hε*	6.84	24Tyr Hε*

^aNOE assignments made based on 2D ¹H NMR TOCSY and ROESY experiments on a 5 mM solution of **pA-EWK** in 20 mM sodium phosphate buffer (pH 7) at room temperature using a 500 MHz Varian INOVA spectrometer.

Table 2-16. NOE assignments for pA-AWK.^a

δ (ppm) F1	Assignment F1	δ (ppm) F2	Assignment F2	Integrated Peak Volume
8.202	12Trp NH	4.558	12Trp H α	45082
8.199	12Trp NH	3.42	12Trp H β *	98442
8.197	12Trp NH	1.496	16Lys H δ a	68964
4.567	12Trp H α	3.418	12Trp H β *	48292
3.419	12Trp H β *	7.65	12Trp H ϵ 3	63723
7.309	12Trp H δ 1	3.987	16Lys H α	27897
7.311	12Trp H δ 1	1.229	16Lys H γ *	20237
7.308	12Trp H δ 1	1.488	16Lys H δ a	57456
7.311	12Trp H δ 1	2.746	16Lys H ϵ a	14343
10.199	12Trp H ϵ 1	7.503	12Trp H ζ 2	28154
7.643	12Trp H ϵ 3	4.557	12Trp H α	66906
7.643	12Trp H ϵ 3	3.42	12Trp H β *	95049
7.659	12Trp H ϵ 3	3.987	16Lys H α	38341
7.643	12Trp H ϵ 3	3.988	16Lys H α	40545
7.659	12Trp H ϵ 3	1.488	16Lys H δ a	32298
7.645	12Trp H ϵ 3	1.485	16Lys H δ a	43349
7.242	12Trp H η 2	1.497	16Lys H δ a	16534
7.498	12Trp H ζ 2	1.496	16Lys H δ a	25264
7.514	12Trp H ζ 2	2.744	16Lys H ϵ a	13413
7.5	12Trp H ζ 2	2.741	16Lys H ϵ a	10851
7.513	12Trp H ζ 2	2.821	16Lys H ϵ b	7922
7.499	12Trp H ζ 2	2.821	16Lys H ϵ b	8683
7.174	12Trp H ζ 3	4.548	12Trp H α	25074
7.162	12Trp H ζ 3	4.548	12Trp H α	119207
7.163	12Trp H ζ 3	3.979	16Lys H α	139294
8.029	16Lys NH	4.568	12Trp H α	12374
8.028	16Lys NH	3.986	16Lys H α	59601
3.986	16Lys H α	7.651	12Trp H ϵ 3	38664
2.955	24Tyr H β a	7.168	24Tyr H δ *	81553
2.935	24Tyr H β a	7.169	24Tyr H δ *	85563
2.911	24Tyr H β a	7.167	24Tyr H δ *	44295
3.097	24Tyr H β b	7.169	24Tyr H δ *	113163
7.174	24Tyr H δ *	2.934	24Tyr H β a	166166
7.163	24Tyr H δ *	2.934	24Tyr H β a	191533
7.174	24Tyr H δ *	3.109	24Tyr H β b	141517
7.163	24Tyr H δ *	3.108	24Tyr H β b	163778

^aNOE assignments made based on 2D ¹H NMR TOCSY and ROESY experiments on a 5 mM solution of pA-AWK in 20 mM sodium phosphate buffer (pH 7) at room temperature using a 500 MHz Varian INOVA spectrometer.

Table 2-17. NOE assignments for **pA-EWA**.^a

δ (ppm) F1	Assignment F1	δ (ppm) F2	Assignment F2	Integrated Peak Volume
8.179	8Glu NH	4.109	8Glu H α	158061
8.179	8Glu NH	8.063	9Ala NH	213495
8.169	8Glu NH	4.245	9Ala H α	336937
4.110	8Glu H α	8.182	8Glu NH	151497
2.258	8Glu H β b	8.185	8Glu NH	38077
1.994	8Glu H γ b	7.272	12Trp H δ 1	10361
8.055	9Ala NH	1.983	8Glu H γ b	135067
8.214	12Trp NH	4.515	12Trp H α	158247
8.213	12Trp NH	3.406	12Trp H β *	283306
8.213	12Trp NH	7.283	12Trp H δ 1	78650
7.281	12Trp H δ 1	8.213	12Trp NH	58320
7.280	12Trp H δ 1	4.513	12Trp H α	275739
7.284	12Trp H δ 1	1.312	16Ala H β *	42052
7.282	12Trp H δ 1	4.127	8Glu H α	108287
7.285	12Trp H δ 1	2.168	8Glu H β a	32106
7.284	12Trp H δ 1	2.26	8Glu H β b	29132
7.284	12Trp H δ 1	1.984	8Glu H γ b	28920
10.175	12Trp H ϵ 1	7.477	12Trp H ζ 2	171082
10.177	12Trp H ϵ 1	2.165	8Glu H β a	13937
10.176	12Trp H ϵ 1	2.261	8Glu H β b	17440
10.177	12Trp H ϵ 1	1.983	8Glu H γ b	12427
7.607	12Trp H ϵ 3	4.516	12Trp H α	106629
7.620	12Trp H ϵ 3	3.405	12Trp H β a	174701
7.608	12Trp H ϵ 3	3.405	12Trp H β a	206571
7.487	12Trp H ϵ 2	10.177	12Trp H ϵ 1	132616
8.096	16Ala NH	3.405	12Trp H β a	125996
8.096	16Ala NH	7.612	12Trp H ϵ 3	83019
1.308	16Ala H β *	8.106	16Ala NH	252348
8.158	24Tyr NH	1.741	23Orn H β b	67715
8.157	24Tyr NH	7.167	24Tyr H δ *	129240
7.173	24Tyr H δ *	2.932	24Tyr H β a	347219
7.163	24Tyr H δ *	2.933	24Tyr H β a	353724
7.173	24Tyr H δ *	3.116	24Tyr H β b	274030
7.163	24Tyr H δ *	3.116	24Tyr H β b	275658

^aNOE assignments made based on 2D ¹H NMR TOCSY and ROESY experiments on a 5 mM solution of **pA-EWA** in 20 mM sodium phosphate buffer (pH 7) at room temperature using a 500 MHz Varian INOVA spectrometer.

Table 2-18. NOEs assignments for **pA-EWK**.^a

δ (ppm) F1	Assignment F1	δ (ppm) F2	Assignment F2	Integrated Peak Volume
8.206	12Trp NH	3.41	12Trp H β *	205808
3.41	12Trp H β *	7.64	12Trp H ϵ 3	81837
7.296	12Trp H δ 1	4.574	12Trp H α	173185
7.298	12Trp H δ 1	4.023	16Lys H α	30352
7.297	12Trp H δ 1	1.479	16Lys H δ *	110836
7.302	12Trp H δ 1	2.792	16Lys H ϵ a	10641
7.302	12Trp H δ 1	2.852	16Lys H ϵ b	10855
7.300	12Trp H δ 1	1.255	16Lys H γ a	16864
7.299	12Trp H δ 1	1.31	16Lys H γ b	16298
7.299	12Trp H δ 1	2.185	8Glu H β a	18624
7.299	12Trp H δ 1	2.284	8Glu H β b	15431
7.299	12Trp H δ 1	2.007	8Glu H γ *	21585
10.193	12Trp H ϵ 1	7.493	12Trp H ζ 2	103103
10.194	12Trp H ϵ 1	1.528	16Lys H δ *	24543
10.194	12Trp H ϵ 1	2.776	16Lys H ϵ a	7219
10.194	12Trp H ϵ 1	1.243	16Lys H γ a	8661
10.194	12Trp H ϵ 1	1.305	16Lys H γ b	7377
10.198	12Trp H ϵ 1	2.182	8Glu H β a	5806
10.195	12Trp H ϵ 1	2.289	8Glu H β b	9212
10.194	12Trp H ϵ 1	2.007	8Glu H γ *	9275
7.646	12Trp H ϵ 3	4.025	16Lys H α	49757
7.634	12Trp H ϵ 3	4.026	16Lys H α	51833
7.646	12Trp H ϵ 3	1.478	16Lys H δ *	91815
7.634	12Trp H ϵ 3	1.478	16Lys H δ *	97478
7.487	12Trp H ζ 2	2.78	16Lys H ϵ a	9140
7.49	12Trp H ζ 2	2.857	16Lys H ϵ b	12797
8.039	16Lys NH	4.579	12Trp H α	43420
2.956	24Tyr H β a	7.169	24Tyr H δ *	99140
2.937	24Tyr H β a	7.17	24Tyr H δ *	102958
3.12	24Tyr H β b	7.17	24Tyr H δ *	66381
3.097	24Tyr H β b	7.17	24Tyr H δ *	134192
7.176	24Tyr H δ *	4.551	24Tyr H α	150166
7.164	24Tyr H δ *	4.551	24Tyr H α	163745
7.175	24Tyr H δ *	2.937	24Tyr H β a	204171
7.164	24Tyr H δ *	2.937	24Tyr H β a	231709
7.175	24Tyr H δ *	3.109	24Tyr H β b	180062
7.164	24Tyr H δ *	3.109	24Tyr H β b	201782

^aNOE assignments made based on 2D ¹H NMR TOCSY and ROESY experiments on a 5 mM solution of **pA-EWK** in 20 mM sodium phosphate buffer (pH 7) at room temperature using a 500 MHz Varian INOVA spectrometer.

2.5 References

1. Donald, J. E.; Kulp, D. W.; DeGrado, W. F., Salt bridges: Geometrically specific, designable interactions. *Proteins-Structure Function and Bioinformatics* **2011**, *79* (3), 898-915.
2. Gong, H. P.; Freed, K. F., Electrostatic Solvation Energy for Two Oppositely Charged Ions in a Solvated Protein System: Salt Bridges Can Stabilize Proteins. *Biophysical Journal* **2010**, *98* (3), 470-477.
3. Salari, R.; Chong, L. T., Desolvation Costs of Salt Bridges across Protein Binding Interfaces: Similarities and Differences between Implicit and Explicit Solvent Models. *Journal of Physical Chemistry Letters* **2010**, *1* (19), 2844-2848.
4. Michael, L. A.; Chenault, J. A.; Miller, B. R., 3rd; Knolhoff, A. M.; Nagan, M. C., Water, shape recognition, salt bridges, and cation-pi interactions differentiate peptide recognition of the HIV rev-responsive element. *J Mol Biol* **2009**, *392* (3), 774-86.
5. Cheng, R. P.; Girinath, P.; Ahmad, R., Effect of lysine side chain length on intra-helical glutamate-lysine ion pairing interactions. *Biochemistry* **2007**, *46* (37), 10528-10537.
6. Anderson, M. A.; Ogbay, B.; Arimoto, R.; Sha, W.; Kisselev, O. G.; Cistola, D. P.; Marshall, G. R., Relative strength of cation-pi vs salt-bridge interactions: the Galpha(340-350) peptide/rhodopsin system. *J Am Chem Soc* **2006**, *128* (23), 7531-41.
7. 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* **2005**, *44* (20), 7553-7558.
8. Bosshard, H. R.; Marti, D. N.; Jelesarov, I., Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings. *Journal of Molecular Recognition* **2004**, *17* (1), 1-16.
9. Pandey, B. K.; Smith, M. S.; Price, J. L., Cys(i)-Lys(i+3)-Lys(i+4) Triad: A General Approach for PEG-Based Stabilization of alpha-Helical Proteins. *Biomacromolecules* **2014**, *15* (12), 4643-4647.
10. Marqusee, S.; Baldwin, R. L., HELIX STABILIZATION BY GLU- ... LYS+ SALT BRIDGES IN SHORT PEPTIDES OF DENOVO DESIGN. *Proceedings of the National Academy of Sciences of the United States of America* **1987**, *84* (24), 8898-8902.
11. Guarracino, D. A.; Chiang, H. J. R.; Banks, T. N.; Lear, J. D.; Hodsdon, M. E.; Schepartz, A., Relationship between salt-bridge identity and 14-helix stability of beta(3)-peptides in aqueous buffer. *Organic Letters* **2006**, *8* (5), 807-810.
12. Arvidsson, P. I.; Rueping, M.; Seebach, D., Design, machine synthesis, and NMR-resolution structure of a beta-heptapeptide forming a salt-bridge stabilised 3(14)-helix in methanol and in water. *Chemical Communications* **2001**, (7), 649-650.
13. Liu, L.; Ousaka, N.; Horie, M.; Mamiya, F.; Yashima, E., Helix-helix inversion of an optically-inactive pi-conjugated foldamer triggered by concentration changes of a single enantiomeric guest leading to a change in the helical stability. *Chem Commun (Camb)* **2016**, *52* (79), 11752-11755.

14. Wimley, W. C.; Gawrisch, K.; Creamer, T. P.; White, S. H., Direct measurement of salt-bridge solvation energies using a peptide model system: implications for protein stability. *Proc Natl Acad Sci U S A* **1996**, *93* (7), 2985-90.
15. Garau, C.; Quinonero, D.; Frontera, A.; Ballester, P.; Costa, A.; Deya, P. M., Anion-pi interactions: must the aromatic ring be electron deficient? *New Journal of Chemistry* **2003**, *27* (2), 211-214.
16. Quinonero, D.; Garau, C.; Frontera, A.; Ballester, P.; Costa, A.; Deya, P. M., Counterintuitive interaction of anions with benzene derivatives. *Chemical Physics Letters* **2002**, *359* (5-6), 486-492.
17. Ngola, S. M.; Kearney, P. C.; Mecozzi, S.; Russell, K.; Dougherty, D. A., A selective receptor for arginine derivatives in aqueous media. Energetic consequences of salt bridges that are highly exposed to water. *J. Am. Chem. Soc.* **1999**, *121* (6), 1192-1201.
18. Chakravarty, S.; Sheng, Z. Z.; Iverson, B.; Moore, B., "eta6"-Type anion-pi in biomolecular recognition. *FEBS Lett* **2012**, *586* (23), 4180-5.
19. Ogiwara, N. L.; Weiss, M. S.; DeGrado, W. F.; Eisenberg, D., The crystal structure of the designed trimeric coiled coil coil-V(a)L(d): Implications for engineering crystals and supramolecular assemblies. *Protein Science* **1997**, *6* (1), 80-88.
20. Shu, J. Y.; Tan, C.; DeGrado, W. F.; Xu, T., New design of helix bundle peptide-polymer conjugates. *Biomacromolecules* **2008**, *9* (8), 2111-7.
21. Smith, M. S.; Billings, W. M.; Whitby, F. G.; Miller, M. B.; Price, J. L., Enhancing a long-range salt bridge with intermediate aromatic and nonpolar amino acids. *Organic & Biomolecular Chemistry* **2017**, *15* (28), 5882-5886.
22. Tsou, L. K.; Tatko, C. D.; Waters, M. L., Simple cation-pi interaction between a phenyl ring and a protonated amine stabilizes an alpha-helix in water. *J. Am. Chem. Soc.* **2002**, *124* (50), 14917-14921.
23. Serrano, L.; Horovitz, A.; Avron, B.; Bycroft, M.; Fersht, A. R., Estimating the contribution of engineered surface electrostatic interactions to protein stability by using double-mutant cycles. *Biochemistry* **1990**, *29* (40), 9343-52.
24. Luisi, D. L.; Snow, C. D.; Lin, J. J.; Hendsch, Z. S.; Tidor, B.; Raleigh, D. P., Surface salt bridges, double-mutant cycles, and protein stability: an experimental and computational analysis of the interaction of the Asp 23 side chain with the N-terminus of the N-terminal domain of the ribosomal protein 19. *Biochemistry* **2003**, *42* (23), 7050-60.
25. Wlodawer, A.; Minor, W.; Dauter, Z.; Jaskolski, M., Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS J* **2008**, *275* (1), 1-21.
26. Nakamura, A.; Ohtsuka, J.; Kashiwagi, T.; Numoto, N.; Hirota, N.; Ode, T.; Okada, H.; Nagata, K.; Kiyohara, M.; Suzuki, E.; Kita, A.; Wada, H.; Tanokura, M., In-situ and real-time growth observation of high-quality protein crystals under quasi-microgravity on earth. *Sci Rep* **2016**, *6*, 22127.
27. Takahashi, S.; Ohta, K.; Furubayashi, N.; Yan, B.; Koga, M.; Wada, Y.; Yamada, M.; Inaka, K.; Tanaka, H.; Miyoshi, H.; Kobayashi, T.; Kamigaichi, S., JAXA protein crystallization in space: ongoing improvements for growing high-quality crystals. *J Synchrotron Radiat* **2013**, *20* (Pt 6), 968-73.
28. Matsumura, H.; Sugiyama, S.; Hirose, M.; Kakinouchi, K.; Maruyama, M.; Murai, R.; Adachi, H.; Takano, K.; Murakami, S.; Mori, Y.; Inoue, T., Approach for growth of high-quality and large protein crystals. *J Synchrotron Radiat* **2011**, *18* (1), 16-9.

29. Moreno, A.; Yokaichiya, F.; Dimasi, E.; Stojanoff, V., Growth and characterization of high-quality protein crystals for X-ray crystallography. *Ann N Y Acad Sci* **2009**, *1161*, 429-36.
30. Burkhard, P.; Stetefeld, J.; Strelkov, S. V., Coiled coils: a highly versatile protein folding motif. *Trends Cell Biol.* **2001**, *11* (2), 82-88.
31. 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* **1993**, *262* (5138), 1401-1407.
32. Shi, Z. S.; Olson, C. A.; Kallenbach, N. R., Cation-pi interaction in model alpha-helical peptides. *J. Am. Chem. Soc.* **2002**, *124* (13), 3284-3291.
33. Shi, Z.; Olson, C. A.; Bell, A. J., Jr.; Kallenbach, N. R., Stabilization of alpha-helix structure by polar side-chain interactions: complex salt bridges, cation-pi interactions, and C-H...O H-bonds. *Biopolymers* **2001**, *60* (5), 366-80.
34. Wishart, D. S.; Bigam, C. G.; Holm, A.; Hodges, R. S.; Sykes, B. D., ¹H, ¹³C and ¹⁵N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J Biomol NMR* **1995**, *5* (1), 67-81.
35. Ciani, B.; Bjelic, S.; Honnappa, S.; Jawhari, H.; Jaussi, R.; Payapilly, A.; Jowitt, T.; Steinmetz, M. O.; Kammerer, R. A., Molecular basis of coiled-coil oligomerization-state specificity. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107* (46), 19850-19855.
36. Spector, S.; Young, P.; Raleigh, D. P., Nativelike structure and stability in a truncation mutant of a protein minidomain: The peripheral subunit-binding domain. *Biochemistry* **1999**, *38* (13), 4128-4136.
37. Woolfson, D. N.; Alber, T., Predicting oligomerization states of coiled coils. *Protein Sci* **1995**, *4* (8), 1596-607.
38. Edelhoch, H., Spectroscopic Determination of Tryptophan and Tyrosine in Proteins*. *Biochemistry (Mosc)*. **1967**, *6* (7), 1948-1954.
39. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallographica Section D-Biological Crystallography* **2010**, *66*, 486-501.
40. Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D-Biological Crystallography* **2010**, *66*, 213-221.
41. Shi, Z. S.; Olson, C. A.; Kallenbach, N. R., Cation-pi interaction in model alpha-helical peptides. *J. Am. Chem. Soc.* **2002**, *124* (13), 3284-3291.

3 MEASURING A HELICAL LONG-RANGE SALT BRIDGE ENHANCED BY NONAROMATIC HYDROPHOBIC AMINO ACIDS.

3.1 Introduction

We demonstrate in Chapter 2 that an arene side chain can enhance the interaction between a Glu and a Lys residue positioned 8 residues apart on the solvent exposed face of a coiled coil. Several computational studies predicted that positioning a positive ion on one face of an aromatic ring and negative ion on the opposing face would have a favorable cooperative interaction, which we also observe in chapter 2.¹⁻⁵ These studies describe the aromatic ring as an “insulator” or a medium that stores information between the charges, in other words the electronics found in the arene ring can be polarized towards the positive ion and away from the negative creating a synergistic three-way interaction. Interestingly, more recent studies predict a similar interaction with hydrocarbons where a cyclic hydrocarbon can mediate a salt bridge by polarizing its electron density (Figure 3-1).

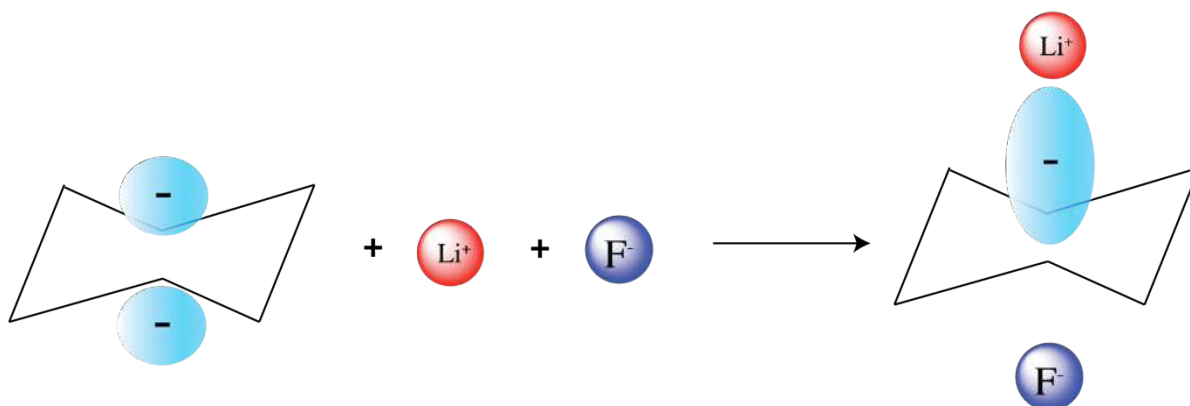


Figure 3-1. Three-way interaction between a cyclohexane, fluorine ion, and a lithium ion. The mechanism is predicted to be induction where the electron density of the ring is polarized towards the lithium cation and away from the fluorine anion.

Elguero et al. was interested in opposing ions interacting with hydrocarbons in order to better understand and design energy storing devices.⁶ Interestingly, they predict that positioning a cation and anion on opposing faces of cyclohexane can promote a cooperative three-way interaction. They model the electron density of a fluorine anion and a lithium cation on opposing faces of a cyclohexane and show that the electron density in the ring polarizes towards the lithium ion and away from the fluorine ion (figure 3-1). Their model predicts a three-way interaction worth about $5.5 \text{ kcal mol}^{-1}$ in cooperative energy between the ions and cyclohexane. An analogous study confirms gas phase interaction between a cesium cation and chlorine anion positioned on opposing faces of a cyclobutane.⁷ Also, a spherical fullene C_{60} crystal structure where a Li^+ cation is trapped within the fullerene shows that there is potential for SbCl_3^- counter ion interaction on the exterior of the sphere.⁸ Due to these studies we wondered if a cyclic nonpolar amino acid could enhance a long range salt bridge between a Glu 10: Lys 18 on the coiled-coil trimeric helix 1CW (Figure 3-2).

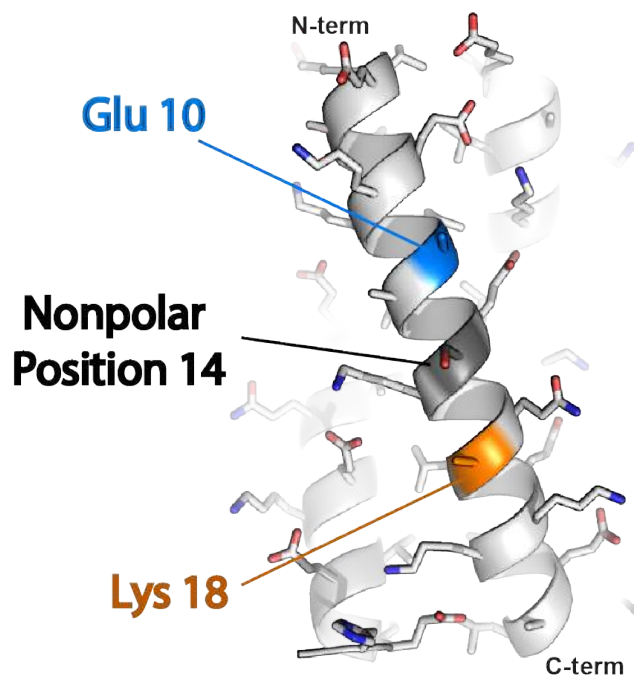


Figure 3-2. Coiled-coil trimeric 1CW with positions 10, 14, and 18 highlighted as the anion, nonpolar, and cation mutation sites.

3.2 Results and discussion

3.2.1 Long range salt-bridge enhanced by nonpolar amino acids.

To measure the extent that a cyclic nonpolar amino acid can enhance a Glu10:Lys18 salt bridge we synthesized the 1CW variant EChaK where position 14 is occupied by a cyclohexylalanine (Cha), position 10 is a Glutamate, and position 18 is a Lysine (Figure 3-2). Circular dichroism and size exclusion chromatography confirm that E(Cha)K along with its sequence variants are coiled-coil trimers (supporting information 3.3.4, 3.3.5). Triple mutant cycle analysis shows that Cha14 enhances the Glu10:Lys18 salt bridge similar to Phe14. This indicates the aromaticity is not required for the three-way interaction between Glu10:Lys18. To confirm this, we applied the Cha long-range salt bridge enhancement to GCN-p1 dimeric model system by

synthesizing EChaKp1 where position 11, 14, and 18 are occupied by Glu, Cha, Lys respectively. In GCN4-p1 Cha14 also enhances the Glu11:Lys18 salt bridge similar to the extent of Phe14 (Table 3-1).

Table 3-1 Cha enhancement of long range salt bridge.^a

Peptide	Sequence	Salt-bridge with Ser14 $\Delta\Delta\Delta G_r$ (kcal/mol)	Salt-bridge with Phe14 or Cha14 $\Delta\Delta\Delta G_r$ (kcal/mol)	Influence of Phe14 or Cha14 $\Delta\Delta\Delta G_r$ (kcal/mol)
EFK	Ac-•••KV E ALEFKVQ K LE•••CONH ₂	0.87 ± 0.06	-0.38 ± 0.07	-1.25 ± 0.09
EChaK	Ac-•••KV E ALE(Cha)KVQ K LE•••CONH ₂	0.87 ± 0.06	-0.47 ± 0.07	-1.34 ± 0.13
EFKp1	Ac-•••VA E LEFKNY K L•••-CONH ₂	1.20 ± 0.07	-0.38 ± 0.07	-1.58 ± 0.10
EChaKp1	Ac-•••VA E LE(Cha)KNY K L•••-CONH ₂	1.20 ± 0.07	-0.09 ± 0.06	-1.30 ± 0.09

^aData are given ± standard error at 30 μM protein concentration in 20 mM sodium phosphate buffer (pH 7). Full sequences can be reported in the supporting information.

We then tested if the size and shape of the hydrophobic residue influenced the interaction. To explore this possibility, we incorporated non-polar Leu at position 14 (Figure 3-3) and assessed its ability to enhance the Glu10-Lys18 salt bridge relative to non-polar Ala. Some of these variants were so stable in 20 mM sodium phosphate buffer (pH 7) as to preclude direct assessment of their folding free energies. To facilitate direct comparison among peptides ELK, and EFK we performed variable temperature CD experiments on ELK, EFK, and their sequence variants in 20 mM sodium phosphate buffer (pH 7) with 1 M urea. The impact of Leu is similar to that of Phe or of Cha, suggesting that the residual Phe-based increase in salt-bridge stability does not require additional atoms beyond the Cδ's of Leu. We also found that norleucine (Nle), norvaline (Nva), Val, and 2-amino-n-butyric acid (Abu) at position 14 are each superior to Ser and Ala, and are slightly better than Phe and Leu in their ability to stabilize the Glu10-Lys18 salt-bridge (Fig 3-3). Together, the observations in Figure 3-1 and chapter 2 suggest that the impact of Phe the Glu10-Lys18 salt-bridge depends only on its non-polar character and not on its shape, size, or surface electrostatic

potential. We also observe that proteinogenic side chains like Leucine and Valine can enhance a long-range salt bridge in the context of a coiled-coil trimeric alpha helix.







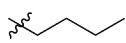
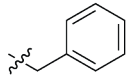
Peptide	Residue 14	Salt-bridge Interaction $\Delta\Delta\Delta G_f$ (kcal/mol)	Influence of Residue 14 $\Delta\Delta\Delta\Delta G_f$ (kcal/mol)
EAK	 CH ₃	0.87 ± 0.11	
ESK		0.20 ± 0.04	-0.67 ± 0.12
EAbuK		-0.52 ± 0.04	-1.39 ± 0.12
EVK		-0.55 ± 0.03	-1.42 ± 0.12
ENvaK		-0.52 ± 0.04	-1.39 ± 0.12
ELK		-0.30 ± 0.06	-1.17 ± 0.13
ENleK		-0.55 ± 0.04	-1.42 ± 0.12
EFK		-0.13 ± 0.03	-1.00 ± 0.12

Figure 3-3. Impact of differing hydrophobic acyclic amino acids on the Glu10:Lys18 salt bridge. Phe 14 is included for reference. Data are given ± standard error at 30 μM protein concentration in 20 mM sodium phosphate buffer (pH 7), and 1M urea. Full sequences along with triple mutant cycle sequence variants are reported in the supporting information.

3.2.2 Crystal structures

We explored the differences in the side chain rotamers between compounds EFK, ELK, ENleK, EVK, ENvaK, and EAbuK (figure 3-4A). To explore this, we measured dihedral angles χ_1 (defined by the atoms N α , C α , C β , and C γ) and χ_2 (C α , C β , C γ , and C δ) for each side-chain, and described each χ_1 and χ_2 in one of three possible ways: g⁺ (positive gauche) for angles between 0° and 120°; g⁻ (negative gauche) for angles between 0° and -120°; and t (trans) for angles

between 120° and 180° or between -120° and -180° . Glu10 occupies the same geometry in variants EFK, ELK, EVK, and ENvaK: while ENleK, and EabuK differ (figure 3-4B). The geometry of the residue 14 position appears to occupy differing rotamers, only Lue14 and Phe14 occupy the t, g⁺ rotamer while the ENleK, EVK, ENvaK, and EAbuK occupy varying rotamers (figure 3-4C). Like residue 14, Lys18 in variants ELK and the EFK variant both occupy the same rotamer (g-, t), while the ENleK, EVK, ENvaK, and EAbuK occupy varying rotamers (figure 3-4D). The caveats described in Chapter 2 for interpreting the rotamers of surface-exposed side-chains in the crystalline state also apply here, making it difficult to correlate the structural details described in Fig. 3-4 with the solution-phase thermodynamic observations shown in Figure 3-3. The Glu10 of all the variants except for EAbuK, and ENleK occupy the g-, g- rotamer. Interestingly, if the crystal packing is taken into account the Glu10 g- g- rotamer of one helix trimer actually packs very close to the Lue14 of an adjacent trimer (Figure 3-5E). Although, this interaction is not applicable to in solution geometry it is tempting to speculate that the crystal packing interface places the Glu10 of an adjacent trimer at a more optimal position to interact in a three-way interaction with Lue14, and Lys18.

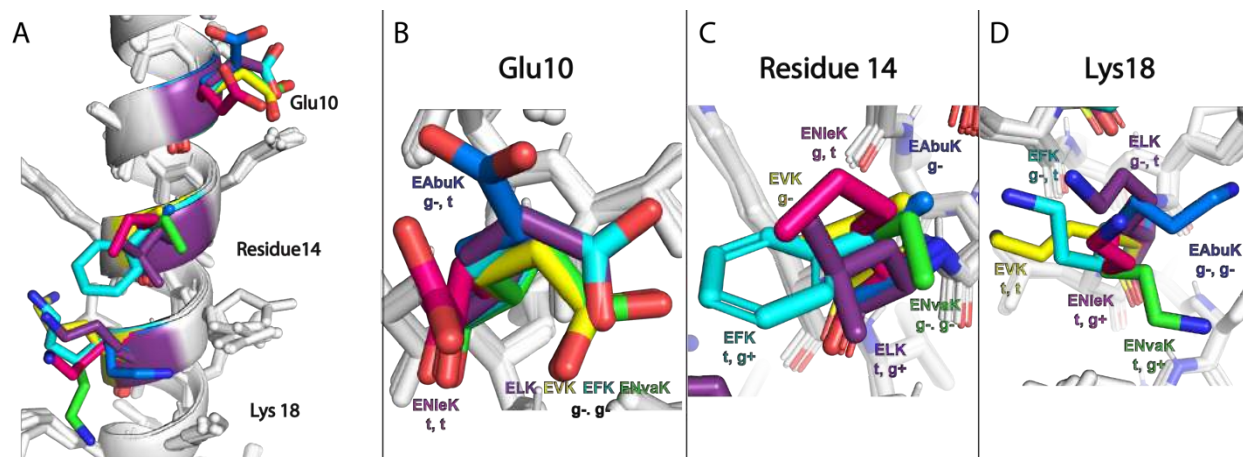


Figure 3-4. (A) overlay of a single helix from EFK, ELK, ENleK, EVK, ENvaK, and EAbuK. (B) Rotamer analysis of Glu10. (C) Rotamer analysis of residue 14. (D) Rotamer analysis of Lys18.

We explored the structural basis for the stabilizing impact of Lue14 on the Glu10:Lys18 salt bridge by crystalizing four variants of peptide ELK, in which Leu vs. Ala is at position 14; Glu vs. Ala is at position 10; and Lys vs. Ala is at position 18, in all possible combinations (i.e., peptides ELK, ELA, ALK, ALA, EAK). These structures are overlaid with the structure of the EFK peptide in (Figure 3-5A). Interestingly, the Glu10 rotamer changes from t, g⁺ in ELA to g⁻, g⁻ in ELK (Figure 3-5B). This rearrangement packs the Glu10 β hydrogens towards the center of Lue14 on an adjacent helix trimer (Figure 3-5 E-F). The Leu14 side chains in peptides ELA, ALK, and ALA adopt the g⁻,t conformation that is most frequently observed for Leu residues within proteins;⁹⁻¹¹ this conformation projects the δ -methyl groups of Leu14 toward the N-terminus of the helix where they are similarly distant from positions 10 and 18. In contrast, Leu14 in ELK adopts the less common t,g⁺ conformation. Interestingly, the t,g⁺ conformation of Leu14 closely resembles that of Phe14 in EFK(Figure 3-5C). The Lysine rotamers in ALK and EAK occupy the g⁻, g⁻ rotamers. In the variants EFK and ELK the Lysine rotamer switches to a g⁻, t rotamer (figure 3-5 D). This switch in rotamer places the Leu14, or Phe14 in close proximity to Lys18: the Leu14 C γ is only 3.9 Å from the Lys18 C δ . We speculate that this

change in distance is due to the close packing of the crystal contact Glu10 to both Phe14, and Leu14 (Figure 5E, G, I). However, the caveats described above for interpreting the rotamers of surface-exposed side-chains in the crystalline state also apply here, making it difficult to correlate the structural details described in Fig. 3-5 with the solution-phase thermodynamic observations shown in Fig. 3-3.

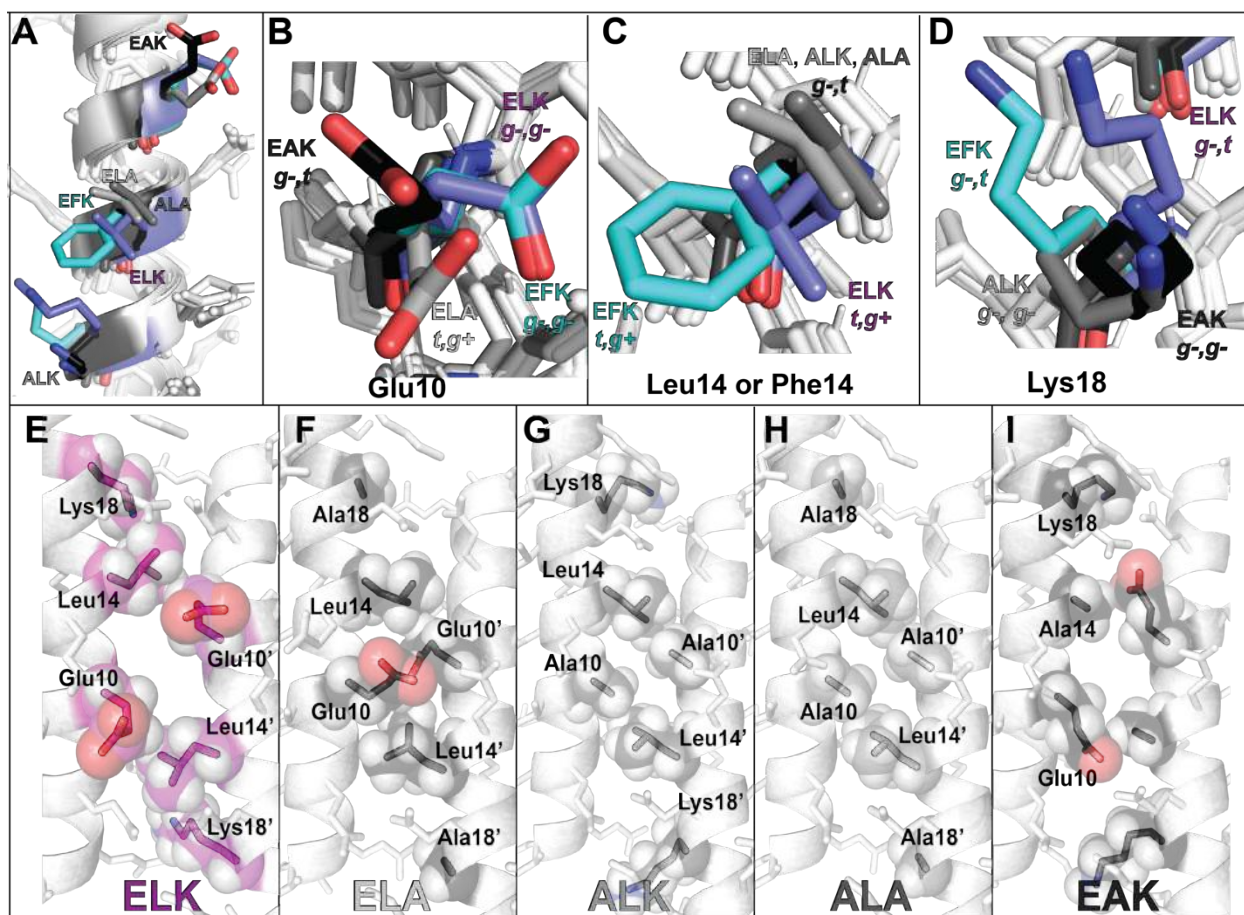


Figure 3-5. (A) Single helix overlay of EFK, ELK, EAK, ALK, ALA, and ELA. (B) Rotamer analysis of Glu10. (C) Rotamer analysis of Leucine14 or Phe 14. (D) Rotamer analysis for Lys18. (E) crystal interface between 2 trimers for ELK, (F) ELA, (G) ALK, (H) ALA, and (I) EAK.

3.2.3 Conclusions

In this chapter, we show that long range salt bridge can contribute significant conformational stability to a coiled-coil alpha helix trimer and dimer if there is a nonpolar amino acid positioned between the cation and anion. We show that the shape size of the amino acid does not matter as long as it is bigger than alanine. The data indicate that both leucine and valine can enhance long range salt bridges in the context of an alpha helix, this suggests a broader impact for cooperative ion-nonpolar interactions in protein structure. Substantial structural data show that the Glu10 in one trimer in the crystal lattice positions itself in close proximity to the leucine 14 on adjacent trimer. Close packing of lysine 18 and leucine 14 is also observed in the presence of Glu10. This is interesting and causes us to speculate that a nonpolar enhanced salt bridge might be more favorable if the Lue enhances salt bridge is positioned at the helical-helical interface.

3.3 Supporting Information

3.3.1 Protein synthesis, characterization, and Purification.

Peptide sequences for **EChak**, **EChakp1**, **ELK**, **ENleK**, **EVK**, **ENvaK**, and **EabuK** along with their sequence variants are shown in table 3-2. **EAK**, **ESK**, and **EFK** data can be found in chapter 2. All peptides were synthesized as C-terminal amides, by microwave-assisted solid-phase peptide synthesis as described in Chapter 2 supporting information section 2.4.1.

Table 3-2 Peptide sequences

Peptide	Sequence
1CW	Ac-EVEALEKKV A ALE C KV Q ALEKKVEALEHGWDGR-CONH ₂
GCN4-P1	Ac-RMKQLEDRVE E LE S KNY H LENEVARLKKLVGER-CONH ₂

EChaK	Ac-EVEALEKKV E ALE Cha KVQ K LEKKVEALEHGWDGR-CONH ₂
EChaA	Ac-EVEALEKKV E ALE Cha KVQ A LEKKVEALEHGWDGR-CONH ₂
AChaK	Ac-EVEALEKKV A ALE Cha KVQ K LEKKVEALEHGWDGR-CONH ₂
AChaA	Ac-EVEALEKKV A ALE Cha KVQ A LEKKVEALEHGWDGR-CONH ₂
EChaK p1	Ac-RMKQLEDRVA E LE Cha KNY K LENEVARLKKLVGER-CONH ₂
EChaA p1	Ac-RMKQLEDRVA E LE Cha KNY A LENEVARLKKLVGER-CONH ₂
AChaK p1	Ac-RMKQLEDRVA E LE Cha KNY K LENEVARLKKLVGER-CONH ₂
AChaA p1	Ac-RMKQLEDRVA E LE Cha KNY A LENEVARLKKLVGER-CONH ₂
ELK	Ac-EVEALEKKVA E LE LK VQ K LEKKVEALEHGWDGR-CONH ₂
ELA	Ac-EVEALEKKVA E LE LK VQ A LEKKVEALEHGWDGR-CONH ₂
ALK	Ac-EVEALEKKVA E LE LK VQ K LEKKVEALEHGWDGR-CONH ₂
ALA	Ac-EVEALEKKVA E LE LK VQ A LEKKVEALEHGWDGR-CONH ₂
ENleK	Ac-EVEALEKKV E ALE Nle KVQ K LEKKVEALEHGWDGR-CONH ₂
ENleA	Ac-EVEALEKKV A ALE Nle KVQ K LEKKVEALEHGWDGR-CONH ₂
ANleK	Ac-EVEALEKKV E ALE Nle KVQ A LEKKVEALEHGWDGR-CONH ₂
ANleA	Ac-EVEALEKKV A ALE Nle KVQ A LEKKVEALEHGWDGR-CONH ₂
EVK	Ac-EVEALEKKV E ALE VK VQ K LEKKVEALEHGWDGR-CONH ₂
EVA	Ac-EVEALEKKV E ALE VK VQ A LEKKVEALEHGWDGR-CONH ₂
AVK	Ac-EVEALEKKV A ALE VK VQ K LEKKVEALEHGWDGR-CONH ₂
AVA	Ac-EVEALEKKV A ALE VK VQ A LEKKVEALEHGWDGR-CONH ₂
ENvaK	Ac-EVEALEKKV E ALE Nva KVQ K LEKKVEALEHGWDGR-CONH ₂
ENvaA	Ac-EVEALEKKV E ALE Nva KVQ A LEKKVEALEHGWDGR-CONH ₂
ANvaK	Ac-EVEALEKKV A ALE Nva KVQ K LEKKVEALEHGWDGR-CONH ₂
ANvaA	Ac-EVEALEKKV A ALE Nva KVQ A LEKKVEALEHGWDGR-CONH ₂
EAbuK	Ac-EVEALEKKV E ALE Abu KVQ K LEKKVEALEHGWDGR-CONH ₂
EAbuA	Ac-EVEALEKKV E ALE Abu KVQ A LEKKVEALEHGWDGR-CONH ₂
AAbuK	Ac-EVEALEKKV A ALE Abu KVQ K LEKKVEALEHGWDGR-CONH ₂
AAbuA	Ac-EVEALEKKV A ALE Abu KVQ A LEKKVEALEHGWDGR-CONH ₂

3.3.2 ESI-TOF mass spec data

Mass spec characterization for **EChak**, **EChaKp1**, **ELK**, **ENleK**, **EVK**, **ENvaK**, and **EabuK** along with their sequence variants are shown below.

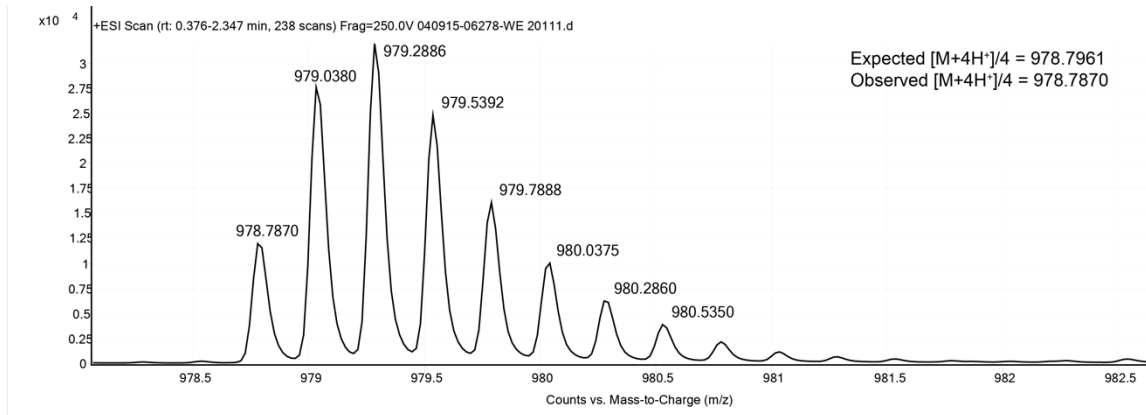


Figure 3-6. ESI TOF spectrum for peptide **EChaK**

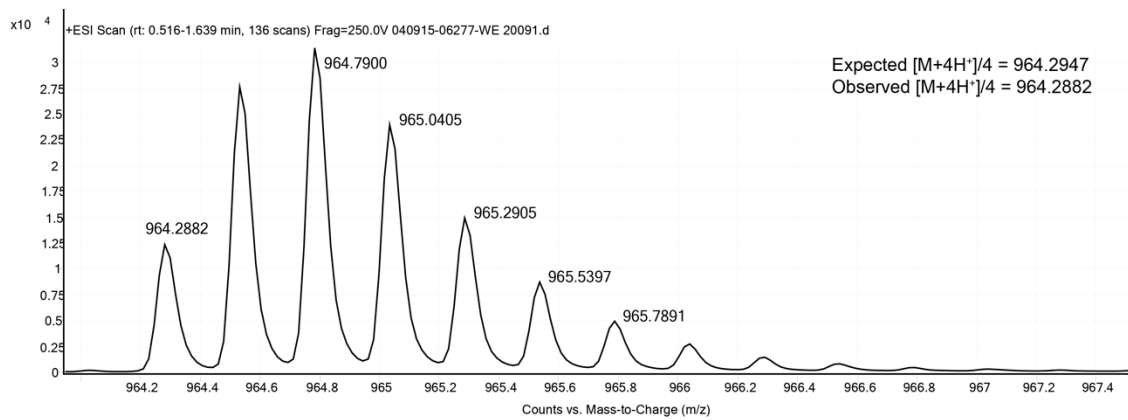


Figure 3-7. ESI TOF spectrum for peptide **ACHaK**

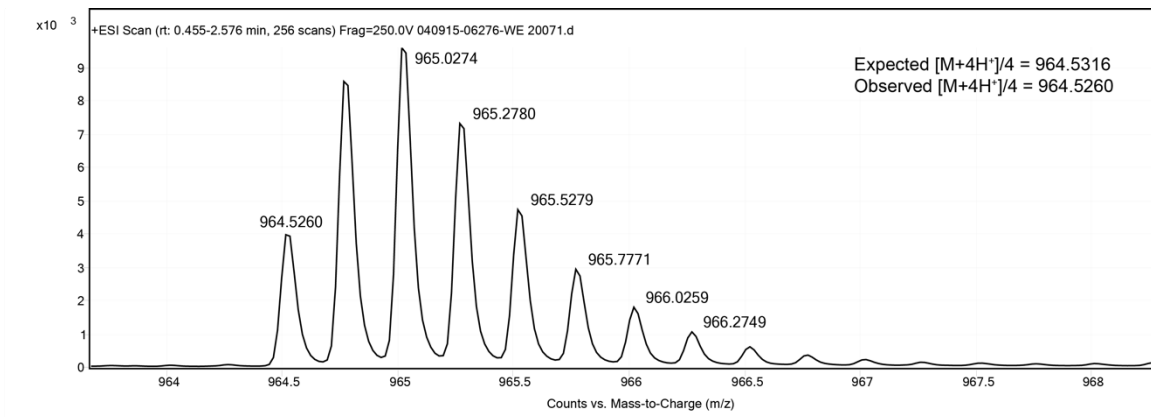


Figure 3-8. ESI TOF spectrum for peptide **EChaA**

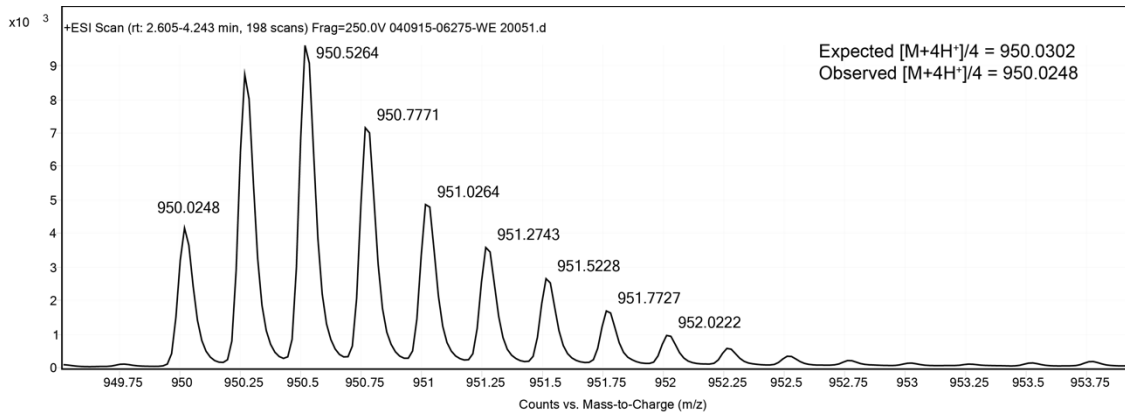


Figure 3-9. ESI TOF spectrum for peptide **ACHaA**

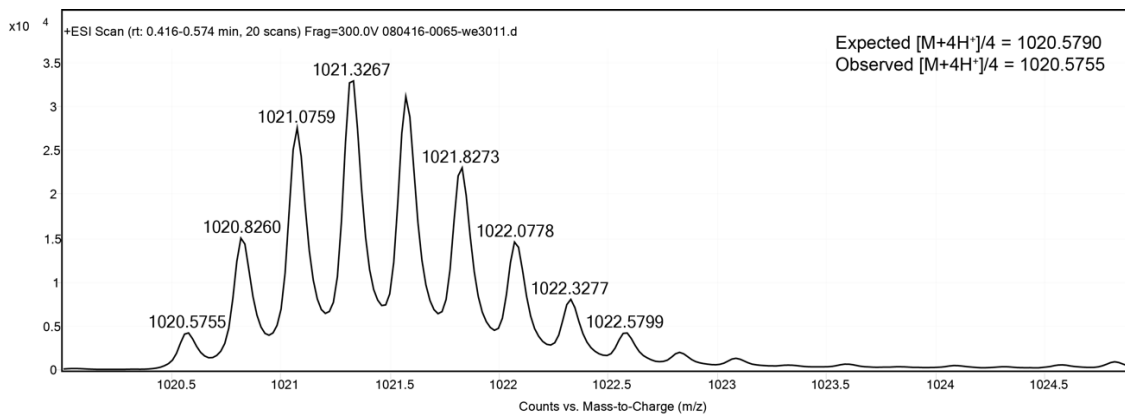


Figure 3-10. ESI TOF data for peptide **EchaK p1**

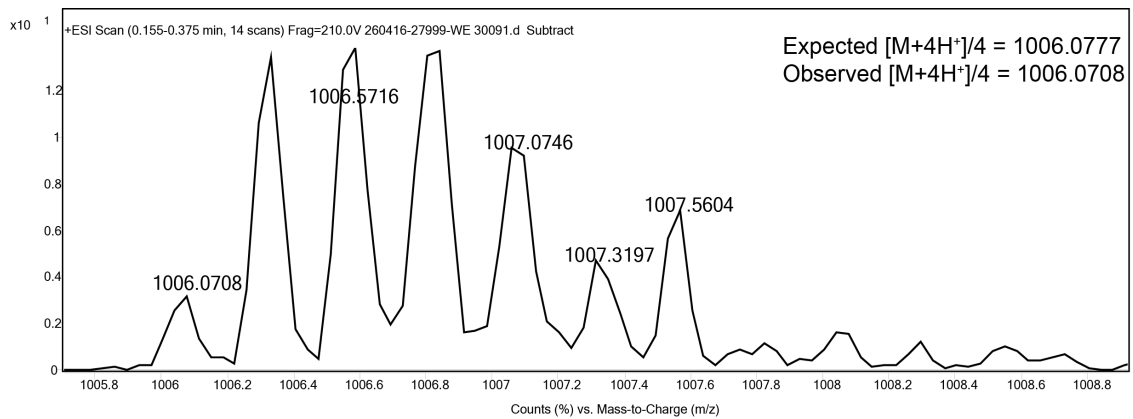


Figure 3-11. ESI TOF data for peptide **AchaK p1**

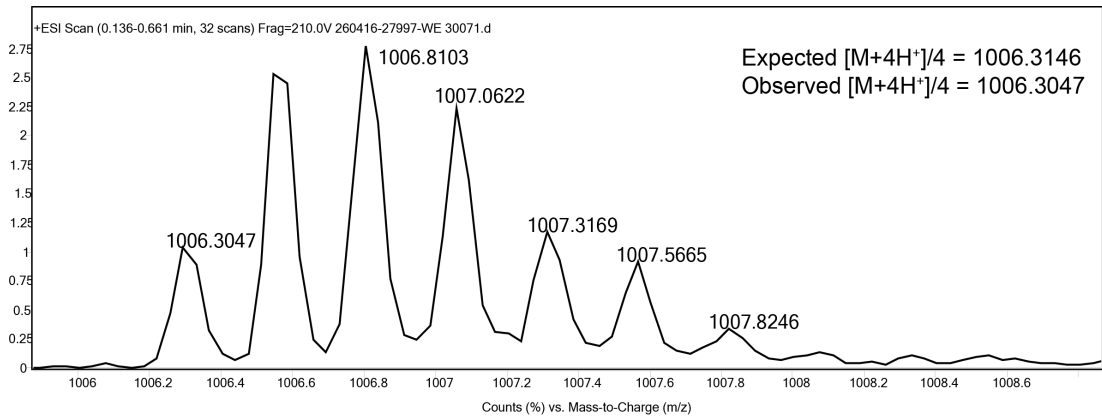


Figure 3-12. ESI TOF data for peptide **EchaA p1**.

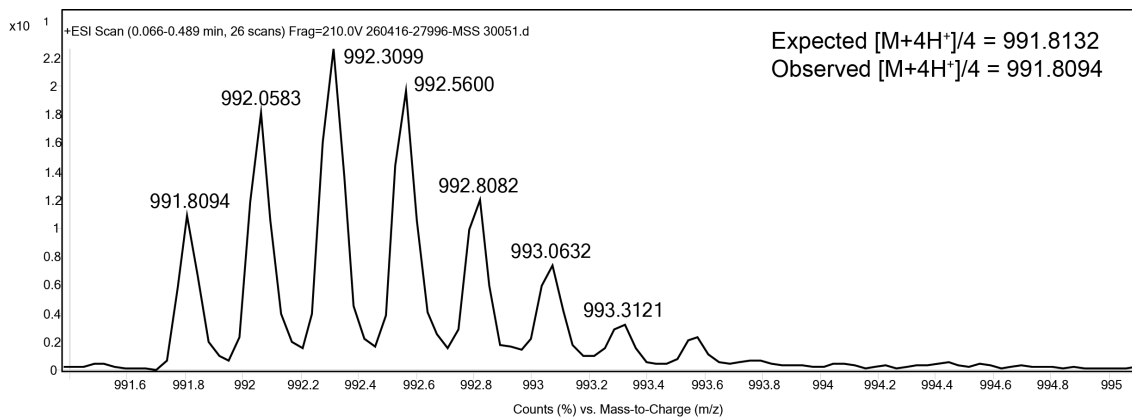


Figure 3-13. ESI TOF spectrum for peptide **AchaA p1**.

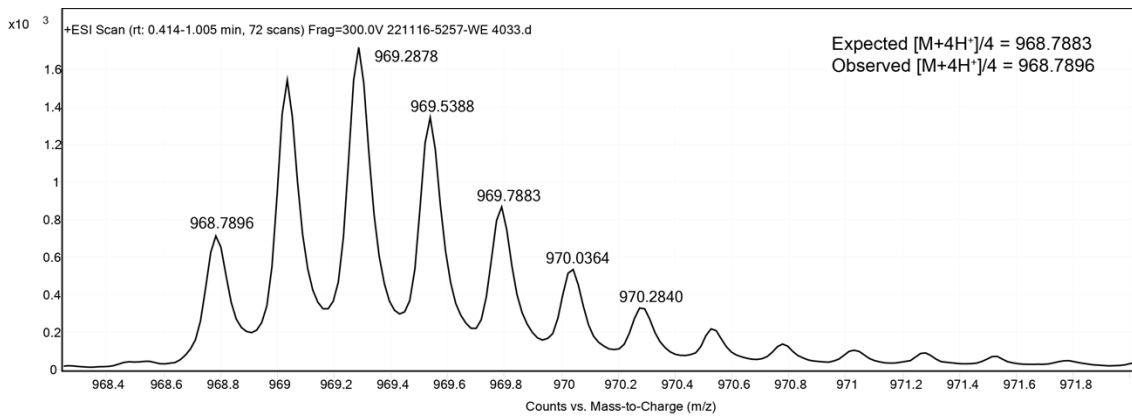


Figure 3-14. ESI TOF spectrum for peptide **ELK**.

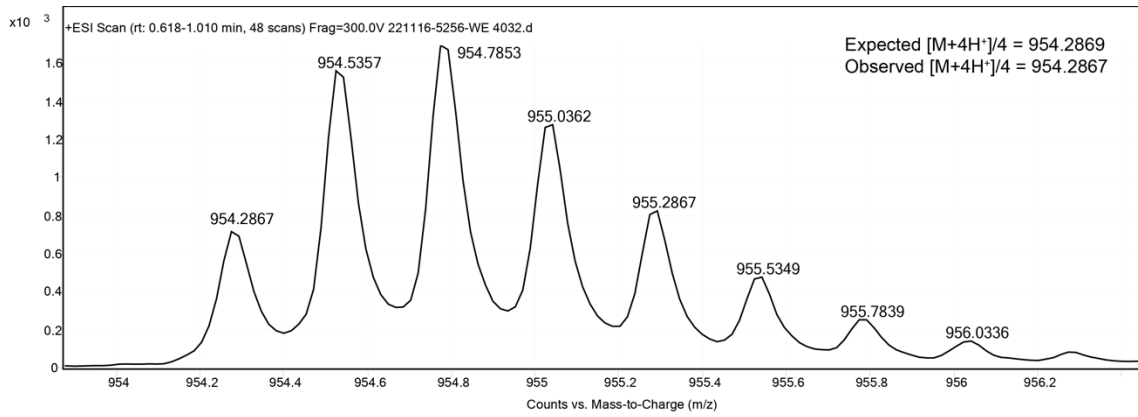


Figure 3-15. ESI TOF spectrum for peptide **ALK**.

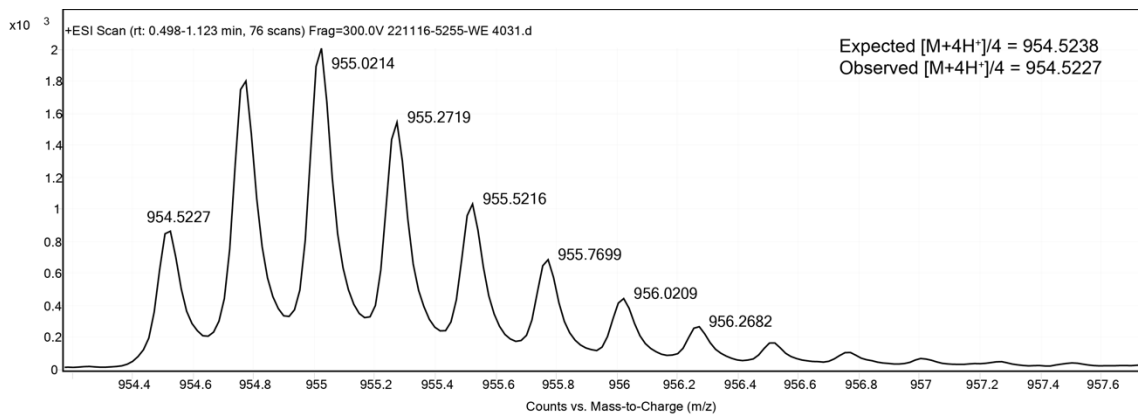


Figure 3-16. ESI TOF spectrum for peptide **ELA**.

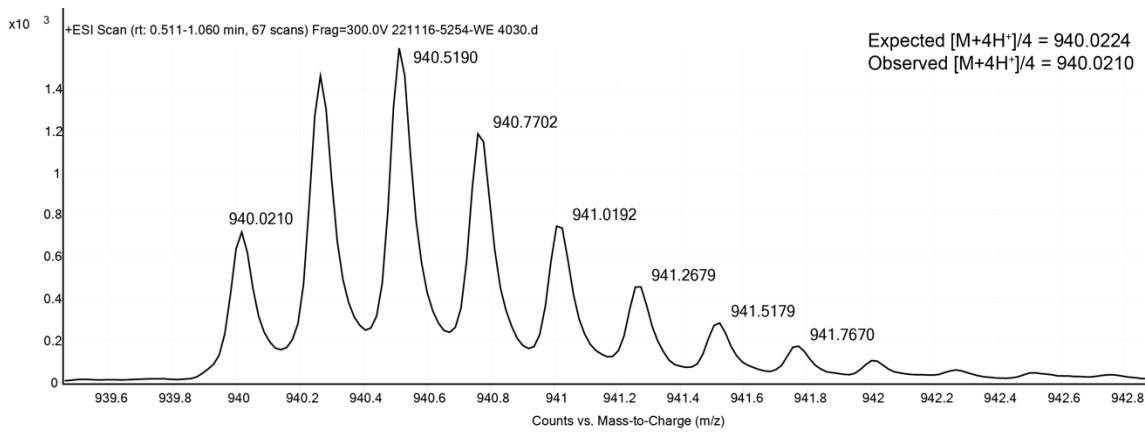


Figure 3-17. ESI TOF spectrum for peptide **ALA**.

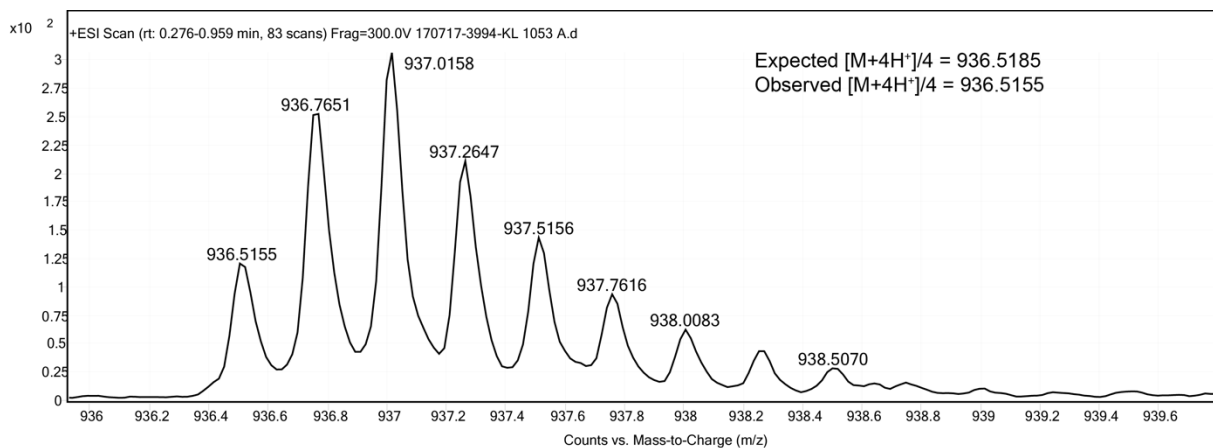


Figure 3-18. ESI TOF spectrum for AVA.

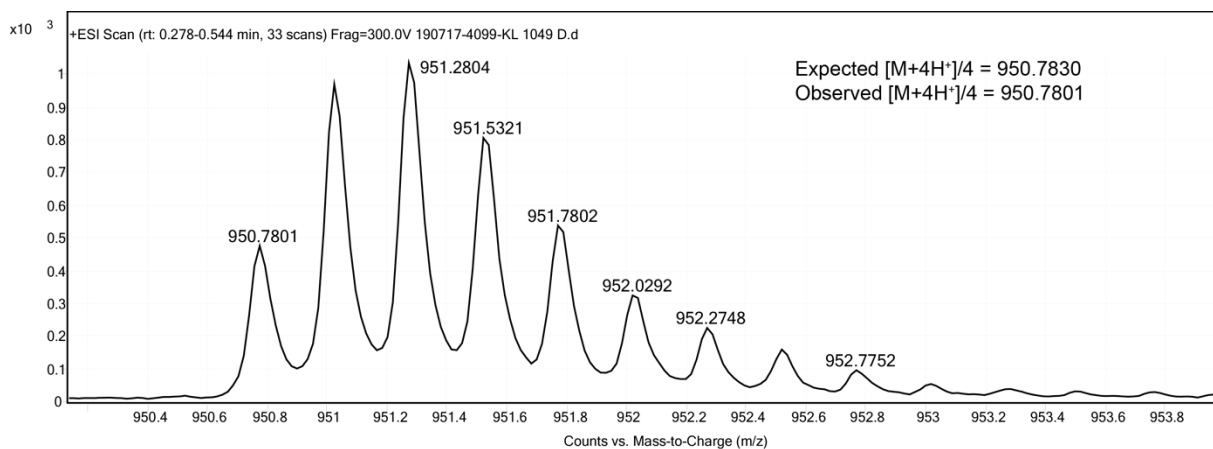


Figure 3-19. ESI TOF spectrum for AVK.

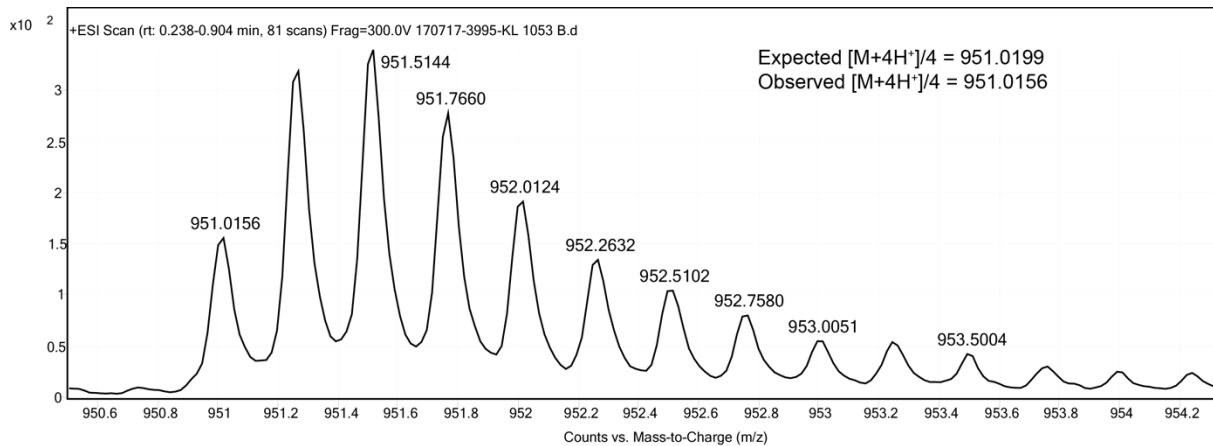


Figure 3-20. ESI TOF spectrum for EVA.

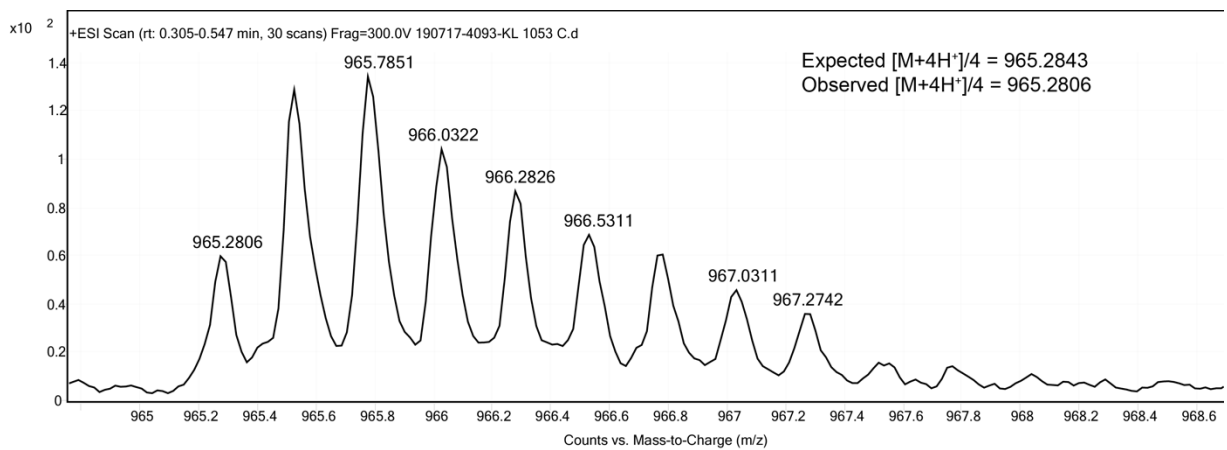


Figure 3-21. ESI TOF spectrum for EVK.

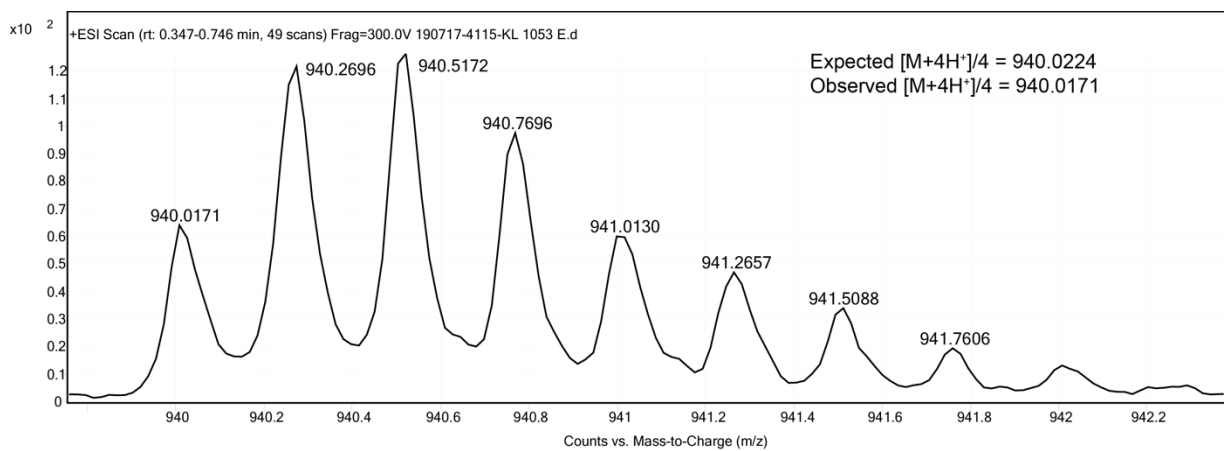


Figure 3-22. ESI TOF spectrum for ANleA.

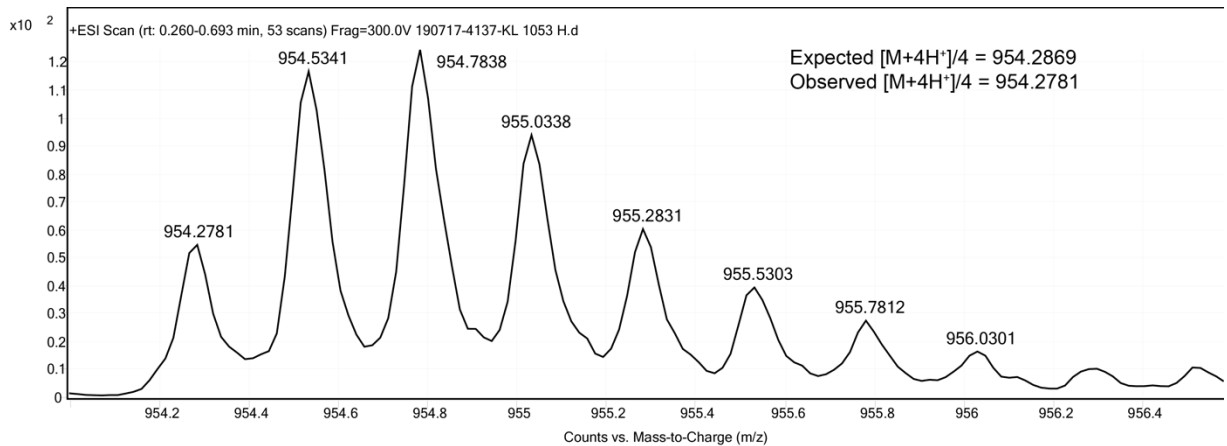


Figure 3-23. ESI TOF spectrum for ANleK.

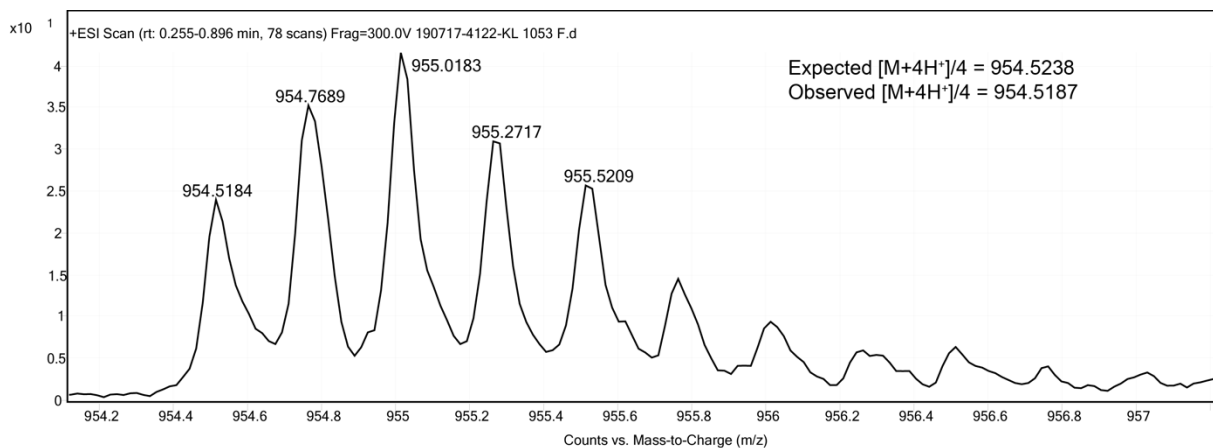


Figure 3-24. ESI TOF spectrum for ENleA.

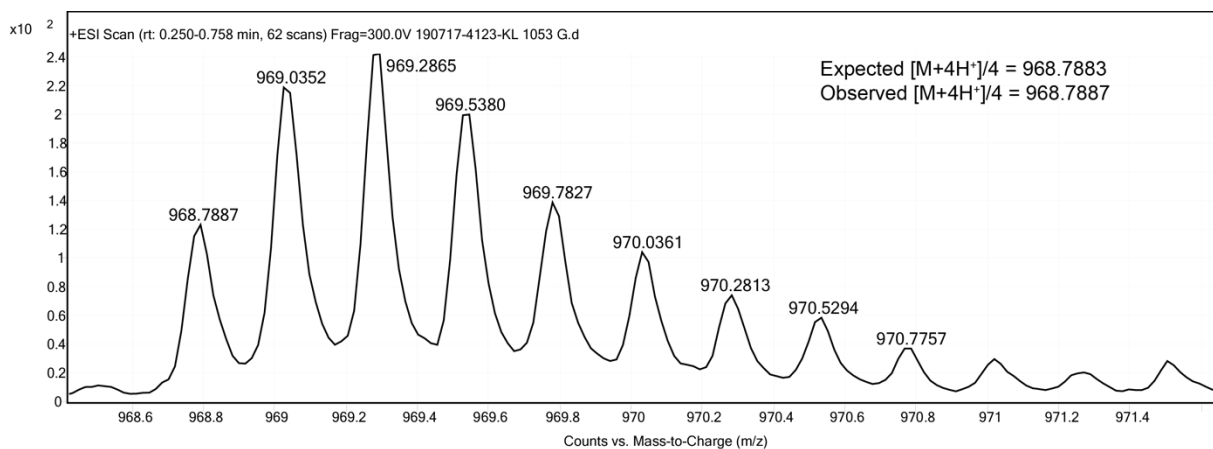


Figure 3-25. ESI TOF spectrum for ENleK.

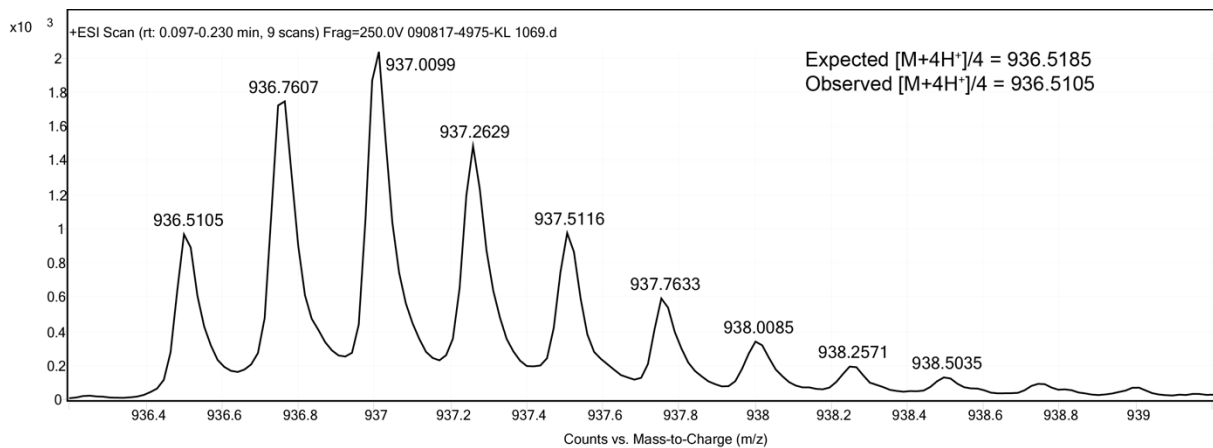


Figure 3-26. ESI TOF spectrum for ANvaA.

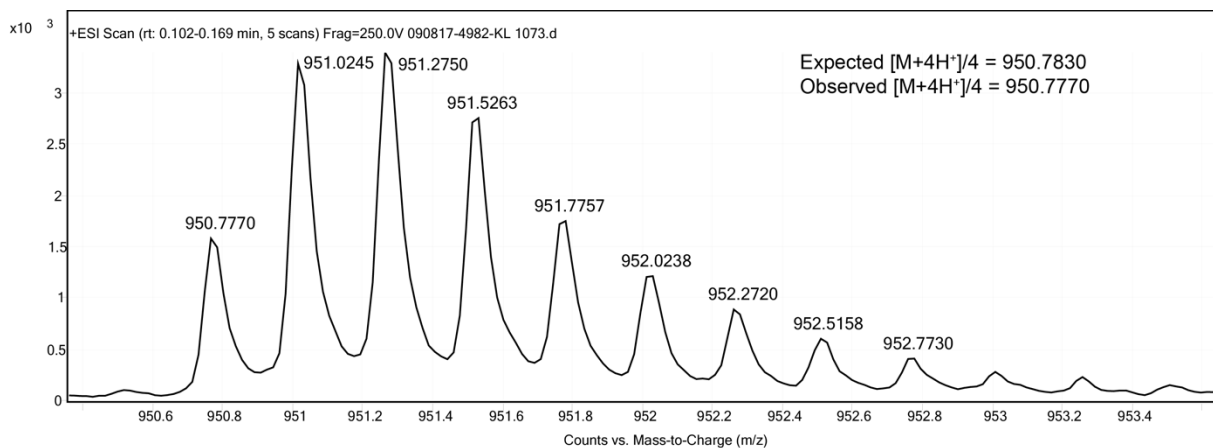


Figure 3-27. ESI TOF spectrum for ANvaK.

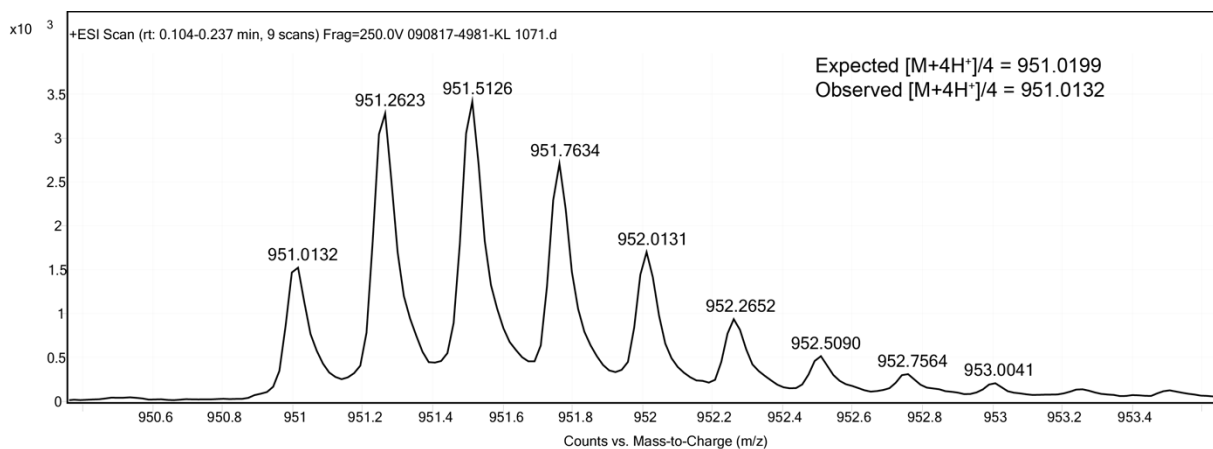


Figure 3-28. ESI TOF spectrum for ENvaA.

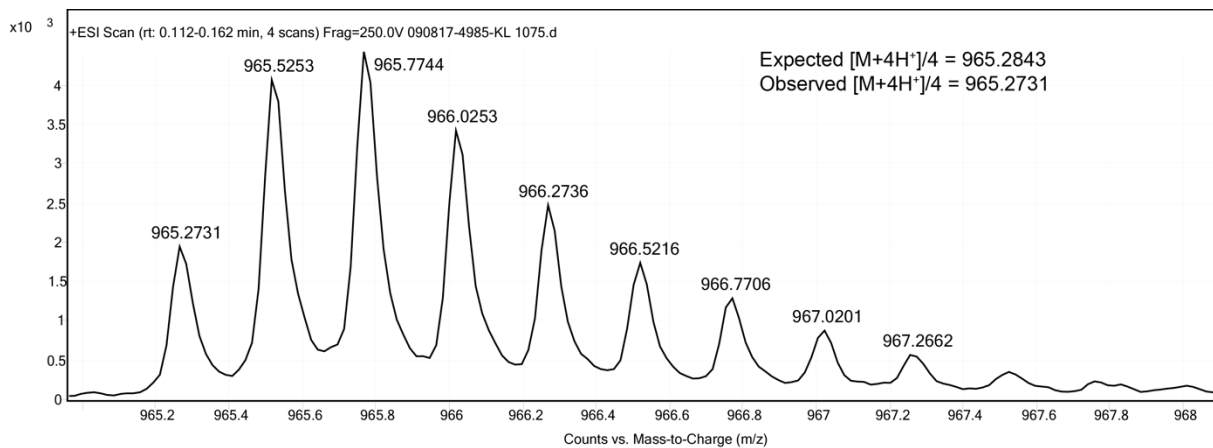


Figure 3-29. ESI TOF spectrum for ENvaK.

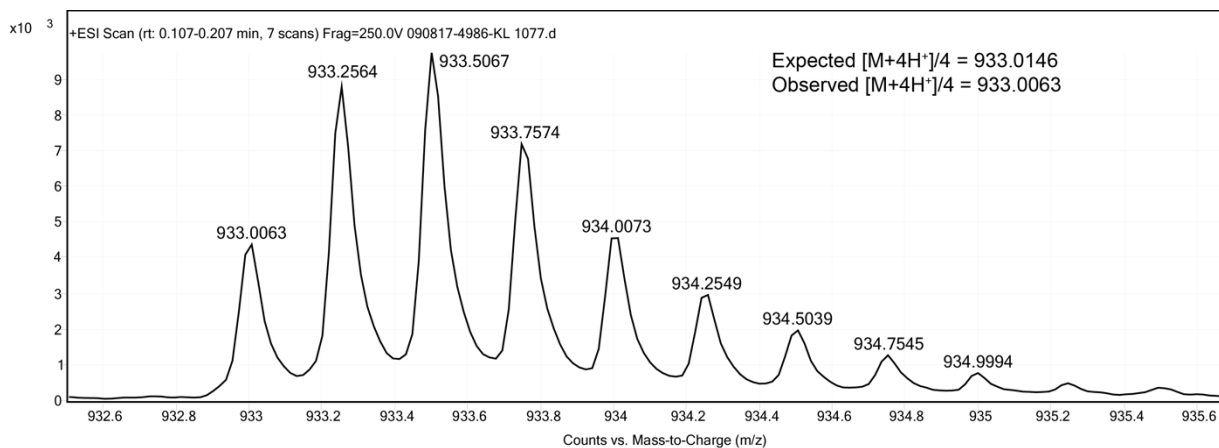


Figure 3-30. ESI TOF spectrum for **AAbuA**.

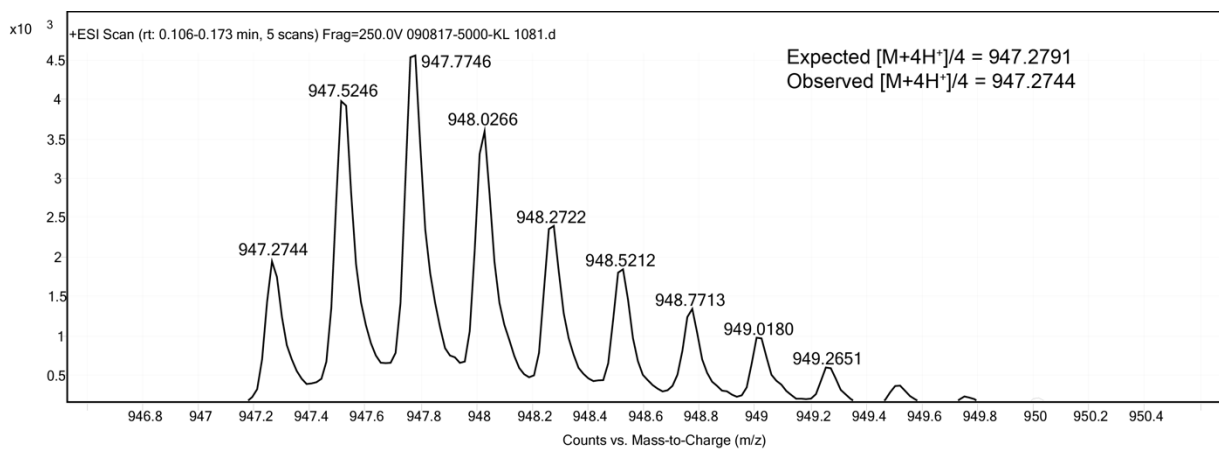


Figure 3-31. ESI TOF spectrum for **AAbuK**.

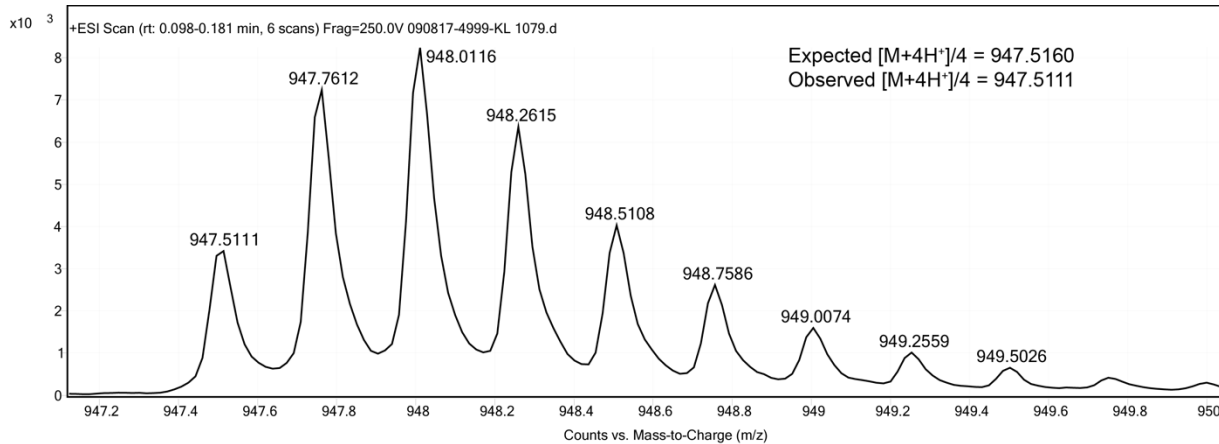


Figure 3-32. ESI TOF spectrum for **EAbuA**.

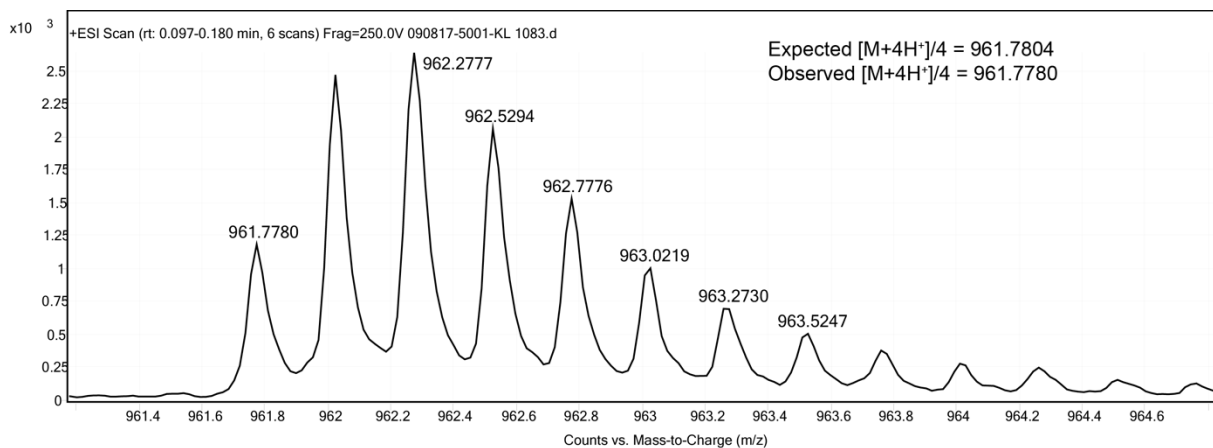


Figure 3-33. ESI TOF spectrum for **EAbuK**.

3.3.3 Analytical HPLC data

HPLC characterization for **EChak**, **EChaKp1**, **ELK**, **ENleK**, **EVK**, **ENvaK**, and **EabuK** along with their sequence variants are shown below.

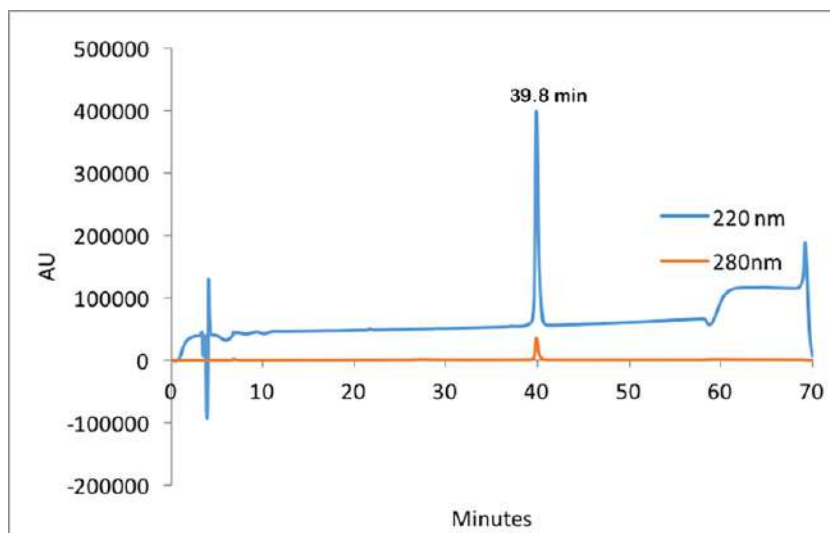


Figure 3-34. Analytical HPLC Data for **EChaK**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

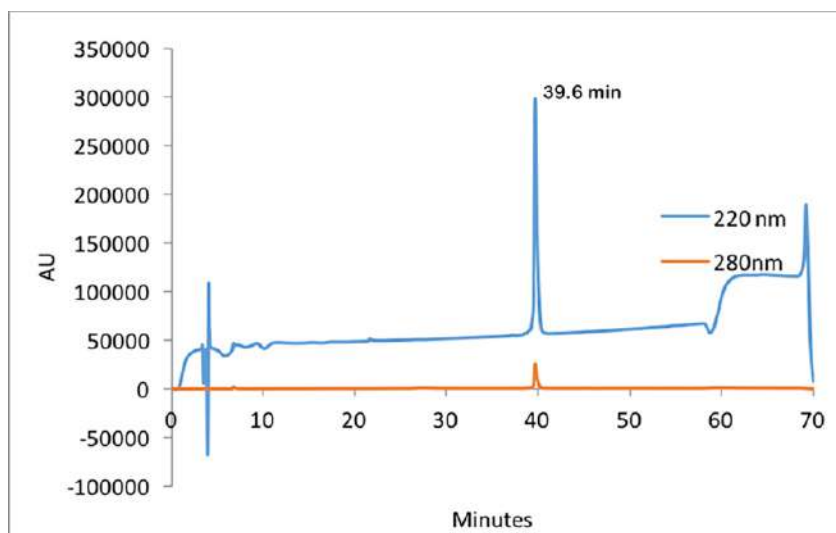


Figure 3-35. Analytical HPLC Data for **AChaK**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

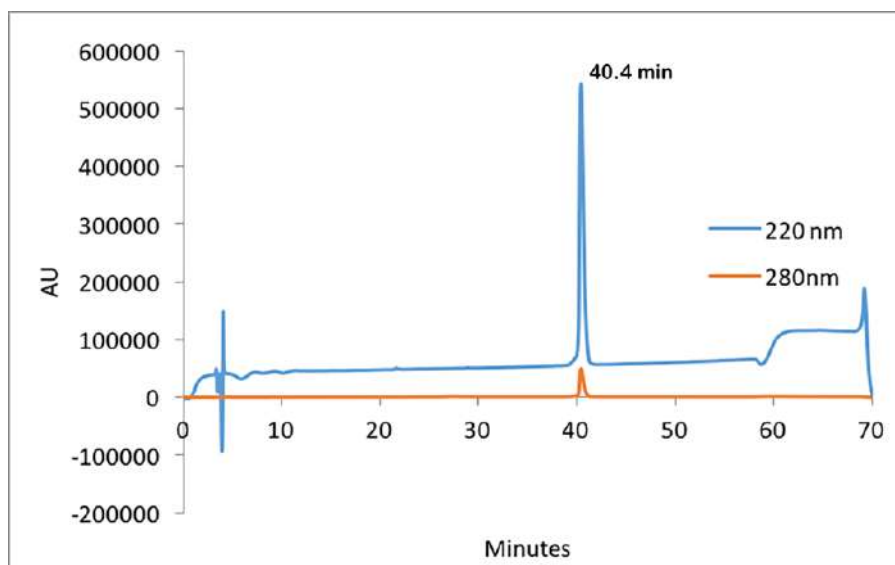


Figure 3-36. Analytical HPLC Data for **EChaA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

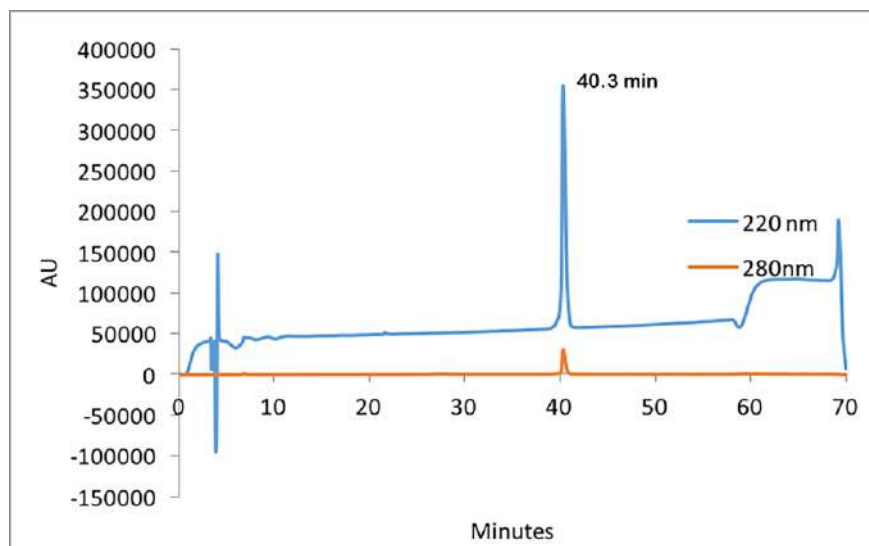


Figure 3-37. Analytical HPLC Data for **AChaA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

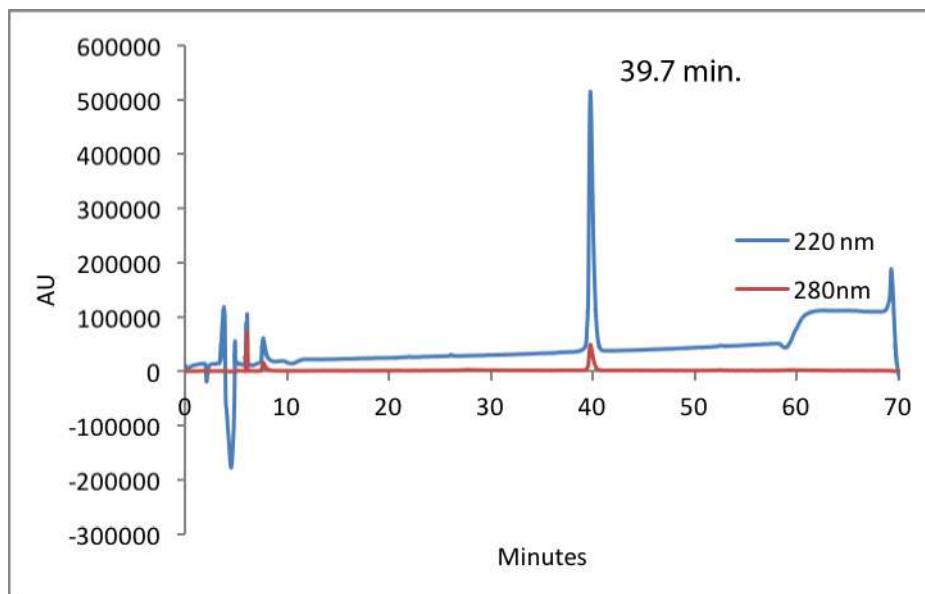


Figure 3-38. Analytical HPLC Data for **ELK**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

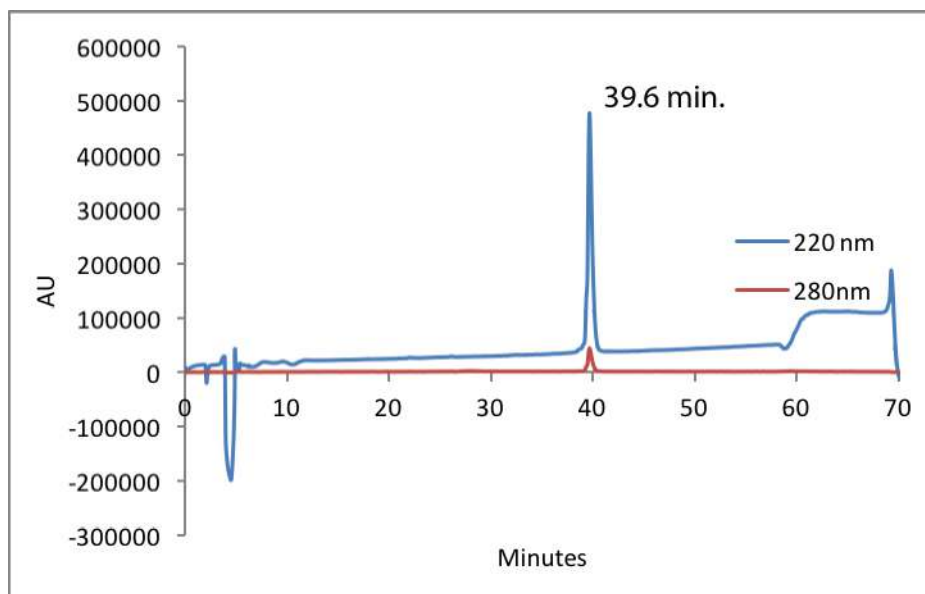


Figure 3-39. Analytical HPLC Data for **ALK**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

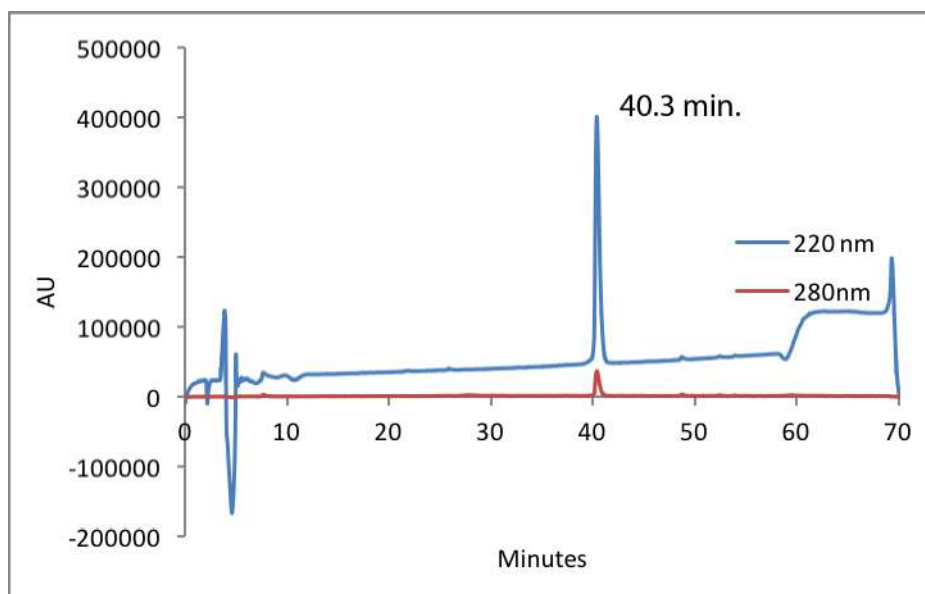


Figure 3-40. Analytical HPLC Data for **ELA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

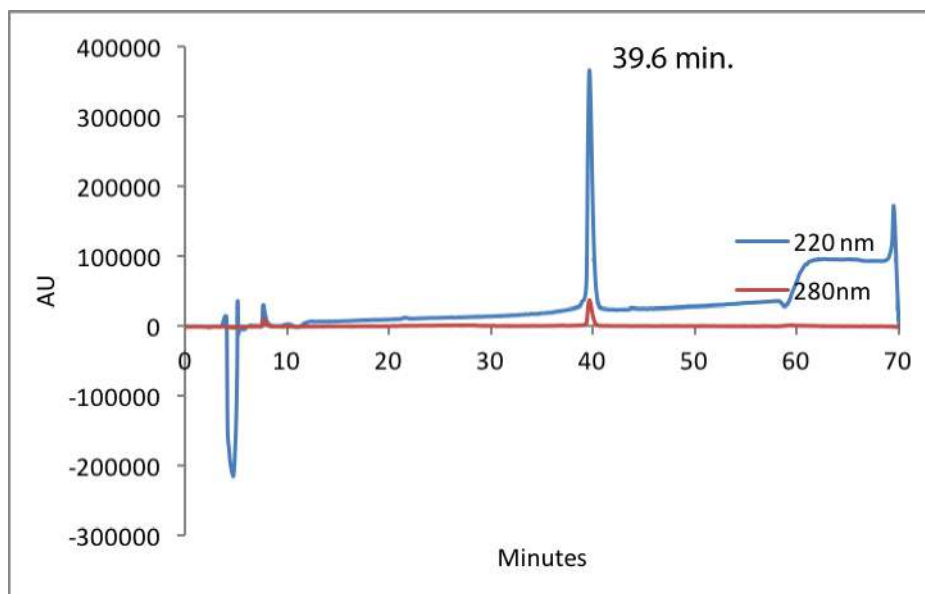


Figure 3-41. Analytical HPLC Data for **ALA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

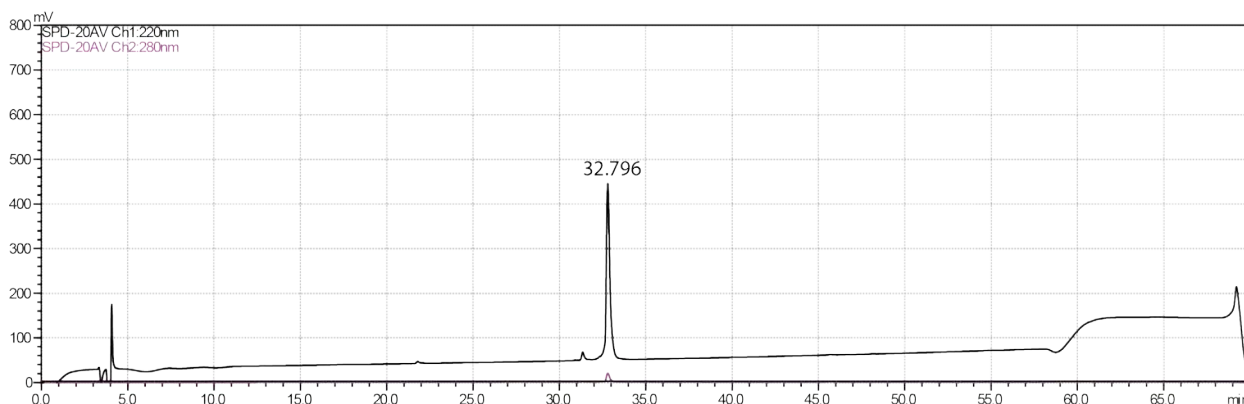


Figure 3-42. Analytical HPLC Data for **EChaK p1**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

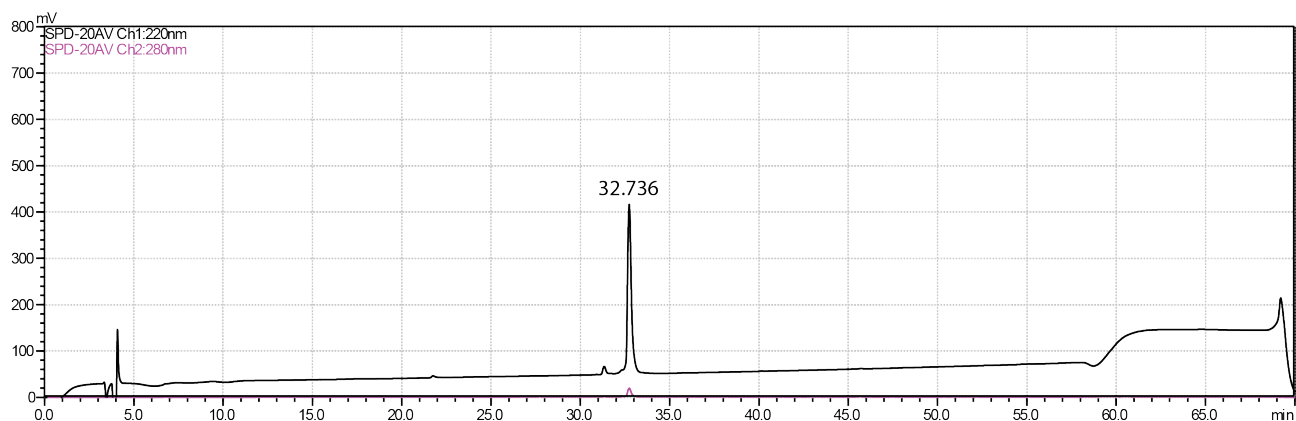


Figure 3-43. Analytical HPLC Data for **AChaK p1**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

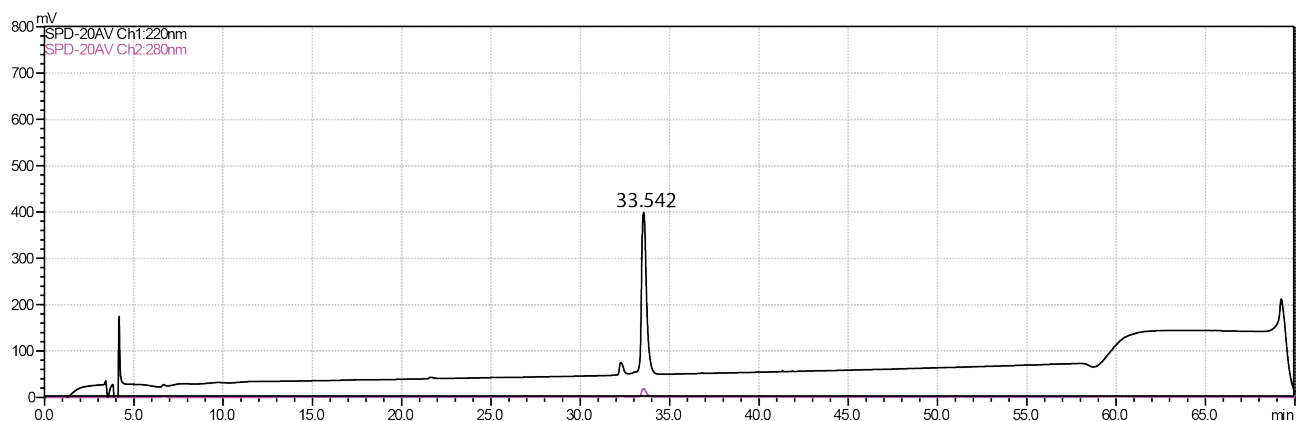


Figure 3-44. Analytical HPLC Data for **EChaA p1**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

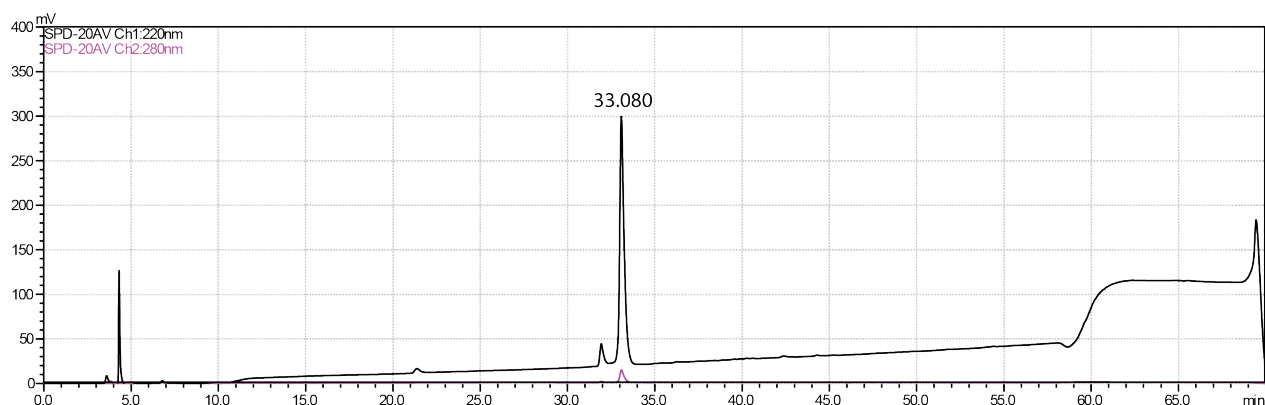


Figure 3-45. Analytical HPLC Data for **AChaA p1**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over

50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

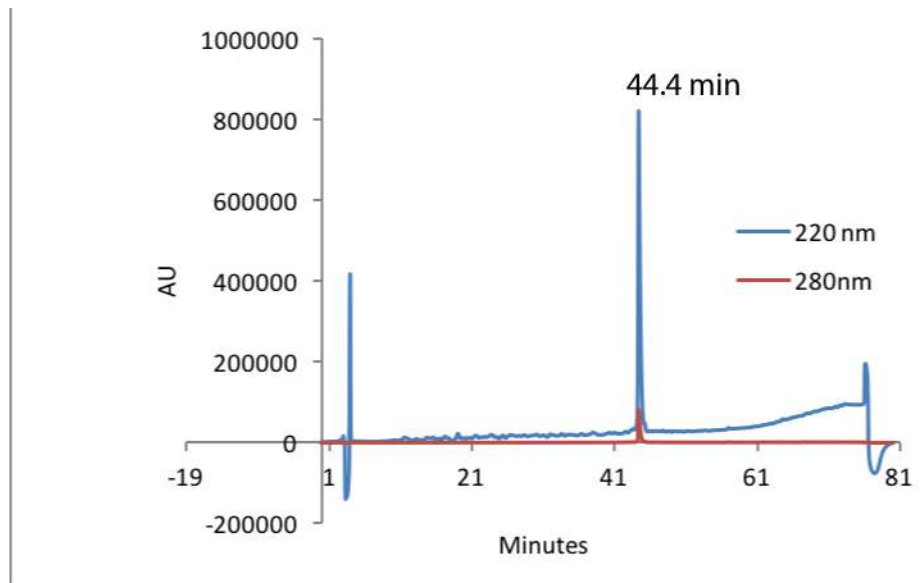


Figure 3-46. Analytical HPLC Data for **AVA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min

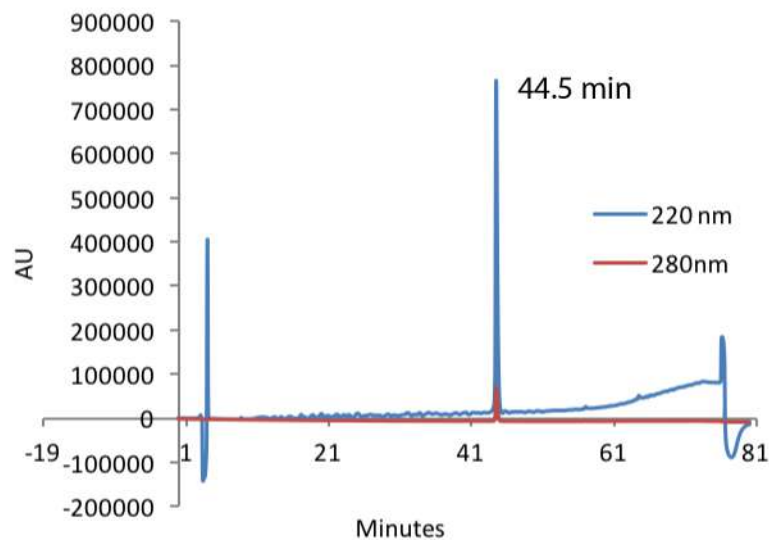


Figure 3-47. Analytical HPLC Data for **EVA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min

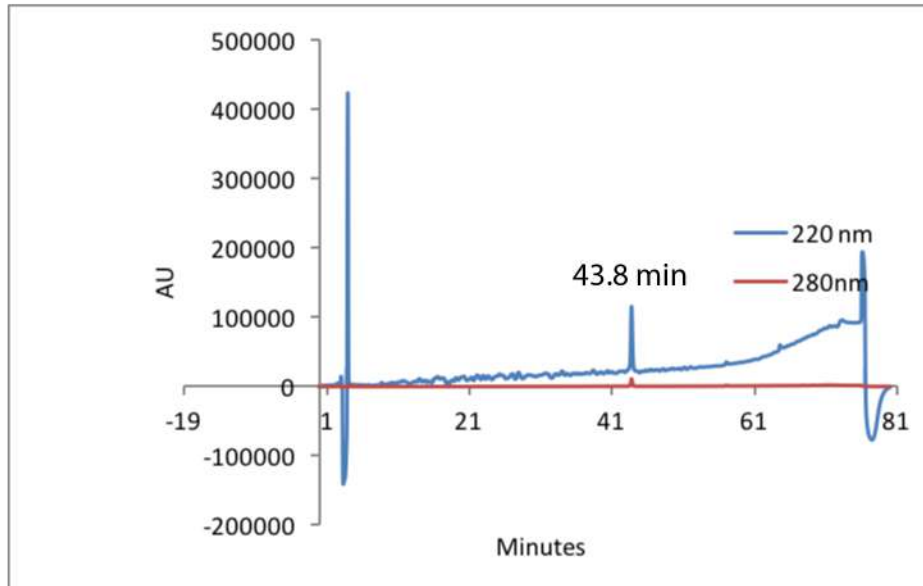


Figure 3-48. Analytical HPLC Data for **AVK**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min

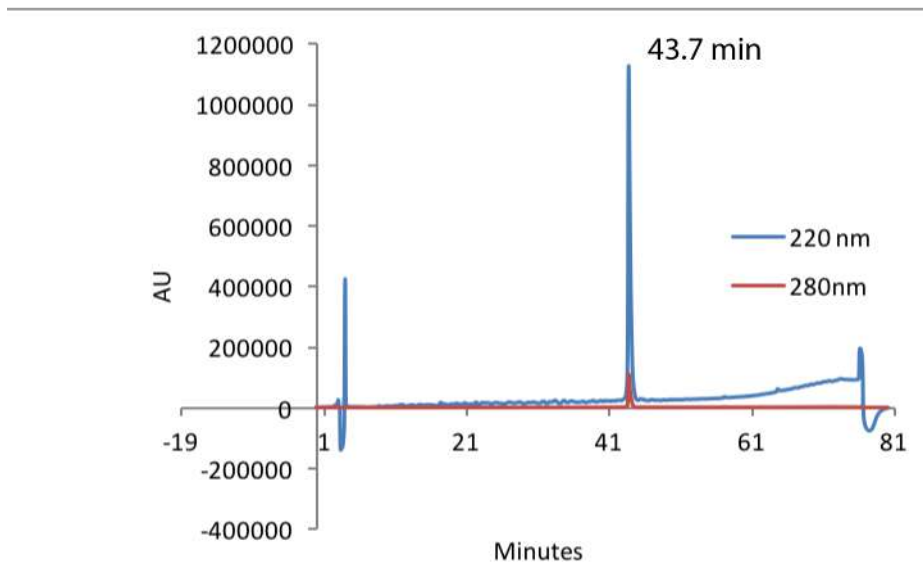


Figure 3-49. Analytical HPLC Data for **EVK**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

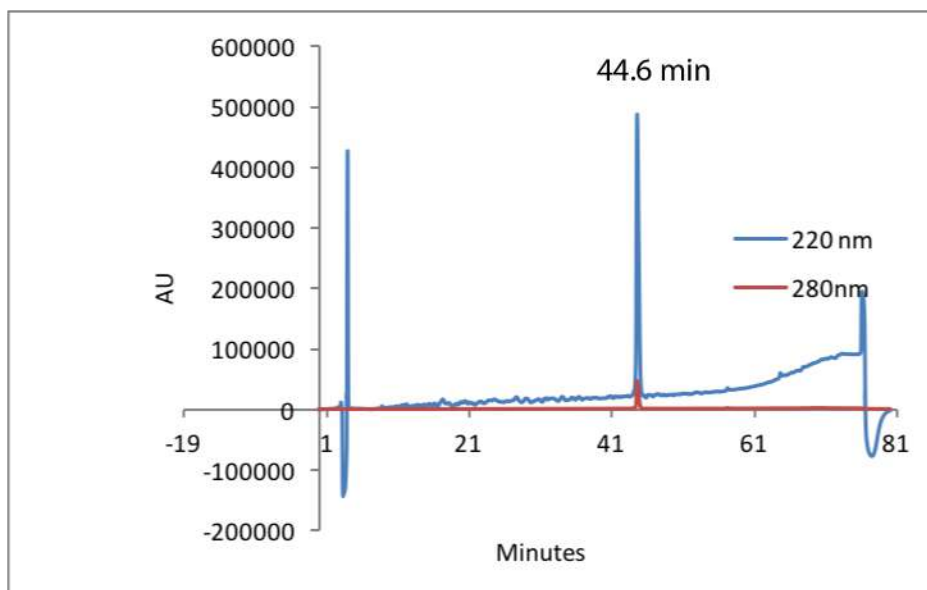


Figure 3-50. Analytical HPLC Data for **ANleA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

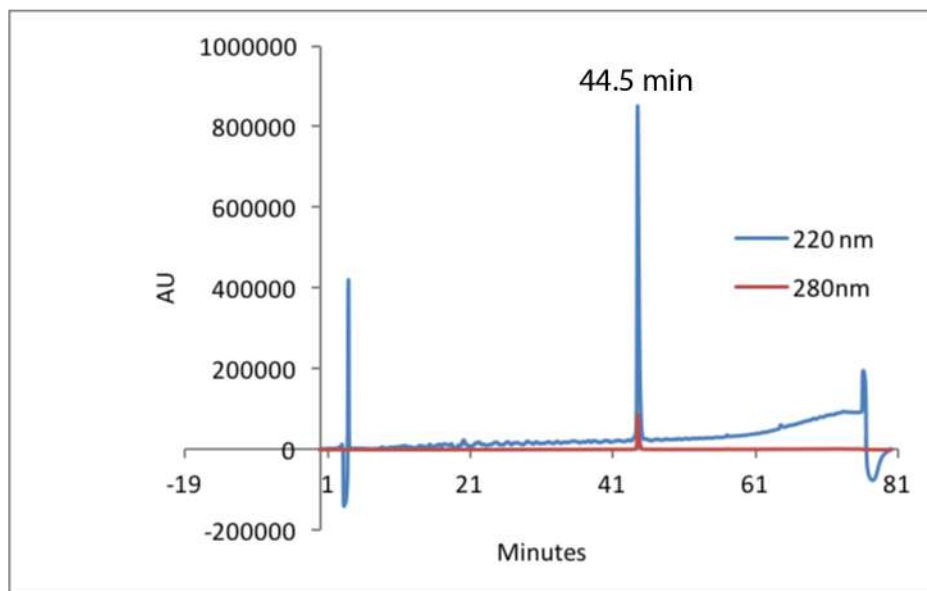


Figure 3-51. Analytical HPLC Data for **ENleA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

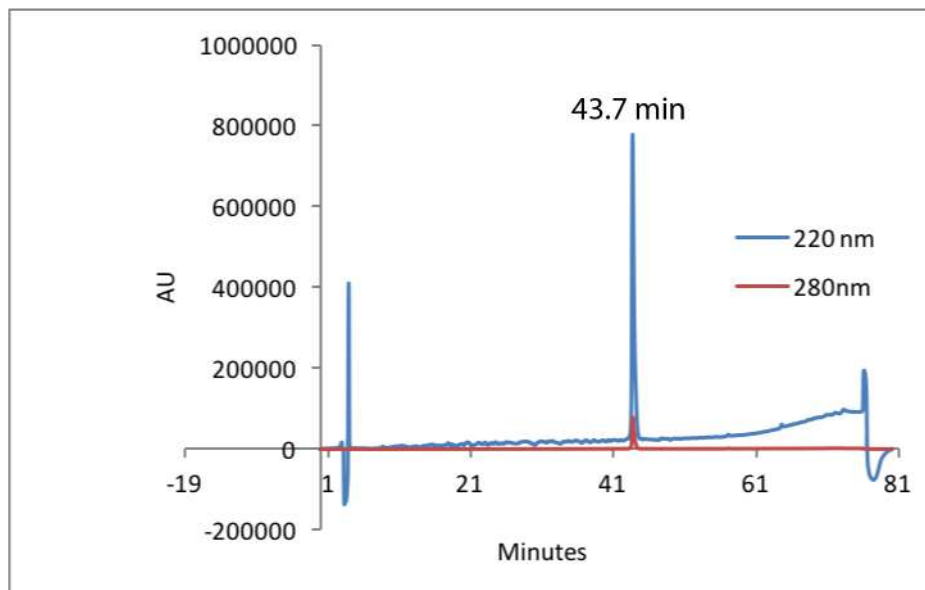


Figure 3-52. Analytical HPLC Data for **ANleK**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

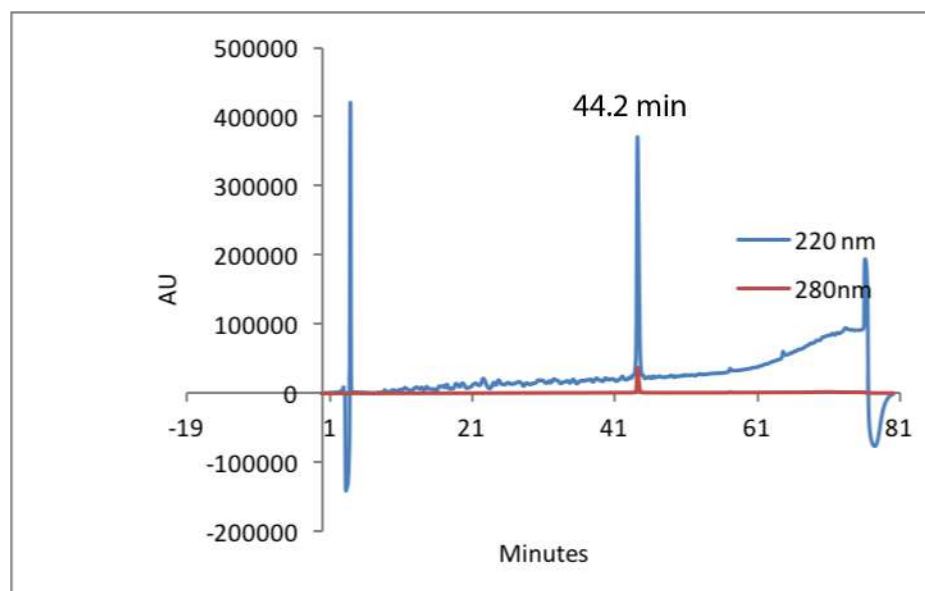


Figure 3-53. Analytical HPLC Data for **ENleK**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

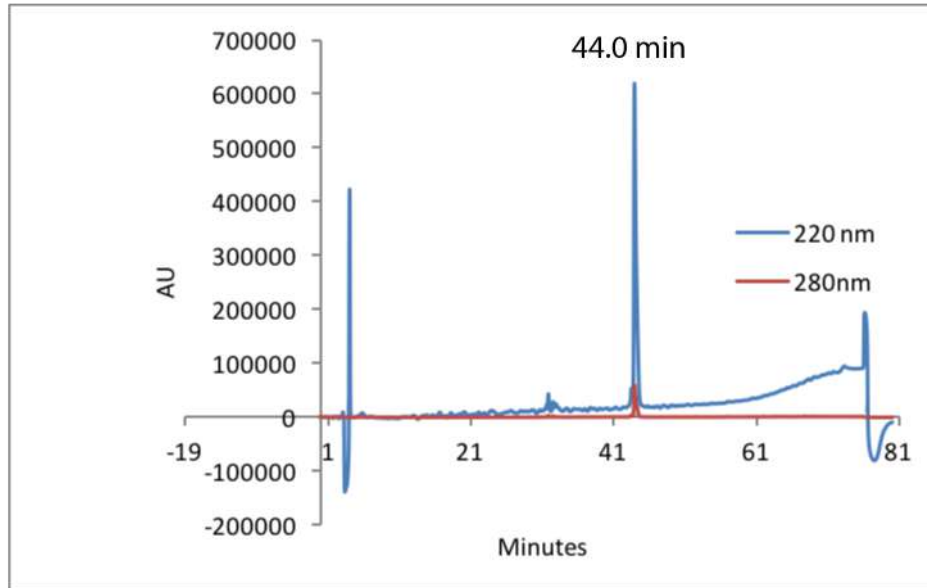


Figure 3-54. Analytical HPLC Data for **ANvaA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

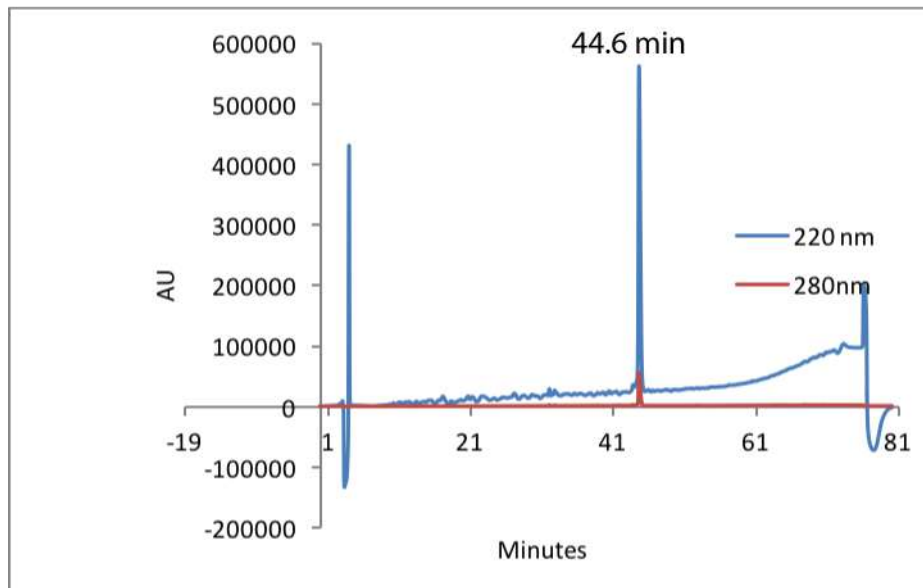


Figure 3-55. Analytical HPLC Data for **ENvaA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

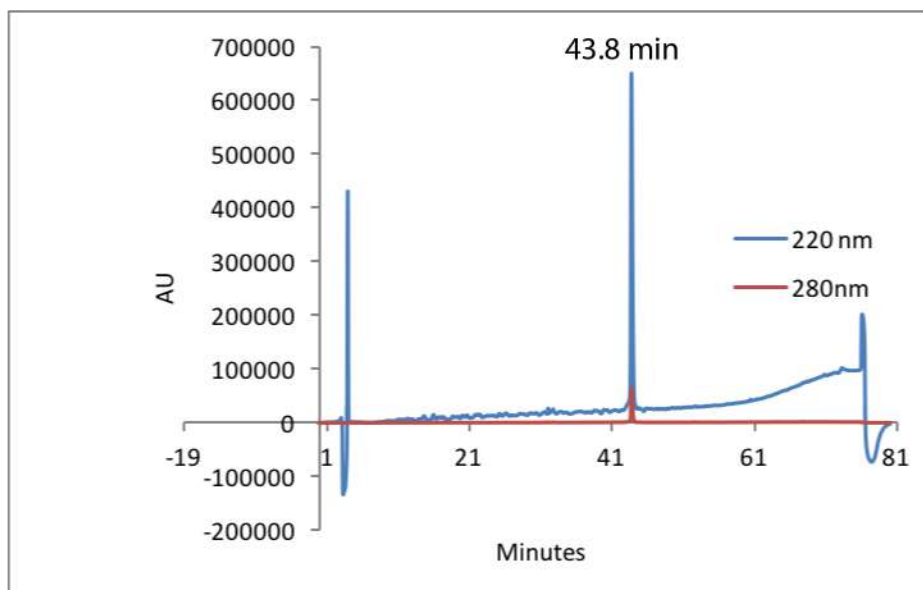


Figure 3-56. Analytical HPLC Data for **ANvaK**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

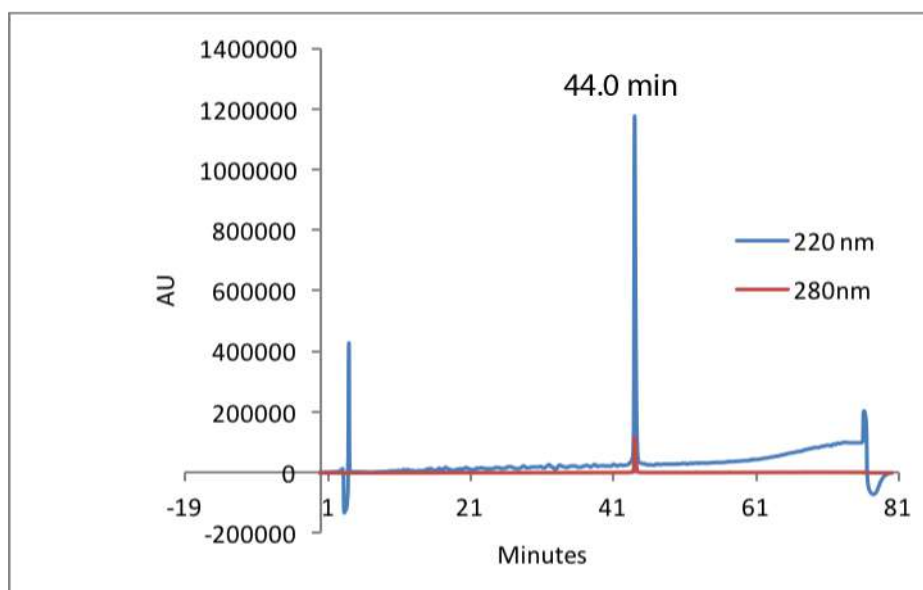


Figure 3-57. Analytical HPLC Data for **ENvaK**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

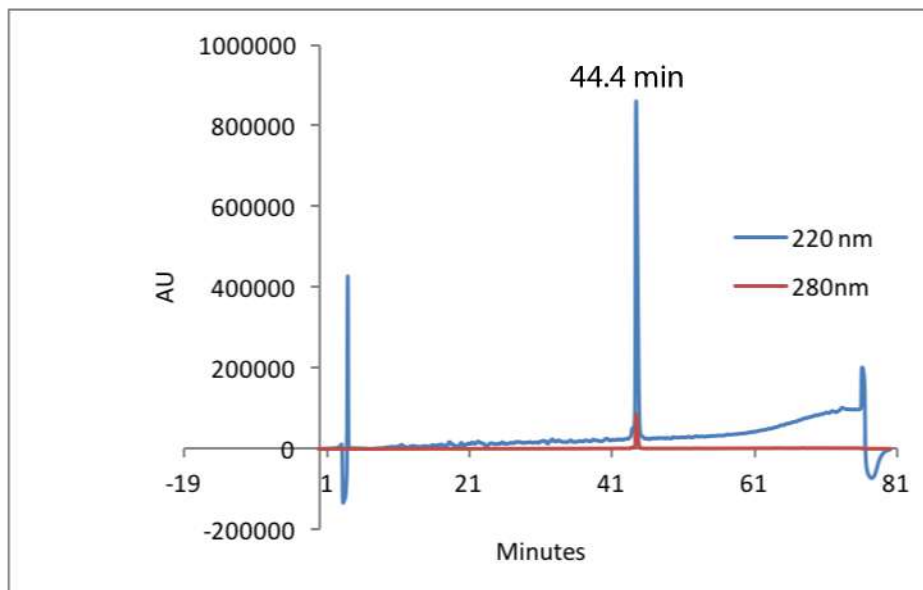


Figure 3-58. Analytical HPLC Data for **AAbuA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

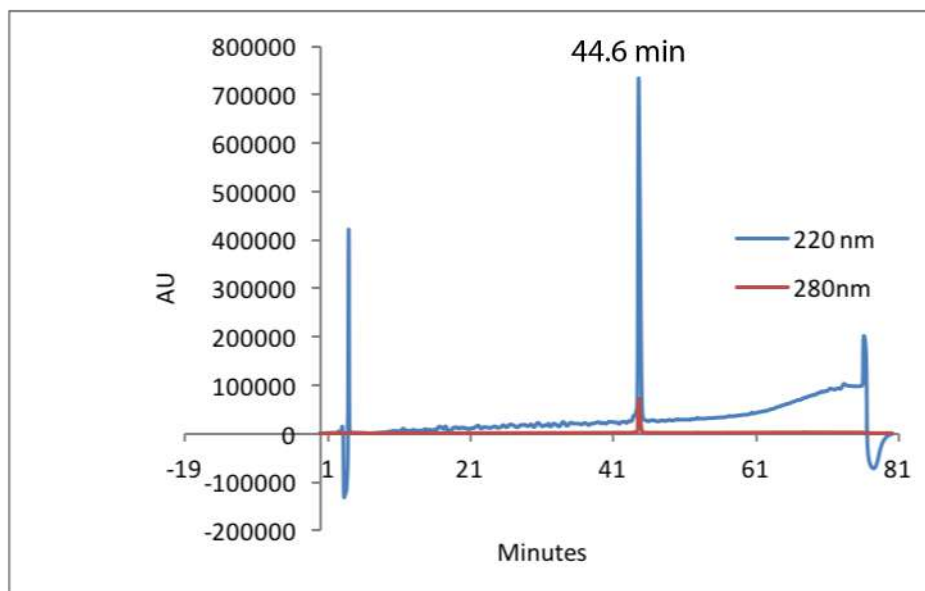


Figure 3-59. Analytical HPLC Data for **EAbuA**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

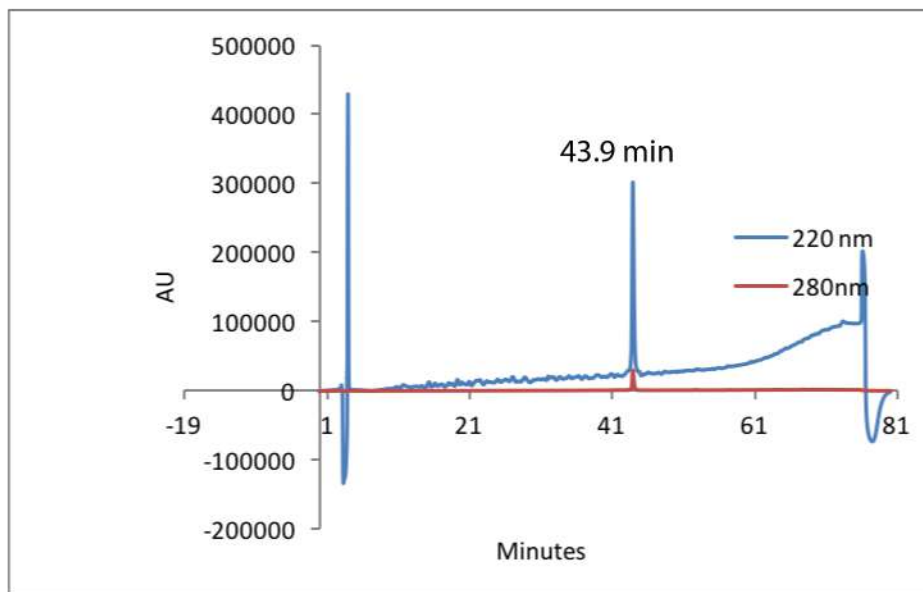


Figure 3-60. Analytical HPLC Data for **AAbuK**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

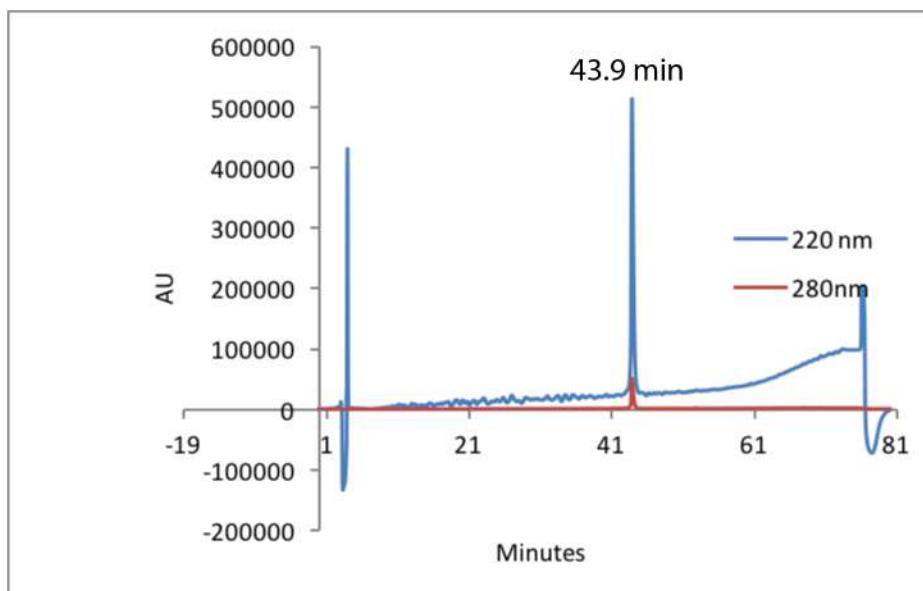


Figure 3-61. Analytical HPLC Data for **EAbuK**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) at 1 mL/min.

3.3.4 Size Exclusion Chromatography

Size exclusion chromatography (SEC) was done on a Shimadzu HPLC instrument using a Zenix-100 SEC analytical column or a Phenomenex Yarra 3u sec-3000 column. Size exclusion chromatography for **EChaK**, **EChaK p1**, **ELK**, **ENleK**, **ENvaK**, **EVK**, and **EAbuK** and their sequence variants were injected as 30 μ M samples in 20mM sodium phosphate (pH 7.0) onto the column on a 25 minute isocratic method of a mobile phase comprised of 20mM sodium phosphate (pH 7.0). and 0.3M NaCl. Each variant was compared to a trimer and dimer standards previously characterized.¹²⁻¹³ Retention times for **EChaK**, **ELK**, **ENleK**, **ENvaK**, **EVK**, and **EAbuK** and their sequence variants were more comparable to the trimer standard than the dimer or monomer standards. Retention time **EChaKp1** and its sequence variants were more comparable to the Dimer standard than the trimer standard.

Table 3-3 Retention times of helical peptides on a Zenix-100 SEC analytical column

Batch 1 (7/27/17)			
Peptide	Calculated MW (Da.)	Retention time (min.)	Association state
Trimer standard	11151	9.15	Trimer
Dimer Standard	8131	10.31	Dimer
AVA	11187	9.26	Trimer
EVA	11358	9.04	Trimer
AVK	11362	9.17	Trimer
EVK	11533	9.17	Trimer
ANleA	11229	9.16	Trimer
ENleA	11400	9.06	Trimer
ANleK	11404	9.19	Trimer
ENleK	11575	9.08	Trimer

Batch 2 (10/10/17)			
Peptide	Calculated MW (Da.)	Retention time (min.)	Association state
Trimer standard	11016	9.04	Trimer
Dimer Standard	8131	10.23	Dimer
ANvaA	11187	9.17	Trimer
ENvaA	11358	9.00	Trimer
ANvaK	11362	9.15	Trimer
ENvaK	11533	9.04	Trimer
AAbuA	11145	9.09	Trimer
EAbuA	11316	8.98	Trimer
AAbuK	11319	9.02	Trimer
EAbuK	11491	9.11	Trimer

Table 3-4 Retention times of helical peptides on a Zenix-100 SEC analytical column

Batch 3			
Peptide	MW (Da)	Retention Time (min)	Association state
1CW(monomer standard)	11199	8.83	Trimer
GCN4(dimer standard)	8131	9.43	Dimer
PSBD36(monomer standard)	4001	10.51	Monomer
AChaA p1	7927	10.34	Dimer/Monomer
EChaA p1	8041	9.83	Dimer
AChaK p1	8043	10.33	Dimer/Monomer
EChaK p1	8157	9.79	Dimer

Batch 4

Peptide	MW (Da)	Retention Time (min)	Association state
1CW(trimer standard)	11199	9.04	Trimer
ELK	11575	8.94	Trimer
19-ALK	11404	8.96	Trimer
19-ELA	11401	9.07	Trimer
1CW(trimer standard)	11199	9.11	Trimer
19-ALA	11229	9.23	Trimer

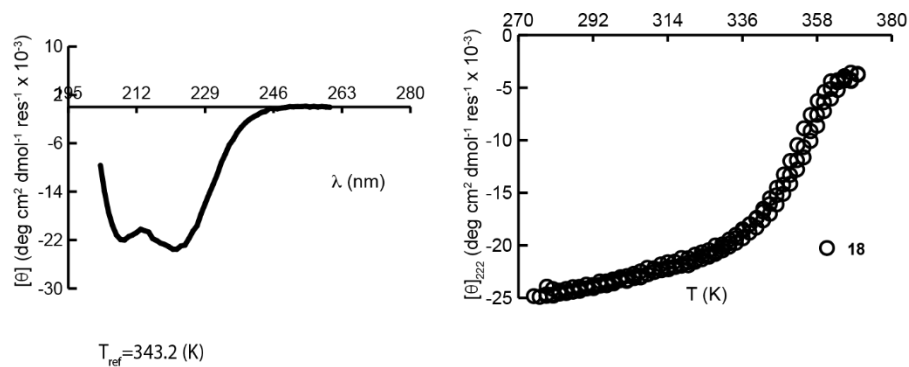
Table 3-5 Retention times of helical peptides on a Phenomenex Yarra 3u sec-3000 column.

Batch 5

Peptide	Calculated MW (Da)	Retention Time (min)	association state
1CW (trimer standard)	11017	10.50	
AChaA	11214	10.44	trimer
EChaA	11385	10.33	trimer
AChaK	11389	10.52	trimer
EChaK	11560	10.41	trimer

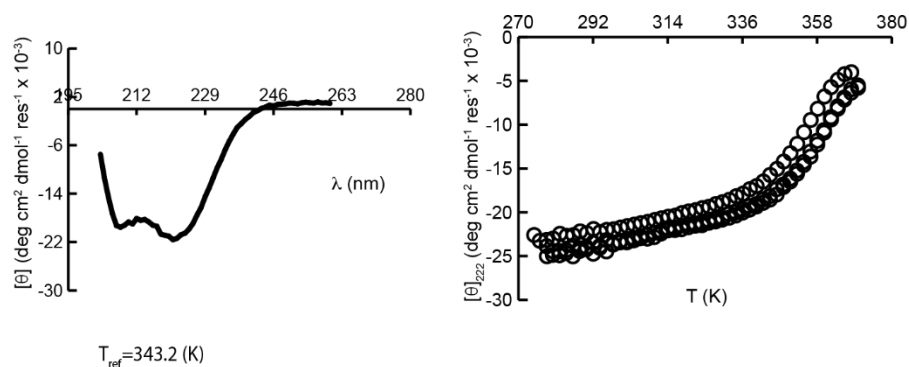
3.3.5 Folded Free energy measurement, Circular Dichroism spectropolarimetry.

Experimental details and the equations used for acquiring the folded free energies for **EChak**, **EChaKp1**, **ELK**, **ENleK**, **EVK**, **ENvaK**, **EabuK**, **EAK**, **ESK**, and **EFK** are the same as for the compounds found in Chapter 2 details can be found in Section 2.4.5. The wavelengths scans and thermodynamic fits a **EChak**, **EChaKp1**, **ELK**, **ENleK**, **EVK**, **ENvaK**, and **EabuK** are found in figures 3-62 to 3-89.



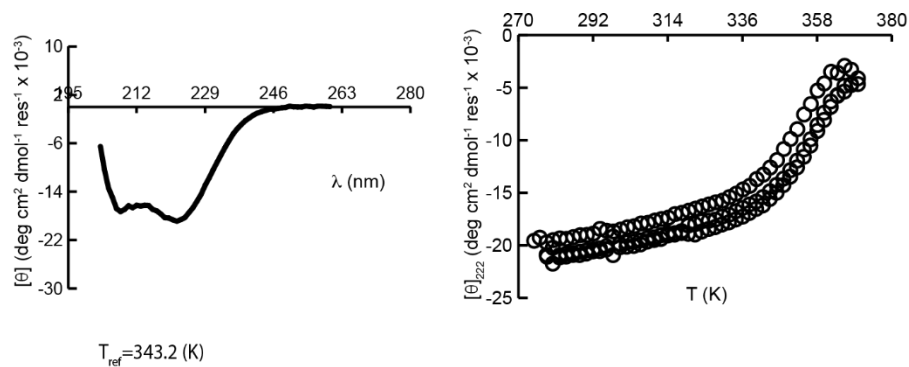
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_1 / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
EChaK	-16.0 ± 0.03	0.19 ± 0.003	0.9995	0.1597

Figure 3-62. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **EChaK** in 20 mM sodium phosphate (pH 7). Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



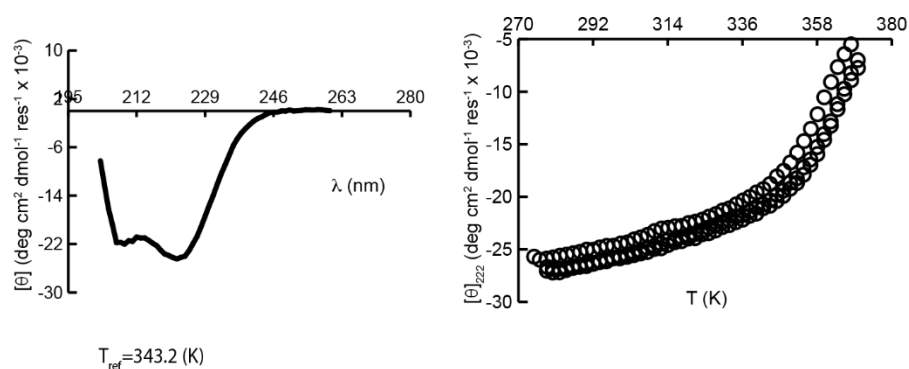
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_1 / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
ACHaK	-17.2 ± 0.07	0.20 ± 0.006	0.9985	0.2219

Figure 3-63. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **ACHaK** in 20 mM sodium phosphate (pH 7). Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



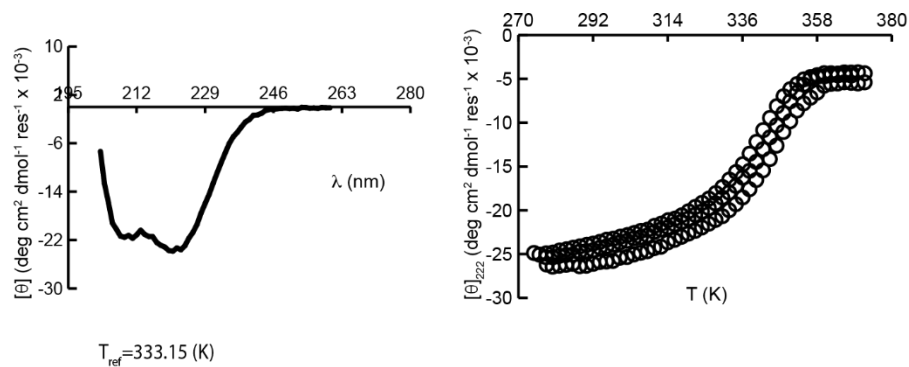
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_1 / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
EChaA	-16.5 ± 0.06	0.20 ± 0.005	0.9988	0.1853

Figure 3-64. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **EChaA** in 20 mM sodium phosphate (pH 7). Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



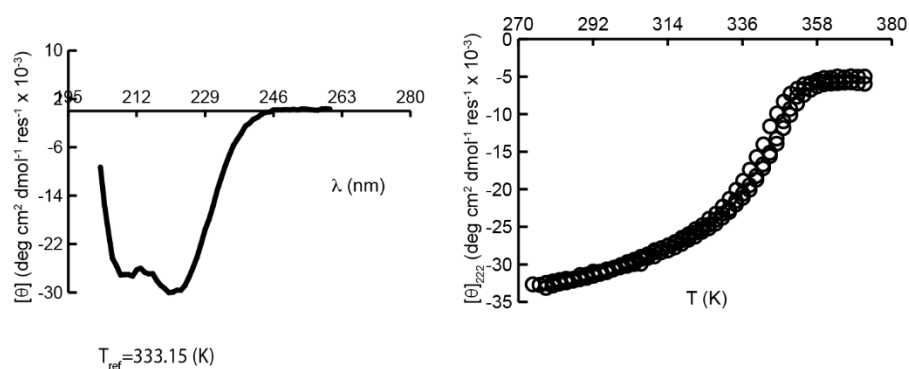
Protein	$\Delta G_f / \text{kcal mol}^{-1}$	$\Delta G_1 / \text{kcal mol}^{-1} \text{K}^{-1}$	R^2	rmsd error
ACHaA	-18.2 ± 0.06	0.19 ± 0.005	0.9992	0.1571

Figure 3-65. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **ACHaA** in 20 mM sodium phosphate (pH 7). Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



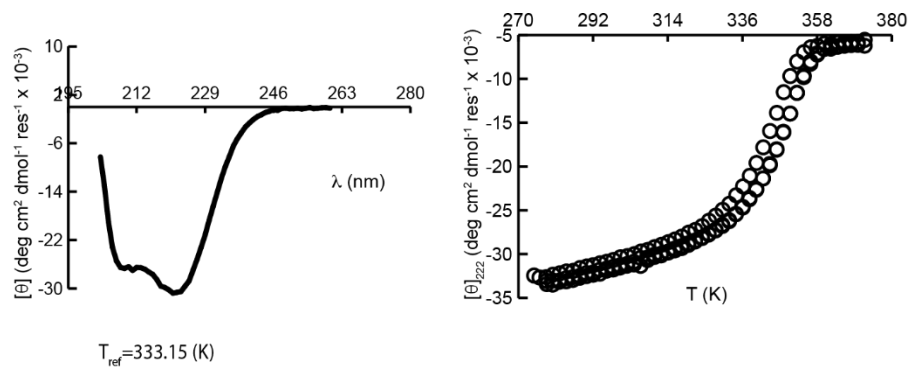
Protein	$\Delta G / \text{kcal mol}^{-1}$	$\Delta G_1 / \text{kcal mol}^{-1} \text{K}^{-1}$	$\Delta G_2 / \text{kcal mol}^{-1} \text{K}^{-2}$	R^2	rmsd error
EChAK p1	-7.97 ± 0.03	0.115 ± 0.002	0.002 ± 0.00005	0.9997	0.1349

Figure 3-66. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **EChAK p1** in 20 mM sodium phosphate (pH 7). Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



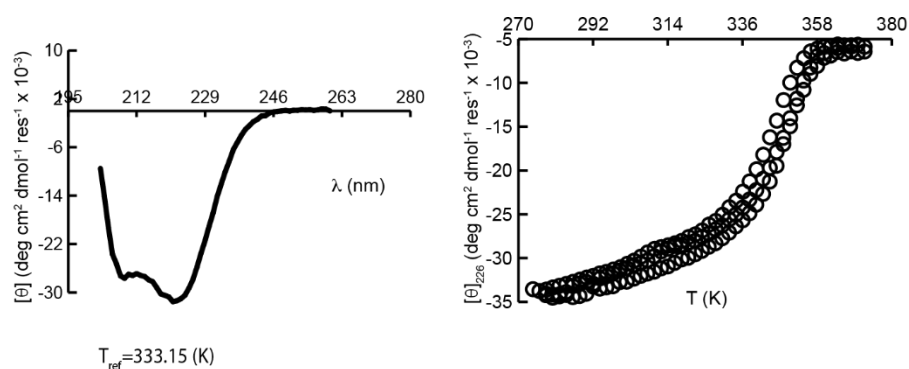
Protein	$\Delta G / \text{kcal mol}^{-1}$	$\Delta G_1 / \text{kcal mol}^{-1} \text{K}^{-1}$	$\Delta G_2 / \text{kcal mol}^{-1} \text{K}^{-2}$	R^2	rmsd error
EChaA p1	-8.19 ± 0.03	0.130 ± 0.002	0.002 ± 0.00004	0.9998	0.13044

Figure 3-67. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **EChaA p1** in 20 mM sodium phosphate (pH 7). Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



Protein	$\Delta G / \text{kcal mol}^{-1}$	$\Delta G_1 / \text{kcal mol}^{-1} \text{K}^{-1}$	$\Delta G_2 / \text{kcal mol}^{-1} \text{K}^{-2}$	R^2	rmsd error
AChaK p1	-8.79 ± 0.02	0.131 ± 0.001	0.002 ± 0.00004	0.9998	0.1280

Figure 3-68. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **AChaK p1** in 20 mM sodium phosphate (pH 7). Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



Protein	$\Delta G / \text{kcal mol}^{-1}$	$\Delta G_1 / \text{kcal mol}^{-1} \text{K}^{-1}$	$\Delta G_2 / \text{kcal mol}^{-1} \text{K}^{-2}$	R^2	rmsd error
AChaA p1	-9.10 ± 0.04	0.137 ± 0.002	0.002 ± 0.0001	0.9997	0.1698

Figure 3-69. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **AChaA p1** in 20 mM sodium phosphate (pH 7). Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

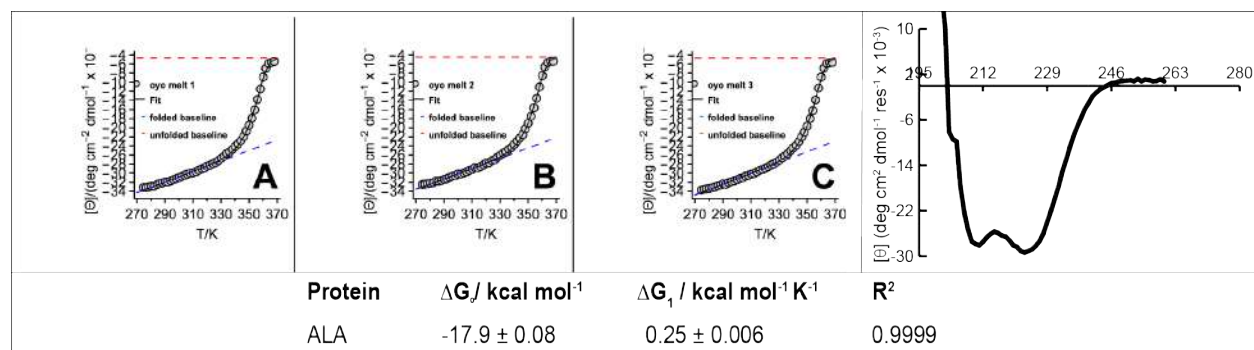


Figure 3-70. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein ALA in 20 mM sodium phosphate (pH 7), and 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

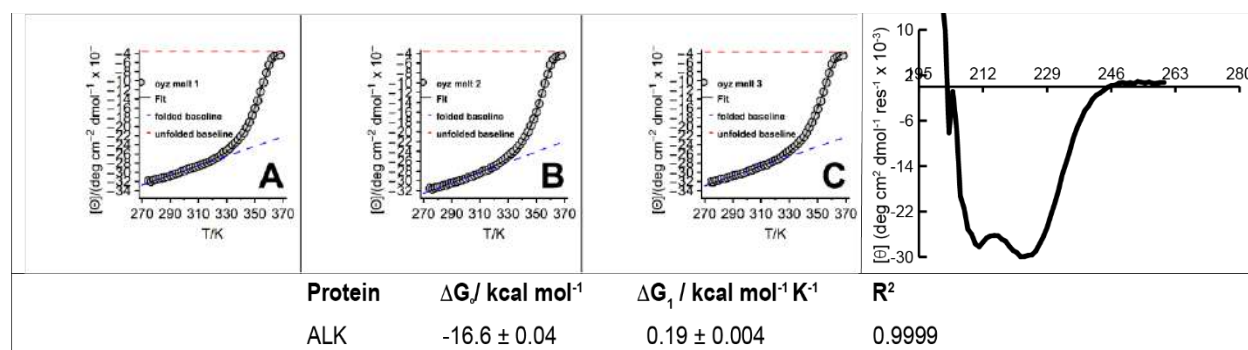


Figure 3-71. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein ALK in 20 mM sodium phosphate (pH 7), and 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

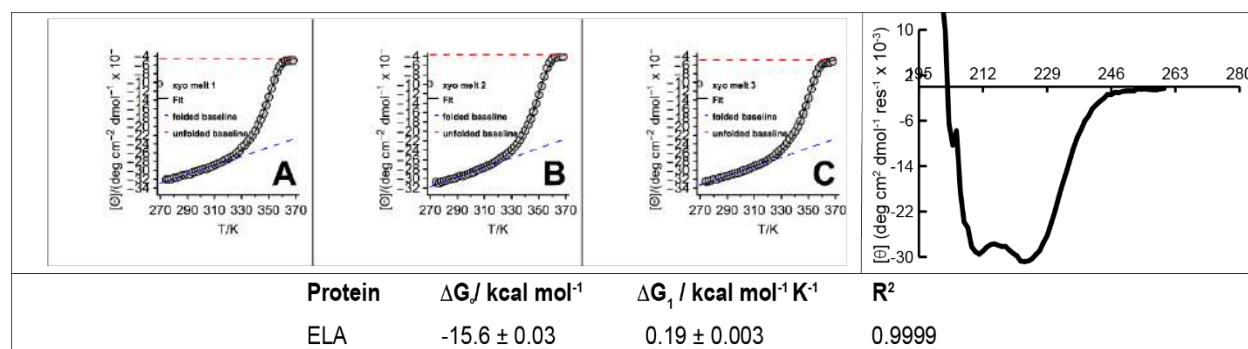


Figure 3-72. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein ELA in 20 mM sodium phosphate (pH 7), and 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

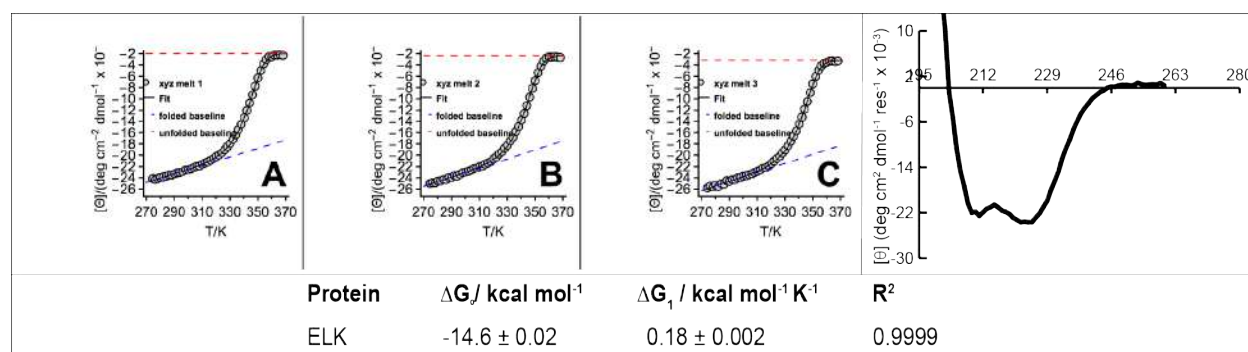


Figure 3-73. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein ELK in 20 mM sodium phosphate (pH 7), and 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

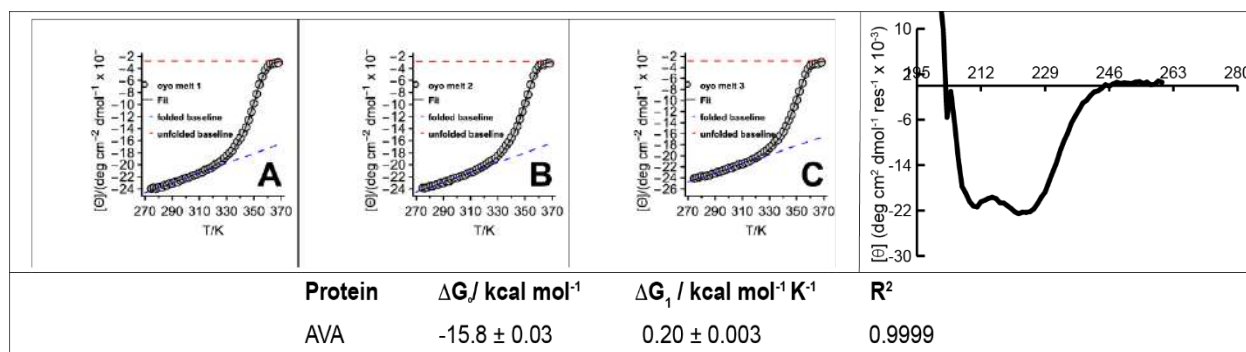


Figure 3-74. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein AVA in 20 mM sodium phosphate (pH 7), and 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

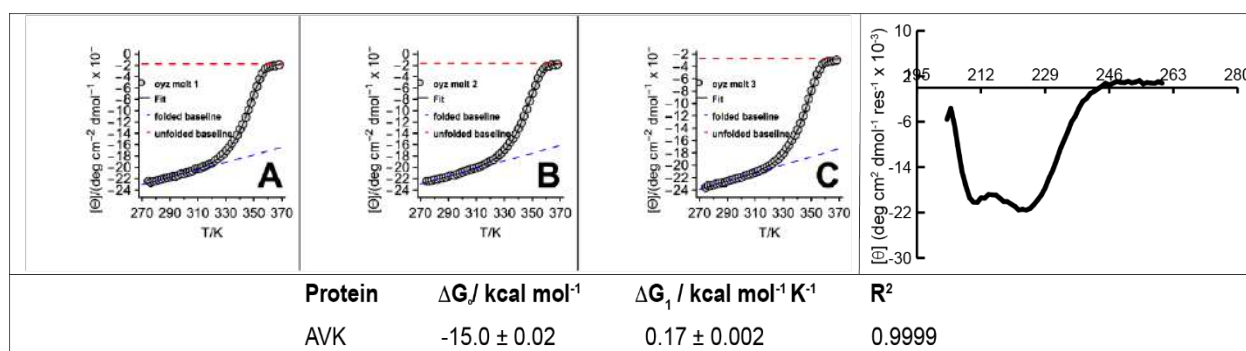


Figure 3-75. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein AVK in 20 mM sodium phosphate (pH 7), and 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

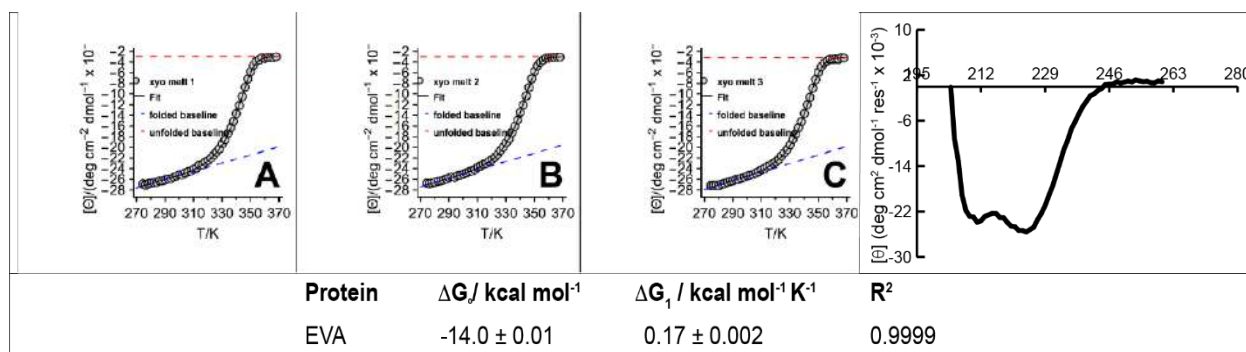


Figure 3-76. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein EVA in 20 mM sodium phosphate (pH 7), and 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

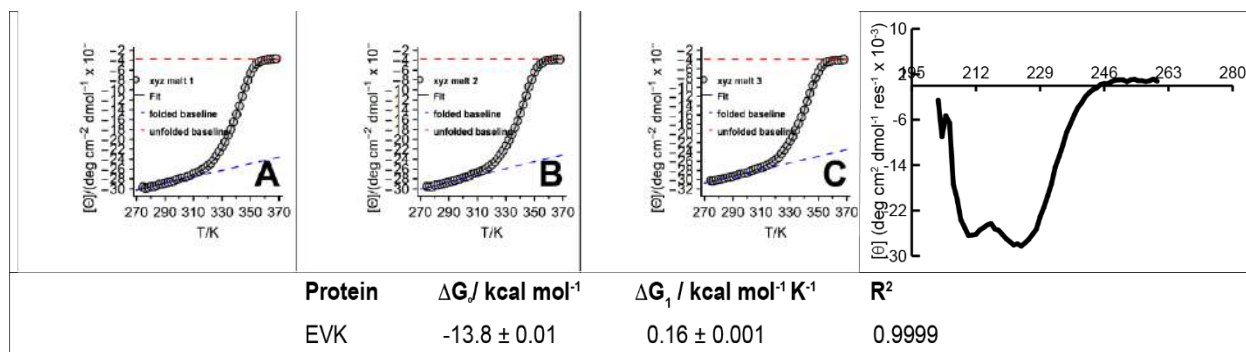


Figure 3-77. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **EVK** in 20 mM sodium phosphate (pH 7), and 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

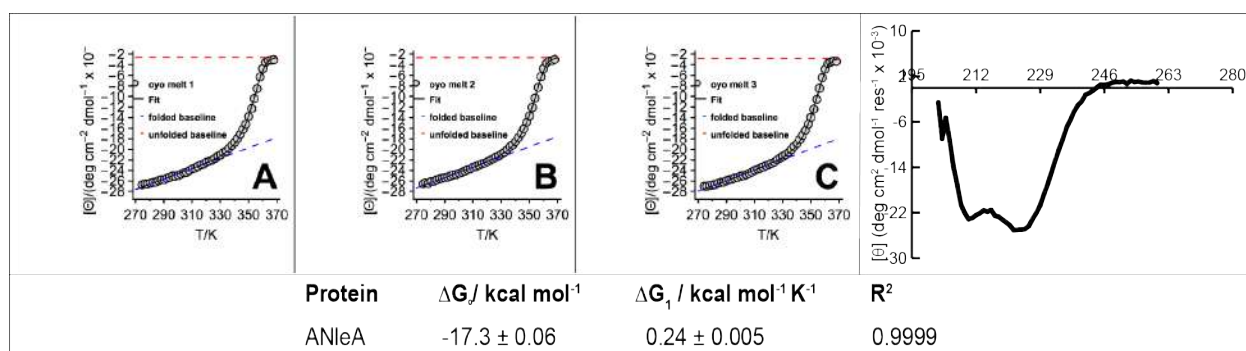


Figure 3-78. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **ANleA** in 20 mM sodium phosphate (pH 7), and 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

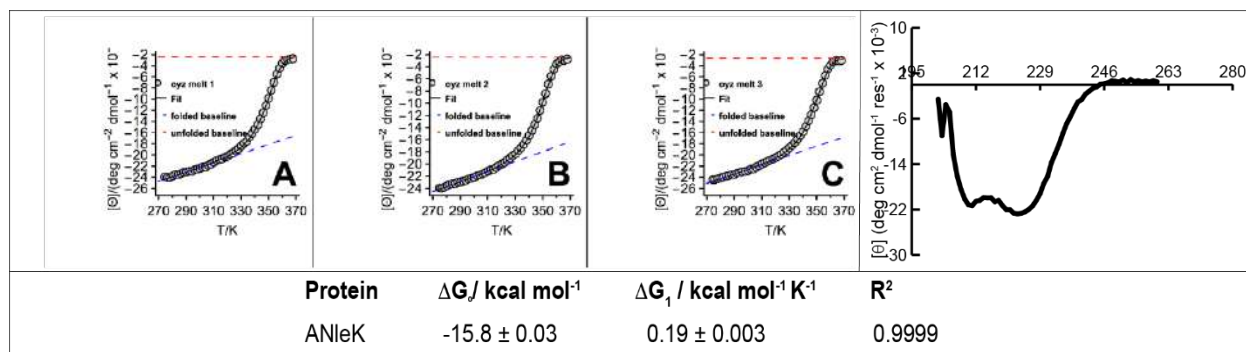


Figure 3-79. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **ANleK** in 20 mM sodium phosphate (pH 7), and 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

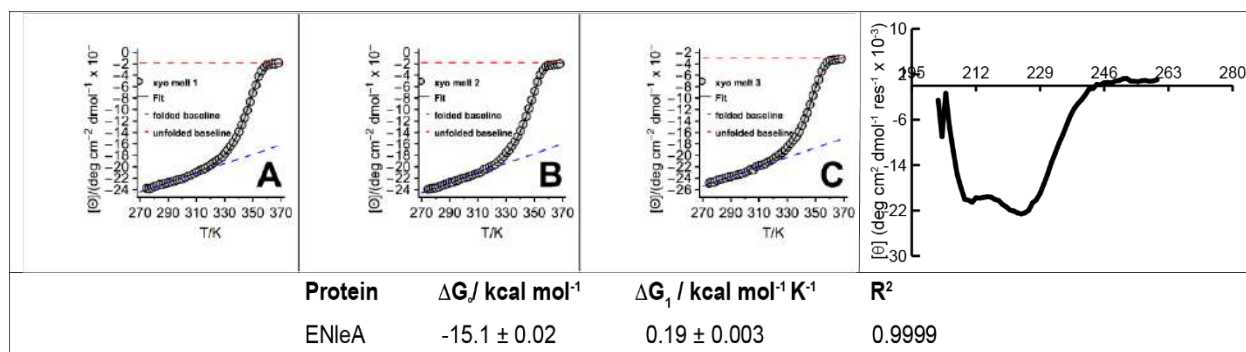


Figure 3-80. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein ENleA in 20 mM sodium phosphate (pH 7), and 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

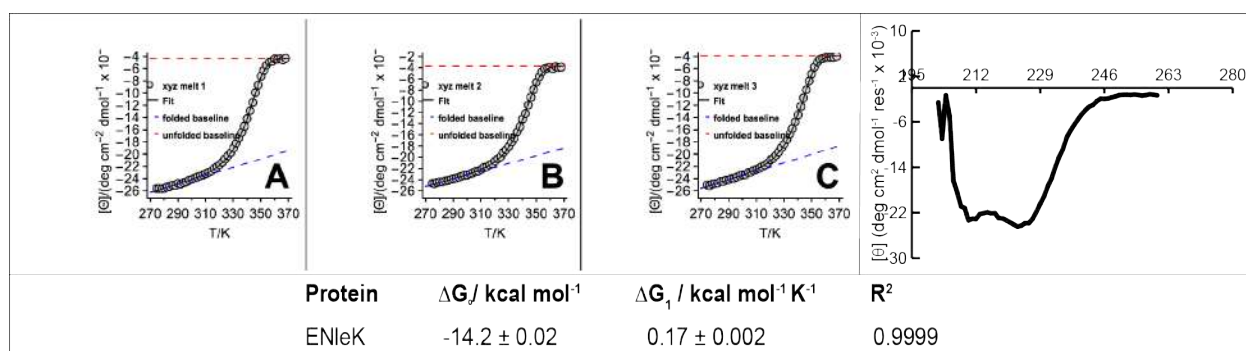


Figure 3-81. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein ENleK in 20 mM sodium phosphate (pH 7), and 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

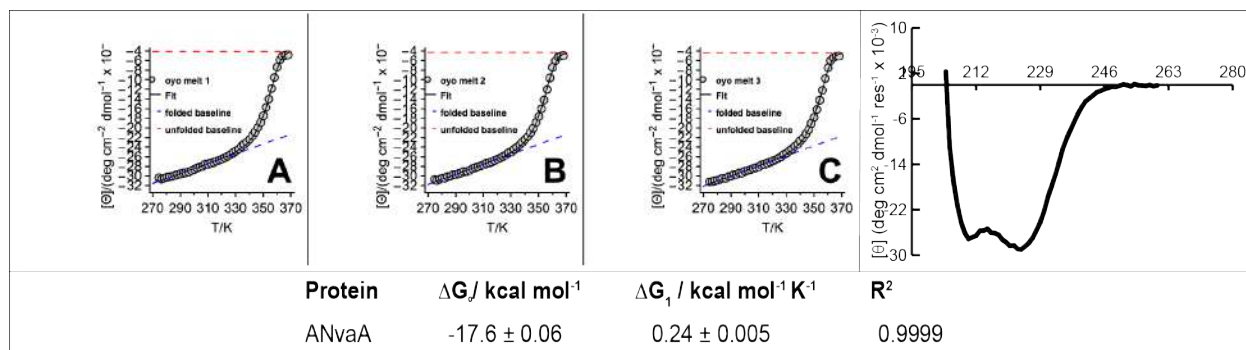


Figure 3-82. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein ANvaA in 20 mM sodium phosphate (pH 7), and 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

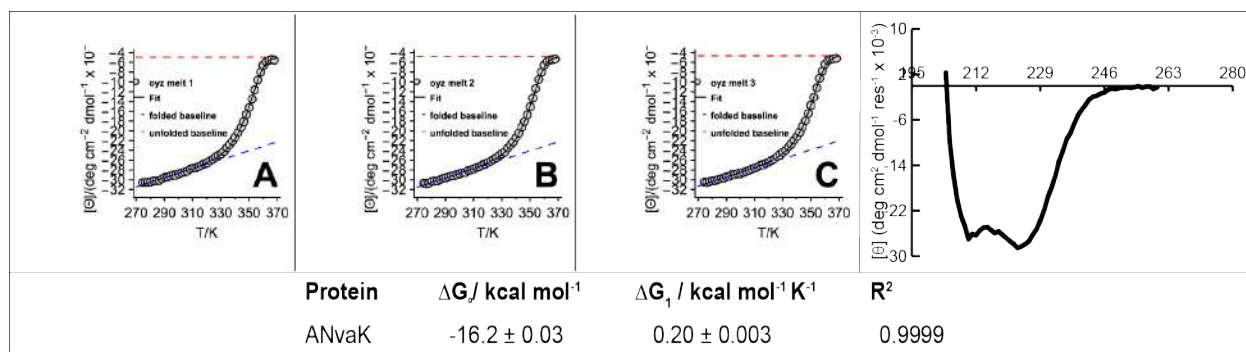


Figure 3-83. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein ANvaK in 20 mM sodium phosphate (pH 7), and 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

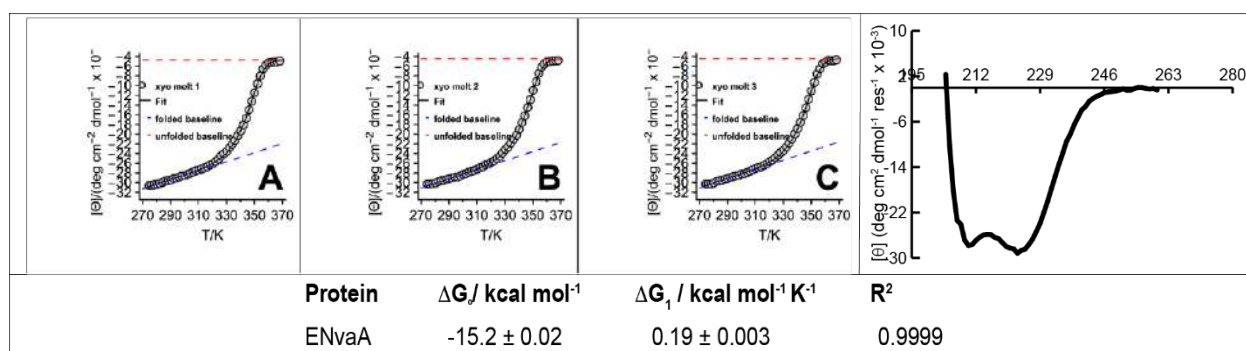


Figure 3-84. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein ENvaA in 20 mM sodium phosphate (pH 7), and 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

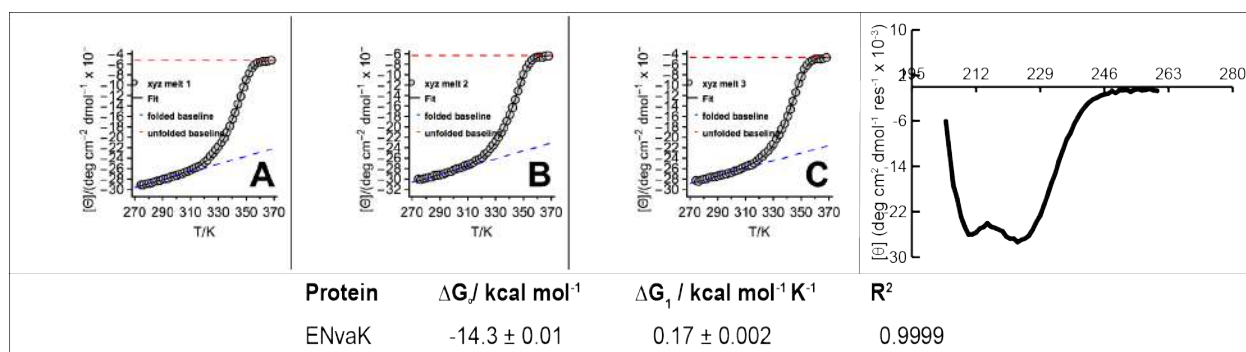


Figure 3-85. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein ENvaK in 20 mM sodium phosphate (pH 7), and 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

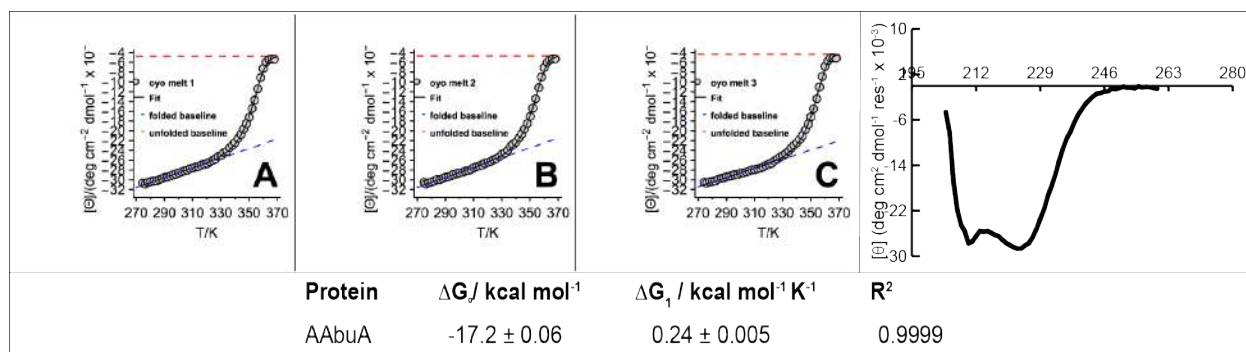


Figure 3-86. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **AAbuA** in 20 mM sodium phosphate (pH 7), and 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

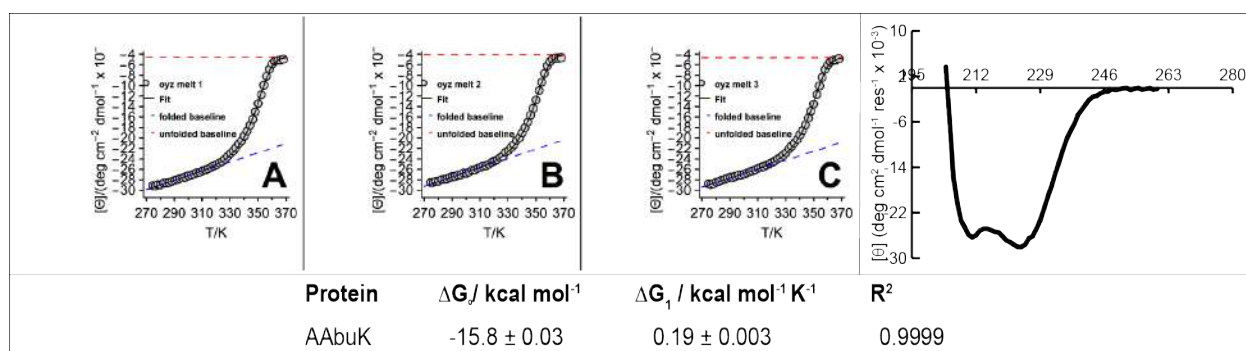


Figure 3-87. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **AAbuK** in 20 mM sodium phosphate (pH 7), and 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

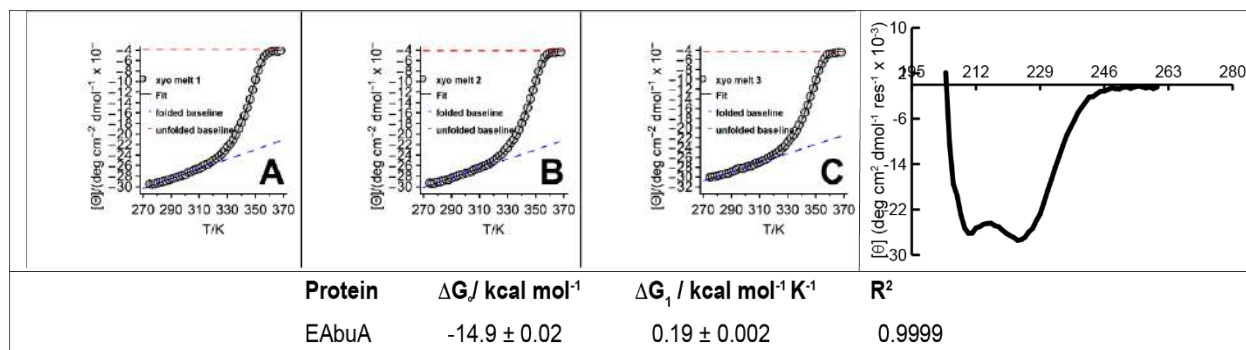


Figure 3-88. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **EAbuA** in 20 mM sodium phosphate (pH 7), and 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

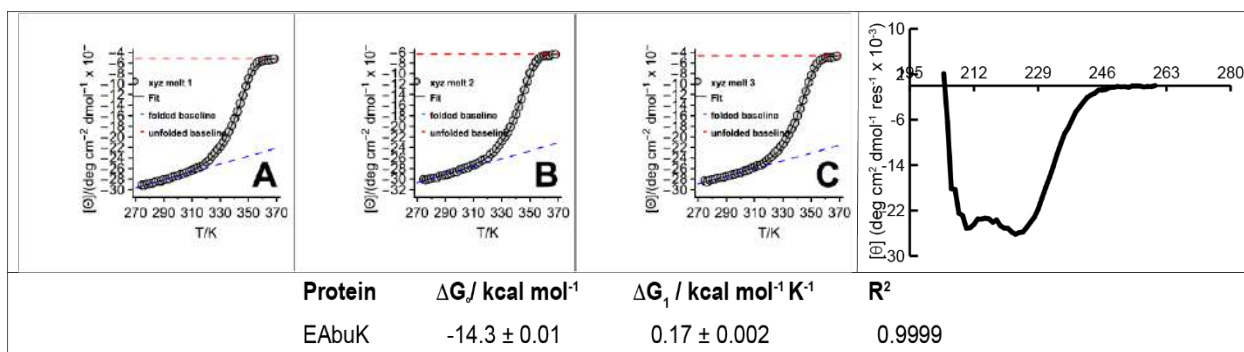


Figure 3-89. CD spectrum (top right) and triplicate variable temperature CD data (top left) for a 30 μM solution of protein **EAbuK** in 20 mM sodium phosphate (pH 7), and 1M urea. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

3.3.6 Single crystal diffraction data.

Freeze dried peptides were dissolved in milli Q water at a concentration of 10 mg/ml and screened against the Rigaku Wizard Cryo 1 and 2 screens. Drops were set with a TPP mosquito liquid handling robot with protein:condition ratios of 1:2, 1:1, and 2:1 at a total drop volume of 300nl. Conditions that grew high quality crystals are summarized in table 3-6. Crystals was harvested in ambient temperatures from sitting well plates with a nylon loop and immediately frozen in liquid N_2 . The crystal was mounted in a stream of cold N_2 and centered in the X-ray beam using a video camera. Low-temperature (100 K) X-ray diffraction data was collected using a MACH3 kappa goniometer coupled to a Bruker Apex II CCD detector with a Bruker-Nonius FR591 rotating anode X-ray source producing $\text{Cu } K_\alpha$ radiation ($\lambda = 1.54178 \text{ \AA}$). The Bruker Proteum-3 suite was used to process (integrate and scale) the data.

Structures were determined by molecular replacement with Phaser (CCP4 program suite) using the coordinates of the a coiled-coil trimer structure previously determined. The initial electron density map indicated two helices each being a helix in a coiled coil trimer; side-chain density was clearly interpretable. Model building was carried out using COOT.¹⁴ Refinement

was performed with Phenix.¹⁵ Data and refinement statistics are summarized in Tables 3-7, 3-8, 3-9, and 3-10.

Table 3-5 Crystal drop conditions

Protein	Precipitant	Buffer	pH	Salt
ALA	40% v/v PEG 300	100 mM Sodium phosphate dibasic/ Citric acid	7.5	0.2M NaCl
ELA	40% v/v PEG 300	100 mM Sodium phosphate dibasic/ Citric acid	4.2	
ALK	40% v/v PEG 300	100 mM Sodium phosphate dibasic/ Citric acid	4.2	
ELK	50% v/v PEG 200	100 mM HEPES free acid/ Sodium hydroxide	7.5	
EAK	50% v/v PEG 200	100 mM Sodium phosphate dibasic/ Potassium phosphate monobasic	6.2	0.2M NaCl
EFK	40% v/v PEG 300	100 mM Sodium phosphate dibasic/ Citric acid	4.2	
EVK	40% v/v PEG 600	100 mM Sodium citrate tribasic/ Citric acid	5.5	
ENvaK	40% v/v PEG 600	100 mM Sodium citrate tribasic/ Citric acid	5.5	
ENleK	40% v/v PEG 600	100 mM Sodium citrate tribasic/ Citric acid	5.5	

Table 3-6 Data and refinement statistics

	ALA	ELA	ALK
Space group	R3	R3	R3
Unit Cell			
a, b, c	37.86, 37.86, 105.85	38.04, 38.04, 104.28	38.18, 38.18, 102.77
α, β, γ	90, 90, 120	90, 90, 120	90, 90, 120
Data collection			
Resolution	31.32-1.90	34.77-1.70	31.48-1.68

R merge (%)	13.8 (34.7)	29.8(26.9)	37.3 (48.5)
I/ σ (I)	7.7 (1.1)	20.2 (2.5)	10.2 (1.2)
completeness	99.2 (92.5)	99.2 (88.3)	99.8 (99.1)
Redundancy	3.7 (2.3)	7.6 (3.5)	3.9 (2.1)
Refinement			
Resolution	31.32-1.90	34.77-1.70	31.48-1.68
# of reflections	4461	6237	6377
R _{work} /R _{free}	15.5/19.3	17.1/21.6	17.3/22.9
# of atoms	636	634	644
Ave B factor	17.4	15.4	15.2
Rmsd			
Bond lengths (Å)	0.007	0.006	0.006
Bond angles (°)	0.69	1.09	0.88

Table 3-7 Data and refinement statistics

	EAK	EFK	ELK
Space group	R3	R3	R3
Unit Cell			
a, b, c	39.01, 39.01, 98.16	38.89, 38.89, 103.47	39.06, 39.06, 101.52
α, β, γ	90, 90, 120	90, 90, 120	90, 90, 120
Data collection			
Resolution	32.72-2.30	34.49-2.15	33.85-1.90
R merge (%)	19.9 (36.4)	19.3 (40.6)	19.1 (53.2)

$I/\sigma(I)$	8.5 (1.6)	9.4 (2.1)	9.9 (1.4)
completeness	99.56 (100)	99.7 (100)	99.1 (94.8)
Redundancy	5.8 (4.4)	7.7 (4.4)	4.9 (2.5)
Refinement			
Resolution	32.72-2.30	34.49-2.15	33.85-1.95
# of reflections	2481	3180	4591
$R_{\text{work}}/R_{\text{free}}$	18.4/24.6	17.6/24.4	17.6/22.7
# of atoms	576	604	620
Ave B factor	28.3	20.5	15.4
Rmsd			
Bond lengths (Å)	0.008	0.008	0.011
Bond angles (°)	1.05	0.96	1.31

Table 3-8 Data and refinement statistics

	EVK	ENVAK	ENLEK
Space group	R3	R3	R3
Unit Cell			
a, b, c	38.14, 38.14, 104.49	38.32, 38.32, 104.59	39.07, 39.07, 113.33
α, β, γ	90, 90, 120	90, 90, 120	90, 90, 120
Data collection			
Resolution	34.84-1.5	34.87-1.70	37.78.-2.20

R merge (%)	15.8 (25.1)	10.5 (34.0)	11.6 (32.0)
I/ σ (I)	18.3 (1.9)	17.5 (1.9)	16.2 (2.1)
completeness	99.9 (97.6)	99.8 (98.3)	99.5 (95.4)
Redundancy	6.1 (2.8)	7.3 (4.7)	4.6 (2.2)
Refinement			
Resolution	34.84-1.5	34.87-1.70	37.78.-2.20
# of reflections	9118	6357	3261
R _{work} /R _{free}	17.5/19.7	16.1/18.9	19.7/26.6
# of atoms	670	667	559
Ave B factor	14.7	15.2	37.7
Rmsd			
Bond lengths (Å)	0.006	0.006	0.008
Bond angles (°)	1.03	1.00	1.03

Table 3-9 Data and refinement statistics

	EABUK
Space group	R3
Unit Cell	
a, b, c	39.37, 39.37, 98.12
α, β, γ	90, 90, 120
Data collection	

Resolution	32.72-1.6
R merge (%)	9.3 (28.4)
I/ σ (I)	16.3 (1.4)
completeness	99.9 (98.7)
Redundancy	3.6 (2.4)
Refinement	
Resolution	32.72-1.6
# of reflections	7495
R _{work} /R _{free}	19.2/21.4
# of atoms	638
Ave B factor	20.28
Rmsd	
Bond lengths (Å)	0.007
Bond angles (°)	1.04

3.4 References

1. Alkorta, I.; Quinonero, D.; Garau, C.; Frontera, A.; Elguero, J.; Deya, P. M., Dual cation and anion acceptor molecules. The case of the (eta(6)-C₆H₆)(eta(C₆F₆)-C-6)Cr(0) complex. *Journal of Physical Chemistry A* **2007**, *111* (16), 3137-3142.
2. Alkorta, I.; Elguero, J., Aromatic systems as charge insulators: Their simultaneous interaction with anions and cations. *Journal of Physical Chemistry A* **2003**, *107* (44), 9428-9433.
3. Alberti, M.; Aguilar, A.; Pirani, F., Cation-pi-anion interaction in alkali ion-benzene-halogen ion clusters. *J Phys Chem A* **2009**, *113* (52), 14741-8.
4. Garau, C.; Quinonero, D.; Frontera, A.; Ballester, P.; Costa, A.; Deya, P. M., Anion-pi interactions: must the aromatic ring be electron deficient? *New Journal of Chemistry* **2003**, *27* (2), 211-214.
5. Ngola, S. M.; Kearney, P. C.; Mecozzi, S.; Russell, K.; Dougherty, D. A., A selective receptor for arginine derivatives in aqueous media. Energetic consequences of salt bridges that are highly exposed to water. *J. Am. Chem. Soc.* **1999**, *121* (6), 1192-1201.
6. Trujillo, C.; Sanchez-Sanz, G.; Alkorta, I.; Elguero, J., Simultaneous Interactions of Anions and Cations with Cyclohexane and Adamantane: Aliphatic Cyclic Hydrocarbons as Charge Insulators. *Journal of Physical Chemistry A* **2011**, *115* (45), 13124-13132.

7. Cochrane, B.; Naumkin, F. Y., Reshaping and linking of molecules in ion-pair traps. *Chemical Physics Letters* **2016**, *643*, 137-141.
8. Aoyagi, S.; Nishibori, E.; Sawa, H.; Sugimoto, K.; Takata, M.; Miyata, Y.; Kitaura, R.; Shinohara, H.; Okada, H.; Sakai, T.; Ono, Y.; Kawachi, K.; Yokoo, K.; Ono, S.; Omote, K.; Kasama, Y.; Ishikawa, S.; Komuro, T.; Tobita, H., A layered ionic crystal of polar Li@C-60 superatoms. *Nature Chemistry* **2010**, *2* (8), 678-683.
9. Towse, C. L.; Akke, M.; Daggett, V., The Dynameomics Entropy Dictionary: A Large-Scale Assessment of Conformational Entropy across Protein Fold Space. *J Phys Chem B* **2017**, *121* (16), 3933-3945.
10. Rysavy, S. J.; Beck, D. A.; Daggett, V., Dynameomics: data-driven methods and models for utilizing large-scale protein structure repositories for improving fragment-based loop prediction. *Protein Sci* **2014**, *23* (11), 1584-95.
11. Scouras, A. D.; Daggett, V., The Dynameomics rotamer library: amino acid side chain conformations and dynamics from comprehensive molecular dynamics simulations in water. *Protein Sci* **2011**, *20* (2), 341-52.
12. Shu, J. Y.; Huang, Y.-J.; Tan, C.; Presley, A. D.; Chang, J.; Xu, T., Amphiphilic Peptide-Polymer Conjugates Based on the Coiled-Coil Helix Bundle. *Biomacromolecules* **2010**, *11* (6), 1443-1452.
13. Ciani, B.; Bjelic, S.; Honnappa, S.; Jawhari, H.; Jaussi, R.; Payapilly, A.; Jowitt, T.; Steinmetz, M. O.; Kammerer, R. A., Molecular basis of coiled-coil oligomerization-state specificity. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107* (46), 19850-19855.
14. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallographica Section D-Biological Crystallography* **2010**, *66*, 486-501.
15. Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D-Biological Crystallography* **2010**, *66*, 213-221.

4 AN ANION- π INTERACTION STRONGLY STABILIZES THE B-SHEET PROTEIN WW.

4.1 Introduction

Interactions between cationic amino acids and the face of an electron rich aromatic amino acids (cation- π) are ubiquitous in protein structure.¹⁻⁵ Experimental and computational work indicate that anions can have analogous interactions with the face of an electron deficient arene.⁶⁻¹² However, the energetic benefit of anion π interactions is a subject of debate,¹³⁻¹⁴ especially for anion- π interactions with the electron-rich amino acid side chains Phe, Tyr, and Trp.¹⁵

Unsurprisingly, one would expect the facewise interaction between an electron-rich arene and an anion to be unfavorable because of electrostatic repulsion,¹¹ though arene polarization can partially compensate for less than optimal electrostatic interactions.⁷ Interestingly, analysis of protein structures in the Protein Data Bank (PDB) show that facewise contacts of anions with the aryl side chains of Phe, Trp, or Tyr do occur despite the debate whether this type of interaction is stabilizing.¹⁵⁻¹⁹

Alternative studies suggest an anion interacting with the electron poor C-H edge of electron rich arene occurs more often and has a more favorable energetic impact.^{7, 18-19} For example, Kallenbach and co-workers found that an *i*-position Glu and an *i* + 4 Phe interaction contributes about -0.5 kcal mol⁻¹ in conformational stability to an alpha helix model peptide; relatively weak

NOEs show that the Glu and Phe side chains are in close proximity, possibly consistent with an interaction between the aryl protons of Phe and the negatively charged oxygens of Glu.²⁰⁻²¹

4.2 Results and discussion

A statistical probe of protein structures in the PDB for facewise anion π interactions showed, among others, two interactions where in a reverse turn an *i*-residue anion interacts with an *i*+2 aromatic residue.¹⁷ The first example, fibronectin is a protein that plays a role in cell adhesion by binding integrin proteins.²² The fibronectin crystal structure (PDB 2QBW) shows that Asp179, found in a reverse turn, is interacting with the face of Tyr181 (Figure 4-1A).²³ β -galactosidase, an enzyme that breaks down polysaccharides, also has a reverse turn in its structure (PDB 1JZ8) where Asp997 interacts with the face of Trp 999 (Figure 4-1B).²⁴⁻²⁵ We wondered if we could assess the energetic contribution of an anion- π interaction with the well-characterized WW domain. The WW domain (WW) is a 33 amino acid protein that has three antiparallel beta sheets connected by two reverse turns.²⁶⁻³¹ The first reverse turn in WW contains positions 16 and 18 whose side chains are in close proximity (Figure 4-1C, PDB 2f21). We thought that placing a negatively charged amino acid at position 16 and an aromatic amino acid at position 18 might facilitate a face-wise anion- π interaction similar to that observed in the context of fibronectin, and β -galactosidase.

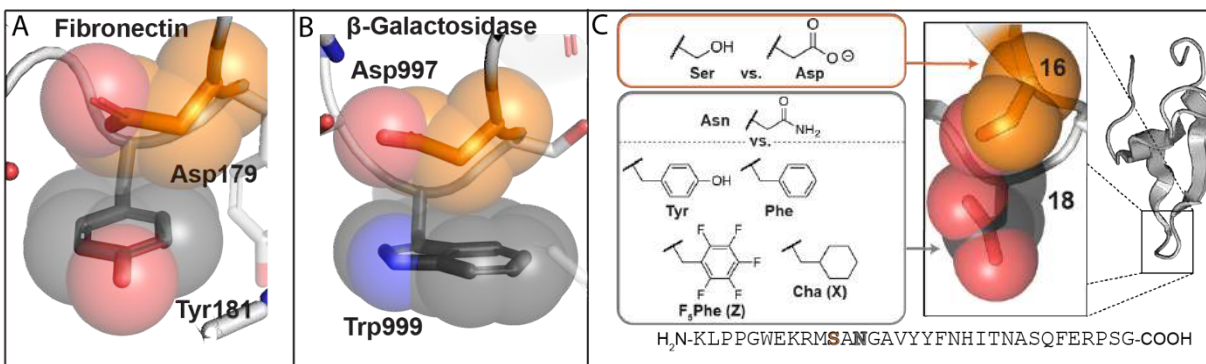


Figure 4-1. (A) Fibronectin face wise anion- π interaction between Asp179 and Tyr181, PDB ID 2QBW. (B) β -galactosidase face wise anion- π interaction between Asp997 and Trp999, PDB ID 1JZ8. (C) Close proximity of residues 16 and 18 in a reverse turn within the Pin WW domain (PDB: 2f21). We made the substitutions shown at the positions indicated.

We tested this hypothesis by making WW variant **DY** in which position 16 is an Asp and position 18 is a Tyr, along with sequence-matched control compounds **SY**, **DN**, and **SN**, where we switched Asp16 to Ser and Tyr18 to Asn, respectively, in all possible combinations. We chose these replacements because Ser16 and Asn18 cannot participate in an anion- π interaction and because they closely resemble the residues that occupy these positions in the parent WW sequence from which these variants were derived, these variants also complete a double mutant cycle allowing us to measure the Asp:Tyr interaction (see Chapter 1.2.1).³²

Variable temperature CD experiments allowed us to assess the conformational stability of **DY** relative to **SY**, **DN**, and **SN** (Table 4-1). Mutating Ser16 to Asp increases WW conformational stability by -0.26 ± 0.04 kcal mol⁻¹ when Asn occupies position 18 (which would preclude any contribution from an anion- π interaction). In contrast, when Tyr occupies position 18, the S16D mutation stabilizes WW by a moderately larger amount ($\Delta\Delta G_f = -0.52 \pm 0.03$ kcal/mol), suggesting that an Asp16-Tyr18 interaction contributes slightly to the stability of **DY** relative to **DN** ($\Delta\Delta\Delta G_f = -0.26 \pm 0.04$ kcal/mol; see Table 4-1).

We wondered if other proteinogenic arenes might more readily engage in a stabilizing interaction with Asp16. To test this hypothesis, we prepared WW variants **SW**, **DW**, **SH**, **DH**, **SF**, and **DF**, in which Trp, His, or Phe have replaced Tyr18. Variable temperature CD data for **SW** were inconsistent with a two-state folding protein, preventing us from determining the strength of the Asp16–Trp18 interaction. Comparison of **SH** and **DH** vs **SN** and **DN** indicates that the Asp16–His18 interaction is only marginally stabilizing ($\Delta\Delta\Delta G_f = -0.12 \pm 0.02$ kcal/mol, Table 4-1). In contrast, the Asp16–Phe18 interaction contributes nearly 5 times more conformational stability than the Asp16–Tyr18 interaction (-1.31 ± 0.05 kcal/mol, compare ΔG_f values for **SF** and **DF** vs **SN** and **DN** in Table 4-1).

These results show that the Asp16–Phe18 interaction stabilizes WW, But the exact orientation of Asp16 relative to Phe18 remains unclear. If the Asp16–Phe18 interaction has a facewise orientation Asp16 packs close to the center Phe18, then replacing Phe with electron-poor pentafluorophenylalanine (F₅Phe; one-letter abbreviation is Z) should result in a more energetically favorable interaction. In contrast, an edgewise interaction of Asp16 with F₅Phe should be much less favorable, due to the electron rich fluorines of F₅Phe repelling the Asp16 carboxylate.

We tested this hypothesis by preparing WW variants **DZ** and **SZ**, in which we replaced Phe18 with F₅Phe. Surprisingly, the Asp16–F₅Phe18 interaction contributes -1.33 ± 0.03 kcal/mol to the stability of **DZ** relative to **SZ**, the same amount at the Asp16:Phe18 interaction. This result potentially indicates an edgewise orientation of Asp16 relative to Phe18 in **DF** and a facewise orientation of Asp16 relative to F₅Phe18 in **DZ** and could indicate that anion– π protein interactions are dynamic and permit local conformational changes that optimize interaction energetics.

Finally, we explored the role of aromaticity in the Asp16–Phe18 interaction by preparing variants **DX** and **SX**, where we incorporated the nonaromatic residue cyclohexylalanine (Cha; one-letter abbreviation is X) at position 18. The interaction of Asp16 with Cha18 does not contribute a significant amount to WW stability (0.08 ± 0.02 kcal/mol), indicating aromaticity of the side chain at position 18 is essential for an anion- π interaction with position 16

Table 4-1. Folded Free energies of WW proteins SN, DN and their Sequence Variants at 60 °C^a

Protein	Position 16	Position 18	T _m (°C)	$\Delta G(\text{kcal/mol})$	$\Delta\Delta G(\text{kcal/mol})$
SN	Ser	Asn	65.7± 0.1	-0.53± 0.01	
DN	Asp	Asn	68.7± 0.1	-0.80± 0.01	
SY	Ser	Tyr	58.3± 0.3	0.15± 0.03	
DY	Asp	Tyr	64.2± 0.2	-0.37± 0.02	-0.26± 0.04
SH	Ser	His	60.8± 0.1	-0.06± 0.01	
DH	Asp	His	65.4± 0.1	-0.45± 0.01	-0.12± 0.02
SF	Ser	Phe	52.8± 0.2	1.05± 0.04	
DF	Asp	Phe	65.9± 0.1	-0.53± 0.01	-1.31± 0.05
SZ	Ser	F ₅ Phe	49.7± 0.1	1.33± 0.03	
DZ	Asp	F ₅ Phe	62.6± 0.1	-0.26± 0.01	-1.33± 0.03
SX	Ser	Cha	58.4± 0.1	0.22± 0.01	

DX	Asp	Cha	61.3± 0.1	-0.13± 0.01	-0.08± 0.02
----	-----	-----	-----------	-------------	-------------

^a Folded free energies are given ± standard error in kcal/mol at 60 °C (333.15 K) in 20 μM sodium Phosphate buffer (pH7). Z = pentafluorophenylalanine; X = cyclohexylalanine

4.3 Conclusions

Here, we have shown that anionic Asp can interact favorably with Phe to increase the stability of the WW domain. Our results complement previous computational predictions^{6, 9-10, 33-35} along with earlier studies of an α -helical model system,²⁰⁻²¹ suggesting that the anion- π interaction should be considered as an important addition to the tool box of noncovalent interactions used to understand protein folding, to design new proteins, and to develop small-molecule effectors of protein function.

4.4 Supporting information

4.4.1 Protein Synthesis

WW peptide variants **SN, DN, SY, DY, SY, DY, SF, DF, DZ, SZ, SX, DX** (sequences shown in main text, Z = pentafluorophenylalanine, X = cyclohexylalanine) were synthesized as C-terminal acids (Fmoc-Gly-Wang resin, EMD Millipore), by microwave-assisted solid-phase peptide synthesis, using a standard Fmoc N α protection strategy as described previously in chapter 2 section 2.4.1.

4.4.2 Protein Purification and Characterization

Immediately prior to purification, the crude protein was dissolved in 1:1 H₂O/MeCN. Proteins were purified by preparative reverse-phase high performance liquid chromatography (HPLC) on a C18 column using a linear gradient of water in acetonitrile with 0.1% v/v TFA. Fractions containing the desired protein product were pooled, frozen, and lyophilized. Proteins were identified by electrospray ionization time of flight mass spectrometry (ESI-TOF); mass spectra appear below in Figures S1–S12. Protein purity was assessed by Analytical HPLC (Figures S13–S24).

ESI-TOF spectra for proteins WW peptide variants **SN**, **DN**, **SY**, **DY**, **SY**, **DY**, **SF**, **DF**, **DZ**, **SZ**, **SX**, **DX** are shown in Figures 4-2–4-11.

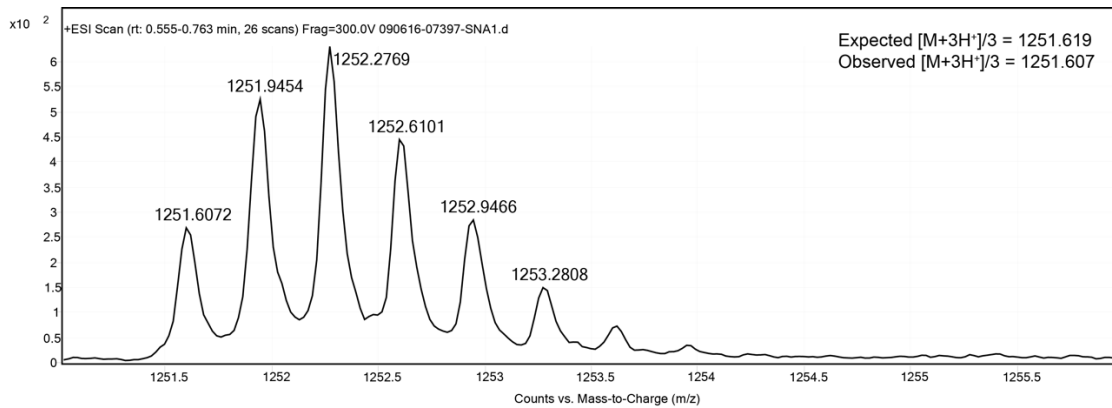


Figure 4-2. ESI TOF spectrum for SN. Expected [M+3H]³⁺/3 = 1251.619 Da. Observed [M+3H]³⁺/3 = 1251.607 Da.

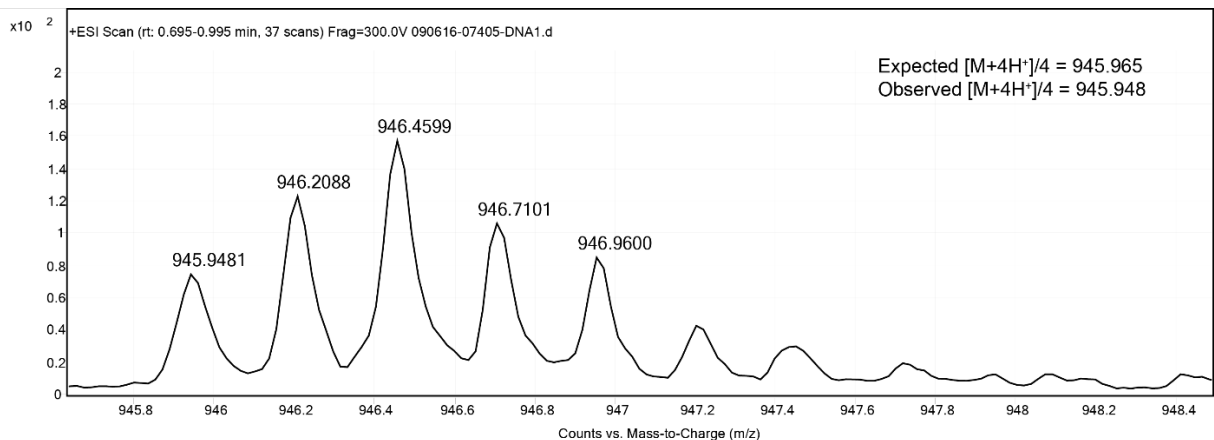


Figure 4-3. ESI TOF spectrum for **DN**. Expected $[M+4H^+]/4 = 945.965$ Da. Observed $[M+4H^+]/4 = 945.948$ Da.

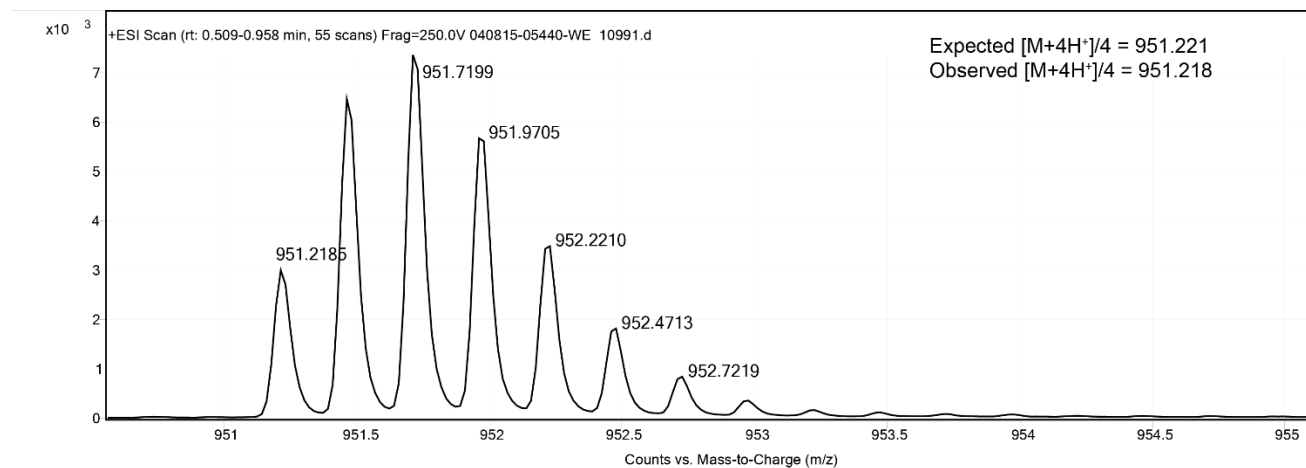


Figure 4-4. ESI TOF spectrum for **SY**. Expected $[M+4H^+]/4 = 951.221$ Da. Observed $[M+4H^+]/4 = 951.218$ Da.

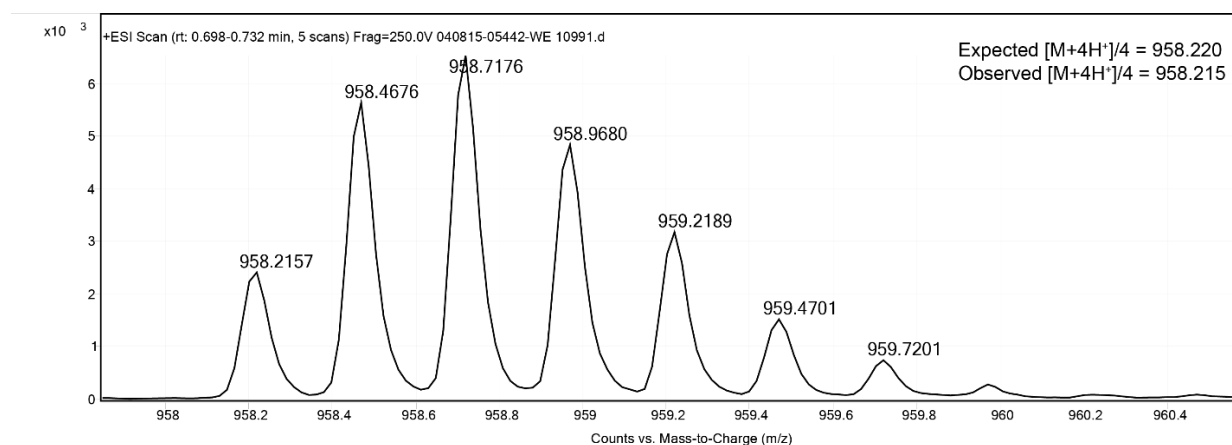


Figure 4-5. ESI TOF spectrum for **DY**. Expected $[M+4H^+]/4 = 958.220$ Da. Observed $[M+4H^+]/4 = 958.215$ Da.

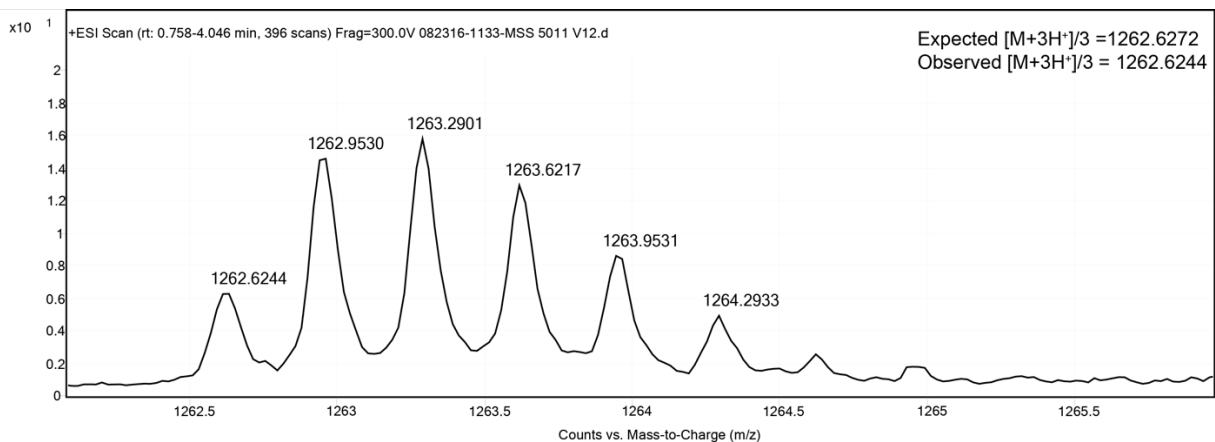


Figure 4-6. ESI TOF spectrum for **SF**. Expected $[M+3H^+]/3 = 1262.6272$ Da. Observed $[M+3H^+]/3 = 1262.6244$ Da.

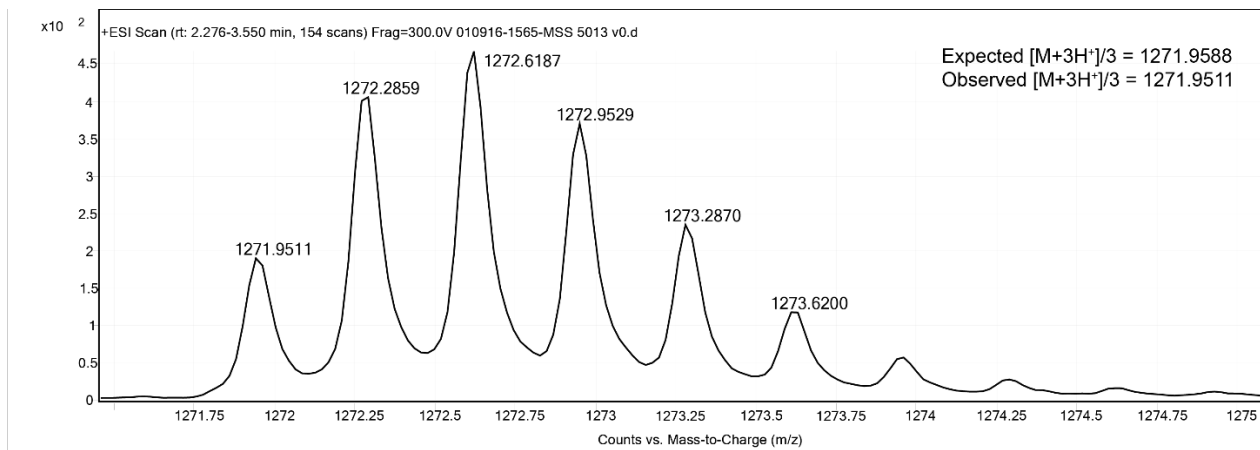


Figure 4-7. ESI TOF spectrum for **DF**. Expected $[M+3H^+]/3 = 1271.9588$ Da. Observed $[M+3H^+]/3 = 1271.9511$ Da.

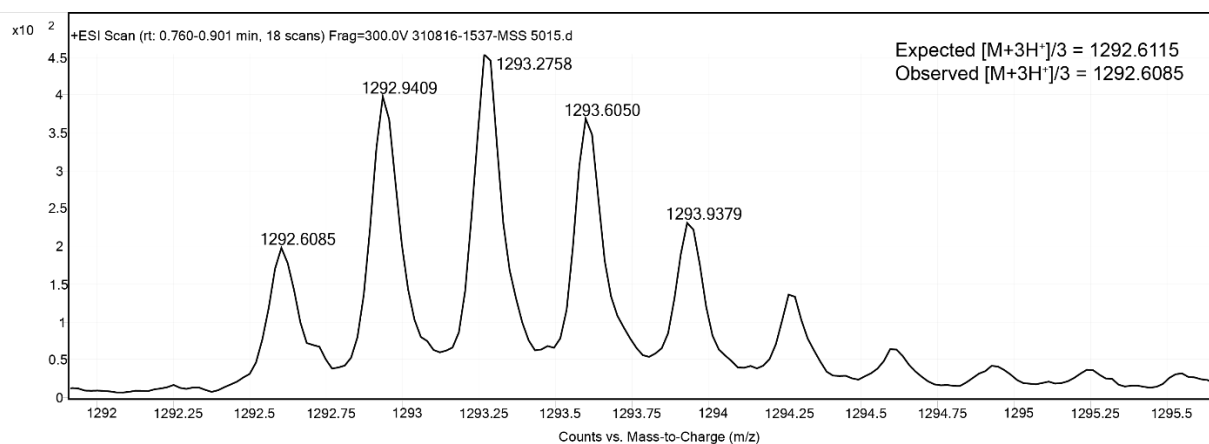


Figure 4-8. ESI TOF spectrum for **SZ**. Expected $[M+3H^+]/3 = 1292.6115$ Da. Observed $[M+3H^+]/3 = 1292.6085$ Da ($Z =$ pentafluorophenylalanine).

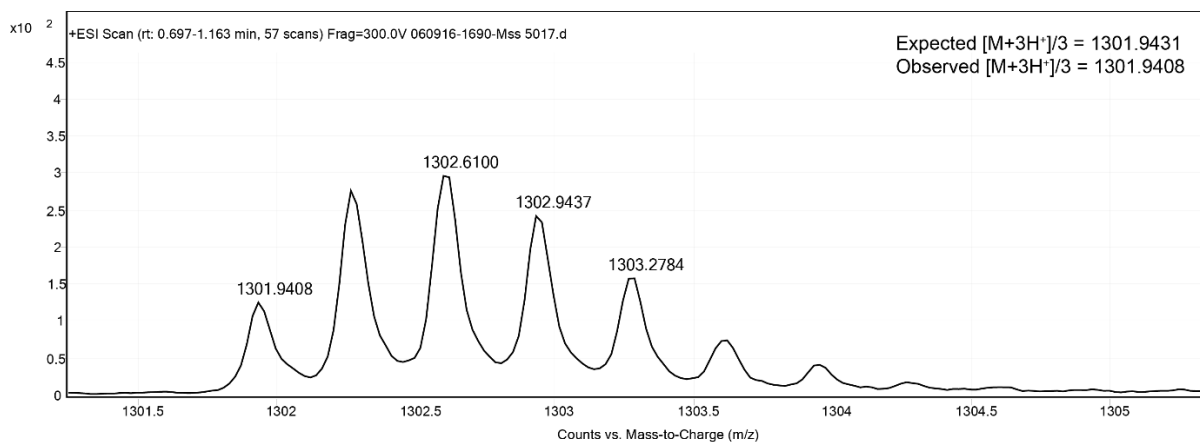


Figure 4-9. ESI TOF spectrum for **DZ**. Expected $[M+3H^+]/3 = 1301.9431$ Da. Observed $[M+3H^+]/3 = 1301.9408$ Da ($Z =$ pentafluorophenylalanine).

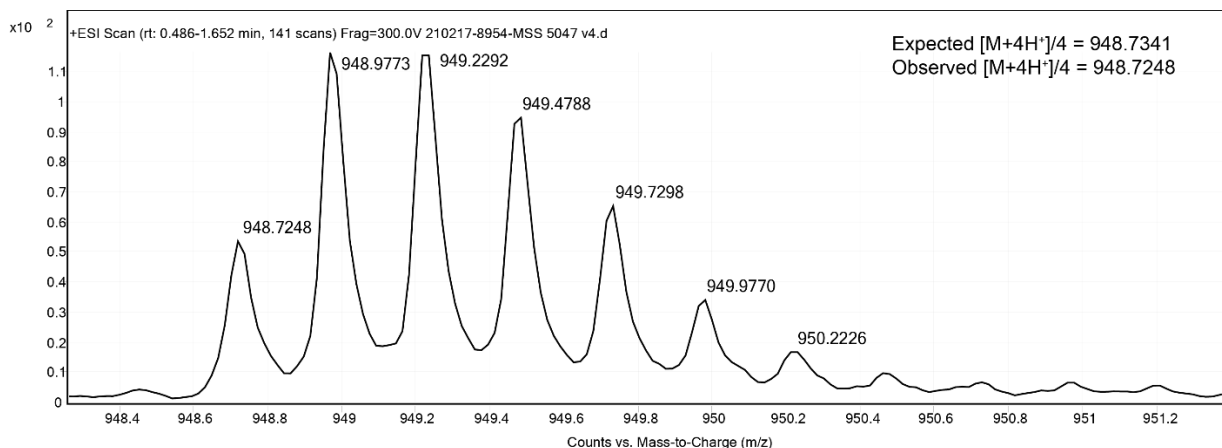


Figure 4-10. ESI TOF spectrum for **SX**. Expected $[M+4H^+]/4 = 948.7341$ Da. Observed $[M+4H^+]/4 = 948.7248$ Da ($X =$ cyclohexylalanine).

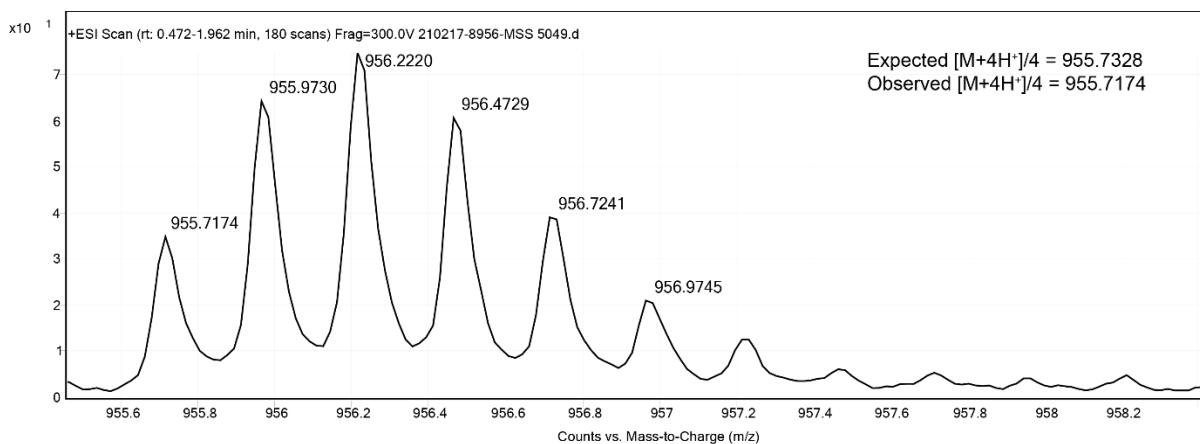


Figure 4-11. ESI TOF spectrum for **DX**. Expected $[M+4H^+]/4 = 955.7328$ Da. Observed $[M+4H^+]/4 = 955.7174$ Da ($X =$ cyclohexylalanine).

HPLC traces for proteins SN, DN, SY, DY, SF, DF, SZ, DZ, SX, DX are shown in Figures 4-12–4-21.

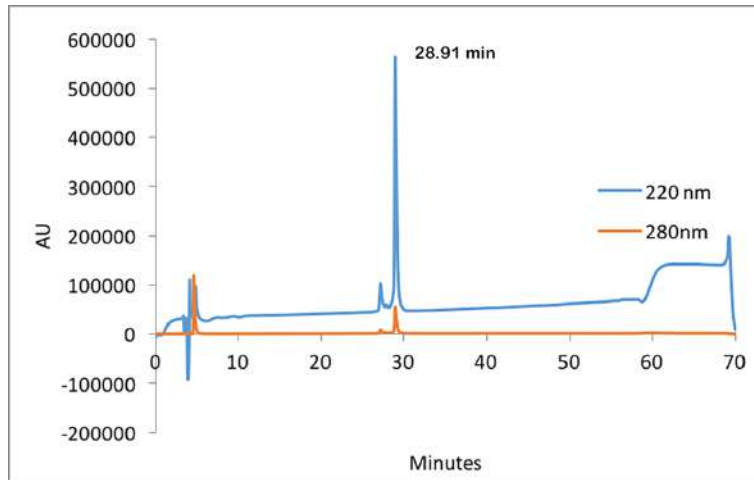


Figure 4-12. Analytical HPLC Data for SN. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

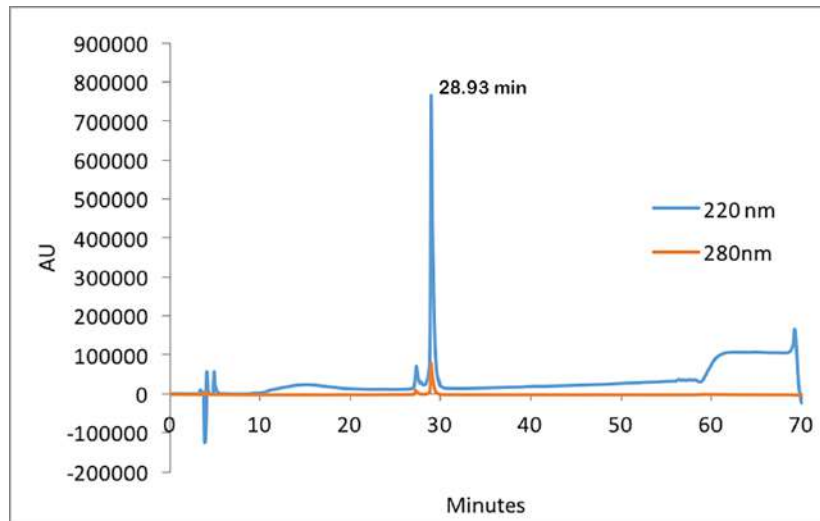


Figure 4-13. Analytical HPLC Data for DN. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

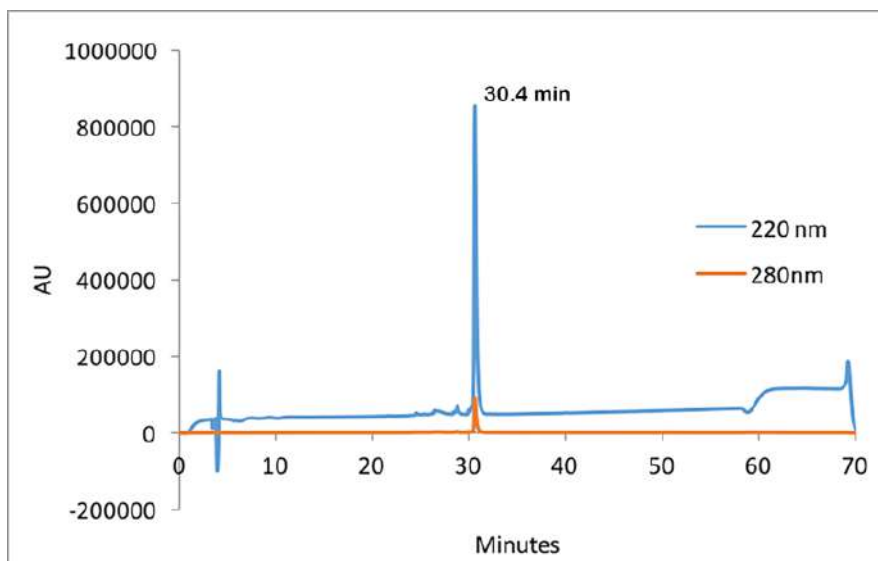


Figure 4-14. Analytical HPLC Data for **SY**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

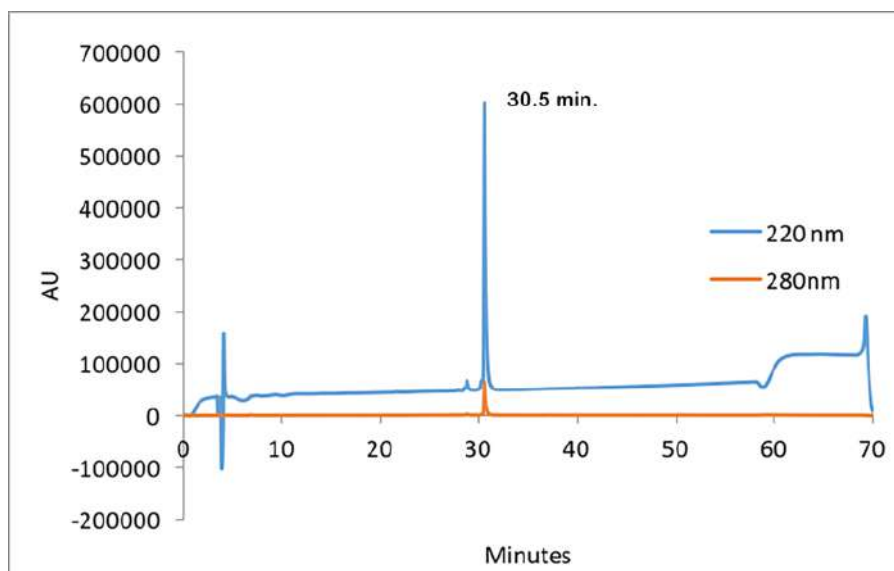


Figure 4-15. Analytical HPLC Data for **DY**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

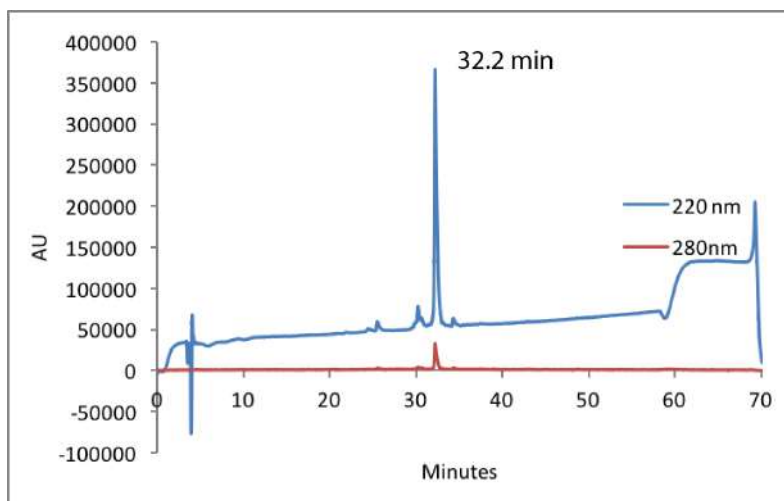


Figure 4-16. Analytical HPLC Data for **SF**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

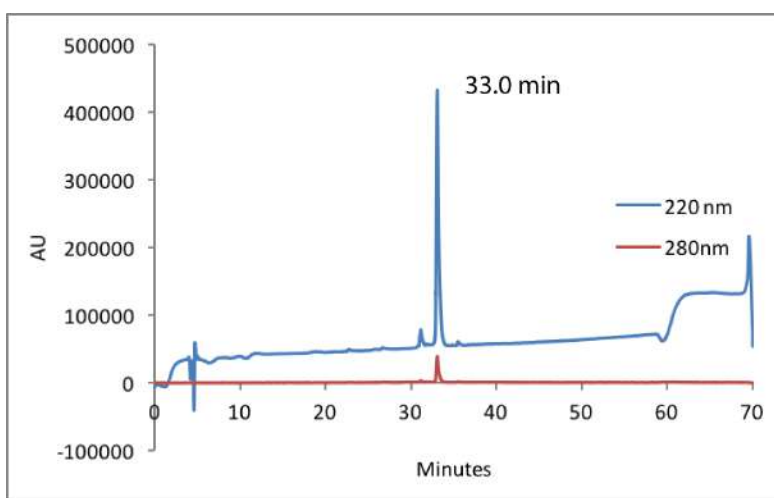


Figure 4-17. Analytical HPLC Data for **DF**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

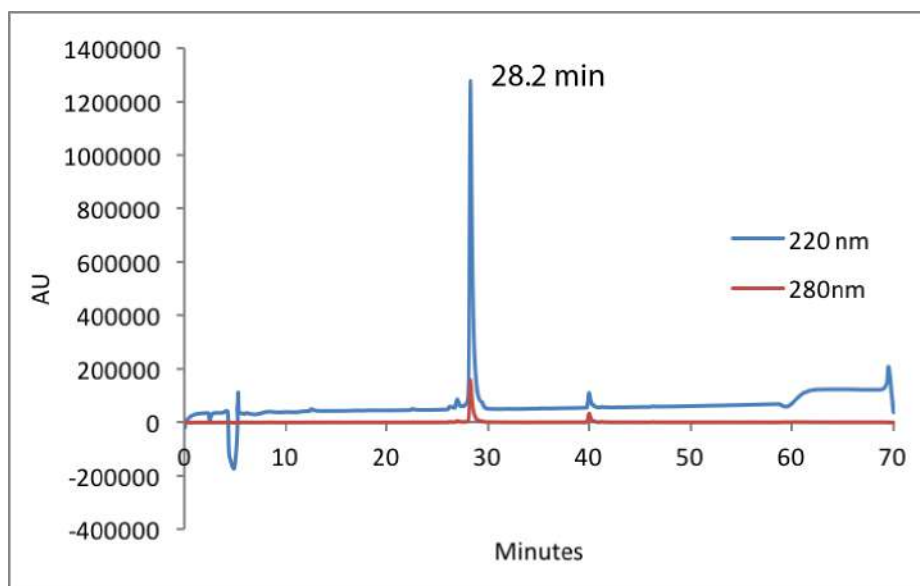


Figure 4-18. Analytical HPLC Data for **SZ**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min. (Z = pentafluorophenylalanine)

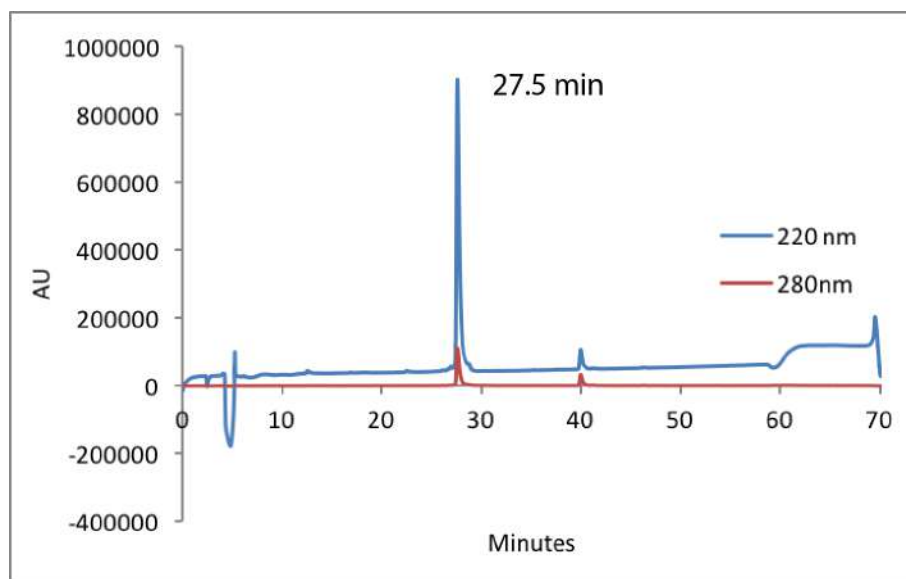


Figure 4-19. Analytical HPLC Data for **DZ**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min. (Z = pentafluorophenylalanine)

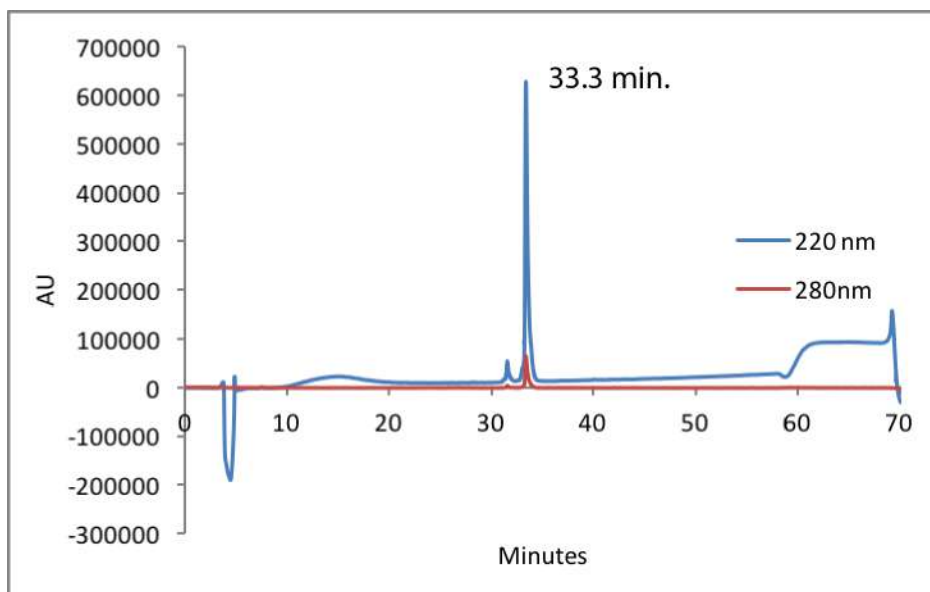


Figure 4-20. Analytical HPLC Data for **SX**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min. (X = cyclohexylalanine)

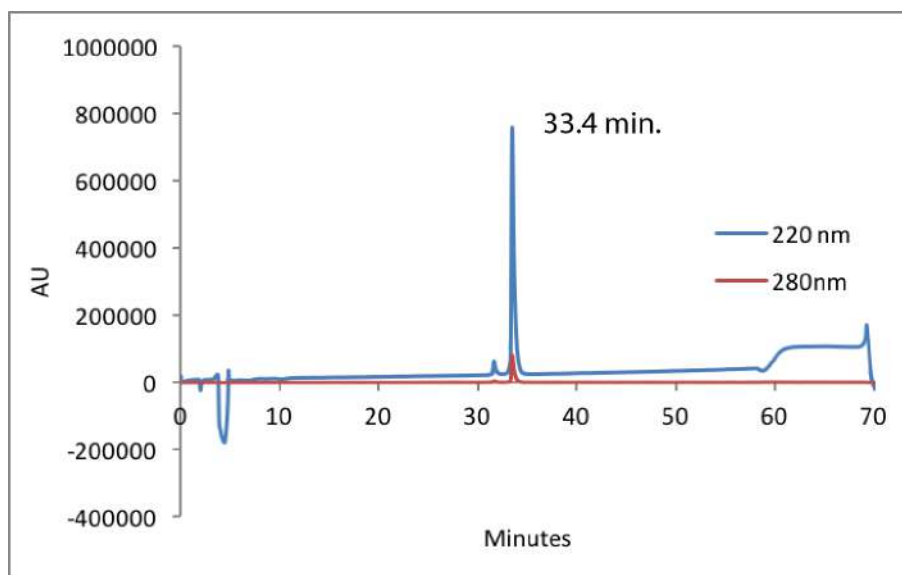


Figure 4-21. Analytical HPLC Data for **DX**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min. (X = cyclohexylalanine)

4.4.3 Circular Dichroism Spectropolarimetry

Measurements were made with an Aviv 420 Circular Dichroism Spectropolarimeter, using quartz cuvettes with a path length of 0.1 cm. Protein solutions were prepared in 20 mM sodium phosphate buffer, pH 7, and protein concentrations were determined spectroscopically based on tyrosine and tryptophan absorbance at 280 nm in 8 M guanidine hydrochloride + 20 mM sodium phosphate ($\epsilon_{\text{Trp}} = 5690 \text{ M}^{-1}\text{cm}^{-1}$, $\epsilon_{\text{Tyr}} = 1280 \text{ M}^{-1}\text{cm}^{-1}$).³⁶ CD spectra of 50 μM solutions of **SN**, **DN**, **SY**, **DY**, **SF**, **DF**, **SZ**, **DZ**, **SX**, and **DX** were obtained from 260 to 200 nm at 25°C. Variable temperature CD data were obtained at least in triplicate by monitoring the molar ellipticity at 227 nm of 50 μM solutions of in 20 mM sodium phosphate (pH 7) from 1 to 95°C at 2 °C intervals, with 120 s equilibration time between data points and 30 s averaging time.

The folding equilibria of **SN**, **DN**, **SY**, **DY**, **SF**, **DF**, **SZ**, **DZ**, **SX**, and **DX** do not involve self-association into quaternary structures; rather, an unfolded monomer proceeds to a folded monomer via a single high-energy transition state, with the position of the equilibrium determined by folding equilibrium constant K_f :



We used the following equation to fit the variable temperature CD data:

$$[\theta] = \frac{(u_0 + u_1 T) + (f_0 + f_1 T) K_f}{1 + K_f} \quad (\text{S11})$$

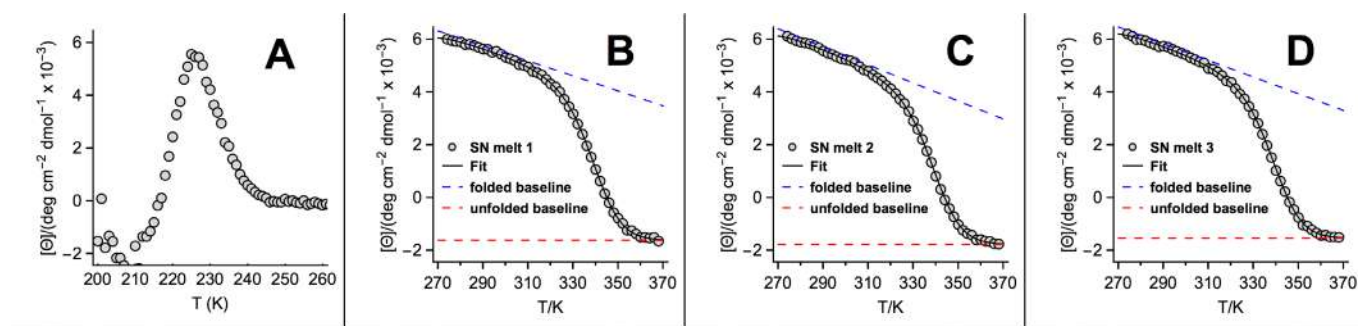
T is the temperature in Kelvin; u_0 and u_1 are the intercept and slope of the post-transition baseline, respectively; f_0 and f_1 are the intercept and slope of the pre-transition baseline, respectively; and K_f is the folding equilibrium constant as defined by equations S12 and S13:

$$K_f = \exp\left(\frac{-\Delta G_f}{RT}\right) \quad (\text{S12})$$

$$\Delta G_f = \frac{\Delta H(T_m) \cdot (T_m - T)}{T_m} + \Delta C_p \cdot (T - T_m - T \cdot \ln \left[\frac{T}{T_m} \right]), \quad (S13)$$

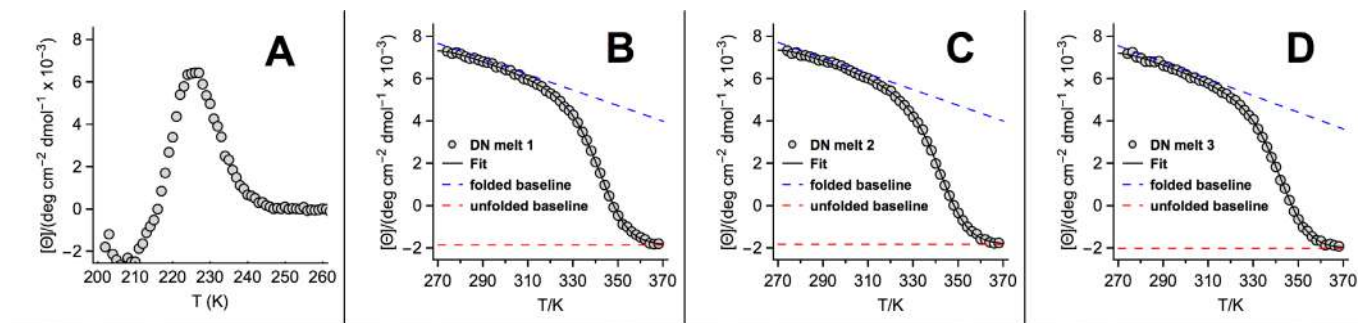
where R is the universal gas constant (0.0019872 kcal/mol/K), and the fit parameters are T_m (the midpoint of the unfolding transition; the temperature at which $\Delta G_f = 0$), $\Delta H(T_m)$, the change in enthalpy upon folding at T_m ; and ΔC_p , the change in heat capacity upon folding. In some cases, standard errors indicated that certain parameters (typically ΔC_p and/or u_1) were unlikely to be significant (p-value > 0.01). When this happened, we fit the data again with modified equations that eliminated the insignificant parameters.

We used parameters from the fits of the variable temperature CD data for **SN, DN, SY, DY, SF, DF, SZ, DZ, SX, and DX** variants to calculate the folding free energy values presented in the main text. CD spectra and variable temperature CD data for **SN, DN, SY, DY, SF, DF, SZ, DZ, SX, and DX** are shown in Figures 4-21–4-40, along with the parameters that were used to generate global fits for each compound. The standard error for each fitted parameter is also shown.



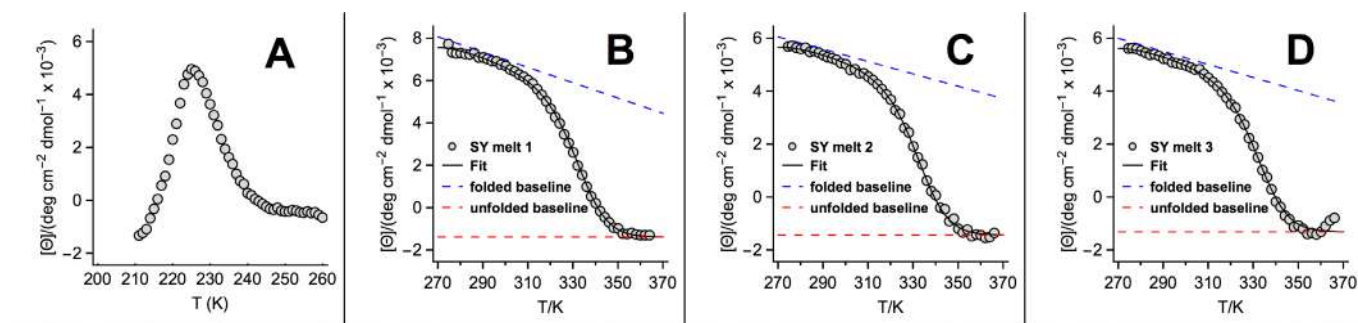
SN Parameters	Values	Units	P-Values
ΔH_f at T_m	-33.6 ± 0.3	kcal/mol	<0.001
T_m	65.7 ± 0.1	°C	<0.001
ΔC_p	-0.67 ± 0.02	kcal/mol/K	<0.001
a1	14.0 ± 0.2	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
b1	-0.0284 ± 0.0007	deg cm ² dmol ⁻¹ K ⁻¹ x 10 ⁻³	<0.001
c1	-1.62 ± 0.01	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
d1	---	---	---
a2	15.6 ± 0.2	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
b2	-0.0341 ± 0.0007	deg cm ² dmol ⁻¹ K ⁻¹ x 10 ⁻³	<0.001
c2	-1.78 ± 0.01	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
d2	---	---	---
a3	15.0 ± 0.2	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
b3	-0.0316 ± 0.0007	deg cm ² dmol ⁻¹ K ⁻¹ x 10 ⁻³	<0.001
c3	-1.54 ± 0.01	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
d3	---	---	---

Figure 4-22. (A) CD spectrum and (B)–(D) variable temperature CD data for 50 μ M solutions of protein SN in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated; $R^2 = 0.9999$.



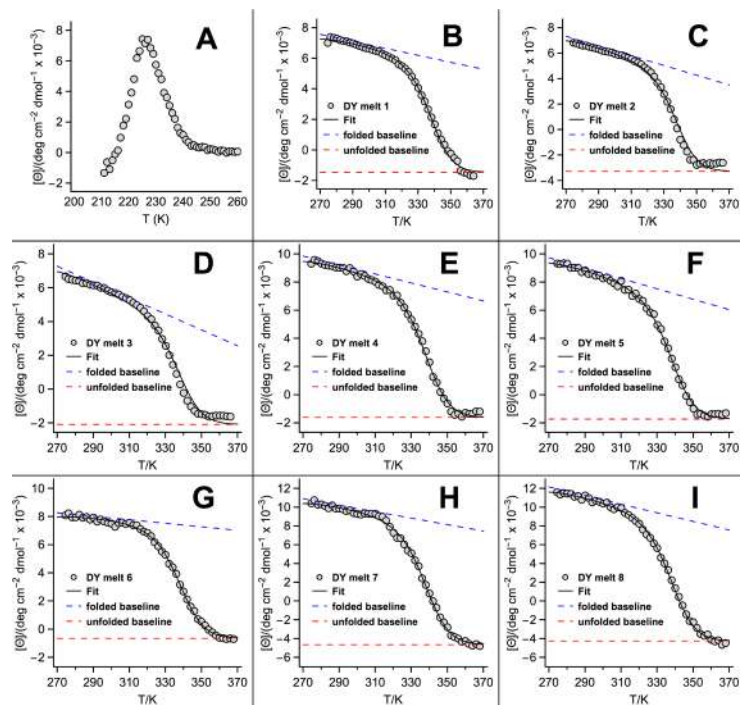
DN Parameters	Values	Units	P-Values
ΔH_f at T_m	-34.2 ± 0.4	kcal/mol	<0.001
T_m	68.7 ± 0.1	$^{\circ}\text{C}$	<0.001
ΔC_p	-0.67 ± 0.02	kcal/mol/K	<0.001
a1	17.6 ± 0.3	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b1	-0.0367 ± 0.0009	$\text{deg cm}^2 \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c1	-1.85 ± 0.02	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d1	---	---	---
a2	17.7 ± 0.3	$\text{deg cm}^{-2} \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b2	-0.0370 ± 0.0009	$\text{deg cm}^{-2} \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c2	-1.82 ± 0.02	$\text{deg cm}^{-2} \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d2	---	---	---
a3	18.1 ± 0.3	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b3	-0.0392 ± 0.0009	$\text{deg cm}^2 \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c3	-2.01 ± 0.02	$\text{deg cm}^{-2} \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d3	---	---	---

Figure 4-23. (A) CD spectrum and (B)–(D) variable temperature CD data for 50 μM solutions of protein **DN** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated; $R^2 = 0.9999$.



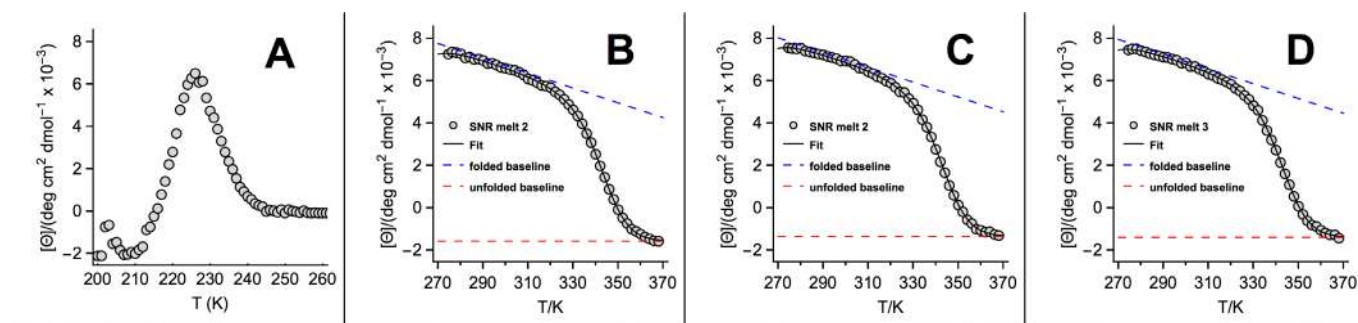
SY Parameters	Values	Units	P-Values
ΔH_f at T_m	-29.7 ± 0.8	kcal/mol	<0.001
T_m	58.4 ± 0.3	°C	<0.001
ΔC_p	-0.65 ± 0.05	kcal/mol/K	<0.001
a1	17.8 ± 1.2	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
b1	-0.036 ± 0.004	deg cm ² dmol ⁻¹ K ⁻¹ x 10 ⁻³	<0.001
c1	-1.38 ± 0.04	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
d1	---	---	---
a2	12.3 ± 1.0	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
b2	-0.023 ± 0.003	deg cm ² dmol ⁻¹ K ⁻¹ x 10 ⁻³	<0.001
c2	-1.43 ± 0.04	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
d2	---	---	---
a3	12.6 ± 0.9	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
b3	-0.025 ± 0.003	deg cm ² dmol ⁻¹ K ⁻¹ x 10 ⁻³	<0.001
c3	-1.31 ± 0.03	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
d3	---	---	---

Figure 4-24. (A) CD spectrum and (B)–(D) variable temperature CD data for 50 μ M solutions of protein SY in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated; $R^2 = 0.9994$.



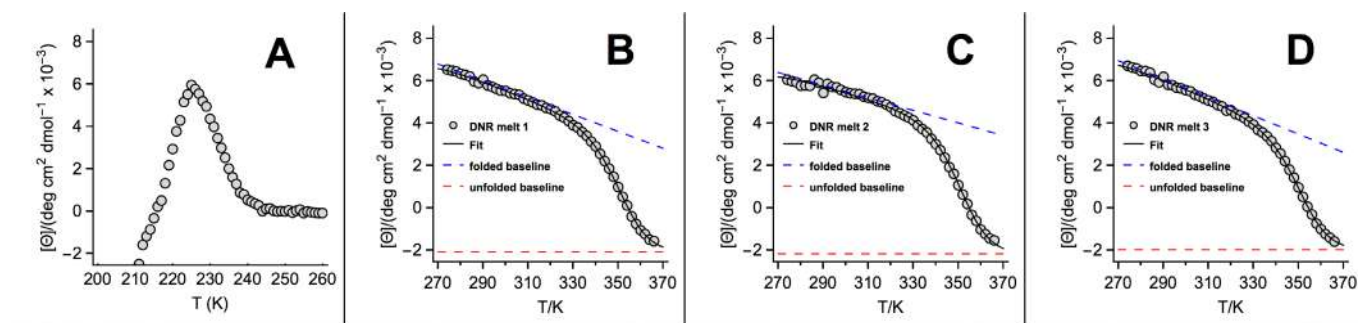
DY Parameters	Values	Units	P-Values
ΔH_f at T_m	-31.7 ± 0.7	kcal/mol	<0.001
T_m	64.2 ± 0.2	$^{\circ}\text{C}$	<0.001
ΔC_p	-0.63 ± 0.05	kcal/mol/K	<0.001
a1	13.7 ± 1.0	$\text{deg cm}^2 \text{dmol}^{-1} \times 10^{-3}$	<0.001
b1	-0.023 ± 0.003	$\text{deg cm}^2 \text{dmol}^{-1} \text{K}^{-1} \times 10^{-3}$	<0.001
c1	-1.46 ± 0.08	$\text{deg cm}^2 \text{dmol}^{-1} \times 10^{-3}$	<0.001
d1	---	---	---
a2	17.6 ± 1.1	$\text{deg cm}^2 \text{dmol}^{-1} \times 10^{-3}$	<0.001
b2	-0.038 ± 0.003	$\text{deg cm}^2 \text{dmol}^{-1} \text{K}^{-1} \times 10^{-3}$	<0.001
c2	-3.27 ± 0.08	$\text{deg cm}^2 \text{dmol}^{-1} \times 10^{-3}$	<0.001
d2	---	---	---
a3	20.0 ± 1.0	$\text{deg cm}^2 \text{dmol}^{-1} \times 10^{-3}$	<0.001
b3	-0.047 ± 0.003	$\text{deg cm}^2 \text{dmol}^{-1} \text{K}^{-1} \times 10^{-3}$	<0.001
c3	-2.09 ± 0.07	$\text{deg cm}^2 \text{dmol}^{-1} \times 10^{-3}$	<0.001
d3	---	---	---
a4	18.4 ± 1.1	$\text{deg cm}^2 \text{dmol}^{-1} \times 10^{-3}$	<0.001
b4	-0.032 ± 0.004	$\text{deg cm}^2 \text{dmol}^{-1} \text{K}^{-1} \times 10^{-3}$	<0.001
c4	-1.61 ± 0.07	$\text{deg cm}^2 \text{dmol}^{-1} \times 10^{-3}$	<0.001
d4	---	---	---
a5	19.7 ± 1.1	$\text{deg cm}^2 \text{dmol}^{-1} \times 10^{-3}$	<0.001
b5	-0.037 ± 0.004	$\text{deg cm}^2 \text{dmol}^{-1} \text{K}^{-1} \times 10^{-3}$	<0.001
c5	-1.72 ± 0.07	$\text{deg cm}^2 \text{dmol}^{-1} \times 10^{-3}$	<0.001
d5	---	---	---
a6	11.6 ± 1.0	$\text{deg cm}^2 \text{dmol}^{-1} \times 10^{-3}$	<0.001
b6	-0.012 ± 0.003	$\text{deg cm}^2 \text{dmol}^{-1} \text{K}^{-1} \times 10^{-3}$	<0.001
c6	-0.65 ± 0.07	$\text{deg cm}^2 \text{dmol}^{-1} \times 10^{-3}$	<0.001
d6	---	---	---
a7	20.2 ± 1.4	$\text{deg cm}^2 \text{dmol}^{-1} \times 10^{-3}$	<0.001
b7	-0.035 ± 0.004	$\text{deg cm}^2 \text{dmol}^{-1} \text{K}^{-1} \times 10^{-3}$	<0.001
c7	-4.65 ± 0.08	$\text{deg cm}^2 \text{dmol}^{-1} \times 10^{-3}$	<0.001
d7	---	---	---
a8	24.6 ± 1.4	$\text{deg cm}^2 \text{dmol}^{-1} \times 10^{-3}$	<0.001
b8	-0.046 ± 0.005	$\text{deg cm}^2 \text{dmol}^{-1} \text{K}^{-1} \times 10^{-3}$	<0.001
c8	-4.25 ± 0.08	$\text{deg cm}^2 \text{dmol}^{-1} \times 10^{-3}$	<0.001
d8	---	---	---

Figure 4-25. (A) CD spectrum and (B)–(I) variable temperature CD data for 50 μM solutions of protein **DY** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated; $R^2 = 0.9988$.



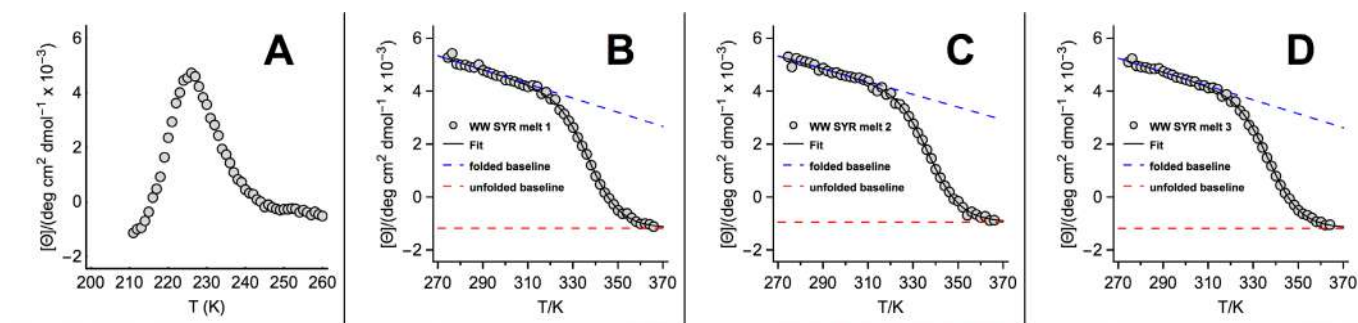
DNR Parameters	Values	Units	P-Values
ΔH_f at T_m	-35.3 ± 1.1	kcal/mol	<0.001
T_m	77.8 ± 0.2	$^{\circ}\text{C}$	<0.001
ΔC_p	-0.61 ± 0.05	kcal/mol/K	<0.001
a1	17.5 ± 0.4	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b1	-0.0397 ± 0.0014	$\text{deg cm}^2 \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c1	-2.08 ± 0.08	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d1	---	---	---
a2	14.4 ± 0.4	$\text{deg cm}^{-2} \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b2	-0.0297 ± 0.0014	$\text{deg cm}^{-2} \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c2	-2.18 ± 0.09	$\text{deg cm}^{-2} \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d2	---	---	---
a3	18.6 ± 0.4	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b3	-0.0433 ± 0.0013	$\text{deg cm}^2 \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c3	-1.98 ± 0.08	$\text{deg cm}^{-2} \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d3	---	---	---

Figure 4-26. (A) CD spectrum and (B)–(D) variable temperature CD data for 50 μM solutions of protein SNR in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated; $R^2 = 0.9999$.



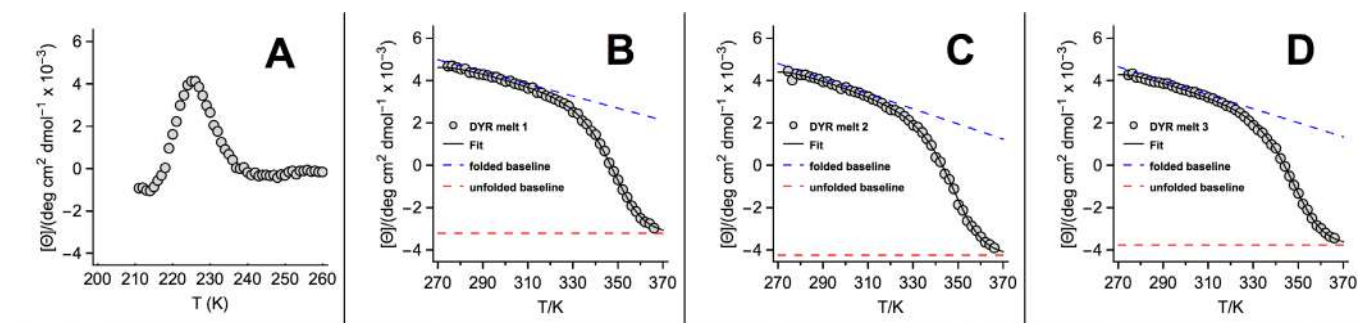
DNR Parameters	Values	Units	P-Values
ΔH_f at T_m	-35.3 ± 1.1	kcal/mol	<0.001
T_m	77.8 ± 0.2	$^{\circ}\text{C}$	<0.001
ΔC_p	-0.61 ± 0.05	kcal/mol/K	<0.001
a1	17.5 ± 0.4	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b1	-0.0397 ± 0.0014	$\text{deg cm}^2 \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c1	-2.08 ± 0.08	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d1	---	---	---
a2	14.4 ± 0.4	$\text{deg cm}^{-2} \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b2	-0.0297 ± 0.0014	$\text{deg cm}^{-2} \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c2	-2.18 ± 0.09	$\text{deg cm}^{-2} \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d2	---	---	---
a3	18.6 ± 0.4	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b3	-0.0433 ± 0.0013	$\text{deg cm}^2 \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c3	-1.98 ± 0.08	$\text{deg cm}^{-2} \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d3	---	---	---

Figure 4-27. (A) CD spectrum and (B)–(D) variable temperature CD data for 50 μM solutions of protein DNR in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated; $R^2 = 0.9997$.



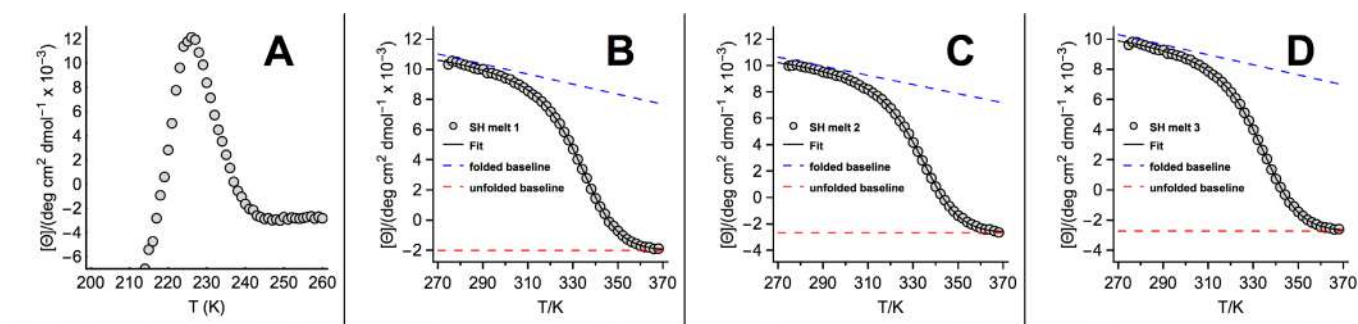
WW SYR Parameters	Values	Units	P-Values
ΔH_f at T_m	-33.0 ± 0.7	kcal/mol	<0.001
T_m	65.1 ± 0.2	$^{\circ}\text{C}$	<0.001
ΔC_p	---	---	---
a1	12.5 ± 0.3	$\text{deg cm}^2 \text{ d mol}^{-1} \times 10^{-3}$	<0.001
b1	-0.027 ± 0.001	$\text{deg cm}^2 \text{ d mol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c1	-1.17 ± 0.03	$\text{deg cm}^2 \text{ d mol}^{-1} \times 10^{-3}$	<0.001
d1	---	---	---
a2	11.8 ± 0.3	$\text{deg cm}^2 \text{ d mol}^{-1} \times 10^{-3}$	<0.001
b2	-0.024 ± 0.001	$\text{deg cm}^2 \text{ d mol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c2	-0.94 ± 0.03	$\text{deg cm}^2 \text{ d mol}^{-1} \times 10^{-3}$	<0.001
d2	---	---	---
a3	12.3 ± 0.3	$\text{deg cm}^2 \text{ d mol}^{-1} \times 10^{-3}$	<0.001
b3	-0.026 ± 0.001	$\text{deg cm}^2 \text{ d mol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c3	-1.17 ± 0.03	$\text{deg cm}^2 \text{ d mol}^{-1} \times 10^{-3}$	<0.001
d3	---	---	---

Figure 4-28. (A) CD spectrum and (B)–(D) variable temperature CD data for 50 μM solutions of protein SYR in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated; $R^2 = 0.9996$.



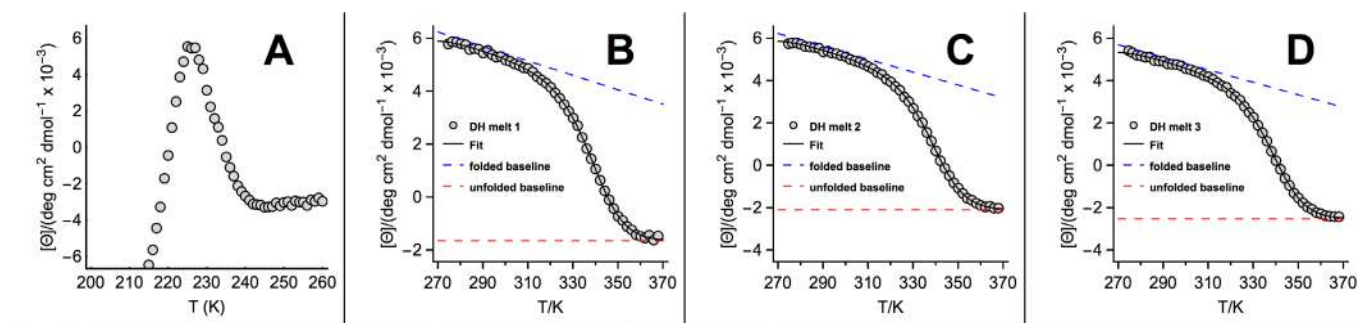
DYR Parameters	Values	Units	P-Values
ΔH_f at T_m	-33.2 ± 0.7	kcal/mol	<0.001
T_m	74.6 ± 0.1	°C	<0.001
ΔC_p	-0.61 ± 0.03	kcal/mol/K	<0.001
a1	12.7 ± 0.4	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
b1	-0.029 ± 0.001	deg cm ² dmol ⁻¹ K ⁻¹ x 10 ⁻³	<0.001
c1	-3.20 ± 0.05	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
d1	---	---	---
a2	14.4 ± 0.4	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
b2	-0.036 ± 0.001	deg cm ² dmol ⁻¹ K ⁻¹ x 10 ⁻³	<0.001
c2	-4.24 ± 0.05	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
d2	---	---	---
a3	13.6 ± 0.4	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
b3	-0.033 ± 0.001	deg cm ² dmol ⁻¹ K ⁻¹ x 10 ⁻³	<0.001
c3	-3.76 ± 0.05	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
d3	---	---	---

Figure 4-29. (A) CD spectrum and (B)–(D) variable temperature CD data for 50 μ M solutions of protein **DYR** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated; $R^2 = 0.9996$.



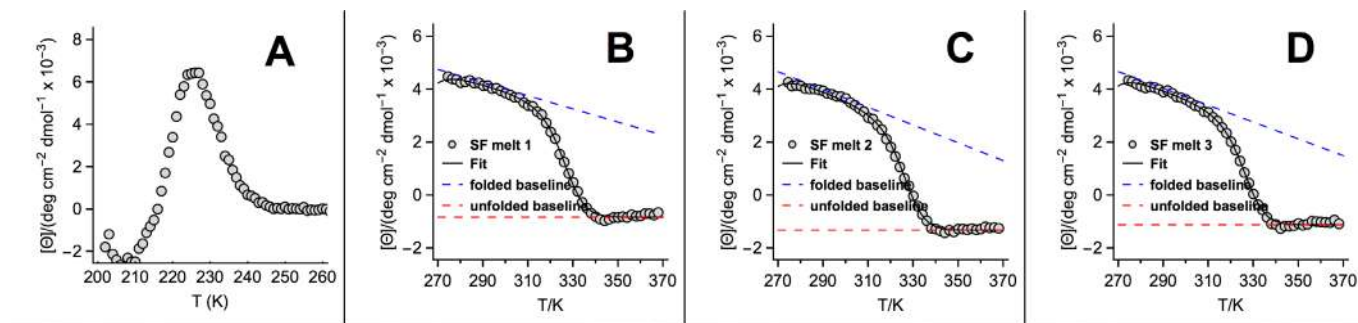
SH Parameters	Values	Units	P-Values
ΔH_f at T_m	-25.4 ± 0.2	kcal/mol	<0.001
T_m	60.8 ± 0.1	$^{\circ}\text{C}$	<0.001
ΔC_p	-0.46 ± 0.02	kcal/mol/K	<0.001
a1	20.0 ± 0.5	$\text{deg cm}^2 \text{ d mol}^{-1} \times 10^{-3}$	<0.001
b1	-0.033 ± 0.001	$\text{deg cm}^2 \text{ d mol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c1	-2.00 ± 0.02	$\text{deg cm}^2 \text{ d mol}^{-1} \times 10^{-3}$	<0.001
d1	---	---	---
a2	20.0 ± 0.5	$\text{deg cm}^2 \text{ d mol}^{-1} \times 10^{-3}$	<0.001
b2	-0.035 ± 0.002	$\text{deg cm}^2 \text{ d mol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c2	-2.65 ± 0.02	$\text{deg cm}^2 \text{ d mol}^{-1} \times 10^{-3}$	<0.001
d2	---	---	---
a3	19.3 ± 0.5	$\text{deg cm}^2 \text{ d mol}^{-1} \times 10^{-3}$	<0.001
b3	-0.033 ± 0.001	$\text{deg cm}^2 \text{ d mol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c3	-2.71 ± 0.02	$\text{deg cm}^2 \text{ d mol}^{-1} \times 10^{-3}$	<0.001
d3	---	---	---

Figure 4-30. (A) CD spectrum and (B)–(D) variable temperature CD data for 50 μM solutions of protein SH in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated; $R^2 = 0.9999$.



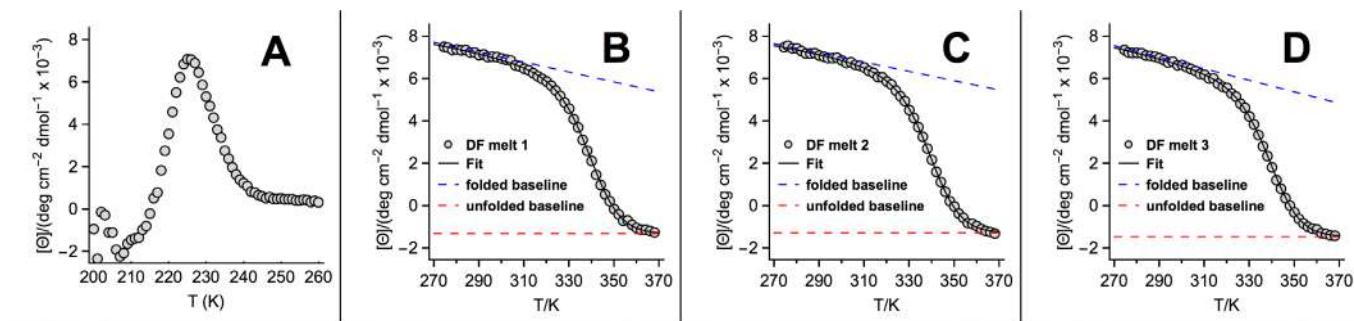
DH Parameters	Values	Units	P-Values
ΔH_f at T_m	-29.7 ± 0.3	kcal/mol	<0.001
T_m	65.4 ± 0.1	$^{\circ}\text{C}$	<0.001
ΔC_p	-0.58 ± 0.02	kcal/mol/K	<0.001
a1	13.6 ± 0.3	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b1	-0.027 ± 0.001	$\text{deg cm}^2 \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c1	-1.65 ± 0.02	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d1	---	---	---
a2	14.4 ± 0.3	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b2	-0.030 ± 0.001	$\text{deg cm}^2 \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c2	-2.09 ± 0.02	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d2	---	---	---
a3	13.6 ± 0.3	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b3	-0.029 ± 0.001	$\text{deg cm}^2 \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c3	-2.52 ± 0.02	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d3	---	---	---

Figure 4-31. (A) CD spectrum and (B)–(D) variable temperature CD data for 50 μM solutions of protein **DH** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated; $R^2 = 0.9999$.



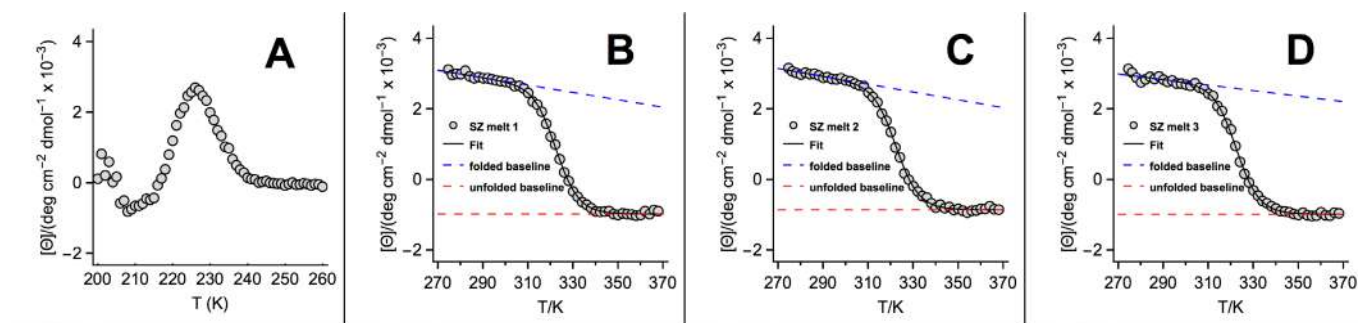
SF Parameters	Values	Units	P-Values
ΔH_f at T_m	-43.0 ± 1.3	kcal/mol	<0.001
T_m	52.8 ± 0.2	°C	<0.001
ΔC_p	-1.21 ± 0.06	kcal/mol/K	<0.001
a1	11.5 ± 0.6	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
b1	-0.025 ± 0.002	deg cm ² dmol ⁻¹ K ⁻¹ x 10 ⁻³	<0.001
c1	-0.82 ± 0.02	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
d1	---	---	---
a2	13.7 ± 0.6	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
b2	-0.034 ± 0.003	deg cm ² dmol ⁻¹ K ⁻¹ x 10 ⁻³	<0.001
c2	-1.32 ± 0.02	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
d2	---	---	---
a3	13.2 ± 0.6	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
b3	-0.032 ± 0.002	deg cm ² dmol ⁻¹ K ⁻¹ x 10 ⁻³	<0.001
c3	-1.12 ± 0.02	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
d3	---	---	---

Figure 4-32. (A) CD spectrum and (B)–(D) variable temperature CD data for 50 μ M solutions of protein SF in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated; $R^2 = 0.9992$.



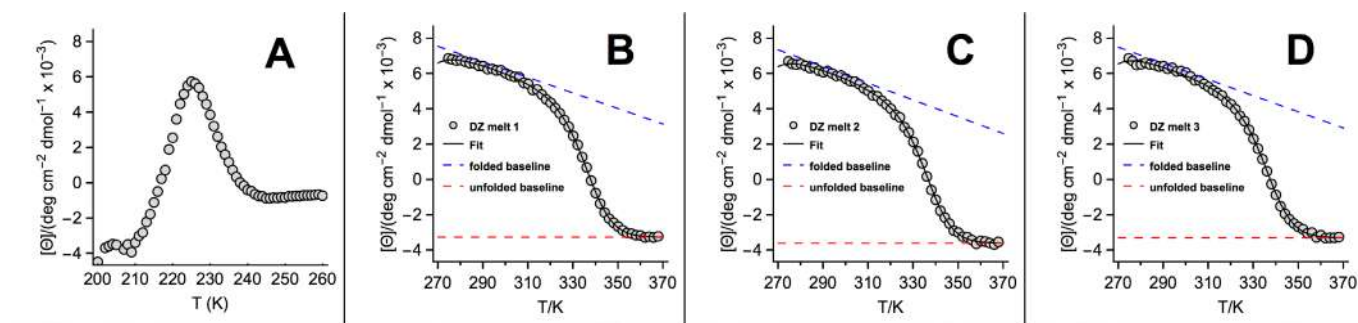
DF Parameters	Values	Units	P-Values
ΔH_f at T_m	-31.9 ± 0.3	kcal/mol	<0.001
T_m	65.9 ± 0.1	$^{\circ}\text{C}$	<0.001
ΔC_p	-0.54 ± 0.05	kcal/mol/K	<0.001
a1	14.0 ± 0.2	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b1	-0.023 ± 0.001	$\text{deg cm}^2 \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c1	-1.31 ± 0.02	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d1	---	---	---
a2	13.5 ± 0.2	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b2	-0.022 ± 0.001	$\text{deg cm}^2 \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c2	-1.27 ± 0.02	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d2	---	---	---
a3	14.9 ± 0.2	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b3	-0.027 ± 0.001	$\text{deg cm}^2 \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c3	-1.46 ± 0.02	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d3	---	---	---

Figure 4-33. (A) CD spectrum and (B)–(D) variable temperature CD data for 50 μM solutions of protein **DF** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated; $R^2 = 0.9999$.



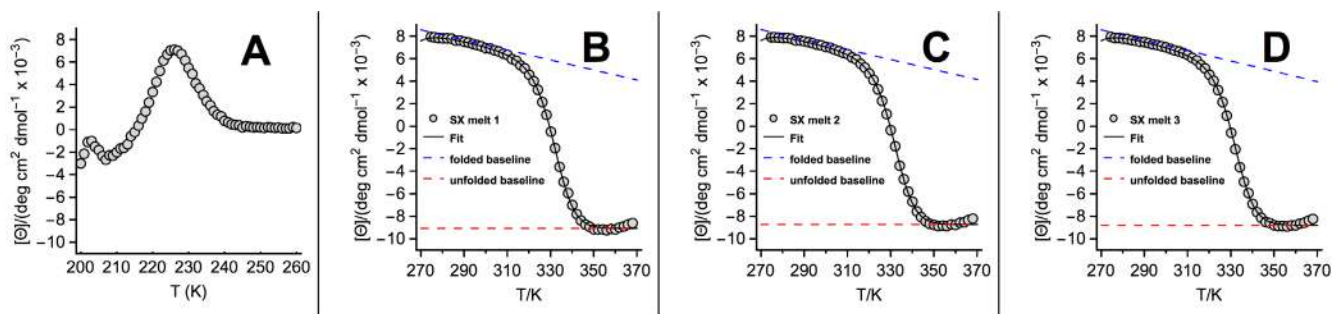
SZ Parameters	Values	Units	P-Values
ΔH_f at T_m	-41.6 ± 0.7	kcal/mol	<0.001
T_m	49.7 ± 0.1	°C	<0.001
ΔC_p	---	---	---
a1	5.9 ± 0.3	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
b1	-0.010 ± 0.001	deg cm ² dmol ⁻¹ K ⁻¹ x 10 ⁻³	<0.001
c1	-0.97 ± 0.01	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
d1	---	---	---
a2	6.1 ± 0.3	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
b2	-0.011 ± 0.001	deg cm ² dmol ⁻¹ K ⁻¹ x 10 ⁻³	<0.001
c2	-0.85 ± 0.01	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
d2	---	---	---
a3	5.1 ± 0.3	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
b3	-0.008 ± 0.001	deg cm ² dmol ⁻¹ K ⁻¹ x 10 ⁻³	<0.001
c3	-0.98 ± 0.01	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
d3	---	---	---

Figure 4-34. (A) CD spectrum and (B)–(D) variable temperature CD data for 50 μ M solutions of protein SZ in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated; $R^2 = 0.9994$.



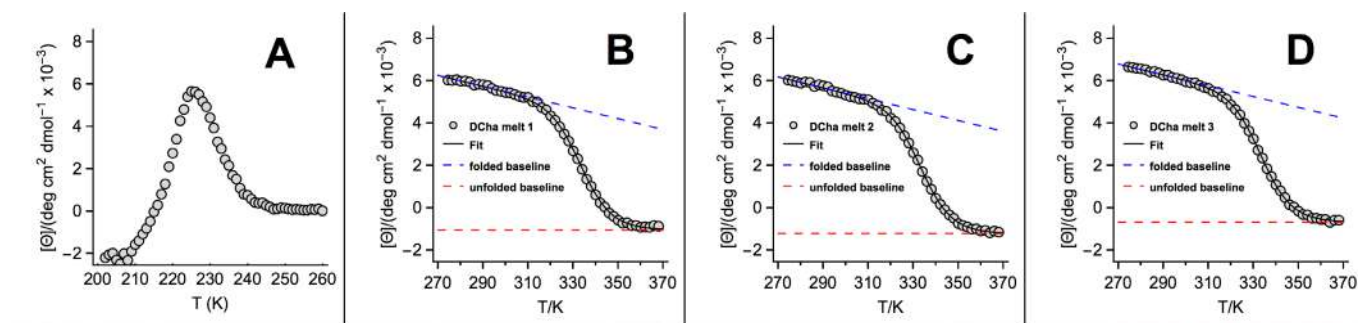
DZ Parameters	Values	Units	P-Values
ΔH_f at T_m	-35.4 ± 0.5	kcal/mol	<0.001
T_m	62.6 ± 0.1	$^{\circ}\text{C}$	<0.001
ΔC_p	-0.82 ± 0.02	kcal/mol/K	<0.001
a1	19.4 ± 0.5	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b1	-0.044 ± 0.002	$\text{deg cm}^2 \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c1	-3.27 ± 0.03	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d1	---	---	---
a2	20.1 ± 0.5	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b2	-0.047 ± 0.002	$\text{deg cm}^2 \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c2	-3.61 ± 0.03	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d2	---	---	---
a3	19.8 ± 0.5	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b3	-0.046 ± 0.002	$\text{deg cm}^2 \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c3	-3.30 ± 0.03	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d3	---	---	---

Figure 4-35. (A) CD spectrum and (B)–(D) variable temperature CD data for 50 μM solutions of protein **DZ** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated; $R^2 = 0.9997$.



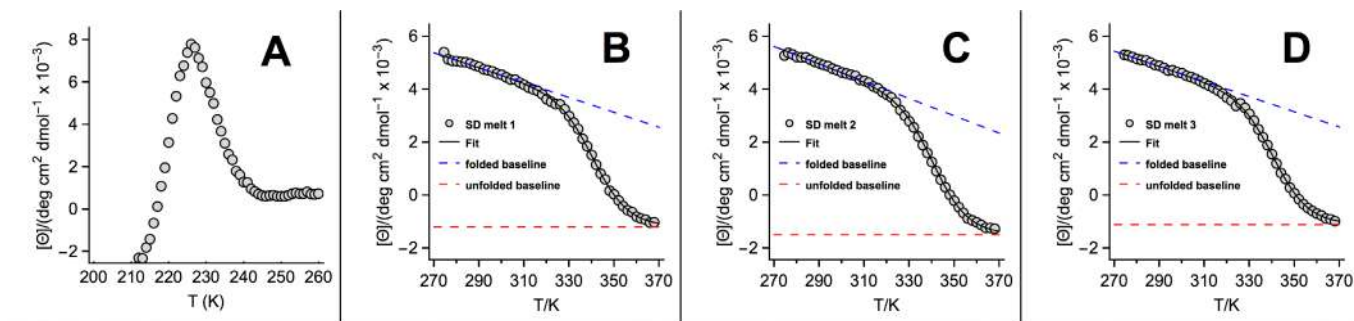
SX Parameters	Values	Units	P-Values
ΔH_f at T_m	-45.1 ± 0.6	kcal/mol	<0.001
T_m	58.4 ± 0.1	$^{\circ}\text{C}$	<0.001
ΔC_p	-1.12 ± 0.03	kcal/mol/K	<0.001
a1	20.6 ± 0.9	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b1	-0.044 ± 0.003	$\text{deg cm}^2 \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c1	-9.06 ± 0.04	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d1	---	---	---
a2	20.5 ± 0.8	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b2	-0.044 ± 0.003	$\text{deg cm}^2 \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c2	-8.72 ± 0.04	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d2	---	---	---
a3	21.0 ± 0.8	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b3	-0.046 ± 0.003	$\text{deg cm}^2 \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c3	-8.78 ± 0.04	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d3	---	---	---

Figure 4-36. (A) CD spectrum and (B)–(D) variable temperature CD data for 50 μM solutions of protein SX in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated. $R^2 = 0.9997$.



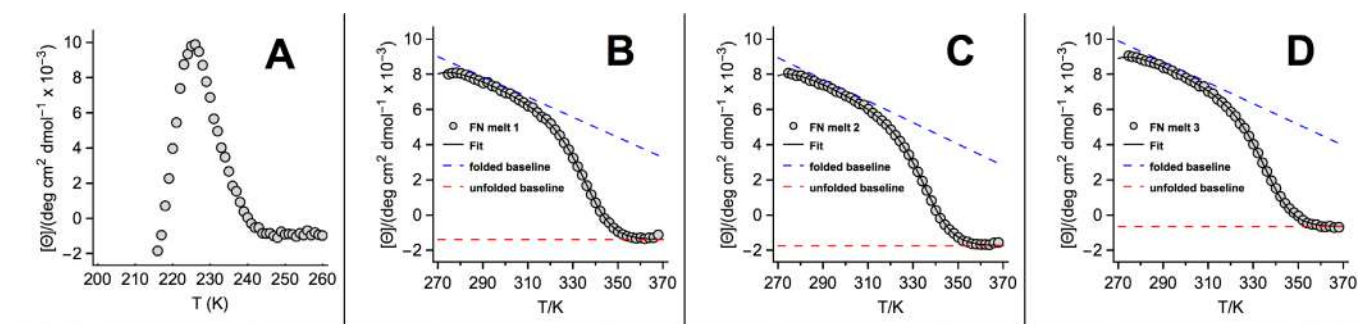
DChA Parameters	Values	Units	P-Values
ΔH_f at T_m	-33.0 ± 0.4	kcal/mol	<0.001
T_m	61.3 ± 0.1	°C	<0.001
ΔC_p	---	---	---
a1	13.1 ± 0.2	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
b1	-0.025 ± 0.001	deg cm ² dmol ⁻¹ K ⁻¹ x 10 ⁻³	<0.001
c1	-1.04 ± 0.02	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
d1	---	---	---
a2	13.1 ± 0.2	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
b2	-0.026 ± 0.001	deg cm ² dmol ⁻¹ K ⁻¹ x 10 ⁻³	<0.001
c2	-1.21 ± 0.02	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
d2	---	---	---
a3	13.7 ± 0.2	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
b3	-0.026 ± 0.001	deg cm ² dmol ⁻¹ K ⁻¹ x 10 ⁻³	<0.001
c3	-0.68 ± 0.02	deg cm ² dmol ⁻¹ x 10 ⁻³	<0.001
d3	---	---	---

Figure 4-37. (A) CD spectrum and (B)–(D) variable temperature CD data for 50 μ M solutions of protein SX in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated. $R^2 = 0.9999$.



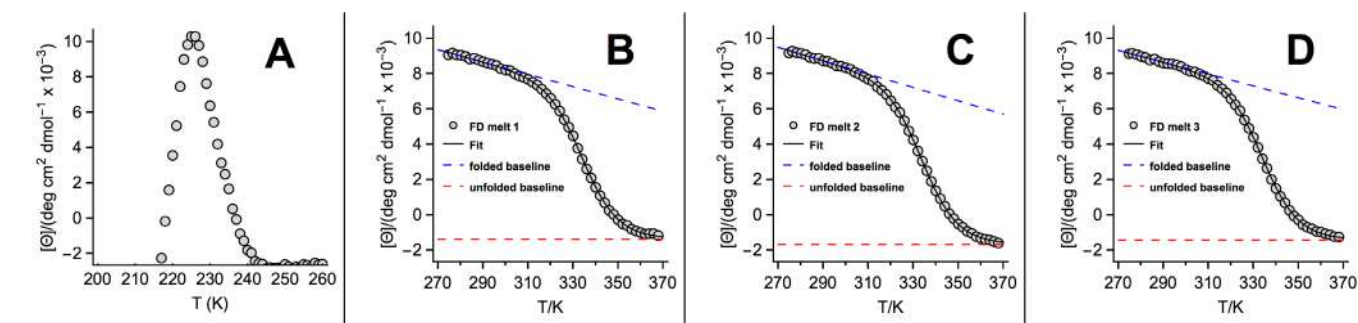
SD Parameters	Values	Units	P-Values
ΔH_f at T_m	-30.5 ± 0.5	kcal/mol	<0.001
T_m	69.4 ± 0.1	$^{\circ}\text{C}$	<0.001
ΔC_p	---	---	---
a1	12.9 ± 0.2	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b1	-0.028 ± 0.001	$\text{deg cm}^2 \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c1	-1.20 ± 0.03	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d1	---	---	---
a2	14.4 ± 0.2	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b2	-0.033 ± 0.001	$\text{deg cm}^2 \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c2	-1.50 ± 0.03	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d2	---	---	---
a3	13.2 ± 0.2	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b3	-0.029 ± 0.001	$\text{deg cm}^2 \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c3	-1.11 ± 0.03	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d3	---	---	---

Figure 4-38. (A) CD spectrum and (B)–(D) variable temperature CD data for 50 μM solutions of protein SD in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated. $R^2 = 0.9998$.



FN Parameters	Values	Units	P-Values
ΔH_f at T_m	-32.2 ± 0.4	kcal/mol	<0.001
T_m	62.1 ± 0.1	$^{\circ}\text{C}$	<0.001
ΔC_p	-0.74 ± 0.02	kcal/mol/K	<0.001
a1	24.5 ± 0.4	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b1	-0.0573 ± 0.0014	$\text{deg cm}^2 \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c1	-1.38 ± 0.02	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d1	---	---	---
a2	25.5 ± 0.4	$\text{deg cm}^{-2} \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b2	-0.0613 ± 0.0014	$\text{deg cm}^{-2} \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c2	-1.73 ± 0.02	$\text{deg cm}^{-2} \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d2	---	---	---
a3	25.9 ± 0.4	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b3	-0.0594 ± 0.0014	$\text{deg cm}^2 \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c3	-0.64 ± 0.02	$\text{deg cm}^{-2} \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d3	---	---	---

Figure 4-39. (A) CD spectrum and (B)–(D) variable temperature CD data for 50 μM solutions of protein FN in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated. $R^2 = 0.9999$.



FD Parameters	Values	Units	P-Values
ΔH_f at T_m	-28.6 ± 0.2	kcal/mol	<0.001
T_m	62.2 ± 0.1	°C	<0.001
ΔC_p	---	---	---
a1	18.7 ± 0.2	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b1	-0.035 ± 0.001	$\text{deg cm}^2 \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c1	-1.38 ± 0.02	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d1	---	---	---
a2	19.7 ± 0.2	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b2	-0.038 ± 0.001	$\text{deg cm}^2 \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c2	-1.68 ± 0.02	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d2	---	---	---
a3	18.4 ± 0.2	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
b3	-0.034 ± 0.001	$\text{deg cm}^2 \text{ dmol}^{-1} \text{ K}^{-1} \times 10^{-3}$	<0.001
c3	-1.43 ± 0.02	$\text{deg cm}^2 \text{ dmol}^{-1} \times 10^{-3}$	<0.001
d3	---	---	---

Figure 4-40. (A) CD spectrum and (B)–(D) variable temperature CD data for 50 μM solutions of protein **FD** in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated. $R^2 = 0.9999$.

4.5 References

1. Dougherty, D. A., Cation- π interactions in chemistry and biology: A new view of benzene, Phe, Tyr, and Trp. *Science* **1996**, *271* (5246), 163-168.
2. Mecozzi, S.; West, A. P.; Dougherty, D. A., Cation- π interactions in aromatics of biological and medicinal interest: Electrostatic potential surfaces as a useful qualitative guide. *Proceedings of the National Academy of Sciences of the United States of America* **1996**, *93* (20), 10566-10571.
3. Gallivan, J. P.; Dougherty, D. A., Cation- π interactions in structural biology. *Proc Natl Acad Sci U S A* **1999**, *96* (17), 9459-64.

4. Tatko, C. D.; Waters, M. L., The geometry and efficacy of cation-pi interactions in a diagonal position of a designed beta-hairpin. *Protein Science* **2003**, *12* (11), 2443-2452.
5. Tsou, L. K.; Tatko, C. D.; Waters, M. L., Simple cation-pi interaction between a phenyl ring and a protonated amine stabilizes an alpha-helix in water. *J. Am. Chem. Soc.* **2002**, *124* (50), 14917-14921.
6. Quinonero, D.; Garau, C.; Rotger, C.; Frontera, A.; Ballester, P.; Costa, A.; Deya, P. M., Anion-pi interactions: Do they exist? *Angewandte Chemie-International Edition* **2002**, *41* (18), 3389-3392.
7. Frontera, A.; Gamez, P.; Mascal, M.; Mooibroek, T. J.; Reedijk, J., Putting anion-pi interactions into perspective. *Angew Chem Int Ed Engl* **2011**, *50* (41), 9564-83.
8. Bauza, A.; Seth, S. K.; Frontera, A., Molecular electrostatic potential and "atoms-in-molecules" analyses of the interplay between pi-hole and lone pair...pi/X-H...pi/metal...pi interactions. *J Comput Chem* **2018**, *39* (9), 458-463.
9. Estarellas, C.; Frontera, A.; Quinonero, D.; Deya, P. M., Anion-pi interactions in flavoproteins. *Chem Asian J* **2011**, *6* (9), 2316-8.
10. Estarellas, C.; Bauza, A.; Frontera, A.; Quinonero, D.; Deya, P. M., On the directionality of anion-pi interactions. *Phys Chem Chem Phys* **2011**, *13* (13), 5696-702.
11. Giese, M.; Albrecht, M.; Rissanen, K., Experimental investigation of anion-pi interactions--applications and biochemical relevance. *Chem Commun (Camb)* **2016**, *52* (9), 1778-95.
12. Giese, M.; Albrecht, M.; Rissanen, K., Anion-pi Interactions with Fluoroarenes. *Chemical Reviews* **2015**, *115* (16), 8867-8895.
13. Wheeler, S. E.; Houk, K. N., Are Anion/pi Interactions Actually a Case of Simple Charge-Dipole Interaction? *Journal of Physical Chemistry A* **2010**, *114* (33), 8658-8664.
14. Gamez, P.; Mooibroek, T. J.; Teat, S. J.; Reedijk, J., Anion binding involving pi-acidic heteroaromatic rings. *Acc Chem Res* **2007**, *40* (6), 435-44.
15. Jones, G. J.; Robertazzi, A.; Platts, J. A., Efficient and accurate theoretical methods to investigate anion-pi interactions in protein model structures. *J Phys Chem B* **2013**, *117* (12), 3315-22.
16. Lucas, X.; Bauza, A.; Frontera, A.; Quinonero, D., A thorough anion-pi interaction study in biomolecules: on the importance of cooperativity effects. *Chemical Science* **2016**, *7* (2), 1038-1050.
17. Chakravarty, S.; Sheng, Z. Z.; Iverson, B.; Moore, B., "eta6"-Type anion-pi in biomolecular recognition. *FEBS Lett* **2012**, *586* (23), 4180-5.
18. Philip, V.; Harris, J.; Adams, R.; Nguyen, D.; Spiers, J.; Baudry, J.; Howell, E. E.; Hinde, R. J., A Survey of Aspartate-Phenylalanine and Glutamate-Phenylalanine Interactions in the Protein Data Bank: Searching for Anion-pi Pairs. *Biochemistry* **2011**, *50* (14), 2939-2950.
19. Jackson, M. R.; Beahm, R.; Duvvuru, S.; Narasimhan, C.; Wu, J.; Wang, H. N.; Philip, V. M.; Hinde, R. J.; Howell, E. E., A preference for edgewise interactions between aromatic rings and carboxylate anions: the biological relevance of anion-quadrupole interactions. *J Phys Chem B* **2007**, *111* (28), 8242-9.
20. Olson, C. A.; Shi, Z.; Kallenbach, N. R., Polar interactions with aromatic side chains in alpha-helical peptides: Ch...O H-bonding and cation-pi interactions. *J Am Chem Soc* **2001**, *123* (26), 6451-2.

21. Shi, Z.; Olson, C. A.; Bell, A. J., Jr.; Kallenbach, N. R., Stabilization of alpha-helix structure by polar side-chain interactions: complex salt bridges, cation-pi interactions, and C-H... leader O H-bonds. *Biopolymers* **2001**, *60* (5), 366-80.
22. Pankov, R.; Yamada, K. M., Fibronectin at a glance. *J Cell Sci* **2002**, *115* (20), 3861-3863.
23. Huang, J.; Koide, A.; Makabe, K.; Koide, S., Design of protein function leaps by directed domain interface evolution. *Proc Natl Acad Sci U S A* **2008**, *105* (18), 6578-83.
24. Juers, D. H.; Heightman, T. D.; Vasella, A.; McCarter, J. D.; Mackenzie, L.; Withers, S. G.; Matthews, B. W., A structural view of the action of Escherichia coli (lacZ) beta-galactosidase. *Biochemistry* **2001**, *40* (49), 14781-14794.
25. Husain, Q., beta Galactosidases and their potential applications: a review. *Crit Rev Biotechnol* **2010**, *30* (1), 41-62.
26. Lawrence, P. B.; Gavrilov, Y.; Matthews, S. S.; Langlois, M. I.; Shental-Bechor, D.; Greenblatt, H. M.; Pandey, B. K.; Smith, M. S.; Paxman, R.; Torgerson, C. D.; Merrell, J. P.; Ritz, C. C.; Prigozhin, M. B.; Levy, Y.; Price, J. L., Criteria for selecting PEGylation sites on proteins for higher thermodynamic and proteolytic stability. *J Am Chem Soc* **2014**, *136* (50), 17547-60.
27. Pandey, B. K.; Smith, M. S.; Torgerson, C.; Lawrence, P. B.; Matthews, S. S.; Watkins, E.; Groves, M. L.; Prigozhin, M. B.; Price, J. L., Impact of Site-Specific PEGylation on the Conformational Stability and Folding Rate of the Pin WW Domain Depends Strongly on PEG Oligomer Length. *Bioconjugate Chemistry* **2013**, *24* (5), 796-802.
28. Dave, K.; Jager, M.; Nguyen, H.; Kelly, J. W.; Gruebele, M., High-Resolution Mapping of the Folding Transition State of a WW Domain. *J Mol Biol* **2016**, *428* (8), 1617-36.
29. Price, J. L.; Culyba, E. K.; Chen, W.; Murray, A. N.; Hanson, S. R.; Wong, C.-H.; Powers, E. T.; Kelly, J. W., N-glycosylation of enhanced aromatic sequons to increase glycoprotein stability. *Biopolymers* **2012**, *98* (3), 195-211.
30. Price, J.; Powers, E.; Kelly, J., N-PEGylation of a Reverse Turn Is Stabilizing in Multiple Sequence Contexts, unlike N-GlcNAcylation. *ACS Chem. Bio.*: 2011; Vol. 6, pp 1188-1192.
31. Jager, M.; Dendle, M.; Kelly, J. W., Sequence determinants of thermodynamic stability in a WW domain--an all-beta-sheet protein. *Protein Sci* **2009**, *18* (8), 1806-13.
32. Jager, M.; Zhang, Y.; Bieschke, J.; Nguyen, H.; Dendle, M.; Bowman, M. E.; Noel, J. P.; Gruebele, M.; Kelly, J. W., Structure-function-folding relationship in a WW domain. *Proc Natl Acad Sci U S A* **2006**, *103* (28), 10648-53.
33. Estarellas, C.; Quinonero, D.; Deya, P. M.; Frontera, A., Anion-pi interactions involving [MX(n)](m-) anions: a comprehensive theoretical study. *Chemphyschem* **2013**, *14* (1), 145-54.
34. Estarellas, C.; Frontera, A.; Quinonero, D.; Deya, P. M., Relevant anion-pi interactions in biological systems: the case of urate oxidase. *Angew Chem Int Ed Engl* **2011**, *50* (2), 415-8.
35. Estarellas, C.; Escudero, D.; Frontera, A.; Quinonero, D.; Deya, P. M., Theoretical ab initio study of the interplay between hydrogen bonding, cation-pi and pi-pi interactions. *Theoretical Chemistry Accounts* **2009**, *122* (5-6), 325-332.
36. Edelhoch, H., Spectroscopic Determination of Tryptophan and Tyrosine in Proteins*. *Biochemistry (Mosc)*. **1967**, *6* (7), 1948-1954.

5 THE CYS_I-LYS_{I+3}-LYS_{I+4} TRIAD: A GENERAL APPROACH FOR PEG-BASED STABILIZATION OF α -HELICAL PROTEINS.

5.1 Introduction

Attaching polyethylene glycol oligomers (PEGs) to protein surface side chains (i.e. PEGylation) is an important strategy for protecting protein drugs from immunogenicity, aggregation, proteolysis, and for increasing their serum half-life.¹⁻³ Early protein PEGylation strategies were generally non-specific, where current chemoselective side-chain modification strategies now allow specific placement of a single PEG at any position on a protein surface.⁴⁻⁸ However, this site-specific protein PEGylation raises the questions: where should we put PEG? Do some PEGylation sites provide more pharmacokinetic protection than others? If so, can we develop rational guidelines for recognizing optimal PEGylation sites in advance, thereby accelerating the development of PEGylated protein drugs?

Because unfolded or misfolded proteins are more aggregation-prone,⁹⁻¹⁰ more susceptible to proteolysis,¹¹⁻¹³ and more readily recognized by antibodies,¹⁴⁻¹⁵ we hypothesize that optimal PEGylation sites should be characterized by the ability of the attached PEG to increase protein conformational stability. To support this hypothesis, the Price lab has demonstrated that PEG-based stabilization of a reverse turn in the β -sheet-containing WW domain of the human protein

Pin 1 is associated with enhanced resistance to proteolysis.¹⁶ However, it is unclear whether PEGylation can stabilize other secondary structures such as α -helices.

Here we explore the energetic impact of PEGylating a trimeric α -helical coiled coil. We find that modifying a solvent-exposed Cys with a PEG-maleimide increases the conformational stability of the coiled coil due to a favorable three-way interaction between the PEG-maleimide and two Lys residues at the $i+3$ and $i+4$ positions, relative to Cys. Analogous positioning of this Cys_{*i*}-Lys_{*i+3*}-Lys_{*i+4*} triad in the C-terminal α -helix of the villin headpiece domain results in similar PEG-based increases to conformational stability. Our findings suggest that the Cys_{*i*}-Lys_{*i+3*}-Lys_{*i+4*} triad is a simple, general, structure based sequence motif for identifying or installing stabilizing maleimide-based PEGylation sites within α -helical proteins.

5.2 Results and Discussion

Early studies showed that N-terminal PEGylation of coiled-coil peptides can promote helical secondary structure;¹⁷⁻¹⁸ but can be either stabilizing¹⁸ or destabilizing¹⁷ to coiled-coil quaternary structure for reasons that are unclear. Xu and coworkers¹⁹ used a 2000 Da PEG-maleimide to modify a centrally located solvent-exposed Cys residue within the designed trimeric coiled coil **1CW** (Figure 5-1).²⁰ The resulting PEGylated peptide **p1CW** was more helical than its non-PEGylated counterpart, and retained the ability to form a cooperatively folded trimeric coiled coil.¹⁹ Small-angle X-ray scattering studies of **p1CW** revealed that the PEG in **p1CW** is more compact than would be expected for a free PEG in solution, possibly due to favorable PEG-peptide interactions.²¹⁻²² Molecular dynamic simulations by two different groups²³⁻²⁴ suggested the possible role of interactions between positively charged Lys residues and electronegative oxygen

atoms of the PEG oligomer in mediating PEG-based increases in helicity. One study implicated Lys7, Lys15, and Lys21 as particularly important.²³

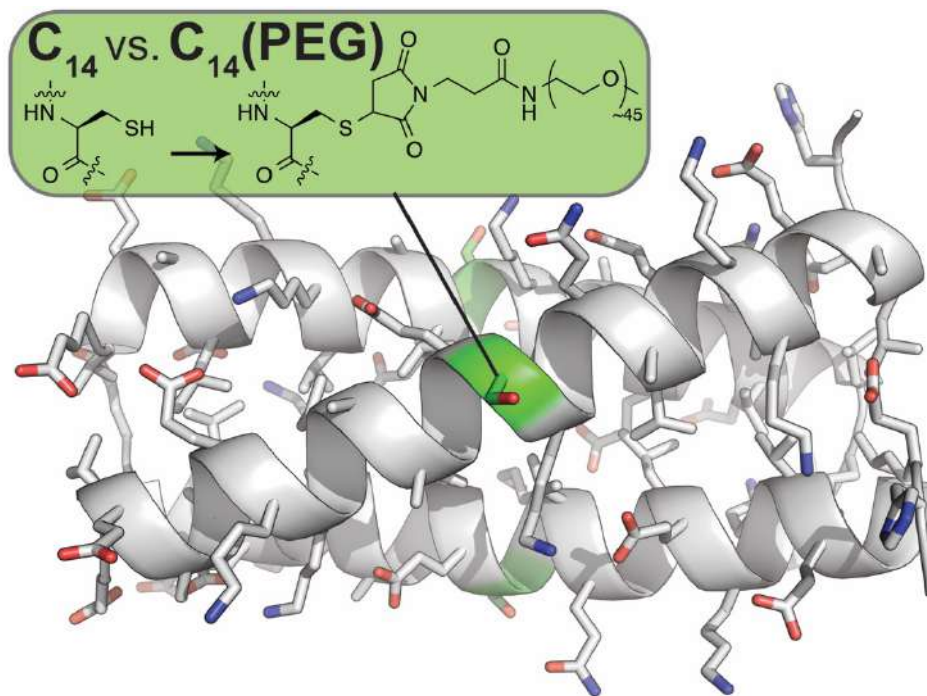


Figure 5-1. Ribbon diagram (rendered in Pymol) of the parent coiled-coil trimer from which **1CW** was derived. Installation of a Cys at position 14 (highlighted in green) facilitates site-specific PEGylation with a 2000 Da PEG maleimide.

If these predictions are accurate, replacing Lys7, Lys15, or Lys21 with Ala (i.e. Ala mutation) should lessen the impact of PEGylation on helicity. To test this hypothesis, we used solid-phase peptide synthesis to prepare peptide **1CW** and alanine mutants **7A**, **15A**, **21A**, in which Lys residues at positions 7, 15, or 21, respectively, have been changed to Ala (see supporting information Table S4). We then prepared PEGylated peptides **p1CW**, **p7A**, **p15A**, and **p21A** by modifying Cys14 in **1CW**, **7A**, **15A**, **21A** with a 2000 Da PEG-maleimide. The CD spectra of these peptides have minima at 222 nm and 208 nm, consistent with α -helical secondary structure and coiled-coil quaternary structure. Analysis of these spectra with the program K2D3²⁵ indicates

that PEGylated **p1CW** is 2% more helical than **1CW** in 20 mM sodium phosphate buffer, pH 7. Interestingly, Ala mutation of Lys₇, Lys₁₅, or Lys₂₁ does not lessen the impact of PEGylation on helicity: **p7A**, **p15A**, and **p21A** are each 6%, 5%, and 9% more helical, respectively, than their non-PEGylated counterparts, a greater increment than we observed for **p1CW** relative to **1CW**. These results suggest that any interactions between PEG and Lys₇, Lys₁₅, or Lys₂₁ do not contribute to the enhanced helicity of **p1CW**.

In principle, Lys-PEG interactions could affect coiled-coil quaternary structural stability, independent of their effect on α -helical secondary structure. To test this hypothesis, we used sedimentation equilibrium experiments to verify the trimeric association state of **7A**, **p7A**, **15A**, **p15A**, **21A**, and **p21A**. We assessed the conformational stability of these compounds by fitting variable temperature CD data to equations derived from a two-state folding model for a coiled-coil trimer (see supporting information for details).

We find that the **p1CW** coiled coil is less stable than **1CW**: a 30 μ M solution of **1CW** has a melting temperature (T_m) of 78.3 °C, whereas a 30 μ M solution of **p1CW** has a T_m of 76.8 °C, corresponding to decrease in stability of 0.26 ± 0.07 kcal mol⁻¹ at the reference temperature of 70°C. Ala mutation at position 7 does not substantially change the impact of PEGylation on coiled-coil stability, suggesting that PEG-Lys₇ interactions do not play a significant role. In contrast, Ala mutation at positions 15 and 21 makes PEGylation even more unfavorable than observed for **p1CW** (see supporting information Table S4, compare **p15A** vs. **15A**, $\Delta\Delta G_f = 0.44 \pm 0.03$ kcal mol⁻¹; compare **p21A** and **21A**, $\Delta\Delta G_f = 0.58 \pm 0.03$), suggesting that interactions between the PEG-maleimide and Lys₁₅ or Lys₂₁ do contribute to the stability of **p1CW**, though not enough to stabilize **p1CW** relative to **1CW**. We wondered whether these Lys/PEG-maleimide interactions might have a greater stabilizing impact if the Lys residues were located closer in space to

PEGylated Cys₁₄. To test this hypothesis, we generated peptides **17K**, **18K**, and their PEGylated counterparts (**p17K** and **p18K**), in which wild-type residues at *i*+3 position 17 or *i*+4 position 18, respectively, have been changed to Lys (see supporting information Table S4). We chose these positions because they are only one helical turn away from Cys₁₄, along the same solvent-exposed face of the helix. We also prepared **17/18K** and PEGylated **p17/18K**, in which Lys residues occupy both positions 17 and 18.

CD spectra and sedimentation equilibrium experiments confirm that **17K**, **p17K**, **18K**, **p18K**, **17/18K**, and **p17/18K** each form coiled-coil trimers. Maleimide-based PEGylation has no impact on the conformational stability of **p17K** and **p18K** relative to **17K** and **18K**, respectively (see supporting information Table S5). In contrast, the **p17/18K** coiled coil is -0.85 ± 0.06 kcal mol⁻¹ more stable than the **17/18K** coiled coil. Because there are three PEG oligomers in each trimeric **p17/18K** coiled coil, we can infer that each PEG-maleimide imparts ~ -0.28 kcal/mol to coiled-coil stability. Peptides **1CW**, **p1CW**, **17K**, **p17K**, **18K**, **p18K**, **17/18K**, and **p17/18K** comprise a triple mutant cycle. As we have done previously,²⁶ we can use variable temperature CD data for these compounds to parse the stabilizing impact of maleimide-based PEGylation on **p17/18K** into several components, including the intrinsic impact of the PEG-maleimide on coiled-coil stability; two-way interactions between Lys₁₇ or Lys₁₈ and the PEG-maleimide; and a three-way interaction among the PEG-maleimide, Lys₁₇, and Lys₁₈ (each component represents the total contribution of the three PEG-maleimides).

Triple mutant cycle analysis reveals that maleimide-based PEGylation of Cys₁₄ is intrinsically destabilizing ($\Delta\Delta G_f = 0.26$ kcal mol⁻¹), but this effect is offset by favorable Lys/PEG-maleimide interactions in **p17/18K** (Figure 5-2). Two-way interactions between Lys₁₇ or Lys₁₈ and the PEG-maleimide contribute a total of ~ -0.27 and -0.22 kcal mol⁻¹, respectively, to coiled-

coil quaternary structural stability. An additional $-0.62 \text{ kcal mol}^{-1}$ of synergistic stabilization comes from a three-way interaction among the PEG-maleimide, Lys₁₇, and Lys₁₈, suggesting that optimal PEG-based stabilization requires the presence of Lys residues at both the $i+3$ and $i+4$ positions relative to the PEGylated i -position Cys.

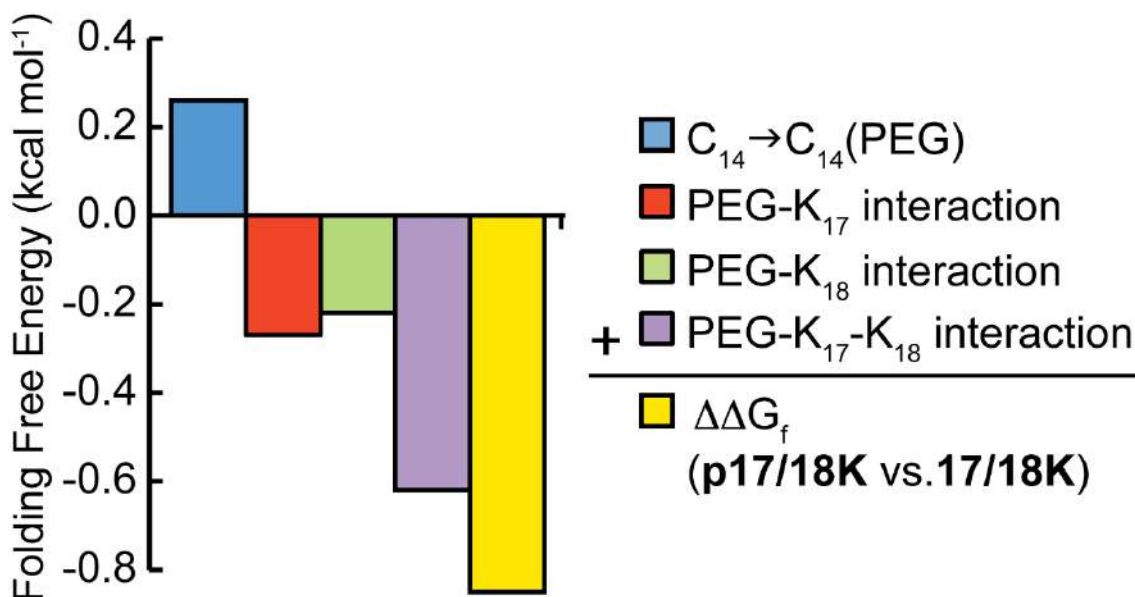


Figure 5-2. Triple mutant cycle analysis of **1CW**, **p1CW 17K**, **p17K, 18K**, **p18K 17/18K**, and **p17/18K**: parsing of PEG-based stabilization of **p17/18K** relative to **17/18K** into the intrinsic impact of maleimide-based PEGylation along with interactions among the PEG-maleimide, Lys₁₇, and Lys₁₈.

We wondered whether the length of PEG affects the increase in conformational stability associated with attaching the PEG-maleimide to the Cys_{*i*}-Lys_{*i*+3}-Lys_{*i*+4} triad in **p17/18K**. To test this hypothesis, we prepared peptides **p17/18K-1kDa** and **m17/18K**, in which a 1000 Da PEG-maleimide and a non-PEGylated maleimide have been attached to the Cys at position 14 in **17/18K**, respectively (Figure 5-3). Variable temperature CD experiments reveal that **p17/18K-1kDa** is slightly more stable than its counterpart **p17/18K** (which harbors a 2000 Da PEG), though the exact difference in stability is challenging to extract from the CD data because **p17/18K-1kDa**

does not reach a fully unfolded conformation even at the highest temperature sampled. We estimate that **p17/18K-1kDa** is -1.2 ± 0.1 kcal mol⁻¹ more stable than non-PEGylated **17/18K** and is -0.33 ± 0.06 kcal mol⁻¹ more stable than **p17/18K**. Surprisingly, non-PEGylated maleimide-modified **m17/18K** is the most stable variant of all; we estimate that **m17/18K** is -2.3 ± 0.1 kcal mol⁻¹ more stable than non-PEGylated **17/18K** and is -1.4 ± 0.1 kcal mol⁻¹ more stable than **p17/18K** (again, these stability differences are estimates as **m17/18K** does not reach a fully unfolded conformation at the highest temperature sampled).

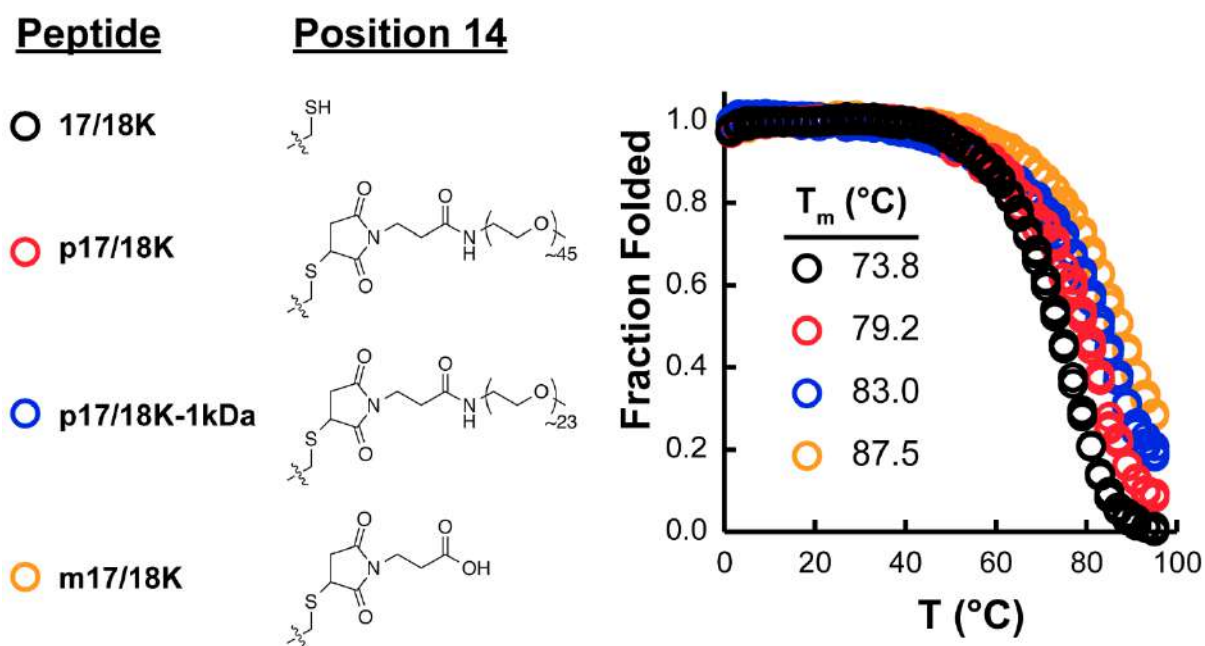


Figure 5-3. Variable temperature CD data for **17/18K**, **p17/18K**, **p17/18K-1kDa** (which harbors a 1000 Da PEG-maleimide), and **m17/18K** (which has a Cys-linked maleimide, but no PEG).

Notwithstanding predictions of previous molecular dynamics simulations to the contrary,^{25,26} these results strongly suggest that the increased conformational stability of **p17/18K** relative to non-PEGylated **17/18K** does not come from favorable interactions between Lys₁₇,

Lys₁₈, and the electronegative oxygen atoms of PEG. Instead, the stabilization appears to come from favorable interactions between Lys₁₇, Lys₁₈, and the maleimide group, possibly between the cationic ε-ammonium groups of Lys₁₇ and Lys₁₈ and the carbonyl oxygen atoms of the maleimide. Adding PEG to the maleimide decreases the stability of these interactions in a length dependent manner, with longer oligomers leading to less substantial stabilization, possibly by increasing the conformational entropy of the PEG-functionalized maleimide group. However, we note that attaching the 2000 Da PEG-maleimide to the Cys_{*i*}-Lys_{*i*+3}-Lys_{*i*+4} triad in **17/18K** is still a stabilizing modification overall ($\Delta\Delta G_f = -0.85 \pm 0.06$ kcal mol⁻¹), suggesting that the Cys_{*i*}-Lys_{*i*+3}-Lys_{*i*+4} triad is a better site for PEG-maleimide modification than locations that lack the *i*+3 and *i*+4 Lys residues.

We wondered whether attaching the 2000 Da PEG-maleimide to the Cys_{*i*}-Lys_{*i*+3}-Lys_{*i*+4} triad in other helical contexts might result in similar stabilization. To test this hypothesis, we turned to a truncated variant of the villin headpiece domain (**VHP**), an extensively characterized 35-residue protein in which three α-helices pack around a core composed of non-polar and aromatic amino acids (Figure 5-4A,B).²⁷⁻²⁹ We began with destabilized variant **10/12L**, instead of wild-type **VHP** (Figure 5-4A),²⁸ because we feared the high thermal stability of **VHP** would make characterizing PEG-based stabilization difficult. The third turn of the C-terminal helix of **10/12L** contains two adjacent Lys residues: Lys₂₉ and Lys₃₀. Replacing Gln₂₆ with Cys gives protein **VHP1**, which harbors the complete Cys_{*i*}-Lys_{*i*+3}-Lys_{*i*+4} triad. We generated PEGylated **pVHP1** by modifying Cys₂₆ in **VHP1** with a 2000 Da PEG-maleimide.

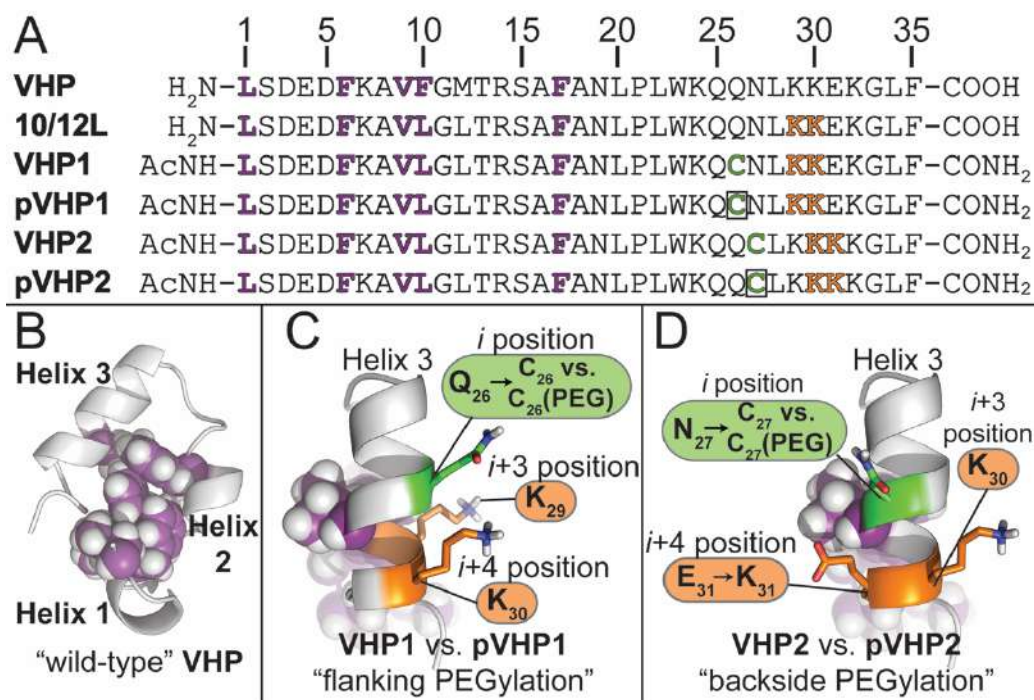


Figure 5-4. (A) Sequences of proteins **VHP**, **10/12L**, **VHP1**, **pVHP1**, **VHP2**, and **pVHP2**; box denotes Cys modified with a 2000 Da PEG-maleimide (B) Ribbon diagram of **VHP** (PDB ID: 1VII), with hydrophobic core side-chain atoms shown as spheres. (C) and (D) Ribbon diagrams showing placement of the Cys_{*i*}-Lys_{*i*+3}-Lys_{*i*+4} triad at flanking and backside positions, respectively, in the C-terminal helix of **VHP1** and **VHP2**.

The T_m of PEGylated **pVHP1** is 3.4 ± 0.1 °C higher than that of unmodified **VHP1**, corresponding to an increase in conformational stability of -0.18 ± 0.01 kcal mol⁻¹. This result demonstrates maleimide-based PEGylation of the Cys_{*i*}-Lys_{*i*+3}-Lys_{*i*+4} triad can stabilize proteins other than **p17/18K**. However, the increased stability of **pVHP1** relative to **VHP1** is somewhat lower than we observed previously for **p17/18K** relative to **17/18K** (~ -0.28 kcal/mol per triad), possibly because the Cys_{*i*}-Lys_{*i*+3}-Lys_{*i*+4} triad in **pVHP1** flanks the hydrophobic core of the protein (Figure 5-4C), where steric clashes with PEG might limit stabilization.

We explored this possibility by shifting the Cys_{*i*}-Lys_{*i*+3}-Lys_{*i*+4} triad one position toward the C-terminal end of the helix to generate **VHP2** and PEGylated **pVHP2**. The Cys_{*i*}-Lys_{*i*+3}-Lys_{*i*+4}

triad in these variants occupies the solvent-exposed “backside” of the helix, as far from the hydrophobic core as possible (Figure 5-4D). Maleimide-based PEGylation increases the T_m of **pVHP2** by 9.1 ± 0.4 °C relative to **VHP2**, corresponding to an increase in conformational stability of -0.29 ± 0.01 kcal mol⁻¹, an increment that more closely matches what we observed in **p17/18K**, suggesting that modifying the Cys_{*i*}-Lys_{*i*+3}-Lys_{*i*+4} triad with a PEG-maleimide is a general way to stabilize α -helical proteins.

5.3 Conclusions

We have shown that favorable interactions between an *i* position PEG-maleimide and two *i*+3 and *i*+4 position Lys residues stabilizes two different α -helical proteins. Our results suggest that the Cys_{*i*}-Lys_{*i*+3}-Lys_{*i*+4} triad is a simple sequence motif for identifying or installing stabilizing maleimide-based PEGylation sites within α -helical proteins. Many important therapeutic proteins contain α -helices (e.g. erythropoietin, granulocyte colony stimulating factor, interferon α , etc.); application of the Cys_{*i*}-Lys_{*i*+3}-Lys_{*i*+4} in these contexts should accelerate the development of PEGylated protein drugs with enhanced conformational stability (and therefore enhanced resistance to proteolysis, aggregation, and immunogenicity). Moreover, the fact that modifying the Cys_{*i*}-Lys_{*i*+3}-Lys_{*i*+4} triad with maleimide alone results in even greater increases to conformational stability suggests that a Cys_{*i*}-Lys_{*i*+3}-Lys_{*i*+4} triad, positioned carefully in an α -helix could be an ideal site for using maleimide chemistry to immobilize helical proteins on solid supports or to functionalize them with fluorophores or drug conjugates.

5.4 Contributions

Contributions of Brijesh Pandey, a former post-doctoral researcher in the Price lab, were significant to this Chapter; When the data were published Mason and Brijesh shared first authorship.³⁰ Brijesh demonstrated the PEG based Helix stability in the 1CW model system (Figures 5-1, 5-2). Mason demonstrated PEG-based helix stability in the VHP system and did the length dependence study showing that the maleimide linker contributed the most stability to the 1CW coiled coil (Figure 5-3, 5-4).

5.5 Supporting Information

5.5.1 Protein Synthesis

Peptides **1CW**, **7A**, **15A**, **21A**, **17K**, **18K**, **17/18K**, **VHP1**, **VHP2**, and their PEGylated counterparts **p1CW**, **p7A**, **p15A**, **p21A**, **p17K**, **p18K**, **p17/18K**, **pVHP1**, and **pVHP2** (Table S1) were synthesized as C-terminal amides, by microwave-assisted solid-phase peptide synthesis, using a standard Fmoc N α protection strategy as described previously in chapter 2 section 2.4.1.¹⁶

We prepared PEGylated proteins **p1CW**, **p7A**, **p15A**, **p21A**, **p17K**, **p18K**, **p17/18K**, **pVHP1**, and **pVHP2** by functionalizing the Cys₁₄ in proteins **1CW**, **7A**, **15A**, **21A**, **17K**, **18K**, **17/18K**, **VHP1**, **VHP2** with a 2 kDa monomethoxy PEG maleimide (average PEG M.W. = 2000 Da; Jenkem Technology, ZZ099P356) according to the following general protocol: Following initial purification of 25 μ mol crude **1CW** variant, half of the resulting purified material (assumed to be \sim 12.5 μ mol) was dissolved in 2mL of milli-Q water. The maleimide-functionalized 2 kDa monomethoxy PEG (62.5 μ mol, 5 eq) was then added to the solution and stirred at room temperature for 2 hours. The resulting solution was then stored at -20°C until subjected to further purification and characterization. We used a similar approach to prepare peptide **p17/18K-1kDa**, except we used a 1 kDa monomethoxy PEG-maleimide.

We prepared peptide **m17/18K** by functionalizing Cys₁₄ in peptide **17/18K** with 3-maleimidopropionic acid (M.W. = 169.135, AK scientific) according the following procedure:

500uL of purified **17/18K** (195uM) was diluted to a 1mL solution of 20mM tris (pH = 6.9), approximately 20eq. of maleimidopropionic acid (0.32mg), and 10 eq. of TCEP HCl (0.27mg, Thermo Scientific). The resulting solution was stirred at room temperature for 24 hours and then subjected to reverse phase HPLC chromatography.

5.5.2 Protein Purification and Characterization

Immediately prior to purification, the crude protein was dissolved in 1:1 H₂O/MeCN. Proteins were purified by preparative reverse-phase high performance liquid chromatography (HPLC) on a C18 column using a linear gradient of water in acetonitrile with 0.1% v/v TFA. Fractions containing the desired protein product were pooled, frozen, and lyophilized. Proteins were identified by electrospray ionization time of flight mass spectrometry (ESI-TOF); mass spectra appear below in Figures S1–S29. Protein purity was assessed by Analytical HPLC (Figures S30–S49).

ESI-TOF spectra for proteins **1CW**, **7A**, **15A**, **21A**, **17K**, **18K**, **17/18K**, **VHP1**, **VHP2**, their PEGylated counterparts **p1CW**, **p7A**, **p15A**, **p21A**, **p17K**, **p18K**, **p17/18K**, **pVHP1**, **pVHP2**, **p17/18K-1kDa**, and **m17/18K** are shown in Figures 5-5–5-33.

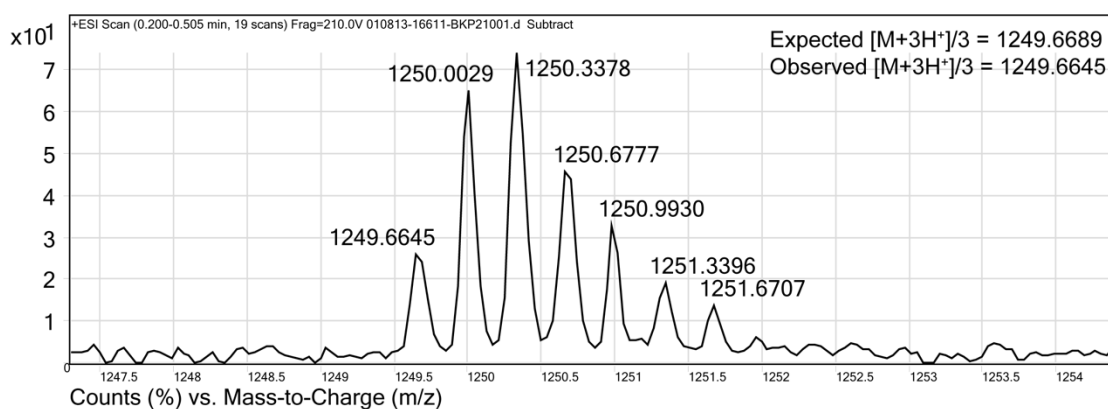


Figure 5-5. ESI TOF spectrum for **1CW**. Expected $[M+3H]^{3+}/3 = 1249.6689$ Da. Observed $[M+3H]^{3+}/3 = 1249.6645$ Da.

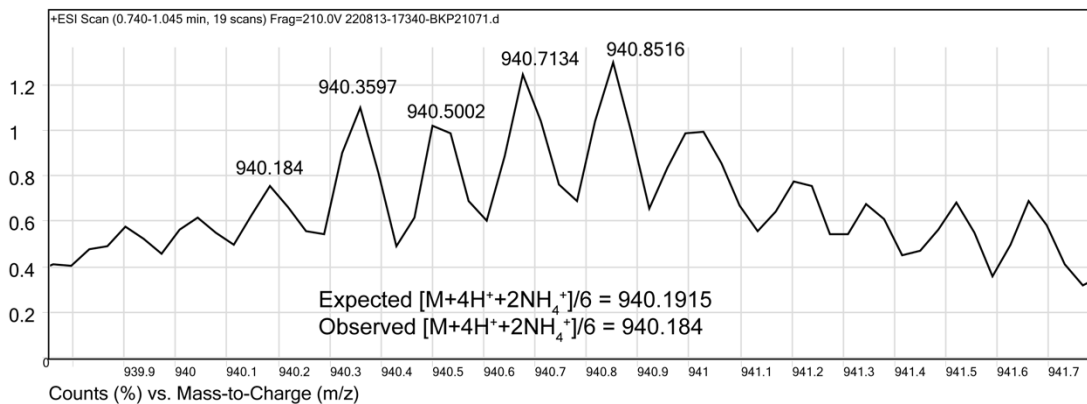


Figure 5-6. ESI TOF spectrum for **p1CW** with 38 PEG units. Expected $[M+4H^++2NH_4^+]/6 = 940.1915$ Da. Observed $[M+4H^++2NH_4^+]/6 = 940.184$ Da.

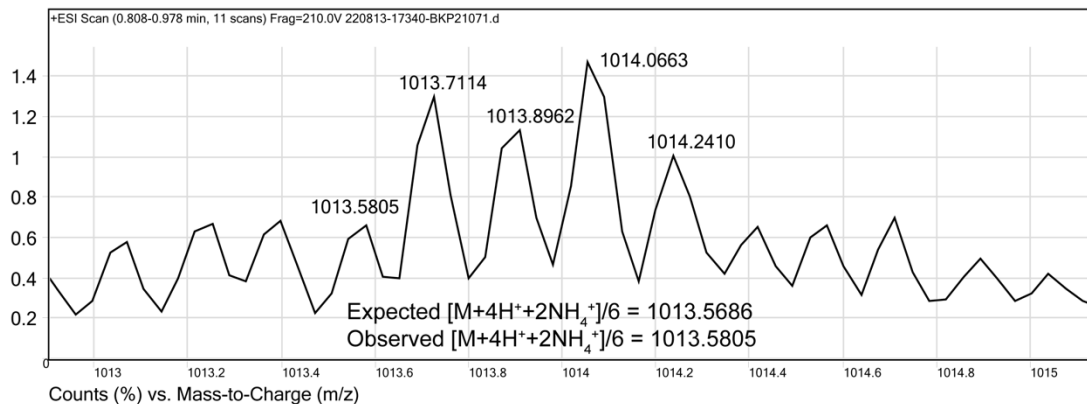


Figure 5-7. ESI TOF spectrum for **p1CW** with 48 PEG units. Expected $[M+4H^++2NH_4^+]/6 = 1013.5686$ Da. Observed $[M+4H^++2NH_4^+]/6 = 1013.58$ Da.

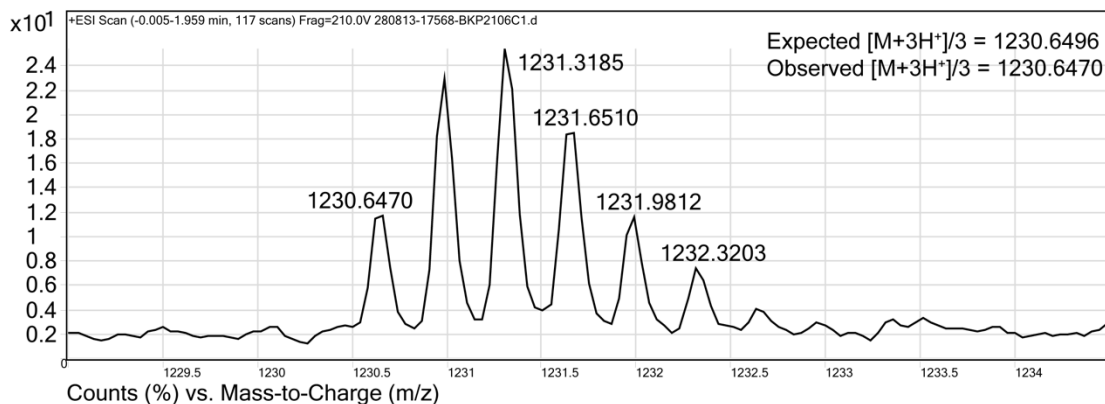


Figure 5-8. ESI TOF spectrum for **7A**. Expected $[M+3H]^3+/3 = 1230.6469$ Da. Observed $[M+3H]^3+/3 = 1230.6470$ Da.

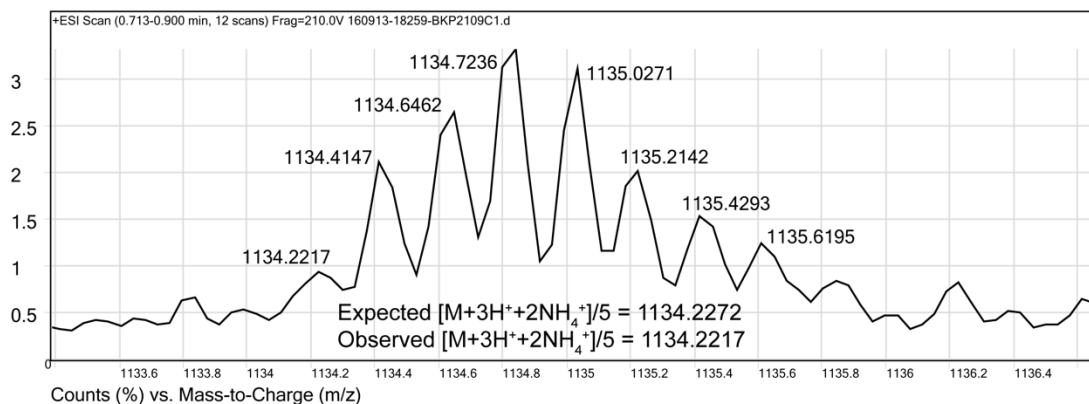


Figure 5-9. ESI TOF spectrum for **p7A** with 40 PEG units. Expected $[M+3H^++2NH_4^+]/5 = 1134.2272$ Da. Observed $[M+3H^++2NH_4^+]/5 = 1134.2217$ Da.

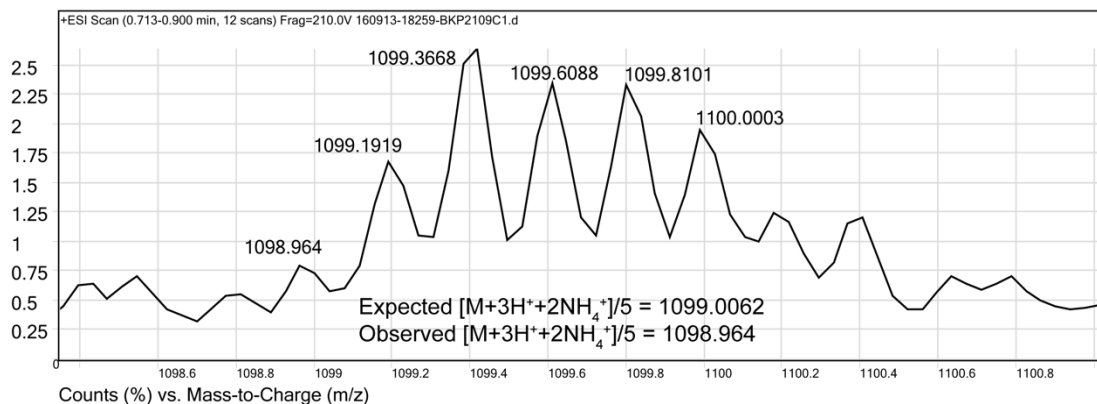


Figure 5-10. ESI TOF spectrum for **p7A** with 36 PEG units. Expected $[M+3H^++2NH_4^+]/5 = 1099.0062$ Da. Observed $[M+3H^++2NH_4^+]/5 = 1098.964$ Da.

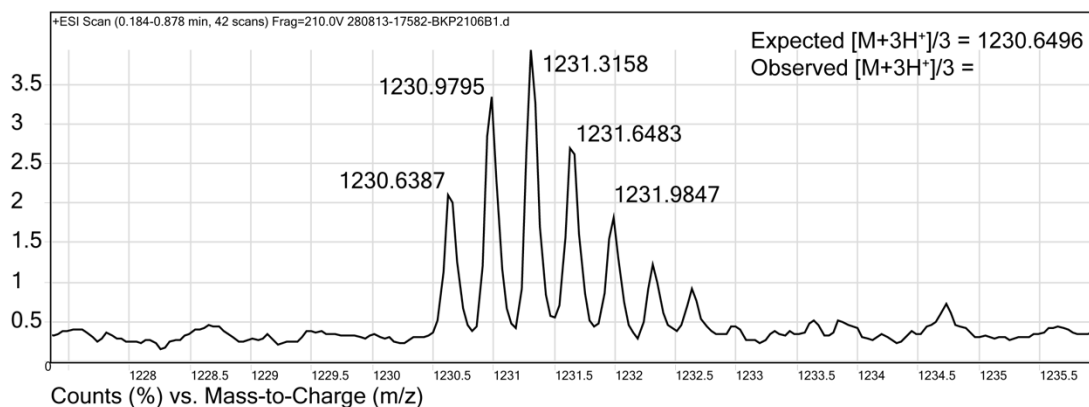


Figure 5-11. ESI TOF spectrum for **15A**. Expected $[M+3H]^3+/3 = 1230.6469$ Da. Observed $[M+3H]^3+/3 = 1230.6387$ Da.

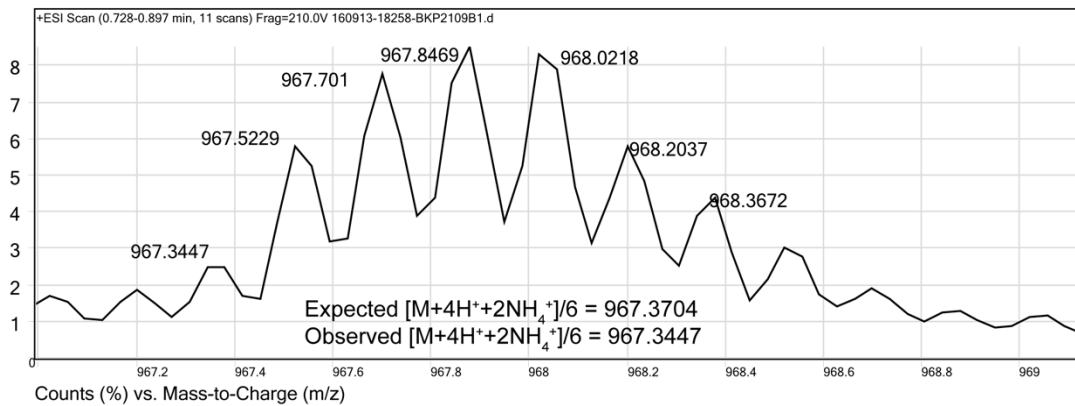


Figure 5-12. ESI TOF spectrum for **p15A** with 43 PEG units. Expected $[M+4H^++2NH_4^+]/6 = 967.3704$ Da. Observed $[M+4H^++2NH_4^+]/6 = 967.3447$ Da.

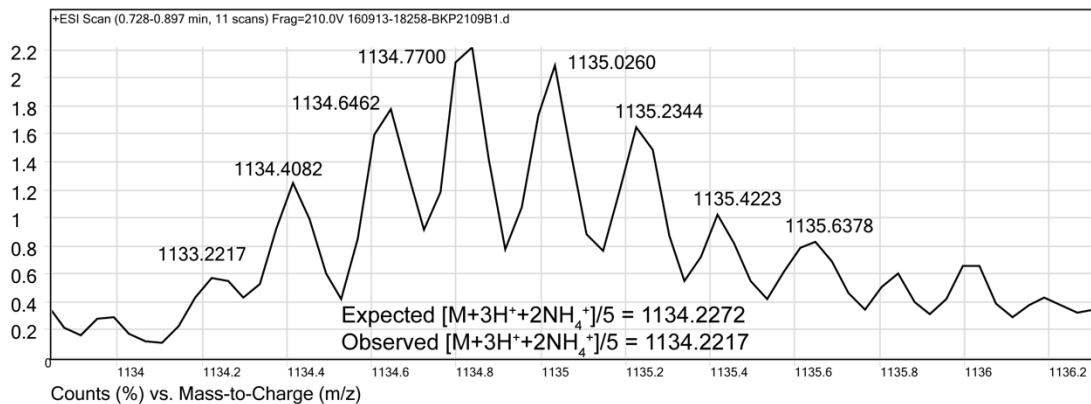


Figure 5-13. ESI TOF spectrum for **p15A** with 40 PEG units. Expected $[M+3H^++2NH_4^+]/5 = 1134.2272$ Da. Observed $[M+3H^++2NH_4^+]/5 = 1134.2217$ Da.

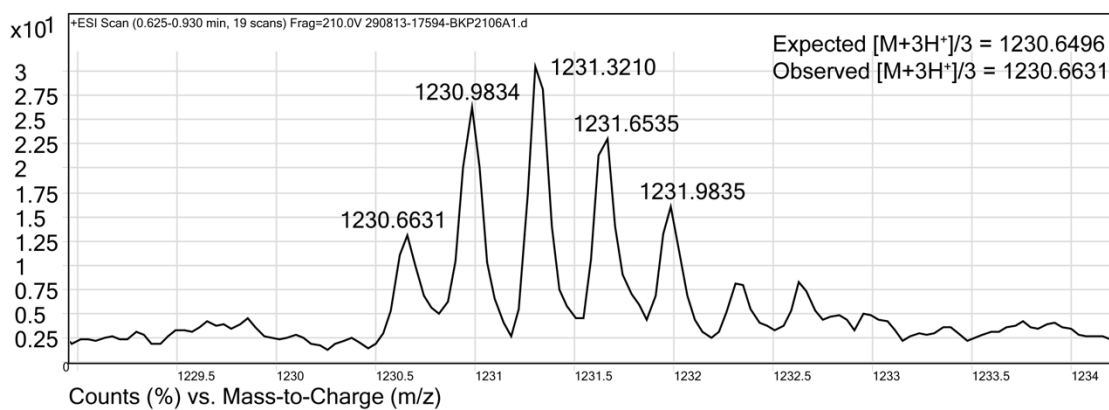


Figure 5-14. ESI TOF spectrum for **K21A**. Expected $[M+3H]^3+/3 = 1230.6469$ Da. Observed $[M+3H]^3+/3 = 1230.6631$ Da.

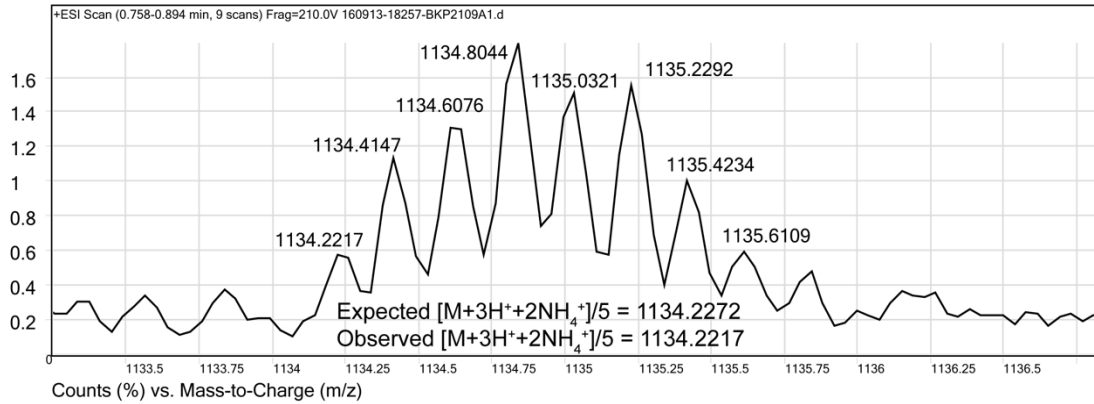


Figure 5-15. ESI TOF spectrum for **p21A** with 40 PEG units. Expected $[M+3H^++2NH_4^+]/5 = 1134.2272$ Da. Observed $[M+3H^++2NH_4^+]/5 = 1134.2217$ Da.

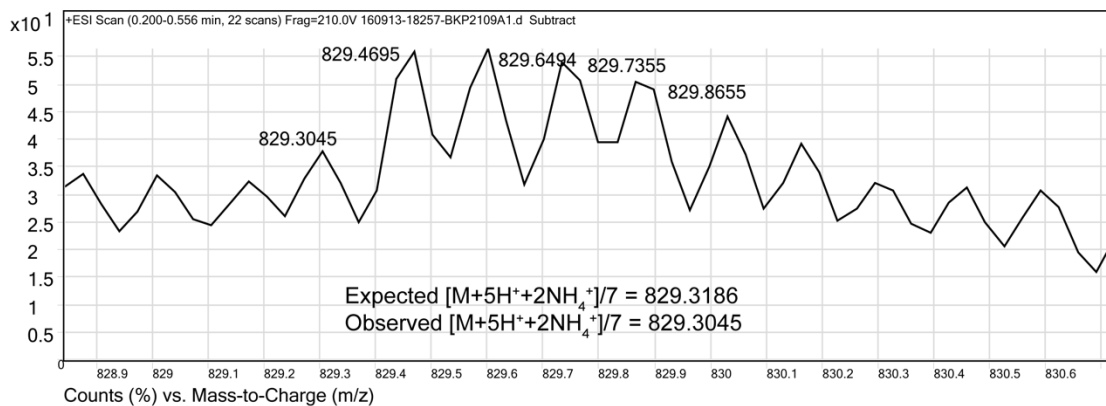


Figure 5-16. ESI TOF spectrum for **p21A** with 43 PEG units. Expected $[M+5H^++2NH_4^+]/7 = 829.3186$ Da. Observed $[M+5H^++2NH_4^+]/7 = 829.3045$ Da.

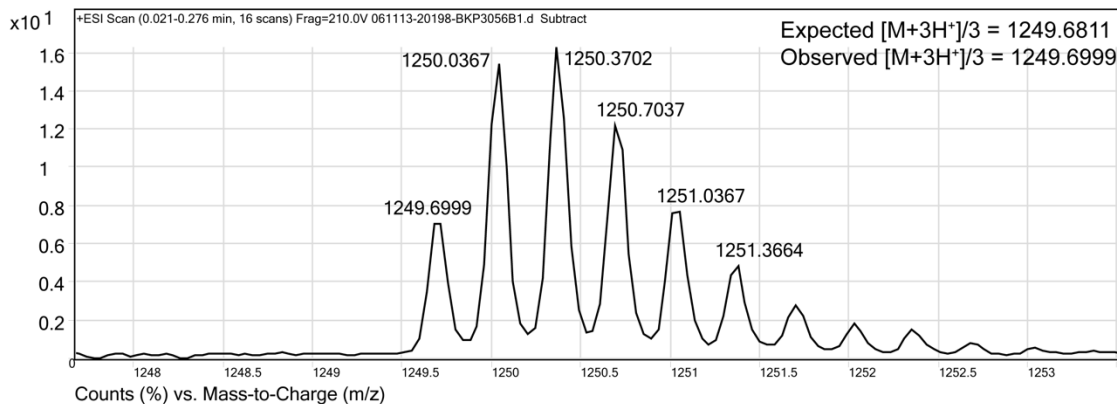


Figure 5-17. ESI TOF spectrum for **17K**. Expected $[M+3H]^3+/3 = 1249.6811$ Da. Observed $[M+3H]^3+/3 = 1249.6999$ Da.

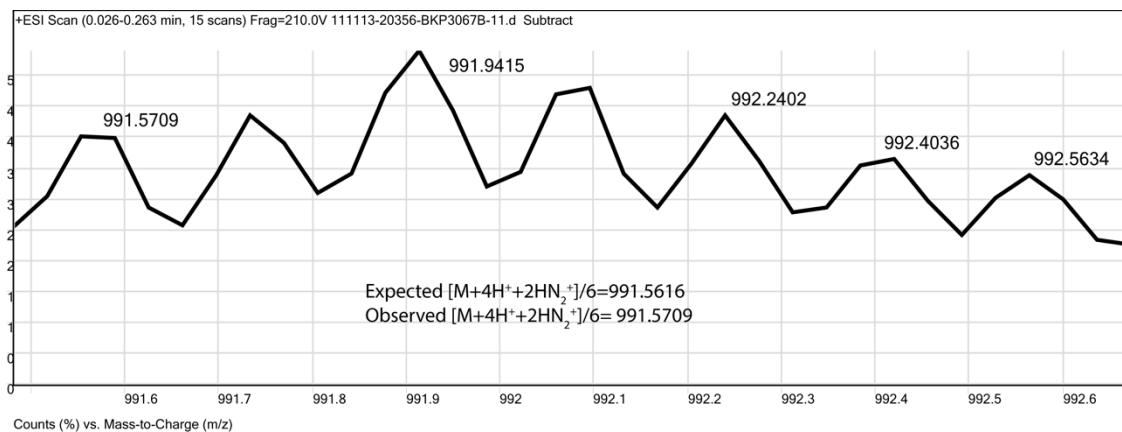


Figure 5-18. ESI TOF spectrum for **p17K** with 45 PEG units. Expected $[M+4H^++2NH_4^+]/6 = 991.5616$ Da. Observed $[M+4H^++2NH_4^+]/6 = 991.5709$ Da.

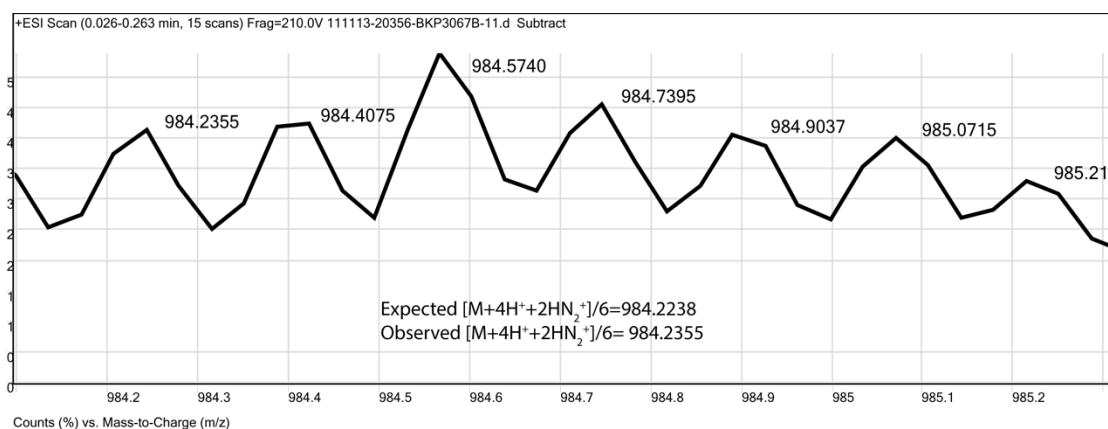


Figure 5-19. ESI TOF spectrum for **p17K** with 44 PEG units. Expected $[M+4H^++2NH_4^+]/6 = 984.2238$ Da. Observed $[M+4H^++2NH_4^+]/6 = 984.2355$ Da.

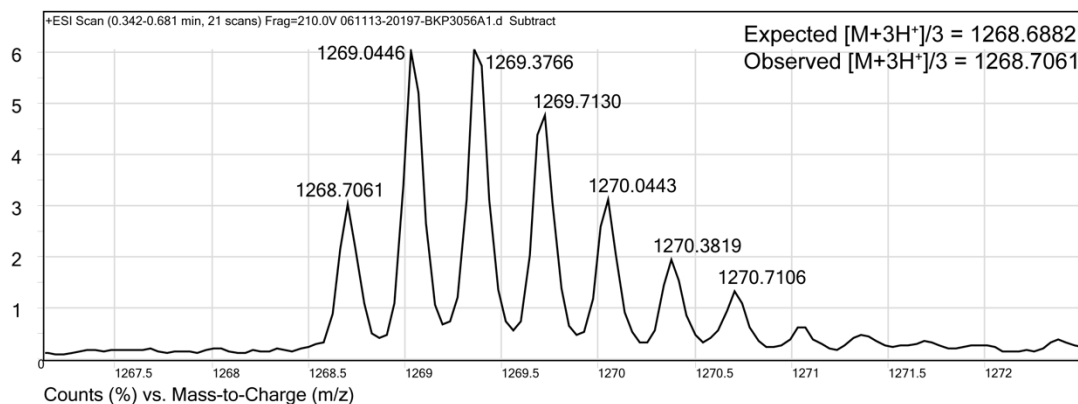


Figure 5-20. ESI TOF spectrum for **18K**. Expected $[M+3H]^{3+}/3 = 1268.6882$ Da. Observed $[M+3H]^{3+}/3 = 1268.7061$ Da.

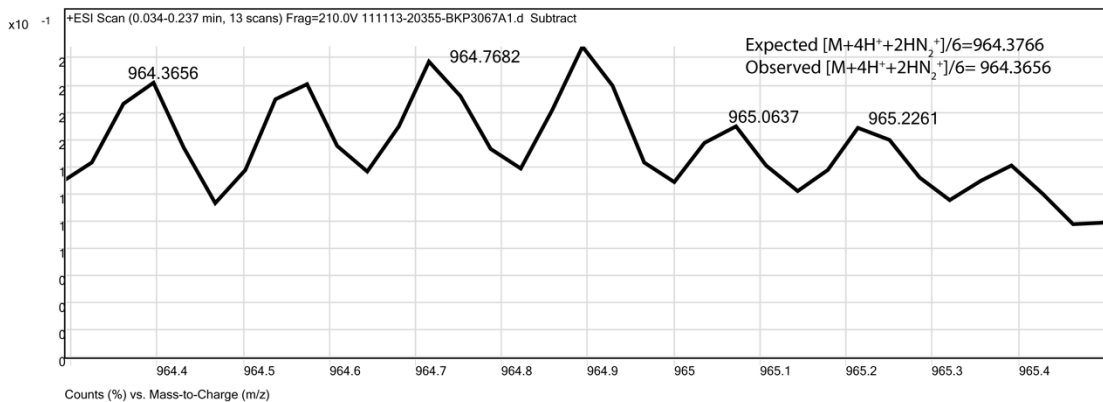


Figure 5-21. ESI TOF spectrum for **p18K** with 40 PEG units. Expected $[M+4H^++2NH_4^+]/6 = 964.3766$ Da. $[M+4H^++2NH_4^+]/6 = 964.3656$ Da.

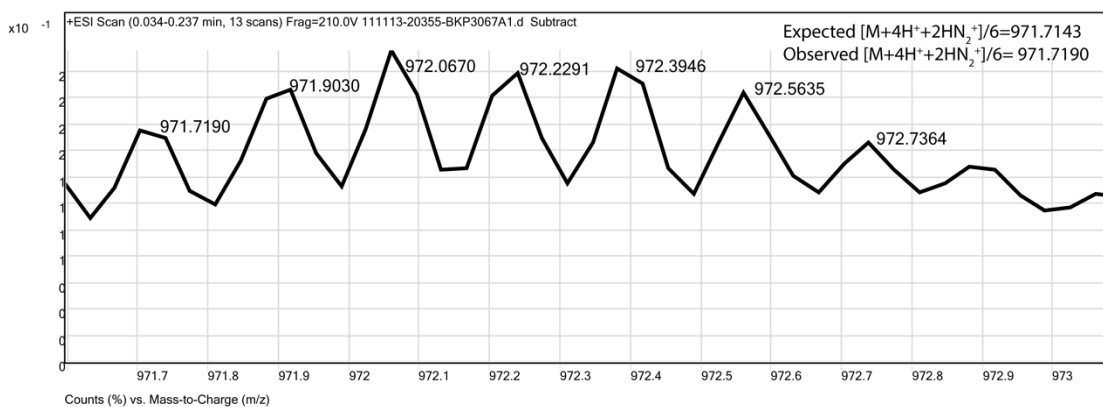


Figure 5-22. ESI TOF spectrum for **p18K** with 41 PEG units. Expected $[M+4H^++2NH_4^+]/6 = 971.7143$ Da. $[M+4H^++2NH_4^+]/6 = 971.7190$ Da.

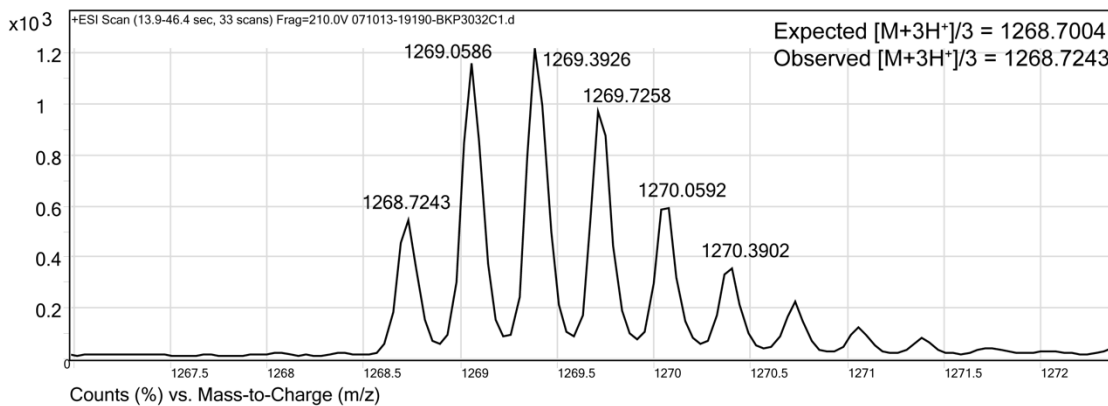


Figure 5-23. ESI TOF spectrum for **17K/18K**. Expected $[M+3H]^3+/3 = 1268.7004$ Da. Observed $[M+3H]^3+/3 = 1268.7243$ Da.

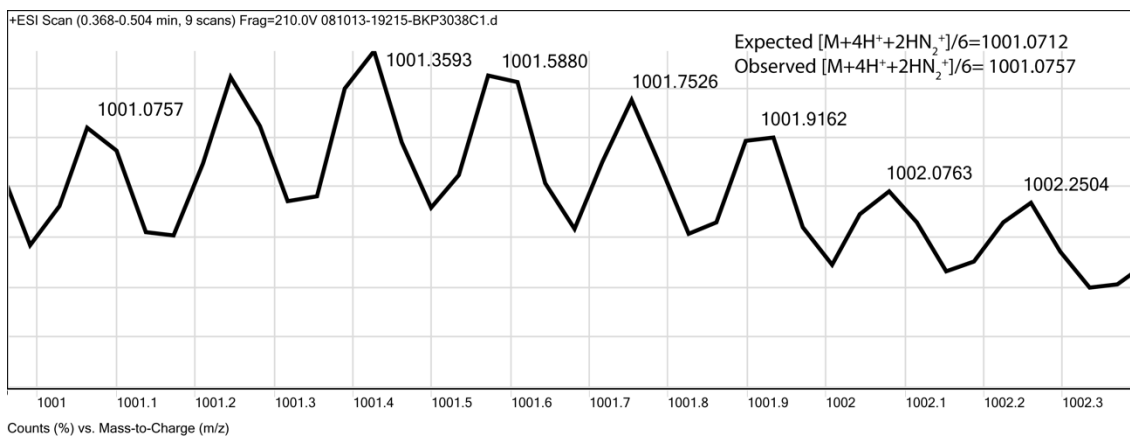


Figure 5-24. ESI TOF spectrum for **p17K/18K** with 45 PEG units. Expected $[M+4H^++2NH_4^+]/6 = 1001.0712$ Da. Observed $[M+4H^++2NH_4^+]/6 = 1001.0757$ Da.

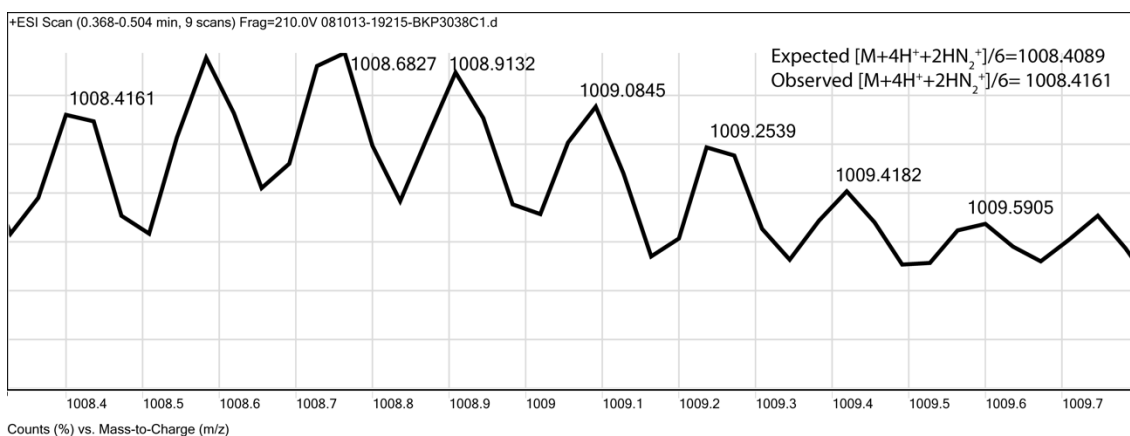


Figure 5-25. ESI TOF spectrum for **p17K/18K** with 46 PEG units. Expected $[M+4H^++2NH_4^+]/6 = 1008.4089$ Da. Observed $[M+4H^++2NH_4^+]/6 = 1008.4161$ Da.

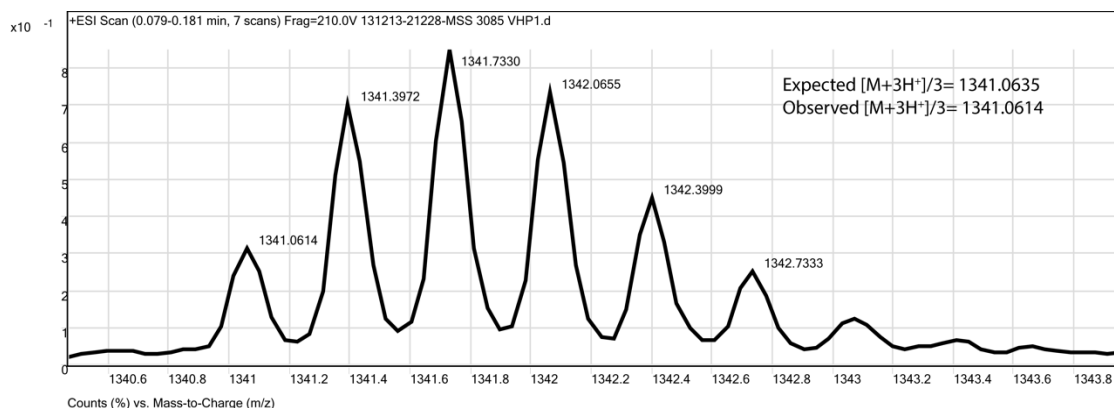


Figure 5-26. ESI TOF spectrum for **VHP1**. Expected $[M+3H]^{3+}/3 = 1341.0635$ Da. Observed $[M+3H]^{3+}/3 = 1341.0614$.

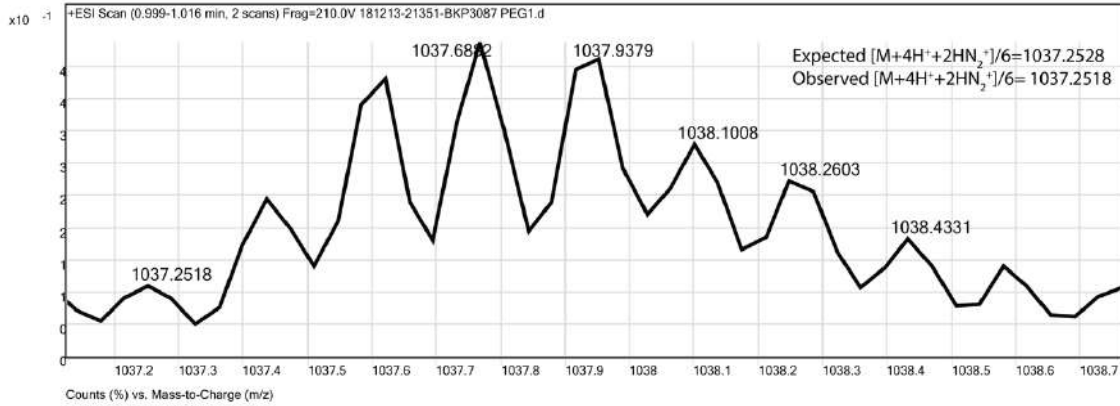


Figure 5-27. ESI TOF spectrum for **pVHP1** with a 45 unit PEG. Expected $[M+4H^++2NH_4^+]/6 = 1037.2528$ Da. Observed $[M+4H^++2NH_4^+]/6 = 1037.2518$ Da.

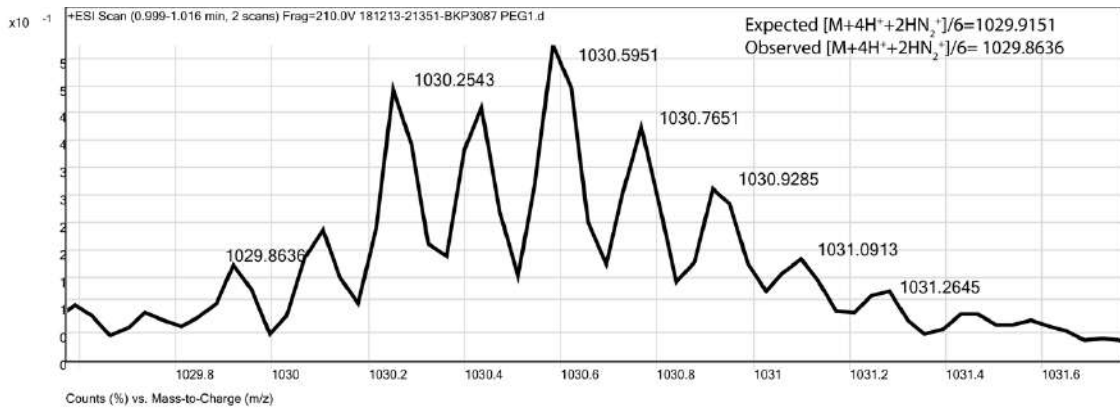


Figure 5-28. ESI TOF spectrum for **pVHP1**. Expected with a 44 unit PEG. Expected $[M+4H^++2NH_4^+]/6 = 1029.9151$ Da. Observed $[M+4H^++2NH_4^+]/6 = 1029.8636$ Da.

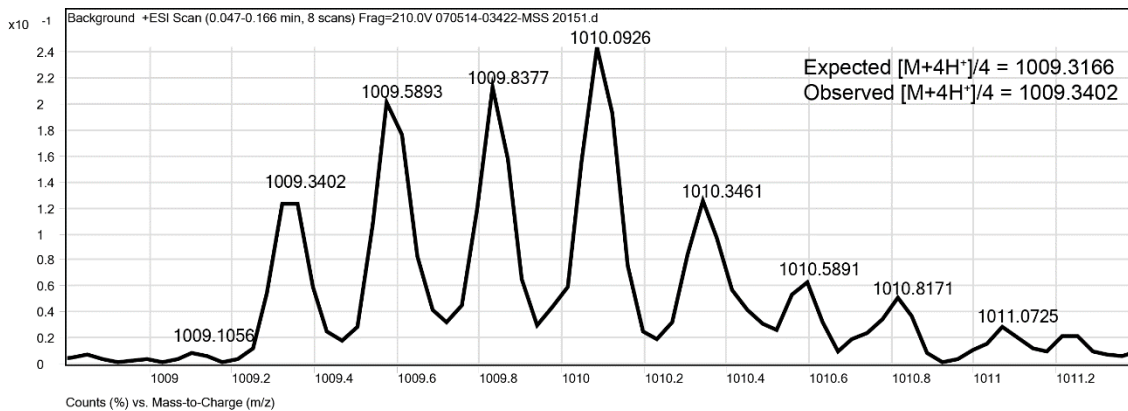


Figure 5-29. ESI TOF spectrum for **VHP2**. Expected $[M+4H^+]/4 = 1009.3166$ Da. Observed $[M+4H^+]/4 = 1009.3402$ Da.

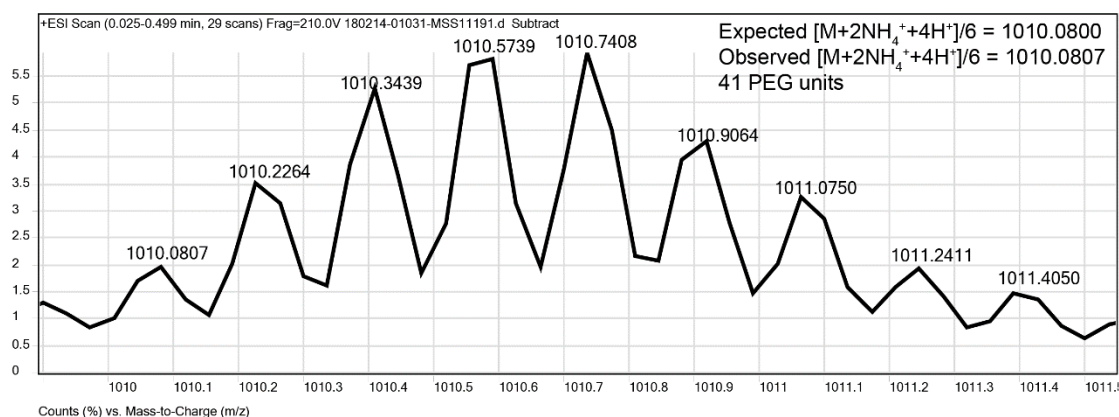


Figure 5-30. ESI TOF spectrum for **pVHP2** with 41 PEG units. Expected $[M+4H^++2NH_4^+]/6 = 1010.0800$ Da. Observed $[M+4H^++2NH_4^+]/6 = 1010.0807$ Da.

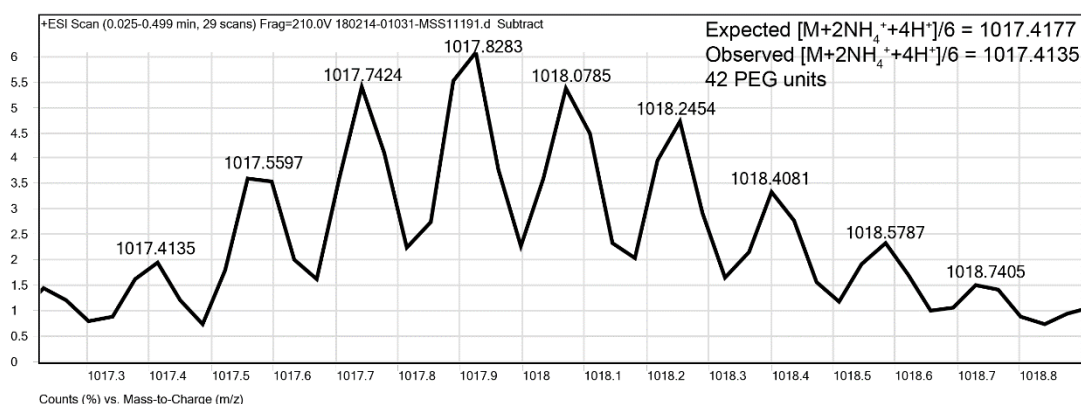


Figure 5-31. ESI TOF spectrum for **pVHP2** with 42 PEG units. Expected $[M+4H^++2NH_4^+]/6 = 1017.4177$ Da. Observed $[M+4H^++2NH_4^+]/6 = 1017.4135$ Da.

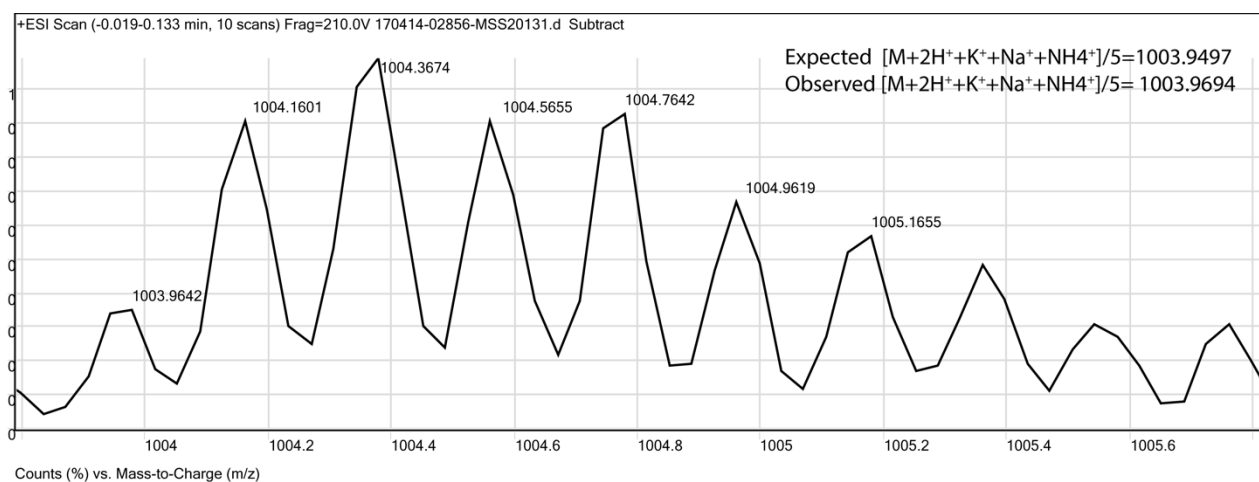


Figure 5-32. ESI TOF spectrum for **p17/18K-1kDa**. Expected $[M+2H^++K^++Na^++NH_4^+]/5 = 1003.9497$ Da. Observed $[M+2H^++K^++Na^++NH_4^+]/5 = 1003.9694$ Da.

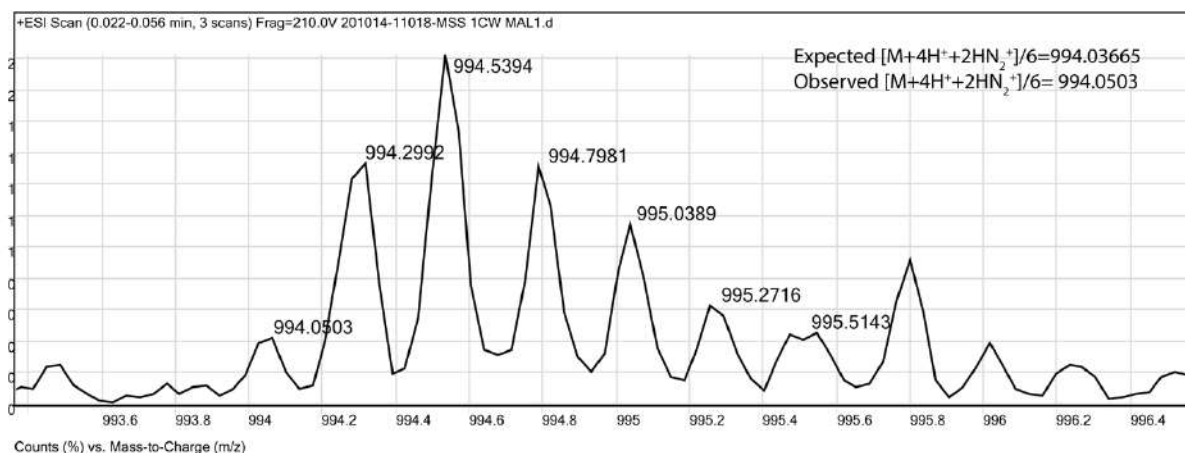


Figure 5-33. ESI TOF spectrum for **m17/18K**. Expected $[M+4H^+]/4 = 994.03665$ Da. Observed $[M+4H^+]/4 = 994.0503$ Da.

HPLC traces for proteins **1CW**, **7A**, **15A**, **21A**, **17K**, **18K**, **17/18K**, **VHP1**, **VHP2**, and their PEGylated counterparts **p1CW**, **p7A**, **p15A**, **p21A**, **p17K**, **p18K**, **p17/18K**, **pVHP1**, **pVHP2**, **p17/18K-1kDa**, and **m17/18K** are shown in Figures 5-34–5-53.

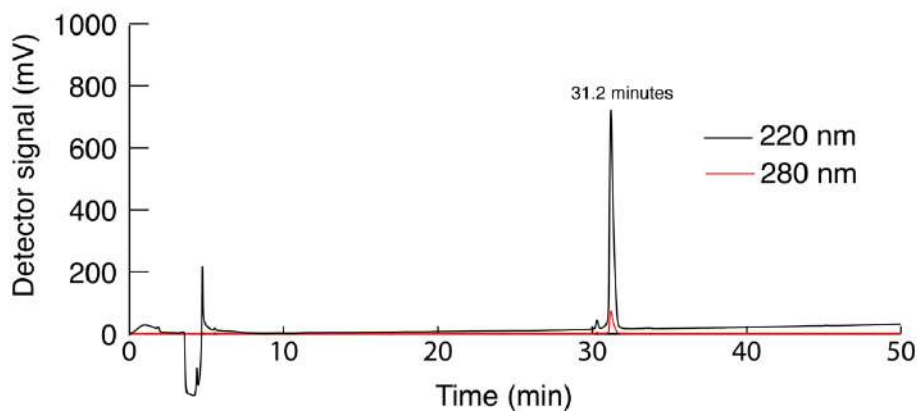


Figure 5-34. Analytical HPLC Data for **1CW**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

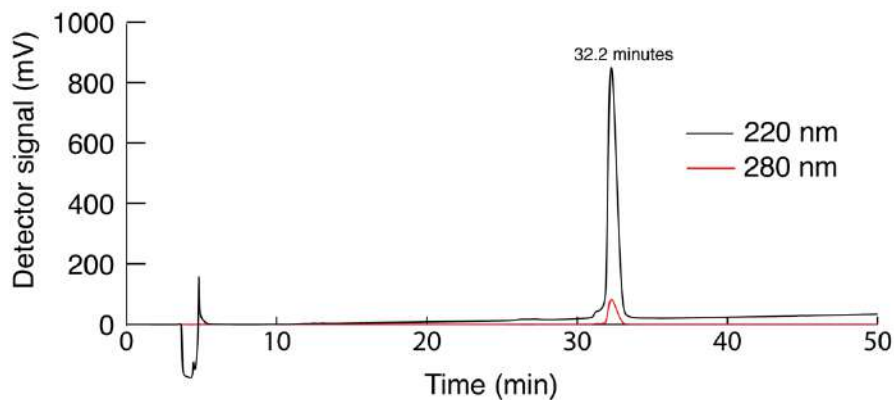


Figure 5-35. Analytical HPLC Data for **p1CW**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

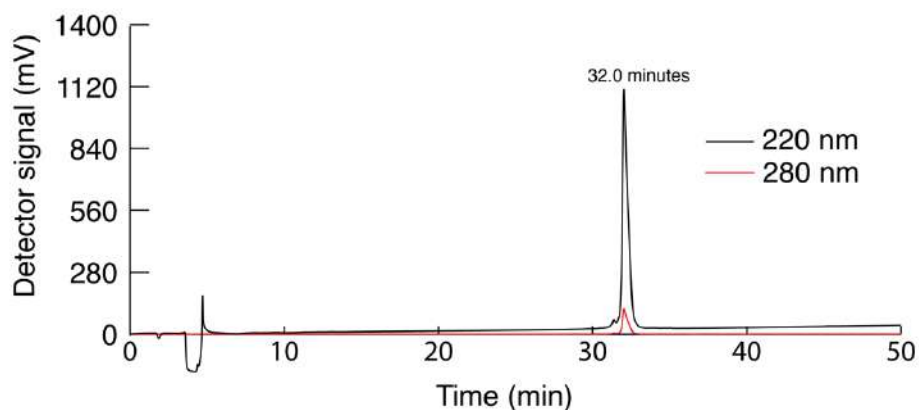


Figure 5-36. Analytical HPLC Data for **7A**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

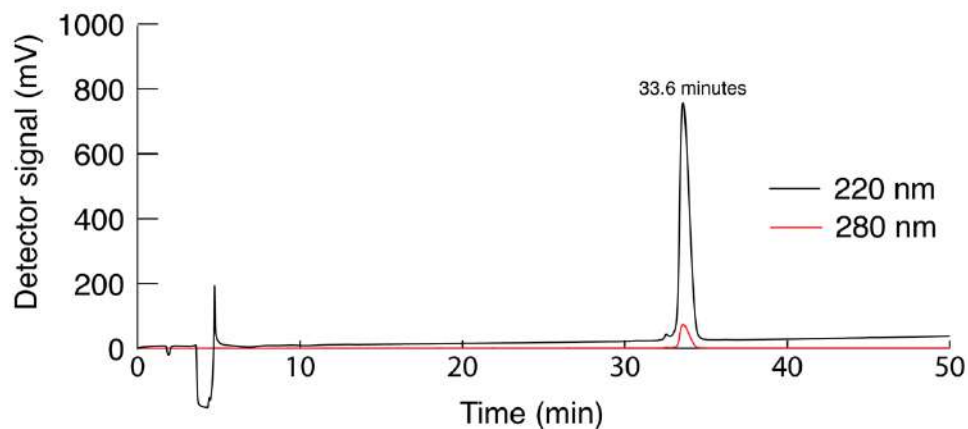


Figure 5-37. Analytical HPLC Data for **p7A**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

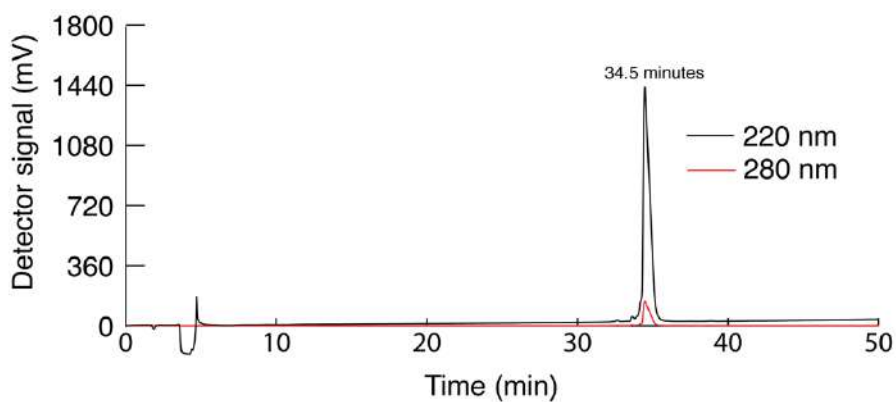


Figure 5-38. Analytical HPLC Data for **15A**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

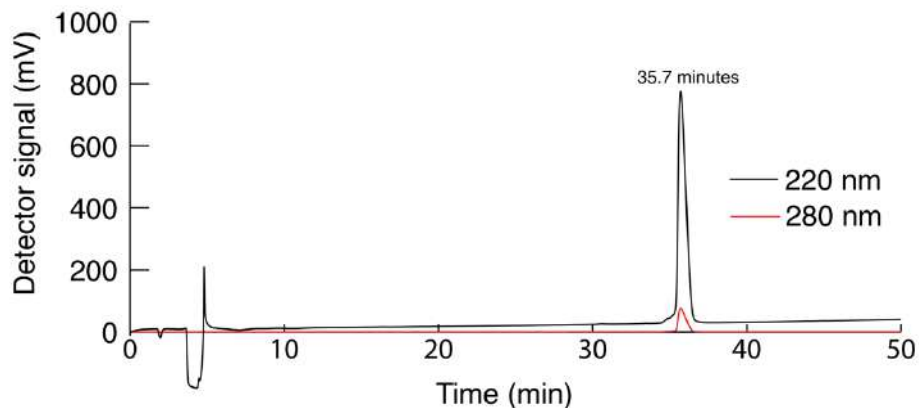


Figure 5-39. Analytical HPLC Data for **p15A**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

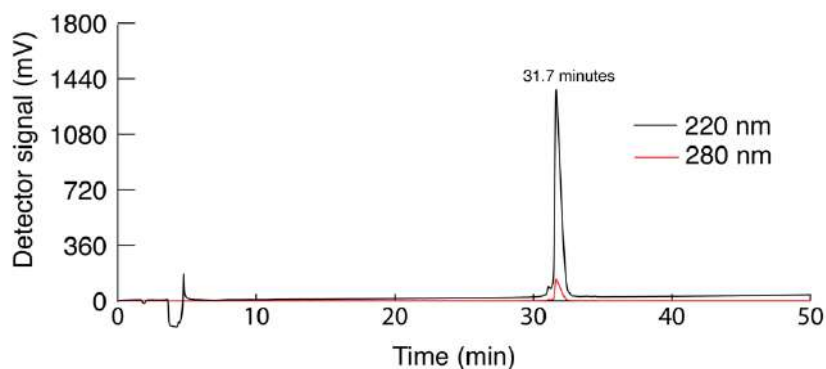


Figure 5-40. Analytical HPLC Data for **21A**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

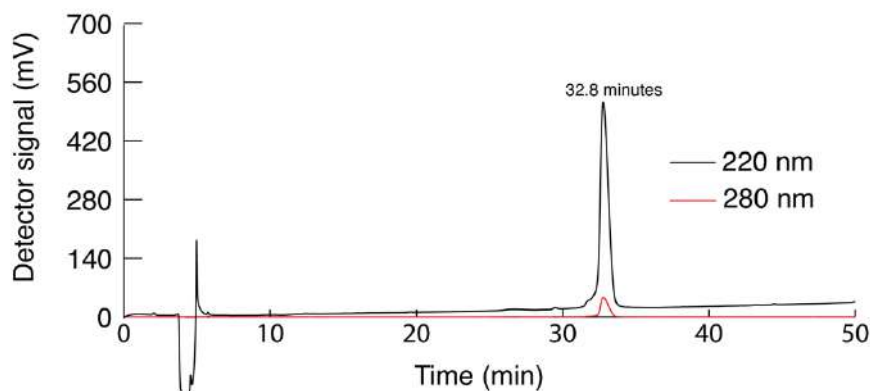


Figure 5-41. Analytical HPLC Data for **p21A**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

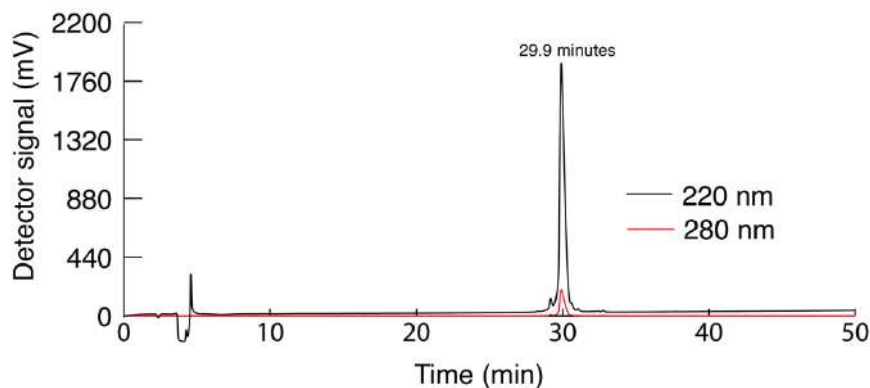


Figure 5-42. Analytical HPLC Data for **17K**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

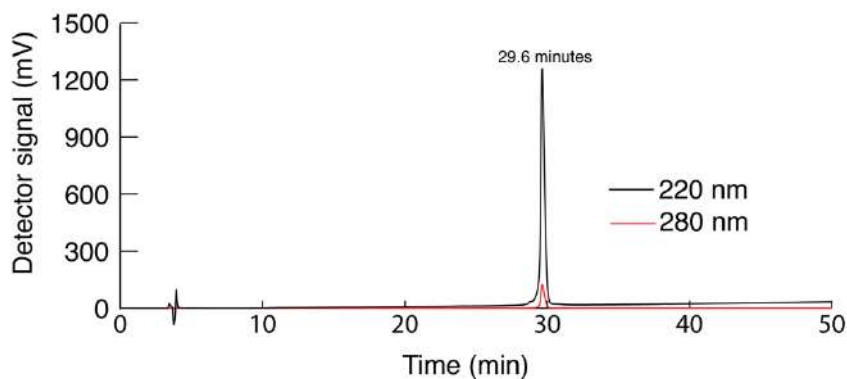


Figure 5-43. Analytical HPLC Data for **p17K**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

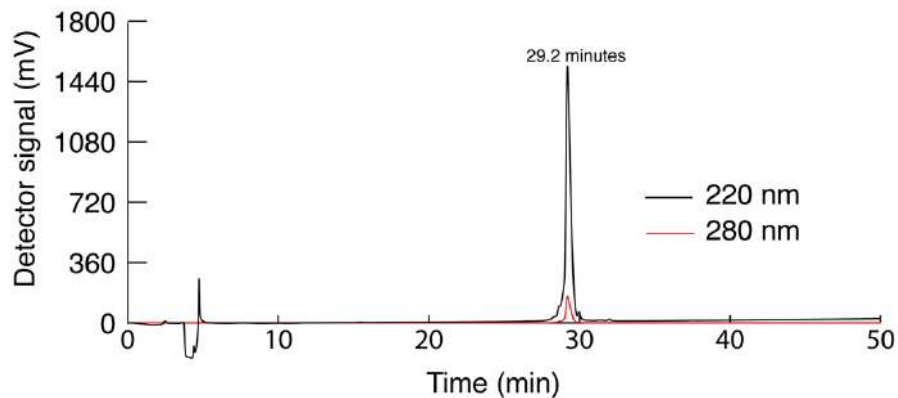


Figure 5-44. Analytical HPLC Data for **18K**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

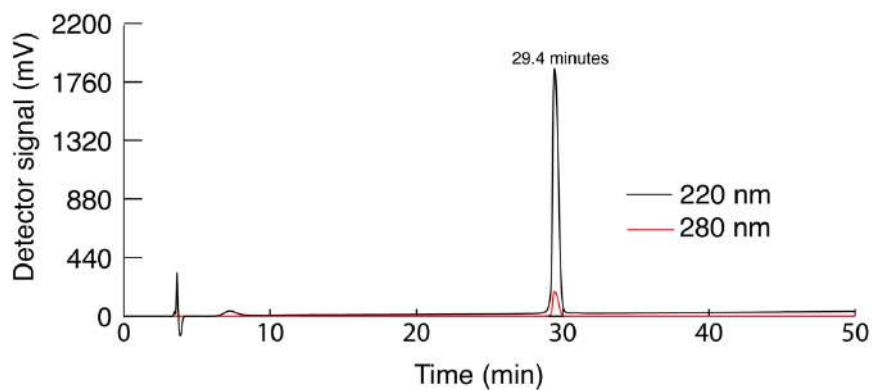


Figure 5-45. Analytical HPLC Data for **p18K**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

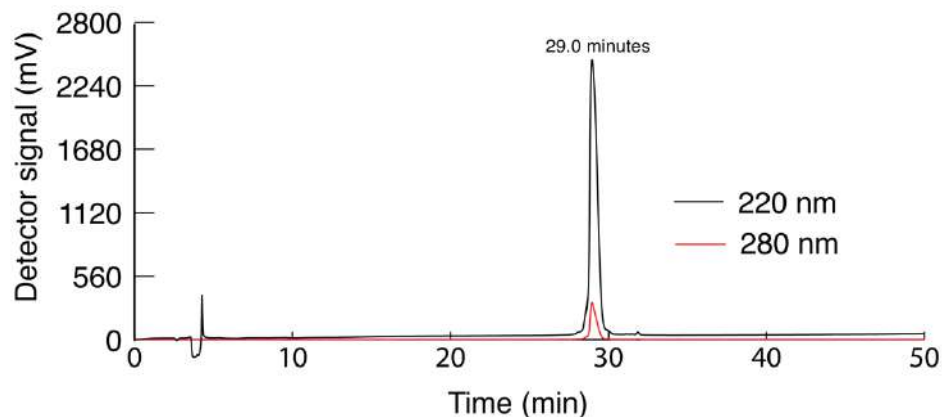


Figure 5-46. Analytical HPLC Data for **17K/18K**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

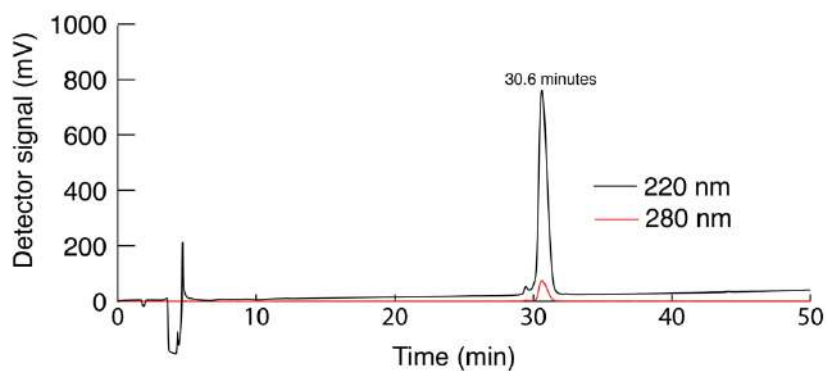


Figure 5-47. Analytical HPLC Data for **p17K/18K**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

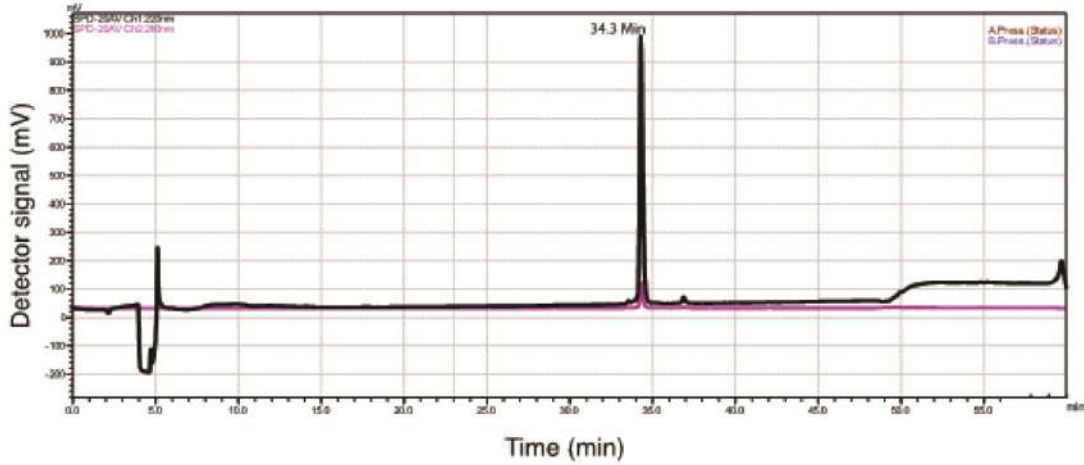


Figure 5-48. Analytical HPLC Data for **VHP1**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

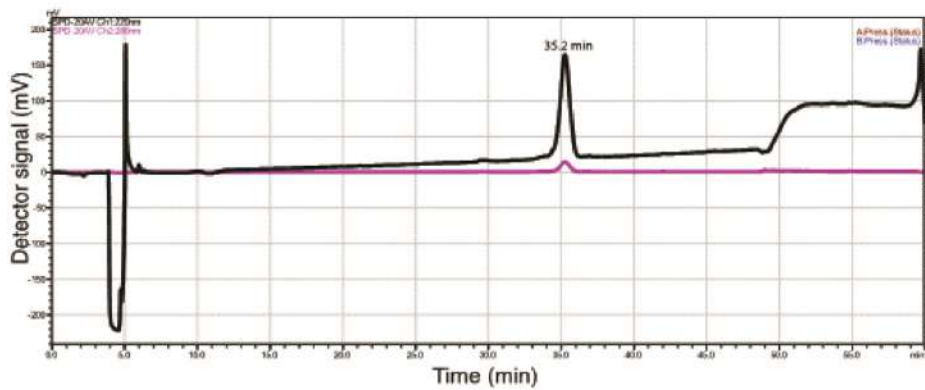


Figure 5-49. Analytical HPLC Data for **pVHP1**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

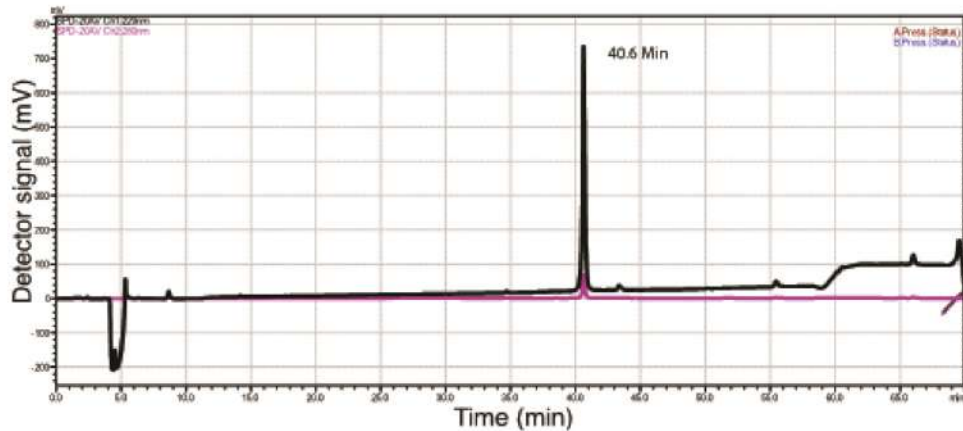


Figure 5-50. Analytical HPLC Data for **VHP2**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

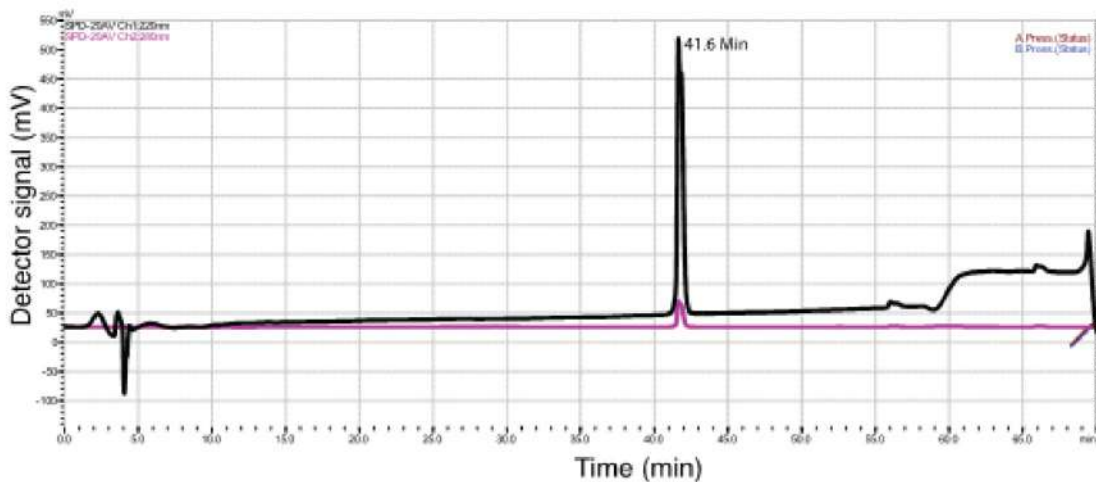


Figure 5-51. Analytical HPLC Data for **pVHP2**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

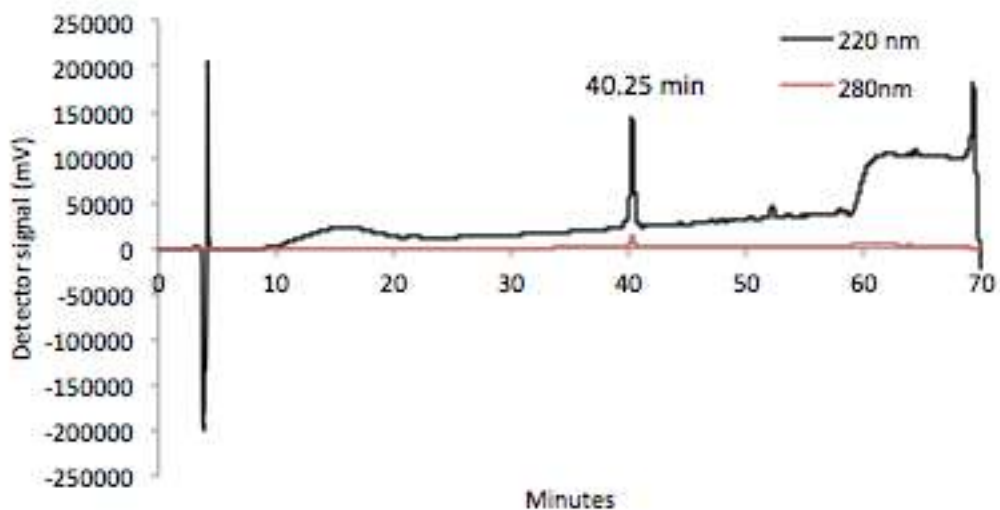


Figure 5-52. Analytical HPLC Data for p17/18K-1kDa. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

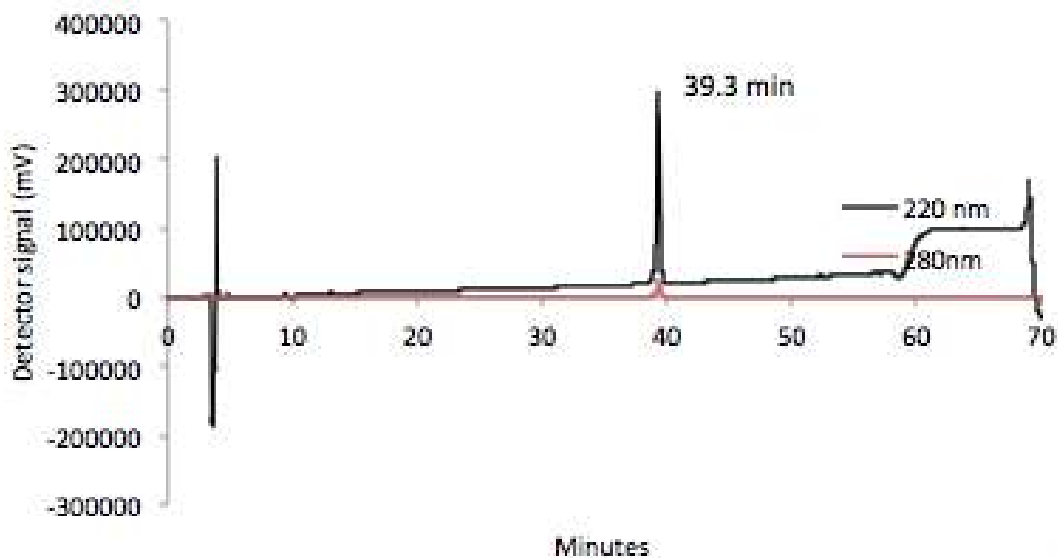


Figure 5-53. Analytical HPLC Data for m17/18K. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H₂O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

5.5.3 Variable Temperature CD experiments

Details and equations used for the Thermodynamic measurements of **1CW**, **p1CW**, **7A**, **p7A**, **15A**, **p15A**, **21A**, **p21A**, **17K**, **p17K**, **18K**, **p18K**, **17K/18K**, **p17K/18K**, **p17K/18K-1kDa** and **m17/18K** can be found in Chapter 2 section 2.4.5. Details and equations used for the Thermodynamic measurements for **VHP1**, **pVHP1**, **VHP2**, and **pVHP2** (50 μ M) can be found in chapter 4 section 4.4.3.

We used the parameters generated by fitting the variable temperature CD data for 1CW and VHP variants to calculate the folding free energy values presented in the main text. CD spectra and variable temperature CD data for **1CW**, **p1CW**, **7A**, **p7A**, **15A**, **p15A**, **21A**, **p21A**, **17K**, **p17K**, **18K**, **p18K**, **17K/18K**, **p17K/18K**, **p17/18K-1kDa**, **m17/18K**, **VHP1**, **pVHP1**, **VHP2**, **pVHP2** are shown in Figures 5-54–5-64, along with the parameters that were used to generate global fits for each compound. The standard error for each fitted parameter is also shown.

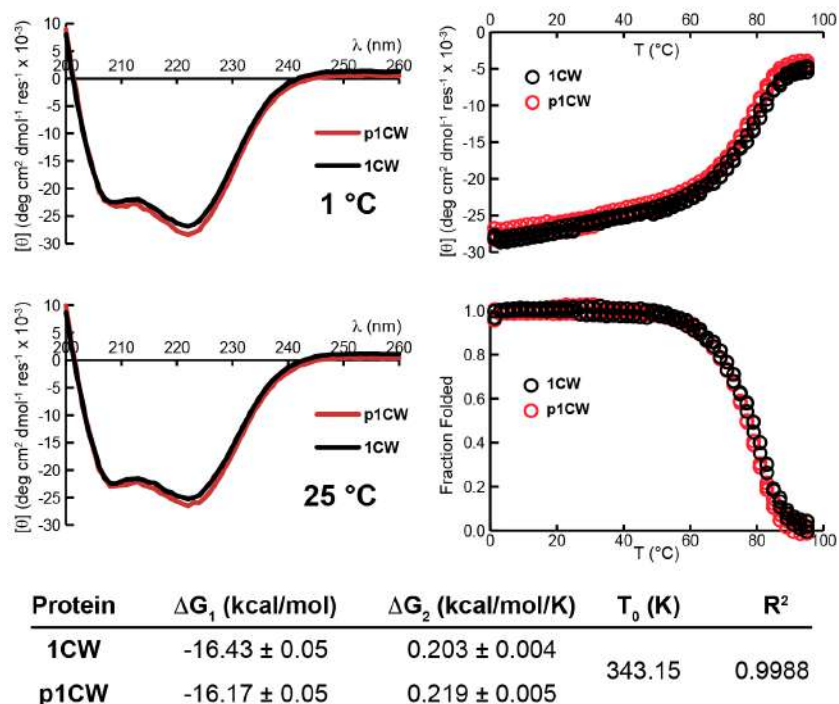


Figure 5-54. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μ M solutions of proteins **1CW** (black) and **p1CW** (red) in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

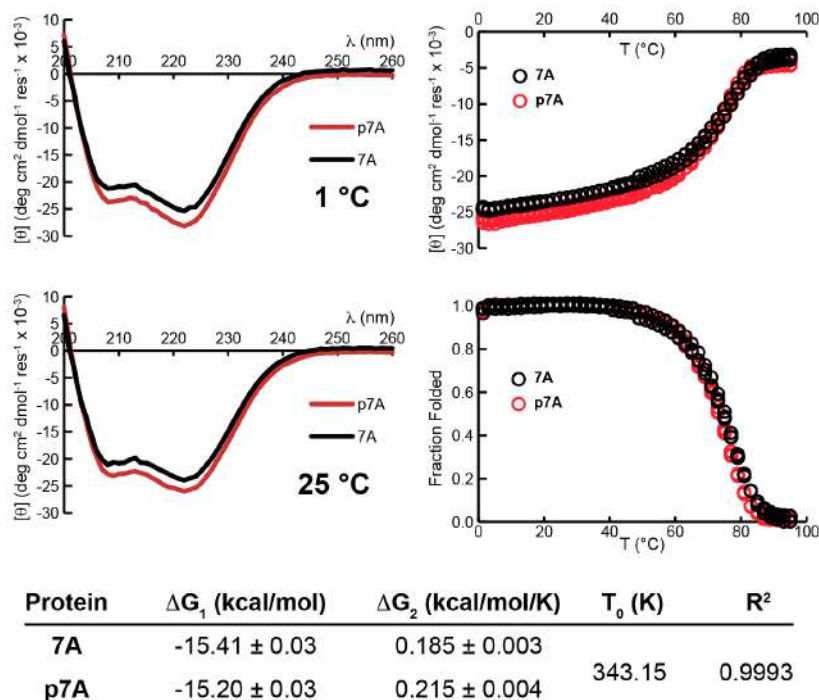
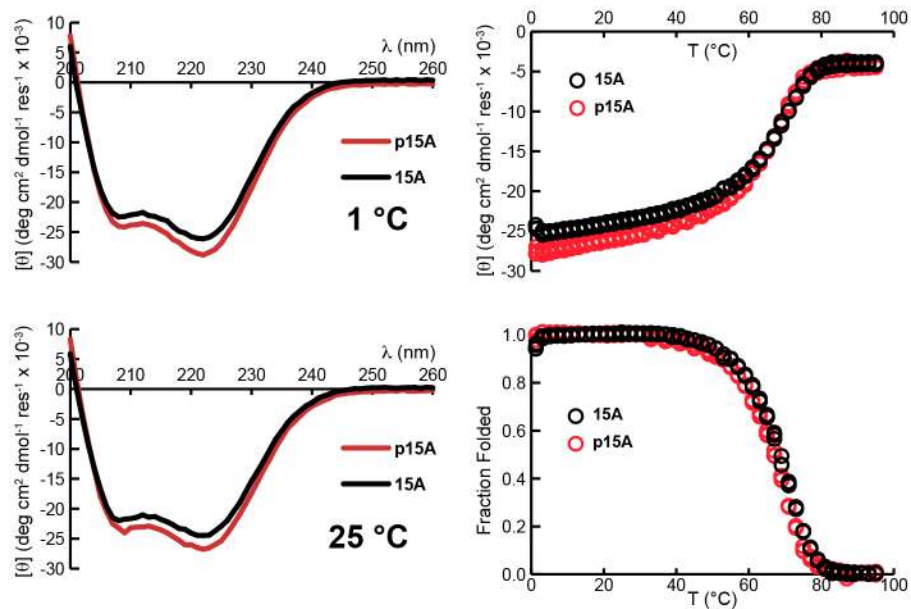


Figure 5-55. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μ M solutions of proteins **7A** (black) and **p7A** (red) in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



Protein	ΔG_1 (kcal/mol)	ΔG_2 (kcal/mol/K)	T_0 (K)	R^2
15A	-14.01 ± 0.02	0.223 ± 0.003	343.15	0.9995
p15A	-13.57 ± 0.02	0.213 ± 0.003		

Figure 5-56. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μ M solutions of proteins **15A** (black) and **p15A** (red) in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

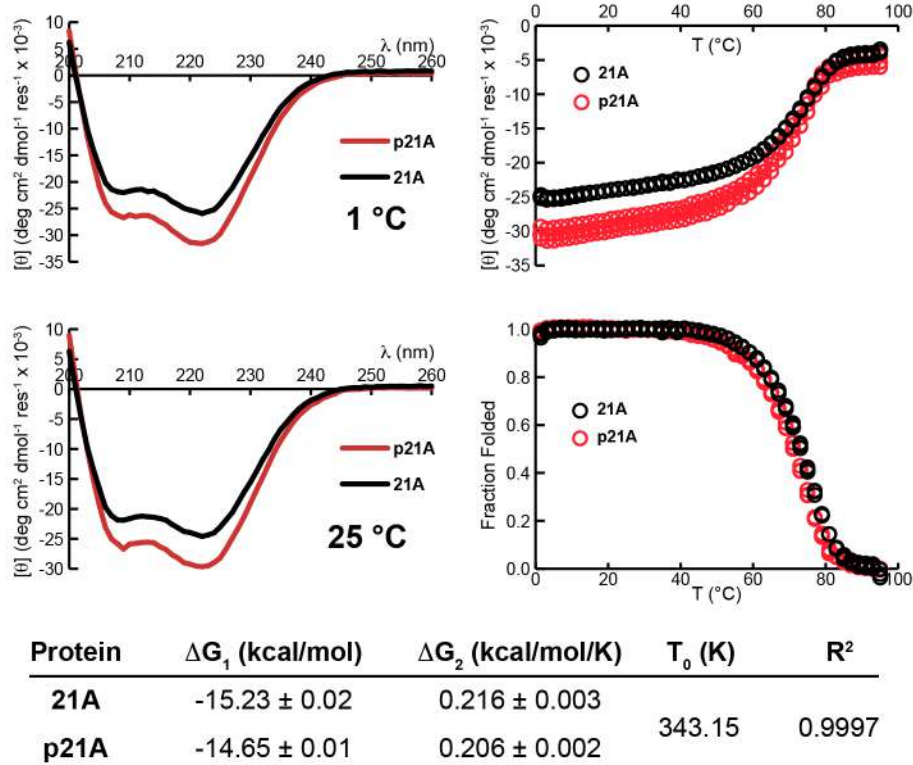


Figure 5-57. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μ M solutions of proteins **21A** (black) and **p21A** (red) in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

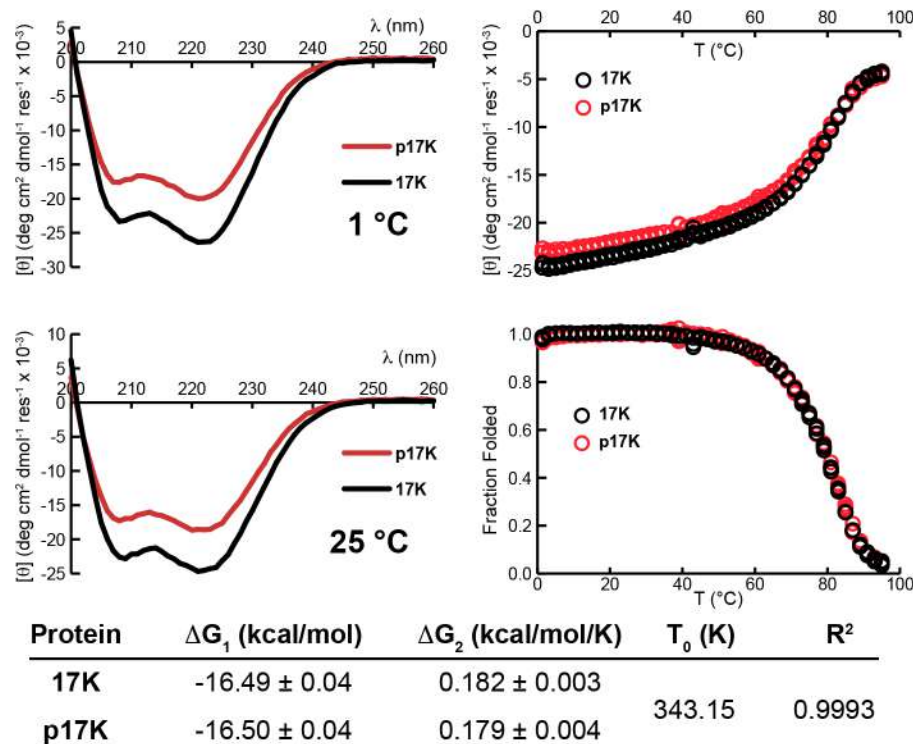


Figure 5-58. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **17K** (black) and **p17K** (red) in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

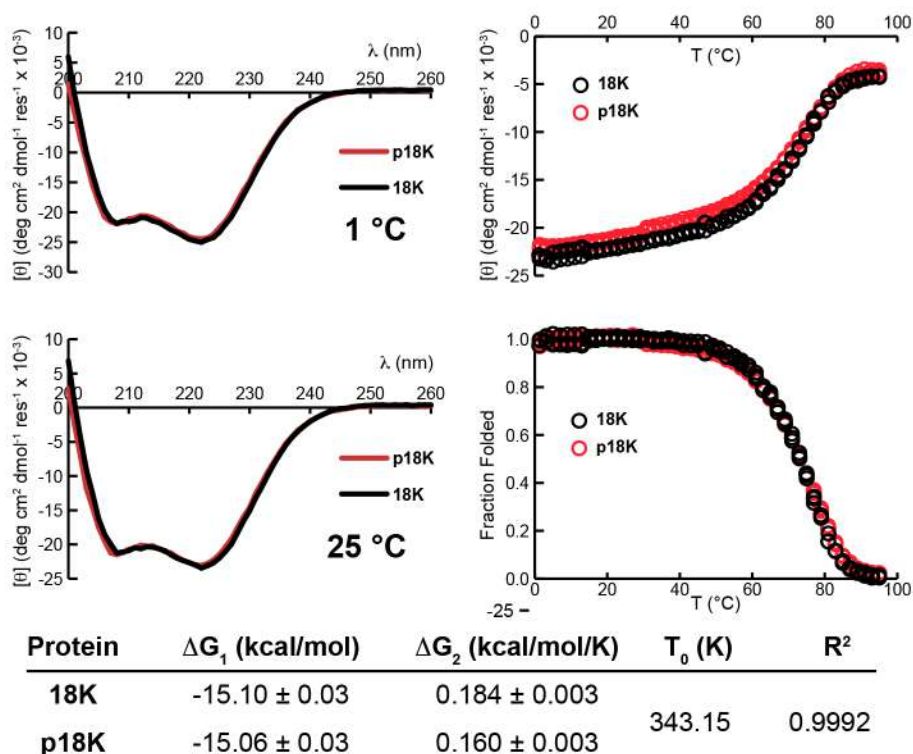


Figure 5-59. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μM solutions of proteins **18K** (black) and **p18K** (red) in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

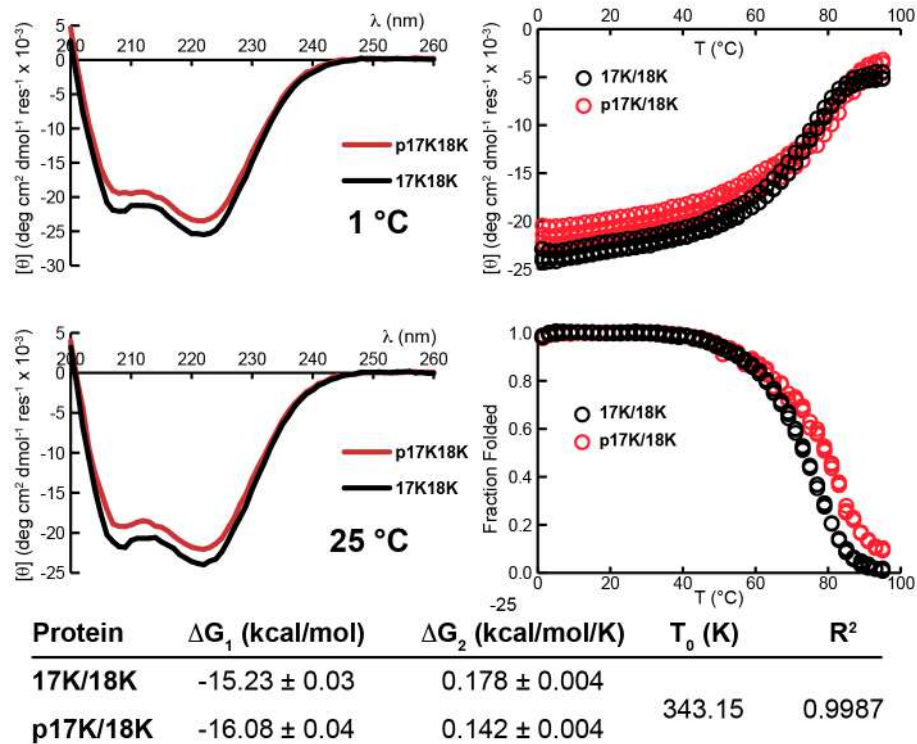


Figure 5-60. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μ M solutions of proteins **17K/18K** (black) and **p17K/18K** (red) in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.

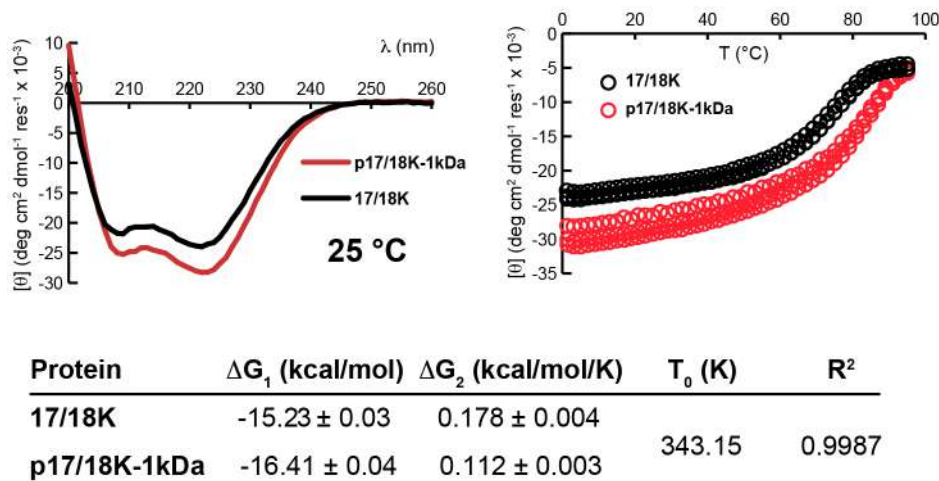
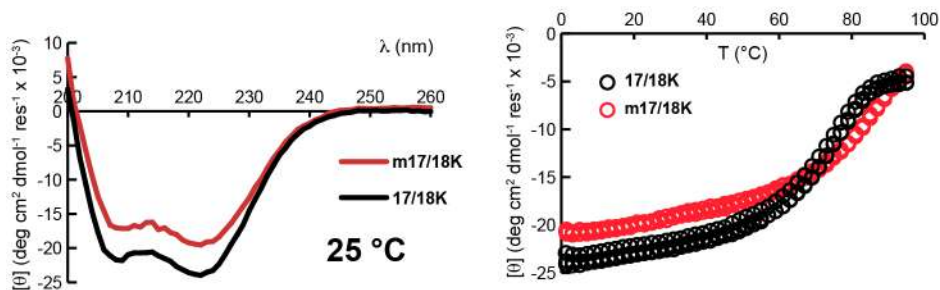
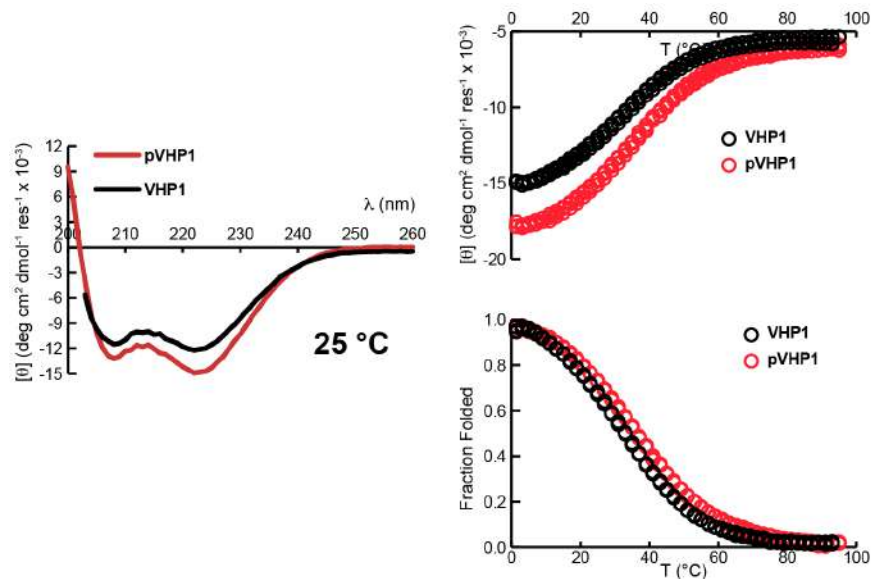


Figure 5-61. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μ M solutions of proteins **17K/18K** (black) and **p17K/18K-1kDa** (red) in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



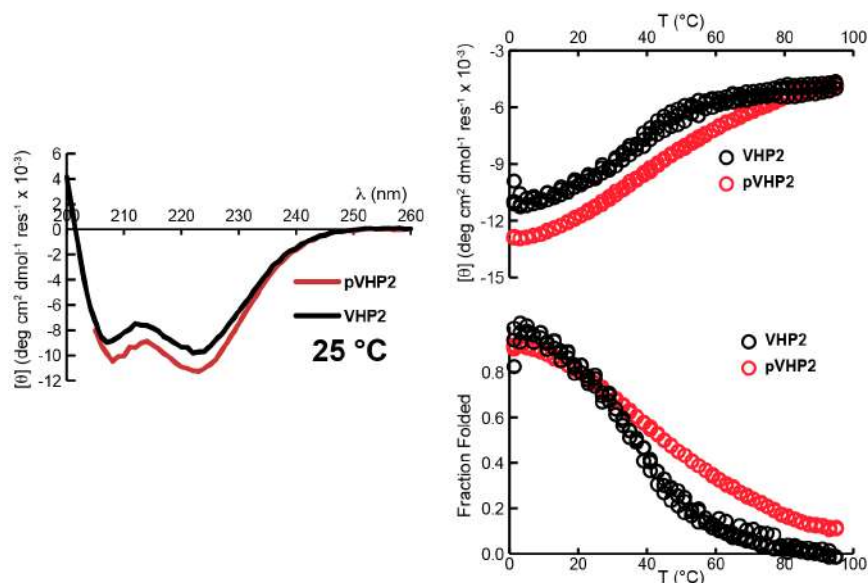
Protein	ΔG_1 (kcal/mol)	ΔG_2 (kcal/mol/K)	T_0 (K)	R^2
17/18K	-15.23 ± 0.03	0.178 ± 0.004	343.15	0.9987
m17/18K	-17.49 ± 0.09	0.135 ± 0.006		

Figure 5-62. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 30 μ M solutions of proteins **17K/18K** (black) and **m17K/18K** (red) in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S6–S9 are also shown, with standard errors as indicated.



Protein	ΔG_1 (kcal/mol)	T_m (K)	R^2
VHP1	0.0566 ± 0.0004	305.9 ± 0.1	0.9997
pVHP1	0.0529 ± 0.0003	309.3 ± 0.1	0.9998

Figure 5-63. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 50 μM solutions of proteins **VHP1** (black) and **pVHP1** (red) in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S11–S13 are also shown, with standard errors as indicated.



Protein	ΔG_1 (kcal/mol)	T_m (K)	R^2
VHP2	0.056 ± 0.002	309.5 ± 0.3	0.9947
pVHP2	0.0317 ± 0.0004	318.6 ± 0.2	0.9996

Figure 5-64. CD spectra (lines, top left), variable temperature CD data (circles, top right) for 50 μM solutions of proteins **VHP2** (black) and **pVHP2** (red) in 20 mM sodium phosphate, pH 7. Parameters used to fit the variable temperature CD data to equations S11–S13 are also shown, with standard errors as indicated.

5.5.4 Sedimentation Equilibrium experiments

Sedimentation equilibrium experiments were performed using a Beckman XLA ultracentrifuge. Peptide solutions were loaded into 1.2 cm cells, and sedimentation was monitored by absorbance at 280 nm. Sedimentation equilibria were analyzed at several speeds ranging from 12 to 45 krpm, and data were collected at 0.001 cm intervals along the length of the cell at each speed. After changing speeds, samples were allowed to equilibrate for 24 hours at the new speed prior to data collection, which was also repeated 2 hours later to verify that the

consecutive data sets were superimposable, signaling that the system was at equilibrium.

Apparent molecular weight was determined by least-squares regression of the equilibrium radial absorbance data using MS Excel and Solver. Data were fit to models either for a single species, or for an equilibrium between monomer and n -mer (equations S14 and S15, respectively):

$$A(r) = \text{baseline} + A_o \exp \left[\frac{M(1-\nu\rho)\omega^2(r^2-r_o^2)}{2RT} \right] \quad (\text{S14})$$

$$A(r) = A_o \exp \left[\frac{M(1-\nu\rho)\omega^2(r^2-r_o^2)}{2RT} \right] + KA_o^n \exp \left[\frac{nM(1-\nu\rho)\omega^2(r^2-r_o^2)}{2RT} \right] \quad (\text{S15})$$

In equation S14, $A(r)$ is the absorbance (at 280 nm) at radial position r (in cm); A_o is the absorbance (at 280 nm) at an arbitrary reference radial position r_o ; M is the apparent molecular weight of the peptide; ν is the partial specific volume of the peptide (in $\text{cm}^3 \text{g}^{-1}$); ρ is the density of the sample (calculated using SEDNTERP; $\rho = 1.00046 \text{ g cm}^{-3}$ for a 20 mM sodium phosphate solution at pH 7); ω is the radial velocity (in s^{-1}) of the rotor; R is the universal gas constant ($8.3144 \times 10^7 \text{ erg mol}^{-1} \text{ K}^{-1}$); T is the temperature (in Kelvin), and “baseline” is a correction for absorbance resulting from non-sedimenting components of the sample.

The variables in equation S15 have the same meaning, except A_o is the absorbance (at 280 nm) due to monomer at the arbitrary reference radial position; K is the equilibrium constant for the association between the monomer and the n -mer; n is the aggregation number for the species in equilibrium with the monomer; M is the molecular weight of the monomer, assuming appropriately ionized side-chains at pH 7.

The partial specific volumes (ν) for each non-PEGylated 1CW variant were calculated following the method of Durschlag and Zipper³¹ and are shown in Table 5-1. For PEGylated 1CW variants, we estimated ν according to the following equation (S16):

$$v = \left[\frac{v_1 M_1 + v_2 M_2}{M_1 + M_2} \right] \quad (\text{S16})$$

where, v_1 is the partial specific volume of a 1CW variant modified at Cys14 with a maleimide-PEG containing only one PEG unit; M_1 is the calculated molecular weight of this peptide; v_2 is the partial specific volume of a 44-unit PEG oligomer (~44 PEG units; $v_2 = 0.836 \text{ cm}^3 \text{ g}^{-1}$, measured previously for a 2000 Da PEG³²); and M_2 is the molecular weight of the 44-unit PEG oligomer (1938 Da).

Table 5-1. Molecular Weight and Partial Specific Volumes for 1CW variants and their PEGylated counterparts.

Peptide	Calculated M (g mol ⁻¹)	Calculated v (cm ³ g ⁻¹)	Observed M single species fit (g mol ⁻¹)	M _{Obs} /M _{calc}	Aggregation state
p1CW	5909.8	0.7687	15.7 ± 0.1 × 10 ³	2.7	trimer
7A	3687.1	0.7327	11.9 ± 0.1 × 10 ³	3.2	trimer
p7A	5851.7	0.7675	13.5 ± 0.1 × 10 ³	2.3	trimer
15A	3687.1	0.7327	10.5 ± 0.3 × 10 ³	2.8	trimer
p15A	5851.7	0.7675	14.0 ± 0.1 × 10 ³	2.4	trimer
21A	3687.1	0.7327	9.2 ± 0.1 × 10 ³	2.5	trimer
p21A	5851.7	0.7675	14.2 ± 0.1 × 10 ³	2.4	trimer
17K	3746.3	0.7389	10.2 ± 0.1 × 10 ³	2.7	trimer
p17K	5910.8	0.7711	16.1 ± 0.2 × 10 ³	2.7	trimer
18K	3803.3	0.7375	10.6 ± 0.1 × 10 ³	2.8	trimer
p18K	5967.9	0.7699	14.7 ± 0.1 × 10 ³	2.5	trimer
17K/18K	3804.4	0.7412	11.3 ± 0.1 × 10 ³	3.0	trimer
p17K/18K	5968.9	0.7723	15.3 ± 0.1 × 10 ³	2.6	trimer

As a rule, equilibrium sedimentation data were fit to the single species model first; the equilibrium models were attempted if the single species fit was unsatisfactory. The quality of a particular fit was judged based on the R^2 value (closer to 1 is better), the fit standard deviation (lower is better), and the appearance of the residual plots, which show the difference between the data and the fit at each radial position (smaller and more random deviations from residual = 0 are better). If the single-species and equilibrium models were of similar quality according to these criteria, we used the simpler model to describe the peptides shown in Table 5-1. The best model selected for each peptide according to these criteria is highlighted with a magenta box in Figures 5-65–5-76. In each case, alternative models are also presented to show how we arrived at the conclusions presented in Table 5-1.

In most cases, the observed apparent molecular weights shown in Table 5-1 are smaller than the molecular weights expected for a trimer composed of three copies of a **1CW** variant or its PEGylated counterpart. These differences may be a result of charged-based non-ideality, in which charge-charge repulsion decreases the apparent molecular weight of peptide assemblies by shifting the equilibrium distribution of the peptide away from the bottom of the cell. The difference may also reflect errors in the calculated values of density (ρ) or partial specific volume (v) that were used in each fit. In most cases, a small 3-6% change in ρ or v could account for the deviation of the experimental apparent molecular weight from the calculated molecular weight of the trimer. For PEGylated compounds, the difference may also reflect the polydisperse nature of the PEG oligomer; some protein-PEG conjugates will have PEG oligomers that are longer or shorter than the average nominal 45-unit length. We note that previously published sedimentation equilibrium experiments from Xu and coworkers¹⁹ gave observed molecular weights for **1CW** and **p1CW** of 10177 Da and 17147 Da. These values are respectively 9% and

3% smaller than expected; Xu and coworkers nonetheless conclude that **1CW** and **p1CW** are helix-bundle trimers.

The smaller size of the observed molecular weights relative to the calculated trimer molecular weights in Table 5-1 could also be consistent with **1CW** variants that are equilibrating between two or more distinct species under the conditions of the experiment. To test this hypothesis, we attempted to fit the sedimentation data for each variant to a monomer-dimer, monomer-trimer, and monomer-tetramer equilibrium model according to equation S15. In so doing, we used the monomer molecular weights listed in Table 5-1 as constants, along with $n = 2, 3, \text{ or } 4$ as desired. In each case, the monomer-dimer and monomer-tetramer fits were of substantially lower quality than the monomer-trimer fit, which itself was of slightly lower quality than the single species fit. Given the potential uncertainty that results from charge-based non-ideality; errors in ρ and v ; and a polydisperse PEG conjugate, we conclude that the single species fit and monomer-trimer equilibrium fit lead to the same conclusion: that each of the **1CW** variants shown in Table 5-1 forms a helix-bundle trimer in solution.

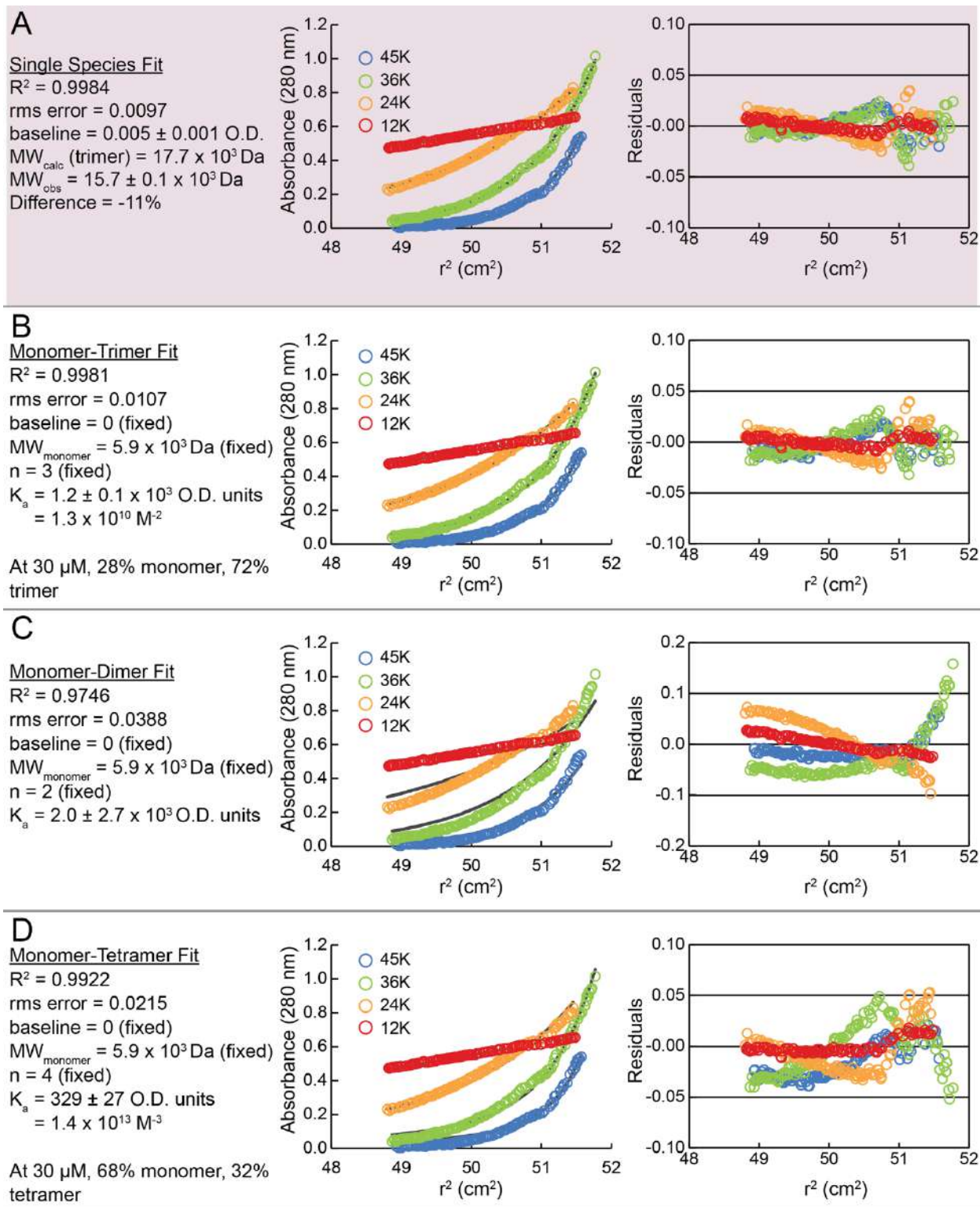


Figure 5-65. Sedimentation equilibrium data for p1CW in 20 mM sodium phosphate, pH 7. Data were fit to four different models; the best-fitting model is highlighted in magenta.

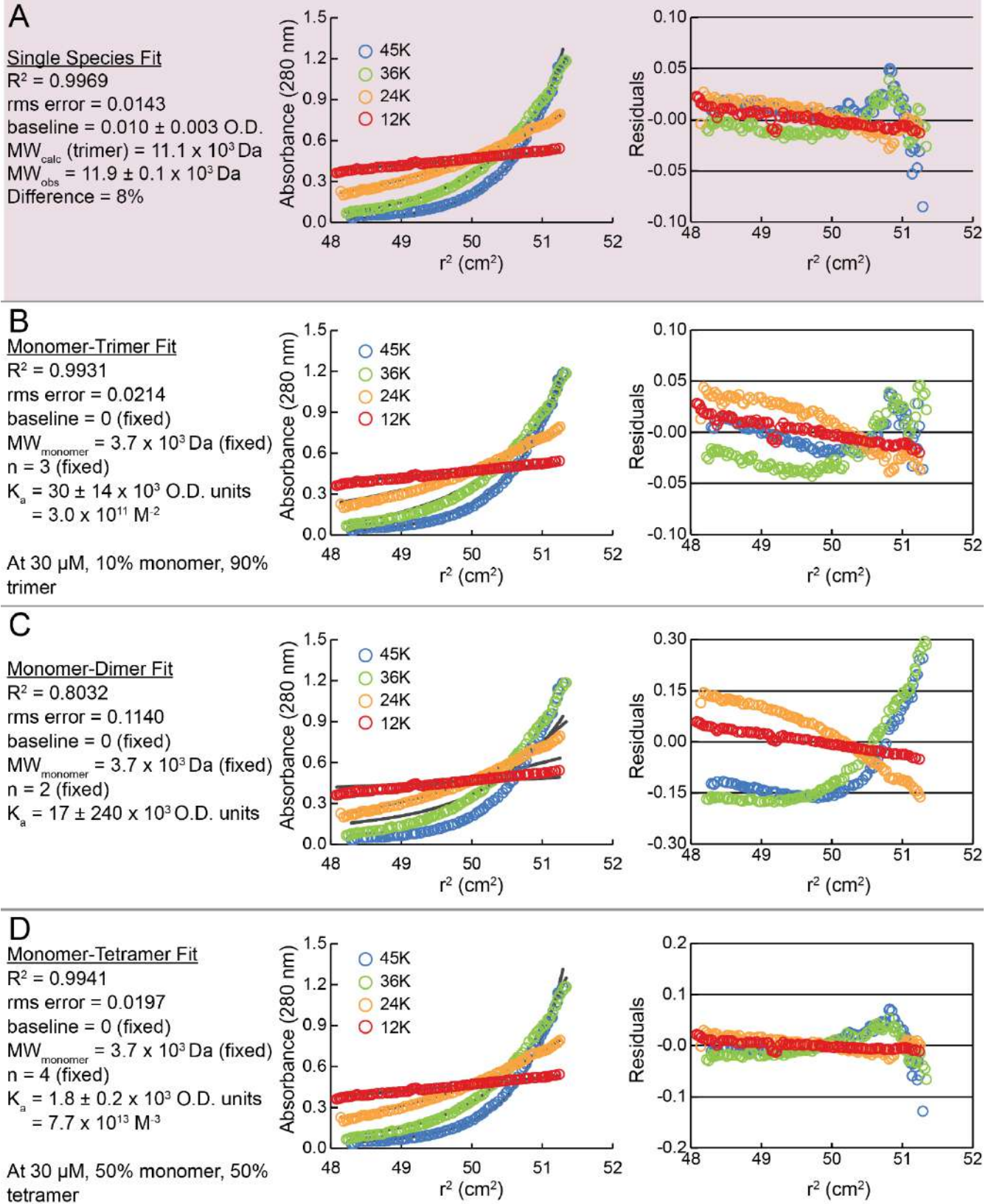


Figure 5-66. Sedimentation equilibrium data for 7A in 20 mM sodium phosphate, pH 7. Data were fit to four different models; the best-fitting model is highlighted in magenta.

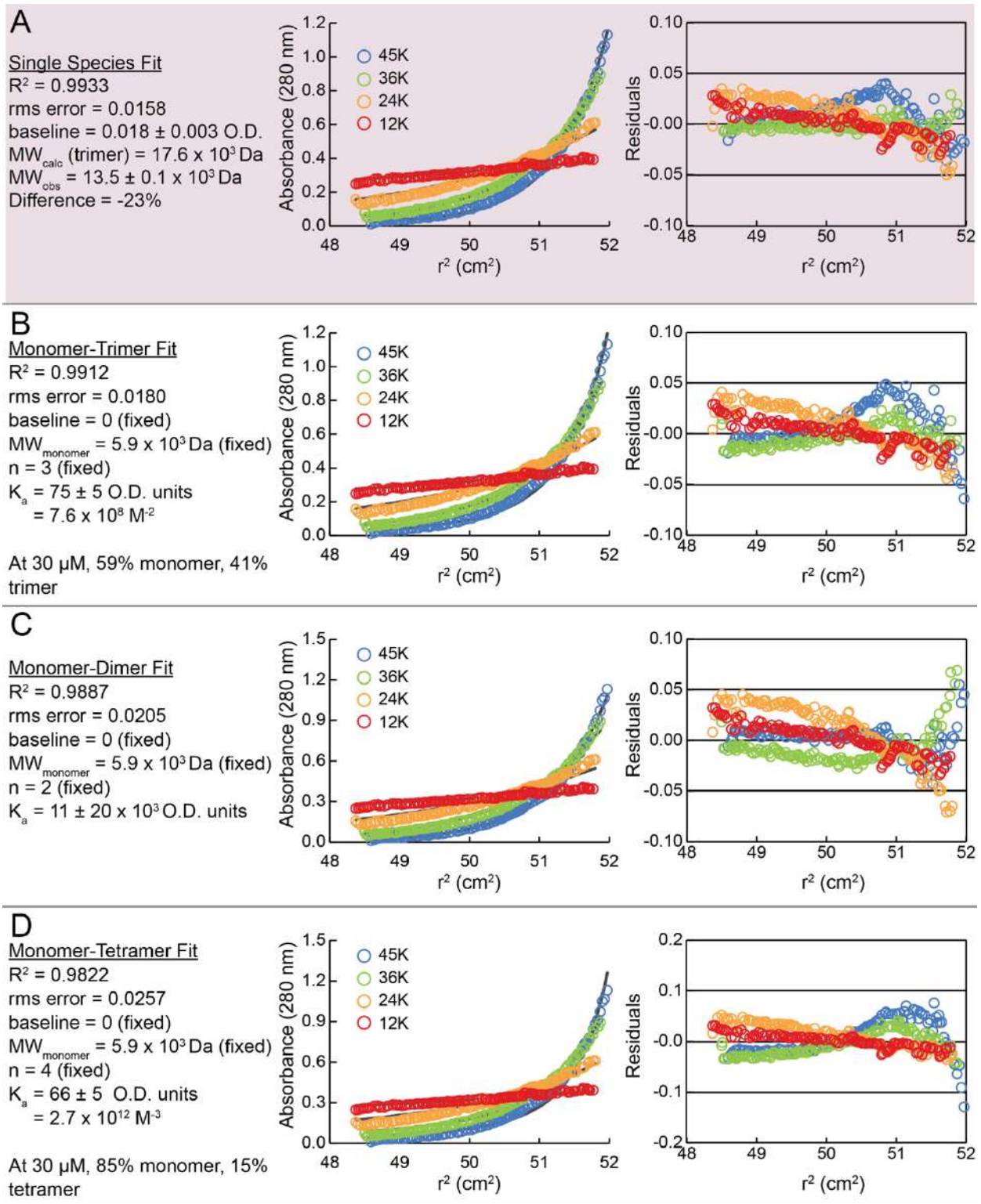


Figure 5-67. Sedimentation equilibrium data for p7A in 20 mM sodium phosphate, pH 7. Data were fit to four different models; the best-fitting model is highlighted in magenta.

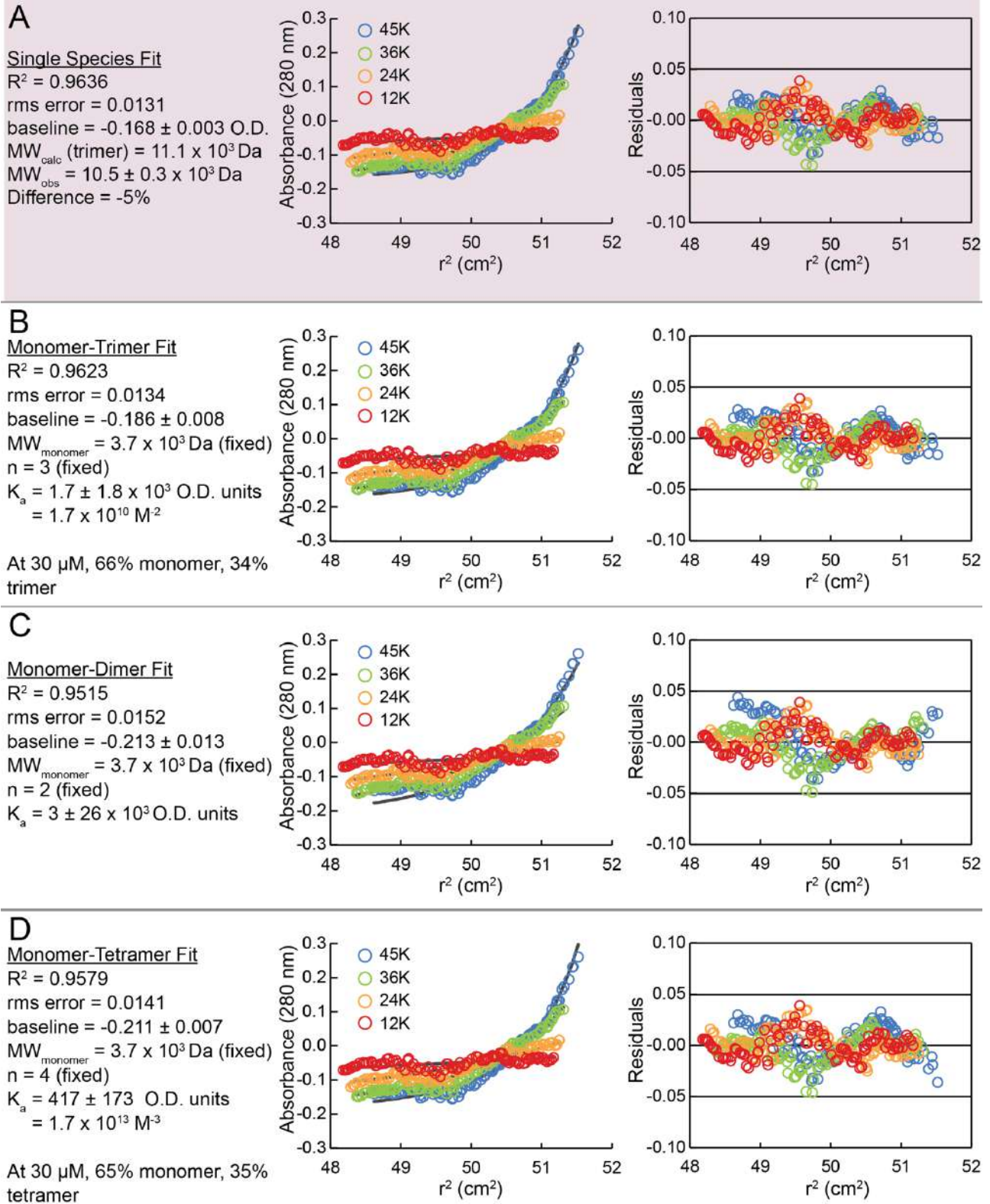


Figure 5-68. Sedimentation equilibrium data for 15A in 20 mM sodium phosphate, pH 7. Data were fit to four different models; the best-fitting model is highlighted in magenta.

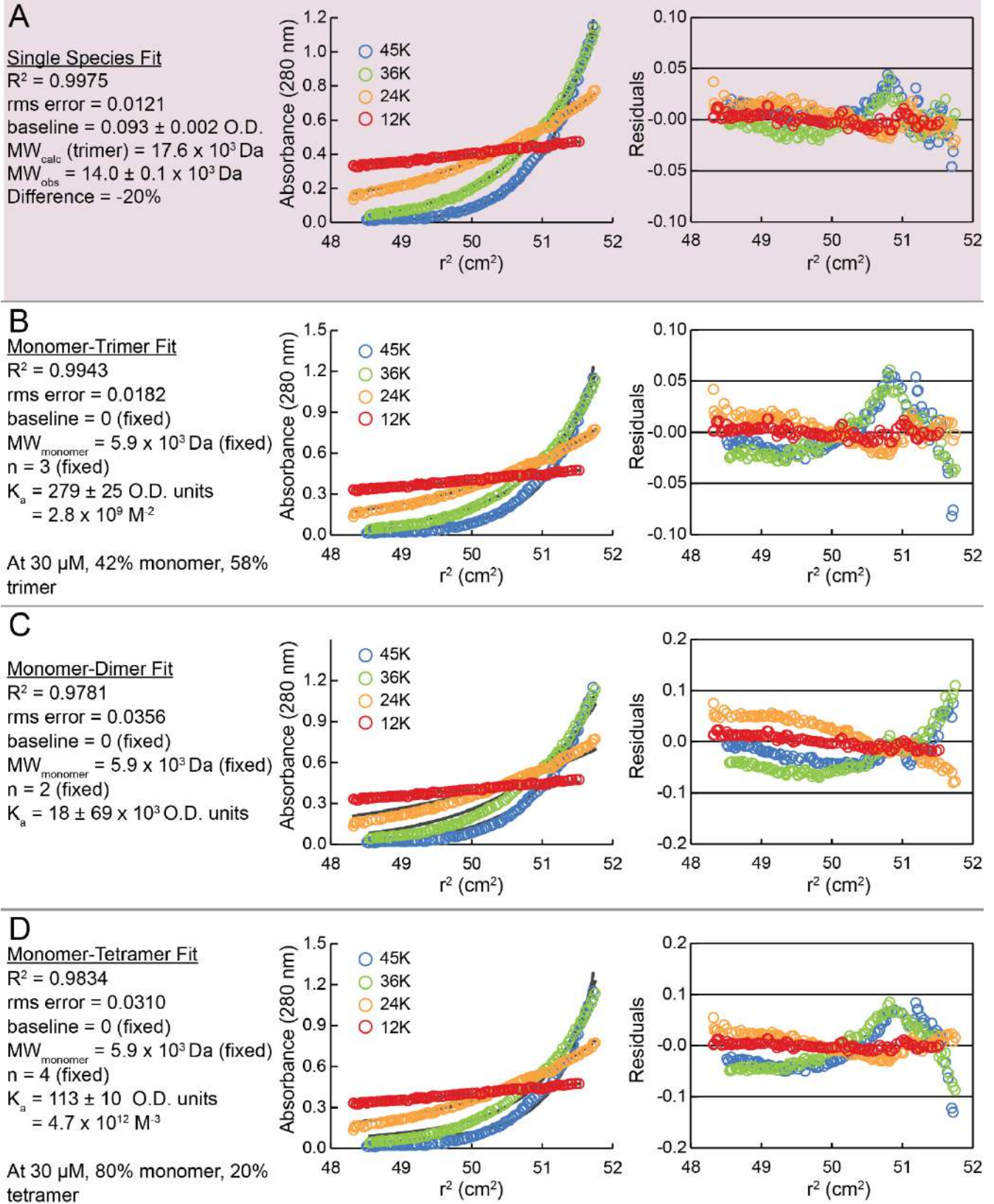


Figure 5-69. Sedimentation equilibrium data for **p15A** in 20 mM sodium phosphate, pH 7. Data were fit to four different models; the best-fitting model is highlighted in magenta.

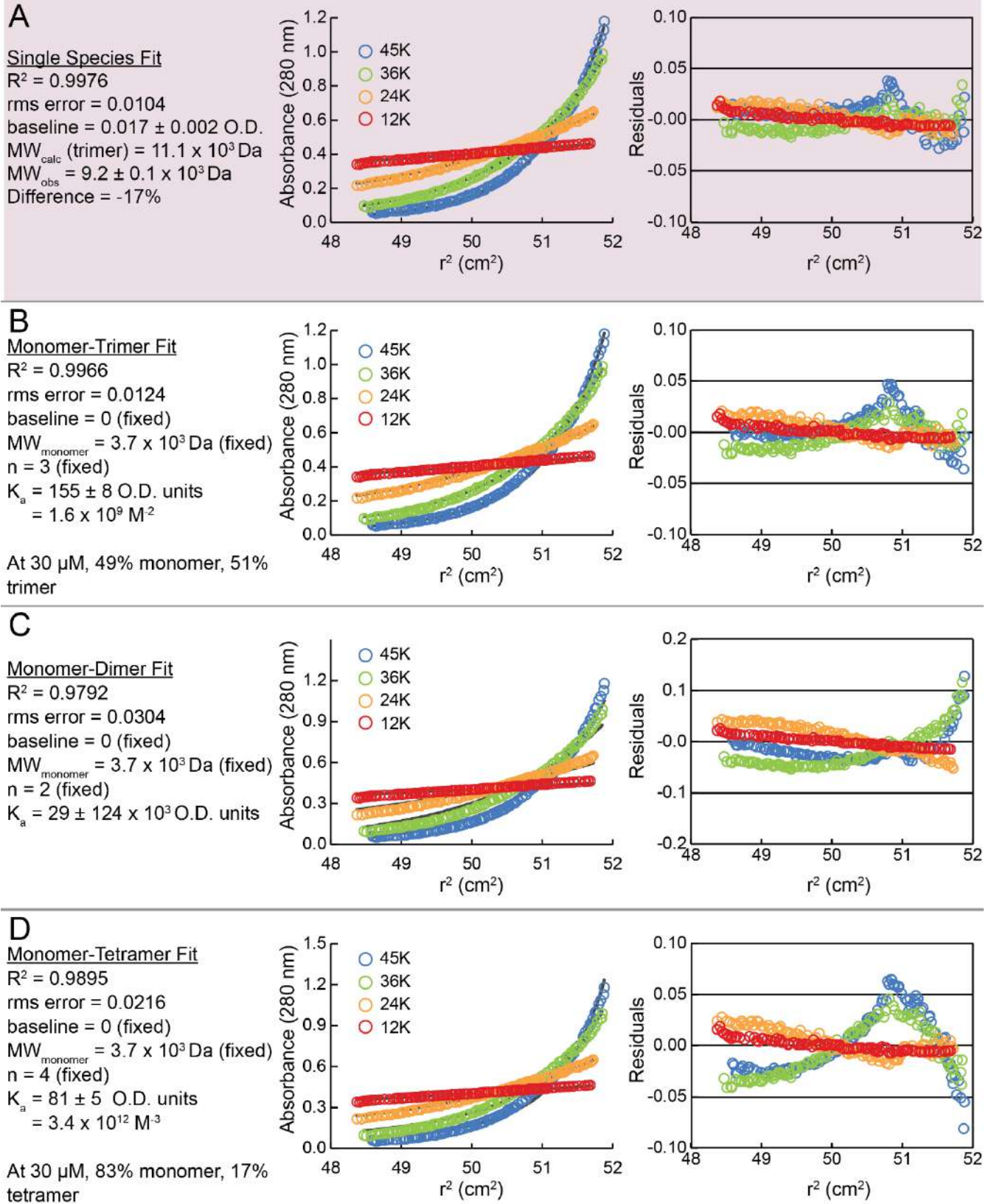


Figure 5-70. Sedimentation equilibrium data for **21A** in 20 mM sodium phosphate, pH 7. Data were fit to four different models; the best-fitting model is highlighted in magenta.

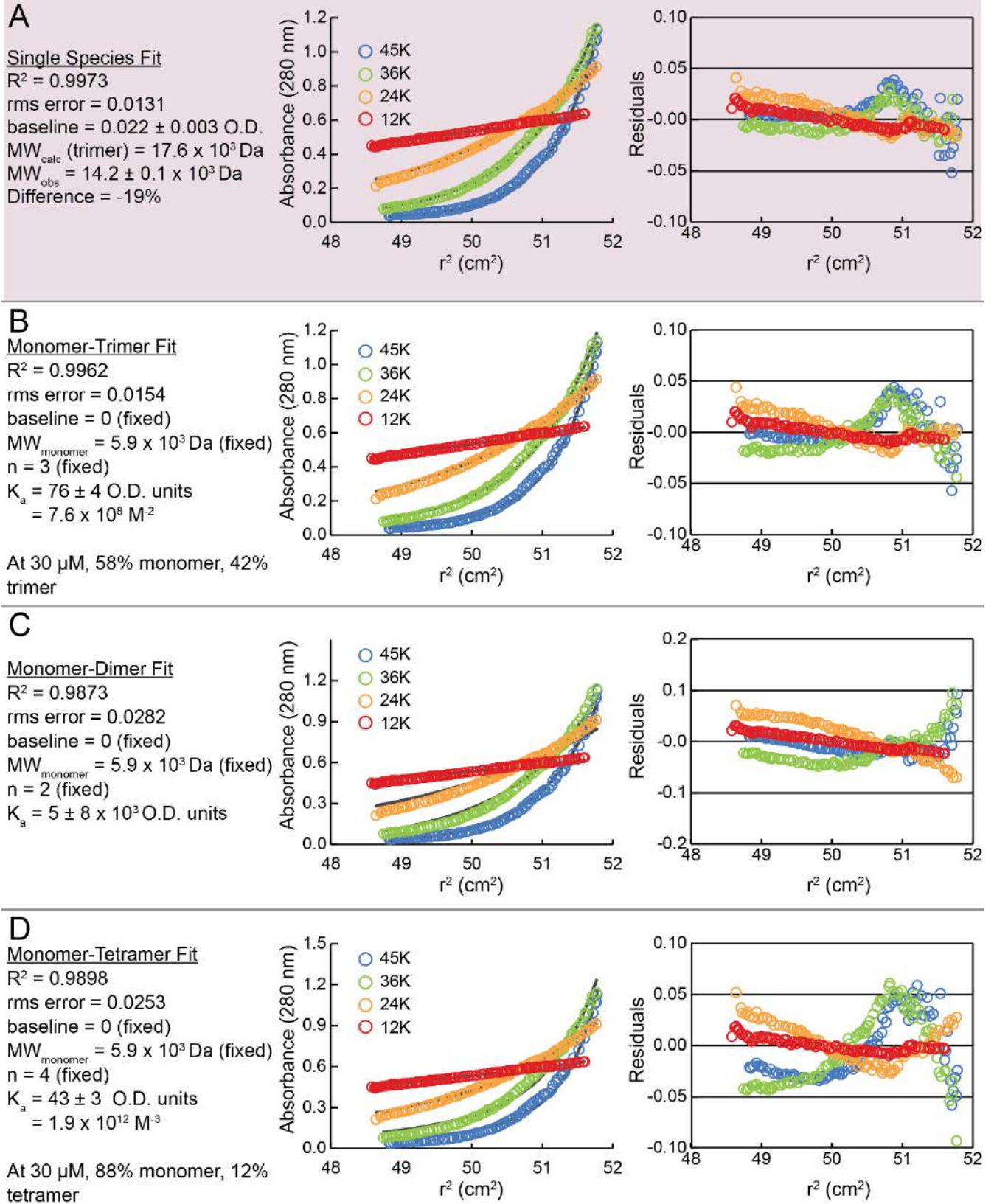


Figure 5-71. Sedimentation equilibrium data for **p21A** in 20 mM sodium phosphate, pH 7. Data were fit to four different models; the best-fitting model is highlighted in magenta.

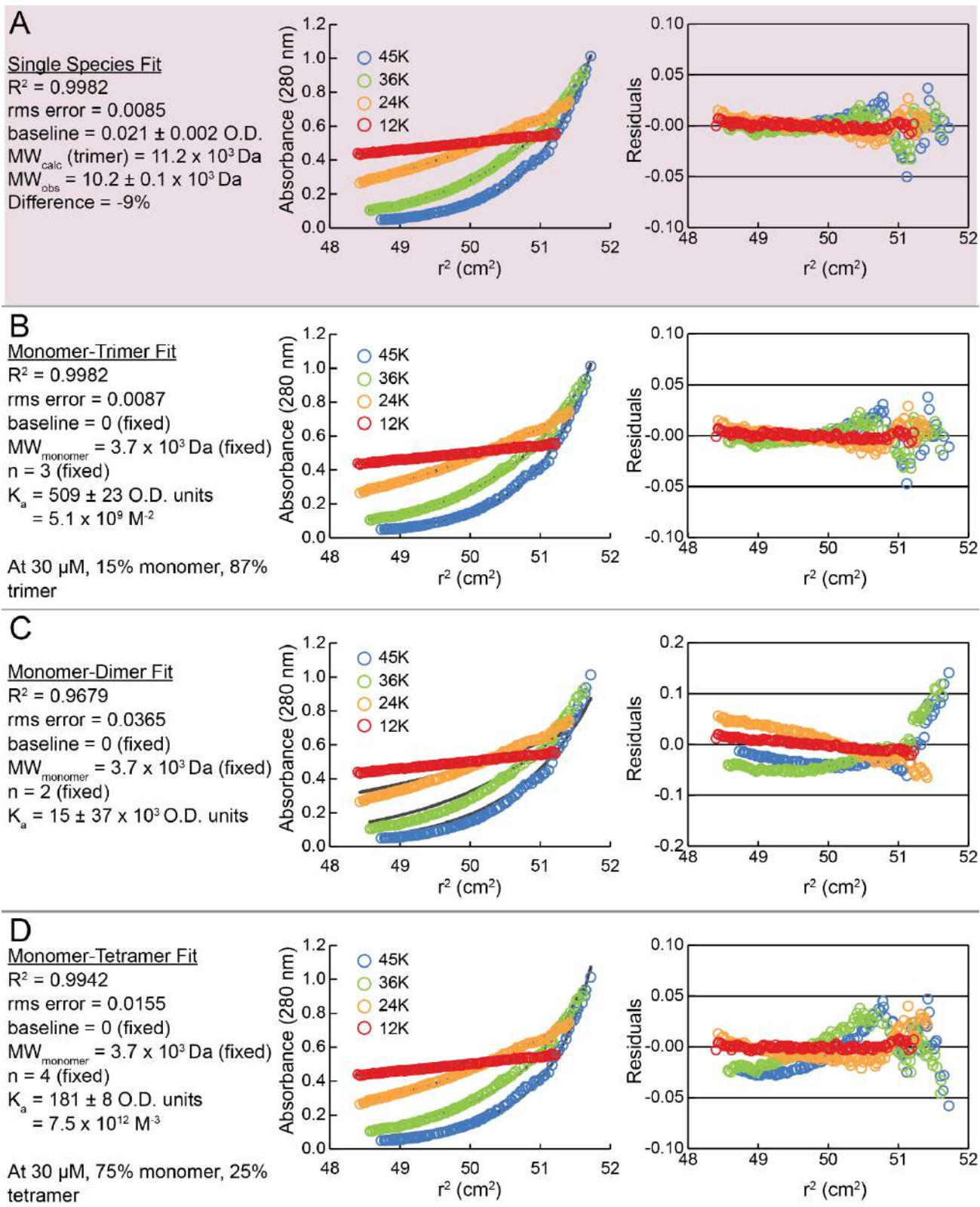


Figure 5-72. Sedimentation equilibrium data for **17K** in 20 mM sodium phosphate, pH 7. Data were fit to four different models; the best-fitting model is highlighted in magenta.

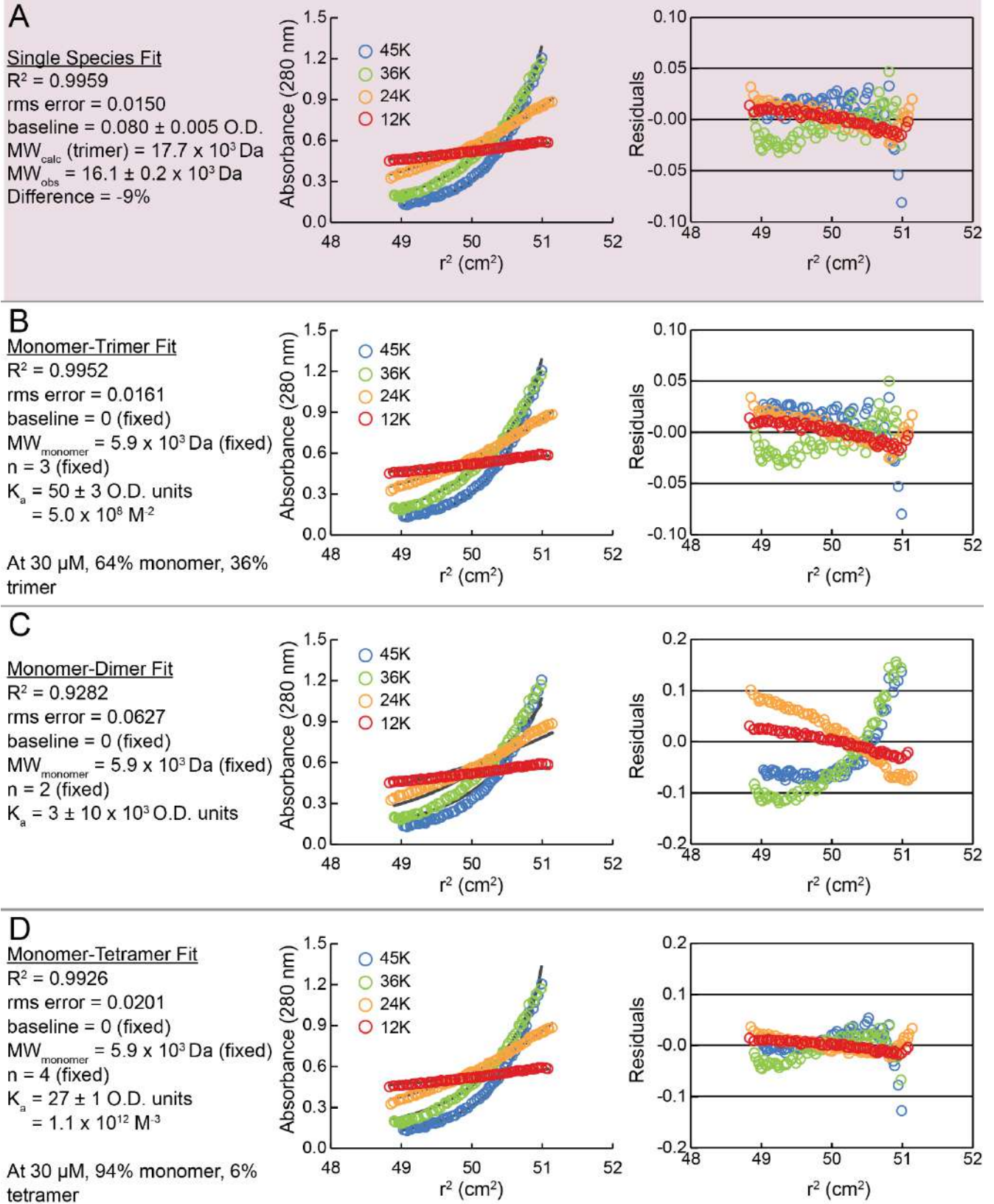


Figure 5-73. Sedimentation equilibrium data for **p17K** in 20 mM sodium phosphate, pH 7. Data were fit to four different models; the best-fitting model is highlighted in magenta.

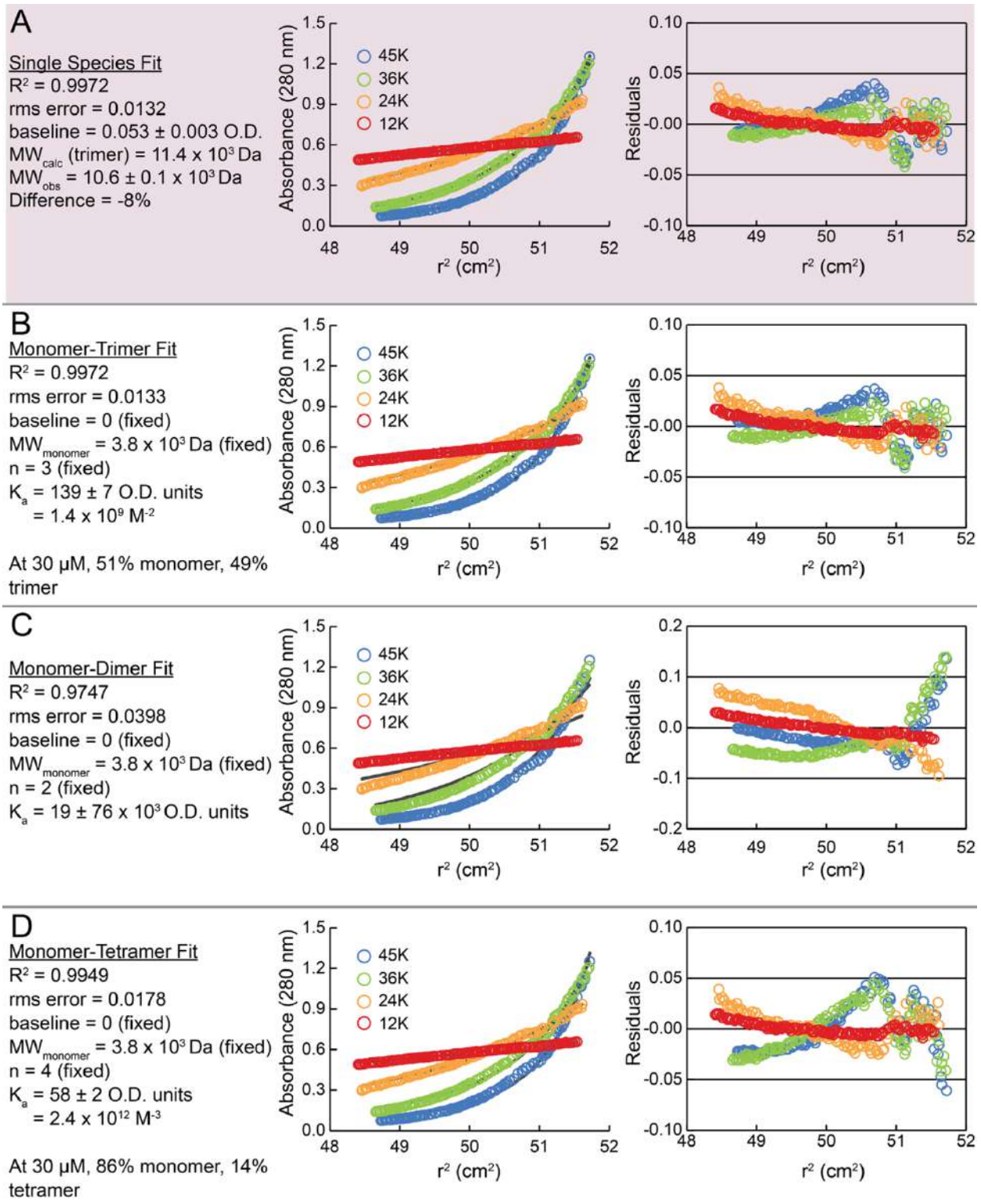


Figure 5-74. Sedimentation equilibrium data for 18K in 20 mM sodium phosphate, pH 7. Data were fit to four different models; the best-fitting model is highlighted in magenta.

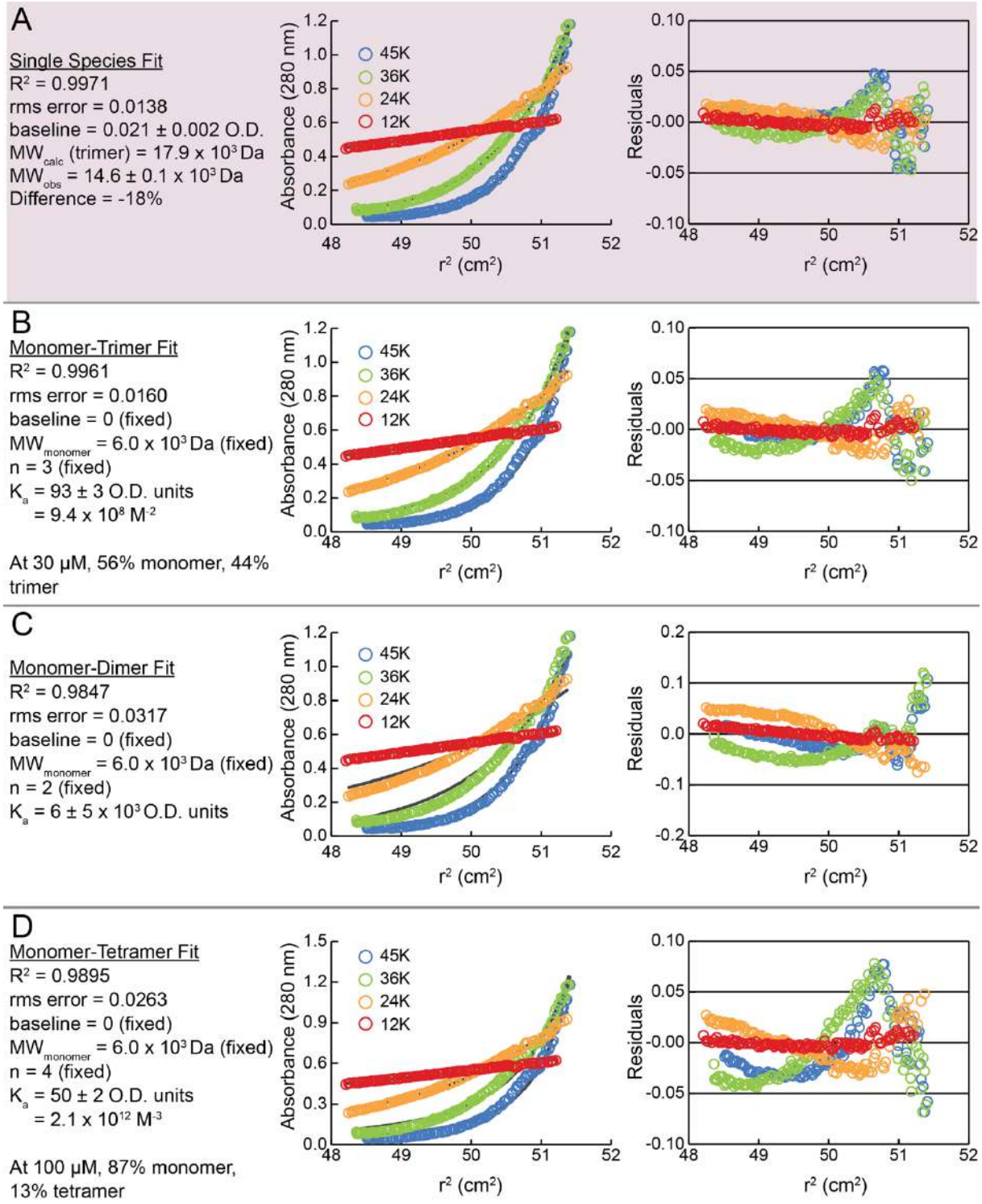


Figure 5-75. Sedimentation equilibrium data for **p18K** in 20 mM sodium phosphate, pH 7. Data were fit to four different models; the best-fitting model is highlighted in magenta.

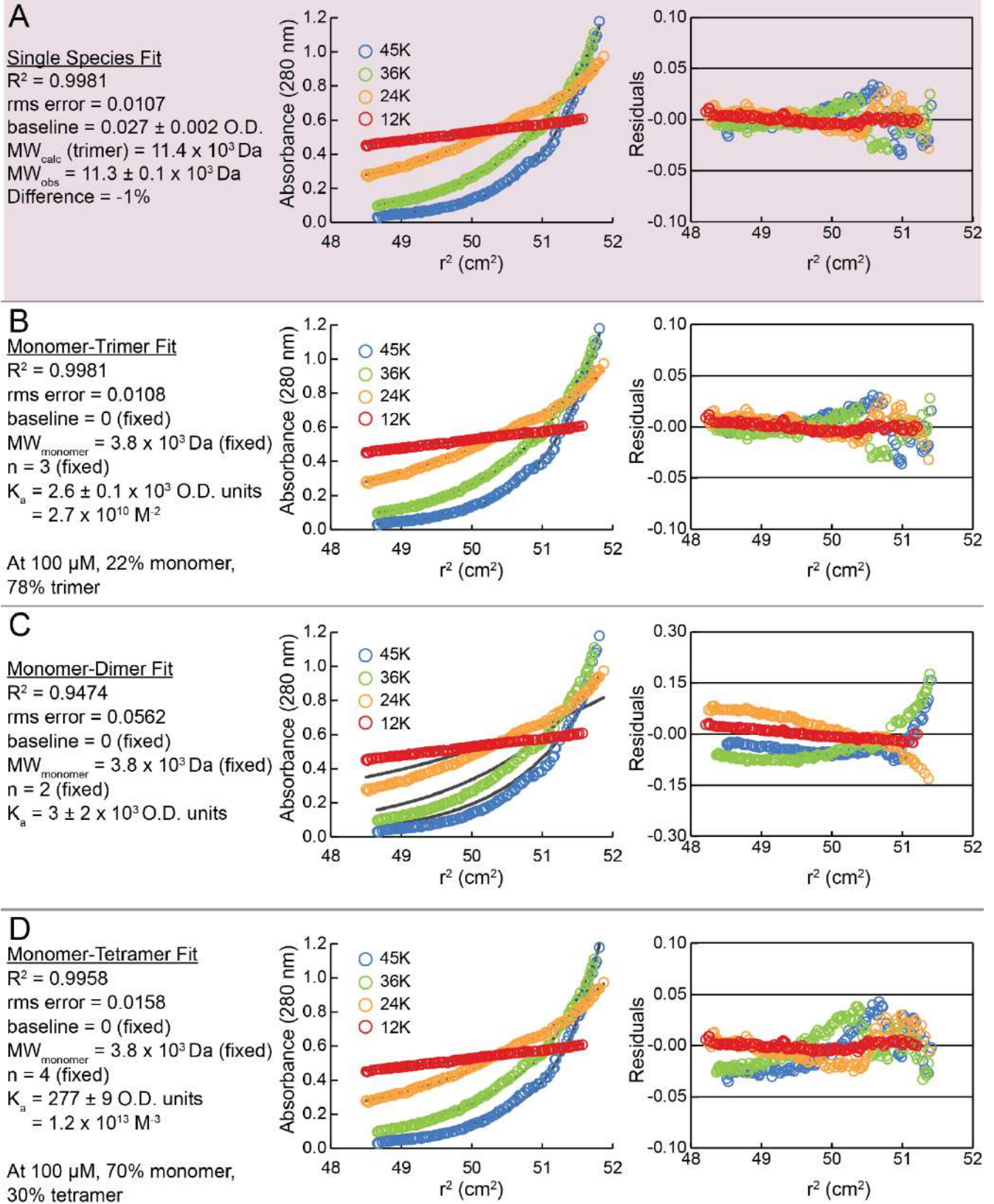


Figure 5-76. Sedimentation equilibrium data for 17K/18K in 20 mM sodium phosphate, pH 7. Data were fit to four different models; the best-fitting model is highlighted in magenta.

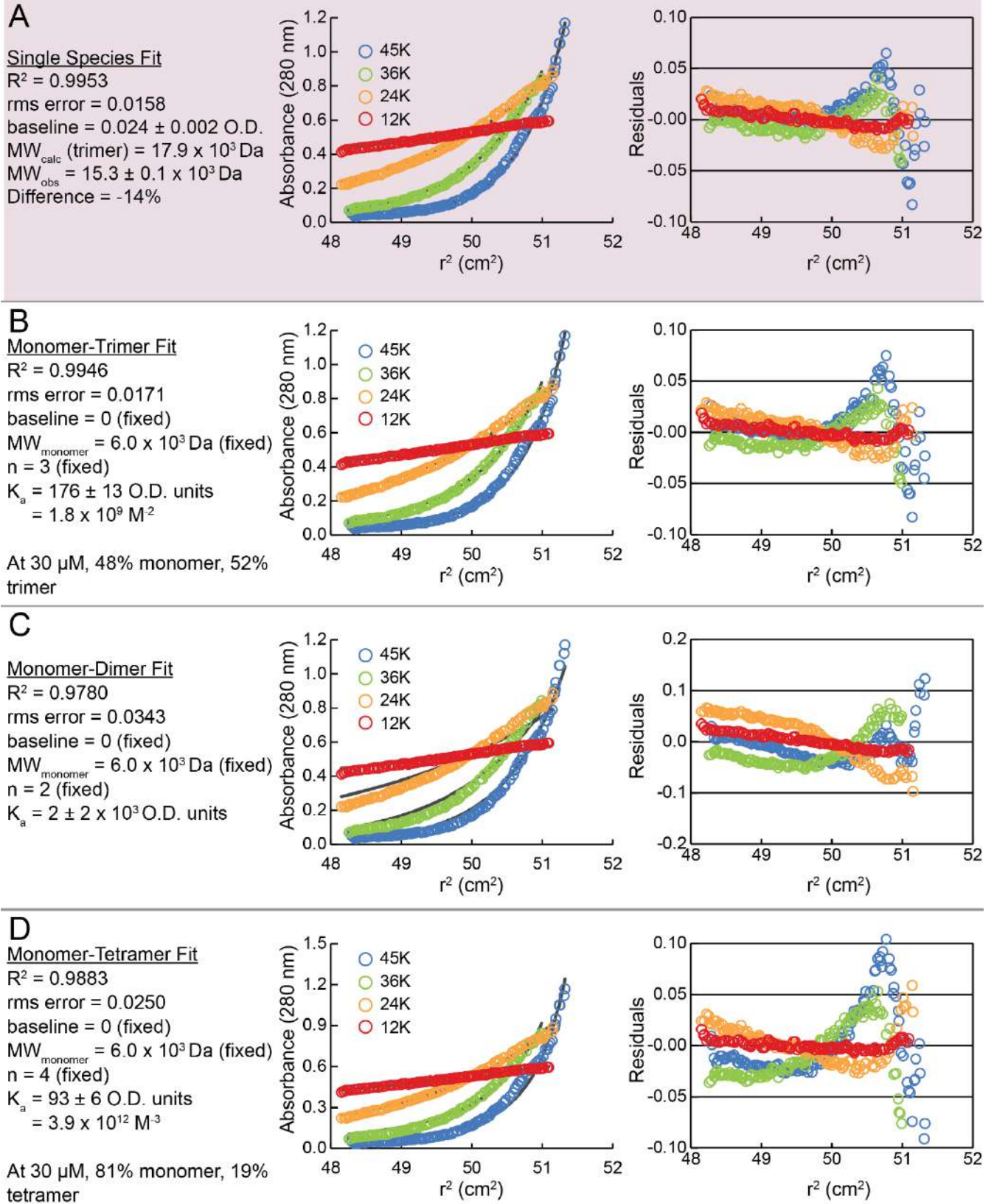


Figure 5-77. Sedimentation equilibrium data for **p17K/18K** in 20 mM sodium phosphate, pH 7. Data were fit to four different models; the best-fitting model is highlighted in magenta.

5.5.5 Triple Mutant Cycle Analysis

As described in the main text, we used least-squares regression to fit the thermodynamic data for **17/18K**, **p17/18K**, **17K**, **p17K**, **18K**, **p18K**, **1CW**, and **p1CW** to the following equation:

$$\begin{aligned} \Delta G_f = & \Delta G_f^o + (\Delta\Delta G_{C14C(PEG)} \cdot W_{14C(PEG)}) + (\Delta\Delta G_{Q17K} \cdot W_{17K}) + (\Delta\Delta G_{A18K} \cdot W_{18K}) \\ & + (\Delta\Delta\Delta G_{14C(PEG),17K} \cdot W_{14C(PEG)} \cdot W_{17K}) + (\Delta\Delta\Delta G_{14C(PEG),18K} \cdot W_{14C(PEG)} \cdot W_{18K}) \\ & + (\Delta\Delta\Delta G_{17K,18K} \cdot W_{17K} \cdot W_{18K}) \\ & + (\Delta\Delta\Delta\Delta G_{14C(PEG),17K,18K} \cdot W_{14C(PEG)} \cdot W_{17K} \cdot W_{18K}) \end{aligned} \quad (S17)$$

where ΔG_f is the folding free energy of a given variant of **1CW**; ΔG_f^o is the folding free energy of **1CW**; $\Delta\Delta G_{C14C(PEG)}$, $\Delta\Delta G_{Q17K}$, and $\Delta\Delta G_{A18K}$ describe the intrinsic energetic consequences of the Cys₁₄ to Cys₁₄(PEG), Gln₁₇ to Lys₁₇, and Ala₁₈ to Lys₁₈ mutations, respectively;

$\Delta\Delta\Delta G_{14C(PEG),17K}$, $\Delta\Delta\Delta G_{14C(PEG),18K}$, and $\Delta\Delta\Delta G_{17K,18K}$, describe the two-way interaction energies between Cys₁₄(PEG) and Lys₁₇, between Cys₁₄(PEG) and Lys₁₈, and between Lys₁₇ and Lys₁₈, respectively, relative to the corresponding two-way interaction energies in **1CW** between Cys₁₄ and Gln₁₇, between Cys₁₄ and Ala₁₈, and between Gln₁₇ and Ala₁₈, respectively;

$\Delta\Delta\Delta\Delta G_{14C(PEG),17K,18K}$ describes the three-way interaction energy between Cys₁₄(PEG), Lys₁₇, and Lys₁₈ relative to the corresponding three-way interaction energy in **1CW** between Cys₁₄, Gln₁₇, and Ala₁₈; $W_{14C(PEG)}$ is 0 when position 14 is Cys or 1 when it is Cys(PEG); $W_{17K} = 0$ when position 17 is Gln or 1 when it is Lys; $W_{18K} = 0$ when position 18 is Ala or 1 when it is Lys. Parameters obtained from this analysis are given in Table 5-2

Table 5-2. Triple mutant cycle analysis of folding free energy data for **1CW** variants at 70 °C (338.15 K).

	Energetic Contribution (kcal mol ⁻¹)
ΔG_f°	-16.43
$\Delta\Delta G_{C14C(PEG)}$	0.26
$\Delta\Delta G_{Q17K}$	-0.06
$\Delta\Delta G_{A18K}$	1.33
$\Delta\Delta\Delta G_{14C(PEG),17K}$	-0.27
$\Delta\Delta\Delta G_{14C(PEG),18K}$	-0.22
$\Delta\Delta\Delta G_{17K,18K}$	0.06
$\Delta\Delta\Delta\Delta G_{14C(PEG),17K,18}$	-0.57

Table 5-3. CD spectra and variable temperature CD data for PEGylated and non-PEGylated variants of **1CW** at 70 °C.^a

Peptide	Sequence	$[\theta]_{222}$	% helicity	T _m (°C)	ΔG_f (kcal mol ⁻¹)	$\Delta\Delta G_f$ (kcal mol ⁻¹)
1CW	Ac-EVEALEKKVAALECKVQALEKKVEALEHGWDGR-CONH ₂	-25.2 ± 0.3	67	78.3	-16.43 ± 0.05	
p1CW	Ac-EVEALEKKVAALE <u>C</u> KVQALEKKVEALEHGWDGR-CONH ₂	-26.5 ± 0.4	69	76.8	-16.17 ± 0.05	0.26 ± 0.07
7A	Ac-EVEALEAKVAALECKVQALEKKVEALEHGWDGR-CONH ₂	-23.9 ± 0.4	62	74.5	-15.41 ± 0.03	
p7A	Ac-EVEALEAKVAALE <u>C</u> KVQALEKKVEALEHGWDGR-CONH ₂	-26.0 ± 0.3	68	73.1	-15.20 ± 0.03	0.21 ± 0.04
15A	Ac-EVEALEKKVAALECAVQALEKKVEALEHGWDGR-CONH ₂	-24.5 ± 0.3	65	68.5	-14.01 ± 0.02	
p15A	Ac-EVEALEKKVAALE <u>C</u> AVQALEKKVEALEHGWDGR-CONH ₂	-26.8 ± 0.3	70	66.7	-13.57 ± 0.02	0.44 ± 0.03
21A	Ac-EVEALEKKVAALECKVQALEAKVEALEHGWDGR-CONH ₂	-24.6 ± 0.3	67	73.2	-15.23 ± 0.02	
p21A	Ac-EVEALEKKVAALE <u>C</u> KVQALEAKVEALEHGWDGR-CONH ₂	-29.6 ± 0.3	76	71.0	-14.65 ± 0.01	0.58 ± 0.03

| | | | | | |
1 5 10 15 20 25 30

^aData collected at 30 μM peptide concentration in 20 mM sodium phosphate, pH 7 at 25 °C. % helicity estimated using the program K2D3. ΔG_f and $\Delta\Delta G_f$ were calculated at 70 °C. $[\theta]_{222}$ is in units of 10³ deg dmol cm⁻² res⁻¹. C denotes a Cys residue modified with a 2000 Da PEG-maleimide (see Figure 1).

Table 5-3. Variable temperature CD data for **1CW**, **17K**, **18K**, **17/18K**, and their PEGylated counterparts at 70 °C.^a

Peptide	Sequence	T _m (°C)	ΔG _f (kcal mol ⁻¹)	ΔΔG _f (kcal mol ⁻¹)
1CW	Ac-EVEALEKKVAALECKVQALEKKVEALEHGWDGR-CONH ₂	78.3	-16.43 ± 0.05	0.26 ± 0.07
p1CW	Ac-EVEALEKKVAALECKVQALEKKVEALEHGWDGR-CONH ₂	76.8	-16.17 ± 0.05	
17K	Ac-EVEALEKKVAALECKVKALEKKVEALEHGWDGR-CONH ₂	79.3	-16.49 ± 0.04	-0.01 ± 0.06
p17K	Ac-EVEALEKKVAALECKVKALEKKVEALEHGWDGR-CONH ₂	79.5	-16.50 ± 0.04	
18K	Ac-EVEALEKKVAALECKVQKLEKKVEALEHGWDGR-CONH ₂	73.1	-15.10 ± 0.03	0.04 ± 0.04
p18K	Ac-EVEALEKKVAALECKVQKLEKKVEALEHGWDGR-CONH ₂	73.3	-15.06 ± 0.03	
17/18K	Ac-EVEALEKKVAALECKVKKLEKKVEALEHGWDGR-CONH ₂	73.8	-15.23 ± 0.03	-0.85 ± 0.06
p17/18K	Ac-EVEALEKKVAALECKVKKLEKKVEALEHGWDGR-CONH ₂	79.2	-16.08 ± 0.04	

1 5 10 15 20 25 30

^aData are for 30 μM peptide solutions in 20 mM sodium phosphate, pH 7. ΔG_f and ΔΔG_f were calculated at 70 °C. **C** denotes a Cys residue modified with a 2000 Da PEG-maleimide (see Figure 1).

5.6 References

1. Abuchowski, A.; Vanes, T.; Palczuk, N. C.; Davis, F. F., Alteration of Immunological Properties of Bovine Serum-Albumin by Covalent Attachment of Polyethylene-Glycol. *J. Biol. Chem.* **1977**, *252* (11), 3578–3581.
2. Harris, J. M.; Chess, R. B., Effect of pegylation on pharmaceuticals. *Nat. Rev. Drug Discov.* **2003**, *2* (3), 214–221.
3. Veronese, F. M.; Mero, A., The Impact of PEGylation on Biological Therapies. *Biodrugs* **2008**, *22* (5), 315–329.
4. Rosendahl, M. S.; Doherty, D. H.; Smith, D. J.; Carlson, S. J.; Chlipala, E. A.; Cox, G. N., A Long-Acting, Highly Potent Interferon α-2 Conjugate Created Using Site-Specific PEGylation. *Bioconjugate Chem.* **2005**, *16* (1), 200–207.
5. Dirksen, A.; Dawson, P. E., Expanding the scope of chemoselective peptide ligations in chemical biology. *Curr. Opin. Chem. Biol.* **2008**, *12* (6), 760–766.

6. Brocchini, S.; Godwin, A.; Balan, S.; Choi, J.-w.; Zloh, M.; Shaunak, S., Disulfide bridge based PEGylation of proteins. *Adv. Drug Deliv. Rev.* **2008**, *60* (1), 3–12.
7. Cho, H.; Daniel, T.; Buechler, Y. J.; Litzinger, D. C.; Maio, Z.; Putnam, A. M.; Kraynov, V. S.; Sim, B. C.; Bussell, S.; Javahishvili, T.; Kaphle, S.; Viramontes, G.; Ong, M.; Chu, S.; Becky, G. C.; Lieu, R.; Knudsen, N.; Castiglioni, P.; Norman, T. C.; Axelrod, D. W.; Hoffman, A. R.; Schultz, P. G.; DiMarchi, R. D.; Kimmel, B. E., Optimized clinical performance of growth hormone with an expanded genetic code. *Proc. Natl. Acad. Sci. USA* **2011**, *108* (22), 9060–9065.
8. Levine, P. M.; Craven, T. W.; Bonneau, R.; Kirshenbaum, K., Semisynthesis of Peptoid-Protein Hybrids by Chemical Ligation at Serine. *Org. Lett.* **2014**, *16* (2), 512–515.
9. Dobson, C. M., Protein folding and misfolding. *Nature* **2003**, *426* (6968), 884–890.
10. Wang, W.; Roberts, C. J., *Aggregation of Therapeutic Proteins*. Wiley: Hoboken, NJ, USA, 2010.
11. Daniel, R. M.; Cowan, D. A.; Morgan, H. W.; Curran, M. P., A correlation between protein thermostability and resistance to proteolysis. *Biochem. J.* **1982**, *207*, 641–644.
12. Cline, L. L.; Waters, M. L., The Structure of Well-Folded beta-Hairpin Peptides Promotes Resistance to Peptidase Degradation. *Biopolymers* **2009**, *92* (6), 502–507.
13. Ahmad, S.; Kumar, V.; Ramanand, K. B.; Rao, N. M., Probing protein stability and proteolytic resistance by loop scanning: A comprehensive mutational analysis. *Protein Sci.* **2012**, *21* (3), 433–446.
14. Hermeling, S.; Crommelin, D.; Schellekens, H.; Jiskoot, W., Structure-Immunogenicity Relationships of Therapeutic Proteins. *Pharmaceut. Res.* **2004**, *21* (6), 897–903.
15. Ohkuri, T.; Nagatomo, S.; Oda, K.; So, T.; Imoto, T.; Ueda, T., A Protein's Conformational Stability Is an Immunologically Dominant Factor: Evidence That Free-Energy Barriers for Protein Unfolding Limit the Immunogenicity of Foreign Proteins. *J. Immunol.* **2010**, *185* (7), 4199–4205.
16. Pandey, B. K.; Smith, M. S.; Torgerson, C.; Lawrence, P. B.; Matthews, S. S.; Watkins, E.; Groves, M. L.; Prigozhin, M. B.; Price, J. L., Impact of Site-Specific PEGylation on the Conformational Stability and Folding Rate of the Pin WW Domain Depends Strongly on PEG Oligomer Length. *Bioconjugate Chem.* **2013**, *24* (5), 796–802.
17. Vandermeulen, G. W. M.; Tziatzios, C.; Klok, H.-A., Reversible Self-Organization of Poly(ethylene glycol)-Based Hybrid Block Copolymers Mediated by a De Novo Four-Stranded α -Helical Coiled Coil Motif. *Macromolecules* **2003**, *36* (11), 4107–4114.
18. Pechar, M.; Kopečková, P.; Joss, L.; Kopeček, J., Associative diblock copolymers of poly(ethylene glycol) and coiled-coil peptides. *Macromol. Biosci.* **2002**, *2* (5), 199–206.
19. Shu, J. Y.; Tan, C.; DeGrado, W. F.; Xu, T., New Design of Helix Bundle Peptide–Polymer Conjugates. *Biomacromolecules* **2008**, *9*, 2111–2117.
20. Ogihara, N.; Weiss, M.; DeGrado, W.; Eisenberg, D., The crystal structure of the designed trimeric coiled coil coil-VaLd: implications for engineering crystals and supramolecular assemblies. *Protein Sci.* **1997**, *6* (1), 80–88.
21. Shu, J. Y.; Lund, R.; Xu, T., Solution Structural Characterization of Coiled-Coil Peptide–Polymer Side-Conjugates. *Biomacromolecules* **2012**, *13* (6), 1945–1955.
22. Lund, R.; Shu, J.; Xu, T., A Small-Angle X-ray Scattering Study of α -helical Bundle-Forming Peptide-Polymer Conjugates in Solution: Chain Conformations. *Macromolecules* **2013**, *46* (4), 1625–1632.
23. Jain, A.; Ashbaugh, H. S., Helix Stabilization of Poly(ethylene glycol)–Peptide Conjugates. *Biomacromolecules* **2011**, *12* (7), 2729–2734.

24. Hamed, E.; Xu, T.; Keten, S., Poly(ethylene glycol) Conjugation Stabilizes the Secondary Structure of α -Helices by Reducing Peptide Solvent Accessible Surface Area. *Biomacromolecules* **2013**, *14* (11), 4053–4060.
25. Louis-Jeune, C.; Andrade-Navarro, M. A.; Perez-Iratxeta, C., Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. *Proteins: Struct., Funct., Bioinf.* **2012**, *80* (2), 374–381.
26. Price, J. L.; Powers, E. T.; Kelly, J. W., N-PEGylation of a Reverse Turn is Stabilizing in Multiple Sequence Contexts unlike N-GlcNAcylation. *ACS Chem. Biol.* **2011**, *6*, 1188–1192.
27. McKnight, C. J.; Matsudaira, P. T.; Kim, P. S., NMR structure of the 35-residue villin headpiece subdomain. *Nat. Struct. Mol. Biol.* **1997**, *4* (3), 180–184.
28. Frank, B. S.; Vardar, D.; Buckley, D. A.; McKnight, C. J., The role of aromatic residues in the hydrophobic core of the villin headpiece subdomain. *Protein Sci.* **2002**, *11* (3), 680–687.
29. Lei, H. X.; Wu, C.; Liu, H. G.; Duan, Y., Folding free-energy landscape of villin headpiece subdomain from molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA* **2007**, *104* (12), 4925–4930.
30. Pandey, B. K.; Smith, M. S.; Price, J. L., Cys(i)-Lys(i+3)-Lys(i+4) Triad: A General Approach for PEG-Based Stabilization of α -Helical Proteins. *Biomacromolecules* **2014**, *15* (12), 4643–4647.
31. Durchschlag, H.; Zipper, P., Calculation of the partial volume of organic compounds and polymers. *Prog. Colloid Polym. Sci.* **1994**, *94*, 20–39.
32. Bhat, R.; Timasheff, S. N., Steric exclusion is the principal source of the preferential hydration of proteins in the presence of polyethylene glycols. *Protein Sci* **1992**, *1* (9), 1133–1143.