Stony Brook University

Development of Fluorescence Probes of Protein Folding

A Dissertation Presented

by

Matthew Douglas Watson

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Chemistry

Stony Brook University

August 2017

Stony Brook University

The Graduate School

Matthew Douglas Watson

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Daniel P. Raleigh, PhD – Dissertation Advisor
Professor, Department of Chemistry

Lisa M. Miller, PhD – Chairperson of Defense
Affiliated Associate Professor, Department of Chemistry

Kathlyn A. Parker, PhD – Third Member
Distinguished Professor, Department of Chemistry

Markus Seeliger, PhD – Outside Member
Assistant Professor, Department of Pharmacological Sciences

This dissertation is accepted by the Graduate School

Charles Taber

Dean of the Graduate School

Abstract of the Dissertation

Development of Fluorescence Probes of Protein Folding

by

Matthew Douglas Watson

Doctor of Philosophy

in

Chemistry

Stony Brook University

2017

Many proteins depend on a stable, well-defined three-dimensional structure to perform biological functions. Protein folding is the process through which a polypeptide chain rearranges to adopt the native structure encoded in its amino acid sequence. The high intrinsic time resolution and signal-to-noise make fluorescence spectroscopy an ideal approach for protein folding experiments. However, interpretation of intrinsic tryptophan fluorescence changes is complicated by multiple fluorescence quenching mechanisms and solvent interactions. This work describes the use of selenomethionine (M_{Se}), the selenium analogue of methionine as a quencher of tryptophan and 4-cyanophenylalanine (F_{CN}) fluorescence to follow protein and peptide folding. The introduction of a quencher simplifies the interpretation of fluorescence changes in both amino acids and allows for the examination of specific side chain interactions.

The approach was extended to the study of protein-protein interactions by incorporation of F_{CN} and M_{Se} into the monomeric units of a heterodimeric coiled coil. The fluorescence signal intensity allows for the detection of coiled coil formation at lower protein concentrations than what

is accessible by standard circular dichroism (CD) methods. In addition, it is shown that the fluorescence quenching system can also be used to rapidly and accurately determine the K_D of the coiled coil interaction.

The structural revolution of the past several decades has generated a vast amount of data on protein structure but has had comparatively less impact on our understanding of the origins of protein stability. An analysis of published stability data was carried out examining length-dependent thermodynamic properties. A clear correlation with chain length is observed for ΔH, ΔS and ΔC_p. Although $\Delta G°$ at 298 K of individual proteins cannot be accurately determined using this model, predictions for the thermal stability of whole proteomes are possible. Existing datasets were significantly expanded and differences between proteins from mesophilic and thermophilic organisms were examined. The large dataset also permitted the reassessment of the existence of convergence temperatures in proteins and an analysis of thermodynamic mutation data was used to predict thermal shifts due to ligand binding.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

$[CC]$	Concentration of coiled coil monomers
$[F]$	Concentration of free folded protein
$[F_{CN}]$	Concentration of *p*-cyanophenylalanine
$[L]$	Concentration of free ligand
$[L]_T$	Total concentration of ligand
$[M_{Se}]$	Concentration of selenomethionine
$[P]$	Total concentration of protein
$[Tyr]$	Concentration of tyrosine
$[U]$	Concentration of unfolded protein
$[\theta]_{222}$	Molar ellipticity at 222 nm
$[\theta]_C$	Molar ellipticity at 222 nm for a random coil
$[\theta]_H$	Molar ellipticity at 222 nm for a 100% helical peptide
$\langle(\Delta N)^2\rangle$	Chain length variance
$\langle N\rangle$	Mean chain length
a_d	Baseline slope in the unfolded state
a_n	Baseline slope in the folded state
b_C	Intercept of chain-length dependent heat capacity change upon unfolding
b_d	Baseline intercept in the unfolded state
b_H	Intercept of chain-length dependent enthalpy change upon unfolding
b_n	Baseline intercept in the folded state
b_S	Intercept of chain-length dependent entropy change upon unfolding
CC-19$F_{CN}A_N^{3.5}$	3.5-heptad acidic coiled coil monomer with *p*-cyanophenylalanine at position 19
CC-19W$A_N^{3.5}$	3.5-heptad acidic coiled coil monomer with tryptophan at position 19
CC-20$M_{Se}B_N^{3.5}$	3.5-heptad basic coiled coil monomer with selenomethionine at position 20
CD	Circular dichroism

CTL9	C-terminal domain of ribosomal protein L9
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
F_{291}	Fluorescence intensity at 291 nm
F_{CN}	*p*-cyanophenylalanine
f_h	Fractional helical content
F_{max}	Maximum fluorescence intensity
F_{min}	Minimum fluorescence intensity
Fmoc	9-fluorenylmethyloxycarbonyl
FRET	Förster resonance energy transfer
HFIP	1,1,1,3,3,3-hexafluoro-2-propanol
HP35	35-residue fragment of the villin helical headpiece subdomain
HP36	36-residue fragment of the villin headpiece helical subdomain
HPLC	High pressure liquid chromatography
K_D	Dissociation constant
K_D^M	Dissociation constant at T_M
K_U	Equilibrium constant of unfolding
LC-ESI-TOFMS	Liquid chromatography electrospray ionization time-of-flight mass spectrometry
m	Concentration dependence of tyrosine fluorescence intensity at 291 nm
MAD	Multiwavelength anomalous diffraction
MALDI-TOFMS	Matrix-assisted laser desorption-ionization time-of-flight mass spectrometry
m_C	Slope of chain-length dependent heat capacity change upon unfolding
MD	Molecular dynamics
MEM	Maximum entropy method
m_H	Slope of chain-length dependent enthalpy change upon unfolding
m_S	Intercept of chain-length dependent entropy change upon unfolding
M_{Se}	Selenomethionine

M_{SeO}	Selenomethionine selenoxide
N	Number of amino acid residues
NMR	Nuclear magnetic resonance
$P(N)$	Chain length distribution
R	Gas constant
r	Distance between FRET donor and acceptor
R_0	Förster radius
SASA	Solvent accessible surface area
T	Temperature
T_C	Midpoint of the cold unfolding transition
TCSPC	Time-correlated single photon counting
TFA	Trifluoroacetic acid
T_H	Enthalpic convergence temperature
T_M	Midpoint of the thermal unfolding transition
T_{max}	Temperature of maximum stability
TOF/TOF	Time-of-flight/time-of-flight
T_R	Reference temperature
T_S	Entropic convergence temperature
TSA	Thermal shift assay
UV	Ultraviolet
$\Delta C_p°$	Heat capacity change upon unfolding
ΔG	Gibbs free energy change upon unfolding
$\Delta G°$	Gibbs free energy change upon unfolding under standard conditions
$\Delta G_L°$	Gibbs free energy change upon unfolding under standard conditions in the presence of ligand
ΔH	Enthalpy change upon unfolding
$\Delta H°$	Enthalpy change upon unfolding under standard conditions

ΔH_M	Enthalpy change upon unfolding at T_M
ΔH_R	Enthalpy change upon unfolding at T_R
ΔS	Entropy change upon unfolding
$\Delta S°$	Entropy change upon unfolding under standard conditions
ΔSASA	Solvent accessible surface area change upon unfolding
ΔS_M	Entropy change upon unfolding at T_M
ΔS_R	Entropy change upon unfolding at T_R
ΔT_M	Change in the midpoint of thermal unfolding
$\Delta\Delta G°$	Change in Gibbs free energy change upon unfolding under standard conditions
$\Delta\Delta G_L°$	Change in Gibbs free energy change upon unfolding under standard conditions in the presence of ligand
θ	Spectroscopic observable

Acknowledgments

I would like to thank my advisor, Professor Daniel Raleigh for his insight and guidance throughout my time at Stony Brook University. It has been a great privilege to learn and work in his laboratory.

I would also like to thank the other members of my committee, Professor Lisa Miller and Professor Kathlyn Parker, for their advice and encouragement. Thank you to Professor Markus Seeliger for serving as my outside member.

I would also like to thank our collaborator, Dr. Osman Bilsel for his time and expertise in fluorescence data collection. His knowledge and insight in the interpretation of the data were invaluable. Thanks to Professor Bob Matthews and members of his laboratory, who generously provided access to their laboratory at UMass.

Thank you to all members of the Raleigh laboratory who I have worked alongside: Bing Shan, Humeyra Taskent, Fanling Meng, Peter Marek, Shifeng Xiao, Harris Noor, Vadim Patsalo, Ping Cao, Wenli Meng, Ling-Hsien Tu, Hui Wang, Yuan Chen, Bowu Luan, Amy Wong, Ivan Peran, Xiaoxue Zhang, Natalie Stenzoski, Rehana Akter, Zack Ridgway, Junjie Zou and Daeun Noh. Special thanks to Humeyra, who served as my mentor in the lab, as well as Ivan and Junjie with whom I collaborated directly. Thanks also to Jeremy Monroe, who found time in his undergraduate schedule to make a major contribution to the work in this dissertation.

Thank you to Dr. Bela Ruzsicska for his assistance with mass spectrometry, Katherine Hughes and all other members of the Chemistry department main office for their assistance in administrative matters.

Finally, I would like to thank my family and friends for their support.

Publications

[1] Watson, M. D., Peran, I., Zou, J., Bilsel, O., and Raleigh, D. P. (2017) Selenomethionine quenching of tryptophan fluorescence provides a simple probe of protein structure, *Biochemistry 56*, 1085-1094.

[2] Brar, G. S., Barrow, B. M., Watson, M., Griesbach, R., Choung, E., Welch, A., Ruzsicska, B., Raleigh, D. P., and Zraika, S. (2017) Neprilysin is required for angiotensin-(1-7)'s ability to enhance insulin secretion via its proteolytic activity to generate angiotensin-(1-2), *Diabetes 66*, 2201-2212.

[3] Peran, I., Watson, M. D., Bilsel, O., and Raleigh, D. P. (2016) Selenomethionine, p-cyanophenylalanine pairs provide a convenient, sensitive, non-perturbing fluorescent probe of local helical structure, *Chem Commun (Camb) 52*, 2055-2058.

[4] Watson, M. D., Peran, I., and Raleigh, D. P. (2016) A non-perturbing probe of coiled coil formation based on electron transfer mediated fluorescence quenching, *Biochemistry 55*, 3685-3691.

[5] Wong, A. G., Wu, C., Hannaberry, E., Watson, M. D., Shea, J. E., and Raleigh, D. P. (2016) Analysis of the amyloidogenic potential of pufferfish (takifugu rubripes) islet amyloid polypeptide highlights the limitations of thioflavin-T assays and the difficulties in defining amyloidogenicity, *Biochemistry 55*, 510-518.

[6] Marine, J. E., Song, S., Liang, X., Watson, M. D., and Rudick, J. G. (2015) Bundle-forming alpha-helical peptide-dendron hybrid, *Chem Commun (Camb) 51*, 14314-14317.

[7] Peran, I., Oudenhoven, T., Woys, A. M., Watson, M. D., Zhang, T. O., Carrico, I., Zanni, M. T., and Raleigh, D. P. (2014) General strategy for the bioorthogonal incorporation of strongly absorbing, solvation-sensitive infrared probes into proteins, *J Phys Chem B 118*, 7946-7953.

1. Background

1.1. Protein Folding

Protein folding is the process through which the amino acid monomers that make up a protein or peptide adopt a stable geometric configuration (Figure 1.1). Although some, particularly larger proteins may have multiple low-energy conformations, the target of the folding process is generally the biologically active structure, termed the native state. In most evolutionarily successful proteins, the native state is also the lowest energy conformation. In rare cases, some proteins are cleaved from a longer, folded pro-sequence, leaving the mature protein trapped in a metastable state that is not the most thermodynamically stable structure, but is required for activity.[1, 2] Understanding the protein folding process is fundamental to understanding the action of evolved proteins, as well as in the *de novo* design of functional proteins. The processes through which proteins fold are also closely tied to protein misfolding processes, which have implications for research on diseases such as Alzheimers, Parkinsons and type II diabetes.

The fundamental 'problem' of protein folding is summarized by the work of two scientists, Christian Anfinsen and Cyrus Levinthal. In the 1950s, Anfinsen demonstrated that bovine pancreatic nuclease, a protein which contains 4 disulfide bridges, that had been reduced to a non-functional state could be returned to the functional native state by oxidizing the sample in "native buffer" (in contrast, oxidation under denaturing conditions led to a random assortment of proteins with native and non-native disulfide bonds).[3] This deceptively simple experiment proved that the native structure of a protein is encoded in the primary sequence—the order of amino acid monomers. In 1969 the field was further complicated by Levinthal's estimation of the time

required for a protein to reach its native state by a random walk through configurational space. To obtain a conservative estimate, a protein can be approximated as a chain of monomers without side chains, each possessing two dihedral angles able to independently sample three rotamers. Using this simple model, it is apparent that even for a small protein of only fifty residues there exist 3^{100} unique conformations. If such a protein were able to sample a new conformation every 100 fs—the timescale of bond rotation—the time required to sample every possible conformation of a fifty residue chain would be seventeen orders of magnitude longer than the age of the universe.[4] Taking these two observations, as well as the fact that proteins fold on very short timescales into consideration,[5-7] it is clear that proteins cannot explore every possible conformation of residues in order to arrive at their native state.

A rough consensus model of the protein folding 'funnel' has emerged, which describes the folding process in terms of a conformational landscape that slopes towards the center—the native state—effectively guiding the protein to the lowest free energy configuration (Figure 1.2). The simplest representations feature a smooth landscape, which would allow an unfolded protein to reach the native state by many different paths. More complex models explain the existence of well-defined folding pathways through 'roughness' in the funnel—local maxima and minima which guide proteins along specific routes between folded and unfolded conformations. There are several models proposed for folding mechanisms. In the hydrophobic collapse model hydrophobic residues that are poorly solvated by water collapse while simultaneously forming secondary structure.[8] By contrast, the framework model proposes that secondary structure is formed rapidly followed by hydrophobic collapse.[9-12] Yet a third model, the nucleation-condensation model holds that initial hydrophobic collapse forms a 'nucleus' that induces further rearrangement to form secondary structure.[13] Local minima along a folding pathway may represent folding

intermediates—stable, defined structures that may rearrange to the native state or convert back to the unfolded ensemble. The details of the process, however, remain sketchy, at best. In particular, the definition of the unfolded ensemble under equilibrium conditions and the order of folding events with respect to hydrophobic versus electrostatic and hydrogen bonding interactions remain to be clearly elucidated.

1.2. Thermodynamics of Protein Folding

Overall protein stability, described by the free energy of unfolding $\Delta G(T)$ at any temperature T can be subdivided into the enthalpy $\Delta H(T)$ and entropy $\Delta S(T)$ of unfolding.

$$\Delta G(T) = \Delta H(T) + T\Delta S(T) \tag{1.1}$$

Where $\Delta G(T)$ is the free energy of unfolding at a temperature T, and $\Delta H(T)$ and $\Delta S(T)$ are the enthalpy and entropy of unfolding at T. The thermodynamic standard state is taken to be 1 atmosphere pressure at 298.15 K with molar units used for concentrations and all activity coefficient are assumed to be equal to 1.0. In the context of protein folding, the enthalpy of unfolding describes the energy change due to differences in protein-solvent, protein-protein and solvent-solvent interactions between the native state and the unfolded ensemble. The entropy of unfolding represents contributions from both the chain entropy, which opposes folding, and the solvation entropy, which stabilizes the folded state over some range. Both ΔH and ΔS of unfolding are temperature dependent, described by the change in heat capacity upon unfolding, $\Delta C_p°$ which is large and positive for the folding process.

$$\Delta H(T) = \Delta H(T_R) + \Delta C_p°(T - T_R) \tag{1.2}$$

$$\Delta S(T) = \Delta S(T_R) + \Delta C_p° \ln\left(\frac{T}{T_R}\right) \tag{1.3}$$

Where T_R is any convenient reference temperature in K, and ΔH_R and ΔS_R are the enthalpy and entropy of unfolding at T_R. $\Delta C_p°$ is assumed to be constant over the entire experimentally accessible temperature range. Thus, the temperature dependent free energy of unfolding can be described by the Gibbs-Helmholtz equation (1.4)

$$\Delta G(T) = \left[\Delta H_R + \Delta C_p°(T - T_R)\right] + T\left[\Delta S_R + \Delta C_p° \ln\left(\frac{T}{T_R}\right)\right] \quad (1.4)$$

Where $\Delta G(T)$ is the free energy change upon unfolding at any temperature T, T_R is any convenient reference temperature in K, T is the temperature in K, ΔH_R and ΔS_R are the changes in enthalpy and entropy upon unfolding at T_R, and $\Delta C_p°$ is the change in heat capacity upon unfolding. It is apparent from a plot of $\Delta H(T)$, $T\Delta S(T)$ and $\Delta G(T)$ *versus* temperature that there are two temperatures where $\Delta G(T)$ is zero (Figure 1.3). The higher of these two temperatures is the midpoint of thermal unfolding, T_M. At this temperature $\Delta H(T)$ is equal to $T_M \Delta S(T)$, hence equation (1.4) can be rewritten as the modified Gibbs-Helmholtz equation (1.5)

$$\Delta G(T) = \Delta H_M \left(1 - \frac{T}{T_M}\right) + \Delta C_p° \left(T - T_M - T \ln\left(\frac{T}{T_M}\right)\right) \quad (1.5)$$

Where $\Delta G(T)$ is the free energy change upon unfolding, T_M is the midpoint of the thermal unfolding transition in K, T is the temperature in K, ΔH_M is the change in enthalpy upon unfolding at T_M, and $\Delta C_p°$ is the change in heat capacity upon unfolding. The relationship between the linear temperature dependence of $\Delta H(T)$ and the nonlinear temperature dependence of $\Delta S(T)$ determines the stability range and maximum stability of any protein.

1.3. Mechanisms of Protein Denaturation

Proteins can be unfolded from their native state by the addition of chemical denaturants or heat energy. One model of the effects of chemical denaturants such as urea and guanidinium

hydrochloride is that they affect the stability of proteins by altering protein-solvent interactions. The properties of proteins that fold in an aqueous environment resemble those of a polymer in a poor solvent: energetically unfavorable interactions between water molecules and the protein chain are avoided when the chain collapses, resulting in a larger number of intramolecular interactions relative to the number of protein-solvent interactions. In this model, chemical denaturants reduce the unfavorability of protein-solvent interactions, stabilizing the unfolded ensemble relative to the native state and resulting in a shift in the folding equilibrium towards the unfolded ensemble.[14, 15]

A second model of these effects is that denaturants interact directly with the protein chain. If the interaction of the denaturants with the protein is more favorable than their interactions with water, the protein folding equilibrium will be shifted towards the unfolded ensemble, maximizing the surface area of the protein and hence the number of protein-denaturant interactions.[16-20] In either case, the effect of chemical denaturants is to stabilize the protein unfolded ensemble.

Thermal or heat denaturation exerts its effects by increasing the entropic cost of folding to a well-defined structure relative to the enthalpic change. The temperature dependence of protein stability can be accurately modeled by the modified Gibbs-Helmholtz equation (1.5). A thermal denaturation monitored by a spectroscopic change can be modeled by equation (1.6)

$$\theta = \frac{a_n + b_n T + (a_d + b_d T) e^{\left(-\frac{\Delta G(T)}{RT}\right)}}{1 + e^{\left(-\frac{\Delta G(T)}{RT}\right)}} \qquad (1.6)$$

Where θ is any spectroscopic signal that reports on the foldedness of the protein, a_n and b_n are the intercept and slope of the folded state baseline, respectively, a_d and b_d are the intercept and slope of the denatured ensemble baseline, respectively, T is the temperature in K, R is the gas constant and $\Delta G(T)$ is the free energy change upon unfolding at T determined from the modified Gibbs-Helmholtz equation (1.5).

1.4. Length-Dependent Properties of Proteins

Advances in methods for the study of protein thermodynamics and kinetics as well as structure prediction and determination have made it possible to gain a deep understanding of proteins on an individual basis. While access to this type of knowledge has revolutionized drug development and protein design, it does not necessarily translate well to the study of organism-scale protein science, such as signaling networks. Unlike more reductionist approaches, systems biology can be employed to model and predict complex systems using more coarse-grained models of protein structure and stability than would be useful in the study of an isolated protein. Perhaps the simplest model for proteins are as simple polymers the properties of which depend solely on chain length. Although such a model is not useful in rational drug design, averaged over an entire genome it can be used to make surprisingly accurate predictions about an organism.[21]

Both $\Delta H(T)$ and $\Delta S(T)$ of protein unfolding can be described as linear values with respect to chain length. In addition, $\Delta C_p°$ of unfolding has also been shown to increase linearly with chain length, being strongly correlated with the change in solvent accessible surface area. Given sufficient data, these linear relationships, together with the Gibbs-Helmholtz equation (1.4), allow the stability of a hypothetical protein to be calculated at any temperature. Of course, such a calculation assumes that the thermodynamic properties of all amino acids are identical, which is clearly not the case. In addition, the simple model does not consider other stabilizing effects, such as disulfide bonds. Consequently, these simple linear relationships are unlikely to result in an accurate calculation of stability for any individual real protein.

Recent advances in sequencing technology have resulted in an explosion of readily available sequence data, in some cases for whole genomes. Considerable effort has been expended towards predicting structure and function using such information, with varying degrees of success.

Given the linear dependence of $\Delta H(T)$, $\Delta S(T)$ and $\Delta C_p°$ on chain length, predicting the stability of these uncharacterized proteins appears to be a more accessible goal. Although such predictions are unlikely to be highly accurate for a single protein, averaged over an entire sequenced genome, they can be used to model protein stability for whole organisms. These models could help reveal the origins of the high thermal stability of proteins that evolved in organisms inhabiting the environment around volcanic vents on the ocean floor. The evolutionary biology of such extremophiles is potentially of interest in understanding evolution on a primitive Earth.

1.5. Fluorescence Spectroscopy

The rapidity of the folding process severely limits the ability of most common spectroscopic techniques to provide insight into its dynamics. Nuclear magnetic resonance (NMR), one of the most important methods for protein studies cannot probe changes on a relevant timescale, except for exceptional cases or for studies conducted at high pressures.[22] Imaging techniques such as X-ray crystallography and cryo-electron microscopy can only collect data on static systems, which in the case of proteins limits them to study of the native state. Fluorescence spectroscopy is the most popular approach to studying fast protein dynamics. The technique is attractive for its high time resolution and benefits from high sensitivity arising from the intensity of the fluorescence signal. Equilibrium fluorescence measurements are facile, rapid and easily interpreted making them attractive for routine experiments as well as in high-throughput applications. Specialized techniques such as stopped-flow can easily be monitored by fluorescence, obtaining information on rapid structural changes in proteins. Time-resolved fluorescence techniques monitoring the decay rate of the fluorophore excited state can provide information on the internal structure and dynamics of a protein on the nanosecond timescale.

Many fluorescence techniques can be modified to incorporate quenching, a variety of mechanisms by which fluorescence intensity is decreased by proximity to other compounds. Förster Resonance Energy Transfer (FRET) has become particularly widespread in biological applications due to the utility of the technique in determination of intra or intermolecular distances. The FRET effect is characterized by non-radiative energy transfer from a fluorophore (the donor) in the excited state to a second functionality (the acceptor) with an absorbance spectrum overlapping with the emission spectrum of the donor. Often described as a "molecular ruler," FRET reports on distance due to the distance dependence of energy transfer efficiency. All FRET pairs have a characteristic Förster distance (R_0) at which the energy transfer efficiency is 50%. As the distance between the donor and acceptor decreases the FRET efficiency approaches 100%, and approaches zero as the distance becomes greater. The distance dependence is non-linear, being more sensitive near R_0, consequently the useful range of FRET is from approximately 0.5 to 1.5 times R_0. This results in FRET being incapable of accurately measuring very short distances (Figure 1.4).

Other quenching mechanisms—intersystem crossing (heavy atom quenching), Dexter electron transfer and photoinduced electron transfer—occur over much shorter distances than FRET, requiring overlap of the fluorophore and quencher molecular orbitals.[23] The resulting quenching effects have a very strong distance dependence, with quenching observed only in the case of contact between the fluorophore and quencher. This strong distance sensitivity makes other quenching mechanisms complementary to FRET, able to report on actual contact where FRET results are ambiguous.

1.6. Incorporation of Unnatural Amino Acids

Unnatural amino acids have been widely used to incorporate functional groups of interest into proteins for studies of structure, folding and dynamics.[24-26] The two major approaches to this are either to incorporate the functionality of interest into the protein during translation (or synthesis) or to ligate the functionality onto a target residue or residues after the protein has been fully formed. Use of solid-phase peptide synthesis makes either approach accessible. An unnatural amino acid can easily be protected for activation in standard coupling chemistry, or a chemically active residue, such as a cysteine can be inserted wherever desired. Peptide synthesis techniques, however, are limited to production of relatively small proteins due to the imperfect yield of the individual coupling reactions. Somewhat longer synthetic constructs can be accessed through the use of native chemical ligation techniques.[27]

The limitations of solid-phase peptide synthesis make the use of natural translational machinery a far more attractive option to produce mutant proteins. However, the very efficiency and precision of the cellular machinery that makes it an attractive tool has traditionally limited the palette of amino acids available to the protein chemist to the twenty natural amino acids. As such, functionalizing proteins with probes has until recently relied upon incorporation of a reactive residue, generally cysteine at the position of interest and ligating the probe to the cysteine post-translationally. This poses several problems; the inherent reactivity of the cysteine residue may lead to undesired modifications within the cell, and labeling proteins with multiple probes cannot be easily accomplished due to the lack of diversity in available chemistries and low yield of disulfide coupling chemistry.

In the past decade, new technologies for the incorporation of unnatural amino acids by native translational machinery have been developed and refined. Despite the exquisite specificity

of biological systems, it has become apparent that several unnatural compounds and functional groups can and will be readily incorporated into biomolecules by cells. Incorporation of selenomethionine by methionine auxotrophs, for example, has long been taken advantage of for X-ray crystallography studies.[28]

Although this sort of metabolic labeling with unnatural amino acids has greatly expanded the horizons of protein studies, the technique still limits the available amino acids to relatively conservative substitutions. In addition, if the amino acid being replaced with an unnatural surrogate occurs in many positions in the wild type protein, these other sites must be mutated to other amino acids to prevent labeling of all of them. This hurdle has recently been overcome by the advent of so-called 21st pair technology, initially developed in the Shultz lab. The 21st pair method involves charging suppressor tRNA which recognizes a stop codon (typically the 'amber' UAG codon) with an unnatural amino acid. As such, whenever the ribosome encounters the specified stop codon the unnatural amino acid will be selectively incorporated at that and only that position.[29] In order for the method to be useful for protein overexpression, an unnatural synthetase that specifically charges the suppressor tRNA with the desired unnatural amino acid must also be evolved to a high degree of specificity and expressed in the cell. The suppressor tRNA and unnatural synthetase hence form the '21st pair'.

The technology allows for the incorporation of a huge variety of unnatural amino acids, and several unnatural synthetases have been evolved. One amino acid that has been the focus of considerable study is L-*p*-cyanophenylalanine (F_{CN}) (Figure 1.5).[30-34] Extensively studied and utilized in FRET pairing studies, F_{CN} can also be used as a vibrational probe by examining the stretching mode of the nitrile group.[35-41]

1.7. The C-terminal Domain of Ribosomal Protein L9

One of the most important considerations in protein folding studies is the selection of model systems. One popular system is the C-terminal domain of ribosomal protein L9 (CTL9). Full length L9 is a 149-residue structural component of the ribosome. It is composed of two globular domains composed of mixed α-helical and β-sheet structure connected by a 33-residue α-helix (Figure 1.6). The precise role of L9 in the ribosome is not well-defined, although it has been implicated in frameshifting during translation.[42] The CTL9 construct used in this study is derived from *Geobacillus stearothermophilus*, a thermophilic bacterium that lives proximal to volcanic vents on the ocean floor.

CTL9 is comprised of the latter 91 residues of L9, incorporating the C-terminal globular domain as well as 16 of the residues that make up the connecting helix (Figure 1.7). CTL9 is of interest in the context of protein folding primarily because it follows a rapid, simple, two-state folding path to form a stable, well defined native structure.[43, 44] The high thermal stability of the construct used in this study facilitates thermal unfolding studies as the unfolded ensemble is nearly entirely unrepresented near room temperature.

Earlier work identified a point mutant of CTL9, I98A which was discovered to undergo cold denaturation at temperatures above 0 °C.[45] Although cold denaturation is predicted by thermodynamics, study of this process has been limited by the fact that it does not occur above the freezing point of water for most proteins. To date most studies of cold denaturation have relied upon the addition of denaturants or studying super-cooled proteins in micelles.[46] I98A-CTL9 therefore offers a unique opportunity to study this poorly understood unfolding process.

1.8. The Villin Headpiece Helical Subdomain

While CTL9 provides a useful model system for folding studies, the sheer complexity of folding makes even a 91-residue protein a large system. Even smaller systems are desirable to study how specific interactions contribute to the overall folding process. The villin headpiece helical subdomain (HP36) is a 36-residue fragment from the C-terminal domain of the F-actin binding protein, villin. HP36 folds rapidly and cooperatively to form a three-helix bundle that is nearly identical to the structure observed in the intact headpiece. The numbering system used here is based upon the residue designations in the full villin protein, the C-terminal methionine residue in HP36 being designated M41. The protein contains four phenylalanine residues, F47, F51, F58 and F76, the first three of which are hypothesized to be important to stabilization of the folded state (Figure 1.8).

1.9. Figures

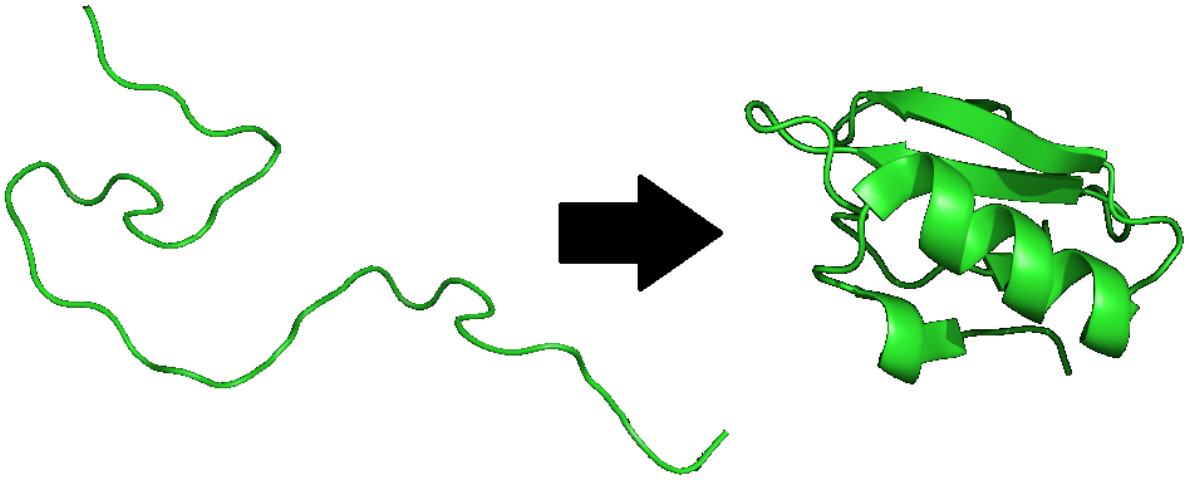


Figure 1.1. Model of two-state protein folding. Disordered polypeptide chains (left) fold to form proteins with a specific, defined structure (right). Rendered in PyMol from PDB structure 1LW6.

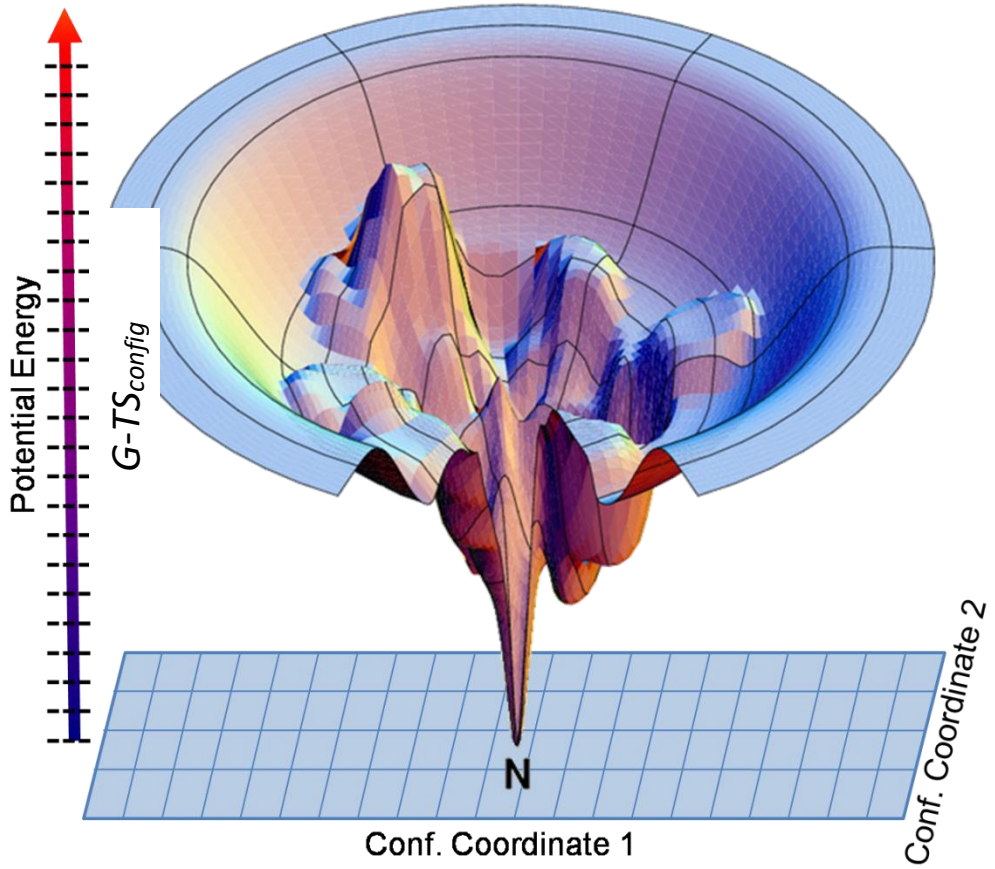


Figure 1.2. Funnel model of protein folding. Proteins do not search conformational space randomly to arrive at their native structure. The energy landscape explorable by a protein (described here by two conformational coordinates) is thought to resemble a rough-surfaced funnel that directs the protein to the lowest free energy conformation (**N**). The width (cross-sectional area) of the funnel at a given height is proportional to the configurational entropy, S_{config}. Adapted from.[47]

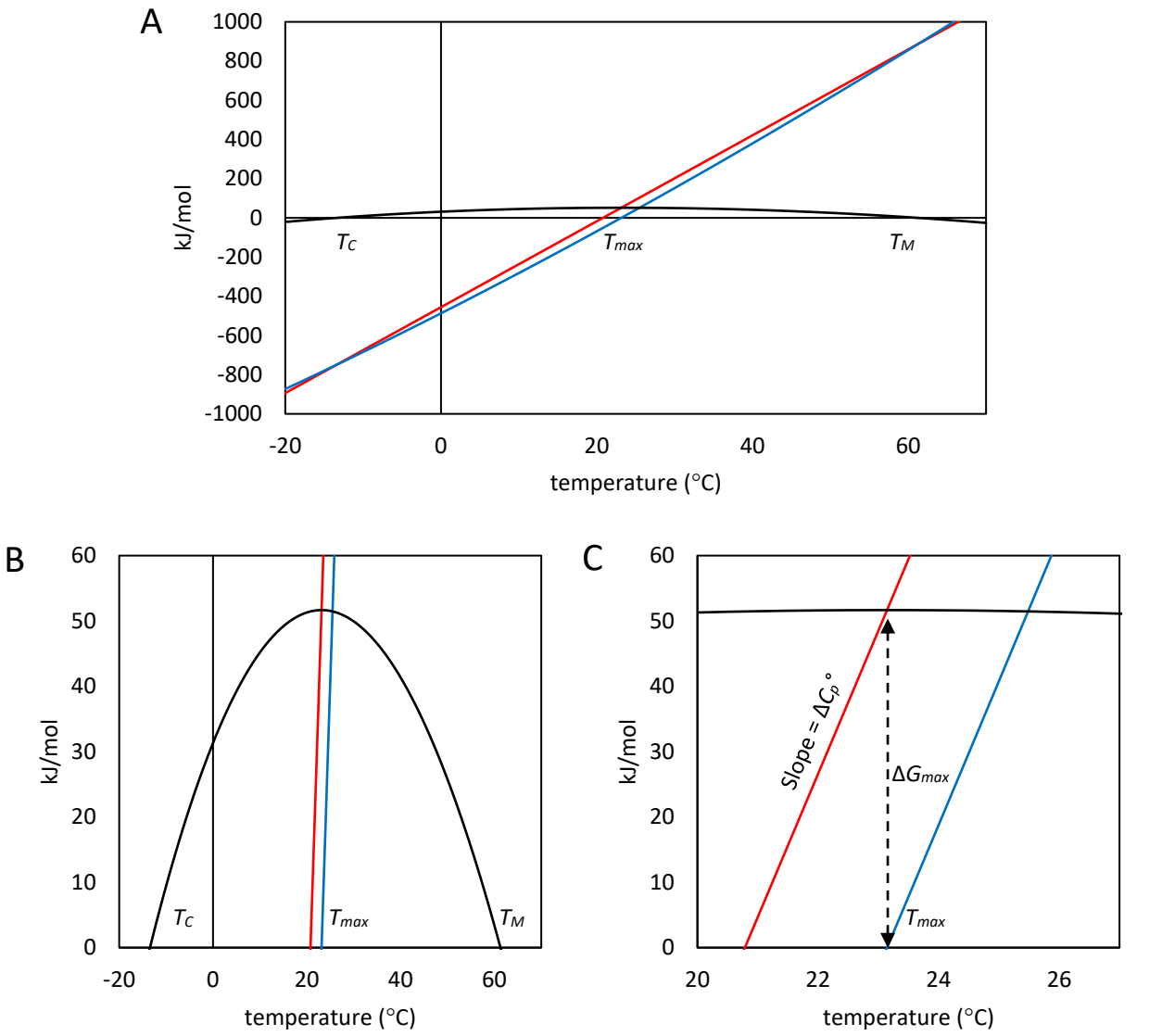


Figure 1.3. Temperature dependence of protein stability. **(A)** $\Delta G(T)$ of unfolding (black) is the difference of $\Delta H(T)$ (red) and $T\Delta S(T)$ (blue). **(B)** Detail of the range of stable folding. There are two temperatures at which $\Delta H(T)$ is equal to $T\Delta S(T)$ and $\Delta G(T)$ is zero: the midpoint of cold denaturation, T_C and the midpoint of thermal denaturation, T_M. Note that $T\Delta S(T)$ is not linear, as is apparent in **(A)** but can appear so over a small range. **(C)** Detail of the point of maximum folding stability. $\Delta G(T)$ reaches its maximum value where $T\Delta S(T)$ is zero, T_{max}.

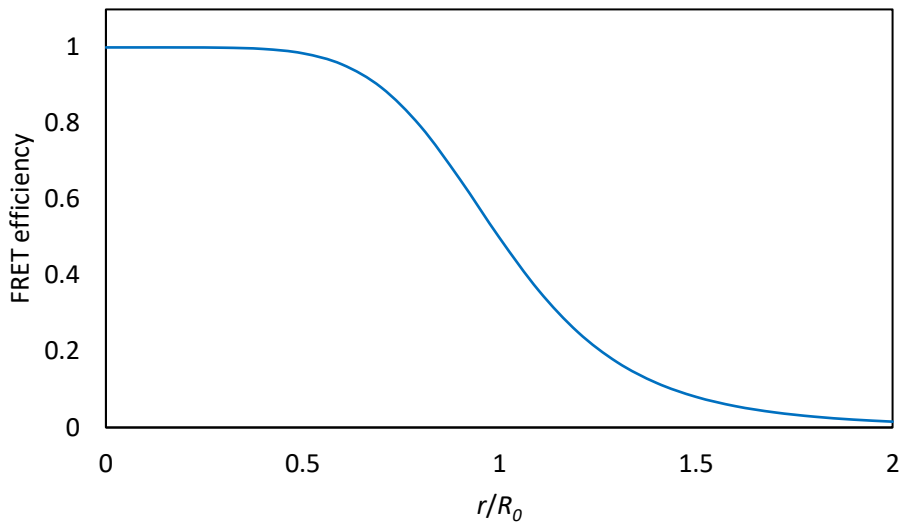


Figure 1.4. FRET efficiency as a function of distance. FRET efficiency scales as $(r/R_0)^{-6}$, where r is the distance between the donor and acceptor, and R_0 is the Förster distance, where FRET efficiency is 50%. This relationship results in a useful range of FRET techniques from approximately 0.5 to 1.5 times R_0.

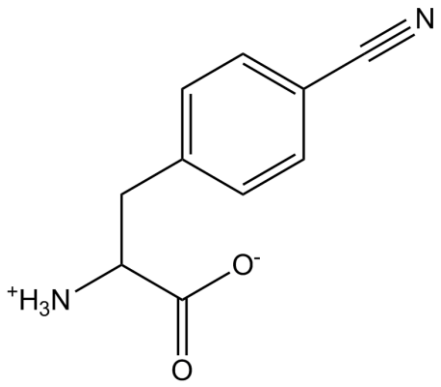


Figure 1.5. Structure of *p*-cyanophenylalanine. F_{CN} can function as a fluorescent probe of the local electrostatic environment.

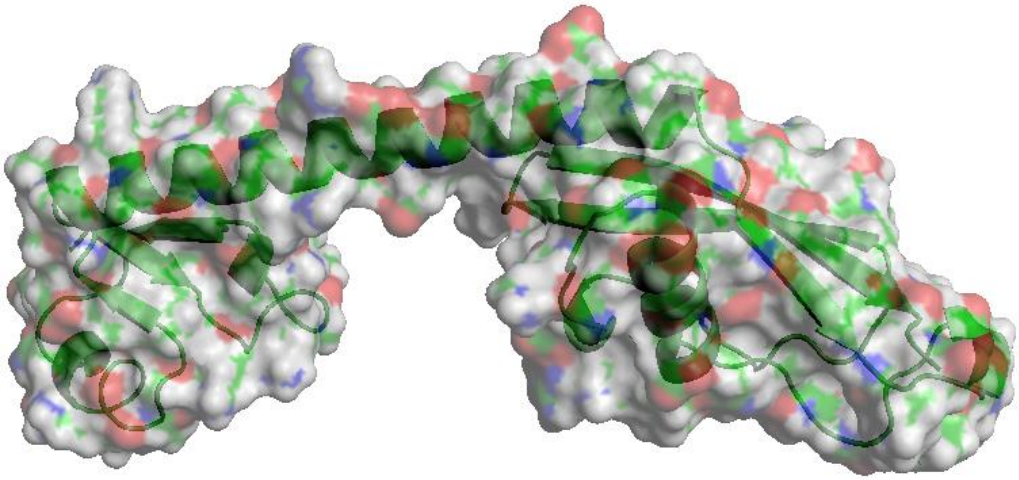
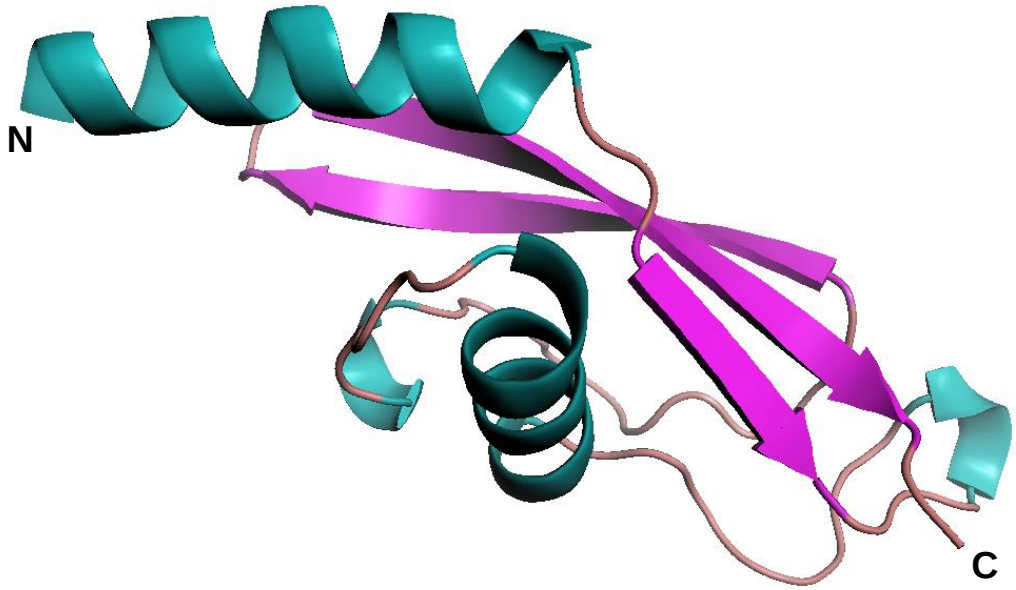


Figure 1.6. Structure of ribosomal protein L9. Cartoon representation of ribosomal protein L9 superimposed with the solvent accessible surface. A long 33-residue α-helix connects two globular domains NTL9 (left) and CTL9 (right). Rendered in PyMol from PDB structure 1DIV.[48]




```
58 60          70          80          90
AA EELANAKKLK EQLEKLTVTI PAKAGEGGRL FGSITSKQIA

   100         110         120         130
   ESLQAQHGLK LDKRKIELAD AIRALGYTNV PVKLHPEVTA

   140
   TLKVHVTEQK
```

Figure 1.7. Structure of CTL9. The N-terminus of the protein is the end of an α-helix and the C-terminal residue is on a dynamic region of the protein near the end of a β-strand. The 1° sequence of CTL9 is shown below. Rendered in PyMol from PDB structure 1DIV.[48]

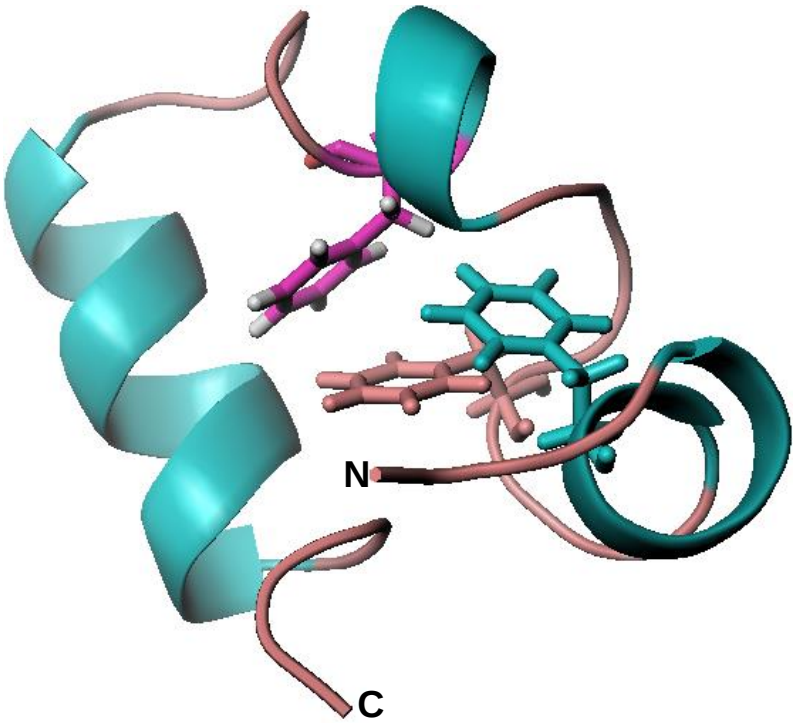


Figure 1.8. Structure of HP36. Three phenylalanine residues that make up the core of HP36. Phe47 is indicated in cyan, Phe51 in pink and Phe58 in magenta. The numbering system used here is based on full-length villin, so the N-terminal Met residue is M41. Rendered in PyMol from PDB structure 1VII.[49]

1.10. References

[1] Sohl, J. L., Jaswal, S. S., and Agard, D. A. (1998) Unfolded conformations of alpha-lytic protease are more stable than its native state, *Nature 395*, 817-819.

[2] Baker, D., Sohl, J. L., and Agard, D. A. (1992) A protein-folding reaction under kinetic control, *Nature 356*, 263-265.

[3] Anfinsen, C. B. (1973) Principles that govern the folding of protein chains, *Science 181*, 223-230.

[4] Levinthal, C. (1969) How to Fold Graciously, In *Proceedings of a meeting held at Allerton House* (J. T. P. Debrunnder, E. M., Ed.), pp 22-24, University of Illinois Press, Monticello, Illinois.

[5] Hart, T., Hosszu, L. L., Trevitt, C. R., Jackson, G. S., Waltho, J. P., Collinge, J., and Clarke, A. R. (2009) Folding kinetics of the human prion protein probed by temperature jump, *Proc Natl Acad Sci U S A 106*, 5651-5656.

[6] Ensign, D. L., Kasson, P. M., and Pande, V. S. (2007) Heterogeneity even at the speed limit of folding: large-scale molecular dynamics study of a fast-folding variant of the villin headpiece, *J Mol Biol 374*, 806-816.

[7] Kubelka, J., Chiu, T. K., Davies, D. R., Eaton, W. A., and Hofrichter, J. (2006) Sub-microsecond protein folding, *J Mol Biol 359*, 546-553.

[8] Dill, K. A., Bromberg, S., Yue, K., Fiebig, K. M., Yee, D. P., Thomas, P. D., and Chan, H. S. (1995) Principles of protein folding--a perspective from simple exact models, *Protein Sci 4*, 561-602.

[9] Magg, C., and Schmid, F. X. (2004) Rapid collapse precedes the fast two-state folding of the cold shock protein, *J Mol Biol 335*, 1309-1323.

[10] Ratner, V., Amir, D., Kahana, E., and Haas, E. (2005) Fast collapse but slow formation of secondary structure elements in the refolding transition of E. coli adenylate kinase, *J Mol Biol 352*, 683-699.

[11] Magg, C., Kubelka, J., Holtermann, G., Haas, E., and Schmid, F. X. (2006) Specificity of the initial collapse in the folding of the cold shock protein, *J Mol Biol 360*, 1067-1080.

[12] Arai, M., Kondrashkina, E., Kayatekin, C., Matthews, C. R., Iwakura, M., and Bilsel, O. (2007) Microsecond hydrophobic collapse in the folding of Escherichia coli dihydrofolate reductase, an alpha/beta-type protein, *J Mol Biol 368*, 219-229.

[13] Daggett, V., and Fersht, A. R. (2003) Is there a unifying mechanism for protein folding?, *Trends in Biochemical Sciences 28*, 18-25.

[14] Creighton, T. E. (1991) Stability of folded conformations, *Current Opinion in Structural Biology 1*, 5-16.

[15] Bennion, B. J., and Daggett, V. (2003) The molecular basis for the chemical denaturation of proteins by urea, *Proc Natl Acad Sci U S A 100*, 5142-5147.

[16] Schellman, J. A. (1987) Selective binding and solvent denaturation, *Biopolymers 26*, 549-559.

[17] Makhatadze, G. I., and Privalov, P. L. (1992) Protein interactions with urea and guanidinium chloride. A calorimetric study, *J Mol Biol 226*, 491-505.

[18] Courtenay, E. S., Capp, M. W., and Record, M. T., Jr. (2001) Thermodynamics of interactions of urea and guanidinium salts with protein surface: relationship between solute effects on protein processes and changes in water-accessible surface area, *Protein Sci 10*, 2485-2497.

[19] Felitsky, D. J., and Record, M. T., Jr. (2003) Thermal and urea-induced unfolding of the marginally stable lac repressor DNA-binding domain: a model system for analysis of solute effects on protein processes, *Biochemistry 42*, 2202-2217.

[20] Tobi, D., Elber, R., and Thirumalai, D. (2003) The dominant interaction between peptide and urea is electrostatic in nature: a molecular dynamics simulation study, *Biopolymers 68*, 359-369.

[21] Dill, K. A., Ghosh, K., and Schmit, J. D. (2011) Physical limits of cells and proteomes, *Proc Natl Acad Sci U S A 108*, 17876-17882.

[22] Zhang, Y., Kitazawa, S., Peran, I., Stenzoski, N., McCallum, S. A., Raleigh, D. P., and Royer, C. A. (2016) High Pressure ZZ-Exchange NMR Reveals Key Features of Protein Folding Transition States, *J Am Chem Soc 138*, 15260-15266.

[23] Lakowicz, J. R. (2006) *Principles of Fluorescence Spectroscopy*, 3 ed., Springer.

[24] Muralidharan, V., Cho, J., Trester-Zedlitz, M., Kowalik, L., Chait, B. T., Raleigh, D. P., and Muir, T. W. (2004) Domain-specific incorporation of noninvasive optical probes into recombinant proteins, *J Am Chem Soc 126*, 14004-14012.

[25] Miyake-Stoner, S. J., Miller, A. M., Hammill, J. T., Peeler, J. C., Hess, K. R., Mehl, R. A., and Brewer, S. H. (2009) Probing Protein Folding Using Site-Specifically Encoded Unnatural Amino Acids as FRET Donors with Tryptophan, *Biochemistry 48*, 5953-5962.

[26] Ye, S., Zaitseva, E., Caltabiano, G., Schertler, G. F., Sakmar, T. P., Deupi, X., and Vogel, R. (2010) Tracking G-protein-coupled receptor activation using genetically encoded infrared probes, *Nature 464*, 1386-1389.

[27] Dawson, P. E., Muir, T. W., Clark-Lewis, I., and Kent, S. B. (1994) Synthesis of proteins by native chemical ligation, *Science 266*, 776-779.

[28] W A Hendrickson, J. R. H., and D M LeMaster. (1990) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure., *EMBO J 9*, 1665-1672.

[29] Wang, L., Xie, J., and Schultz, P. G. (2006) Expanding the genetic code, *Annu Rev Biophys Biomol Struct 35*, 225-249.

[30] Tucker, M. J., Oyola, R., and Gai, F. (2006) A novel fluorescent probe for protein binding and folding studies: p-cyano-phenylalanine, *Biopolymers 83*, 571-576.

[31] Tucker, M. J., Oyola, R., and Gai, F. (2005) Conformational Distribution of a 14-Residue Peptide in Solution: A Fluorescence Resonance Energy Transfer Study, *The Journal of Physical Chemistry B 109*, 4788-4795.

[32] Tang, J., Yin, H., Qiu, J., Tucker, M. J., DeGrado, W. F., and Gai, F. (2009) Using two fluorescent probes to dissect the binding, insertion, and dimerization kinetics of a model membrane peptide, *J Am Chem Soc 131*, 3816-3817.

[33] Tucker, M. J., Tang, J., and Gai, F. (2006) Probing the kinetics of membrane-mediated helix folding, *J Phys Chem B 110*, 8105-8109.

[34] Marek, P., Gupta, R., and Raleigh, D. P. (2008) The fluorescent amino acid p-cyanophenylalanine provides an intrinsic probe of amyloid formation, *Chembiochem 9*, 1372-1374.

[35] Miyake-Stoner, S. J., Miller, A. M., Hammill, J. T., Peeler, J. C., Hess, K. R., Mehl, R. A., and Brewer, S. H. (2009) Probing protein folding using site-specifically encoded unnatural amino acids as FRET donors with tryptophan, *Biochemistry 48*, 5953-5962.

[36] Taskent-Sezgin, H., Chung, J., Patsalo, V., Miyake-Stoner, S. J., Miller, A. M., Brewer, S. H., Mehl, R. A., Green, D. F., Raleigh, D. P., and Carrico, I. (2009) Interpretation of p-cyanophenylalanine fluorescence in proteins in terms of solvent exposure and contribution of side-chain quenchers: a combined fluorescence, IR and molecular dynamics study, *Biochemistry 48*, 9040-9046.

[37] Serrano, A. L., Bilsel, O., and Gai, F. (2012) Native state conformational heterogeneity of HP35 revealed by time-resolved FRET, *J Phys Chem B 116*, 10631-10638.

[38] Weeks, C. L., Polishchuk, A., Getahun, Z., Degrado, W. F., and Spiro, T. G. (2008) Investigation of an unnatural amino acid for use as a resonance Raman probe: Detection limits, solvent and temperature dependence of the nuC identical withN band of 4-cyanophenylalanine, *J Raman Spectrosc 39*, 1606-1613.

[39] Liu, J., Strzalka, J., Tronin, A., Johansson, J. S., and Blasie, J. K. (2009) Mechanism of interaction between the general anesthetic halothane and a model ion channel protein, II: Fluorescence and vibrational spectroscopy using a cyanophenylalanine probe, *Biophys J 96*, 4176-4187.

[40] Fafarman, A. T., and Boxer, S. G. (2010) Nitrile bonds as infrared probes of electrostatics in ribonuclease S, *J Phys Chem B 114*, 13536-13544.

[41] Bagchi, S., Boxer, S. G., and Fayer, M. D. (2012) Ribonuclease S dynamics measured using a nitrile label with 2D IR vibrational echo spectroscopy, *J Phys Chem B 116*, 4034-4042.

[42] Naganathan, A., Wood, M. P., and Moore, S. D. (2015) The large ribosomal subunit protein L9 enables the growth of EF-P deficient cells and enhances small subunit maturation, *Plos One 10*, e0120060.

[43] Sato, S., Kuhlman, B., Wu, W. J., and Raleigh, D. P. (1999) Folding of the multidomain ribosomal protein L9: the two domains fold independently with remarkably different rates, *Biochemistry 38*, 5643-5650.

[44] Sato, S., and Raleigh, D. P. (2002) pH-dependent stability and folding kinetics of a protein with an unusual alpha-beta topology: the C-terminal domain of the ribosomal protein L9, *J Mol Biol 318*, 571-582.

[45] Li, Y., Shan, B., and Raleigh, D. P. (2007) The cold denatured state is compact but expands at low temperatures: hydrodynamic properties of the cold denatured state of the C-terminal domain of L9, *J Mol Biol 368*, 256-262.

[46] Van Horn, W. D., Simorellis, A. K., and Flynn, P. F. (2005) Low-temperature studies of encapsulated proteins, *J Am Chem Soc 127*, 13553-13560.

[47] Dill, K. A., and Chan, H. S. (1997) From Levinthal to pathways to funnels, *Nat Struct Biol 4*, 10-19.

[48] Hoffman, D. W., Davies, C., Gerchman, S. E., Kycia, J. H., Porter, S. J., White, S. W., and Ramakrishnan, V. (1994) Crystal structure of prokaryotic ribosomal protein L9: a bi-lobed RNA-binding protein, *EMBO J 13*, 205-212.

[49] McKnight, C. J., Matsudaira, P. T., and Kim, P. S. (1997) NMR structure of the 35-residue villin headpiece subdomain, *Nat Struct Biol 4*, 180-184.

2. Selenomethionine Quenching of Tryptophan Fluorescence Provides a Simple Probe of Protein Structure

2.1. Abstract

Fluorescence spectroscopy, relying on intrinsic protein fluorophores, is one of the most widely used methods for studying protein folding, protein-ligand interactions, and protein dynamics. Tryptophan is usually the fluorophore of choice, given its sensitivity to environment and having the highest quantum yield of the natural amino acids, however changes in tryptophan florescence can be difficult to interpret in terms of specific structural changes. The introduction of quenchers of tryptophan fluorescence can provide information about specific structures, particularly if quenching is short range, however the most commonly employed quencher is histidine, and it is only effective when the imidazole sidechain is protonated, thus limiting the pH range over which this approach can be employed. In addition, histidine is not always a conservative substitution and is likely to be destabilizing if inserted into the hydrophobic core of proteins. Here we illustrate the use of a Trp-selenomethionine (M_{Se}) pair as a specific probe of protein structure. M_{Se} requires close approach to Trp to quench its fluorescence, and this effect can be exploited to design specific probes of α-helix and β-sheet formation. The approach is illustrated using equilibrium and time-resolved fluorescence measurements on designed peptides and globular proteins. M_{Se} is easily incorporated into proteins, provides a conservative replacement for hydrophobic sidechains, and M_{Se} quenching of Trp fluorescence is pH independent. The oxidized form of $M_{Se,}$ selenomethionine selenoxide, is also an efficient quencher of Trp fluorescence.

Note: The material presented in this chapter has been published (Watson, M. D., Peran, I., Zou, J., Bilsel, O. and Raleigh, D. P. (2017) Selenomethionine Quenching of Tryptophan Fluorescence Provides a Simple Probe of Protein Structure *Biochemistry 56*, 1085-1094). This chapter contains direct excerpts from the manuscript that was written by me with additional writing and revisions by the other authors. Junjie Zou performed the molecular dynamics simulations.

2.2. Introduction

Fluorescence spectroscopy is one of the most accessible and widely used methods for studies of proteins. High sensitivity and time resolution make fluorescence methods attractive for studies of protein structure, dynamics, stability and aggregation.[1, 2] Naturally occurring tryptophan residues offer the simplest sensitive, non-perturbative intrinsic fluorophore. In addition to a shift in emission maximum upon exclusion from solvent, Trp undergoes changes in quantum yield, although the molecular basis of these changes can be difficult to interpret as they are determined by a number of factors. One approach towards simplifying Trp fluorescence experiments is introduction of a quencher.[3] Of the naturally occurring amino acids, His is probably the most commonly employed potent quencher of Trp fluorescence. However, the quenching mechanism of the Trp-His pair requires a protonated imidazole ring, restricting the utility of this system to pH ranges below the pK_a of His. Incorporation of a His residue in the interior of a protein can also be destabilizing, especially if it is protonated. A growing body of work has made use of the unnatural amino acid *p*-cyanophenylalanine (F_{CN}), the cyano analogue of Tyr, as an alternative to Trp.[4-9] His can also be used to quench F_{CN} fluorescence, but in this case is only effective above the pK_a of the imidazole ring, unlike in the Trp-His pair.[10]

Selenomethionine (M_{Se}), the selenium analogue of Met, is an effective pH independent quencher of F_{CN} fluorescence, and earlier work suggests that it may also quench Trp fluorescence.[11-14] The F_{CN} quenching effect occurs through an electron transfer mechanism which requires van der Waals contact between the side chains, making the approach complementary to related methods such as Förster Resonance Energy Transfer (FRET) which can be used to probe the distance between two fluorophores, but cannot definitively confirm whether they are in direct contact.[11, 15]

Although the F_{CN}-M_{Se} pair is a useful probe of structure, incorporation of the pair into an expressed protein does present some technical challenges. Methods for the incorporation of F_{CN} using so-called 21st pair technology through the use of an unnatural aminoacyl-tRNA synthetase/tRNA pair have been developed.[7, 8] There is an extensive body of work on the incorporation of M_{Se} in high yield by Met auxotrophic cell lines for multi-wavelength anomalous diffraction (MAD) phasing in X-ray crystallography and ^{77}Se NMR.[16, 17] Simultaneous incorporation of both unnatural amino acids into a single protein has not yet been demonstrated but is possible, however it could result in low protein yields and could present a significant challenge. A more straightforward approach is to exploit the natural fluorophore Trp and to use well documented methods for the incorporation of M_{Se}. The unbranched alkyl side chain of M_{Se} is likely to be less perturbative at many sites in a protein than the polar aromatic ring of His, suggesting that Trp-M_{Se} pairs could provide a widely applicable protein of protein structure.

Here we demonstrate the use of M_{Se} quenching of Trp fluorescence to probe protein structure and to monitor protein unfolding. The fluorophore-quencher pair is incorporated recombinantly into a β-sheet in the C-terminal domain of ribosomal protein L9 (CTL9). The Trp-M_{Se} pair is also studied in two α-helical systems prepared by solid phase peptide synthesis: a synthetic designed 21-residue helical peptide and a 36-residue segment of the villin headpiece helical subdomain (HP36). The results indicate that Trp-M_{Se} pairs can be used to probe local conformational changes, including the formation of specific α-helices and β-sheets at both high and low pH. The single oxidation product of M_{Se}, selenomethionine selenoxide (M_{SeO}), is also shown to be an effective quencher of Trp fluorescence.

2.3. Materials and Methods

2.3.1. Peptide Synthesis

The 21-residue helical peptide and HP36 were synthesized using standard 9-fluorenylmethyloxycarbonyl (Fmoc) chemistry on a CEM Liberty microwave peptide synthesizer. Use of 5-(4'-Fmoc-aminomethyl-3',5-dimethoxyphenol)valeric acid (Fmoc-PAL-PEG-PS) resin afforded an amidated C-terminus on both peptides, and acetic anhydride was used to acetylate the N-terminus of the helical peptide. β-branched residues, prolines, arginines and the C-terminal residue were double coupled. Peptides were deprotected and cleaved from resin in trifluoroacetic acid (TFA) with 5% thioanisole, 3.3% anisole and 3% 1,2-ethanedithiol as scavengers. Resin was removed by filtration and isolated peptides were precipitated in cold diethyl ether and pelleted by centrifugation at 10,000 rcf for 10 min. The supernatant was discarded; peptides were solubilized in 20% (*v*/*v*) acetic acid and lyophilized.

2.3.2. Protein Expression

Mutants of CTL9 containing M_{Se} were expressed in M15MA *E.coli* cells carrying the CTL9 gene on a pQE-80L vector. A pREP4 plasmid was also present to repress "leaky" expression before induction. CTL9 mutants containing M_{Se} were expressed in M9 minimal media. An overnight culture of M15MA-pQE-80L-Y126W/H144M_{Se}-CTL9 was grown in LB rich media treated with ampicillin (200 μg/mL) and kanamycin (35 μg/mL) and added to 0.5 L of M9 minimal media supplemented with the 20 natural amino acids (40 mg/L) and treated with ampicillin (200 μg/mL) and kanamycin (35 μg/mL). Cells were grown at 37 °C to an optical density at 600 nm of 0.8-1, harvested by centrifugation at 5,000 rcf for 15 min and resuspended in 0.5 L of fresh M9 buffer solution treated with antibiotics. Cells were harvested a second time by centrifugation at 5,000 rcf for 15 min, resuspended in 0.5 L of fresh M9 minimal media supplemented with 19 of

the natural amino acids (except methionine) and grown at 37 °C for 20 min. M_{Se} (40 mg/L) was added to the media, cells were grown at 37 °C for an additional 20 min and protein expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. After 4 hr cells were harvested by centrifugation at 5,000 rcf for 15 min. CTL9 mutants without M_{Se} were expressed in LB rich media using standard methods.

2.3.3. Protein and Peptide Purification

Harvested cells were lysed by sonication and protein partially purified from the cell lysate by cation exchange chromatography on a GE Sepharose Fast Flow column. Proteins were purified from the eluate by reversed-phase high performance liquid chromatography (HPLC) on a Higgins Analytical Proto 300 C18 preparative column. Synthetic peptides were purified by HPLC using the same column. A two-buffer A-B gradient system was used where buffer A was 0.1% (*v*/*v*) TFA in water and buffer B was 0.1% (*v*/*v*) TFA in 9:1 acetonitrile:water. All peptides and proteins were eluted with a linear gradient of 20-60% B in 40 min.

2.3.4. Protein Oxidation

Purified proteins were dissolved in 18 MΩ H_2O to a concentration of 1 mg/mL. The solution was acidified by addition of $HClO_4$ to a final concentration of 0.2 mM. Oxidation was initiated by addition of H_2O_2 to a final concentration of 0.005%. Trial experiments, monitored by HPLC indicated that after 4 hours >90% of the protein was in the form of the single oxidation product, selenomethionine selenoxide. No multiple oxidation products were detected. Thus, the reaction was allowed to proceed for 4 hr before dilution with 0.1% TFA in H_2O and the oxidized protein was purified by HPLC.

2.3.5. Mass Spectrometry

Peptides and proteins were characterized by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOFMS) on a Bruker Daltonics autoflex TOF/TOF instrument. 21-residue helical peptide expected mass: 2014.987 Da (mono) observed: 2015.687 Da (mono). N68M_{Se}-HP36 expected mass: 4252.141 Da (mono) observed: 4252.315 Da (mono). N68M-HP36 expected mass: 4204.197 Da (mono) observed: 4207.8 Da (mono). Y126W-CTL9 expected mass: 10004.51 Da (average) observed: 10000.5 Da (average). Y126W/H144M_{Se}-CTL9 expected mass: 10045.46 Da (average) observed: 10045.509 Da (average). Y126W/H144M-CTL9 expected mass: 9998.55 Da (average) observed: 9998.3 Da (average).

Oxidized proteins were characterized by liquid chromatography electrospray ionization time-of-flight mass spectrometry (LC-ESI-TOFMS) on an Agilent G6224a oaTOF instrument. Oxidation products purified by HPLC were further separated on a Jupiter C18-MW-22.m column by a sequence of increasingly steep linear gradients from 0.1% acetic acid in water to 0.1% acetic acid in 9:1 acetonitrile:isopropanol and ionized by electrospray ionization (ESI). Y126W/H144M-CTL9 expected mass: 9998.69 Da (average) observed: 9999.09 Da (average). Y126W/H144M_{ox}-CTL9 expected mass: 10014.69 Da (average) observed: 10014.99 Da (average). Y126W/H144M_{Se}-CTL9 expected mass: 10045.46 Da (average) observed: 10046.01 Da (average). Y126W/H144M_{SeO}-CTL9 expected mass: 10061.46 Da (average) observed: 10062.00 Da (average).

2.3.6. Circular Dichroism Spectroscopy

Circular dichroism (CD) measurements were performed on an Applied Photophysics Chirascan spectrometer. CD wavelength scans were recorded in a 1 mm pathlength quartz cuvette with a thermostated sample holder at 20 °C from 190 to 260 nm in buffer and from 210 to 260 nm

in urea using a 1 nm stepsize. Spectra were recorded as the average of three scans with an averaging time of 0.5 s.

2.3.7. Estimation of Helical Content

The percent of helical structure in the 21-residue helical peptide was calculated from the molar ellipticity at 222 nm, $[\theta]_{222}$ using equation (2.1)

$$f_h = \frac{[\theta]_{222} - [\theta]_C}{[\theta]_H - [\theta]_C} \tag{2.1}$$

Where f_h is the fractional helical content of the sample, $[\theta]_H$ is the molar ellipticity at 222 nm for a 100% helical peptide determined from equation (2.2) and $[\theta]_C$ is the molar ellipticity at 222 nm for a random coil determined from equation (2.3)

$$[\theta]_H = -40000\left(1 - \frac{2.5}{N}\right) + 100T \tag{2.2}$$

$$[\theta]_C = 640 - 45T \tag{2.3}$$

Where N is the number of residues in the peptide and T is the temperature in °C.[18]

2.3.8. Circular Dichroism Monitored Thermal Unfolding

Thermal denaturation curves were acquired on an Applied Photophysics Chrirascan circular dichroism spectrometer. Ellipticity was measured at 222 nm using a 10 mm pathlength quartz cuvette. The temperature was increased from 2 to 94 °C in 2 °C steps with a 120 s equilibration time at each temperature and a 60 s averaging time. Data was fit to equation (2.4)

$$\theta_{222} = \frac{a_n + b_n T + (a_d + b_d T)e^{\left(-\frac{\Delta G(T)}{RT}\right)}}{1 + e^{\left(-\frac{\Delta G(T)}{RT}\right)}} \tag{2.4}$$

Where θ_{222} is the ellipticity at 222 nm in millidegrees, a_n, b_n, a_d and b_d are fitting parameters defining the pre- and post-transition baseline, T is the temperature in K, R is the gas constant in kcal mol^{-1} K^{-1}, and $\Delta G(T)$ is the change in free energy upon unfolding in kcal mol^{-1}, determined from the modified Gibbs-Helmholtz equation (2.5)

$$\Delta G(T) = \Delta H_M \left(1 - \frac{T}{T_M}\right) + \Delta C_p^\circ \left(T - T_M - T \ln\left(\frac{T}{T_M}\right)\right) \quad (2.5)$$

Where $\Delta G(T)$ is the change in free energy upon unfolding in kcal mol^{-1}, ΔH_M is the change in enthalpy upon unfolding at T_M in kcal mol^{-1}, T is the temperature in K, T_M is the midpoint of the thermal unfolding transition in K and ΔC_p° is the change in heat capacity upon unfolding in kcal mol^{-1} K^{-1}. ΔC_p° was set to 1.07 kcal mol^{-1} K^{-1} for all CTL9 constructs and 0.38 kcal mol^{-1} K^{-1} for all HP36 constructs.

2.3.9. Equilibrium Fluorescence

Fluorescence emission spectra were recorded on a Photon Technologies International fluorimeter in a 10 mm x 10 mm quartz cuvette using a slit width of 0.8 mm. Using an excitation wavelength of 280 nm, spectra were recorded from 290 to 450 nm with a step size of 1 nm and an averaging time of 1 s.

2.3.10. Time-resolved Fluorescence

Time-resolved fluorescence experiments were performed at the University of Massachusetts Medical School in Worcester, MA. Decays were measured on a home-built time correlated single photon counting (TCSPC) apparatus built by Dr. Osman Bilsel. The tripled output of a 10 W Verdi (Coherent) pumped Ti:sapphire laser (Coherent Mira) was used to excite Trp at 292 nm. The repetition rate was reduced to 3.8 MHz. The detection utilized a bandpass filter (FF01-357/44, Semrock, Rochester, NY) and Glan-Taylor polarizer at the magic angle. A PMH-100-6 photomultiplier tube connected to an SPC150 photon counting card (Becker-Hickl, Berlin,

Germany) was used for TCSPC. The instrument response was collected using scattered light from a solution of distilled deionized water using a bandpass filter that partially overlaps the excitation wavelength. The peak counts in the instrument response were approximately 30,000 counts. The instrument response exhibits a full width at half-maximum of approximately 200 ps and is primarily determined by the response of the photomultiplier tube. All measurements were collected at 20 ±1 °C. The decay curves were fit using a maximum entropy model (MEM) with reconvolution of the instrument response.[19]

2.3.11. Molecular Dynamics Simulations

The simulations were performed by Junjie Zou and are included in this chapter for completeness and because they aid in the interpretation of the experimental results. The starting structures for HP36 and CTL9 were constructed using the pdb files 1YRF[20] (Note that Chiu *et al*. incorporated an N68H mutation as a fluorescence quencher) and 1DIV[21], respectively. A single Met residue was added to the N-terminus of the structure of HP35 (PDB: 1YRF) to generate HP36, and the mutation H68M-HP36 was made by using Swiss PDB.[22] The PDB file 1DIV is for full length L9 and residues 1 to 57 were deleted to generate CTL9. Tyr 126 was changed to Trp and His 144 was changed to Met. These modified structures were equilibrated by molecular dynamics (MD) simulations with restraints on unmodified residues. Simulations were performed using the Amber software package with the Amber ff14SB force field and TIP3P water.[23, 24] The step size was set to 2 fs. Truncated octahedron boxes with periodic boundary condition were used. The cutoff of non-bonded interactions was set to 8 Å. No ions were included in the simulation. Particle mesh Ewald methods were used to evaluate electrostatic energies.[25] Hydrogen atoms were constrained using the SHAKE algorithm.[26] The temperature was set to 298 K by using a weak-coupling algorithm with the coupling constant set to 1 ps.[27] A constant pressure of 101,325 pascal

controlled by the Berendsen barostat was used.[27] The program Cpptraj was used to calculate distances, dihedral angles and the solvent accessible surface area (SASA).[28]

2.4. Results and Discussion

2.4.1. Tryptophan Selenomethionine Pairs Provide a Fluorescence Probe of β-sheet Formation

Trp and M_{Se} residues were recombinantly incorporated into the C-terminal domain of ribosomal protein L9 (CTL9) in order to examine the utility of the Trp-M_{Se} pair to monitor folding in globular proteins. The N-terminal M_{Se} residue is efficiently cleaved and no product with an additional M_{Se} residue was observed based on MALDI-TOFMS data. The numbering system used here is based upon full length L9 and designates the first residue in CTL9 as residue 58. This 92 residue, mixed α-β protein possesses naturally occurring Tyr and His residues at positions 126 and 144 on the second and third β-strands, respectively. The residues are located in an antiparallel β-sheet with the side chains in van der Waals contact. Three mutants were prepared, one with Tyr 126 mutated to Trp (Y126W-CTL9), another with Tyr 126 and His 144 mutated to Trp and M_{Se}, respectively (Y126W/H144M_{Se}-CTL9), and a third with Tyr 126 and His 144 mutated to Trp and Met, respectively (Y126W/H144M-CTL9) (Figure 2.1A). Circular dichroism (CD) confirmed that all mutants are folded in buffer and fully unfolded in buffer with 9.5 M urea (Figure 2.2). CD monitored thermal denaturation indicated small T_M differences between mutants, ranging from 79.5 °C (Y126W-CTL9), 75.6 °C (Y126W/H144M_{Se}-CTL9), to 74.9 °C (Y126W/H144M-CTL9) (Figure 2.3).

Trp fluorescence is high in the unfolded state at pH 7.5 (20 mM tris, 9.5 M urea) for Y126W/H144M_{Se}-CTL9, and 66% quenched in the folded state (Figure 2.1B). In contrast, the fluorescence intensity of Y126W-CTL9 and Y126W/H144M-CTL9 is higher in the folded state than in the urea unfolded state (9.5 M urea) at pH 7.5 (Figure 2.1C, D). This effect is likely due to

partial burial of Trp 126 in the native state. The solvent accessible surface area (SASA) of Trp 126 in the folded state of Y126W/H144M-CTL9 calculated using trajectories obtained from MD simulations is 106.0 $Å^2$. The SASA of Trp in a reference state corresponding to a fully extended penta-peptide with the sequence from CTL9 (Leu-Gly-Trp-Thr-Asn) is 144.0 $Å^2$. Histidine is known to be a quencher of Trp fluorescence, however the quenching mechanism requires the imidazole ring to be in the protonated state. As expected, significantly enhanced quenching—90% relative to the unfolded state—is observed for Y126W-CTL9 at pH 5.0 where His 144 is protonated (Figure 2.4A). In contrast, no fluorescence quenching was observed for Y126W/H144M-CTL9 at pH 7.5 or 5.0 (Figure 2.4E, F). The pH dependent data illustrates the practical limitations of His as a quencher of Trp fluorescence. In contrast, Trp fluorescence was efficiently quenched by M_{Se} at both pH values. The fluorescence quenching efficiency of M_{Se} was 66% at pH 5.0 and 7.5 relative to the respective unfolded states (Figure 2.4C, D). The pH dependent studies illustrate the advantage of Trp-M_{Se} over Trp-His pairs, namely that quenching is pH independent which in turn makes the interpretation of the data more straightforward.

The dynamic quenching effects in Y126W/H144M_{Se}-CTL9 were examined by conducting fluorescence lifetime measurements in 20 mM tris buffer at pH 7.5 and in the presence of 9.5 M urea under the same conditions (Figure 2.5). Maximum entropy model (MEM) analysis of the data indicates a single lifetime distribution centered around 3 ns in the urea unfolded state and two lifetime distributions centered around and 3 ns and 0.1 ns in the folded state. The two lifetime distributions have similar intensities in the folded state, though the intensity of the 0.1 ns lifetime distribution is slightly higher. Based on the known stability of CTL9, the 3 ns lifetime distribution in the folded state cannot represent unfolded protein; the population of folded molecules is estimated to be in excess of 99.4% under these conditions. Instead this population likely represents

rotameric states of Trp which are not within van der Waals contact of the Se atom and therefore do not experience quenching. The existence of such states is supported by modeling of the different rotameric states of Trp (Figure 2.6).

Long time MD simulations of Y126W/H144M-CTL9 were conducted to test this hypothesis. Since M_{Se} is not parametrized in the Amber force field, Met was used in the simulations. A total of four independent simulations were carried out. In the first and second runs, W126 starts in the rotamer state of χ_1 = -65.0°, χ_2 = -84.9°, which is the rotamer state found in the crystal structure (1DIV). The third run starts with a rotamer state of χ_1 = -71.2°, χ_2 = 94.3°, and the fourth run starts with a rotamer state of χ_1 = 62.1°, χ_2 = 94.4°. The distances between W126 and M144 were measured using the geometric center of Cδ2 and Cε2 of W126, and the sulfur atom of Met for each frame of the simulations (Figure 2.7). The χ_1 and χ_2 dihedral angles were calculated for the first run (Figure 2.8) and show a clear correlation between the rotameric state of W126 and the distance between W126 and M144 (Figure 2.7A). Three clusters of distances, which are centered around 4 Å, 5 Å and 7 Å are observed when the distances between W126 and M144 from all four MD simulations are plotted as a histogram (Figure 2.9). Over the course of a 3 ns simulation with a sampling frequency of 0.002 ns, the distances jumped multiple times between 4 Å and 5 Å (Figure 2.10) which indicates that distances at 4 Å and 5 Å are in rapid exchange; this likely accounts for the experimental lifetime distribution centered around 0.1 ns. The longer lifetime of 3 ns likely corresponds to a distance of 7 Å between W126 and M_{Se}144. The absence of a short lifetime distribution for CTL9 in the presence of 9.5 M urea indicates that Trp does not experience significant quenching under these conditions as the β-sheet is not formed.

2.4.2. Oxidation of Selenomethionine to the Selenoxide Does Not Abolish Quenching

Selenium is more susceptible to oxidation than sulfur, but it is not immediately clear what effect this will have on the quenching efficiency of M_{Se}. Selenomethionine, like methionine, can undergo oxidation to selenomethionine selenoxide (M_{SeO}) and further oxidation to selenomethionine selenone (Figure 2.11A). No spontaneous oxidation of Y126W/H144M_{Se}-CTL9 was detected by LC-ESI-TOFMS following protein expression and purification. To examine the differences in oxidation susceptibility between Met and M_{Se}, Y126W/H144M-CTL9 and Y126W/H144M_{Se}-CTL9 were exposed to oxidizing conditions. Exposure to 0.005% H_2O_2 acidified with 0.2 mM $HClO_4$ for 4 hours was found to oxidize ~90% of Y126W/H144M_{Se}-CTL9 to Y126W/H144M_{SeO}-CTL9, while only oxidizing ~10% of Y126W/H144M-CTL9 based on HPLC monitored by UV absorbance at 220 nm (Figure 2.12). Analysis by LC-ESI-TOFMS found no evidence of oxidation of either protein to the selenone or sulfone. CD confirmed that Y126W/H144M_{SeO}-CTL9 was properly folded in buffer and fully unfolded in buffer with 9.5 M urea (Figure 2.11B). CD monitored thermal denaturation indicated a small change in T_M from 75.6 °C (Y126W/H144M_{Se}-CTL9), to 72.5 °C (Y126W/H144M_{SeO}-CTL9) (Figure 2.3).

After correcting for concentration, the Trp fluorescence of Y126W/H144M_{SeO}-CTL9 was identical to that of Y126W/H144M_{Se}-CTL9 in 9.5 M urea, 20 mM tris buffer at pH 7.4. The quenching efficiency of Y126W/H144M_{SeO}-CTL9 upon folding in 20 mM tris buffer at pH 7.4 was 84%, compared to 66% for Y126W/H144M_{Se}-CTL9 under the same conditions (Figure 2.11C, D). The only moderately higher susceptibility of M_{Se} to oxidation compared to Met, together with the modest differences in quenching efficiency of M_{Se} *vs* M_{SeO} indicate that oxidation is unlikely to be a significant source of error in applications of M_{Se} as a fluorescence quencher, provided oxidation of M_{Se} does not perturb the structure of the protein

2.4.3. Tryptophan Selenomethionine Pairs Can be Used to Follow α-helix Formation in Globular Proteins

To investigate the utility of the Trp-M_{Se} pair as a probe of local α-helical structure in globular proteins, a mutant of the 36-residue villin headpiece helical subdomain (HP36) was prepared. HP36 is a small, autonomously folding three-helix protein that has been used widely as a model system for both computational and experimental studies of protein folding.[20, 29-40] HP36 possesses a naturally occurring Trp residue at position 24. We mutated Asn 28 to M_{Se} to create an *i*, *i*+4 arrangement that places the side chains roughly one turn apart, bringing them into close contact in the helical state. Thus we expect Trp fluorescence to be higher when the peptide is unfolded and lower when it adopts a helical conformation and is quenched by M_{Se} (Figure 2.13A). HP36 is a subdomain of villin, but the numbering system used here designates the first residue of HP36 as residue 1. This construct was designated N28M_{Se}-HP36. CD experiments in native buffer (20 mM acetate at pH 5.0) show that the substitution, as expected, does not perturb the secondary structure. The experimental CD spectrum is very similar to that of wild type HP36 reported in the literature.[33] The spectrum observed for N28M_{Se}-HP36 in buffer with 9.5 M urea indicates that the protein is fully unfolded under these conditions (Figure 2.13B). CD monitored thermal unfolding indicates that the thermal stabilities of both proteins in 10 mM acetate buffer at pH 5.0 are very similar, with T_M values of 71.7 °C (N28M-HP36) and 70.8 °C (N28M_{Se}-HP36) (Figure 2.14). This is not significantly perturbed from the reported T_M of 70.5 °C for the wild type protein in 20 mM phosphate buffer at pH 7.0.[41]

The Trp quenching efficiency of M_{Se} in HP36 was first examined by comparing the equilibrium fluorescence intensity of folded N28M_{Se}-HP36 in 20 mM acetate buffer at pH 5.0 to the fluorescence intensity of the unfolded protein in buffer with 9.5 M urea (Figure 2.13C). Trp

fluorescence is 60% quenched in the folded state relative to the unfolded state. Surprisingly, the quenching efficiency of N28M-HP36 was very similar (Figure 2.13D). The data indicates that, in this system, M_{Se} offers little improvement over Met. Possible explanations include a closer approach of the Met side chain, which is not considered likely, or quenching of Trp fluorescence by other groups in both proteins. There are a number of Lys residues, K25, K30 and K31 near W24 in the primary sequence. MD simulations indicate that the epsilon amino group of K25 makes a close approach to the indole ring of W24; the mean distance from the nitrogen of the K25 sidechain to the geometric center of Cδ2 and Cε2 of the Trp is 6.9 ±1.1 Å during the simulation (Figure 2.15), which is expected to contribute to the quenching of Trp fluorescence. This case highlights the importance of choosing alternative fluorophores in some systems, such as 4-cyanophenylalanine, which is not quenched by Lys in the same manner as Trp. N28M_{Se}-HP36 appears to experience greater fluorescence quenching in the unfolded state than N28M-HP36 (Figure 2.13C, D). This might arise from non-helical interactions of the M_{Se} and Trp residues due to their proximity in primary sequence, transient helical structure even in the presence of urea, or some combination of these effects. Such interactions should also be present in the unfolded state of N28M-HP36 and the difference in quenching between the two proteins could be due to more effective quenching by M_{Se} or by small changes in the unfolded ensemble.

The dynamic quenching effects of M_{Se} on Trp in HP36 were examined by conducting fluorescence lifetime measurements in 20 mM acetate buffer at pH 5.0 and in the presence of 9.5 M urea in the same buffer (Figure 2.16). MEM analysis of the data indicates two lifetime distributions centered around 4 ns and 1 ns in both the folded state and the urea unfolded state. The 4 ns lifetime distribution is the major component of the decay in the urea unfolded state while the 1 ns lifetime distribution is the major component in the folded state.

The presence of a short lifetime component in the urea unfolded state supports the conclusion from the equilibrium fluorescence data that the Trp residue still experiences some quenching under these conditions. The presence of a long lifetime component in the absence of urea could arise from contributions from a small unfolded population or rotameric states of the Trp sidechain that are not within van der Waals contact with the Se atom even in the helical conformation. Based on the known stability of HP36, the small, long-lifetime population in the absence of urea most likely represents an unquenched rotameric state of Trp rather than a non-helical population (Figure 2.17).

2.4.4. Tryptophan Fluorescence Quenching by Selenomethionine Probes α-helix Formation in a Designed Peptide

We next tested the utility of the approach using a designed 21-residue helical peptide (Figure 2.18A). The studies with the peptide also provide additional clues as to the origin of the short and long lifetime components. The synthetic peptide included Ala residues to induce helical structure, Lys residues to improve solubility, and an Asp-Pro pair at the N-terminus that acts as an initiator of helical structure. The N and C-termini were acetylated and amidated, respectively, to increase the stability of the helical state. A Trp residue was incorporated at position 12 and M_{Se} at position 16 to produce the same *i*, *i*+4 configuration used in HP36 (Figure 2.18A). Helical structure was confirmed by CD (Figure 2.18B). Data indicates significant helical structure in 10 mM acetate buffer at pH 5.5, indicated by a local minimum at 222 nm and a maximum at 193 nm. The helical content of the peptide in buffer was estimated to be 38% based on the molar ellipticity at 222 nm (Equation 2.1). In contrast, the spectrum observed in the same buffer with 8 M urea indicates that the peptide is unstructured under these conditions. The ability of M_{Se} to quench Trp fluorescence was examined by comparing the fluorescence intensity of the peptide in buffer and in the presence

of denaturant. Fluorescence is high in the unfolded state (8 M urea) and is 35% quenched under folding conditions (10 mM acetate buffer at pH 5.5) (Figure 2.18C). The smaller magnitude of the effect compared to the 60% quenching efficiency observed for N28M$_{Se}$-HP36 can be attributed to the fact that the peptide is not fully folded in the absence of urea while N28M$_{Se}$-HP36 is. Consequently, the quenching observed for the 21-residue helical peptide represents Trp quenching in only a fraction of molecules in the sample whereas nearly all Trp residues are quenched in a sample of N28M$_{Se}$-HP36 under folding conditions.

The effect of M$_{Se}$ quenching on the Trp fluorescence lifetime was examined by conducting fluorescence lifetime measurements in 10 mM acetate buffer at pH 5.5 and in the presence of 8.0 M urea under the same conditions (Figure 2.19). MEM analysis indicates the presence of two lifetime distributions centered around 3 ns and 1 ns in both the folded state and in the urea unfolded state. The lifetimes of the individual components are similar in urea and in buffer, but the relative intensities are not. Whereas the decay is dominated by the 3 ns lifetime distribution in the urea unfolded state, the 1 ns lifetime distribution is the major component in the folded state.

The results of the MEM analysis of the 21-residue helical protein are broadly similar to the results of the MEM analysis of N28M$_{Se}$-HP36. The most significant difference is that the relative intensity of the short lifetime distribution in the HP36 sample is higher in both the presence and absence of urea compared to the relative intensity of the short lifetime distribution in the helical peptide sample under the same conditions. This effect is more pronounced in the folded state than in the urea unfolded state. This suggests that conformations in which Trp fluorescence is quenched are more highly populated in N28M$_{Se}$-HP36 than in the helical peptide, which is consistent with CD data that indicates higher helicity in the folded state of HP36 than in the helical peptide, as well as equilibrium fluorescence data which shows a higher Trp quenching efficiency in N28M$_{Se}$-

HP36 than in the helical peptide. The relative intensity of the short lifetime distribution is higher in buffer than what might be expected for a peptide which is only 38% helical based on the CD data. However, CD is a global probe of structure and the effect could be due to higher local helicity in the vicinity of the Trp-M_{Se} pair, ie. the ends of the helix are frayed while the core of the helix is more structured.

2.5. Conclusions

The data presented here illustrates the utility of Trp-M_{Se} pairs as non-perturbing fluorescence probes of protein structure. Advantages of the approach include the ease of incorporation of M_{Se}, the conservative nature of the substitution of Leu, Ile, Met and Val by M_{Se}, the pH independence of the observed quenching and the short range nature of the quenching, which facilitates the interpretation of fluorescence changes in terms of specific structural changes, including local conformation differences in the folded state, as exemplified by the sensitivity of the lifetime distributions to the rotameric state of Trp 126 in CTL9. Potential issues include the fact that M_{Se} may not always be a conservative replacement for a surface residue owing to its hydrophobicity. Oxidation of M_{Se} to the selenoxide does not affect quenching, but the increased bulk caused by the additional oxygen atom could make the substitution less conservative in cases where packing is important. Trp-His pairs have been widely used as a fluorescence probe of protein structure and are complementary to the Trp-M_{Se} approach illustrated here. M_{Se} is a more conservative substitution in the hydrophobic core of a protein or in transmembrane helices than His. His on the other hand may be a more conservative replacement for surface exposed hydrophilic residues.

The lifetime of the slow component (~3-4 ns) is similar in all three systems investigated here, but the rapidly decaying component of the curve has a shorter lifetime in CTL9 (~0.1 ns) than in the other systems studied (~1 ns). This is consistent with the closer approach of Trp and M_{Se} in the folded state of Y126W/H144M_{Se}-CTL9 than in the helical peptide or N28M_{Se}-HP36. The positioning of the Trp-M_{Se} pair on the antiparallel β-sheet of CTL9 brings the indole ring of Trp in much closer proximity to the Se atom of M_{Se} than does introducing the pair at *i* and *i*+4 positions in an α-helix, a conclusion supported by examining models of the low energy rotamers

(Figure 2.20). This contrasts with the lifetime data in high urea for the helical peptide and N28M_{Se}-HP36, which indicates that Trp is at least partially quenched by M_{Se} even under denaturing conditions. The different effects observed in the urea unfolded states of the helical systems versus CTL9 are likely caused by transient contacts between the indole ring and Se atom in the helical systems, facilitated by their close proximity in primary sequence, in contrast the M_{Se} and Trp residues are further apart in primary sequence in CTL9. Low levels of residual helical structure could also contribute to the effects observed for the unfolded state of HP36.

The Trp-M_{Se} pair is complementary to the F_{CN}-M_{Se} pair.[11, 13] An advantage of F_{CN} is that it can be selectively excited against a background of Trp and Tyr residues, however incorporation of an F_{CN}-M_{Se} pair into an expressed protein does present more technical challenges than incorporation of a Trp-M_{Se} pair. Interesting future applications could include the incorporation of specific F_{CN}-M_{Se} and Trp-M_{Se} pairs into the same protein as orthogonal fluorescence probes of different structural elements.

In summary, the data presented here demonstrates the utility of Trp-M_{Se} pairs as a straightforward probe of specific elements of protein structure. The cases studied here involve monitoring α-helical and anti-parallel β-sheet structure, but the approach is not limited to these classes of secondary structure; any structure, including a parallel β-sheet or a loop which brings the two residues into van der Waals contact will lead to efficient quenching. The pair could also be used to monitor protein-ligand interactions provided complex formation leads to the close approach of the Se atom and the indole ring.

2.6. Figures

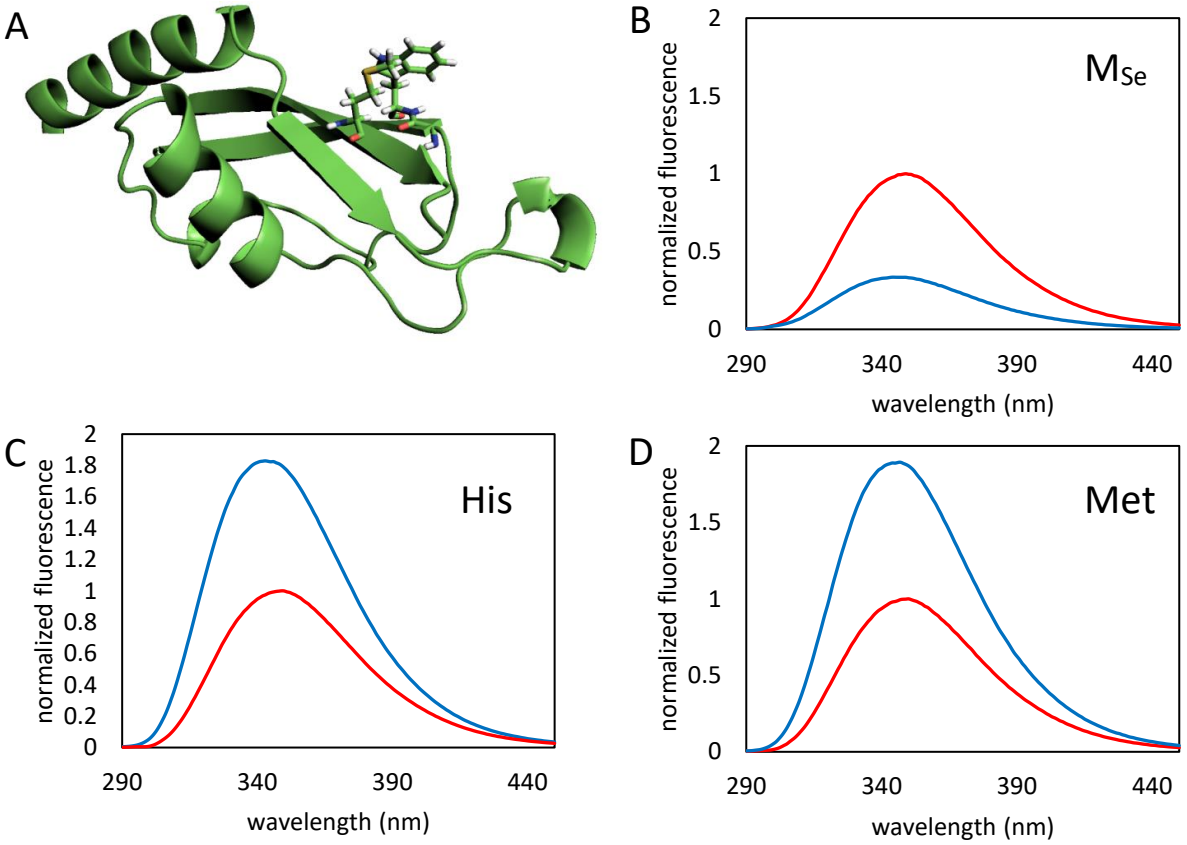


Figure 2.1. The fluorescence of Trp-M_{Se} pairs probe β-sheet formation. **(A)** Ribbon diagram of Y126W/H144M_{Se}-CTL9 based on PDB structure 1DIV showing the proximity of Trp and M_{Se} at positions 126 and 144.[21] **(B)** Fluorescence emission spectra of Y126W/H144M_{Se}-CTL9 at 20 °C in 20 mM tris buffer at pH 7.5 (blue) and in the same buffer with 9.5 M urea (red). **(C)** Fluorescence emission spectra of Y126W-CTL9 at 20 °C in 20 mM tris buffer at pH 7.5 (blue) and in the same buffer with 9.5 M urea (red). **(D)** Fluorescence emission spectra of Y126W/H144M-CTL9 at 20 °C in 20 mM tris buffer at pH 7.5 (blue) and in the same buffer with 9.5 M urea (red).

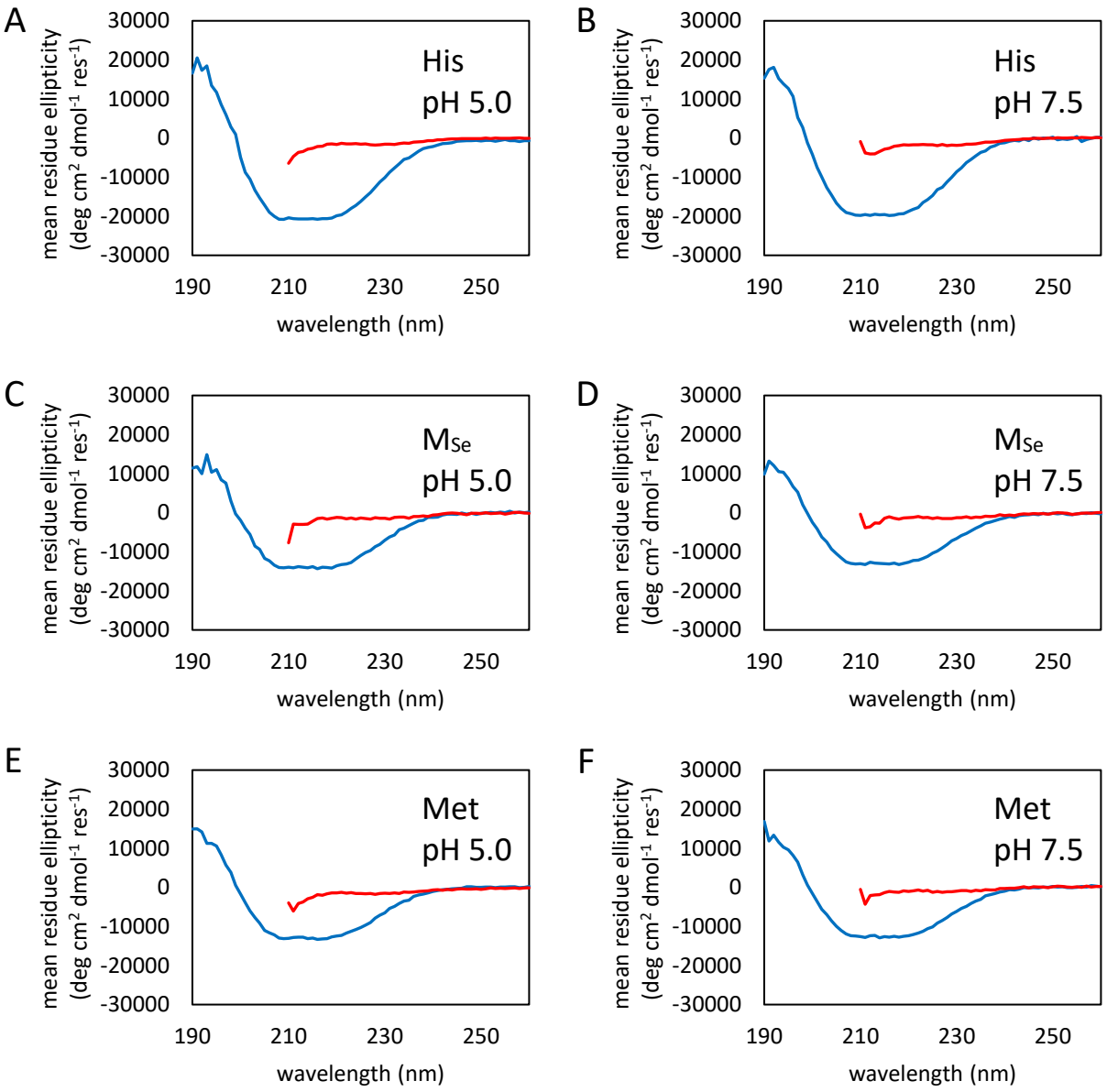


Figure 2.2. CD spectra of CTL9 quenching mutants. **(A)** CD spectra of Y126W-CTL9 at 20 °C in 20 mM acetate buffer at pH 5.0 (blue) and in the same buffer with 9.5 M urea (red). **(B)** CD spectra of Y126W-CTL9 at 20 °C in 20 mM tris buffer at pH 7.5 (blue) and in the same buffer with 9.5 M urea (red). **(C)** CD spectra of Y126W/H144M$_{Se}$-CTL9 at 20 °C in 20 mM acetate buffer at pH 5.0 (blue) and in the same buffer with 9.5 M urea (red). **(D)** CD spectra of Y126W/H144M$_{Se}$-CTL9 at 20 °C in 20 mM tris buffer at pH 7.5 (blue) and in the same buffer with 9.5 M urea (red). **(E)** CD spectra of Y126W/H144M-CTL9 at 20 °C in 20 mM acetate buffer at pH 5.0 (blue) and in the same buffer with 9.5 M urea (red). **(F)** CD spectra of Y126W/H144M-CTL9 at 20 °C in 20 mM tris buffer at pH 7.5 (blue) and in the same buffer with 9.5 M urea (red). The protein concentration in all samples was 25 μM.

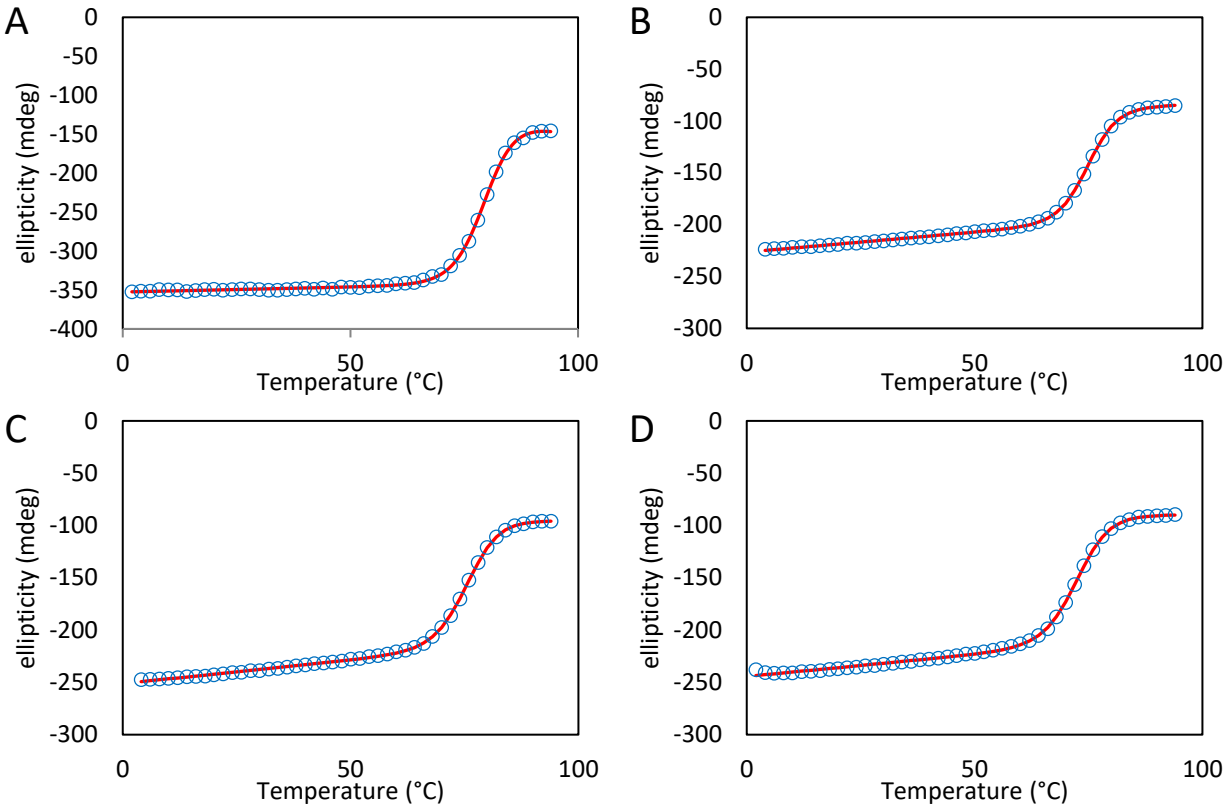


Figure 2.3. Thermal stability of CTL9 constructs in tris buffer (20 mM, pH 7.5). **(A)** Y126W-CTL9, T_M = 79.5 °C **(B)** Y126W/H144M-CTL9, T_M = 74.9 °C **(C)** Y126W/H144M$_{Se}$-CTL9, T_M = 75.6 °C **(D)** Y126W/H144M$_{SeO}$-CTL9, T_M = 72.5 °C. The red line is the best fit to equation (2.4).

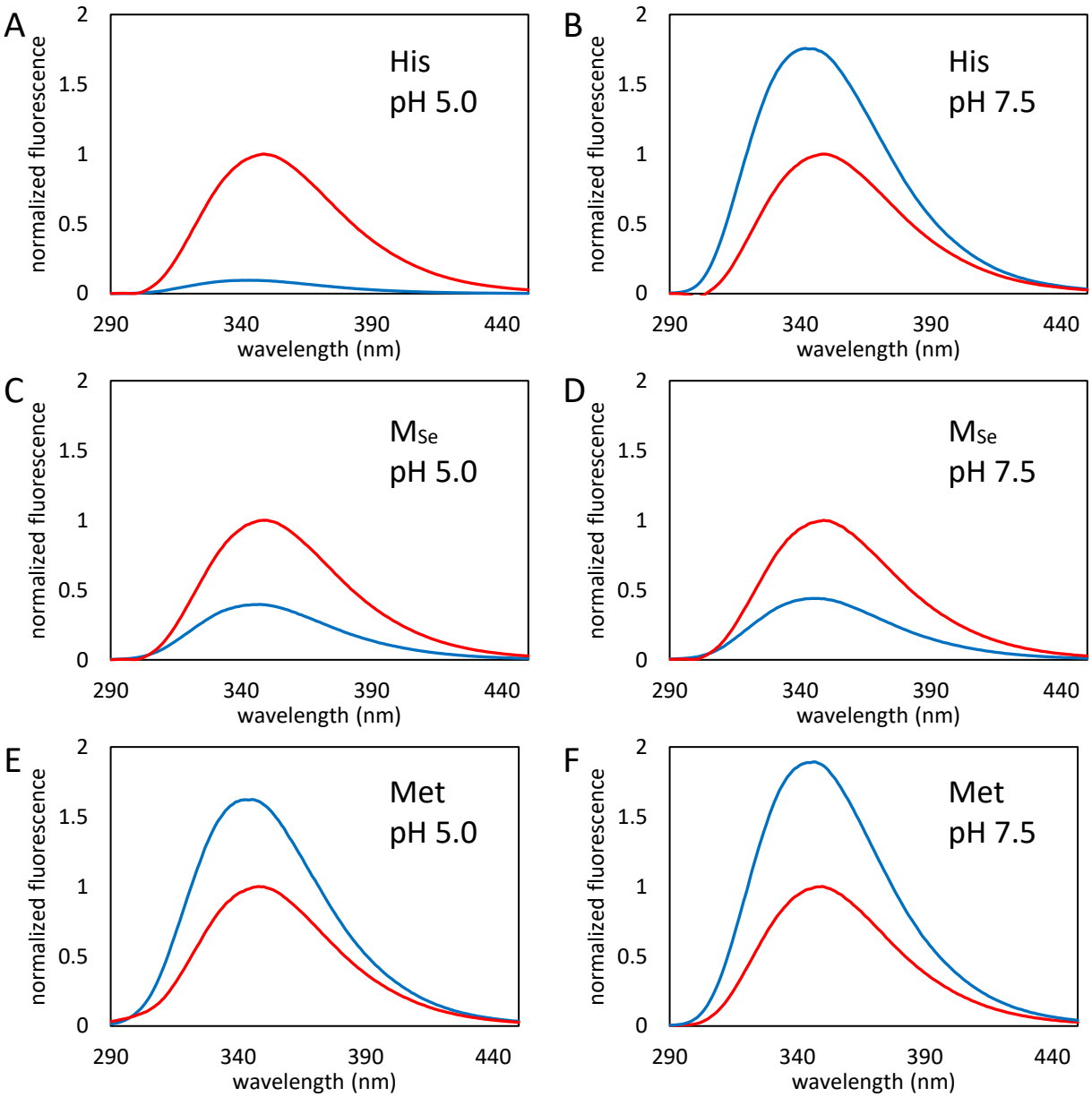


Figure 2.4. Fluorescence emission spectra of CTL9 mutants. **(A)** Fluorescence emission spectra of Y126W-CTL9 at 20 °C in 20 mM acetate buffer at pH 5.0 (blue) and in the same buffer with 9.5 M urea (red). **(B)** Fluorescence emission spectra of Y126W-CTL9 at 20 °C in 20 mM tris buffer at pH 7.5 (blue) and in the same buffer with 9.5 M urea (red). **(C)** Fluorescence emission spectra of Y126W/H144M$_{Se}$-CTL9 at 20 °C in 20 mM acetate buffer at pH 5.0 (blue) and in the same buffer with 9.5 M urea (red). **(D)** Fluorescence emission spectra of Y126W/H144M$_{Se}$-CTL9 at 20 °C in 20 mM tris buffer at pH 7.5 (blue) and in the same buffer with 9.5 M urea (red). **(E)** Fluorescence emission spectra of Y126W/H144M-CTL9 at 20 °C in 20 mM acetate buffer at pH 5.0 (blue) and in the same buffer with 9.5 M urea (red). **(F)** Fluorescence emission spectra of Y126W/H144M-CTL9 at 20 °C in 20 mM tris buffer at pH 7.5 (blue) and in the same buffer with 9.5 M urea (red). The protein concentration in all samples was 25 μM.

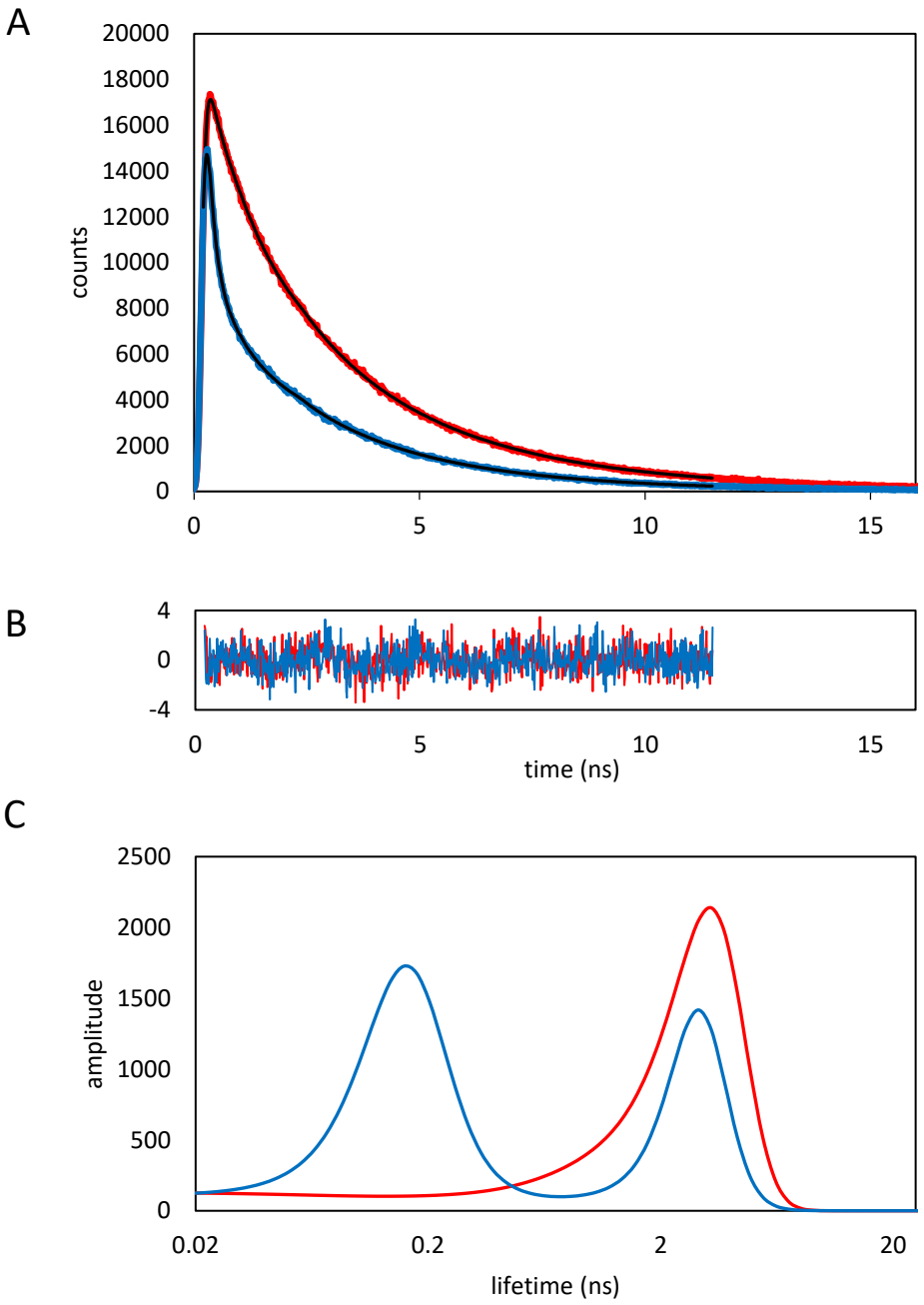


Figure 2.5. Maximum entropy analysis of time-resolved fluorescence decays of Y126W/H144MSe-CTL9 in 20 mM tris buffer at pH 7.5 (blue) and in the same buffer with 9.5 M urea (red). **(A)** Experimentally observed decays fit using a maximum entropy model reconvoluted with the instrument response (black). **(B)** Weighted residuals between the fit and the experimental data. **(C)** Maximum entropy analysis derived plot of the distribution of lifetimes.

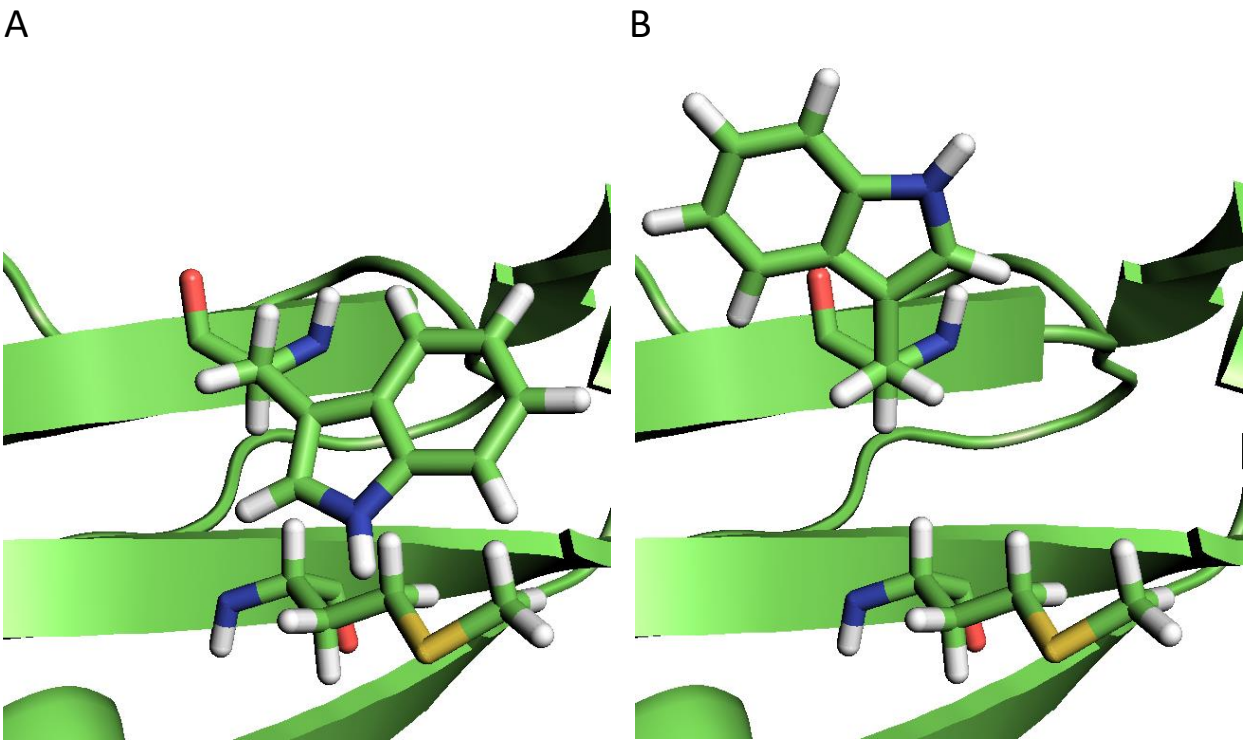


Figure 2.6. Models of the allowed χ_1 rotamers for Y126W/H144MSe-CTL9 based on PDB structure 1DIV.[21] **(A)** The $\chi_1 \approx -60°$ rotamer, which corresponds to the orientation of the phenol ring of Tyr in the crystal structure, packing the indole ring against the M_{Se} sidechain. **(B)** The $\chi_1 \approx 60°$ rotamer, which exposes the Trp sidechain to solvent and moves the indole ring out of van der Waals contact with the Se atom. Models were constructed using the parameters for Met.

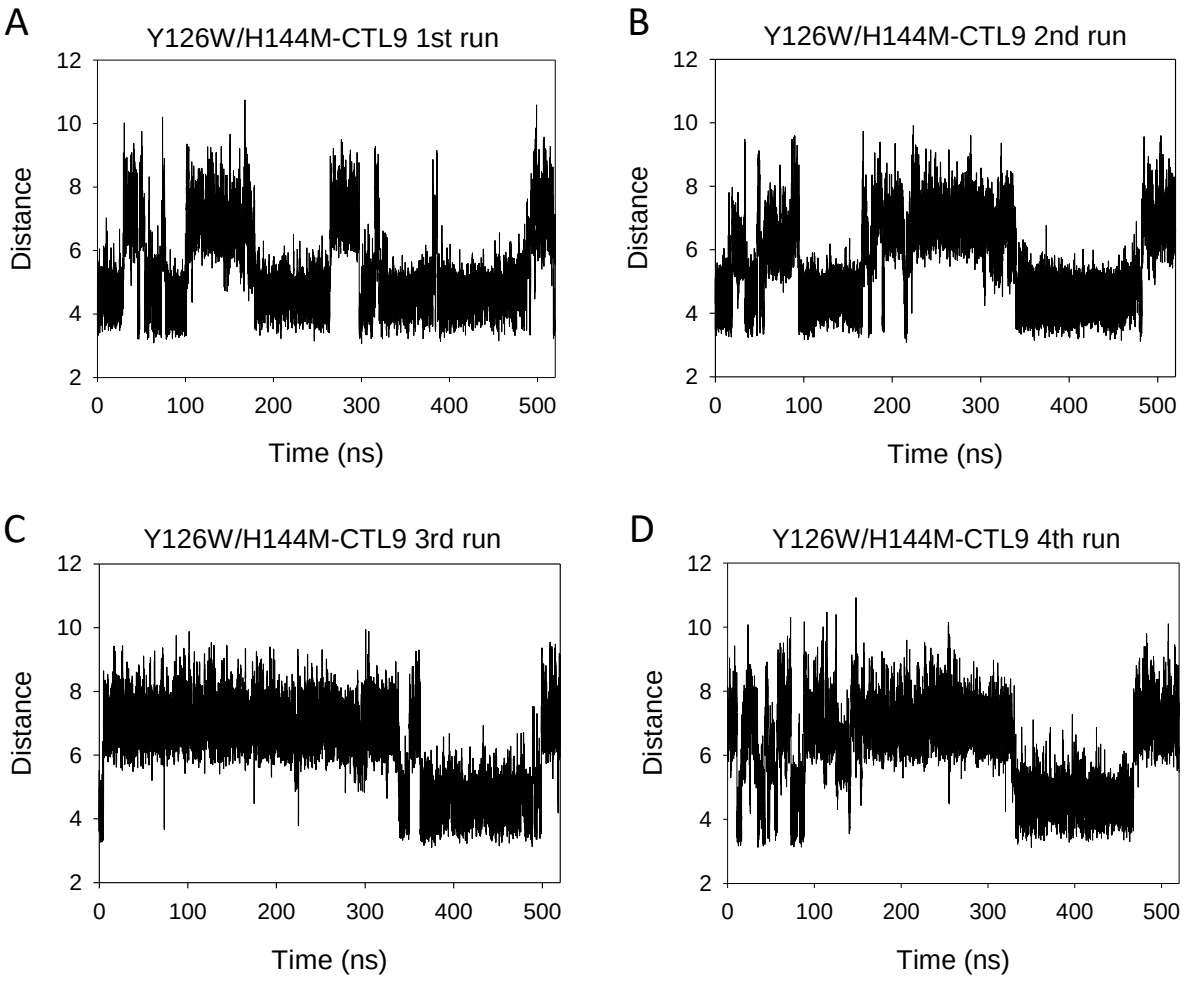


Figure 2.7. MD simulations of the distance between Trp and Met in Y126W/H144M-CTL9. **(A)** and **(B)** Independent MD simulations in which W126 starts with the rotamer χ_1 = -65.0°, χ_2 = -84.9° **(C)** MD simulation in which W126 starts with the rotamer χ_1 = -71.2°, χ_2 = 94.3° **(D)** MD simulation in which W126 starts with the rotamer χ_1 = 62.1°, χ_2 = 94.4°. Distances were measured between the geometric center of Cδ2 and Cε2 of the Trp indole ring and sulfur atom of Met.

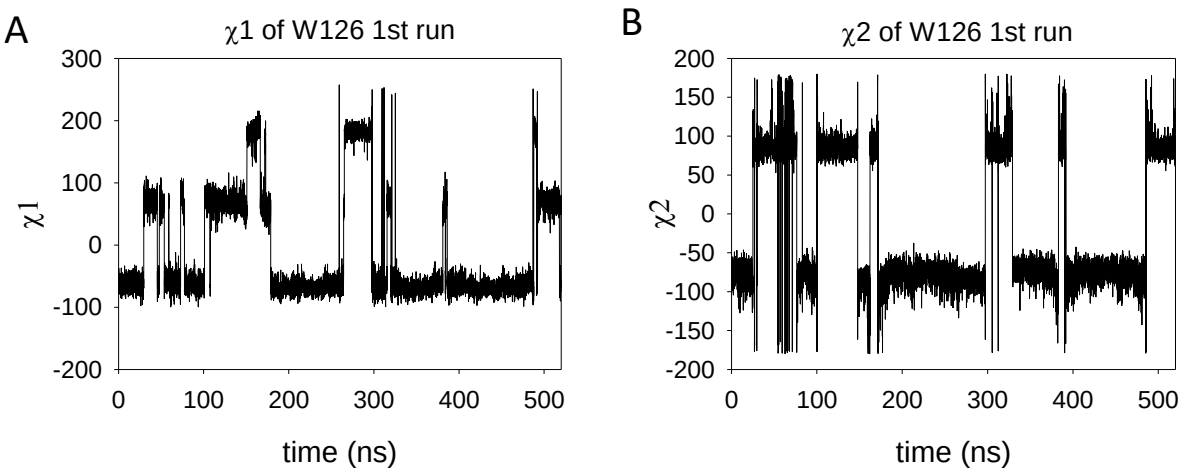


Figure 2.8. MD simulations of the χ_1 and χ_2 dihedral angles of W126 in Y126W/H144M-CTL9 during the 1st run of the MD simulations in which W126 starts with a rotamer state of χ_1 = -65.0°, χ_2 = -84.9° **(A)** The χ_1 angle of W126 in degrees *vs* time. **(B)** The χ_2 angle of W126 in degrees *vs* time.

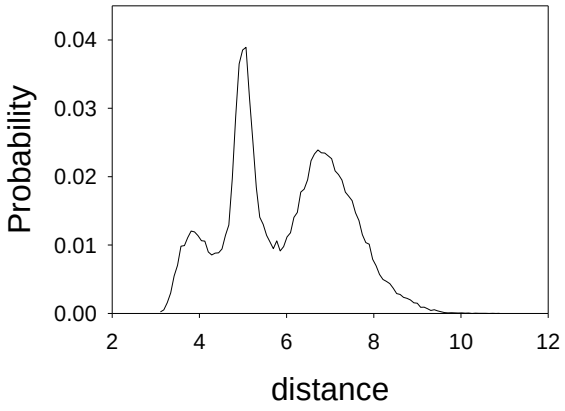


Figure 2.9. Histogram of the distance distribution between W126 and M144 in Y126W/H144M-CTL9 collected from MD simulations. The distance was measured between the geometric center of Cδ2 and Cε2 of the Trp indole ring and the sulfur atom of Met.

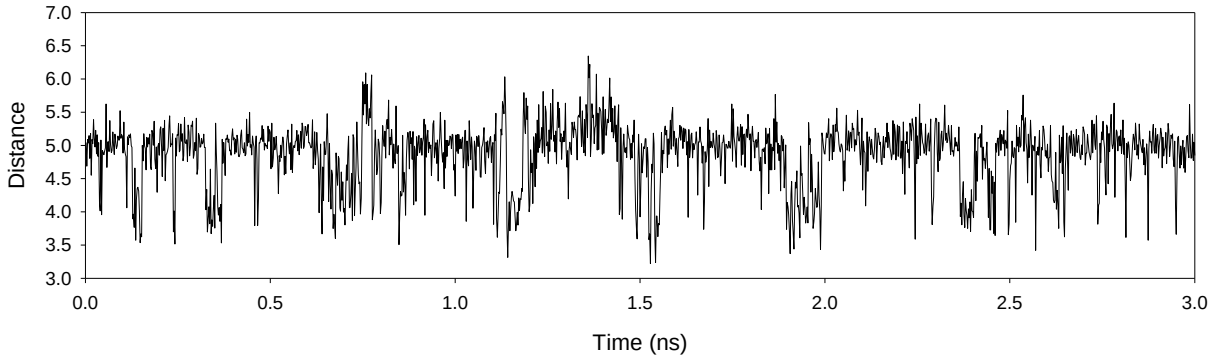


Figure 2.10. The distance between W126 and M144 of Y126W/H144M-CTL9 during a 3 ns MD simulation with a sampling frequency of 0.002 ns.

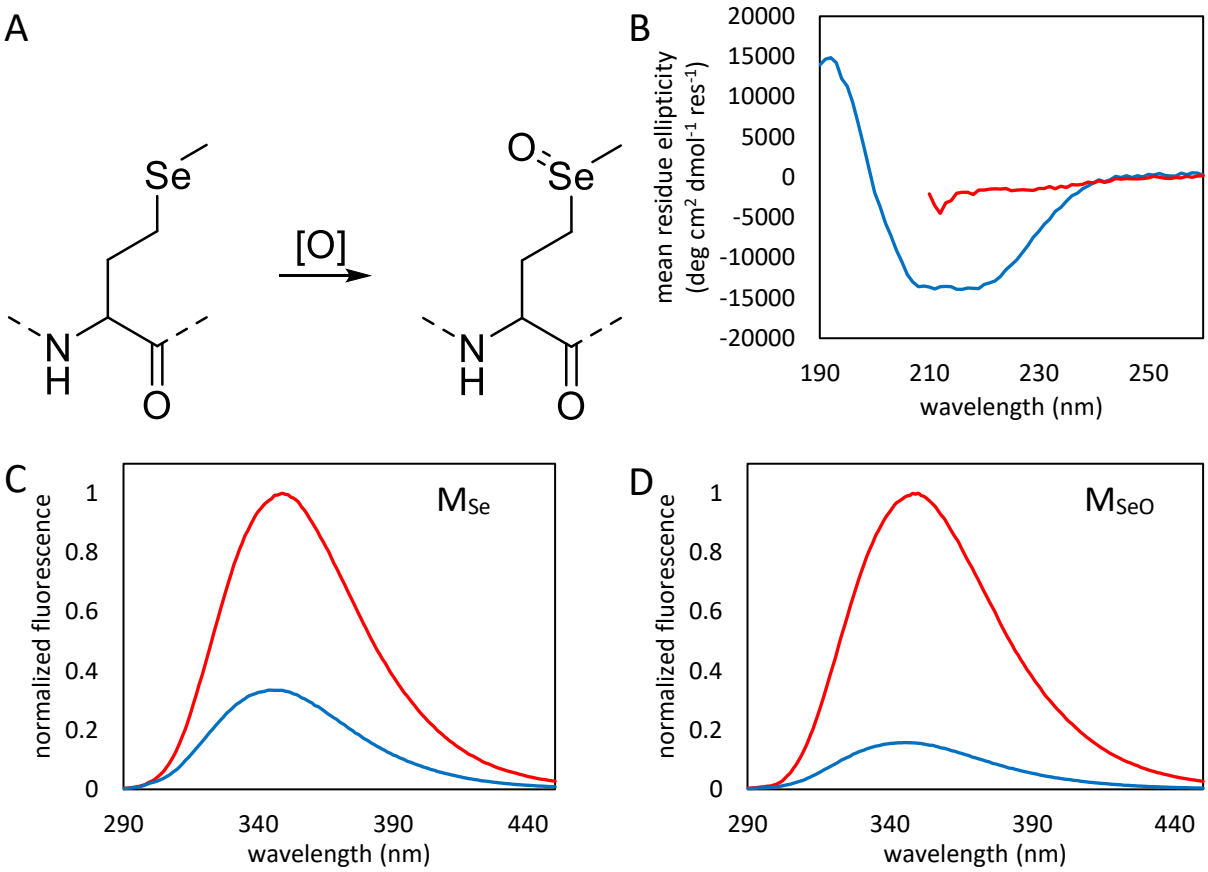


Figure 2.11. Oxidation of M_{Se} to M_{SeO} slightly increases fluorescence quenching efficiency. **(A)** Structure of selenomethionine (M_{Se}) and the oxidation product selenomethionine selenoxide (M_{SeO}). **(B)** CD spectra of Y126W/H144M_{SeO}-CTL9 at 20 °C in 20 mM tris buffer at pH 7.4 (blue) and in the same buffer with 9.5 M urea (red). **(C)** Fluorescence emission spectra of Y126W/H144M_{Se}-CTL9 at 20 °C in 20 mM tris buffer at pH 7.4 (blue) and in the same buffer with 9.5 M urea (red). **(D)** Fluorescence emission spectra of Y126W/H144M_{SeO}-CTL9 at 20 °C in 20 mM tris buffer at pH 7.4 (blue) and in the same buffer with 9.5 M urea (red). Note the decreased native state intensity for the selenoxide sample.

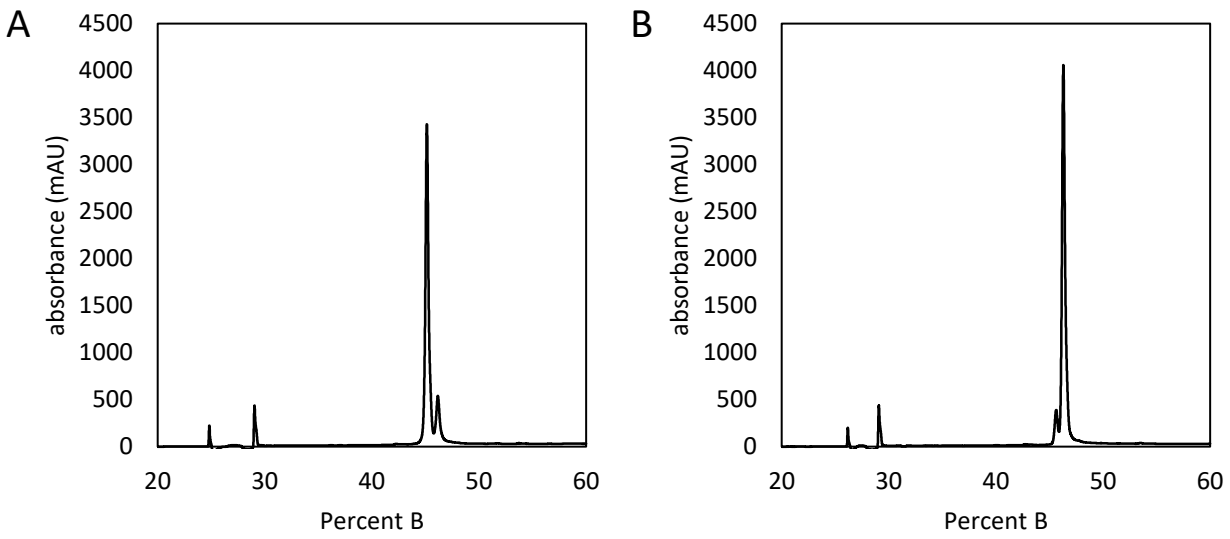


Figure 2.12. Selenomethionine is efficiently oxidized by 0.005% H_2O_2 while methionine is not. **(A)** Preparative HPLC trace of Y126W/H144M_{Se}-CTL9 after oxidation in 0.005% H_2O_2 for 4 hr. The peak centered at 45% B was identified by LC-TOF MS as Y126W/H144M_{SeO}-CTL9 and the peak centered at 46% B was identified as Y126W/H144M_{Se}-CTL9. **(B)** Preparative HPLC trace of Y126W/H144M-CTL9 after oxidation in 0.005% H_2O_2 for 4 hr. The peak centered at 46% B was identified by LC-TOF MS as Y126W/H144M_{ox}-CTL9 and the peak centered at 47% B was

identified as Y126W/H144M-CTL9. The absorbance was monitored at 220 nm.

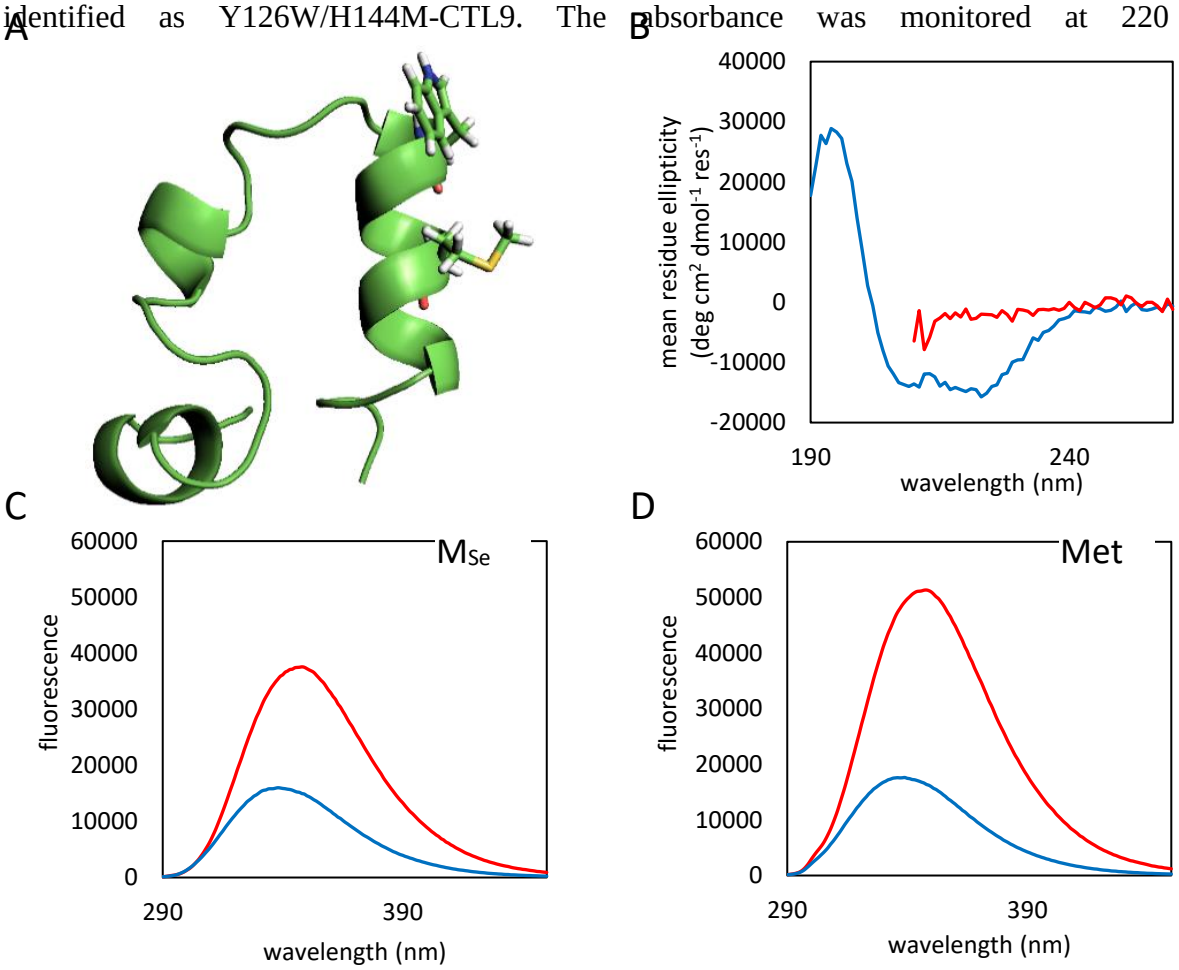


Figure 2.13. The fluorescence of Trp-M_{Se} pairs monitors the folding of helical proteins. **(A)** Ribbon diagram of HP36 based on PDB structure 1VII showing the proximity of Trp and M_{Se} at positions 24 and 28.[30] **(B)** CD spectra of N28M_{Se}-HP36 at 20 °C in 20 mM acetate buffer at pH 5.0 (blue) and in the same buffer with 9.5 M urea (red). **(C)** Fluorescence emission spectra of N28M_{Se}-HP36 at 20 °C in 20 mM acetate buffer at pH 5.0 (blue) and in the same buffer with 9.5 M urea (red). **(D)** Fluorescence emission spectra of N28M-HP36 at 20 °C in 20 mM acetate buffer at pH 5.0 (blue) and in the same buffer with 9.5 M urea (red).

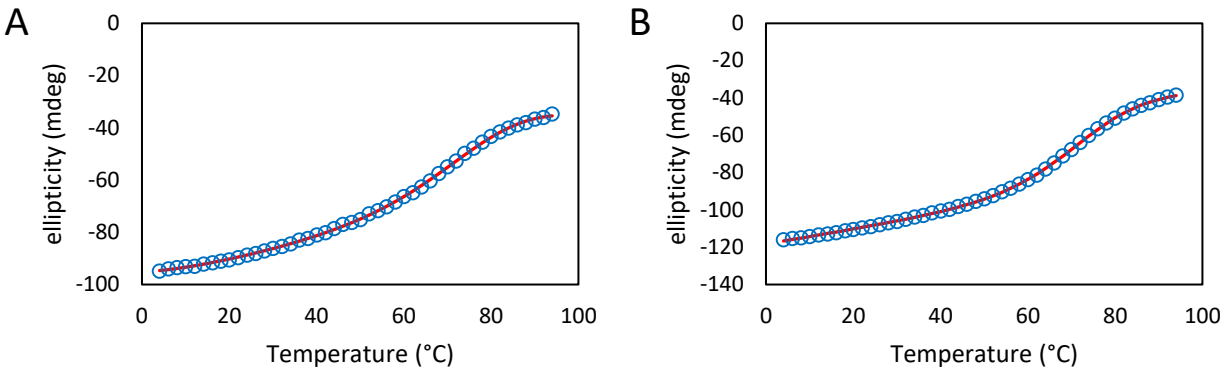


Figure 2.14. Thermal stability of HP36 constructs. **(A)** N28M$_{Se}$-HP36, T_M = 70.8 °C **(B)** N28M-HP36 T_M = 71.8 °C. The red line is the best fit to equation (2.4).

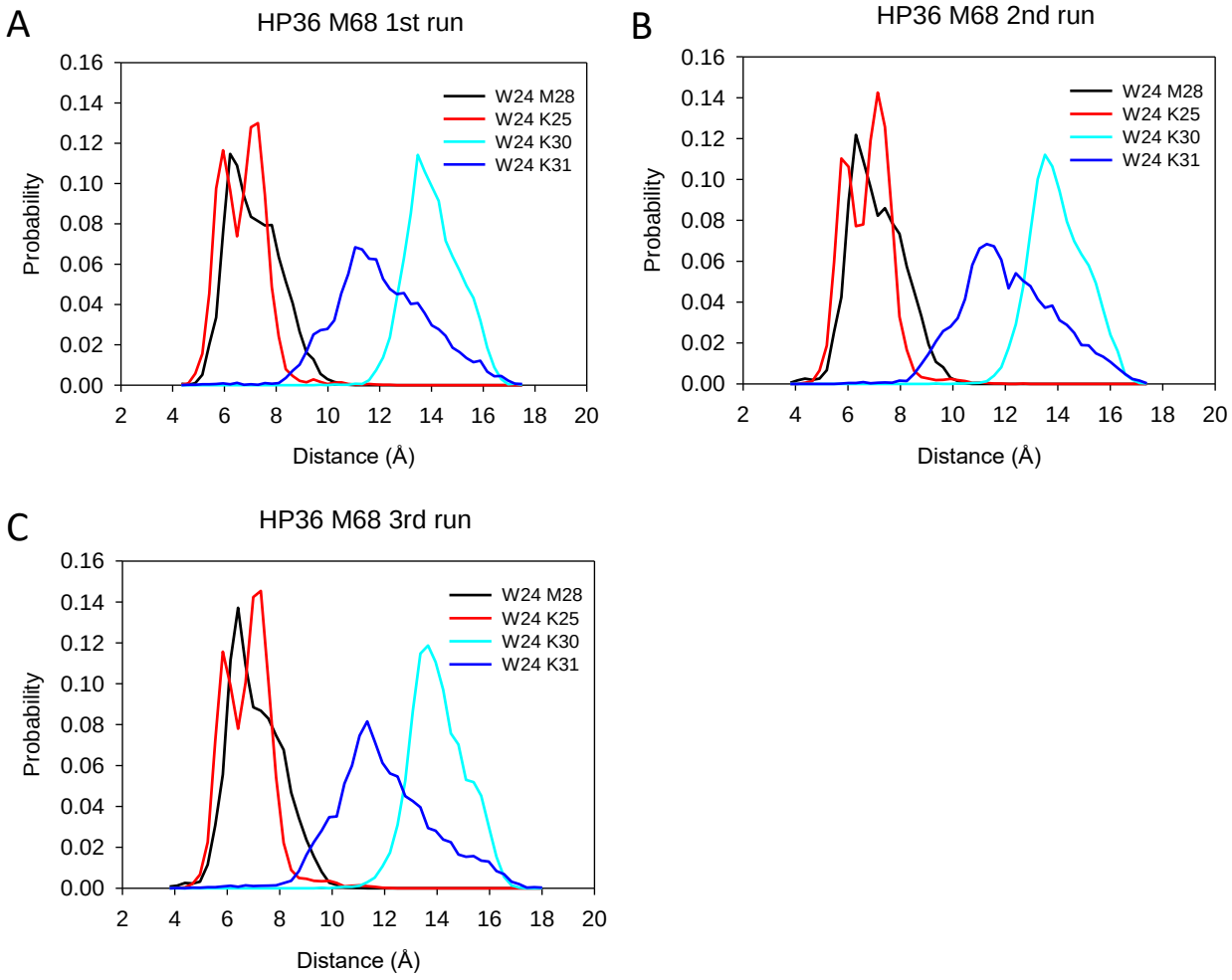


Figure 2.15. Distribution of distances between Trp and various quenchers. Distances between W24/M28, W24/K25, W24/K30 and W24/K31 in N28M-HP36 from the last 300 ns of three independent 400 ns MD simulations are plotted. The simulations used starting structures with different rotamer states of W24. **(A)** χ_1 = 63.2°, χ_2 = 85.5° **(B)** χ_1 = -82.2°, χ_2 = -89.3° **(C)** χ_1 = -177.0°, χ_2 = 75.0°

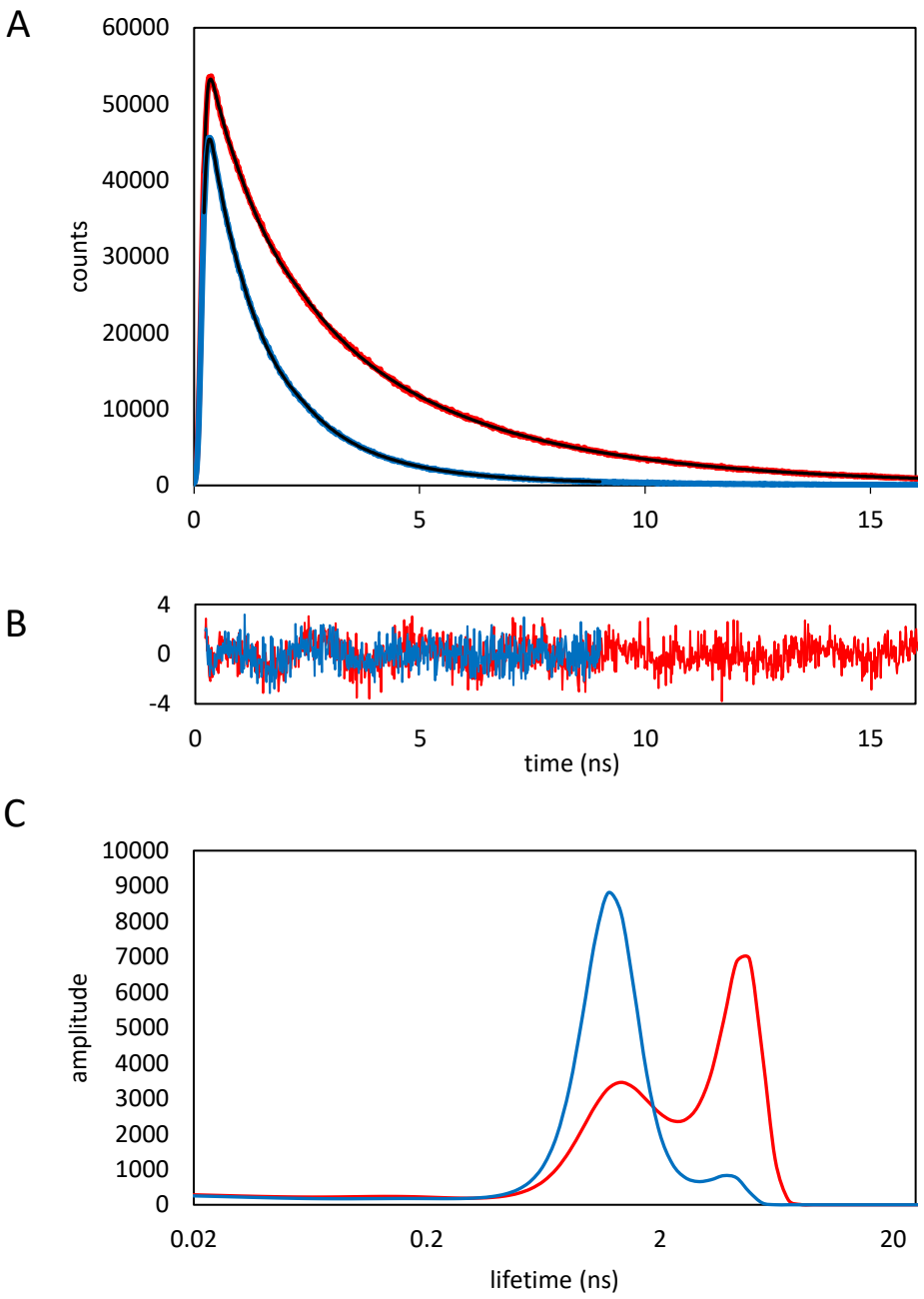


Figure 2.16. Maximum entropy analysis of time-resolved fluorescence decays of N28M$_{Se}$-HP36 in 20 mM acetate buffer at pH 5.0 (blue) and in the same buffer with 9.5 M urea (red). **(A)** Experimentally observed decays fit using a maximum entropy model reconvoluted with the instrument response (black). **(B)** Weighted residuals between the fit and the experimental data. **(C)** Maximum entropy analysis derived plot of the distribution of lifetimes.

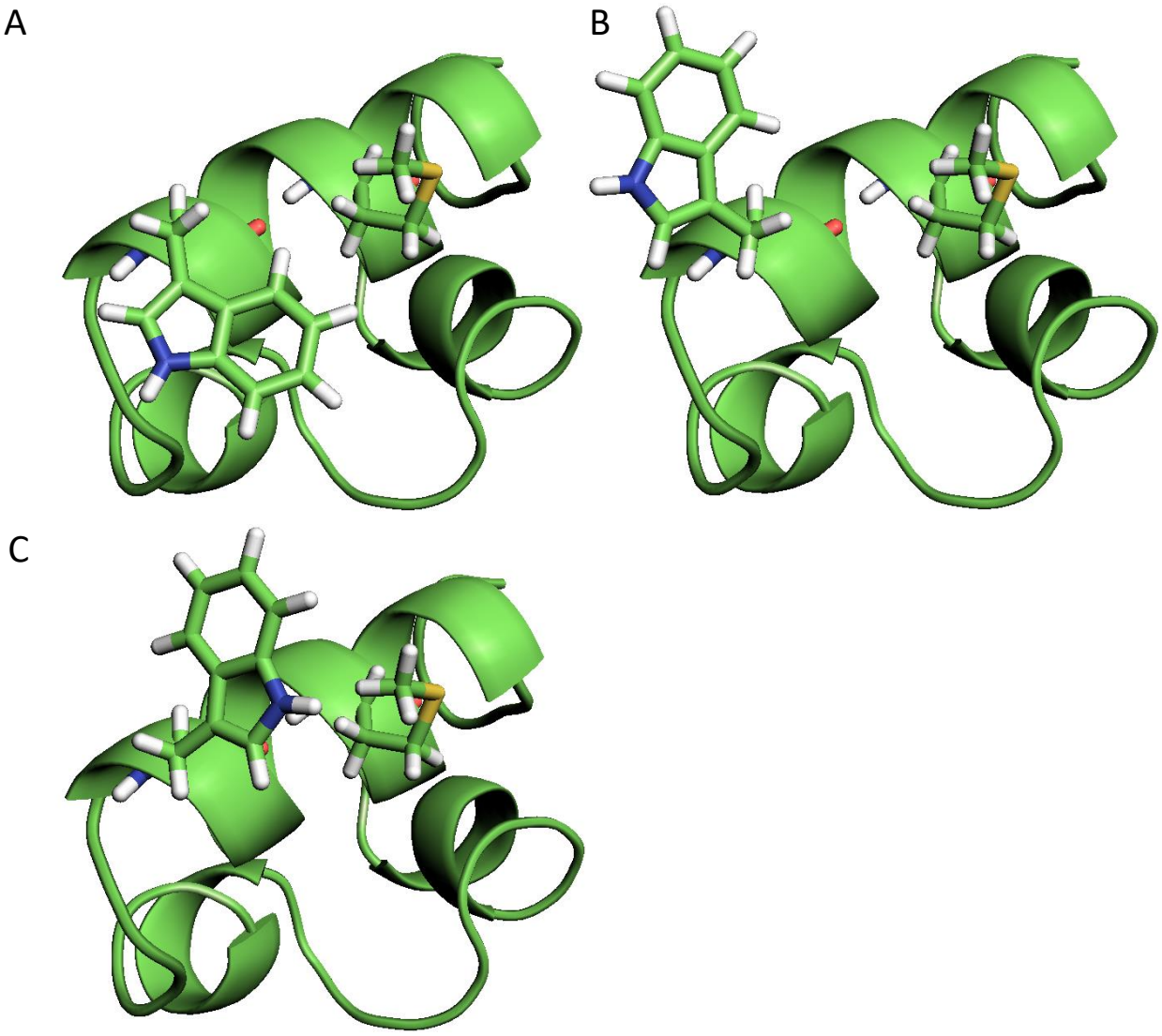


Figure 2.17. Models of the allowed χ_1 rotamers for N28M_{Se}-HP36 based on PDB structure 1VII.[30] **(A)** The $\chi_1 \approx 60°$ rotamer, the orientation from the crystal structure, which may bring the indole ring into transient contact with the Se atom. **(B)** The $\chi_1 \approx -60°$ rotamer, which exposes the Trp sidechain to solvent and moves the indole ring out of van der Waals contact with the Se atom. **(C)** The $\chi_1 \approx 180°$ rotamer, which brings the indole ring into close contact with the Se atom. Models were built using the parameters for Met.

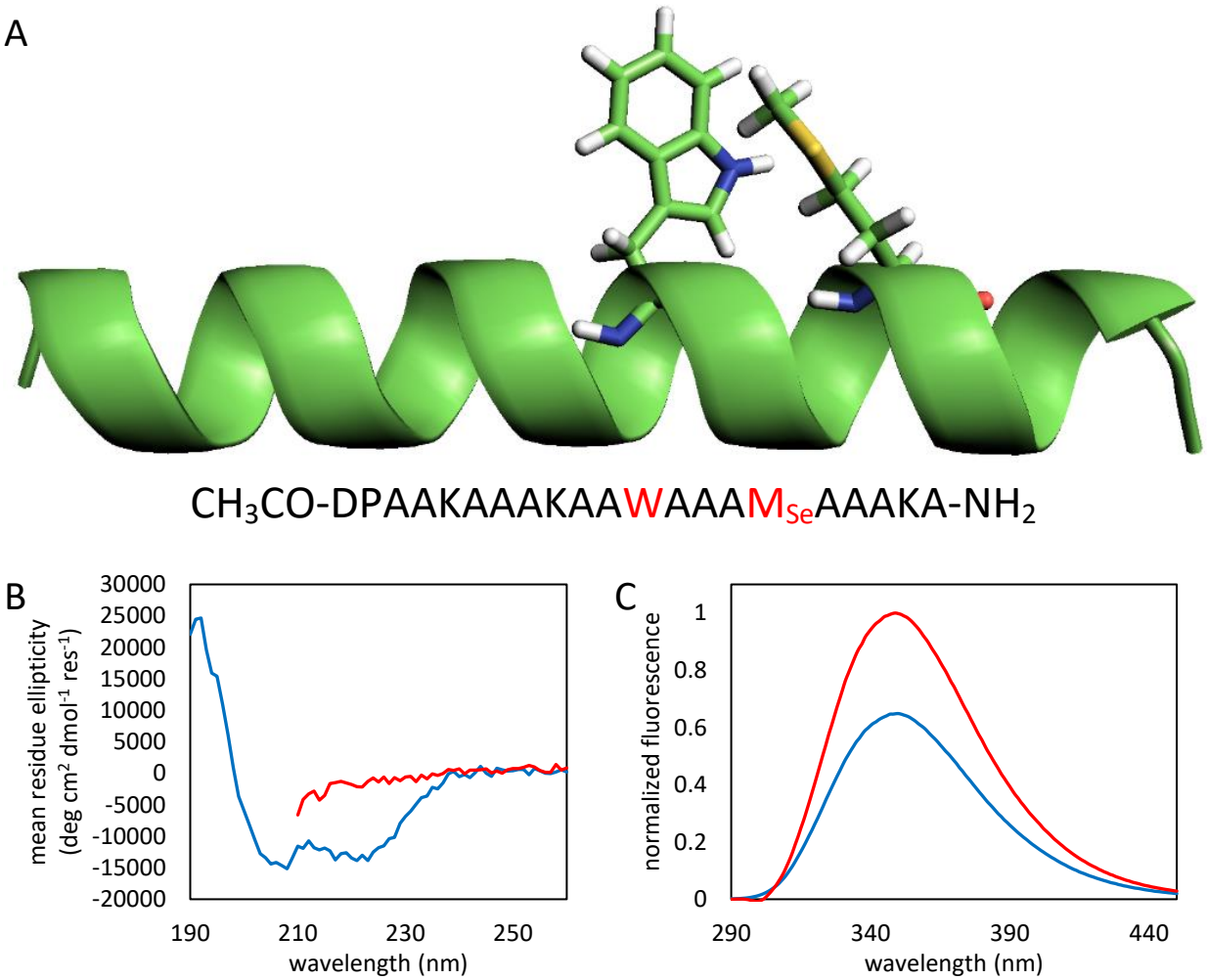


Figure 2.18. The fluorescence of Trp-M_{Se} pairs is sensitive to α-helix formation. **(A)** Sequence and ribbon diagram of an idealized 21-residue helical peptide showing the proximity of Trp and M_{Se} at positions *i* and *i*+4. **(B)** CD spectra of the 21-residue helical peptide at 25 °C in 10 mM acetate buffer at pH 5.5 (blue) and in the same buffer with 8.0 M urea (red). **(C)** Fluorescence emission spectra of the 21-residue helical peptide at 25 °C in 10 mM acetate buffer at pH 5.5 (blue) and in the same buffer with 8.0 M urea (red). The model was built using the parameters for Met.

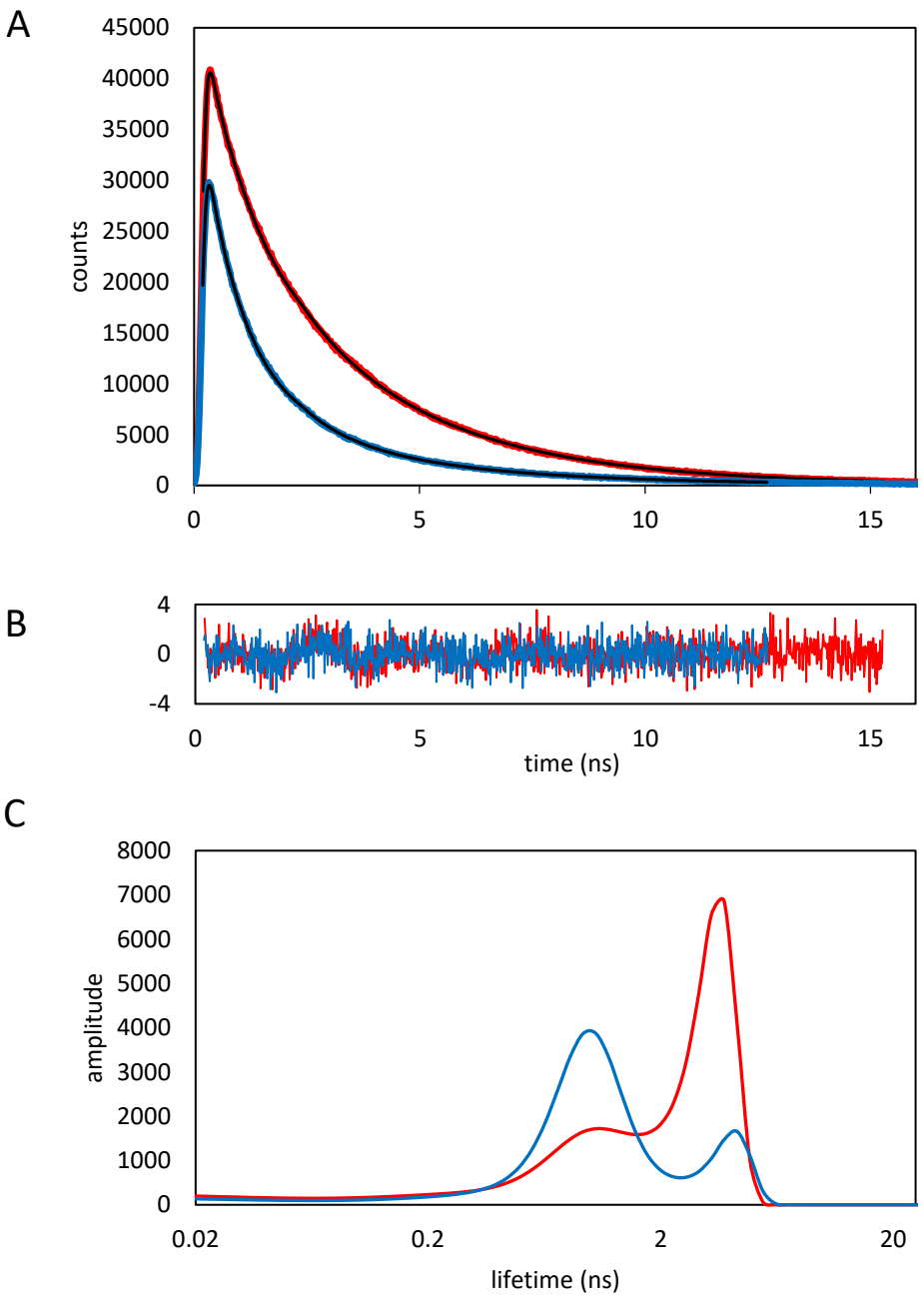


Figure 2.19. Maximum entropy analysis of time-resolved fluorescence decays for the 21-residue helical peptide in 10 mM acetate buffer at pH 5.5 (blue) and in the same buffer with 8.0 M urea (red). **(A)** Experimentally observed decays fit using a maximum entropy model reconvoluted with the instrument response (black). **(B)** Weighted residuals between the fit and the experimental data. **(C)** Maximum entropy analysis derived plot of the distribution of lifetimes.

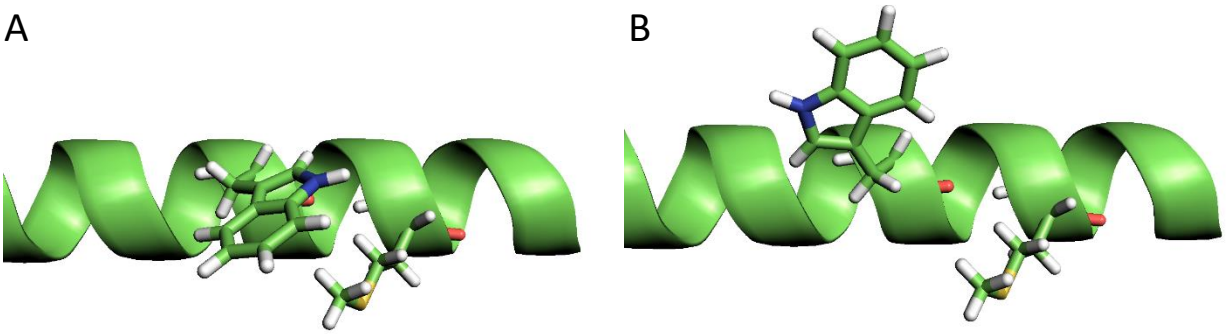


Figure 2.20. Models of the allowed χ_1 rotamers for the 21-residue helical peptide. **(A)** The $\chi_1 \approx 180°$ rotamer, which brings the indole ring into close contact with the Se atom. **(B)** The $\chi_1 \approx -60°$ rotamer, which moves the indole ring out of van der Waals contact with the Se atom. The model was built using the parameters for Met.

2.7. References

[1] Lakowicz, J. R. (2006) *Principles of Fluorescence Spectroscopy*, 3 ed., Springer.

[2] Brown, M. P., and Royer, C. (1997) Fluorescence spectroscopy as a tool to investigate protein interactions, *Curr Opin Biotech 8*, 45-49.

[3] Kubelka, J., Eaton, W. A., and Hofrichter, J. (2003) Experimental tests of villin subdomain folding simulations, *J Mol Biol 329*, 625-630.

[4] Tucker, M. J., Oyola, R., and Gai, F. (2006) A novel fluorescent probe for protein binding and folding studies: p-cyano-phenylalanine, *Biopolymers 83*, 571-576.

[5] Tucker, M. J., Tang, J., and Gai, F. (2006) Probing the kinetics of membrane-mediated helix folding, *J Phys Chem B 110*, 8105-8109.

[6] Aprilakis, K. N., Taskent, H., and Raleigh, D. P. (2007) Use of the novel fluorescent amino acid p-cyanophenylalanine offers a direct probe of hydrophobic core formation during the folding of the N-terminal domain of the ribosomal protein L9 and provides evidence for two-state folding, *Biochemistry 46*, 12308-12313.

[7] Wang, L., Xie, J., and Schultz, P. G. (2006) Expanding the genetic code, *Annu Rev Biophys Biomol Struct 35*, 225-249.

[8] Miyake-Stoner, S. J., Miller, A. M., Hammill, J. T., Peeler, J. C., Hess, K. R., Mehl, R. A., and Brewer, S. H. (2009) Probing protein folding using site-specifically encoded unnatural amino acids as FRET donors with tryptophan, *Biochemistry 48*, 5953-5962.

[9] Rogers, J. M., Lippert, L. G., and Gai, F. (2010) Non-natural amino acid fluorophores for one- and two-step fluorescence resonance energy transfer applications, *Anal Biochem 399*, 182-189.

[10] Taskent-Sezgin, H., Marek, P., Thomas, R., Goldberg, D., Chung, J., Carrico, I., and Raleigh, D. P. (2010) Modulation of p-cyanophenylalanine fluorescence by amino acid side chains and rational design of fluorescence probes of alpha-helix formation, *Biochemistry 49*, 6290-6295.

[11] Mintzer, M. R., Troxler, T., and Gai, F. (2015) p-Cyanophenylalanine and selenomethionine constitute a useful fluorophore-quencher pair for short distance measurements: application to polyproline peptides, *Phys Chem Chem Phys 17*, 7881-7887.

[12] Yuan, T., Weljie, A. M., and Vogel, H. J. (1998) Tryptophan fluorescence quenching by methionine and selenomethionine residues of calmodulin: orientation of peptide and protein binding, *Biochemistry 37*, 3187-3195.

[13] Peran, I., Watson, M. D., Bilsel, O., and Raleigh, D. P. (2016) Selenomethionine, p-cyanophenylalanine pairs provide a convenient, sensitive, non-perturbing fluorescent probe of local helical structure, *Chem Commun (Camb) 52*, 2055-2058.

[14] Watson, M. D., Peran, I., and Raleigh, D. P. (2016) A non-perturbing probe of coiled coil formation based on electron transfer mediated fluorescence quenching, *Biochemistry 55*, 3685-3691.

[15] Sahoo, H., Roccatano, D., Zacharias, M., and Nau, W. M. (2006) Distance distributions of short polypeptides recovered by fluorescence resonance energy transfer in the 10 A domain, *J Am Chem Soc 128*, 8118-8119.

[16] Schaefer, S. A., Dong, M., Rubenstein, R. P., Wilkie, W. A., Bahnson, B. J., Thorpe, C., and Rozovsky, S. (2013) (77)Se enrichment of proteins expands the biological NMR toolbox, *J Mol Biol 425*, 222-231.

[17] Hendrickson, W. A., Horton, J. R., and LeMaster, D. M. (1990) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure, *EMBO J 9*, 1665-1672.

[18] Kuhlman, B., Yang, H. Y., Boice, J. A., Fairman, R., and Raleigh, D. P. (1997) An exceptionally stable helix from the ribosomal protein L9: implications for protein folding and stability, *J Mol Biol 270*, 640-647.

[19] Wu, Y., Kondrashkina, E., Kayatekin, C., Matthews, C. R., and Bilsel, O. (2008) Microsecond acquisition of heterogeneous structure in the folding of a TIM barrel protein, *Proc Natl Acad Sci U S A 105*, 13367-13372.

[20] Chiu, T. K., Kubelka, J., Herbst-Irmer, R., Eaton, W. A., Hofrichter, J., and Davies, D. R. (2005) High-resolution x-ray crystal structures of the villin headpiece subdomain, an ultrafast folding protein, *Proc Natl Acad Sci U S A 102*, 7517-7522.

[21] Hoffman, D. W., Davies, C., Gerchman, S. E., Kycia, J. H., Porter, S. J., White, S. W., and Ramakrishnan, V. (1994) Crystal structure of prokaryotic ribosomal protein L9: a bi-lobed RNA-binding protein, *EMBO J 13*, 205-212.

[22] Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling, *Electrophoresis 18*, 2714-2723.

[23] Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB, *J Chem Theory Comput 11*, 3696-3713.

[24] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of Simple Potential Functions for Simulating Liquid Water, *J Chem Phys 79*, 926-935.

[25] Darden, T., York, D., and Pedersen, L. (1993) Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems, *J Chem Phys 98*, 10089-10092.

[26] Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes, *J Comput Phys 23*, 327-341.

[27] Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A., and Haak, J. R. (1984) Molecular-Dynamics with Coupling to an External Bath, *J Chem Phys 81*, 3684-3690.

[28] Roe, D. R., and Cheatham, T. E., 3rd. (2013) PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data, *J Chem Theory Comput 9*, 3084-3095.

[29] Bi, Y., Cho, J. H., Kim, E. Y., Shan, B., Schindelin, H., and Raleigh, D. P. (2007) Rational design, structural and thermodynamic characterization of a hyperstable variant of the villin headpiece helical subdomain, *Biochemistry 46*, 7497-7505.

[30] McKnight, C. J., Matsudaira, P. T., and Kim, P. S. (1997) NMR structure of the 35-residue villin headpiece subdomain, *Nat Struct Biol 4*, 180-184.

[31] Brewer, S. H., Vu, D. M., Tang, Y., Li, Y., Franzen, S., Raleigh, D. P., and Dyer, R. B. (2005) Effect of modulating unfolded state structure on the folding kinetics of the villin headpiece subdomain, *Proc Natl Acad Sci U S A 102*, 16662-16667.

[32] Xiao, S., Patsalo, V., Shan, B., Bi, Y., Green, D. F., and Raleigh, D. P. (2013) Rational modification of protein stability by targeting surface sites leads to complicated results, *Proc Natl Acad Sci U S A 110*, 11337-11342.

[33] Glasscock, J. M., Zhu, Y., Chowdhury, P., Tang, J., and Gai, F. (2008) Using an amino acid fluorescence resonance energy transfer pair to probe protein unfolding: application to the villin headpiece subdomain and the LysM domain, *Biochemistry 47*, 11070-11076.

[34] Kubelka, J., Chiu, T. K., Davies, D. R., Eaton, W. A., and Hofrichter, J. (2006) Sub-microsecond protein folding, *J Mol Biol 359*, 546-553.

[35] Cellmer, T., Henry, E. R., Kubelka, J., Hofrichter, J., and Eaton, W. A. (2007) Relaxation rate for an ultrafast folding protein is independent of chemical denaturant concentration, *J Am Chem Soc 129*, 14564-14565.

[36] Cellmer, T., Buscaglia, M., Henry, E. R., Hofrichter, J., and Eaton, W. A. (2011) Making connections between ultrafast protein folding kinetics and molecular dynamics simulations, *Proc Natl Acad Sci U S A 108*, 6103-6108.

[37] Reiner, A., Henklein, P., and Kiefhaber, T. (2010) An unlocking/relocking barrier in conformational fluctuations of villin headpiece subdomain, *Proc Natl Acad Sci U S A 107*, 4955-4960.

[38] Neumaier, S., and Kiefhaber, T. (2014) Redefining the dry molten globule state of proteins, *J Mol Biol 426*, 2520-2528.

[39] Gelman, H., and Gruebele, M. (2014) Fast protein folding kinetics, *Q Rev Biophys 47*, 95-142.

[40] Brown, J. W., Farelli, J. D., and McKnight, C. J. (2012) On the unyielding hydrophobic core of villin headpiece, *Protein Sci 21*, 647-654.

[41] Tang, Y., Grey, M. J., McKnight, J., Palmer, A. G., 3rd, and Raleigh, D. P. (2006) Multistate folding of the villin headpiece domain, *J Mol Biol 355*, 1066-1077.

3. Selenomethionine, p-Cyanophenylalanine Pairs Provide a Convenient, Sensitive, Non-Perturbing Fluorescent Probe of Local Helical Structure

3.1. Abstract

The ability to follow secondary structure formation is critical in studies of protein folding, protein-protein interactions and amyloid formation. Circular dichroism is limited to providing global information and is usually less sensitive, in terms of signal to noise, than fluorescence. Here we demonstrate how selenomethionine, the selenium analog of methionine, can be used to follow local secondary structure formation with high sensitivity *via* selective quenching of 4-cyanophenylalanine (F_{CN}). Gai and coworkers have shown that selenomethionine is a quencher of F_{CN} fluorescence. Fluorescence quenching of F_{CN} by selenomethionine was used to detect α-helix formation in a 21-residue designed helical peptide as well as in the 36-residue villin headpiece subdomain. The F_{CN} fluorescence quenching efficiency of selenomethionine is ~80%. Histidine has been used in a similar fashion, but suffers from pH dependent effects and is an effective quencher of F_{CN} fluorescence only when the sidechain is deprotonated. We show that the quenching mechanism of selenomethionine is independent of pH, making it a useful alternative to tryptophan fluorescence quenching by histidine above pH 6.5.

Note: The material presented in this chapter has been published (Peran, I., Watson, M. D., Bilsel, O., and Raleigh, D. P. (2016) Selenomethionine, p-cyanophenylalanine pairs provide a convenient, sensitive, non-perturbing fluorescent probe of local helical structure, *Chem Commun (Camb) 52*, 2055-2058. I.P. and M.D.W were co-first authors). This chapter contains direct excerpts from the manuscript that was written by Ivan Peran and myself with additional writing and revisions by the other authors.

3.2. Introduction

Fluorescence measurements are widely employed in studies of protein dynamics, folding, stability, and aggregation.[1, 2] Trp has the highest quantum yield of the naturally occurring fluorescent residues in proteins, but its quantum yield depends on a variety of factors, and proteins often contain multiple Trp residues. Both factors can make structural interpretation of Trp fluorescence changes ambiguous. Trp-His pairs have been used to probe secondary structure and rely on the quenching of Trp fluorescence by the His sidechain.[3] However, only the protonated form of the His sidechain is an effective quencher of Trp fluorescence, limiting the approach to pH values at which the imidazole group is protonated. The covalent attachment of fluorescent dyes is another popular approach, particularly for use as Förster resonance energy transfer (FRET) pairs, however the method requires selective attachment of two dyes and often requires the introduction of Cys mutations. Furthermore, while typical dyes are very bright, they can perturb the properties of the protein of interest as they are usually built around large polyaromatic cores. In many cases FRET pairs have large R_0 values making it difficult to probe smaller local changes in structure. For these applications short range quenchers are desired.[4] A simple non-perturbing approach which involves a fluorophore that can be selectively excited and which provides easily interpreted structural information would be a useful addition to the arsenal of fluorescent methods.

We demonstrate that selenomethionine (M_{Se}) and *p*-cyanophenylalanine (F_{CN}) (Figure 3.1A) can be used as a minimally perturbative fluorescent probe of protein structure. The pair has been used to examine short oligoproline peptides.[5] F_{CN} is the cyano analogue of Tyr. The residue can be incorporated into proteins recombinantly using 21st pair technology or by solid phase peptide synthesis and represents a conservative substitution for Tyr or Phe.[6-11] F_{CN} fluorescence can be selectively excited in the presence of Tyr and Trp.[6] The quantum yield of F_{CN} is controlled

by solvation; fluorescence is high when the cyano group forms a hydrogen bond in a polar protic solvent and is low when it is buried in a hydrophobic environment.[6-12] F_{CN} fluorescence is also quenched by deprotonated His and Lys as well as *via* FRET to Tyr or Trp.[6, 7, 10, 11, 13] The fluorescence lifetime decay of free F_{CN} is single exponential and has been reported to be 7.5 ns for a G-F_{CN}-G tripeptide.[5] In contrast, the fluorescence decay of Trp is multi-exponential, even for simple peptides. M_{Se} is the selenium analogue of Met and has been widely used in multi-wavelength anomalous diffraction (MAD) phasing for X-ray crystallography and has seen some applications as an NMR probe.[14] The residue can be easily incorporated into proteins in very high yield *via* recombinant expression and is also compatible with solid phase peptide synthesis. M_{Se} quenches F_{CN} fluorescence *via* electron transfer.[5] The short range of the quenching effect suggests that F_{CN}-M_{Se} pairs could be used to design fluorescence-based probes of local secondary structure. Here we illustrate its use to monitor helical structure.

3.3. Materials and Methods

3.3.1. Peptide Synthesis and Purification

Peptides were synthesized on a 0.1 mmol scale using 9-fluoronylmethoxycarbonyl (Fmoc) chemistry on a CEM Liberty microwave peptide synthesizer. An amidated C-terminus was incorporated using a 5-(4'-Fmoc-aminomethyl-3',5-dimethoxyphenol)valeric acid (Fmoc-PAL-PEG-PS) resin. An acetylated N-terminus was incorporated into the designed helical peptide using acetic anhydride. For the 21-residue helical peptide, the C-terminal alanine and the proline were double coupled. In HP36, all β-branched amino acids along with the C-terminal Phe and the Pro were double coupled. A standard trifluoroacetic acid (TFA) protocol with thioscavengers was used to cleave the peptides from the resin. Crude peptides were dissolved in 20% acetic acid (*v*/*v*) and lyophilized.

3.3.2. Protein and Peptide Purification

Crude peptides were purified by reverse-phase high performance liquid chromatography (HPLC) using a Higgins Analytical Proto 300 C18 preparative column (10 mm X 250 mm). A two buffer gradient system was used where buffer A consisted of 0.1% TFA (*v*/*v*) in H_2O, and buffer B consisted of 9:1 acetonitrile:H_2O with 0.1% TFA (*v*/*v*). Purified dry peptides were dissolved in hexafluoro-2-propanaol (HFIP) at a concentration of 10 mg/mL and allowed to sit overnight at room temperature. This step is included to remove residual scavengers. Peptides in HFIP were diluted into 18 MΩ water and purified by HPLC using the same protocols used for the first purification step. Analytical HPLC was used to check the purity of the peptides.

3.3.3. Mass Spectrometry

Peptide and proteins were characterized by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOFMS) on a Bruker Daltonics autoflex TOF/TOF

instrument. 21-residue helical peptide expected mass: 2002.0 Da (mono) observed: 2001.0 Da (mono). W24F_{CN}/N28M_{Se}-HP36 expected mass: 4237.1 Da (mono) observed: 4237.2 Da (mono).

3.3.4. NMR Spectroscopy

W24F_{CN}/N28M_{Se}-HP36 was dissolved in a 100% D_2O solution containing 20 mM sodium acetate (pre-exchanged with D_2O) at pD 4.6 (uncorrected pH reading). The protein concentration was 750 μM. 1D ^1H NMR spectra were recorded on a Bruker 700 MHz spectrometer at 25 °C. 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) (0.5 mM) was used as an internal reference. The data was analyzed using the commercially available software package Mnova 7.

3.3.5. Circular Dichroism Spectroscopy

CD wavelength spectra were recorded on an Applied Photophysics Chirascan instrument. For wavelength scans, the dry 21-residue helical peptide was dissolved in a 10 mM sodium acetate buffer at pH 5.5 with or without 8 M urea for a final peptide concentration of 25 μM. Measurements were performed at 25 °C. Dry HP36 was dissolved in 20 mM sodium acetate buffer with or without 10 M urea at pH 5.0 for a final peptide concentration of 25 μM. Data was collected at 25 °C. The concentration of urea was determined by measuring the refractive index. The concentration of the peptides and proteins was estimated using the F_{CN} absorbance measured at 280 nm based on an extinction coefficient (ε) of 850 M^{-1} cm^{-1}. For the helical peptide, the fraction of helical structure was determined using the measured mean residue ellipticity at 222 nm, $[\theta]_{222}$, from the following expression:[15]

$$f_h = \frac{[\theta]_{222} - [\theta]_C}{[\theta]_H - [\theta]_C} \tag{3.1}$$

Where f_h is the fractional helical content of the sample, $[\theta]_H$ corresponds to the mean residue ellipticity at 222 nm for 100% helical peptide and $[\theta]_C$ corresponds to the value for a random coil:

$$[\theta]_H = -40{,}000\times\left(1-\frac{2.5}{N}\right)+100\times T \tag{3.2}$$

$$[\theta]_C = 640-45\times T \tag{3.3}$$

Where N is the number of residues in the peptide and T is the temperature in °C.

For CD monitored thermal unfolding, the HP36 variant was dissolved in 20 mM sodium acetate and 150 mM NaCl at pH 5.0 for a final peptide concentration of 25 μM. Thermal denaturation data was fit to equation (3.4)

$$\theta_{222} = \frac{a_n + b_n T + (a_d + b_d T)e^{\left(-\frac{\Delta G(T)}{RT}\right)}}{1+e^{\left(-\frac{\Delta G(T)}{RT}\right)}} \tag{3.4}$$

Where θ_{222} is the ellipticity at 222 nm in millidegrees, a_n, b_n, a_d and b_d are fitting parameters defining the pre- and post-transition baseline, T is the temperature in K, R is the gas constant in kcal mol^{-1} K^{-1}, and $\Delta G(T)$ is the change in free energy upon unfolding in kcal mol^{-1}, determined from the modified Gibbs-Helmholtz equation (3.5)

$$\Delta G(T) = \Delta H_M\left(1-\frac{T}{T_M}\right)+\Delta C_p^{\circ}\left(T-T_M-T\ln\left(\frac{T}{T_M}\right)\right) \tag{3.5}$$

Where T_M is the midpoint temperature, T is the temperature, ΔH_M is the change in enthalpy upon unfolding at T_M, and ΔC_p° is the change in heat capacity, which was set to 0.38 kcal mol^{-1} deg^{-1}, as determined by previous experiments.[16]

3.3.6. Steady-State Fluorescence

Fluorescence emission spectra were measured using a Photon Technologies fluorimeter. The same samples used to collect the CD data were used. Measurements were made at 25 °C. For pH dependent measurements, HP36 was dissolved in 20 mM phosphate buffer for a final peptide

concentration of 17 μM. F_{CN} fluorescence was excited at 240 nm and spectra were recorded from 265 to 400 nm. Experiments were performed at 20 °C.

3.3.7. Time-resolved Fluorescence

Time-resolved experiments were performed at the University of Massachusetts Medical School in Worcester, MA. 21-residue helical peptide was dissolved in 10 mM sodium acetate buffer at pH 5.5 with or without 8 M urea. HP36 was dissolved in 20 mM sodium acetate buffer at pH 5.0 with or without 10 M urea. Final peptide and protein concentrations were 25 μM. Decays were measured on a home-built time correlated single photon counting (TCSPC) apparatus. The details of the apparatus are described in the Materials and Methods section of chapter 2. All measurements were collected at 20 ±1 °C. The decay curves were analyzed by fitting to a two-exponential model reconvoluted with the measured instrument response function.[12] Decays were also fit using a maximum entropy model (MEM Laplace Inversion version 2.0c developed by Prof Bilsel).[12]

3.4. Results and Discussion

As a first test of the proposed approach we designed a 21-residue α-helical polypeptide (Figure 3.1B, C). M_{Se} is prone to oxidation, however, we did not detect significant oxidation products after incubating the peptide in buffer (Figure 3.2). Reducing agents can be used and oxygen excluded by degassing if this is an issue. The polypeptide contains an F_{CN} residue at position 16 and an M_{Se} residue at position 12. The *i*, *i*+4 separation brings the two residues into proximity in the α-helical state. Circular dichroism (CD) shows that the peptide is helical in buffer at 25 °C (Figure 3.1D) and is much less structured in 8 M urea. The mean residue ellipticity at 222 nm is -12,600 deg cm^2 $dmol^{-1}$ in buffer which corresponds to an estimated helical content of 38% based on equation (3.1). In the urea unfolded state the mean residue ellipticity at 222 nm is -1,400 deg cm^2 $dmol^{-1}$. F_{CN} fluorescence is high in the urea unfolded state, but is reduced twofold in the α-helical state (Figure 3.1E). We next conducted time-resolved fluorescence lifetime measurements. The integrated area under the fluorescence decay curve for the folded state is much smaller than for the unfolded state, consistent with quenching of the F_{CN} fluorescence in the helical state. The fluorescence decay for the folded state is fit by two components with lifetimes of 4.12 and 0.93 ns with relative amplitudes of 0.73 and 0.27, respectively (Figure 3.3). The multi-exponential decay suggests the presence of two populations with different separations between the sidechains of the two residues. Bi-exponential decays have been reported for an M_{Se}-F_{CN} dipeptide indicating the effect is likely local in origin.[5] The fluorescence decay of the urea unfolded state is fit by two components with lifetimes of 5.72 and 1.12 ns with relative amplitudes of 0.86 and 0.14, respectively. Analysis of the data using a maximum entropy approach yields very similar time constants and relative amplitudes (Figure 3.4).

We next examined the applicability of the pair to follow helix formation in a globular protein. The 36-residue villin headpiece helical subdomain (HP36) was chosen as a model system (Figure 3.5A). HP36 is a three-helix protein which contains a single Trp at position 24 in the third helix.[17-19] The helical subdomain is part of the villin protein and the numbering used here denotes the first residue in our construct as residue 1. The domain has been widely used for studies of protein folding, dynamics and stability.[20-29] We replaced Trp-24 with F_{CN} and residue 28 with M_{Se}; this construct was designated W24F_{CN}/N28M_{Se}-HP36. These sites are located on the surface of the protein on the C-terminal helix. The CD spectrum of the construct is very similar to that reported for wild-type HP36 indicating that these surface substitutions do not perturb the structure.[20] The hydrophobic core of HP36 contains three closely packed Phe residues, leading to characteristic ring current shifted resonances in the ^{1}H-NMR spectrum which is indicative of the folded state.[19, 29, 30] These peaks are observed in the ^{1}H-NMR spectrum of W24F_{CN}/N28M_{Se}-HP36, providing additional evidence that the substitutions do not perturb the fold (Figure 3.6). W24F_{CN}/N28M_{Se}-HP36, like the wild-type, exhibits a sigmoidal thermal unfolding transition (Figure 3.7). The *i*, *i*+4 spacing of residues leads to efficient quenching in the folded state, but not in the urea unfolded state. The F_{CN} fluorescence intensity is six-fold less in the folded state relative to the urea unfolded state (Figure 3.5C). The larger change in fluorescence observed for HP36 compared to the 21-residue helical peptide reflects the fact that HP36 is fully folded in buffer, while the designed helical peptide is only partially structured. Thus, a significant fraction of the helical peptide is unfolded and does not experience effective fluorescence quenching. Fluorescence lifetime studies confirm that there are significant differences in F_{CN} fluorescence between the folded and unfolded states; the integrated area under the time-resolved decay is much less for the folded state relative to the urea unfolded state. The folded state exhibits a bi-exponential decay with time constants of

4.96 and 0.72 ns. The relative amplitudes of the components are 0.68 and 0.32, respectively. The urea unfolded state also exhibits a bi-exponential decay with time constants of 6.72 and 1.21 ns, and relative amplitudes of 0.87 and 0.13, respectively (Figure 3.8). Analysis of the data using a maximum entropy approach yields very similar time constants and relative amplitudes (Figure 3.9).

In principle F_{CN} fluorescence quenching can be used to monitor thermally induced protein unfolding. However, like Trp, the quantum yield of F_{CN} is temperature dependent and decreases with increasing temperature (Figure 3.10). This will lead to potential problems with pre- and post-transition baselines, and the ability to unambiguously detect the protein unfolding transition will depend on the magnitude of the fluorescence change due to unfolding.[1, 3]

His-Trp pairs have been used as probes of α-helical structure and rely on the ability of a protonated His sidechain to quench Trp fluorescence, but this approach is limited to pH values where a significant fraction of the imidazole sidechain is protonated.[3, 31-33] The quenching of F_{CN} fluorescence by His is also pH dependent; in the case of F_{CN}, a deprotonated His sidechain is an effective quencher, but a protonated His sidechain is not.[13] This is not an issue with the M_{Se} approach, and effective quenching is observed at both high and low pH (Figure 3.11).

3.5. Conclusions

In summary, we have demonstrated that F_{CN}-M_{Se} pairs provide a simple probe of local helical structure in polypeptides and globular proteins. The F_{CN}-M_{Se} pair offers several advantages including the ability to selectively excite F_{CN} fluorescence in the presence of Tyr and Trp, the pH independent response, and the conservative nature of the substitution. In the present case, we illustrated the approach by developing local probes of α-helical structure, but the methodology could also probe β-sheet formation. For example, residues located across from each other on two adjacent β-strands will be close in the native state of a protein, but much more distant in the unfolded state and should experience a significant change in fluorescence. The approach is best suited to solvent exposed sites for the F_{CN} residue since burial of the F_{CN} sidechain can quench the fluorescence independent of any M_{Se} effect. However, the choice of a surface site also ensures that the substitution will be minimally perturbing. As noted earlier, F_{CN} fluorescence can be quenched by deprotonated His and Lys and by FRET to Tyr or Trp. Interaction with a deprotonated Lys is unlikely given its pK_a, but the presence of a His, Tyr or Trp residue that is in close proximity to the F_{CN} site should be avoided. We anticipate that the approach will be useful for studies of protein folding and dynamics, protein-protein interactions, and protein aggregation. The approach described here is complementary to the use of thioamides as fluorescence quenchers.[34, 35] Thioamides offer a quencher localized to the backbone while M_{Se} provides a sidechain based quencher. Thioamides are commonly incorporated by native chemical ligation while M_{Se} is incorporated with standard auxotrophic strains.

3.6. Figures

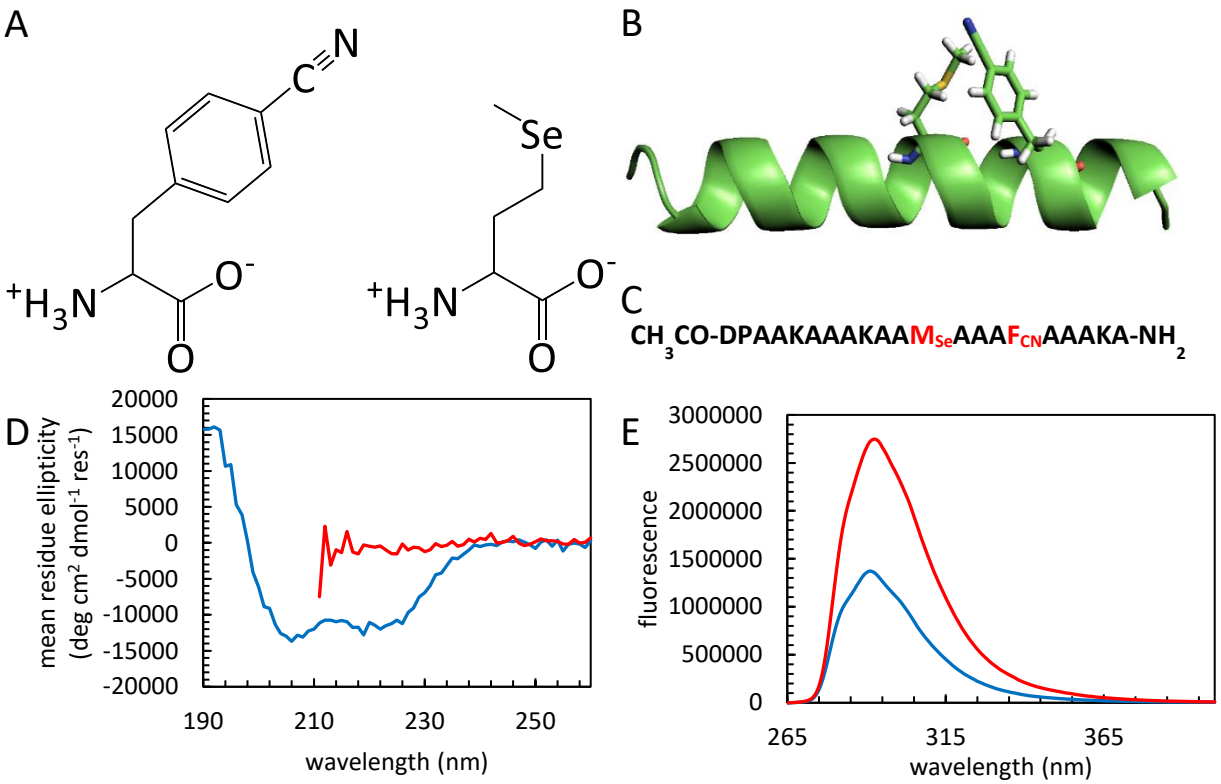


Figure 3.1. F_{CN}-M_{Se} pairs effectively detect helical structure. **(A)** Structure of *p*-cyanophenylalanine (F_{CN}) and selenomethionine (M_{Se}). **(B)** Ribbon diagram of an idealized helix showing the interaction of F_{CN} and M_{Se} where the residues are located at positions *i* and *i*+4. **(C)** Sequence of the designed helical peptide. The M_{Se} and F_{CN} residues are coloured red. **(D)** CD spectra of the peptide in buffer (blue) and in 8 M urea (red). **(E)** Fluorescence emission spectra of the α-helical state (blue) and the 8 M urea unfolded state (red). Experiments were conducted in 10 mM sodium acetate at pH 5.5 and 25 °C. The concentration of peptide in the samples was 25 µM.

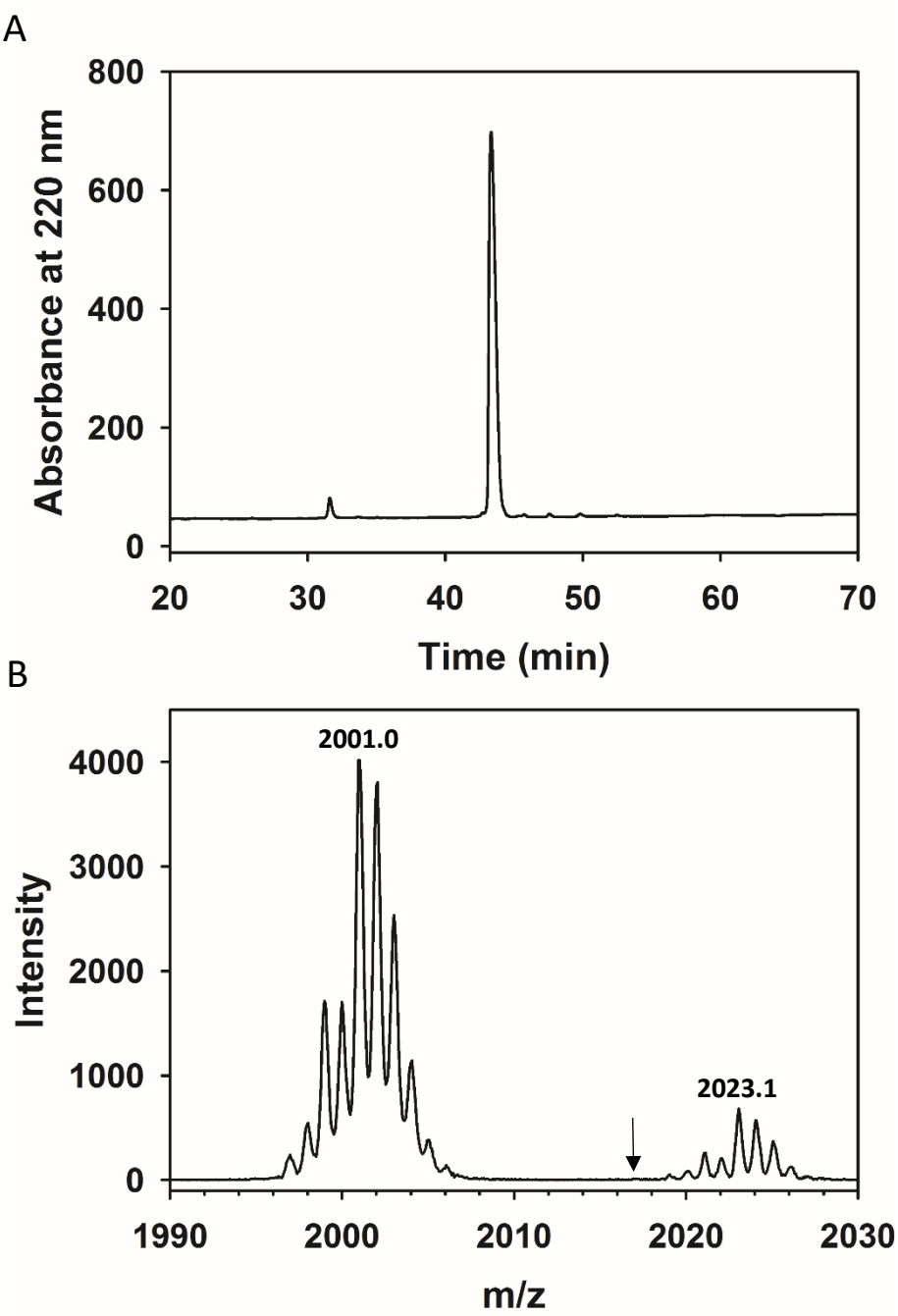


Figure 3.2. M_{Se} is not oxidized under normal experimental conditions. **(A)** Analytical HPLC trace of the helical peptide after 72 hours in buffer. A single peak is observed with a retention time of 43.3 minutes which corresponds to unmodified material. No additional peptide peak is observed. **(B)** MALDI-TOFMS spectrum of the helical peptide after 72 hours in buffer. The arrow indicates where a peak would appear if the selenomethionine was oxidized. The set of peaks with lower intensity centered near 2023.1 Da are the sodium adduct. The peptide was incubated in 10 mM sodium acetate at pH 5.5 and room temperature.

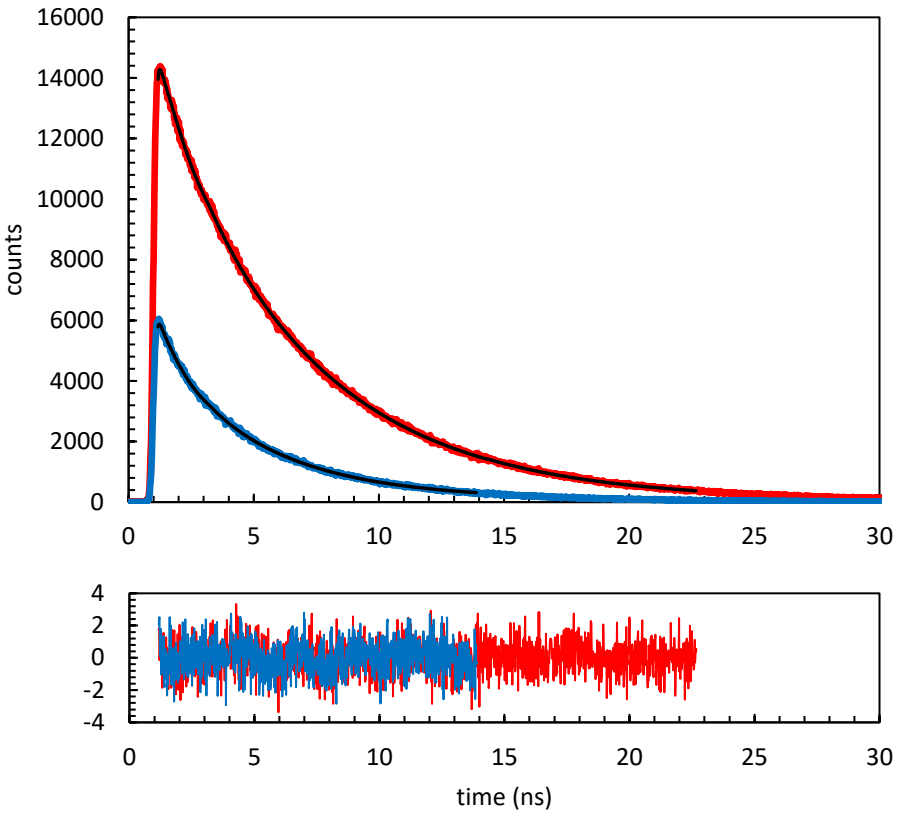


Figure 3.3. Multiexponential fitting of time-resolved fluorescence decays for the 21-residue helical peptide. Time-resolved fluorescence decays for the α-helical state (blue) and the urea unfolded state (red). Decays were fit using two exponentials. The residuals are also plotted.

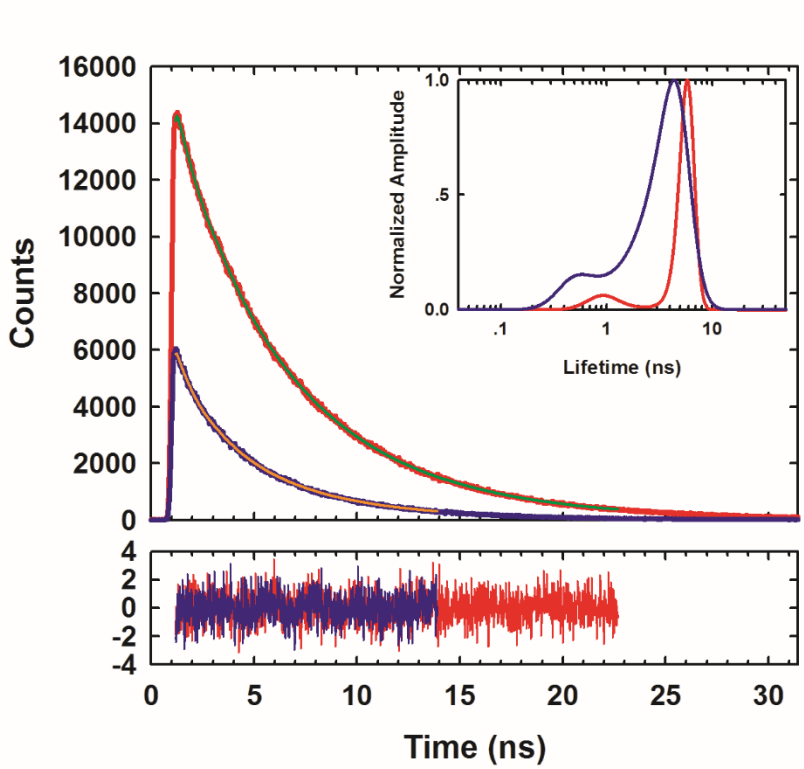


Figure 3.4. Maximum entropy analysis of time-resolved fluorescence decays for the 21-residue helical peptide. Time-resolved fluorescence decays for peptide in buffer (blue) and in 8 M urea (red) fit using a maximum entropy model. For the helical state in buffer the fits yield time constants of 4.37 ns and 0.59 ns for the slow and fast components, respectively, with relative amplitudes of 0.87 and 0.13, respectively. In 8 M urea, the fit yields time constants of 5.84 and 0.93 ns for the slow and fast components, respectively, and relative amplitudes are 0.94 and 0.06, respectively. Shown in the inset is the lifetime distribution. Residuals are plotted below.

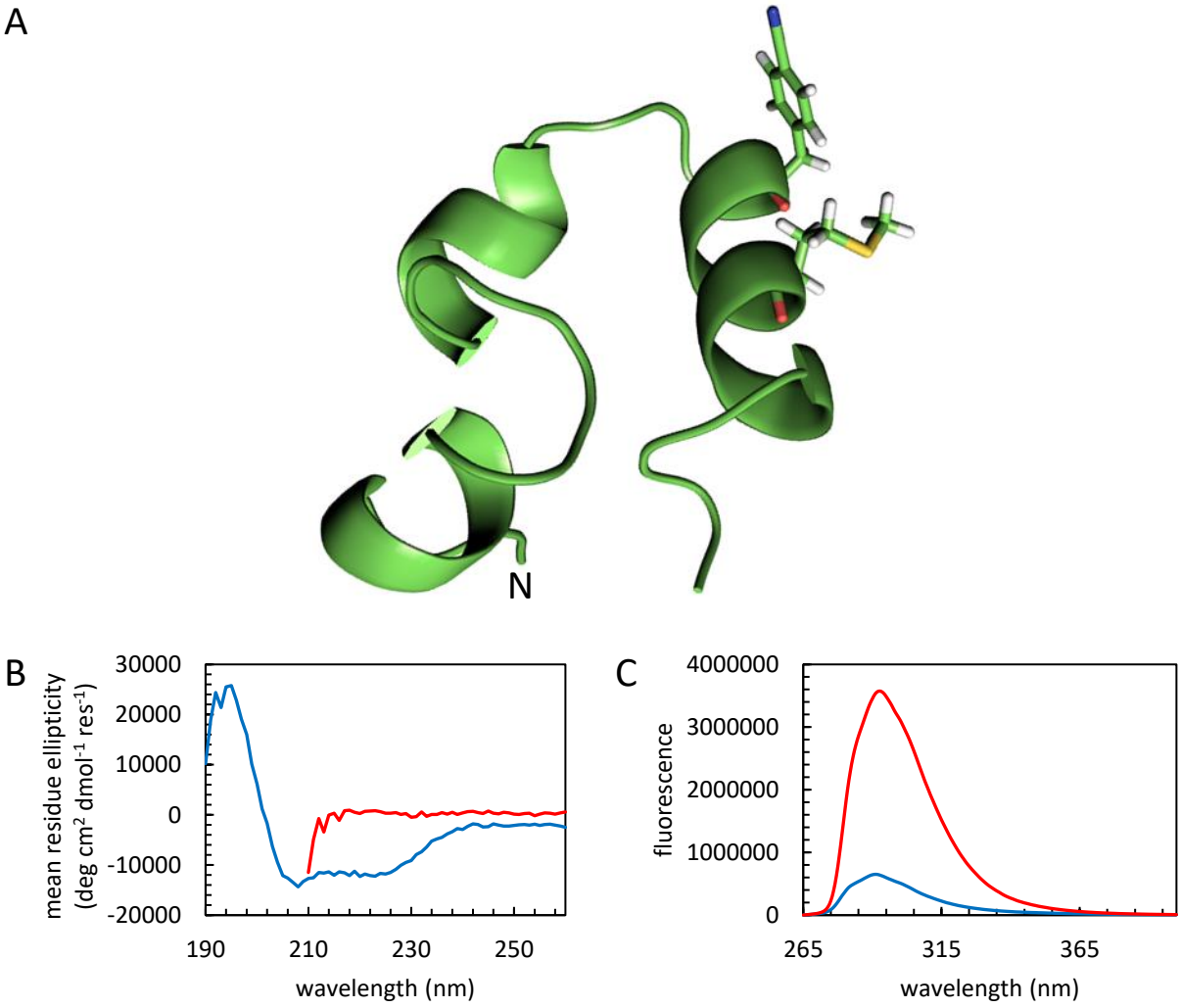


Figure 3.5. F_{CN}-M_{Se} pairs monitor the folding of helical proteins. **(A)** Ribbon diagram of HP36 showing the location of the F_{CN} and M_{Se} residues. The N-terminus is labelled. **(B)** CD spectra of the protein in buffer (blue) and in 10 M urea (red). **(C)** Fluorescence emission spectra of the protein in buffer (blue) and in 10 M urea (red). Experiments were performed in 20 mM sodium acetate at pH 5.0 and 25 °C. Protein concentration was 25 µM. The model was constructed using the parameters for Met.

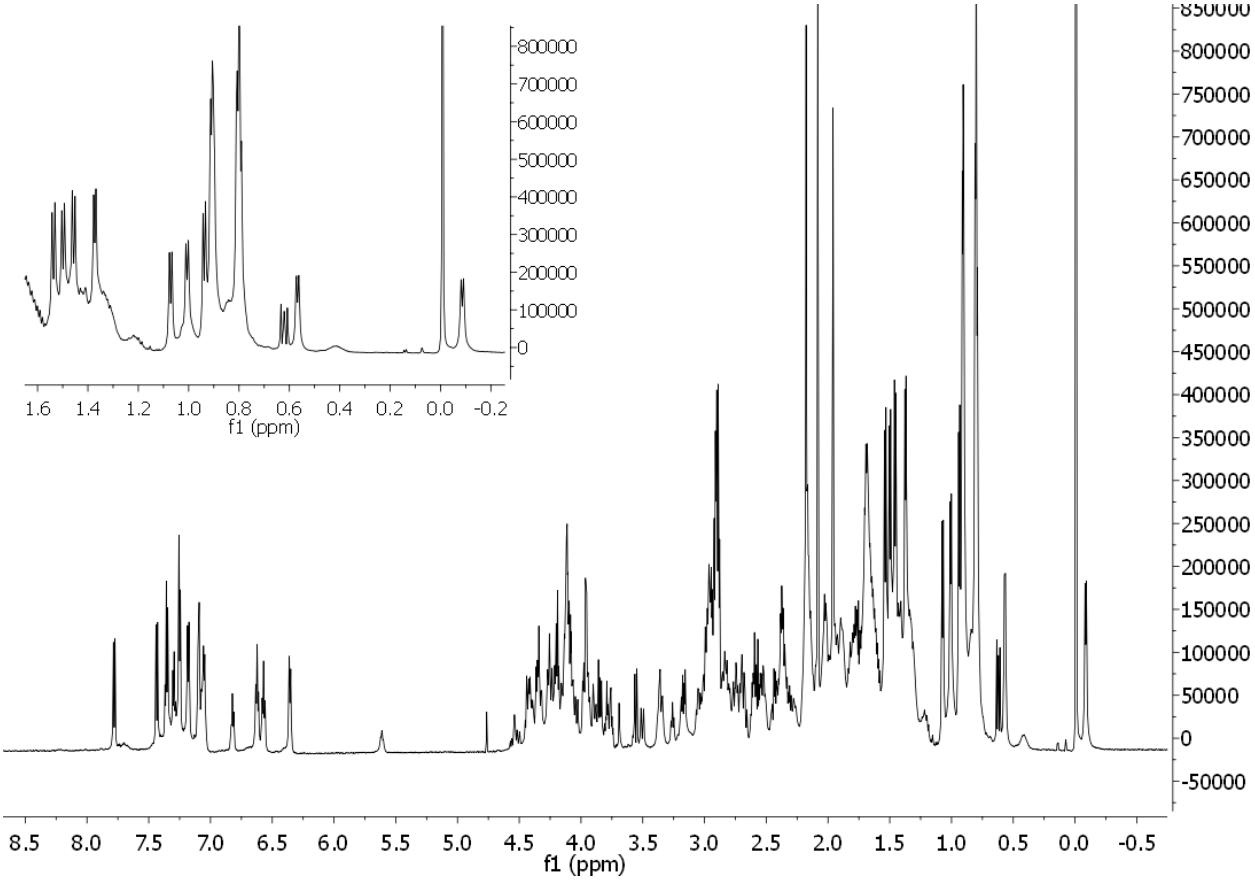


Figure 3.6. 1D ^{1}H-NMR spectrum of the M_{Se}, F_{CN} variant of HP36. The spectrum was recorded in D_2O at pD 4.6 (uncorrected pH reading) in 20 mM sodium acetate (pre-exchanged with D_2O) at 25 °C. The sharp peak at 0.00 ppm is the chemical shift standard. Note the characteristic ring current shifted methyl resonances and the distinctive up-field aromatic resonance. These peaks are indicative of the HP36 fold. The inset shows the upfield methyl resonances. We did not detect any resolved ^{77}Se-^{1}H J-couplings in the one-dimensional proton spectrum, but this is not surprising given the resolution of the one-dimensional experiment.

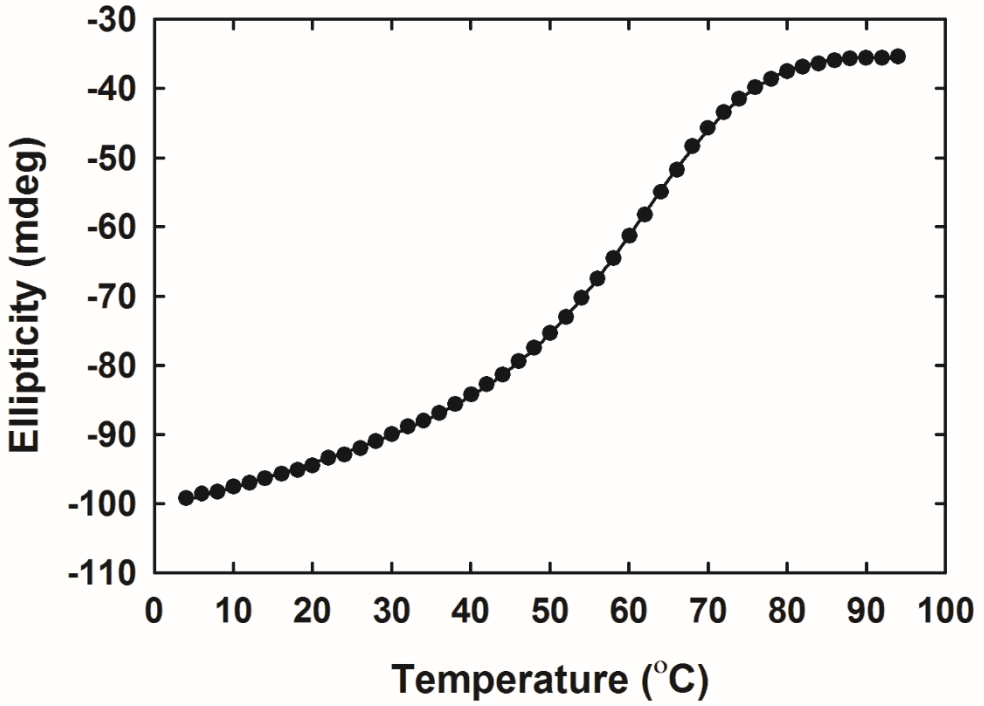


Figure 3.7. CD monitored thermal unfolding of the M_{Se}, F_{CN} variant of HP36 recorded in 20 mM sodium acetate and 150 mM NaCl at pH 5.0. The solid line is the best fit for a two-state unfolding transition (Equation 3.4). The T_M of the M_{Se} F_{CN} variant is 62 °C under these conditions. The reported T_M for the synthetic wild-type sequence with an amidated C-terminus is 70.5 °C in 10 mM phosphate.[36] Prior studies have shown that replacing Trp-24 with Leu or Ala destabilized the domain by 12 to 16 °C, so the change observed with the $W24F_{CN}$, $N24M_{Se}$ mutant is relatively modest.[37]

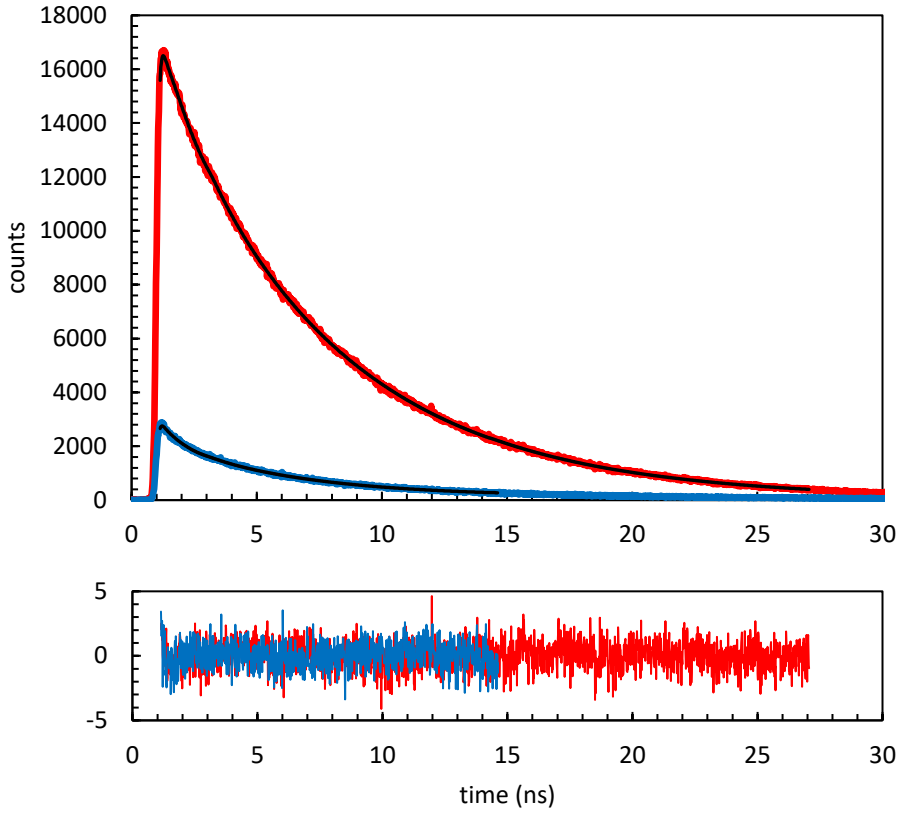


Figure 3.8. Multiexponential fitting of time-resolved fluorescence decays for W24F_{CN}/N28M_{Se}-HP36. Time-resolved fluorescence decays for the protein in buffer (blue) and in 10 M urea (red). Decays were fit using two exponentials. Residuals are plotted below.

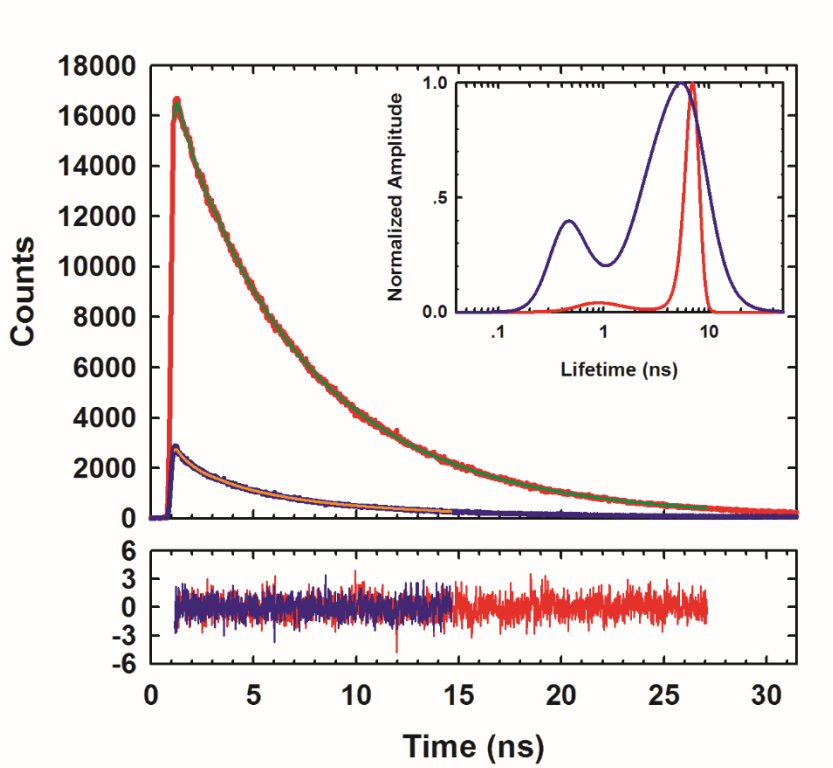


Figure 3.9. Maximum entropy analysis of time-resolved fluorescence decays for $W24F_{CN}/N28M_{Se}$-HP36. Time-resolved fluorescence decays for the villin helical headpiece subdomain (HP36) protein in buffer (blue) and in 10 M urea (red) fit using a maximum entropy model. For the folded protein in buffer, the fit gave time constants of 5.38 ns for the slow component and 0.47 ns for the fast component with relative amplitudes of 0.72 and 0.28, respectively. For unfolded protein in 10 M urea, the time constants are 6.91 ns for the slow component and 0.89 for the fast component with relative amplitudes of 0.96 and 0.04, respectively. Shown in the inset is the lifetime distribution. Residuals are plotted below.

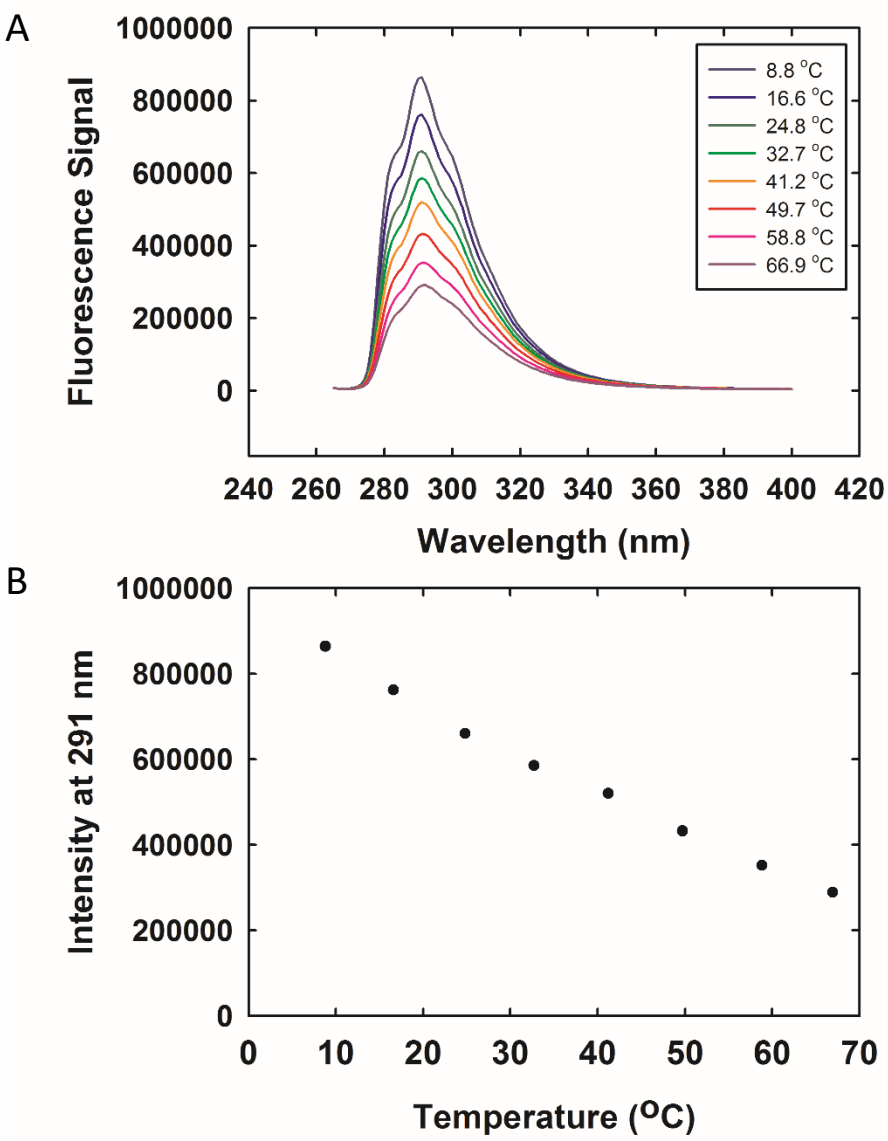


Figure 3.10. Temperature dependence of *p*-cyanophenylalanine fluorescence. **(A)** Fluorescence emission spectra of the free amino acid at different temperatures. **(B)** Plot of the intensity at the emission maximum versus temperature. Experiments were performed in 10 mM sodium acetate at pH 5.5 and 20 °C. The concentration of F_{CN} was 25 µM.

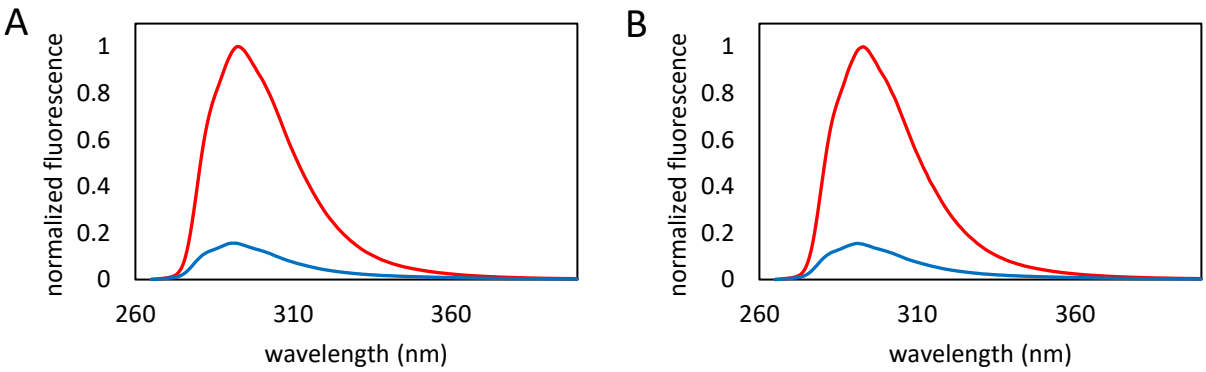


Figure 3.11. M_{Se} quenching of F_{CN} is pH independent. Fluorescence emission spectra in buffer (blue) and in 10 M urea (red) at **(A)** pH 5.0 **(B)** pH 8.5. Experiments were conducted in 20 mM sodium phosphate at 20 °C. Protein concentration was 17 µM.

3.7. References

[1] Lakowicz, J. R. (2006) *Principles of Fluorescence Spectroscopy*, 3 ed., Springer.

[2] Brown, M. P., and Royer, C. (1997) Fluorescence spectroscopy as a tool to investigate protein interactions, *Curr Opin Biotech 8*, 45-49.

[3] Kubelka, J., Eaton, W. A., and Hofrichter, J. (2003) Experimental tests of villin subdomain folding simulations, *J Mol Biol 329*, 625-630.

[4] Sahoo, H., Roccatano, D., Zacharias, M., and Nau, W. M. (2006) Distance distributions of short polypeptides recovered by fluorescence resonance energy transfer in the 10 A domain, *J Am Chem Soc 128*, 8118-8119.

[5] Mintzer, M. R., Troxler, T., and Gai, F. (2015) p-Cyanophenylalanine and selenomethionine constitute a useful fluorophore-quencher pair for short distance measurements: application to polyproline peptides, *Phys Chem Chem Phys 17*, 7881-7887.

[6] Tucker, M. J., Oyola, R., and Gai, F. (2006) A novel fluorescent probe for protein binding and folding studies: p-cyano-phenylalanine, *Biopolymers 83*, 571-576.

[7] Tucker, M. J., Tang, J., and Gai, F. (2006) Probing the kinetics of membrane-mediated helix folding, *J Phys Chem B 110*, 8105-8109.

[8] Aprilakis, K. N., Taskent, H., and Raleigh, D. P. (2007) Use of the novel fluorescent amino acid p-cyanophenylalanine offers a direct probe of hydrophobic core formation during the folding of the N-terminal domain of the ribosomal protein L9 and provides evidence for two-state folding, *Biochemistry 46*, 12308-12313.

[9] Wang, L., Xie, J., and Schultz, P. G. (2006) Expanding the genetic code, *Annu Rev Biophys Biomol Struct 35*, 225-249.

[10] Miyake-Stoner, S. J., Miller, A. M., Hammill, J. T., Peeler, J. C., Hess, K. R., Mehl, R. A., and Brewer, S. H. (2009) Probing protein folding using site-specifically encoded unnatural amino acids as FRET donors with tryptophan, *Biochemistry 48*, 5953-5962.

[11] Rogers, J. M., Lippert, L. G., and Gai, F. (2010) Non-natural amino acid fluorophores for one- and two-step fluorescence resonance energy transfer applications, *Anal Biochem 399*, 182-189.

[12] Serrano, A. L., Troxler, T., Tucker, M. J., and Gai, F. (2010) Photophysics of a Fluorescent Non-natural Amino Acid: p-Cyanophenylalanine, *Chem Phys Lett 487*, 303-306.

[13] Taskent-Sezgin, H., Marek, P., Thomas, R., Goldberg, D., Chung, J., Carrico, I., and Raleigh, D. P. (2010) Modulation of p-cyanophenylalanine fluorescence by amino acid side chains and rational design of fluorescence probes of alpha-helix formation, *Biochemistry 49*, 6290-6295.

[14] Schaefer, S. A., Dong, M., Rubenstein, R. P., Wilkie, W. A., Bahnson, B. J., Thorpe, C., and Rozovsky, S. (2013) (77)Se enrichment of proteins expands the biological NMR toolbox, *J Mol Biol 425*, 222-231.

[15] Kuhlman, B., Yang, H. Y., Boice, J. A., Fairman, R., and Raleigh, D. P. (1997) An exceptionally stable helix from the ribosomal protein L9: implications for protein folding and stability, *J Mol Biol 270*, 640-647.

[16] Bi, Y. (2008) PhD Thesis, In *Department of Chemistry*, Stony Brook University, Stony Brook, NY 11794-3400.

[17] Bi, Y., Cho, J. H., Kim, E. Y., Shan, B., Schindelin, H., and Raleigh, D. P. (2007) Rational design, structural and thermodynamic characterization of a hyperstable variant of the villin headpiece helical subdomain, *Biochemistry 46*, 7497-7505.

[18] Chiu, T. K., Kubelka, J., Herbst-Irmer, R., Eaton, W. A., Hofrichter, J., and Davies, D. R. (2005) High-resolution x-ray crystal structures of the villin headpiece subdomain, an ultrafast folding protein, *Proc Natl Acad Sci U S A 102*, 7517-7522.

[19] McKnight, C. J., Matsudaira, P. T., and Kim, P. S. (1997) NMR structure of the 35-residue villin headpiece subdomain, *Nat Struct Biol 4*, 180-184.

[20] Brewer, S. H., Vu, D. M., Tang, Y., Li, Y., Franzen, S., Raleigh, D. P., and Dyer, R. B. (2005) Effect of modulating unfolded state structure on the folding kinetics of the villin headpiece subdomain, *Proc Natl Acad Sci U S A 102*, 16662-16667.

[21] Xiao, S., Patsalo, V., Shan, B., Bi, Y., Green, D. F., and Raleigh, D. P. (2013) Rational modification of protein stability by targeting surface sites leads to complicated results, *Proc Natl Acad Sci U S A 110*, 11337-11342.

[22] Glasscock, J. M., Zhu, Y., Chowdhury, P., Tang, J., and Gai, F. (2008) Using an amino acid fluorescence resonance energy transfer pair to probe protein unfolding: application to the villin headpiece subdomain and the LysM domain, *Biochemistry 47*, 11070-11076.

[23] Kubelka, J., Chiu, T. K., Davies, D. R., Eaton, W. A., and Hofrichter, J. (2006) Sub-microsecond protein folding, *J Mol Biol 359*, 546-553.

[24] Cellmer, T., Henry, E. R., Kubelka, J., Hofrichter, J., and Eaton, W. A. (2007) Relaxation rate for an ultrafast folding protein is independent of chemical denaturant concentration, *J Am Chem Soc 129*, 14564-14565.

[25] Cellmer, T., Buscaglia, M., Henry, E. R., Hofrichter, J., and Eaton, W. A. (2011) Making connections between ultrafast protein folding kinetics and molecular dynamics simulations, *Proc Natl Acad Sci U S A 108*, 6103-6108.

[26] Reiner, A., Henklein, P., and Kiefhaber, T. (2010) An unlocking/relocking barrier in conformational fluctuations of villin headpiece subdomain, *Proc Natl Acad Sci U S A 107*, 4955-4960.

[27] Neumaier, S., and Kiefhaber, T. (2014) Redefining the dry molten globule state of proteins, *J Mol Biol 426*, 2520-2528.

[28] Gelman, H., and Gruebele, M. (2014) Fast protein folding kinetics, *Q Rev Biophys 47*, 95-142.

[29] Brown, J. W., Farelli, J. D., and McKnight, C. J. (2012) On the unyielding hydrophobic core of villin headpiece, *Protein Sci 21*, 647-654.

[30] Wang, M., Tang, Y., Sato, S., Vugmeyster, L., McKnight, C. J., and Raleigh, D. P. (2003) Dynamic NMR line-shape analysis demonstrates that the villin headpiece subdomain folds on the microsecond time scale, *J Am Chem Soc 125*, 6032-6033.

[31] Chen, Y., and Barkley, M. D. (1998) Toward understanding tryptophan fluorescence in proteins, *Biochemistry 37*, 9976-9982.

[32] VanGilst, M., and Hudson, B. S. (1996) Histidine-tryptophan interactions in T4 lysozyme: 'Anomalous' pH dependence of fluorescence, *Biophys Chem 63*, 17-25.

[33] Willaert, K., and Engelborghs, Y. (1991) The Quenching of Tryptophan Fluorescence by Protonated and Unprotonated Imidazole, *Eur Biophys J 20*, 177-182.

[34] Wissner, R. F., Batjargal, S., Fadzen, C. M., and Petersson, E. J. (2013) Labeling proteins with fluorophore/thioamide Forster resonant energy transfer pairs by combining unnatural amino acid mutagenesis and native chemical ligation, *J Am Chem Soc 135*, 6529-6540.

[35] Goldberg, J. M., Batjargal, S., and Petersson, E. J. (2010) Thioamides as fluorescence quenching probes: minimalist chromophores to monitor protein dynamics, *J Am Chem Soc 132*, 14718-14720.

[36] Tang, Y., Grey, M. J., McKnight, J., Palmer, A. G., 3rd, and Raleigh, D. P. (2006) Multistate folding of the villin headpiece domain, *J Mol Biol 355*, 1066-1077.

[37] Xiao, S., and Raleigh, D. P. (2010) A critical assessment of putative gatekeeper interactions in the villin headpiece helical subdomain, *J Mol Biol 401*, 274-285.

4. A Non-Perturbing Probe of Coiled Coil Formation Based on Electron Transfer Mediated Fluorescence Quenching

4.1. Abstract

Coiled coils are abundant in nature, occurring in ~3% of proteins across sequenced genomes and are found in proteins ranging from transcription factors to structural proteins. The motif continues to be an important model system for understanding protein-protein interactions and is finding increased use in bio-inspired materials and in synthetic biology. Knowledge of the thermodynamics of self-assembly, particularly the dissociation constant K_D, is essential for the application of designed coiled coils and for understanding the in vivo specificity of natural coiled coils. Standard methods for measuring K_D typically rely on concentration dependent circular dichroism (CD). Fluorescence methods are an attractive alternative, however Trp is rarely found in an interior position of a coiled coil and appending unnatural fluorophores can perturb the system. We demonstrate a simple, non-perturbing method to monitor coiled coil formation using p-cyanophenylalanine (F_{CN}) and selenomethionine (M_{Se}), the Se analogue of Met. F_{CN} fluorescence can be selectively excited and is effectively quenched by electron transfer with M_{Se}. Both F_{CN} and M_{Se} represent minimally perturbing substitutions in coiled coils. M_{Se} quenching of F_{CN} fluorescence is shown to offer a non-perturbing method for following coiled coil formation and for accurately determining dissociation constants. The method is validated using a designed heterodimeric coiled coil. The K_D deduced by fluorescence monitored titration is in excellent agreement with the value deduced from concentration dependent CD measurements to within the uncertainty of the measurement. However, the fluorescence approach requires less protein, is less

time consuming, can be applied to lower concentrations and could be applied to high throughput screens.

Note: The material presented in this chapter has been published (Watson, M. D., Peran, I., and Raleigh, D. P. (2016) A Non-Perturbing Probe of Coiled Coil Formation Based on Electron Transfer Mediated Fluorescence Quenching *Biochemistry 55*, 3685-3691). This chapter contains direct excerpts from the manuscript that was written by me with additional writing and revisions by the other authors.

4.2. Introduction

Protein-protein interactions are fundamental to the understanding of both normal protein function and many deleterious processes associated with disease. The coiled coil is one of the most common and important oligomerization motifs found in biology. This well-defined and extensively characterized fold comprises two or more α-helices twisted into a left-handed superhelix stabilized by a hydrophobic core.[1] Coiled coils are abundant in nature, occurring on average in ~3% of proteins across sequenced genomes and are found in proteins ranging from transcription factors to structural proteins.[2] They continue to be important model systems for understanding protein-protein interaction motifs, have been the focus of a large body of work on *de novo* protein design and are finding increased use in bio-inspired materials and synthetic biology.[3-9]

Coiled coil systems are also useful as test systems for the development of techniques for the study of protein-protein interactions. The system lends itself particularly well to the development of methods for non-invasive determination of dissociation constants (K_D), with the extent of association controllable by temperature, concentration of the individual peptides or the presence of denaturing agents.[8] Determination of K_D is also an important concern in the *de novo* design of coiled coil systems, which are an attractive target for functional protein engineering.[4, 5]

Current methods for determining the K_D of coiled coils rely on the use of circular dichroism (CD); either by globally fitting thermal melting curves or by using the midpoint of the thermal unfolding transition (T_M) at several concentrations to determine the K_D.[10, 11] These methods are time consuming, relying on multiple melting experiments with precise control of concentration and peptide ratio required. They are also difficult to implement in a high throughput fashion and are not well suited for systems which do not undergo reversible thermal unfolding. Furthermore,

the sensitivity of CD limits the lowest concentration that can be studied. Fluorescence determination of K_D is the standard in protein-protein and protein-small molecule binding studies and could offer better ease of use and potentially higher accuracy than current methods for coiled coil systems. The potential for improvements in accuracy, ease of implementation and sensitivity make the application of fluorescence-based methods to coiled coil binding studies an attractive prospect, but current methods unfortunately suffer from drawbacks.

Tryptophan fluorescence is the obvious choice for an intrinsic fluorophore, owing to its relatively high quantum yield and sensitivity to the local environment. Trp fluorescence is widely used in studies of protein folding and protein-protein interactions as burial of a Trp sidechain leads to a shift in the emission maximum and often changes in quantum yield. However, the polyaromatic side chain of Trp is quite large relative to the coiled coil interface and may not be well-tolerated in these systems. Indeed, Trp is rarely found at interior positions in coiled coils.[12] Large, bright fluorophores such as Alexa dyes can be attached to the termini of coiled coils and assembly monitored by fluorescence resonance energy transfer (FRET), but are likely to perturb the system through aromatic-aromatic and hydrophobic interactions.

A smaller, unnatural fluorescent amino acid *p*-cyanophenylalanine (F_{CN}) offers a promising alternative. F_{CN} is structurally analogous to Tyr, with the hydroxyl group replaced by a nitrile functionality. The amino acid is smaller and hence less likely to perturb the coiled coil interface than tryptophan, and has a quantum yield and excitation and emission maxima that make it attractive for protein fluorescence applications. F_{CN} can be inserted into proteins either recombinantly, using an evolved aminoacyl-tRNA synthetase/tRNA pair or by solid phase peptide synthesis using standard Fmoc chemistry.[13-15] F_{CN} has two absorbance maxima at 233 and 280 nm, allowing it to be selectively excited in the presence of Trp or Tyr at 240 nm, which corresponds to

a minimum of Trp absorbance.[14, 15] F_{CN} fluorescence is modulated by solvation, increasing when the cyano group forms a hydrogen bond with the solvent and decreasing in aprotic, hydrophobic environments. F_{CN} is not expected to make a significant contribution to the CD signal at 222 nm and is likely to be less of a concern than Trp in this regard, as Trp has a more intense absorbance at 220 nm.[14]

Recent work has shown that F_{CN} fluorescence is strongly quenched by selenomethionine (M_{Se}), the selenium analogue of methionine, through an electron transfer mechanism.[16, 17] The sensitivity and short range of electron transfer mediated fluorescence quenching make it a valuable tool for protein structure studies, and Trp-His pairs have been extensively used in this manner.[18] However, His quenching is pH sensitive and requires the imidazole ring to be in the protonated state. In contrast, the quenching of F_{CN} by M_{Se} is independent of pH, extending the utility of the pair beyond the pH range accessible by protonated His.[16] Robust methods for the recombinant incorporation of M_{Se} have been developed because of its use in multi-wavelength anomalous dispersion (MAD) phasing in X-ray crystallography.[19] M_{Se} can also be incorporated into proteins using standard solid phase peptide synthesis methods. In this work we demonstrate the utility of the F_{CN}-M_{Se} pair to detect coiled coil formation in a sensitive, non-perturbing fashion and to measure the K_D of the interaction.

4.3. Materials and Methods

4.3.1. Peptide Synthesis and Purification

Peptides were synthesized using standard 9-fluorenylmethyloxycarbonyl (Fmoc) chemistry on a CEM Liberty microwave peptide synthesizer. The C-terminus of each peptide was amidated through the use of 5-(4'-Fmoc-aminomethyl-3',5-dimethoxyphenol)valeric acid (Fmoc-PAL-PEG-PS) resin. The N-termini of peptides were acetylated with acetic anhydride. Peptides were cleaved from resin in a cleavage cocktail of 5% thioanisole (*v*/*v*), 3.3% anisole (*v*/*v*) and 3% 1,2-ethanedithiol (*v*/*v*) in trifluoroacetic acid (TFA), filtered to remove resin and precipitated in cold ether. Peptides were pelleted by centrifugation at 10,000 rcf for 10 min, the supernatant was discarded, peptides were solubilized in 20% acetic acid (*v*/*v*) and lyophilized. Peptides were redissolved in 0.1% TFA (*v*/*v*) in 18 MΩ H_2O and purified by reverse-phase high performance liquid chromatography (HPLC) on a Higgins Analytical Proto 300 C18 10 μm preparative column (250x20 mm). A two buffer gradient system was used where buffer A was 0.1% TFA (*v*/*v*) in 18 MΩ H_2O and buffer B was 0.1% TFA (*v*/*v*) in 9:1 acetonitrile:H_2O. Purity of the peptides was checked by HPLC on a Higgins Analytical Proto 300 C18 5 μm analytical column (250x4.6 mm).

4.3.2. Mass Spectrometry

Peptides were characterized by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOFMS) on a Bruker Daltonics autoflex TOF/TOF instrument. CC-$19F_{CN}A_N^{3.5}$ expected mass: 2907.588 Da (mono) observed: 2908.553 Da (mono). CC-$19WA_N^{3.5}$ expected mass: 2922.521 Da (mono) observed: 2923.394 Da (mono). CC-$20M_{Se}B_N^{3.5}$ expected mass: 2958.726 Da (mono) observed: 2959.409 Da (mono).

4.3.3. CD Measurements

CD wavelength scans were acquired on an Applied Photophysics Chirascan circular dichroism spectrometer. Measurements were performed in a 1 mm pathlength quartz cuvette from 190-260 nm in buffer or from 210-260 nm in urea using a 1 nm step size. Spectra were recorded as the average of three scans with an averaging time of 0.5 s. Thermal denaturation curves were collected at 222 nm using a 1 mm pathlength quartz cuvette for the 200 and 100 μM samples, and a 10 mm pathlength quartz cuvette for the 50, 20 and 10 μM samples. The temperature was ramped from 2-94 °C in 2 °C steps with 120 s equilibration times and a 60 s averaging time. Thermal unfolding data was fit to equation (4.1).

$$\theta_{222} = \frac{a_n + b_n T + (a_d + b_d T) e^{\left(-\frac{\Delta G(T)}{RT}\right)}}{1 + e^{\left(-\frac{\Delta G(T)}{RT}\right)}} \tag{4.1}$$

Where θ_{222} is the ellipticity at 222 nm in millidegrees, a_n, b_n, a_d and b_d are fitting parameters defining the pre- and post-transition baseline, T is the temperature in K, R is the gas constant in kcal mol^{-1} K^{-1}, and $\Delta G(T)$ is the change in free energy upon unfolding in kcal mol^{-1}, determined from the modified Gibbs-Helmholtz equation (4.2).

$$\Delta G(T) = \Delta H_M \left(1 - \frac{T}{T_M}\right) + \Delta C_p^\circ \left(T - T_M - T \ln\left(\frac{T}{T_M}\right)\right) \tag{4.2}$$

Where $\Delta G(T)$ is the change in free energy upon unfolding in kcal mol^{-1}, ΔH_M is the change in enthalpy upon unfolding at T_M in kcal mol^{-1}, T is the temperature in K, T_M is the midpoint of the thermal unfolding transition in K and ΔC_p° is the change in heat capacity upon unfolding in kcal mol^{-1} K^{-1}.

4.3.4. Fluorescence Measurements

Fluorescence emission spectra were acquired on a Photon Technologies International fluorimeter using a 10x10 mm quartz cuvette and a slit width of 1 mm (F_{CN} experiments) or 0.9

mm (Trp experiments). Spectra were recorded as the average of two scans (dimerization experiments) or four scans (K_D determination experiments) from 260-400 nm with an excitation wavelength of 240 nm (F_{CN} experiments) or from 300-450 nm with an excitation wavelength of 280 nm (Trp experiments), a step size of 1 nm and an averaging time of 0.5 s.

4.3.5. Determination of Dissociation Constant by Circular Dichroism

A dimer sample solution of CC-19$F_{CN}A_N^{3.5}$ and CC-20$M_{Se}B_N^{3.5}$ (100 μM each) in phosphate buffer (10 mM, pH 7.4) was prepared by mixing solutions of CC-19$F_{CN}A_N^{3.5}$ (200 μM, determined by absorbance at 280 nm) and CC-20$M_{Se}B_N^{3.5}$ (200 μM, determined by absorbance at 280 nm) in phosphate buffer. Additional dimer sample solutions at 100, 50, 20 and 10 μM total peptide concentration were prepared by serial dilution with phosphate buffer. The value of K_D was determined using the method of Marky and Breslauer as implemented for coiled coils by Woolfson and coworkers.[10, 11] T_M was measured as a function of total peptide concentration [*CC*] (note that [*CC*] denotes the total peptide concentration, not the concentration of the dimer) by fitting individual thermal melts to equation (4.1). Using the van 't Hoff equation leads to a linear relationship between $1/T_M$ and [*CC*], equation (4.3).

$$\frac{1}{T_M} = \frac{R}{\Delta H^\circ}\ln[CC] + \frac{\Delta S^\circ - R\ln 4}{\Delta H^\circ} \qquad (4.3)$$

Where T_M is the midpoint of the thermal unfolding transition in K, ΔS° is the change in entropy upon unfolding in kcal mol^{-1} K^{-1}, R is the gas constant in kcal mol^{-1} K^{-1}, [*CC*] is the total concentration of coiled coil peptides in M and ΔH° is the change in enthalpy upon unfolding in kcal mol^{-1}. At T_M the fraction of molecules which are folded is 0.5, hence

$$\frac{1}{K_D^M} = \frac{4}{[CC]} \qquad (4.4)$$

Where K_D^M is the K_D at T_M. These relationships allow one to determine the total concentration of peptide, and thus the K_D, that would give a T_M equal to the temperature of interest.

4.3.6. Determination of Dissociation Constant by Fluorescence

A sample solution of CC-19$F_{CN}A_N^{3.5}$ (412 nM) was prepared by diluting a stock solution of CC-19$F_{CN}A_N^{3.5}$ in H_2O (82 μM, determined by absorbance at 280 nm) with phosphate buffer (10 mM, pH 7.4). A dimer solution of CC-19$F_{CN}A_N^{3.5}$ and CC-20$M_{Se}B_N^{3.5}$ (403 nM and 64 μM, respectively) was prepared by diluting a stock solution of CC-20$M_{Se}B_N^{3.5}$ in H_2O (2987 μM, determined by absorbance at 280 nm) with the solution of CC-19$F_{CN}A_N^{3.5}$ (412 nM) in phosphate buffer. Additional dimer sample solutions were prepared at CC-20$M_{Se}B_N^{3.5}$ concentrations of 32000, 16000, 8000, 4000, 2000, 1000, 512, 256, 128, 64, 32, 16, 8, 4, 2 and 1 nM by serial dilution with the 412 nM CC-19$F_{CN}A_N^{3.5}$ solution. The slight increase from 403 to 412 nM in the CC-19$F_{CN}A_N^{3.5}$ concentration over the course of the dilution was corrected for in the fitting procedure. The value of K_D was determined by plotting the fluorescence intensity at 291 nm versus the concentration of CC-20$M_{Se}B_N^{3.5}$ and fitting to equation (4.5).

$$F_{291} = m[Tyr] + F_{max} + (F_{min} - F_{max})\left(\frac{([M_{Se}]+[F_{CN}]+K_D)-\sqrt{([M_{Se}]+[F_{CN}]+K_D)^2-4[F_{CN}][M_{Se}]}}{2[F_{CN}]}\right) \quad (4.5)$$

Where F_{291} is the fluorescence intensity at 291 nm, m is the concentration dependence of the fluorescence intensity of tyrosine at 291 nm, $[Tyr]$ is the concentration of tyrosine in nM (equal to the concentration of CC-20$M_{Se}B_N^{3.5}$ in the present case), F_{max} is the maximal fluorescence intensity at 291 nm, F_{min} is the minimum fluorescence intensity at 291 nm, $[M_{Se}]$ is the concentration of CC-20$M_{Se}B_N^{3.5}$ in nM, $[F_{CN}]$ is the concentration of CC-19$F_{CN}A_N^{3.5}$ in nM, and K_D is the dissociation constant of the coiled coil dimer in nM. The term $m[Tyr]$ is a correction for the contribution of the Tyr to the total observed fluorescence.

4.4. Results and Discussion

4.4.1. Design of Coiled Coils

The coiled coil system used in this study was based on the parallel heterodimeric coiled coils designed by Thomas *et al.*[10] The pair was formed by two 27-residue peptides encompassing 3.5 heptad repeats with a hydrophobic interface composed of Leu and Ile residues (Figure 4.1). Heterodimerization was enforced by charge complementarity: one of the peptides was acidic, rich in Asp and the other basic, rich in Lys. The expected net charges on the two peptides at pH 7.0 are -4 and +9, respectively. A single Asn residue at position 13 in each peptide locked the system into a parallel geometry stabilized by Asn-Asn H-bonding. The short range of M_{Se} quenching requires the Se atom to closely approach the aromatic ring of the F_{CN} residue. This could theoretically be accomplished by introducing the pair of residues at either the ***g*** and ***a*** or ***d*** and ***e*** positions in the heptad. A model based on PDB structure 2ZTA suggests that the ***ga*** arrangement offers slightly better geometry for the interaction, so F_{CN} was introduced at position 19 in the acidic peptide and M_{Se} was introduced at position 20 in the basic peptide.[1] Measurements on an idealized model of this system suggest a distance between the center of the Se atom (the M_{Se} residue was modeled as a Met and the distance is to the S of the Met sidechain) and the center of the F_{CN} ring of 4.6 Å (Figure 4.1). The fractional solvent accessibility of the F_{CN} side chain in the dimer is 70% relative to an extended peptide, and the cyano group is fully exposed to solvent in the model structure. A single Tyr residue was incorporated at position 4 in the basic (M_{Se}) peptide to allow the concentration of the peptide to be determined by UV absorbance. The site was chosen to place the Tyr side chain as far as possible from the F_{CN} residue to limit FRET effects.[15] Measurements based on an idealized model of this system suggest a distance between the centers of the aromatic rings

of the Tyr and F_{CN} residues of 26 Å. The N termini of the peptides were acetylated and the C termini amidated to increase stability of the dimer.

4.4.2. The Designed Coiled Coil Adopts Helical Structure

Coiled coil formation was characterized by circular dichroism (CD) spectroscopy. Individual peptides (35 μM) in phosphate buffer (10 mM, pH 7.4) show little α-helical structure as indicated by the molar ellipticity at 222 nm relative to the value at 200 nm (Figure 4.2). The fluorescent (acidic) and quenching (basic) peptides appear to adopt very similar disordered conformations as estimated by CD. In contrast, the CD spectrum of a solution of both peptides (17.5 μM each) in the same buffer reveals significant helical structure, indicated by intense negative molar ellipticity at 222 and 208 nm as well as intense positive molar ellipticity at 193 nm. Denaturing the samples with urea (10 M) abolished the helical CD signal for both the monomeric peptides and the heterodimeric coiled coil (Figure 4.3). These results indicate that α-helix formation is strongly tied to heterodimeric association for these peptides, and that the introduction of F_{CN} and M_{Se} at the peptide interface does not prevent dimerization.

4.4.3. p-Cyanophenylalanine Fluorescence Quenching Provides a Sensitive Probe of Coiled Coil Formation

The quenching efficiency of M_{Se} in the coiled coil system was examined by comparing the emission spectrum of the F_{CN} monomer (35 μM) in phosphate buffer (10 mM, pH 7.4) to the spectrum of a sample that contained both peptides (35 μM each) in the same buffer (Figure 4.4). We avoided the use of chloride salts since chloride can quench F_{CN} fluorescence.[20] F_{CN} was excited at 240 nm to minimize any contribution from the Tyr residue in CC-20$M_{Se}B_N$[3.5]. Fluorescence intensity at the emission maximum (291 nm) was 81% lower in the dimer sample compared to the F_{CN} peptide alone, indicating a high quenching efficiency. The quenching effect was abolished by

denaturing the coiled coil with urea (10 M), returning the intensity to the same value observed for the isolated F_{CN} peptide in either buffer or 10 M urea (Figure 4.5).

4.4.4. Tryptophan Fluorescence is Much Less Sensitive to Coiled Coil Formation

Although the use of F_{CN} appears quite promising, we also investigated the use of Trp as the fluorophore in the coiled coil system. The peptides used for this pair were identical to those that made up the F_{CN}-M_{Se} pair with the only change being the replacement of F_{CN} with Trp; this peptide was designated CC-19W$A_N^{3.5}$ (Figure 4.6). Measurements on a model based on PDB structure 2ZTA suggest a distance between the center of the Se atom, (modeled using Met) and the center of the indole ring of 5.0 Å.[1] The Trp-M_{Se} heterodimer was characterized by CD spectroscopy (Figure 4.7). As with the F_{CN}-M_{Se} pair, the monomeric peptides appeared to be largely unstructured, with only weak helical structure as indicated by CD. In contrast, strong negative ellipticity was observed at 222 and 208 nm as well as intense positive ellipticity at 193 nm in a mixed peptide solution, as was observed for the F_{CN}-M_{Se} pair, consistent with coiled coil formation. Denaturing the samples with urea (10 M) abolished the helical CD signal for both the monomeric peptides and the heterodimeric coiled coil (Figure 4.8).

The fluorescence quenching efficiency of the Trp-M_{Se} pair was also examined, but only a weak 14% quenching effect was observed in the mixed peptide solution (35 μM each) compared to a 35 μM solution of CC-19W$A_N^{3.5}$ (Figure 4.9). The weaker quenching could result from less optimal geometry in the Trp-M_{Se} coiled coil, lower intrinsic quenching efficiency of M_{Se} for Trp relative to F_{CN}, the quenching effect being partially offset by the intrinsic increase in Trp fluorescence due to partial burial, proximity of the indole ring to a Lys sidechain in both the folded and unfolded states as Lys can quench Trp fluorescence, or some combination of these effects.

4.4.5. Fluorescence Monitored Titrations Provide a Convenient and Accurate Method to Determine K_D

The dissociation constant for the heterodimer was first determined using a standard CD method for coiled coils. Solutions of the heterodimer were prepared in phosphate buffer (10 mM, pH 7.4) at 202, 101, 49, 19 and 8 μM (1:1 peptide concentration ratio, concentration determined by absorbance at 280 nm). CD-monitored thermal denaturation experiments were performed for each of these samples and the unfolding transition midpoint (T_M) was determined using standard methods (Figure 4.10). The K_D was calculated using the method of Marky and Breslauer.[11] $1/T_M$ was plotted versus the natural logarithm of the total concentration of coiled coil peptides and fit to equation (4.3).

Analysis of the thermally induced unfolding of coiled coils has often assumed that the $\Delta C_p°$ is zero or small.[10] The linear van 't Hoff plot is consistent with this, however the value of $\Delta C_p°$ can also enter into the analysis of the fits of the individual melting curves. For samples that have a high T_M, ie. samples with high total peptide concentration [*CC*], the post-transition baseline may not always be well defined. In this case, the apparent T_M derived from the non-linear fit to the melting curve can vary if $\Delta C_p°$ is fixed at different values. This likely reflects the fact that curves with a poorly defined post-transition baseline are being fit with seven variables; two for each baseline, T_M, ΔH_M, and $\Delta C_p°$. This potential complication is less of an issue when the pre- and post-transition baselines are well defined (Figure 4.11).

We analyzed the individual thermal unfolding curves using a range of fixed values of $\Delta C_p°$ to generate apparent values of T_M as a function of [*CC*]. Using these values in equation (4.3) leads to a modest variation in K_D. Using $\Delta C_p° = 0$ kcal mol^{-1} K^{-1} to fit the thermal melts leads to a K_D of 24.7 nM while a value of 0.20 kcal mol^{-1} K^{-1} leads to a K_D of 10.3 nM. An alternative approach is

to perform a global analysis of all five melting curves, allowing $\Delta C_p°$ to be determined as a fitting variable that is constant across all experiments. The value of $\Delta C_p°$ determined in this analysis was 0.58 kcal mol^{-1} K^{-1}, which leads to a K_D value of 1.67 nM (Figure 4.12). This value of $\Delta C_p°$ is in excellent agreement with the theoretical value of 0.55 kcal mol^{-1} K^{-1} predicted by the empirical relationship between $\Delta C_p°$ and the number of residues in a protein.[21] The value of K_D determined using the theoretical $\Delta C_p°$ value of 0.55 kcal mol^{-1} K^{-1} is 2.27 nM. Given that the changes in the peptide sequences likely result in a slight perturbation of the dimer interface, these values are in good agreement with the K_D of 5.15 nM determined for the parent peptides by Thomas *et al.*[10]

Having measured K_D by CD based methods we then tested if the F_{CN}-M_{Se} pair can be exploited to give an accurate determination of K_D. Fluorescence emission of F_{CN} was monitored at a fixed concentration of CC-19$F_{CN}A_N^{3.5}$ as the concentration of CC-20$M_{Se}B_N^{3.5}$ was varied (Figure 4.13). The fluorescence intensity at 291 nm (F_{291}) was plotted against the concentration of CC-20$M_{Se}B_N^{3.5}$, [M_{Se}], and a typical binding isotherm was observed for CC-20$M_{Se}B_N^{3.5}$ below 3 μM. At high CC-20$M_{Se}B_N^{3.5}$ concentrations the F_{CN} fluorescence is effectively quenched and the weaker Tyr fluorescence contributes significantly to the observed intensity at 291 nm, possibly from limited excitation of Tyr at 240 nm. This leads to an increase in the observed fluorescence at high concentrations of the M_{Se} peptide. This effect is trivial to model and correct for with a linear function dependent on the concentration of the M_{Se} peptide. The plot of F_{291} versus [M_{Se}] was fit to equation (4.5) with m, F_{max}, F_{min} and K_D as fitting parameters. The value of K_D determined using this procedure was 16.7 nM, which is in excellent agreement with the values determined by CD assuming $\Delta C_p°$ values of 0.0 to 0.2 kcal mol^{-1} K^{-1} and is reasonable agreement to the value determined by the global analysis of the CD-monitored thermal melts. We also fit the fluorescence data using a truncated data set which excluded data collected for CC-20$M_{Se}B_N^{3.5}$ concentrations

above 2 μM. This choice restricts the data to a range where the post-transition baseline is constant. The data were fit to a modified version of equation (4.5) which did not include the Tyr correction term. This analysis yielded a K_D of 14.8 nM, which is in excellent agreement with the value determined using the more robust method.

4.5. Conclusions

The data presented here demonstrates the utility of F_{CN}-M_{Se} pairs to quantitatively monitor coiled coil formation. The pair has the advantage of being sensitive and non-perturbing. A third advantage is that the detection limit is clearly lower than is possible with CD. This will be of particular use when one needs to determine K_D values for strongly binding systems. The fluorescence based approach also avoids any difficulties due to irreversible thermal unfolding. The fluorescence quenching method also avoids the problems associated with the large number of fitting variables inherent in the CD method used here. In particular, varying the value of $\Delta C_p°$ (which can be determined, but requires additional experiments) from 0 to 0.58 kcal mol^{-1} K^{-1} changes the value of K_D from 24.7 to 1.67 nM. Although this level of uncertainty may not be an issue in all applications, the fluorescence quenching method avoids it altogether. This aspect of the CD fitting problem can be circumvented by taking the second derivative of very high quality thermal unfolding data to determine T_M. We were unable to use this method to analyze our data because of signal-to-noise issues, but the fluorescence method again avoids this problem.

The M_{Se}-F_{CN} approach also suffers from some potential disadvantages relative to the CD based approaches. M_{Se} is expected to be more prone to oxidation than Met, although no oxidation was detected in the system studied here, and the fluorescence quenching properties of oxidized M_{Se} derivatives have not been investigated for F_{CN}, although we have shown that oxidized M_{Se} is an effective quencher of Trp fluorescence (Chapter 2). An oxidized M_{Se} residue is larger than M_{Se} and thus might affect packing. Reducing agents and degassed buffers can be used if oxidation is a potential concern. There are some additional practical considerations if F_{CN} fluorescence quenching is employed. F_{CN} fluorescence is quenched by Cl^- ions, thus high concentrations of Cl^- containing buffers should be avoided.

Fluorescence-based methods are well suited to high throughput assays, while CD generally is not. The F_{CN}-M_{Se} pair offers a fluorescence probe that is amenable to screening libraries of coiled coils. For example, a single F_{CN} containing peptide could be screened against a library of M_{Se} containing peptides. Another useful feature is that, unlike for Trp-His pairs, the quenching effect is independent of pH.[16] In principle the F_{CN}-M_{Se} pair can be used to follow thermally induced dissociation, however F_{CN} fluorescence has a significant intrinsic temperature dependence which may complicate or compromise the analysis. The data analyzed in this work was collected using peptides prepared by solid phase peptide synthesis, but both F_{CN} and M_{Se} can be incorporated into proteins using well established recombinant methods, thus the methodology is not limited to synthetic peptides. The use of F_{CN}-M_{Se} pairs is not limited to coiled coils and can be used to quantitatively monitor other protein-protein interactions and self-assembly processes.

4.6. Figures

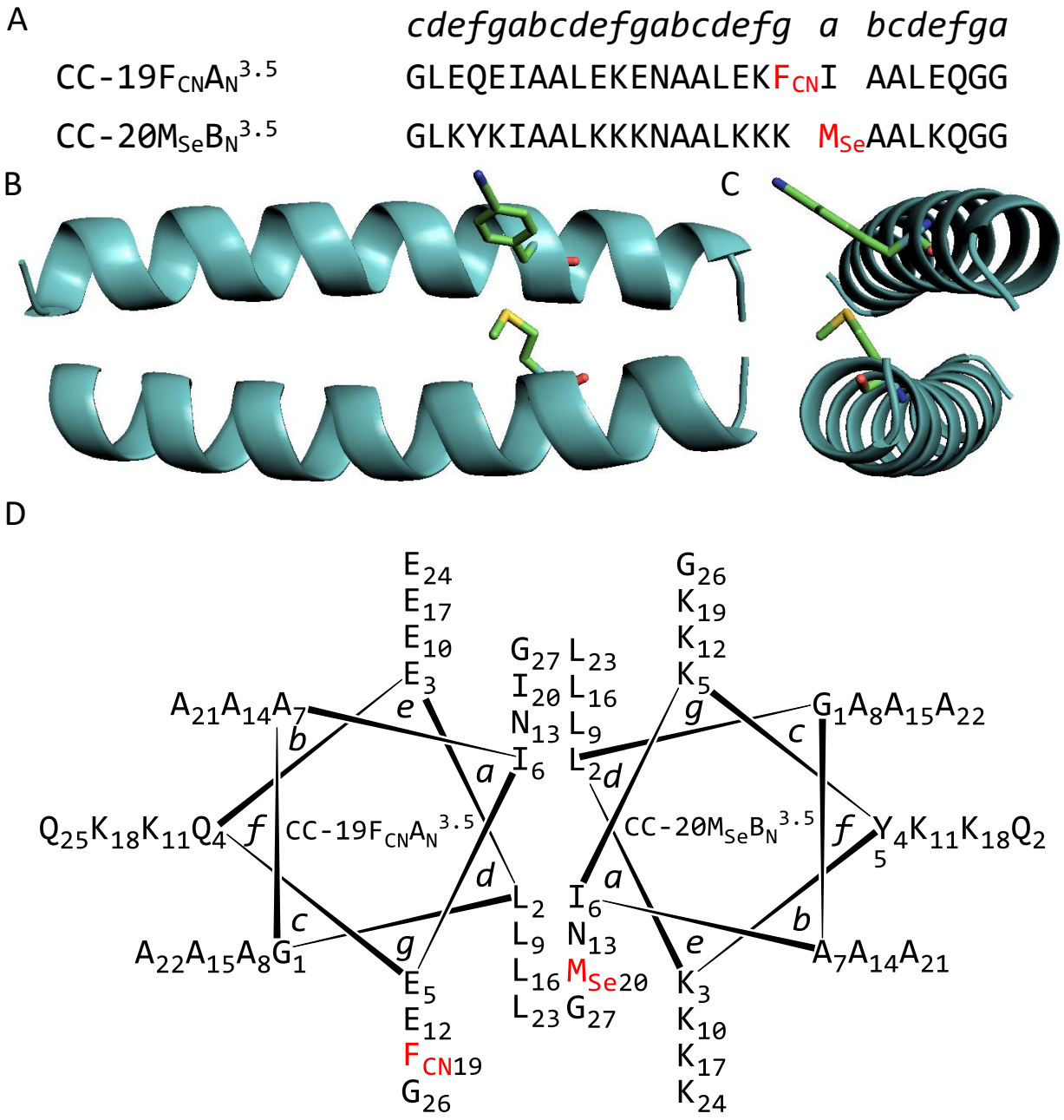


Figure 4.1. Design of the F_{CN}-M_{Se} coiled coil. **(A)** Sequences and heptad register of the coiled coil peptides. These constructs were designated CC-19$F_{CN}A_N^{3.5}$ and CC-20$M_{Se}B_N^{3.5}$ (coiled coil, acidic or basic, Asn, 3.5 heptad repeats) after the system used by Thomas *et al.*[10] F_{CN} denotes *p*-cyanophenylalanine and M_{Se} denotes selenomethionine. **(B)** A model of the coiled coil dimer based on PDB structure 2ZTA showing the F_{CN} and M_{Se} residues in stick format.[1] **(C)** A view of the structure displayed in **(B)** rotated 90°. **(D)** Helical-wheel diagram showing the heptad repeats. The heptad register designations ***a-g*** correspond to the heptad register in **(A)**.

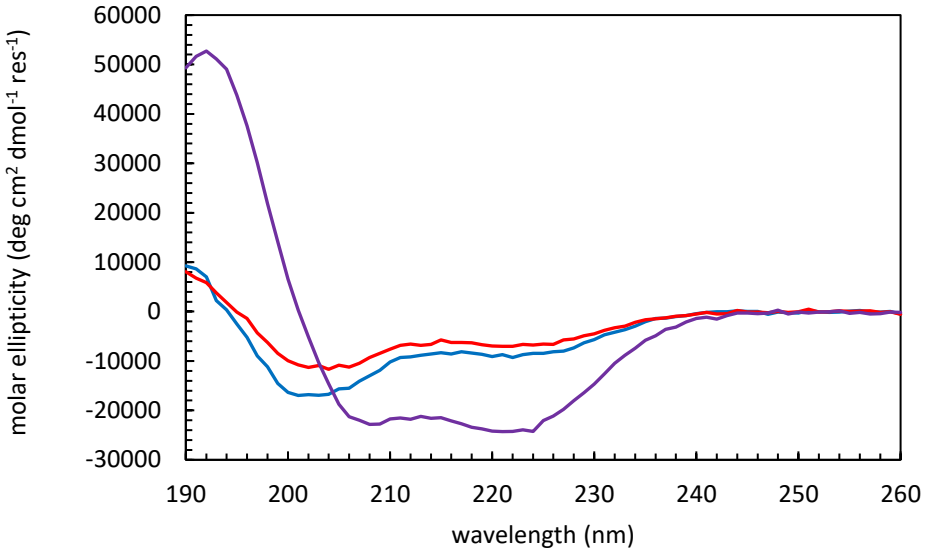


Figure 4.2. F_{CN} and M_{Se} peptides form a coiled coil . CC-19$F_{CN}A_N^{3.5}$ (red), CC-20$M_{Se}B_N^{3.5}$ (blue) and the dimer (purple) in phosphate buffer (10 mM, pH 7.4) at 20 °C with a 35 μM total peptide concentration.

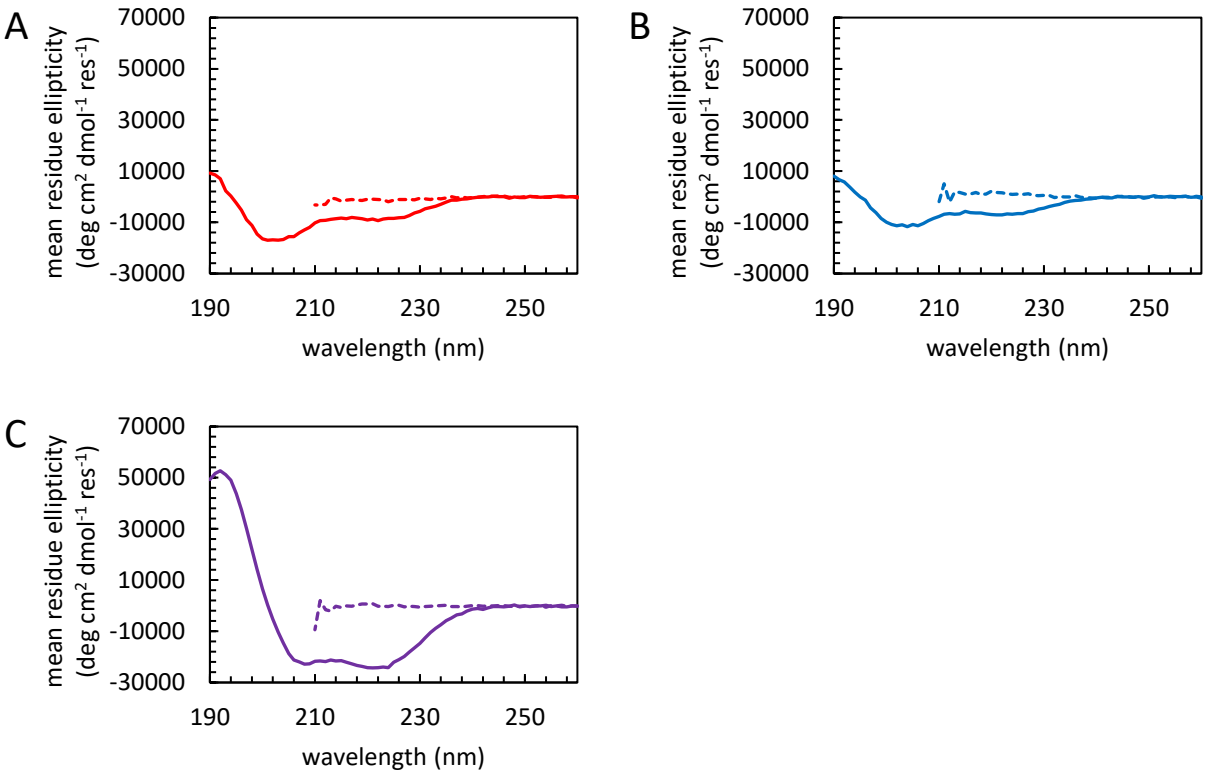


Figure 4.3. The F_{CN}-M_{Se} coiled coil dissociates in urea. **(A)** Circular dichroism spectra of CC-19$F_{CN}A_N^{3.5}$ (35 μM) at 20 °C in phosphate buffer (10 mM, pH 7.4, solid red) and in urea (10 M, dashed red). **(B)** Circular dichroism spectra of CC-20$M_{Se}B_N^{3.5}$ (35 μM) at 20 °C in phosphate buffer (10 mM, pH 7.4, solid blue) and in urea (10 M, dashed blue). **(C)** Circular dichroism spectra of a mixed solution of CC-19$F_{CN}A_N^{3.5}$ and CC-20$M_{Se}B_N^{3.5}$ (1:1, 35 μM total peptide concentration) at 20 °C in phosphate buffer (10 mM, pH 7.4, solid purple) and in urea (10 M, dashed purple).

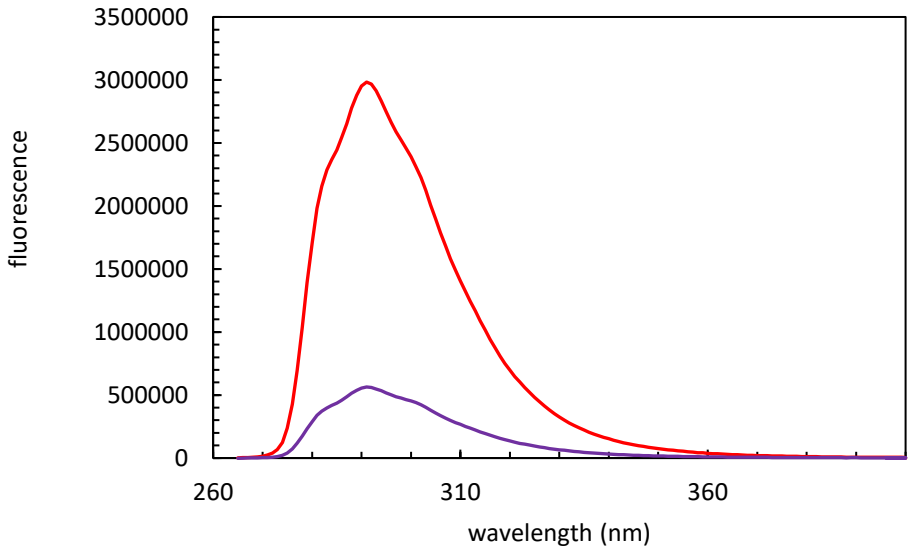


Figure 4.4. F_{CN}-M_{Se} pairs monitor coiled coil formation. Fluorescence emission spectra of CC-19$F_{CN}A_N^{3.5}$ (red) and the dimer with CC-20$M_{Se}B_N^{3.5}$ (purple) in phosphate buffer (10 mM, pH 7.4) at 20 °C with a CC-19$F_{CN}A_N^{3.5}$ concentration of 35 μM. The excitation wavelength was 240 nm.

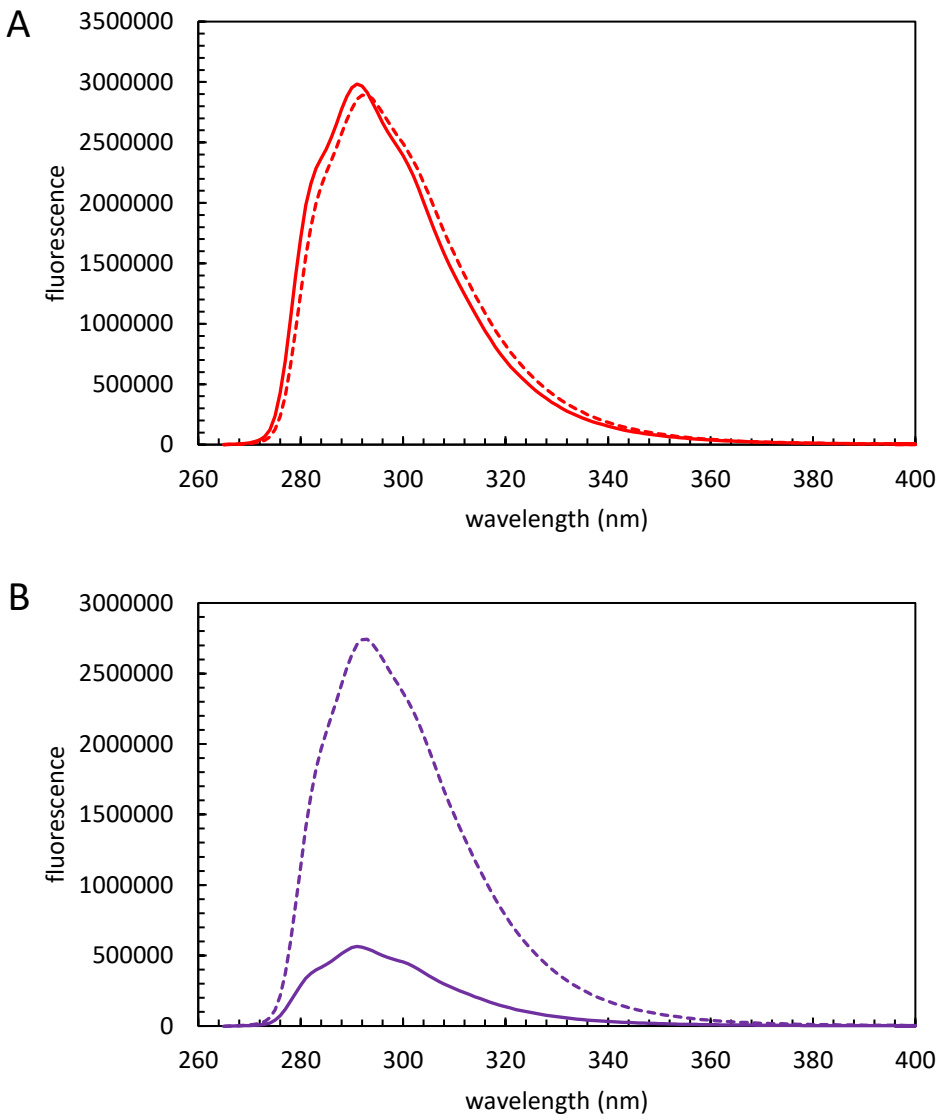


Figure 4.5. F_{CN}-M_{Se} pairs monitor the urea-induced dissociation of coiled coils. **(A)** Fluorescence emission spectra of CC-19$F_{CN}A_N^{3.5}$ (35 μM) at 20 °C in phosphate buffer (10 mM, pH 7.4, solid red) and in urea (10 M, dashed red). The excitation wavelength used was 240 nm. **(B)** Fluorescence emission spectra of a mixed solution of CC-19$F_{CN}A_N^{3.5}$ and CC-20$M_{Se}B_N^{3.5}$ (1:1, 70 μM total peptide concentration) at 20 °C in phosphate buffer (10 mM, pH 7.4, solid purple) and in urea (10 M, dashed purple). The excitation wavelength used was 240 nm.

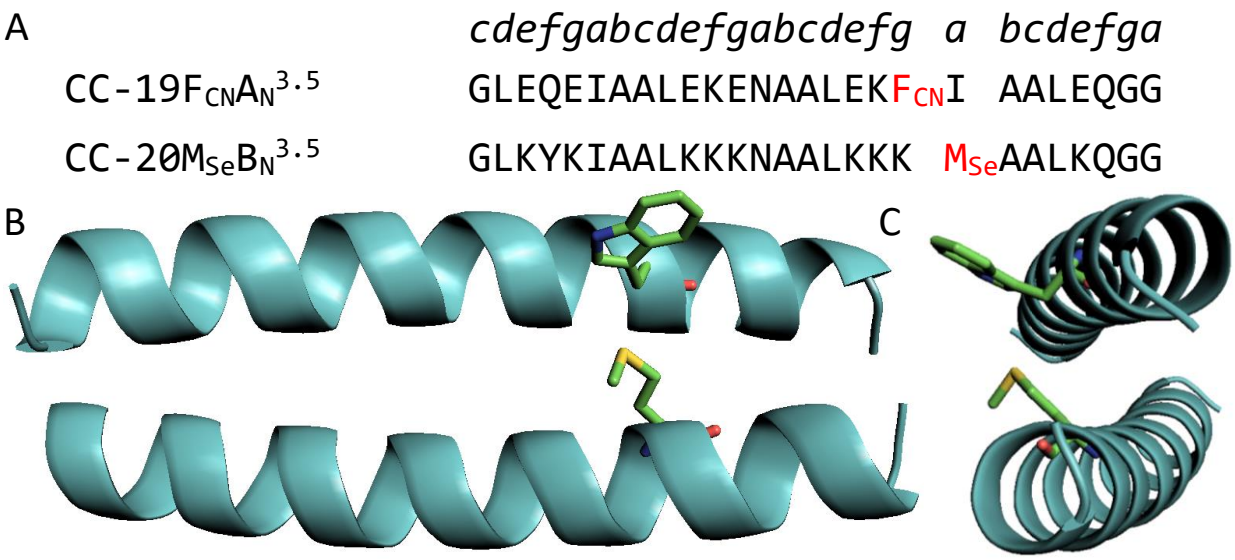


Figure 4.6. Design of the Trp-M_{Se} coiled coil. **(A)** Sequences and heptad register of the CC-19W$A_N^{3.5}$ and CC-20$M_{Se}B_N^{3.5}$ coiled coil peptides. M_{Se} denotes selenomethionine. **(B)** A model of the coiled coil dimer based on PDB structure 2ZTA showing the Trp and M_{Se} residues in stick format.[1] **(C)** A view of the structure displayed in **(B)** rotated 90°.

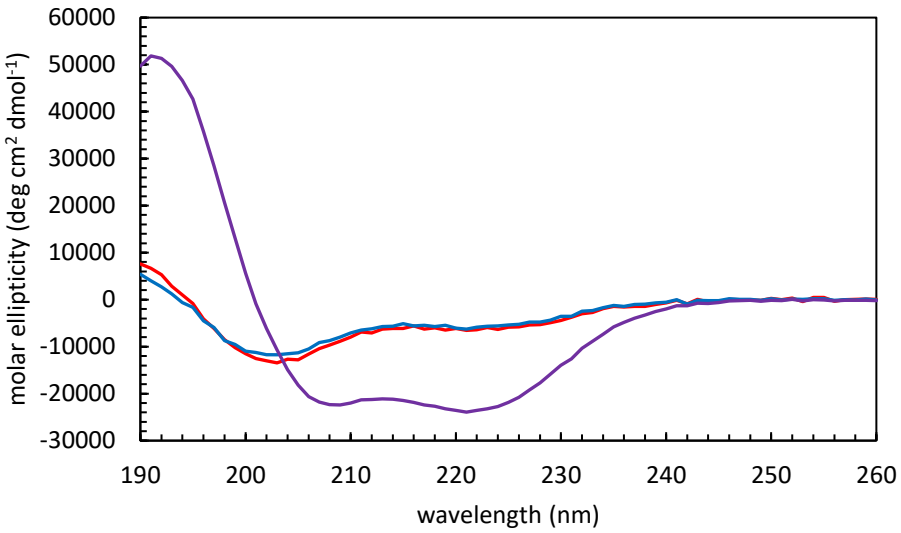


Figure 4.7. Trp and M_{Se} peptides form a coiled coil. Circular dichroism spectra of CC-19W$A_N^{3.5}$ (red), CC-20$M_{Se}B_N^{3.5}$ (blue) and the dimer (purple) in phosphate buffer (10 mM, pH 7.4) at 20 °C with a 35 μM total peptide concentration.

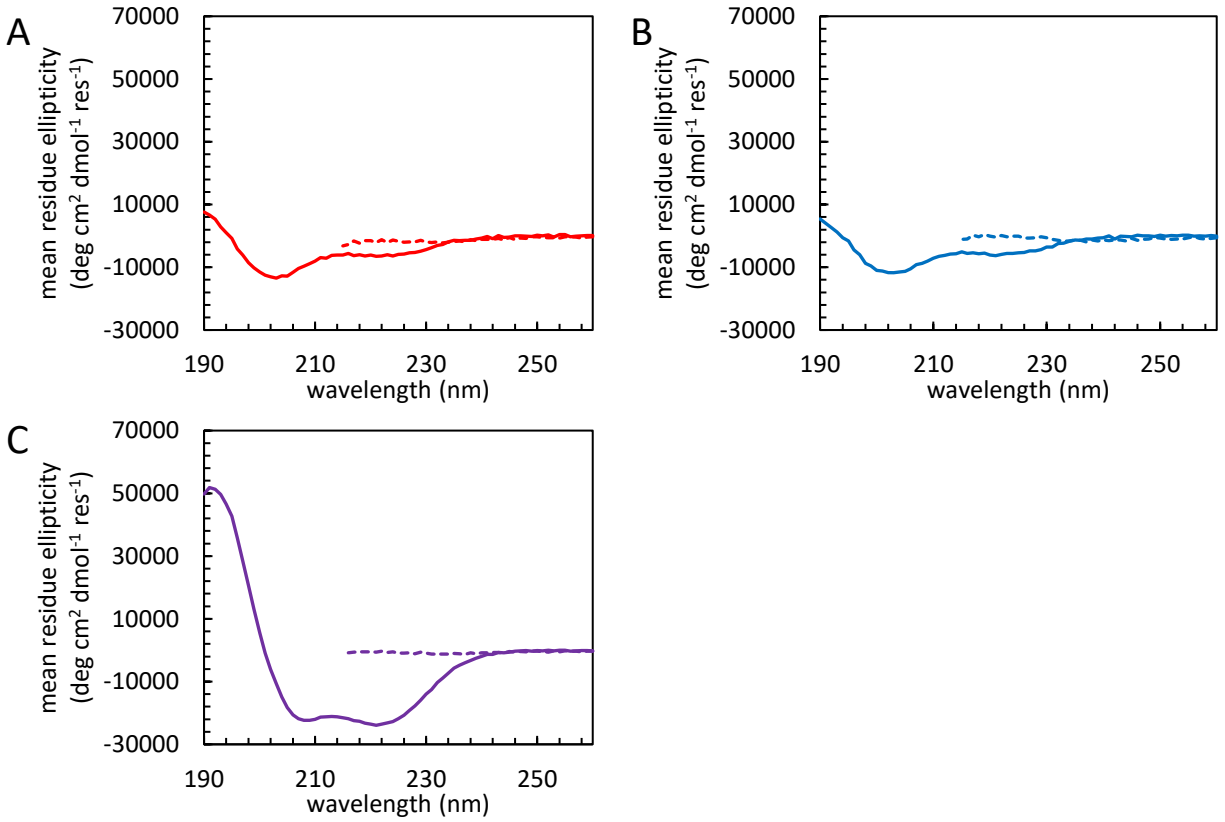


Figure 4.8. The Trp-M_{Se} coiled coil dissociates in urea. **(A)** Circular dichroism spectra of CC-19W$A_N^{3.5}$ (35 μM) at 20 °C in phosphate buffer (10 mM, pH 7.4, solid red) and in urea (10 M, dashed red). **(B)** Circular dichroism spectra of CC-20$M_{Se}B_N^{3.5}$ (35 μM) at 20 °C in phosphate buffer (10 mM, pH 7.4, solid blue) and in urea (10 M, dashed blue). **(C)** Circular dichroism spectra of a mixed solution of CC-19W$A_N^{3.5}$ and CC-20$M_{Se}B_N^{3.5}$ (1:1, 35 μM total peptide concentration) at 20 °C in phosphate buffer (10 mM, pH 7.4, solid purple) and in urea (10 M, dashed purple).

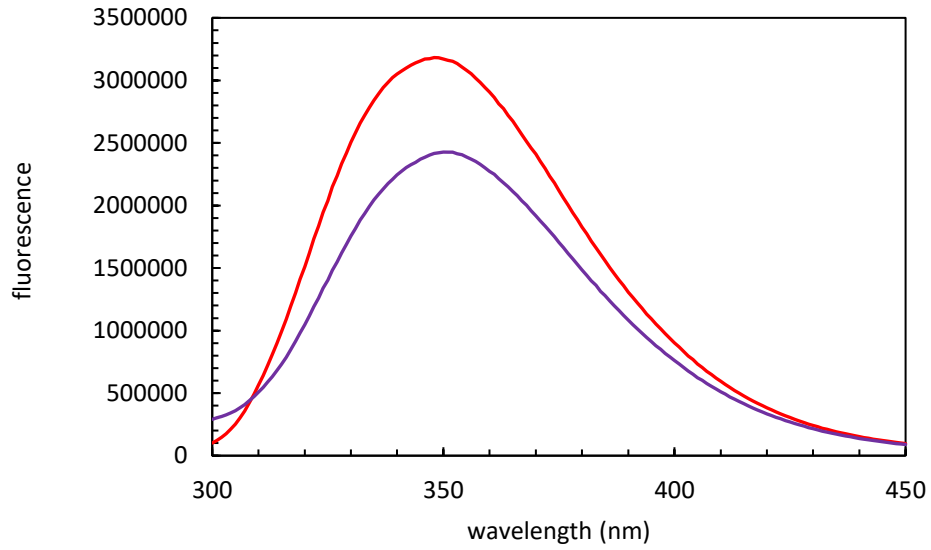


Figure 4.9. Trp-M_{Se} pairs do not monitor coiled coil formation. Fluorescence emission spectra of CC-19W$A_N^{3.5}$ (red) and the dimer with CC-20$M_{Se}B_N^{3.5}$ (purple) in phosphate buffer (10 mM, pH 7.4) at 20 °C with a CC-19W$A_N^{3.5}$ concentration of 35 μM. The excitation wavelength was 280 nm.

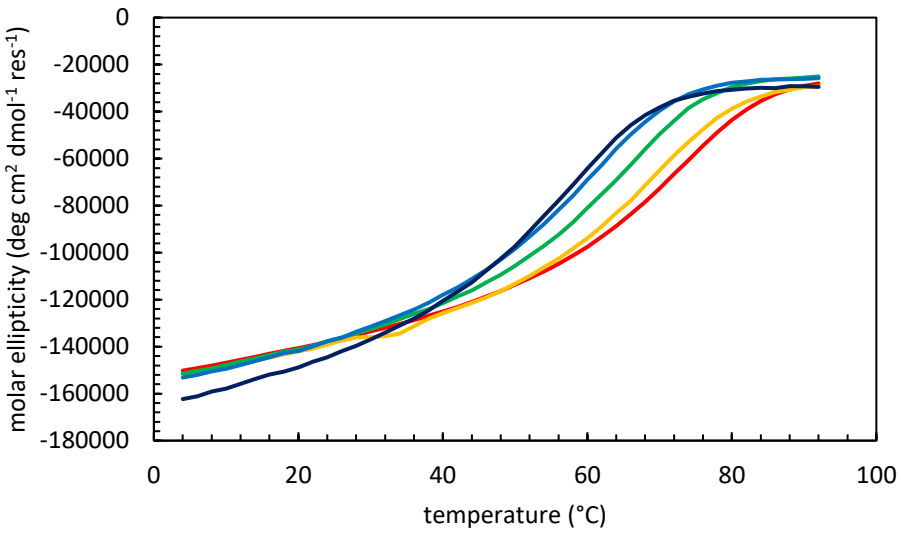


Figure 4.10. Circular dichroism monitored thermal unfolding of F_{CN}-M_{Se} coiled coil. Thermal unfolding curves for the CC-19$F_{CN}A_N^{3.5}$/CC-20$M_{Se}B_n^{3.5}$ dimer at 202 μM (red), 101 μM (orange), 49 μM (green), 19 μM (blue) and 8 μM (indigo). Experiments were performed in phosphate buffer (10 mM, pH 7.4) at 20 °C with a monomer ratio of 1:1. Curves were fit to equation (4.1) to determine T_M, the midpoint of the unfolding transition.

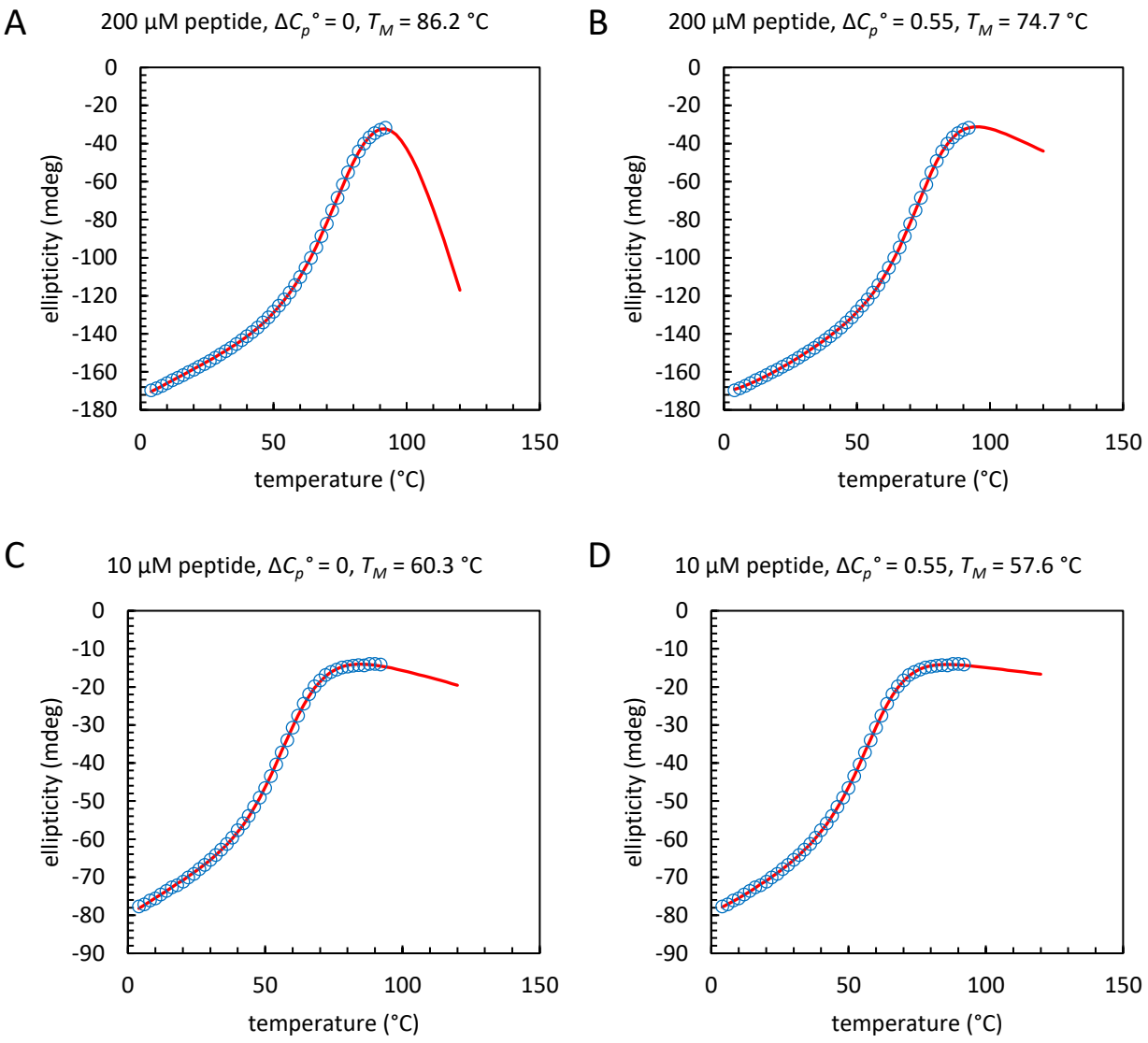


Figure 4.11. Dependence of apparent T_M values on the choice of $\Delta C_p°$ illustrated for [*CC*] = 200 µM and [*CC*] = 10 µM. **(A)** Fit to the 200 µM data with $\Delta C_p°$ = 0 kcal mol^{-1} K^{-1} yields a T_M of 86.2 °C, but also leads to a nonphysical post-transition baseline. **(B)** Fit to the 200 µM data with $\Delta C_p°$ = 0.55 kcal mol^{-1} K^{-1} yields T_M = 74.7 °C with a more reasonable post-transition baseline. **(C)** and **(D)** display fits to melting data collected at [*CC*] = 10 µM. The choice of $\Delta C_p°$ has less dramatic effects.

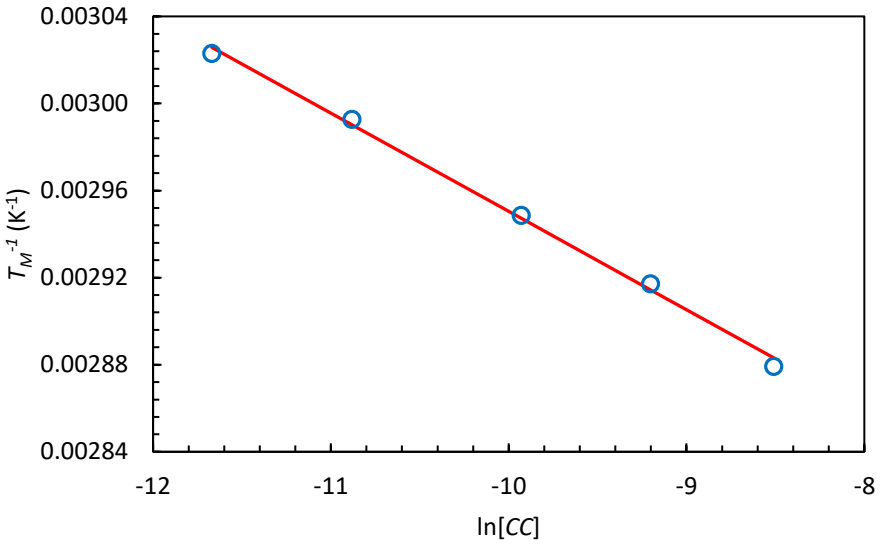


Figure 4.12. Determination of K_D from CD monitored thermal unfolding curves. The inverse of the thermal unfolding transition midpoints (T_M) determined using a $\Delta C_p°$ value of 0.58 kcal mol^{-1} K^{-1} for 1:1 mixtures of CC-19F$_{CN}$A$_N^{3.5}$ and CC-20M$_{Se}$B$_N^{3.5}$ at total peptide concentrations of 202, 101, 49, 19 and 8 μM plotted against the natural logarithm of the peptide concentration in M (circles). The solid line is a linear fit of the data to equation (4.3), R^2 = 0.9970, p = 6.88·10^{-5}.

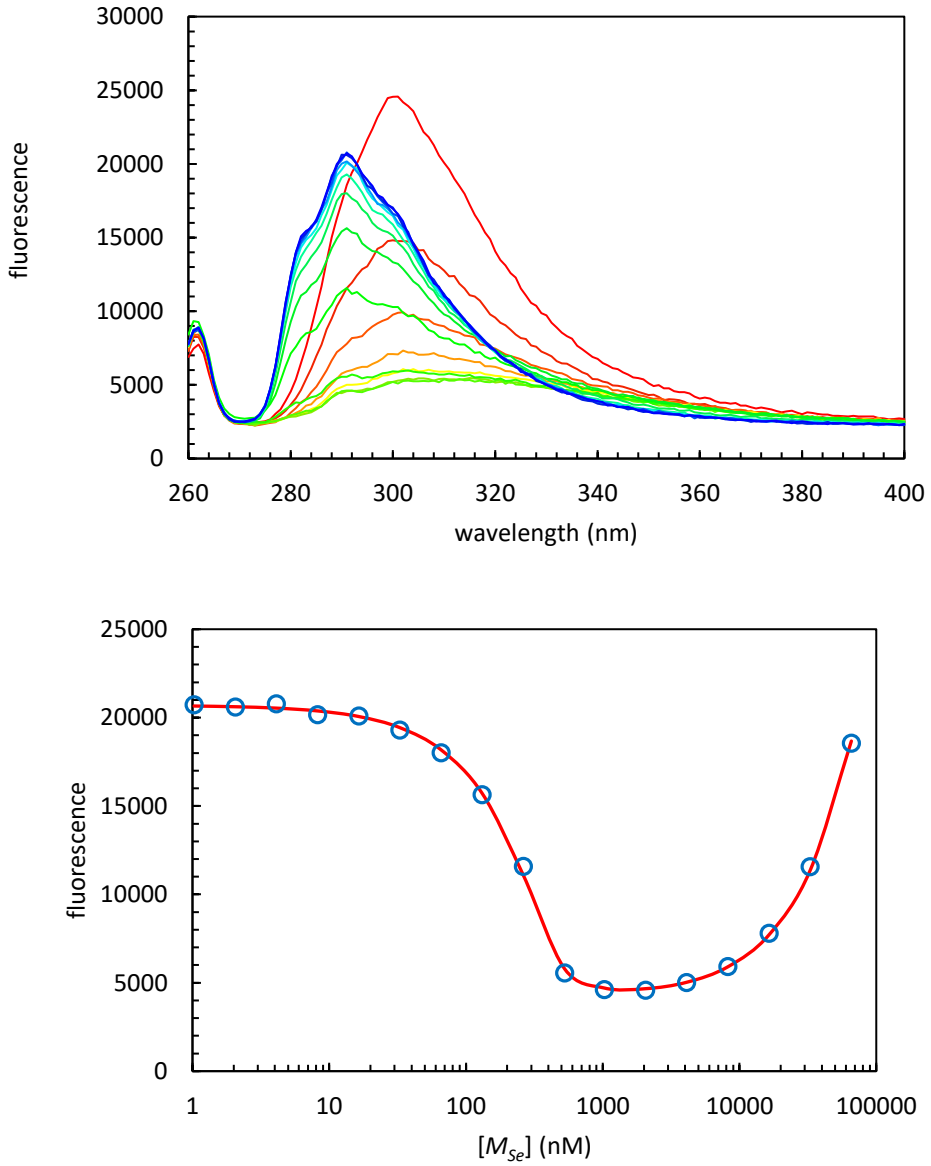


Figure 4.13. Determination of K_D by fluorescence monitored titration. **(A)** Fluorescence emission spectra of CC-19$F_{CN}A_N^{3.5}$ (412 nM) in phosphate buffer (10 mM, pH 7.4) at 20 °C in the presence of increasing concentrations of the CC-20$M_{Se}B_N^{3.5}$ (blue to red). The excitation wavelength was 240 nm. **(B)** The fluorescence intensity at 291 nm is plotted against the concentration of CC-20$M_{Se}B_N^{3.5}$ (open circles) and fit to equation (4.5).

4.7. References

[1] Oshea, E. K., Klemm, J. D., Kim, P. S., and Alber, T. (1991) X-Ray Structure of the Gcn4 Leucine Zipper, a 2-Stranded, Parallel Coiled Coil, *Science 254*, 539-544.

[2] Rackham, O. J., Madera, M., Armstrong, C. T., Vincent, T. L., Woolfson, D. N., and Gough, J. (2010) The evolution and structure prediction of coiled coils across all genomes, *J Mol Biol 403*, 480-493.

[3] Oneil, K. T., and Degrado, W. F. (1990) A Thermodynamic Scale for the Helix-Forming Tendencies of the Commonly Occurring Amino-Acids, *Science 250*, 646-651.

[4] Negron, C., and Keating, A. E. (2014) A set of computationally designed orthogonal antiparallel homodimers that expands the synthetic coiled-coil toolkit, *J Am Chem Soc 136*, 16544-16556.

[5] Root, M. J., Kay, M. S., and Kim, P. S. (2001) Protein design of an HIV-1 entry inhibitor, *Science 291*, 884-888.

[6] Boyle, A. L., and Woolfson, D. N. (2011) De novo designed peptides for biological applications, *Chem Soc Rev 40*, 4295-4306.

[7] Woolfson, D. N. (2005) The design of coiled-coil structures and assemblies, *Adv Protein Chem 70*, 79-+.

[8] Morais, A. C., and Ferreira, S. T. (2005) Folding and stability of a coiled-coil investigated using chemical and physical denaturing agents: comparative analysis of polymerized and non-polymerized forms of alpha-tropomyosin, *Int J Biochem Cell Biol 37*, 1386-1395.

[9] Wagner, D. E., Phillips, C. L., Ali, W. M., Nybakken, G. E., Crawford, E. D., Schwab, A. D., Smith, W. F., and Fairman, R. (2005) Toward the development of peptide nanofilaments and nanoropes as smart materials, *Proc Natl Acad Sci U S A 102*, 12656-12661.

[10] Thomas, F., Boyle, A. L., Burton, A. J., and Woolfson, D. N. (2013) A set of de novo designed parallel heterodimeric coiled coils with quantified dissociation constants in the micromolar to sub-nanomolar regime, *J Am Chem Soc 135*, 5161-5166.

[11] Marky, L. A., and Breslauer, K. J. (1987) Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves, *Biopolymers 26*, 1601-1620.

[12] Woolfson, D. N., and Alber, T. (1995) Predicting oligomerization states of coiled coils, *Protein Sci 4*, 1596-1607.

[13] Miyake-Stoner, S. J., Miller, A. M., Hammill, J. T., Peeler, J. C., Hess, K. R., Mehl, R. A., and Brewer, S. H. (2009) Probing protein folding using site-specifically encoded unnatural amino acids as FRET donors with tryptophan, *Biochemistry 48*, 5953-5962.

[14] Tucker, M. J., Oyola, R., and Gai, F. (2005) Conformational distribution of a 14-residue peptide in solution: a fluorescence resonance energy transfer study, *J Phys Chem B 109*, 4788-4795.

[15] Taskent-Sezgin, H., Chung, J., Patsalo, V., Miyake-Stoner, S. J., Miller, A. M., Brewer, S. H., Mehl, R. A., Green, D. F., Raleigh, D. P., and Carrico, I. (2009) Interpretation of p-cyanophenylalanine fluorescence in proteins in terms of solvent exposure and contribution of side-chain quenchers: a combined fluorescence, IR and molecular dynamics study, *Biochemistry 48*, 9040-9046.

[16] Peran, I., Watson, M. D., Bilsel, O., and Raleigh, D. P. (2016) Selenomethionine, p-cyanophenylalanine pairs provide a convenient, sensitive, non-perturbing fluorescent probe of local helical structure, *Chem Commun (Camb) 52*, 2055-2058.

[17] Mintzer, M. R., Troxler, T., and Gai, F. (2015) p-Cyanophenylalanine and selenomethionine constitute a useful fluorophore-quencher pair for short distance measurements: application to polyproline peptides, *Phys Chem Chem Phys 17*, 7881-7887.

[18] Lakowicz, J. R. (2006) *Principles of fluorescence spectroscopy*, 3rd ed., Springer, New York.

[19] Hendrickson, W. A. (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation, *Science 254*, 51-58.

[20] Pazos, I. M., Roesch, R. M., and Gai, F. (2013) Quenching of p-Cyanophenylalanine Fluorescence by Various Anions, *Chem Phys Lett 563*, 93-96.

[21] Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding, *Protein Sci 4*, 2138-2148.

5. Length-Dependent Stability of Proteins: Agreement Between *In Vitro* and *In Vivo* Assays and Insensitivity of Thermal Shift Assays in Large Proteins

5.1. Abstract

Despite advances in protein structural determination and prediction, it remains challenging to accurately predict protein stability. Chain-length dependent stability models have been reasonably successful at reproducing *in vitro* data and have been demonstrated to accurately predict the temperature dependent growth rate of cells. However, existing models classify proteins as mesophilic or thermophilic based upon their denaturation temperature (T_M), and recent *in vivo* experiments suggest that they do not reflect the stability of proteins in the cellular environment. An expanded dataset of 174 proteins that fold in a two-state manner was compiled, with classification based on the source organism rather than T_M. Linear relationships of ΔH, ΔS and the change in heat capacity upon unfolding ($\Delta C_p°$) with respect to the chain length (N) were used to reparametrize the chain length dependent stability model. The new parametrization was found to more accurately reproduce the results of *in vivo* experiments and predicts more complex differences between proteins from mesophilic and thermophilic organisms. Chain length dependent stability relationships also have implications for thermal shift assays. Thermal shift assays are used to identify drug leads by measuring changes in T_M resulting from the binding of ligands that stabilize protein structure. Thermal shift assays are attractive for their generality and simplicity, but are not well-understood at a quantitative level. A dataset stability and T_M data for 1124 mutations of 16 proteins was compiled and used to show that there is an inverse relationship between N and the expected thermal shift. The results warrants caution in the design and application of thermal shift assays.

5.2. Introduction

Advances in structural determination over the past several decades have revolutionized understanding of the form and function of biomolecules and especially proteins. Most notable has been the advent of structure-based drug discovery and design. Protein structural determination has also played a key role in understanding many of the determinants of protein stability, such as hydrophobic clusters and cores. On the other hand, structure determination has had comparatively less impact on the ability to predict protein stability.

Simultaneously, the sequencing revolution has produced a vast quantity of data on the primary sequences of proteins and entire genomes. This data is now used extensively to predict protein structure and function based on homology. Although sequence data contains less information than a structure, it is very useful in systems biology analysis, which relies on relatively simple inputs to map networks of interactions. Indeed, sequence data is already used to predict binding partners for uncharacterized proteins.[1] Comparatively less effort has gone into modeling of protein stability based upon sequence, despite proteostasis being a key component of understanding protein interactions *in vivo* and despite the importance of protein stability in aggregation and protein misfolding. The ability to rapidly predict protein stability allows calculation of proteome stability and their response to thermal stress.

Protein stability models also have implications for drug discovery methods, particularly thermal shift assays (TSA), which measures changes in thermal denaturation temperatures (T_M) to detect ligand binding. Briefly, TSA is based on an increase in protein stability in the presence of a ligand which binds preferentially to the folded state; the increase in stability results in an increase in T_M. However, the quantitative relationship between changes in T_M and ligand affinity is not well understood. The method is attractive for its broad applicability and amenability to high-throughput

screening, especially fluorescence-based TSA, which uses nonspecific, extrinsic dyes to detect protein unfolding. Recently, methods for cellular thermal shift assays have been developed, offering the promise of high-throughput *in vivo* drug screening.[2-5] The largest thermal shifts are known to be observed at low protein concentrations and significant molar excess of ligand. TSA sensitivity is also expected to depend on protein size, although this relationship has not been closely examined.

To make predictions regarding protein stability based on of large amounts of low-complexity sequence data a simple model is necessary. It has been noted that many fundamental properties of proteins are strongly correlated with chain length (N).[6] The development of the differential scanning calorimeter by Privalov in the 1960's made possible the determination of high quality data for the change in heat capacity upon unfolding ($\Delta C_p°$) of proteins. It has since been shown that $\Delta C_p°$ is related to the change in solvent accessible surface area upon unfolding ($\Delta SASA$), and that $\Delta SASA$ and therefore $\Delta C_p°$ is strongly correlated with chain length.[7, 8] Since the thermal dependence ΔH and ΔS is determined by $\Delta C_p°$, these values are also expected to correlate with chain length, as has been shown.[8]

These relationships were employed by Dill *et al* to develop a model for the chain length dependent stability of proteins.[9] Separate parameterizations of the model were determined for mesophilic proteins and thermophilic proteins from a dataset of protein thermodynamic data assembled by Sawle and Ghosh.[10] The distinction between mesophilic and thermophilic proteins was based on the biophysical properties of proteins rather than upon the organism from which they were derived. Proteins were classified by the T_M of each protein, with the cutoff temperature chosen to minimize the least-square error for the linear dependence of ΔH, ΔS and $\Delta C_p°$ on N. By incorporating data for the chain length distribution of fully sequenced genomes, the model predicts

the temperature dependent growth rate of six mesophilic and six thermophilic organisms with reasonable accuracy. However, the parameterization may not reflect the distribution of protein stability in organisms because it artificially reclassifies some mesophilic proteins as thermophiles. Small proteins tend to have high T_M values owing to their small $\Delta C_p°$, thus the Sawle and Ghosh approach can lead to an artificial reduction in short chain length protein from the mesophile dataset and an artificial enrichment of small proteins in the "thermophile" dataset.

Recently, a major analysis of *in vivo* protein stabilities based on limited proteolysis mass spectrometry presented experimental data to test the predictions of the chain-length dependent stability model.[11] The model was found to disagree with experiment in several key details, calling into question the applicability of *in vitro* stability data to proteins *in vivo*. Most importantly, temperature-induced cell death in *E. coli* was found to result from the denaturation of a small number of critical proteins which showed no chain-length bias. This differs from the prediction of the chain-length dependent stability model, which proposed that cellular collapse was triggered by a proteome catastrophe in which many small proteins unfold at the same temperature.

We extend and reassess the dataset of protein thermodynamic data, removing non-cooperative proteins and reclassifying proteins based on the source organism rather than a T_M cutoff. Reparameterizations of the chain-length dependent stability model for mesophilic and thermophilic proteins are found to more accurately describe the behavior of protein unfolding *in vivo*. Analysis of Gibbs free energy plots allows an assessment of protein stabilization mechanisms in thermophilic organisms. In addition, the model is extended with a separate dataset of 1124 mutants of 16 proteins that makes it possible to predict ligand-binding dependent thermal shifts.

5.3. Materials and Methods

Data was gathered from the ProTherm database as well as literature sources.[12] For this analysis, selection criteria required that each entry have data for T_M, ΔH_M and $\Delta C_p°$. Values for ΔS_M were determined by dividing ΔH_M by T_M. Entries were screened to exclude proteins that did not exhibit thermodynamically two-state and reversible unfolding, as well as proteins that were stabilized by cofactors or stabilizing agents such as sulfate ion. Several homodimeric proteins are included; for these proteins, the chain length, N is given as the number of residues in the cooperative unit. Proteins were classified based upon the optimal growth temperature of the source organism into mesophilic and thermophilic proteins.

5.4. Results and Discussion

5.4.1. Design of an Expanded Dataset

Starting from the dataset of 116 proteins assembled by Sawle and Ghosh (an expansion of the earlier dataset of 63 proteins assembled by Robertson and Murphy) a dataset of 174 proteins was created.[8, 10] Data for ΔH, ΔS and $\Delta C_p°$ of 150 mesophilic proteins were plotted against chain length (N) (Figure 1). ΔH_M and ΔS_M values were extrapolated to a common reference temperature, T_R using equations (1) and (2)

$$\Delta H_R = \Delta H_M + \Delta C_p°(T_R - T_M) \quad (1)$$

$$\Delta S_R = \Delta S_M + \Delta C_p° \ln\left(\frac{T_R}{T_M}\right) \quad (2)$$

The use of a reference temperature is necessary to obtain a linear correlation with N; however, the choice of T_R has no special significance. Any convenient reference temperature may be used. It is apparent from equations (1) and (2) that as the magnitude of the difference between T_R and T_M

increases the data becomes more dependent upon $\Delta C_p°$. A T_R value of 273 K was chosen to maximize R^2 of the thermodynamic parameters with respect to N while limiting the $\Delta C_p°$ dependence of the plots.

5.4.2. There are No Convergence Temperatures for ΔH and ΔS

Previous analyses have chosen so-called convergence temperatures, T_S and T_H as reference temperatures.[8-10, 13] Convergence temperatures are points at which ΔH per residue and ΔS per residue were believed to converge to a common value for all proteins.[14-17] Originally hypothesized based on the limited protein thermodynamic data available at the time, subsequent analyses have not supported the existence of such behavior in proteins.[8] While the use of T_S and T_H as reference temperatures does not compromise the analysis, we believe that it is worth avoiding the use of these temperatures altogether. Our dataset also allows for more rigorous scrutiny of convergence temperatures. Plots of ΔH per residue and ΔS per residue *versus* T indicate that although these values do approach convergence near 323 K, this behavior is only apparent over a very broad temperature range (Figure 2). It is interesting to speculate that this may be less reflective of evolutionary pressure on early ancestral proteins, but simply reflects properties of polyamides in water. These favorable thermodynamic properties are likely one reason that all life relies on proteins to perform cellular functions; their thermodynamics are such that their stability falls over a physiologically relevant temperature.

5.4.3. Thermodynamic Parameters Correlate with Chain Length

Linear correlations of ΔH_R, ΔS_R and $\Delta C_p°$ with N are observed for mesophilic proteins described by equations (3-5)

$$\Delta H_R(N) = (m_H N + b_H)\ \text{kJ/mol} \qquad (3)$$

$$\Delta S_R(N) = (m_S N + b_S)\ \text{kJ/mol} \quad (4)$$

$$\Delta C_p^\circ(N) = (m_C N + b_C)\ \text{kJ/mol} \quad (5)$$

While the results are similar to those based on the dataset of Sawle and Ghosh, they differ significantly (Table 1).[9] Equations (3-5) can be combined with the parameters in Table 1 to calculate the length and temperature dependence of the free energy of unfolding (ΔG) for an ideal mesophilic protein, defined as a protein for which ΔH, ΔS and ΔC_p° are given by equations (3-5) exactly. The resulting relationship is given by equation (6)

$$\Delta G(N,T) = \Delta H_R + \Delta C_p^\circ(T - T_R) - T\Delta S_R - T\Delta C_p^\circ \ln\left(\frac{T}{T_R}\right) \quad (6)$$

A plot of ΔG° *vs N* shows a much weaker correlation than the plots of other thermodynamic properties. The poor correlation is largely a consequence of the choice to plot the data at 298 K. It is apparent from Figure 3A that there are two temperatures at which there is no dependence of ΔG on chain length. At these temperatures, all of the variation in experimental ΔG data is due to deviations of individual proteins from the idealized model. 298 K is nearly midway between these temperatures, and consequently offers the strongest positive correlation of ΔG with *N*. However, if ΔG *vs N* is plotted outside the range bracketed by these temperatures, a negative correlation is obtained that improves at temperatures further from 298 K. In the same manner as the plots of ΔH_R and ΔS_R, this is a result of increasing the contribution of ΔC_p° and reducing the contribution from ΔH_M and ΔS_M. For example, very strong correlations are observed if T_R is arbitrarily chosen as 1000 K. Nonetheless, the chain length dependence of ΔG° is clearly of more interest than at temperatures outside the experimentally accessible range.

Equation (6) was used to generate a plot of ΔG for mesophilic proteins *vs* T for several different values of N (Figure 3A). This plot indicates that larger proteins are more stable near physiological temperatures. The smaller $\Delta C_p°$ of shorter proteins results in a broader region of stability with higher thermal unfolding and lower cold unfolding temperatures. Transforming ΔG into fraction folded indicates that mesophilic proteins begin to unfold at approximately 50 °C, nearly independent of chain length (Figure 3B). This result differs from the report of Dill *et al*, who found that small proteins begin to undergo unfolding at significantly lower temperatures than larger proteins.[9] Our results are in agreement with experimental evidence that it is not the unfolding of small proteins that triggers cellular collapse at temperatures just above the optimal growth temperature, but rather the unfolding of a small number of essential proteins that shows no dependence on chain length.[11] Experimental evidence further suggests that shorter proteins tend to retain their structure at higher temperatures than larger proteins, which is also in agreement with our analysis.

5.4.4. Thermophilic Proteins Behave Differently Than Mesophilic Proteins

A similar analysis of ΔH_R, ΔS_R and $\Delta C_p°$ *vs* N for 24 thermophilic proteins was also performed (Figure 4). The results indicate that the chain length dependent thermodynamic properties of thermophilic proteins differ from those of their mesophilic counterparts, as described by the parameters in Table 2. The derived parameters also differ from those derived from the dataset of Sawle and Ghosh, although there are not any obvious systematic trends in the differences. The correlation of $\Delta G°$ with N is better than that observed for mesophilic proteins. This strong chain length dependence for $\Delta G°$ of thermophilic proteins is also apparent from Figure 5A. Large thermophilic proteins ($N > 100$) are predicted to have higher maximum stabilities than mesophilic proteins of similar size. Conversely, smaller thermophilic proteins are predicted to be

have lower maximum stabilities than mesophilic proteins of the same length. Thermophilic proteins are also shifted towards higher temperatures of maximum stability (T_{max}) and exhibit less dependence of T_{max} on N (Figure 6). As a result of a lower dependence of $\Delta C_p°$ on N, thermophilic proteins are also stable over a significantly larger temperature range than mesophilic proteins.

A plot of fraction folded versus T for different N suggests a very different relationship between resistance to thermal denaturation and chain length in thermophiles (Figure 5B). In contrast to the negative correlation between T_M and N observed for mesophilic proteins, large thermophilic proteins remain well folded at higher temperatures than shorter thermophilic proteins. These differences between thermophiles and mesophilic are supported by experimental data.[11]

5.4.5. Prediction of Proteome Stability and Implications for Heat Stress

Although our analysis shows that $\Delta G°$ of individual proteins cannot be accurately determined using a simple chain-length dependent model, previous work has demonstrated that the model can be averaged over entire proteomes to make reasonable predictions regarding the heat stress behavior of cells.[9] The distribution of chain lengths for fully sequenced genomes can be approximated by a gamma distribution (7)[18]

$$P(N) = \frac{N^{\alpha-1} e^{-\frac{N}{\theta}}}{\Gamma(\alpha).\theta^{\alpha}} \tag{7}$$

where α and θ are parameters defining the chain length distribution, determined from equations (8) and (9)

$$\langle N \rangle = \alpha\theta \tag{8}$$

$$\langle (\Delta N)^2 \rangle = \alpha\theta^2 \tag{9}$$

where $\langle N \rangle$ and $\langle (\Delta N)^2 \rangle$ are the mean chain length and variance for a given proteome. Using equations (3-6) and (7) a distribution of protein stabilities can be plotted at a given T. The predicted stability distributions predicted by the dataset compiled by Sawle and Ghosh and the reparameterization are shown in Figure 7. The narrower distribution of the reparameterization is again in agreement with experimental results.[11] However, experimental data does not support the existence of a pronounced tail towards high stabilities that persists in the reparameterization. This is likely an artifact resulting from the model being limited to predicting the stability of single domains. The long tail of highly stable proteins therefore represents nonexistent giant, single domains that are in reality large, multidomain proteins. Domains likely fold independently or quasi-independently in very large proteins, resulting in a lower overall stability and narrowing the stability distribution. The largest single domain protein in the dataset is 456 residues.

5.4.6. The Relationship Between $\Delta\Delta G°$ and T_M Scales Inversely with Chain Length; This Has Important Implications for Thermal Shift Assays

The chain length dependence of protein stability also has implications for the relationship between maximum stability and T_M. Using the parameters for chain length dependent stability determined by Robertson and Murphy, Rees and Robertson demonstrated an inverse relationship between the change in T_M per change in maximum stability and N.[19] To test this relationship, a new dataset of 1124 mutants of 16 proteins was compiled for which T_M and $\Delta G°$ data was available. These data were fitted to a linear equation to extract a value of $\Delta T_M/\Delta\Delta G°$ for each protein (Figure 8). Values of $\Delta T_M/\Delta\Delta G°$ were plotted against $1/N$ and fitted to a linear equation with a fixed intercept at the origin (Figure 9). This relationship between chain length the change in T_M predicted for a given stability change at 298 K is described by equation (10)

$$\frac{\Delta T_M}{\Delta\Delta G^\circ} = \frac{90.43}{N} \tag{10}$$

The inverse correlation between chain length and the effect of stabilization at 298 K on T_M is apparent, demonstrating that the thermal stability of small proteins is more sensitive to changes in stability than in larger proteins. This is a consequence of the broad stability curves of small proteins resulting from the small ΔC_p° values of these systems.

More than predicting T_M for mutant proteins, equation (10) is a general relationship between stability and T_M. Thus, it can also be employed to predict changes in T_M observed upon the binding of a ligand to the folded or unfolded state of a protein of a given length. Since ligand binding alters stability, ΔG° of folding in the absence of ligand is given by equation (11)

$$\Delta G^\circ = -RT \ln \frac{[F]}{[U]} \tag{11}$$

where R is the gas constant, $[F]$ is the equilibrium concentration of folded protein and $[U]$ is the equilibrium concentration of unfolded protein. In the presence of a ligand that binds exclusively to the folded state, equation (11) becomes equation (12)

$$\Delta G_L^\circ = \Delta G^\circ - RT \ln\left(1 + \frac{[L]}{K_D}\right) \tag{12}$$

where ΔG_L° is the Gibbs free energy of folding at 298 K in the presence of ligand, $[L]$ is the concentration of free ligand and K_D is the dissociation constant for ligand binding to the folded state. The ΔG° term can be dropped to give equation (13)

$$\Delta\Delta G_L^\circ = -RT \ln\left(1 + \frac{[L]}{K_D}\right) \tag{13}$$

where $\Delta\Delta G_L°$ is the change in the Gibbs free energy of folding at 298 K upon the introduction of ligand, which is related to ΔT_M by the linear relationship from the mutational dataset analysis (equation 10). It is common in ligand binding assays to use a large molar excess of ligand such that the total ligand concentration $[L]_T$ is much greater than the total protein concentration $[P]$. In this limiting case $[L] \approx [L]_T$ and the binding curve can be approximated by a linear equation. However, it is straightforward to derive the quadratic solution for $[L]$ which is valid at all $[L]_T$ and $[P]$ (14)

$$[L] = \left(\frac{-([P]-[L]_T+K_D+K_DK_U)+\sqrt{([P]-[L]_T+K_D+K_DK_U)^2-4(-K_D-K_DK_U)}}{2}\right) \quad (14)$$

where K_U is the protein unfolding equilibrium constant in the absence of ligand. The value of K_U can be determined from the chain-length dependent stability equation (6), however this term can generally be excluded except in cases of proteins that are unstable in the absence of ligand (large K_U) or very weakly binding ligands (large K_D).

Equations (10), (13) and (14) can be combined to calculate the expected thermal shift for any N, $[P]$, $[L]_T$ and K_D (Figure 10). Several important considerations for thermal shift assays can be inferred from these plots. Firstly, that it is necessary for the ligand to be in molar excess for a significant thermal shift to be observed. Secondly, that thermal shift assays are not particularly sensitive to μM or weaker binding affinities without employing high ligand concentrations. Most importantly, there is a strong inverse correlation between protein size and the magnitude of the expected thermal shift (Figure 11). In general, a protein can be expected to exhibit half the thermal shift experienced by a protein that is half as large. The model is validated by thermal shift assays from the literature, which report similar thermal shifts to those predicted.[20, 21] It is important to note that the thermal shift model is based upon the chain-length dependent stability and

relationship between ΔT_M and $\Delta\Delta G°$ for single domain proteins. Large, multidomain proteins may exhibit much larger thermal shifts than predicted if the assay monitors only the unfolding of a small ligand binding domain. In addition, experimental thermal shift assays may actually be monitoring the conversion of monomeric folded proteins into an aggregated oligomeric unfolded form. In this case the equilibrium is more complicated.

5.5. Conclusions

An existing dataset of thermodynamic properties for single domain proteins was expanded and refined. In addition to increasing the number of proteins, improvements to the dataset included the removal of proteins that are not thermodynamically two-state and proteins that are stabilized by cofactors or ionic species. Proteins were reclassified as either mesophilic or thermophilic based on the source organism, rather than the T_M values of individual proteins. These changes allowed for the creation of more robust models for the chain length dependent stability of proteins in mesophiles and thermophiles.

The large dataset made it possible to reassess the concept of convergence temperatures, which have been used in several recent publications. These temperatures were thought to be points at which the enthalpy and entropy per residue converged to a single value across all proteins. The results do not support the existence of a true convergence temperature. This result agrees with the conclusions of earlier reviews.[8]

The reparametrized models were validated by comparison to published experimental data. In agreement with experiment, the model predicts that there is no chain length dependence to the loss of stability experienced by *E. coli* proteins at temperatures just above the optimal growth temperature. In addition, the model correctly predicts that shorter proteins show greater thermal

stability in mesophilic organisms. The model for thermophilic proteins did not mirror experimental data as closely, but was markedly different from the model for mesophiles. This contrasts with the previous parametrizations, which predicted broadly similar behavior for mesophilic and thermophilic proteins. A comparison of predicted behavior of mesophilic and thermophilic proteins makes it possible to examine the strategies used by thermophiles to adapt the thermal stability of their proteome to high temperature environments. Thermophilic proteins appear to be stabilized by three key changes. Firstly, the maximum stability of large proteins is significantly higher in thermophiles than in mesophiles. Secondly, proteins derived from thermophiles exhibit a generally lower change in heat capacity upon unfolding, broadening their stability curves. Thirdly, the temperature of maximum stability is shifted to higher temperatures for thermophilic proteins of all lengths. These conclusions should be viewed as preliminary since the dataset for thermophiles is limited in size. Clearly an important goal for the future is to collect more experimental data on thermophilic proteins.

In combination with proteome length distribution data for *E. coli*, the reparametrized model predicts a protein stability distribution that is in good agreement with experiment. Few proteins are expected to be significantly less stable than the average protein, and the average protein is slightly more stable than in the previous parameterization. The changes have not altered the prediction of a long tail at the stable end of the distribution, but this is likely a result of the assumption inherent to the model that all proteins are composed of a single, cooperatively folding domain. Correcting this necessitates either changes to the model, or domain length data for the entire proteome.

A second dataset of mutational stability data for 1124 mutants of 16 proteins was analyzed to derive a relationship between changes in $\Delta G°$ and T_M as a function of chain length. This

relationship was used to model thermal stability changes upon ligand binding. The model predicts several important considerations for the design of thermal shift assays. Firstly, that thermal shift assays are largely insensitive to μM or weaker binding affinities. Secondly, that ligand concentration must be in molar excess of the protein concentration for a shift to be clearly observable. Most importantly, that there is a strong inverse correlation between chain length and the magnitude of the thermal shift, with the thermal shift being halved for every doubling of the protein length. Thus, thermal shift assays will be less sensitive for large proteins.

5.6. Figures

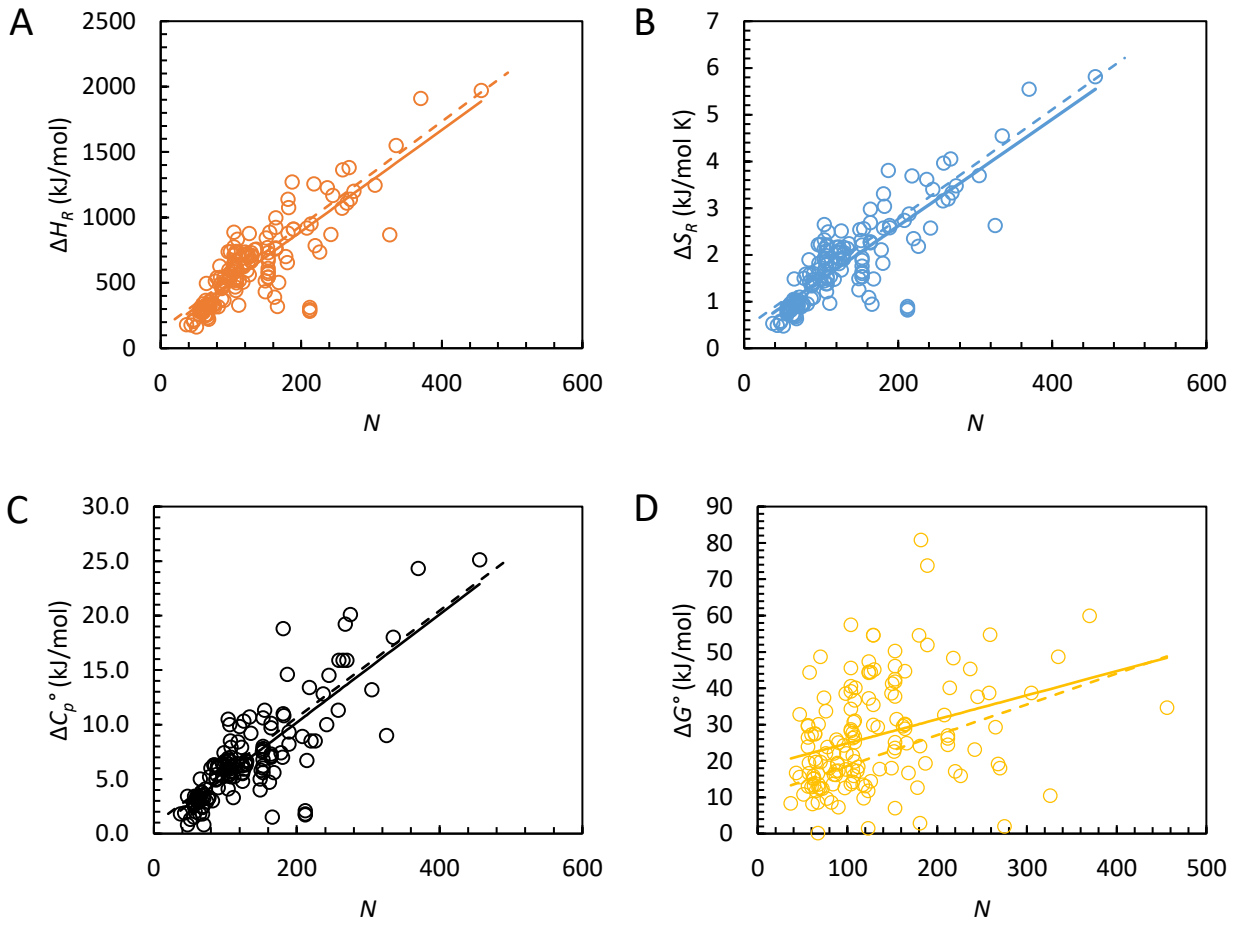


Figure 5.1. Chain length dependence of thermodynamic properties in mesophiles. Plots of **(A)** ΔH_R, **(B)** ΔS_R, **(C)** ΔC_p° and **(D)** ΔG° versus chain length for proteins derived from mesophilic organisms. A total of 150 proteins are included. Linear regression for the new data (solid lines) and from the Sawle and Ghosh analysis (dashed lines).

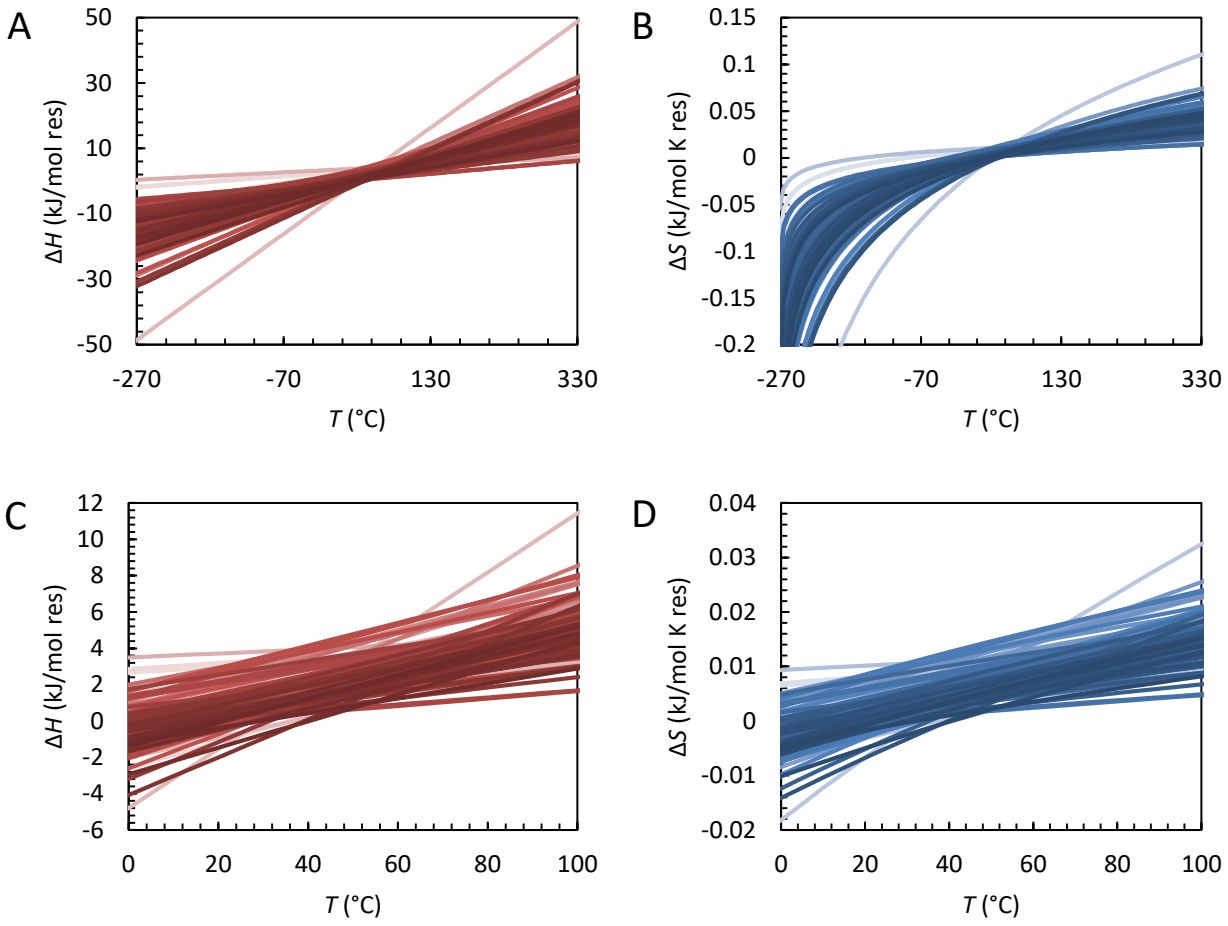


Figure 5.2. There are no convergence temperatures for Δ*H* and Δ*S*. **(A)** Δ*H* per residue and **(B)** Δ*S* per residue for all proteins do approach a common value near 50 °C, however a single convergence temperature cannot be identified. This is more clearly shown in panels **(C)** and **(D)** which display an expanded view.

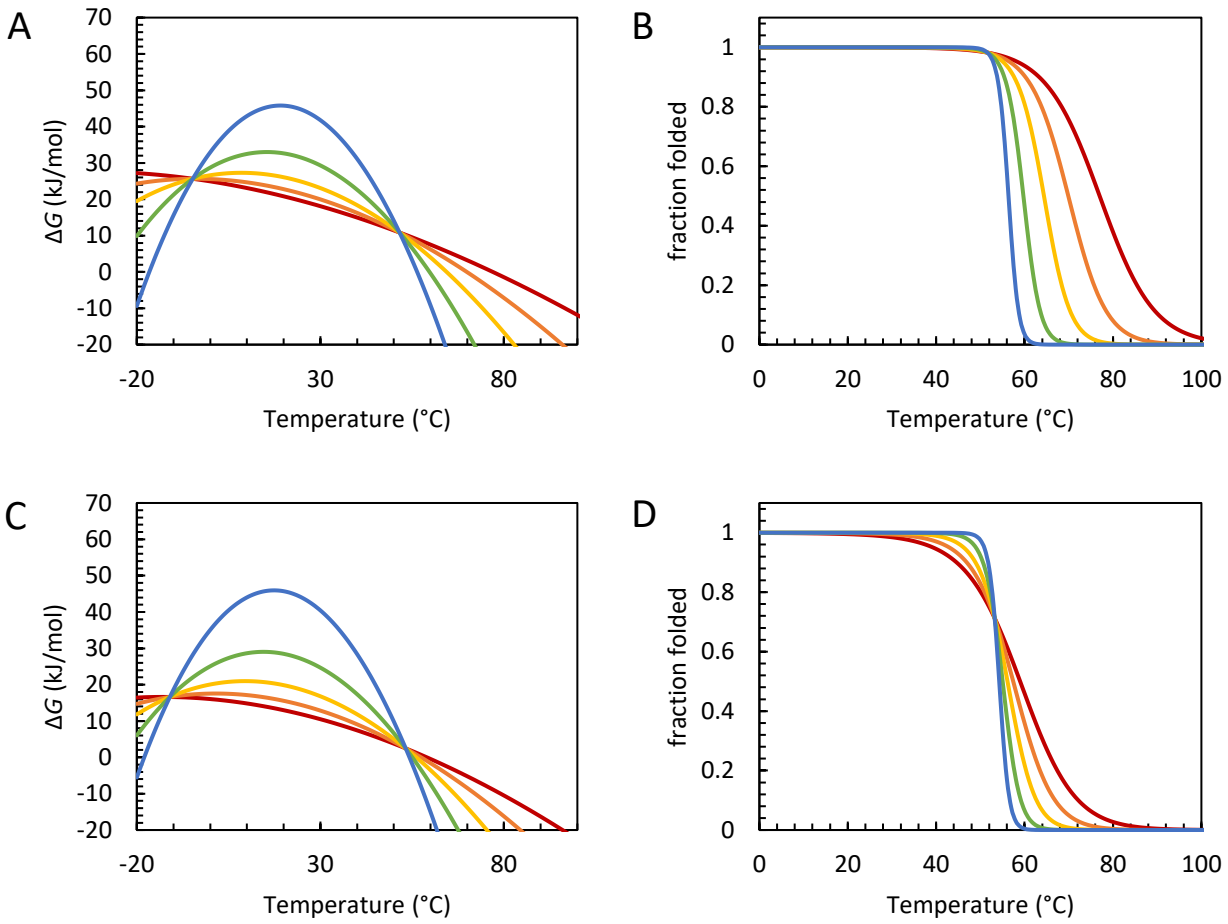


Figure 5.3. Predicted temperature dependent stability for mesophiles. **(A)** Stability curves and **(B)** plots of fraction folded predicted by analysis of the expended database for mesophilic proteins of length 20 (red), 50 (orange), 100 (yellow), 200 (green) and 400 (blue). **(C)** Stability curves and **(D)** plots of fraction folded predicted by the analysis of Sawle and Ghosh for mesophilic proteins. Color coding is the same as in **(A)**. The new analysis suggests that in mesophilic organisms all proteins begin to unfold at approximately the same temperature, however the transition is more cooperative for larger proteins and small proteins are stable over a broader temperature range. In contrast, the analysis of Sawle and Ghosh predicts that smaller proteins begin to unfold at lower temperatures.

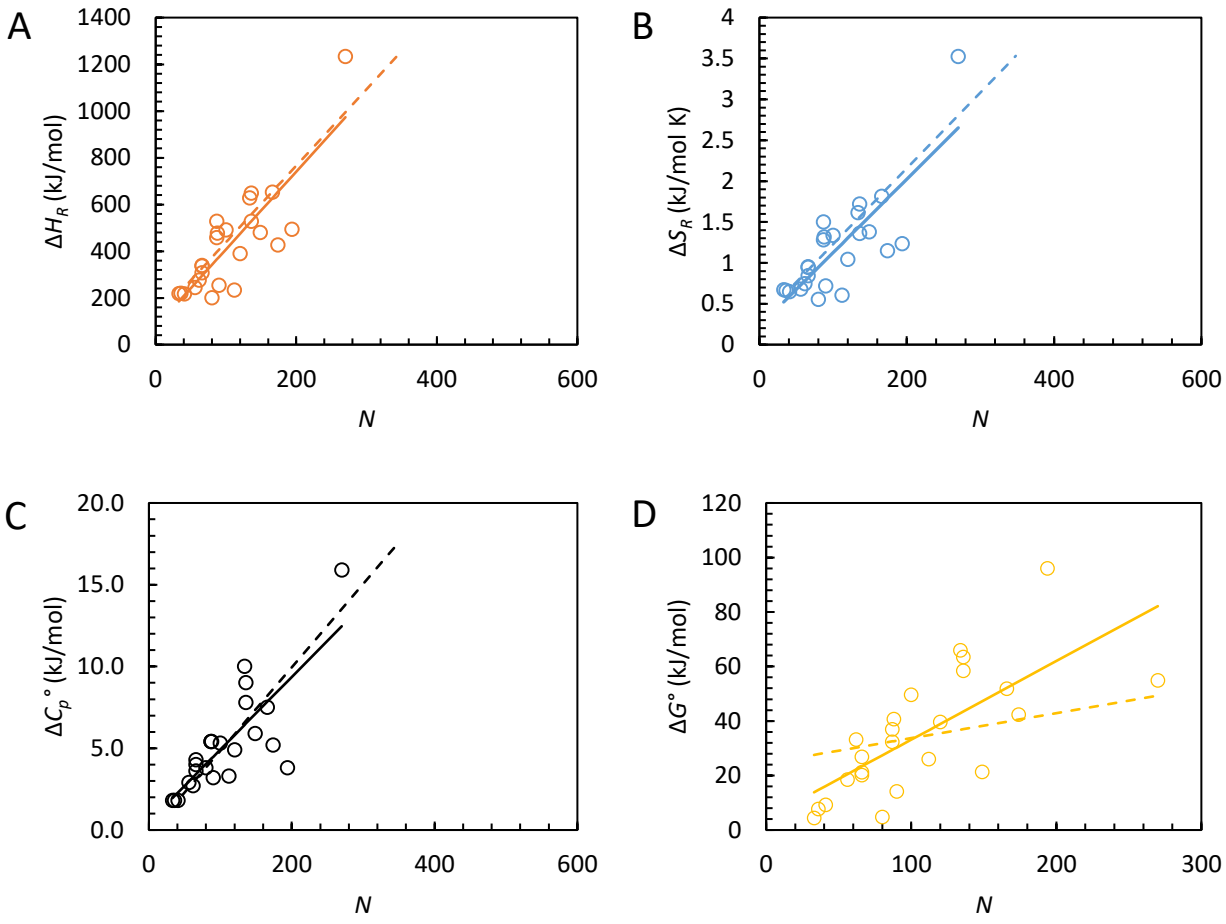


Figure 5.4. Chain length dependence of thermodynamic properties in thermophiles. Plots of **(A)** ΔH_R, **(B)** ΔS_R, **(C)** ΔC_p° and **(D)** ΔG° versus chain length for proteins derived from thermophilic organisms. A total of 24 proteins are included. Linear regression for the new data (solid lines) and from the Sawle and Ghosh analysis (dashed lines).

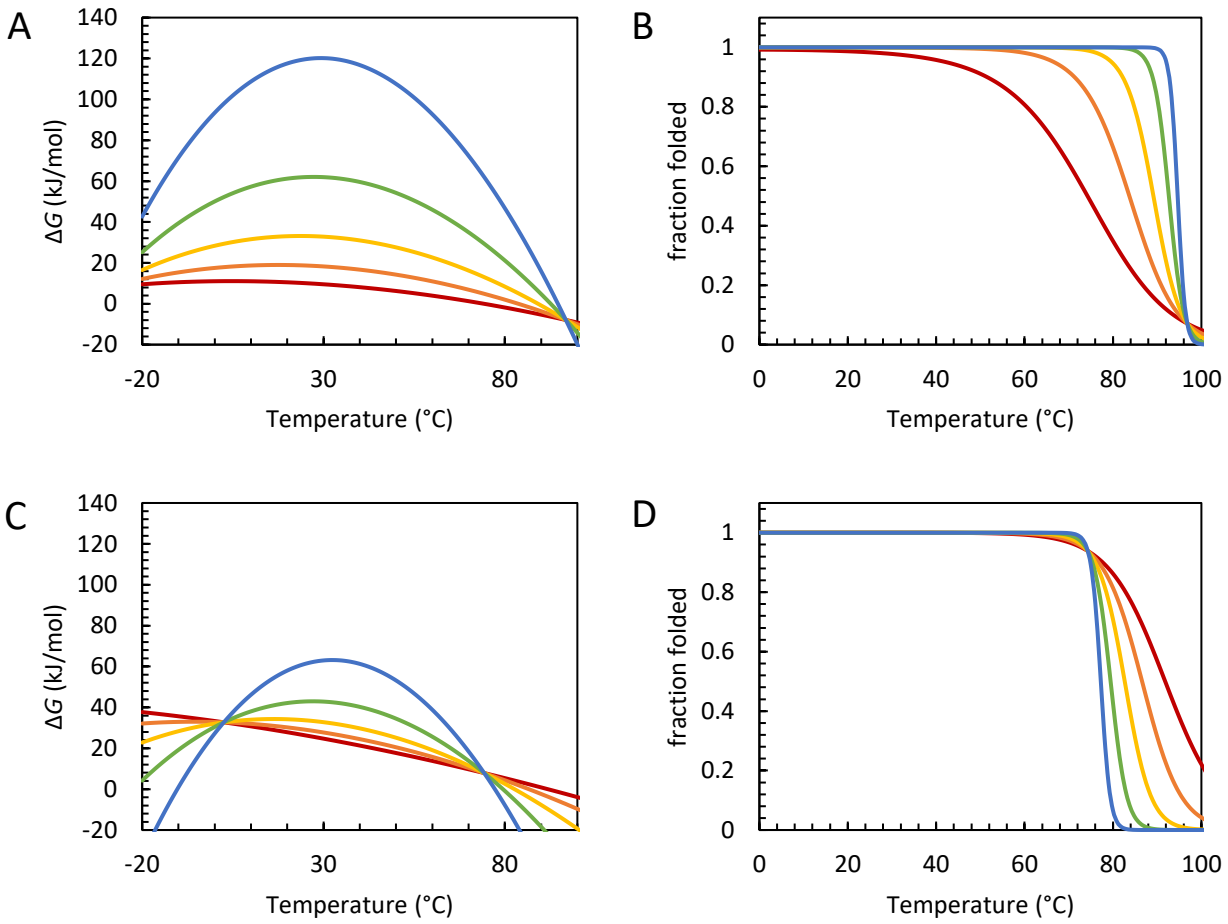


Figure 5.5. Predicted temperature dependent stability for thermophiles. **(A)** Stability curves and **(B)** plots of fraction folded predicted by analysis of the expended database for thermophilic proteins of length 20 (red), 50 (orange), 100 (yellow), 200 (green) and 400 (blue). **(C)** Stability curves and **(D)** plots of fraction folded predicted by the analysis of Sawle and Ghosh for thermophilic proteins. Color coding is the same as in **(A)**. The new analysis predicts very different behavior in thermophilic and mesophilic proteins. Large thermophilic proteins are predicted to be much more stable than mesophilic proteins of all lengths. In addition, stability curves are broadened relative to mesophiles. In contrast, the analysis of Sawle and Ghosh predicts similar behavior in mesophilic and thermophilic proteins. Thermal stabilization is predicted to be the result of a uniform shift to higher temperatures of maximum stability (T_{max}).

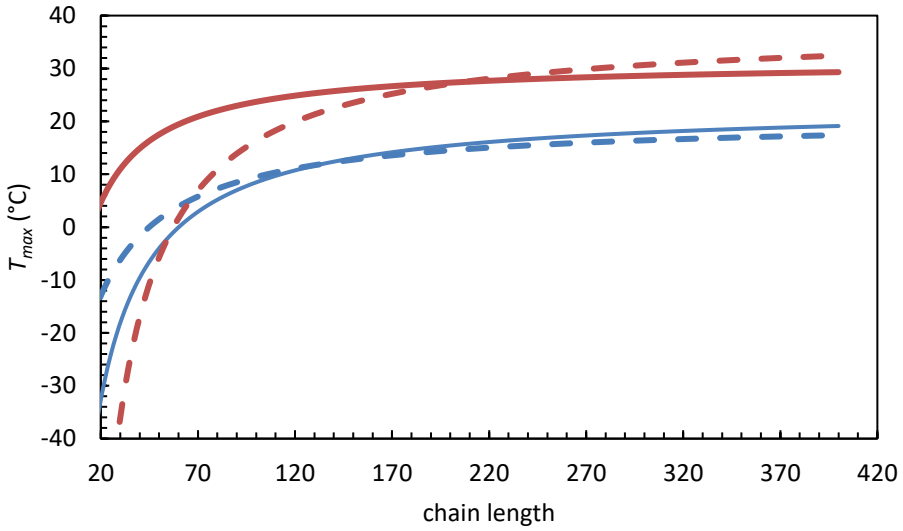


Figure 5.6. Chain-length dependence of the temperature of maximum stability (T_{max}) in proteins from mesophiles (blue) and thermophiles (red) determined from parameters derived from the Sawle and Ghosh (dashed lines) and new (solid lines) analyses. Both analyses predict that T_{max} is higher in thermophilic proteins. The analysis of Sawle and Ghosh predicts that T_{max} depends more strongly on chain length in thermophilic organisms.

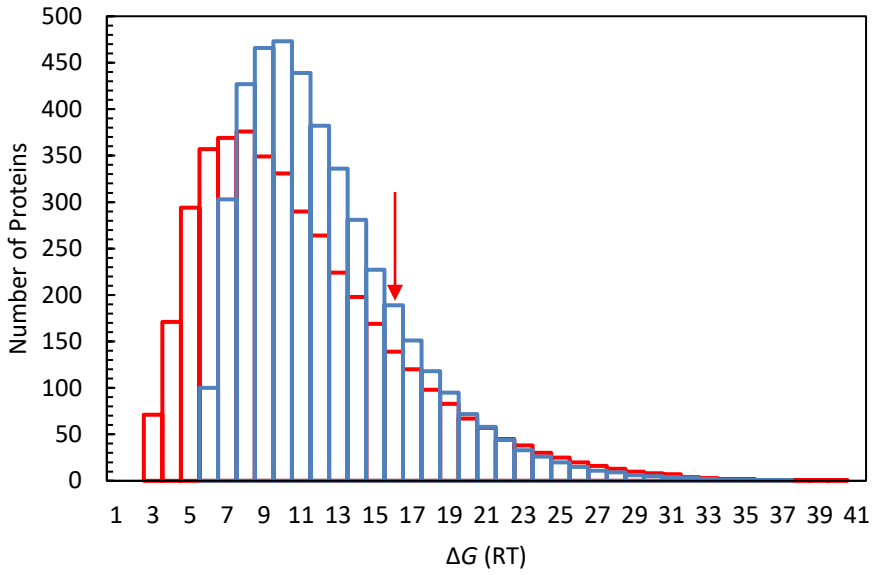
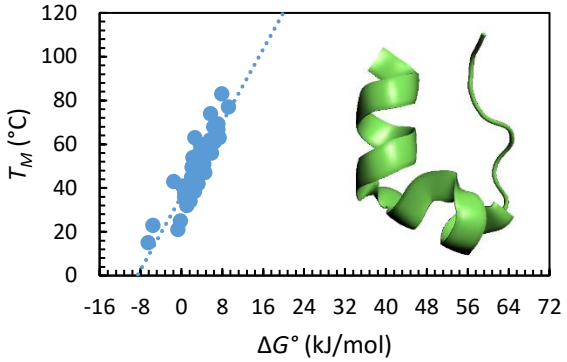
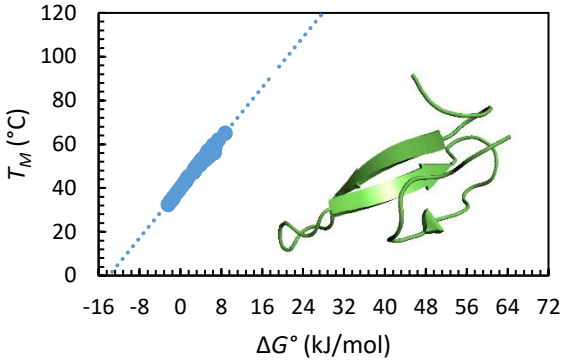


Figure 5.7. Predicted distributions of stability for the proteome of *E. coli* at 37 °C using the chain-length dependent stability parameters predicted by the dataset from Sawle and Ghosh for mesophiles (red) and the modified dataset for mesophiles (blue). The red arrow indicates the predicted stability of the largest protein in the dataset (456 residues).

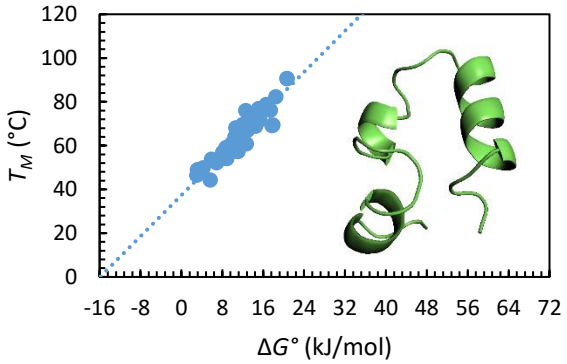
Trp-Cage



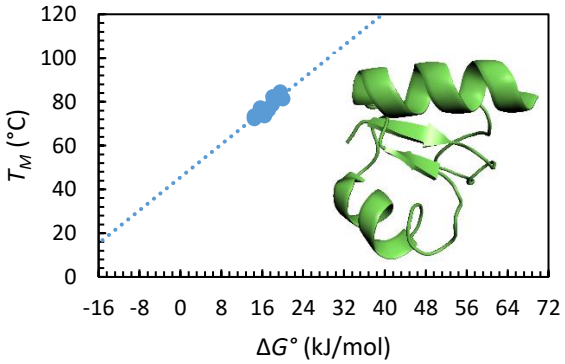
WW domain of hPin1



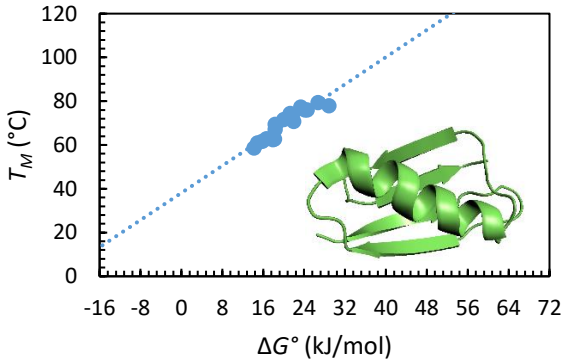
HP36



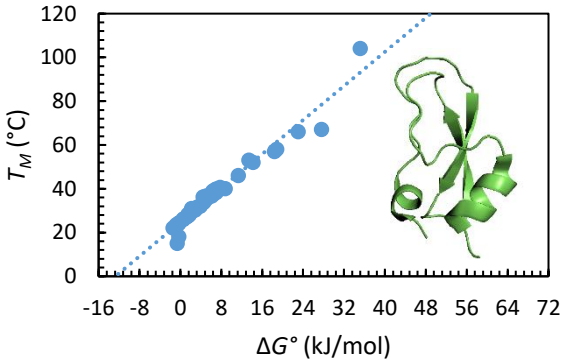
NTL9

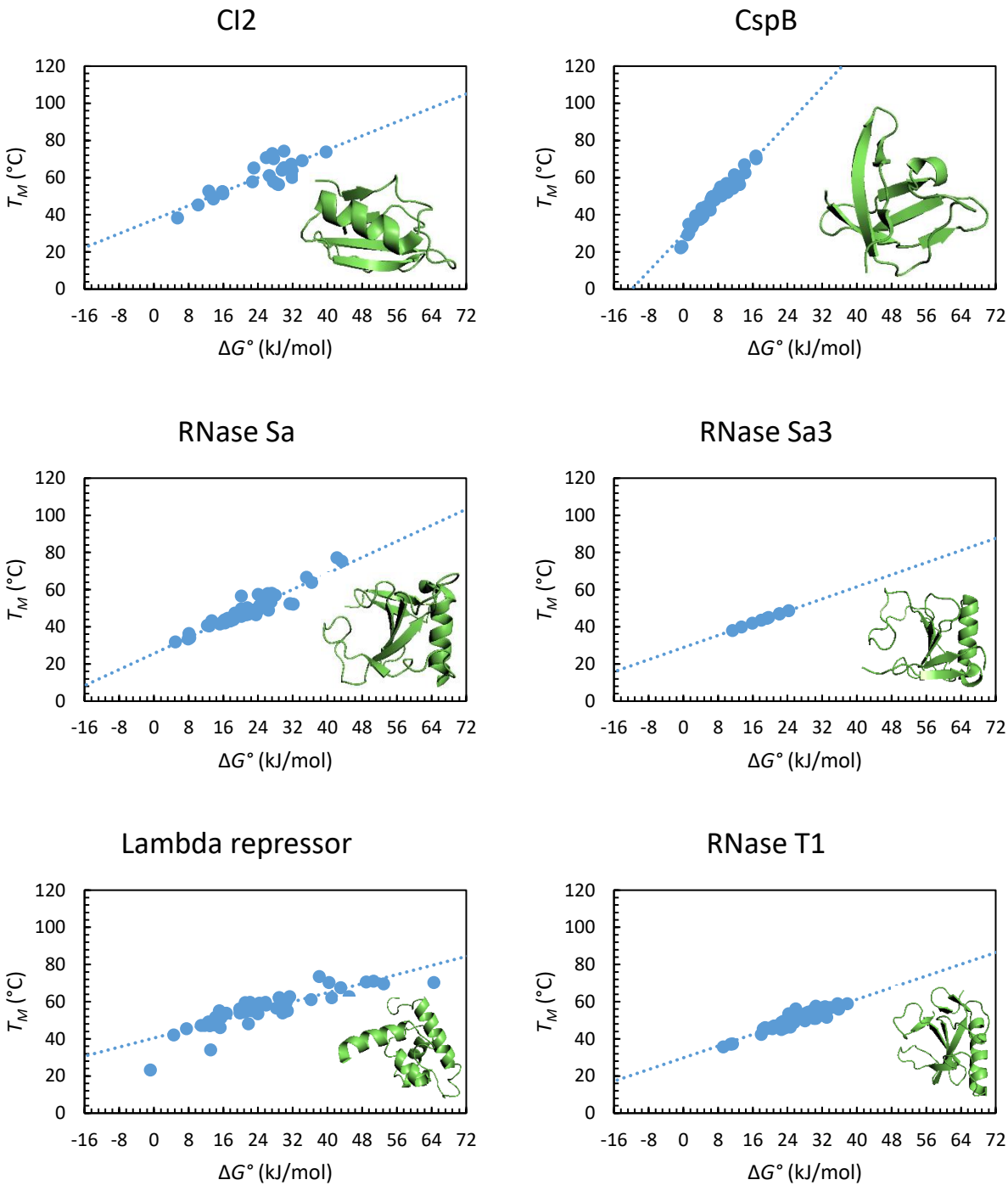


B1 domain of Protein G



BPTI




CI2
CspB
RNase Sa
RNase Sa3
Lambda repressor
RNase T1
T_M (°C)
ΔG° (kJ/mol)
0
20
40
60
80
100
120
-16 -8 0 8 16 24 32 40 48 56 64 72

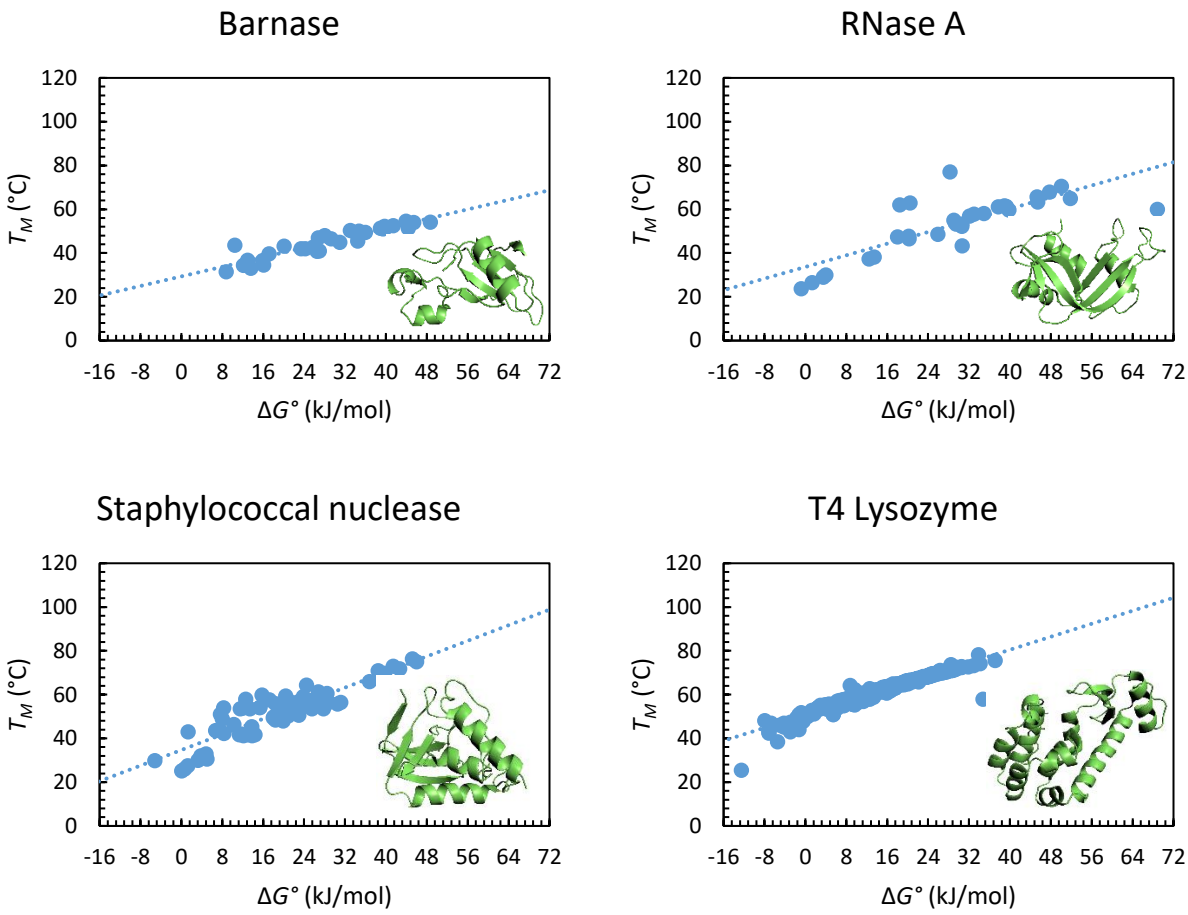


Figure 5.8. Plots of T_M *vs* $\Delta G°$ for mutants of all proteins included in the analysis

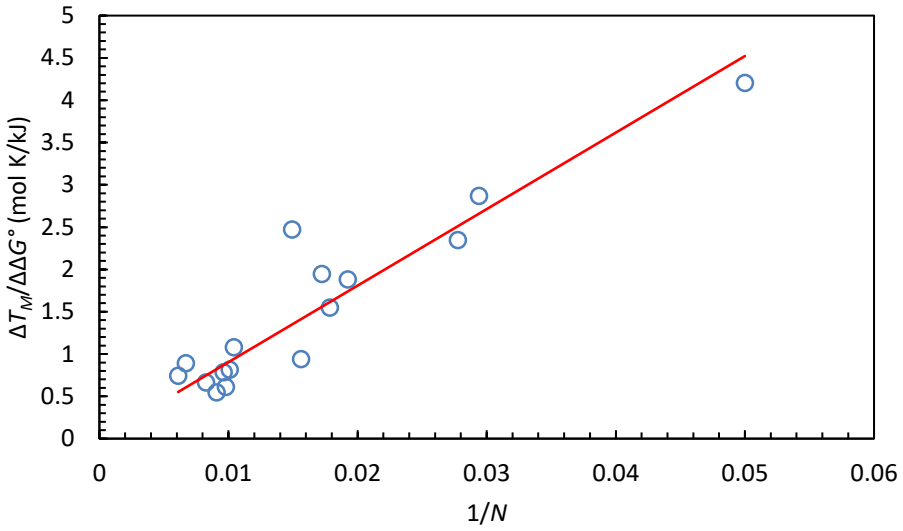


Figure 5.9. The slope of T_M *vs.* $\Delta G°$ plots correlates linearly with $1/N$ ($R^2 = 0.8765$)

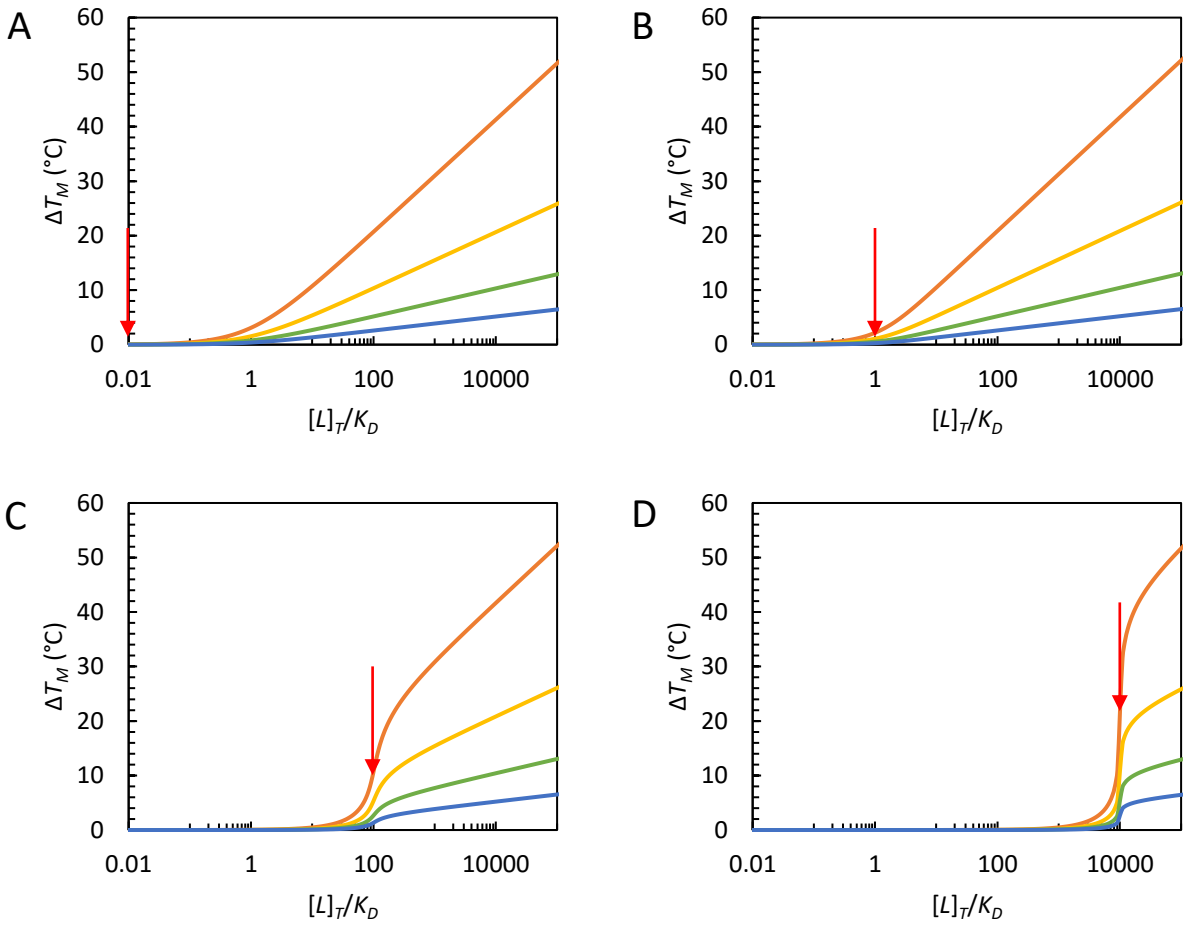


Figure 5.10. The thermal shift resulting from binding of a ligand with a K_D of **(A)** 1 mM **(B)** 10 μM **(C)** 100 nM and **(D)** 1 nM. Individual curves represent proteins of length 50 (orange), 100 (yellow), 200 (green) and 400 (blue). Red arrows indicate $[L]_T$ = 10 μM, which also corresponds to the protein concentration used to generate the plots. This concentration was chosen to agree with experimental methods from the literature.[20, 21]

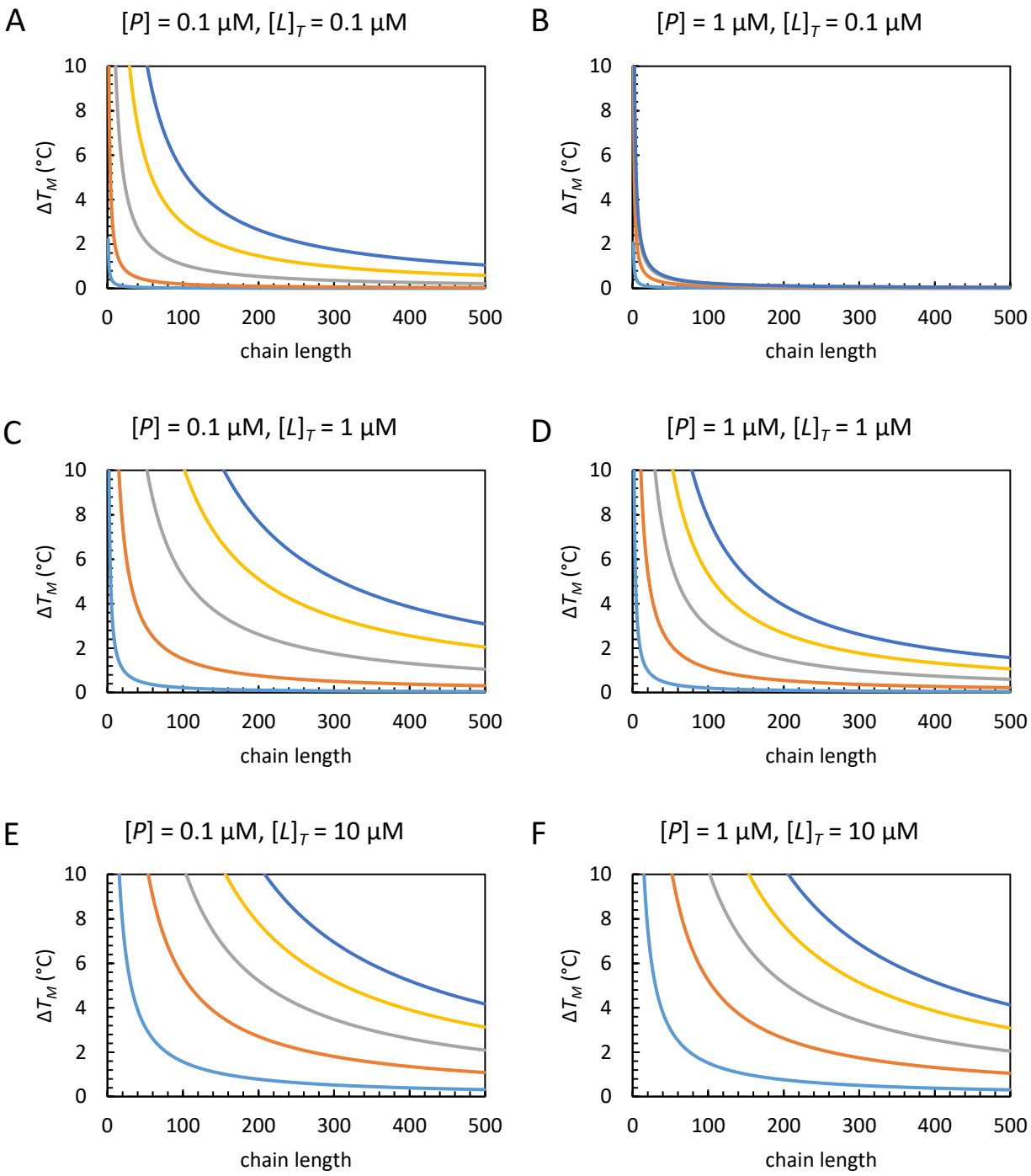


Figure 5.11. Chain length dependence of the thermal shift for a ligand with a K_D of 10 µM (light blue), 1 µM (orange), 100 nM (gray), 10 nM (yellow) and 1 nM (dark blue). Curves were calculated at total protein concentrations of 0.1 and 1 µM and total ligand concentrations of 0.1, 1 and 10 µM.

5.7. Tables

Table 5.1. Thermodynamic parameters for equations (3-5) derived from mesophilic protein data in the Sawle and Ghosh and expanded datasets.

Parameter	Sawle and Ghosh	Reparametrization
m_H (kJ/mol)	3.98	3.88
b_H (kJ/mol)	142	117
m_S (kJ/mol)	0.0117	0.0114
b_S (kJ/mol)	0.420	0.325
m_C (kJ/mol)	0.0491	0.0496
b_C (kJ/mol)	0.835	0.270

Table 5.2. Thermodynamic parameters for equations (3-5) derived from thermophilic protein data in the Sawle and Ghosh and expanded datasets.

Parameter	Sawle and Ghosh	Reparametrization
m_H (kJ/mol)	3.27	3.33
b_H (kJ/mol)	112	75.3
m_S (kJ/mol)	0.00929	0.00899
b_S (kJ/mol)	0.299	0.225
m_C (kJ/mol)	0.0510	0.0443
b_C (kJ/mol)	0.254	0.486

Table 5.3. Thermodynamic parameters for mesophilic proteins. Values of T_M, ΔH_M, ΔS_M, $\Delta C_p°$ and $\Delta G°$ are in units of K, kJ/mol, kJ/mol K, kJ/mol and kJ/mol, respectively.

Protein	N	T_M	ΔH_M	ΔS_M	$\Delta C_p°$	$\Delta G°$	Organism	S-S bonds	PDB
WW domain of FBP28[22]	37	334.1	110.5	0.331	1.8	8.3	homo sapiens	0	-
Hevein[23]	43	363.2	159.0	0.438	1.9	16.8	hevea brasiliensis	4	1HEV
GA module of PAB (K5I/K39V mutant)[24]	47	372.2	213.4	0.573	3.4	15.5	finegoldia magna	0	1PRB
GA module of PAB (K5I mutant)[24]	47	369.0	200.8	0.544	0.8	32.8	finegoldia magna	0	1PRB
lac repressor headpiece[25]	51	338.2	118.0	0.349	1.3	10.9	escherichia coli	0	1LCD
GCN4[26]	56	343.2	259.4	0.756	1.5	29.5	saccharomyces cerevisiae	0	2ZTA
B1 of protein G[27]	56	360.7	258.0	0.715	2.6	29.8	streptococcus sp. GX7805	0	1PGB
OMTKY3[28]	56	358.4	240.2	0.670	2.6	26.4	lophura nycthemera	3	2OVO
B2 of protein G[27]	56	352.6	238.0	0.675	2.9	23.9	streptococcus	0	1PGX

Protein	N	T_M	ΔH_M	ΔS_M	$\Delta C_p°$	$\Delta G°$	Organism	S-S bonds	PDB
OMTKY3[29]	56	340.0	164.0	0.482	2.7	13.0	lophura nycthemera	3	2OVO
Bergerac D48G mutant[30]	57	347.2	225.0	0.648	3.4	19.4	gallus gallus	0	1BK2
α-spectrin SH3 domain[31]	57	339.0	197.0	0.581	2.8	16.6	gallus gallus	0	1SHG
BPTI[32]	58	373.7	300.0	0.803	2.0	44.3	bos taurus	3	5PTI
Sem5[33]	60	352.3	271.1	0.770	3.3	27.2	caenorhabditis elegans	0	1KFZ
Itk (SH3 domain)[34]	60	342.0	178.0	0.520	3.4	12.8	mus musculus	0	2RNA
Engrailed Homeodomain[35]	61	325.3	138.1	0.424	2.9	8.1	drosophila melanogaster	0	2JWT
α-spectrin[36]	62	339.0	197.0	0.581	3.4	15.0	gallus gallus	0	1SHG
Bergerac-SH3[30]	62	339.0	197.0	0.581	3.1	15.9	gallus gallus	0	1SHG
Abl[36]	63	341.5	194.0	0.568	3.3	15.2	homo sapiens	0	1BBZ

Protein	N	T_M	ΔH_M	ΔS_M	$\Delta C_p°$	$\Delta G°$	Organism	S-S bonds	PDB
Tec (SH3 domain)[37]	63	344.0	169.0	0.491	2.9	13.3	mus musculus	0	1GL5
CI2[38]	64	347.0	280.0	0.807	3.3	27.5	hordeum vulgare	0	1COA
Fyn (SH3 domain)[36]	64	343.6	233.0	0.678	3.3	20.5	homo sapiens	0	1SHF
Tumor suppressor P53 tetramerization domain[39]	64	358.0	220.1	0.615	1.8	27.5	homo sapiens	0	1TUP
Bergerac SHH mutant[30]	65	345.3	230.6	0.668	3.5	19.6	gallus gallus	0	2OAW
Type III Antifreeze[40]	65	319.8	228.4	0.714	5.0	11.8	macrozoarces americanus	0	1MSI
Btk (SH3 domain)[37]	67	353.0	196.0	0.555	3.1	16.5	homo sapiens	0	1AWX
Cold Shock B[41]	67	326.6	154.0	0.471	3.8	8.6	bacillus subtilis	0	1CSP
CspB[42]	67	327.2	58.4	0.178	3.8	0.1	bacillus subtilis	0	1CSP
Trypsin inhibitor-V[43]	68	344.1	267.8	0.778	1.8	30.2	cucurbita maxima	1	1TIN

Protein	N	T_M	ΔH_M	ΔS_M	$\Delta C_p°$	$\Delta G°$	Organism	S-S bonds	PDB
Trypsin inhibitor-V (C3S/C48S)[43]	68	322.4	209.2	0.649	2.5	13.5	cucurbita maxima	0	1TIN
Protein W[44]	68	344.5	152.3	0.442	2.4	12.7	bacteriophage lambda	0	2L6Q
Eglin C[45]	70	358.8	312.7	0.872	0.8	48.6	hirudo medicinalis	0	2TEC
Bergerac SHA mutant[30]	70	339.1	237.5	0.700	3.5	19.8	gallus gallus	0	1SHG
Cold Shock A[46]	70	330.2	181.0	0.548	3.2	12.4	escherichia coli	0	1MJC
Phage 434 Cro[47]	72	330.2	195.0	0.591	4.1	12.4	bacteriophage	0	1ZUG
tendamistat[48]	74	366.2	306.5	0.837	2.9	37.5	streptomyces tendae	2	3AIT
ubiquitin[49]	76	363.2	302.0	0.832	3.3	33.6	homo sapiens	0	1UBQ
Ubiquitin[50]	76	344.0	242.0	0.704	3.2	22.0	saccharomyces cerevisiae	0	1OTR
acyl carrier protein (apo)[51]	77	325.9	160.2	0.492	3.3	9.6	escherichia coli	0	1ACP

Protein	N	T_M	ΔH_M	ΔS_M	$\Delta C_p°$	$\Delta G°$	Organism	S-S bonds	PDB
plasminogen K4 domain[52]	78	335.2	315.0	0.940	5.2	23.8	homo sapiens	3	1PMK
Lambda-Repressor 6-85[53]	80	330.4	284.5	0.861	6.0	18.0	enterobacteria phage lambda	0	1LMB
cytochrome c551[54]	82	320.3	157.7	0.492	3.0	8.6	pseudomonas aeruginosa	0	451C
C-terminal domain of Protein S[55]	84	317.9	281.0	0.884	6.3	13.6	myxococcus xanthus	0	1PRR
Histidine Ec[56]	85	336.6	317.1	0.942	6.2	22.0	escherichia coli	0	1POH
Hpr[57]	85	337.8	268.4	0.795	6.3	16.2	escherichia coli	0	1POH
HPr[58]	87	347.0	247.7	0.714	4.9	17.3	bacillus subtilis	0	2HPR
Histidine Oi[59]	87	331.2	246.9	0.745	5.0	16.1	oceanobacillus iheyensis	0	-
Histidine Bs[59]	88	346.0	322.2	0.931	5.4	25.7	bacillus subtilis	0	1SPH
barstar[60]	89	343.1	291.6	0.850	6.2	19.0	bacillus amyloliquefaciens	0	1BTA

Protein	N	T_M	ΔH_M	ΔS_M	$\Delta C_p°$	$\Delta G°$	Organism	S-S bonds	PDB
N-terminal domain of Protein S[55]	89	340.8	291.0	0.854	6.1	19.4	myxococcus xanthus	0	1NPS
cytochrome b5 (soluble domain)[61]	90	343.2	332.0	0.967	6.0	25.0	bacillus subtilis	0	1CYO
TNfn3[62]	90	329.9	279.5	0.847	6.1	17.3	homo sapiens	0	1TEN
ApoCytochrome b5 (apo)[63]	90	321.5	149.3	0.464	4.2	7.2	leporidae	0	-
Rnase Sa[64]	96	321.6	407.5	1.267	6.4	24.2	streptomyces aureofaciens	1	1RGG
Rnase Sa2[64]	97	314.3	286.2	0.911	5.3	12.5	streptomyces aureofaciens	1	-
stefin A[65]	98	364.0	473.0	1.300	7.4	38.6	homo sapiens	0	1CYV
CT-Acp[66]	98	327.0	290.0	0.887	6.1	17.6	homo sapiens	0	2VH7
Rnase Sa3[64]	99	320.4	391.6	1.222	6.6	22.1	streptomyces aureofaciens	1	1MGR
ONC[67]	104	361.0	530.0	1.468	6.0	57.4	rana pipiens	4	1ONC

Protein	N	T_M	ΔH_M	ΔS_M	$\Delta C_p°$	$\Delta G°$	Organism	S-S bonds	PDB
K25-RNase T1[68]	104	330.1	470.0	1.424	4.1	39.0	aspergillus oryzae	2	8RNT
Q25-Ribonuclease T1[68]	104	327.3	465.0	1.421	5.4	34.4	aspergillus oryzae	2	1RN1
Cytochrome c[69]	104	351.2	447.3	1.274	6.4	40.5	bos taurus	0	2B4Z
RNase T1[64]	104	324.8	442.2	1.362	6.9	28.6	aspergillus oryzae	2	9RNT
cyt c (horse)[70]	104	358.2	440.5	1.230	5.3	45.4	equus caballus	0	1HRC
RNase T1[71]	104	333.9	387.0	1.159	6.7	28.2	aspergillus oryzae	2	9RNT
Ribosomal protein L30E[72]	104	318.9	319.0	1.000	10.5	13.6	saccharomyces cerevisiae	0	1CN7
Epithelial cadherin Domain 2[73]	106	327.5	397.5	1.214	5.9	27.7	mus musculus	0	1EDH
Cytochrome b562[53]	106	340.2	393.3	1.156	10.0	21.3	escherichia coli	0	1QPU
ECAD2[73]	106	327.0	376.6	1.152	5.4	26.2	mus musculus	0	1EDH

Protein	N	T_M	ΔH_M	ΔS_M	$\Delta C_p°$	$\Delta G°$	Organism	S-S bonds	PDB
arc repressor[74]	106	327.2	297.1	0.908	6.7	17.5	enterobacteria phage P22	0	1ARR
Arc repressor-st11 (homodimer)[75]	106	330.7	236.4	0.715	5.5	14.2	bacteriophage P22	0	1PAR
Trp Repressor[76]	107	363.5	383.3	1.054	6.1	30.7	escherichia coli	0	2OZ9
subtilisin inhibitor[77]	107	323.4	313.0	0.968	8.5	15.9	streptomyces albogriseolus	2	3SIC
barnase[78]	108	328.3	486.0	1.481	5.7	36.6	bacillus amyloliquefaciens	0	1BNI
thioredoxin[79]	108	360.3	411.0	1.141	7.0	31.0	escherichia coli	1	2TRX
thioredoxin[80]	108	356.6	407.3	1.142	7.9	26.4	escherichia coli	1	2TRX
cyt c (yeast isozyme 1)[81]	108	328.6	360.2	1.096	5.7	25.1	saccharomyces cerevisiae	0	1YCC
cyt c (yeast isozyme 1) (C102A iso- 1)[82]	108	329.4	292.5	0.888	5.2	19.9	saccharomyces cerevisiae	0	1YCC
barnase[83]	109	326.9	546.0	1.670	6.2	40.1	bacillus amyloliquefaciens	0	1BNJ

Protein	N	T_M	ΔH_M	ΔS_M	$\Delta C_p°$	$\Delta G°$	Organism	S-S bonds	PDB
Prion Protein (121-231)[84]	111	338.0	214.2	0.634	3.3	17.3	mus musculus	1	1AG2
cyt c (yeast isozyme 2)[82]	112	327.7	282.0	0.861	5.2	18.3	saccharomyces cerevisiae	0	1YEA
S16 (Meso)[85]	117	332.2	244.0	0.735	6.4	13.4	chlamydia pneumoniae	0	-
Hisactophilin[86]	118	326.7	234.3	0.717	8.4	9.7	dictyostelium discoideum	0	1HCE
CheY (Meso)[87]	120	328.0	294.8	0.899	9.8	13.1	bacillus subtilis	0	3CHY
α-lactalbumin (apo)[88]	123	316.5	276.1	0.873	7.9	11.7	bos taurus	4	1HFZ
α-lactalbumin (apo)[89]	123	301.2	147.0	0.488	6.2	1.5	bos taurus	4	1HFZ
RNase B[90]	124	337.4	520.0	1.541	5.6	47.3	bos taurus	4	1RBB
RNase A[91]	124	336.0	499.6	1.487	5.5	44.2	bos taurus	4	5RSA
RNaseA[92]	124	337.2	481.0	1.426	4.8	44.5	bos taurus	4	3RN3

Protein	N	T_M	ΔH_M	ΔS_M	$\Delta C_p°$	$\Delta G°$	Organism	S-S bonds	PDB
RNaseA[93]	124	335.1	457.3	1.365	6.3	37.2	bos taurus	4	3RN3
ROP[94]	126	344.2	580.0	1.685	10.3	44.4	escherichia coli	0	1RPR
Histone[95]	126	322.6	263.2	0.816	5.9	14.4	gallus gallus	0	1HIO
Lysozyme (chicken)[96]	129	349.2	550.3	1.576	6.6	54.6	gallus gallus	4	1HEL
lysozyme (hen)[97]	129	351.2	537.0	1.529	6.3	54.6	gallus gallus	4	1LYS
lysozyme (hen)[98]	129	337.2	435.0	1.290	6.4	35.4	gallus gallus	4	1LYS
lysozyme (hen)[99]	129	328.2	429.0	1.307	6.7	29.9	gallus gallus	4	1LYS
lysozyme (human)[100]	130	342.0	502.5	1.470	6.6	45.1	homo sapiens	4	1LZ1
rHFoB[101]	134	348.0	484.9	1.393	10.7	29.3	methanobacterium formicicum	0	-
Staphylococcus nuclease[102]	136	327.2	336.8	1.029	9.2	17.6	staphylococcus aureus	0	1STN

Protein	N	T_M	ΔH_M	ΔS_M	$\Delta C_p°$	$\Delta G°$	Organism	S-S bonds	PDB
Odorant Binding Protein[103]	149	342.2	391.2	1.143	4.0	38.6	sus scrofa	2	1A3Y
PsbQ[104]	149	338.0	256.5	0.759	5.0	18.0	spinacia oleracea	0	1VYK
α-sarcin[105]	150	325.6	569.0	1.748	5.8	41.2	aspergillus giganteus	2	1DE3
N1 Cellulose-Binding Domain of Endoglucanase CenC[106]	152	322.5	391.4	1.214	7.5	22.6	cellulomonas fimi	1	1ULO
Metmyoglobin (horse)[107]	153	354.7	548.0	1.545	7.8	50.2	equus caballus	0	1YMB
Metmyoglobin (rat)[107]	153	357.0	473.0	1.325	6.2	46.1	rattus norvegicus	0	-
Metmyoglobin (raccoon)[107]	153	355.3	465.0	1.309	6.8	41.7	procyon lotor	0	-
myoglobin (whale)[108]	153	353.3	460.0	1.302	10.6	23.6	physeter catodon	0	1MBO
Metmyoglobin (opossum)[107]	153	353.5	435.0	1.231	5.6	42.5	didelphis	0	-
Metmyoglobin (armadillo)[107]	153	352.0	417.0	1.185	6.0	37.7	armadillo	0	-

Protein	N	T_M	ΔH_M	ΔS_M	$\Delta C_p°$	$\Delta G°$	Organism	S-S bonds	PDB
Metmyoglobin (carp)[107]	153	341.3	370.0	1.084	7.6	25.1	cyprinus carpio	0	-
IL-1β[109]	153	310.2	227.0	0.732	8.0	7.0	homo sapiens	0	6I1B
RNaseH (Meso)[110]	155	339.0	502.1	1.481	11.3	31.5	escherichia coli	0	2RN2
β-Lactoglobulin[111]	162	359.1	325.9	0.908	4.7	29.5	bos taurus	2	2Q2M
T4 Lysozyme (Pseudo WT)[112]	164	337.6	581.0	1.721	9.7	44.7	enterobacteria phage T4 sensu lato	0	1L63
T4 lysozyme[113]	164	324.4	507.0	1.563	10.1	30.1	enterobacteria phage T4 sensu lato	0	2LZM
T4 lysozyme (S44[A])[112]	164	327.3	442.0	1.350	7.1	30.0	enterobacteria phage T4 sensu lato	0	1L63
T4 lysozyme (A42K)[112]	164	327.4	440.0	1.344	7.0	30.0	enterobacteria phage T4 sensu lato	0	1L63
T4 lysozyme (K48[A])[112]	164	327.6	430.0	1.313	7.3	28.8	enterobacteria phage T4 sensu lato	0	1L63
Erythropoietin[114]	166	333.8	259.4	0.777	1.5	24.9	homo sapiens	2	1BUY

Protein	N	T_M	ΔH_M	ΔS_M	$\Delta C_p°$	$\Delta G°$	Organism	S-S bonds	PDB
flavodoxin[115]	168	330.3	264.0	0.799	5.6	16.7	anabaena PCC7119	0	1FTG
C-terminal Ada protein[116]	178	316.8	284.0	0.896	7.4	12.6	escherichia coli	0	1SFE
Chymotrypsin Inhibitor[117]	180	357.9	548.1	1.531	7.0	54.5	schizolobium parahyba	2	-
Kunitz type soybean trypsin inhibitor[118]	181	332.0	428.9	1.292	11.0	24.1	glycine max	2	1AVU
Orosomucoid[119]	181	331.0	350.0	1.057	18.8	2.9	homo sapiens	2	3APU
SRBP[120]	182	351.0	836.8	2.384	10.8	80.8	homo sapiens	3	1RBP
Diphtheria toxin subunit A[121]	187	317.4	460.2	1.450	14.6	19.2	corynebaterium diphteriae	0	1DDN
DsbA (reduced)[122]	189	350.0	720.9	2.060	8.2	73.9	escherichia coli	0	1A23
DsbA (oxidized)[122]	189	341.6	618.8	1.812	9.3	51.9	escherichia coli	1	1A23
met repressor[123]	208	326.4	505.0	1.547	8.9	32.6	escherichia coli	0	1CMB

Protein	N	T_M	ΔH_M	ΔS_M	$\Delta C_p°$	$\Delta G°$	Organism	S-S bonds	PDB
Prion protein (VRQ)[124]	212	344.5	251.1	0.729	2.1	27.0	ovis aries	1	1UW3
Prion protein (ARQ)[124]	212	343.3	241.0	0.702	1.8	26.1	ovis aries	1	1UW3
Prion protein (ARR)[124]	212	341.9	229.3	0.671	1.7	24.4	ovis aries	1	1UW3
Glucanase[125]	214	322.5	611.7	1.897	6.7	40.0	bacillus macerans	1	1CPM
Endoglucanase 3[126]	218	332.0	707.1	2.130	13.4	48.3	trichoderma reesei	6	1H8V
ADK (Yeast)[127]	220	320.7	340.0	1.060	8.5	17.1	saccharomyces cerevisiae	0	1AKY
Subtilisin Inhibitor[8]	226	323.4	312.5	0.967	8.5	15.9	streptomyces albogriseolus	2	3SSI
α-chymotrypsin[128]	237	330.2	677.8	2.053	12.8	45.4	bos taurus	5	5CHA
ApoLipoprotein A1[129]	242	327.0	410.0	1.254	10.0	23.1	homo sapiens	0	2A01
α-chymotrypsinogen[130]	245	335.2	619.2	1.848	14.5	37.6	bos taurus	5	2CGA

Protein	N	T_M	ΔH_M	ΔS_M	$\Delta C_p°$	$\Delta G°$	Organism	S-S bonds	PDB
dephospho-EIN[131]	258	330.1	585.8	1.775	11.3	38.7	escherichia coli	0	1ZYM
Carbonic anhydrase[132]	259	337.2	795.0	2.358	15.9	54.7	bos taurus	0	1G6V
ESBL[133]	265	340.0	585.8	1.723	15.9	29.3	bacillus lichenformis	0	4BLM
Tryptophan synthase alpha-subunit[134]	268	327.3	503.8	1.539	19.2	19.1	escherichia coli	0	1WQ5
Cellulase CenA[135]	270	329.6	447.7	1.358	15.9	18.0	cellulomonas fimi	2	1GU3
subtilisin BPN'[136]	275	331.7	370.0	1.115	20.1	2.0	bacillus amyloliquefaciens	0	2ST1
arabinose binding protein[137]	305	326.7	635.1	1.944	13.2	38.6	escherichia coli	0	1ABE
Seed coat soybean peroxidase (apo)[138]	326	311.0	308.0	0.990	9.0	10.4	glycine max	4	1FHF
Diphtheria toxin subunit B[121]	335	331.1	795.0	2.401	18.0	48.6	corynebaterium diphteriae	0	1DDN
Maltose Binding Protein[139]	370	336.0	1010.0	3.006	24.3	59.9	escherichia coli	0	1OMP

Protein	N	T_M	ΔH_M	ΔS_M	$\Delta C_p°$	$\Delta G°$	Organism	S-S bonds	PDB
Hsp70 (Mge1p)[140]	456	325.0	765.3	2.355	25.1	34.6	saccharomyces cerevisiae	1	4B9Q

Table 5.4. Thermodynamic parameters for thermophilic proteins. Values of T_M, ΔH_M, ΔS_M, $\Delta C_p°$ and $\Delta G°$ are in units of K, kJ/mol, kJ/mol K, kJ/mol and kJ/mol, respectively.

Protein	*N*	T_M	ΔH_M	ΔS_M	$\Delta C_p°$	$\Delta G°$	Organism	S-S bonds	PDB
Peripheral Subunit-Binding Domain 36[141]	36	321.2	126.8	0.395	1.8	7.6	geobacillus stearothermophilus	0	2PDD
Peripheral Subunit-Binding Domain 41[141]	41	326.2	132.6	0.406	1.8	9.2	geobacillus stearothermophilus	0	2PDD
NTL9[142]	56	358.8	202.9	0.566	2.9	18.6	geobacillus stearothermophilus	0	2HBB
Sso7d[143]	62	372.0	274.0	0.737	2.7	33.2	sulfolobus solfataricus	0	1SSO
Sac7d[144]	66	364.1	274.9	0.755	3.6	26.9	sulfolobus acidocaldarius	0	1SAP
Cold Shock Tm[145]	66	355.2	261.9	0.737	4.3	21.2	thermotoga maritima	0	1G6P
Cold Shock Bc[146]	66	350.1	245.0	0.700	4.0	20.1	bacillus caldolyticus	0	1C9O
cytochrome c552[54]	80	360.5	153.6	0.426	3.8	4.9	hydrogenobacter thermophilus	0	1AYG
Histidine Bh[59]	87	355.5	364.0	1.024	5.4	32.1	bacillus halodurans	0	-

Protein	N	T_M	ΔH_M	ΔS_M	$\Delta C_p°$	$\Delta G°$	Organism	S-S bonds	PDB
Histidine containing protein (HPr)[147]	87	348.8	397.5	1.140	5.4	36.9	streoptococcus thermophilus	0	-
Histidine Bst[59]	88	361.3	414.2	1.146	5.4	40.5	geobacillus stearothermophilus	0	1Y4Y
HU DNA binding protein[148]	90	350.7	183.0	0.522	3.2	14.1	thermotoga maritima	0	1B8Z
Ribosomal protein L30E[72]	100	367.0	459.0	1.251	5.3	49.5	thermococcus celer	0	1H7M
S16 (Thermo)[85]	112	384.2	270.0	0.703	3.3	26.0	aquifex aeolicus	0	3BN0
CheY (Thermo)[87]	120	374.0	395.3	1.057	4.9	39.6	thermotoga maritima	0	1TMY
rHPyA1[101]	134	387.3	771.5	1.992	10.0	66.0	pyrococcus GB-3a	0	-
rHMfA[101]	136	377.2	686.6	1.820	9.0	63.5	methanothermus fervidus	0	1B67
rHMfB[101]	136	386.0	628.9	1.629	7.8	58.4	methanothermus fervidus	0	1BFM
Glutamate Dehydrogenase Domain II[149]	149	342.6	301.2	0.879	5.9	21.4	thermotoga maritima	0	1B26

Protein	N	T_M	ΔH_M	ΔS_M	$\Delta C_p°$	$\Delta G°$	Organism	S-S bonds	PDB
RNaseH (thermo)[110]	166	359.0	548.1	1.527	7.5	51.7	thermus thermophilus	0	1RIL
O6-methyguanine-DNA methyltransferase[116]	174	371.6	419.0	1.128	5.2	42.3	thermococcus kodakarensis	0	1MGT
Ssh10B[150]	194	398.2	590.0	1.482	3.8	95.9	sulfolobus shibattae	0	1Y9X
Cellulase E2[135]	270	345.4	795.0	2.302	15.9	54.8	thermomonospora fusca	2	1TF4

Table 5.5. Mutants of Trp-cage.[151-155]

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
TC12b	36	1.3	Biochemistry (2014) 53, 6011-6021
TC11b1	40	2.4	Biochemistry (2014) 53, 6011-6021
TC9b	51	4.4	Biochemistry (2014) 53, 6011-6021
R16Norvaline (nva)	36	0.6	Biochemistry (2014) 53, 6011-6021
control (NAUYUQWLKDGGPSSGRAA)	23	-5.6	Biochemistry (2014) 53, 6011-6021
TC10b	41	1.6	Biochemistry (2014) 53, 6011-6021
TC10b P19A	15	-6.5	Biochemistry (2014) 53, 6011-6021
TC10b S14A	21	-0.7	Biochemistry (2014) 53, 6011-6021
TC10b P17A	46	2.6	Biochemistry (2014) 53, 6011-6021
TC10b P18A	47	4.3	Biochemistry (2014) 53, 6011-6021
TC10b P12W/P18A	40	1.9	Biochemistry (2014) 53, 6011-6021
TC13b	68	6.3	Biochemistry (2014) 53, 6011-6021
TC13b P12W	77	9.2	Biochemistry (2014) 53, 6011-6021
TC13b P12W/P18A	50	2.1	Biochemistry (2014) 53, 6011-6021
tr-TC16b	60	5.4	Biochemistry (2014) 53, 6011-6021
tr-TC16b R16nva	50	2.1	Biochemistry (2014) 53, 6011-6021
TC16b	74	5.7	Biochemistry (2014) 53, 6011-6021
TC16b R16nva	63	2.6	Biochemistry (2014) 53, 6011-6021
TC16b P19A	43	-1.5	Biochemistry (2014) 53, 6011-6021
TC10b	56	5.3	Biochemistry (2011) 50, 1143-1152
TC10b	40	1.9	Biochemistry (2011) 50, 1143-1152

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
D9E	57	4.6	Biochemistry (2011) 50, 1143-1152
D9E	41	0.2	Biochemistry (2011) 50, 1143-1152
D9E/R16E	56	3.6	Biochemistry (2011) 50, 1143-1152
D9E/R16E	25	-0.2	Biochemistry (2011) 50, 1143-1152
D9E/R16K	60	3.7	Biochemistry (2011) 50, 1143-1152
R16K	54	2.3	Biochemistry (2011) 50, 1143-1152
D9E/R16O	52	2.9	Biochemistry (2011) 50, 1143-1152
R16O	45	2.0	Biochemistry (2011) 50, 1143-1152
R16O	38	0.6	Biochemistry (2011) 50, 1143-1152
TC10b P17A	42	3.3	Protein Eng. Des. Sel. (2008) 21, 171-185
TC10b P17G	32	1.1	Protein Eng. Des. Sel. (2008) 21, 171-185
TC10b P18L	34	1.7	Protein Eng. Des. Sel. (2008) 21, 171-185
TC10b P18A	47	4.6	Protein Eng. Des. Sel. (2008) 21, 171-185
TC10b P12A	43	3.1	Protein Eng. Des. Sel. (2008) 21, 171-185
TC10b P12A/P18A	39	2.6	Protein Eng. Des. Sel. (2008) 21, 171-185
TC5b WT	46	3.3	J. Am. Chem. Soc. (2011) 133, 18750-18759
TC10b G10 (D)A	67	7.1	J. Am. Chem. Soc. (2011) 133, 18750-18759
TC10b (D)N	56	5.9	J. Am. Chem. Soc. (2011) 133, 18750-18759
TC10b (D)Q	69	7.1	J. Am. Chem. Soc. (2011) 133, 18750-18759
TC10b K8A	61	6.4	Org. Biomol. Chem. (2008) 6, 4287-4289
TC10b S13A	63	7.4	Org. Biomol. Chem. (2008) 6, 4287-4289
TC10b G15(D)A	62	5.6	Org. Biomol. Chem. (2008) 6, 4287-4289

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
TC16b	83	7.9	Org. Biomol. Chem. (2008) 6, 4287-4289

Table 5.6. Mutants of hPin1 WW domain.[156]

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
hPin1 L7A	37.8	-0.66	Protein Sci. (2009) 18, 1806-1813
hPin1 L7I	49.3	2.99	Protein Sci. (2009) 18, 1806-1813
hPin1 L7V	44.0	1.27	Protein Sci. (2009) 18, 1806-1813
hPin1 L7NVa	48.8	2.79	Protein Sci. (2009) 18, 1806-1813
hPin1 P8A	47.4	2.51	Protein Sci. (2009) 18, 1806-1813
hPin1 P8G	47.7	2.72	Protein Sci. (2009) 18, 1806-1813
hPin1 P9A	56.0	5.76	Protein Sci. (2009) 18, 1806-1813
hPin1 P9G	53.1	4.49	Protein Sci. (2009) 18, 1806-1813
hPin1 G10A	49.0	3.00	Protein Sci. (2009) 18, 1806-1813
hPin1 W11F	35.0	-1.52	Protein Sci. (2009) 18, 1806-1813
hPin1 E12A	52.6	4.53	Protein Sci. (2009) 18, 1806-1813
hPin1 K13A	59.6	6.45	Protein Sci. (2009) 18, 1806-1813
hPin1 R14A	39.2	-0.28	Protein Sci. (2009) 18, 1806-1813
hPin1 M15A	51.8	4.08	Protein Sci. (2009) 18, 1806-1813
hPin1 S16A	54.0	4.70	Protein Sci. (2009) 18, 1806-1813
hPin1 S16G	47.6	2.69	Protein Sci. (2009) 18, 1806-1813
hPin1 R17A	58.8	6.53	Protein Sci. (2009) 18, 1806-1813
hPin1 R17G	57.3	5.64	Protein Sci. (2009) 18, 1806-1813
hPin1 S18A	58.4	6.69	Protein Sci. (2009) 18, 1806-1813
hPin1 S18G	56.5	6.64	Protein Sci. (2009) 18, 1806-1813
hPin1 S19A	57.1	5.98	Protein Sci. (2009) 18, 1806-1813
hPin1 S19G	56.0	6.58	Protein Sci. (2009) 18, 1806-1813

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
hPin1 G20A	48.9	2.94	Protein Sci. (2009) 18, 1806-1813
hPin1 R21A	50.9	3.85	Protein Sci. (2009) 18, 1806-1813
hPin1 R21G	51.5	3.74	Protein Sci. (2009) 18, 1806-1813
hPin1 V22A	54.2	5.48	Protein Sci. (2009) 18, 1806-1813
hPin1 Y23A	33.9	-2.03	Protein Sci. (2009) 18, 1806-1813
hPin1 Y23L	45.3	1.61	Protein Sci. (2009) 18, 1806-1813
hPin1 Y23F	52.8	4.39	Protein Sci. (2009) 18, 1806-1813
hPin1 Y24F	51.4	3.77	Protein Sci. (2009) 18, 1806-1813
hPin1 Y24W	52.9	4.22	Protein Sci. (2009) 18, 1806-1813
hPin1 F25A	32.5	-2.39	Protein Sci. (2009) 18, 1806-1813
hPin1 F25L	42.5	0.83	Protein Sci. (2009) 18, 1806-1813
hPin1 F25Y	62.0	7.45	Protein Sci. (2009) 18, 1806-1813
hPin1 N26D	36.0	-1.31	Protein Sci. (2009) 18, 1806-1813
hPin1 H27A	57.7	6.04	Protein Sci. (2009) 18, 1806-1813
hPin1 H27G	50.5	3.71	Protein Sci. (2009) 18, 1806-1813
hPin1 I28A	54.2	5.05	Protein Sci. (2009) 18, 1806-1813
hPin1 I28G	47.2	2.55	Protein Sci. (2009) 18, 1806-1813
hPin1 T29A	44.3	1.34	Protein Sci. (2009) 18, 1806-1813
hPin1 T29G	34.4	-1.76	Protein Sci. (2009) 18, 1806-1813
hPin1 T29S	50.8	3.85	Protein Sci. (2009) 18, 1806-1813
hPin1 T29D	42.9	0.97	Protein Sci. (2009) 18, 1806-1813
hPin1 N30A	53.3	4.57	Protein Sci. (2009) 18, 1806-1813
hPin1 N30G	65.0	8.76	Protein Sci. (2009) 18, 1806-1813

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
hPin1 A31G	40.9	0.32	Protein Sci. (2009) 18, 1806-1813
hPin1 S32A	56.9	5.45	Protein Sci. (2009) 18, 1806-1813
hPin1 S32G	50.1	3.18	Protein Sci. (2009) 18, 1806-1813
hPin1 Q33A	53.1	4.17	Protein Sci. (2009) 18, 1806-1813
hPin1 W34A	52.9	4.86	Protein Sci. (2009) 18, 1806-1813
hPin1 W34F	58.0	6.12	Protein Sci. (2009) 18, 1806-1813
hPin1 E35A	50.3	3.50	Protein Sci. (2009) 18, 1806-1813
hPin1 R36A	56.7	5.33	Protein Sci. (2009) 18, 1806-1813
hPin1 Δ3,4P37	55.1	5.00	Protein Sci. (2009) 18, 1806-1813
hPin1 S38A	58.8	6.76	Protein Sci. (2009) 18, 1806-1813
hPin1 S38G	58.2	6.21	Protein Sci. (2009) 18, 1806-1813

Table 5.7. Mutants of HP36.[157-165]

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
WT-HP36	73.0	13.47	Shifeng
N68A/K70M-HP36	90.6	20.67	Shifeng
N68A/K70M/P62A-HP36	71.8	15.36	Shifeng
N68A/K70M/W64A-HP36	78.7	16.61	Shifeng
N68A/K70M/W64L-HP36	74.4	14.98	Shifeng
N68A/K70M/W64K-HP36	74.9	15.02	Shifeng
N68A/K70M/P62A/W64L-HP36	67.9	11.46	Shifeng
WT-HP35	70.0	13.81	J. Mol. Biol. (1996) 260, 126–134
N27H-HP36	68.9	14.53	J. Mol. Biol. (2003) 329, 625–630
N27H/F35A-HP36	66.9	12.92	J. Mol. Biol. (2003) 329, 625–630
WT-HP35	69.1	17.70	J. Am. Chem. Soc. (2016) 138, 6498-6505
N19β3N-HP35	44.2	5.65	J. Am. Chem. Soc. (2016) 138, 6498-6505
W23β3W-HP35	59.4	11.21	J. Am. Chem. Soc. (2016) 138, 6498-6505
Q26β3Q-HP35	57.4	10.17	J. Am. Chem. Soc. (2016) 138, 6498-6505
K30β3K-HP35	57.2	11.09	J. Am. Chem. Soc. (2016) 138, 6498-6505
N19ACPC-HP35 (t-(1S,2S)-2-aminocyclopentyl-1-carboxylic acid)	54.0	8.91	J. Am. Chem. Soc. (2016) 138, 6498-6505
W23ACPC-HP35 (t-(1S,2S)-	56.9	8.91	J. Am. Chem. Soc. (2016) 138, 6498-6505

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
2-aminocyclopentyl-1-carboxylic acid)			
Q26ACPC-HP35 (t-(1S,2S)- 2-aminocyclopentyl-1-carboxylic acid)	57.1	9.00	J. Am. Chem. Soc. (2016) 138, 6498-6505
K30APC (t- (3R,4S)-4-aminopyrrolidine-3-carboxylic acid)	69.4	17.91	J. Am. Chem. Soc. (2016) 138, 6498-6505
WT-HP36	61.1	11.13	Biopolymers (2015) 103, 627-637
F51Z-HP36 (4-fluoroPhe)	66.0	11.51	Biopolymers (2015) 103, 627-637
F51X-HP36 (4-methylPhe)	60.8	10.29	Biopolymers (2015) 103, 627-637
F58Z-HP36 (4-fluoroPhe)	60.2	10.00	Biopolymers (2015) 103, 627-637
F58X-HP36 (4-methylPhe)	57.7	8.79	Biopolymers (2015) 103, 627-637
F51Z/F58Z-HP36 (4-fluoroPhe, 4-methylPhe)	63.2	10.67	Biopolymers (2015) 103, 627-637
F51Z/F58X-HP36 (4-flouroPhe, 4-methylPhe)	64.3	11.67	Biopolymers (2015) 103, 627-637
F51X/F58Z-HP36 (4-methylPhe, 4-fluoroPhe)	62.7	11.25	Biopolymers (2015) 103, 627-637
K24Nle-HP36	76.9	15.06	Proc. Natl. Acad. Sci. USA (2005) 102, 7517–7522
A18V-HP35	61.9	11.23	J. Mol. Biol. (2006) 359, 546–553
A18S-HP35	63.9	10.46	J. Mol. Biol. (2006) 359, 546–553
N68A/K70M/F47L-HP36	69.6	11.92	Biochemistry (2009) 48, 4607-4616
N68A/K70M/F51L-HP36	76.0	12.59	Biochemistry (2009) 48, 4607-4616
N68A/K70M/F58L-HP36	60.9	10.96	Biochemistry (2009) 48, 4607-4616
N68A/K70M/F47L/F51L-HP36	68.1	10.63	Biochemistry (2009) 48, 4607-4616
N68A/K70M/F47L/F58L-HP36	48.9	3.14	Biochemistry (2009) 48, 4607-4616
N68A/K70M/F51L/F58L-HP36	46.4	3.01	Biochemistry (2009) 48, 4607-4616

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
N68A/K70M/F47L/F51L/F58L-HP36	49.8	4.27	Biochemistry (2009) 48, 4607-4616
N68A/K70M/W64L-HP36	74.4	14.98	Biochemistry (2009) 48, 4607-4616
WT-HP36	61.1	9.92	Biochemistry (2010) 49, 4255–4263
P62Hyp-HP36 (4R,2S)-4-hydroxyproline	59.0	8.70	Biochemistry (2010) 49, 4255–4263
P62Flp-HP36 (4R,2S)-4-fluoroproline	53.8	5.86	Biochemistry (2010) 49, 4255–4263
P62Mop-HP36 (4R,2S)-4-methoxyproline	56.8	8.08	Biochemistry (2010) 49, 4255–4263
P62hyp-HP36 (4S,2S)-4-hydroxyproline	54.1	8.08	Biochemistry (2010) 49, 4255–4263
P62flp-HP36 (4S,2S)-4-fluoroproline	60.8	12.68	Biochemistry (2010) 49, 4255–4263
P62mop-HP36 (4S,2S)-4-methoxyproline	52.2	6.86	Biochemistry (2010) 49, 4255–4263
WT-HP36	73.0	13.47	Biochemistry (2007) 46, 7497-7505
N68A-HP36	76.1	17.41	Biochemistry (2007) 46, 7497-7505
K70M-HP36	82.2	18.45	Biochemistry (2007) 46, 7497-7505
N68A/K70M-HP36	90.6	20.67	Biochemistry (2007) 46, 7497-7505

Table 5.8. Mutants of NTL9.

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	References
N42FCN/Y25F-NTL9	73.43	14.5	Ivan
N42FCN/Y25W-NTL9	72.28	14.5	Ivan
K2FCN/Y25F-NTL9	76.20	17.2	Ivan
K2FCN/Y25W-NTL9	73.71	16.5	Ivan
K2FCN/Y25F/Q33F-NTL9	81.40	20.0	Ivan
K2FCN/Y25F/Q33W-NTL9	84.47	19.5	Ivan
K2FCN/Y25F/K51F-NTL9	82.22	18.1	Ivan
Y25W-NTL9	77.08	15.6	Ivan
WT-NTL9	78.70	18.0	Ivan

Table 5.9. Mutants of the B1 domain of Protein G.[166]

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
B1 A6A53	79.3	26.69	Structure (1999) 7, 1333-1343
B1 AE53	77.8	28.83	Structure (1999) 7, 1333-1343
B1 A6R	77.2	23.35	Structure (1999) 7, 1333-1343
B1 A6E	76.0	24.18	Structure (1999) 7, 1333-1343
B1 A6E/A53E	75.9	24.60	Structure (1999) 7, 1333-1343
B1 A6E/A53R	74.3	21.25	Structure (1999) 7, 1333-1343
B1 A6K	74.0	21.25	Structure (1999) 7, 1333-1343
B1 A6K/A53E	73.8	21.71	Structure (1999) 7, 1333-1343
B1 A6K/A53R	72.2	21.25	Structure (1999) 7, 1333-1343

Table 5.10. Mutants of BPTI.[167, 168]

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
C14A/C30A/C38A/C51A-BPTI	39.2	7.00	J. Mol. Biol. (1995) 249, 388-397
R1A/C14A/C30A/C38A/C51A-BPTI	35.3	5.14	J. Mol. Biol. (1995) 249, 388-397
P2A/C14A/C30A/C38A/C51A-BPTI	29.4	2.20	J. Mol. Biol. (1995) 249, 388-397
D3A/C14A/C30A/C38A/C51A-BPTI	40.6	7.73	J. Mol. Biol. (1995) 249, 388-397
F4A/C14A/C30A/C38A/C51A/M52L-BPTI	18.0	-0.30	J. Mol. Biol. (1995) 249, 388-397
L6A/C14A/C30A/C38A/C51A-BPTI	34.6	4.71	J. Mol. Biol. (1995) 249, 388-397
E7A/C14A/C30A/C38A/C51A-BPTI	28.0	1.78	J. Mol. Biol. (1995) 249, 388-397
P8A/C14A/C30A/C38A/C51A/M52L-BPTI	36.7	5.54	J. Mol. Biol. (1995) 249, 388-397
P9A/C14A/C30A/C38A/C51A-BPTI	33.7	4.58	J. Mol. Biol. (1995) 249, 388-397
Y10A/C14A/C30A/C38A/C51A-BPTI	30.2	2.18	J. Mol. Biol. (1995) 249, 388-397
T11A/C14A/C30A/C38A/C51A/M52 L-BPTI	39.0	6.74	J. Mol. Biol. (1995) 249, 388-397
G12A/C14A/C30A/C38A/C51A-BPTI	26.1	0.58	J. Mol. Biol. (1995) 249, 388-397
P13A/C14A/C30A/C38A/C51A-BPTI	30.2	2.48	J. Mol. Biol. (1995) 249, 388-397
K15A/C14A/C30A/C38A/C51A-BPTI	36.6	5.02	J. Mol. Biol. (1995) 249, 388-397
R17A/C14A/C30A/C38A/C51A-BPTI	37.3	5.93	J. Mol. Biol. (1995) 249, 388-397
I18A/C14A/C30A/C38A/C51A-BPTI	28.3	1.51	J. Mol. Biol. (1995) 249, 388-397
I19A/C14A/C30A/C38A/C51A-BPTI	24.1	-0.37	J. Mol. Biol. (1995) 249, 388-397
R20A/C14A/C30A/C38A/C51A-BPTI	26.0	0.60	J. Mol. Biol. (1995) 249, 388-397
F22A/C14A/C30A/C38A/C51A-BPTI	24.5	-0.18	J. Mol. Biol. (1995) 249, 388-397
N24A/C14A/C30A/C38A/C51A-BPTI	23.4	-0.83	J. Mol. Biol. (1995) 249, 388-397

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
K26A/C14A/C30A/C38A/C51A-BPTI	39.1	7.16	J. Mol. Biol. (1995) 249, 388-397
G28A/C14A/C30A/C38A/C51A-BPTI	31.9	3.85	J. Mol. Biol. (1995) 249, 388-397
L29A/C14A/C30A/C38A/C51A-BPTI	39.5	6.92	J. Mol. Biol. (1995) 249, 388-397
Q31A/C14A/C30A/C38A/C51A-BPTI	31.7	3.37	J. Mol. Biol. (1995) 249, 388-397
T32A/C14A/C30A/C38A/C51A-BPTI	38.8	6.29	J. Mol. Biol. (1995) 249, 388-397
V34A/C14A/C30A/C38A/C51A-BPTI	30.2	2.85	J. Mol. Biol. (1995) 249, 388-397
Y35A/C14A/C30A/C38A/C51A/M52L-BPTI	30.9	2.19	J. Mol. Biol. (1995) 249, 388-397
G36A/C14A/C30A/C38A/C51A-BPTI	23.8	-0.47	J. Mol. Biol. (1995) 249, 388-397
G37A/C14A/C30A/C38A/C51A-BPTI	22.0	-1.43	J. Mol. Biol. (1995) 249, 388-397
R39A/C14A/C30A/C38A/C51A-BPTI	39.2	6.62	J. Mol. Biol. (1995) 249, 388-397
K41A/C14A/C30A/C38A/C51A-BPTI	36.2	4.43	J. Mol. Biol. (1995) 249, 388-397
R42A/C14A/C30A/C38A/C51A-BPTI	35.6	5.11	J. Mol. Biol. (1995) 249, 388-397
N44A/C14A/C30A/C38A/C51A/M52L-BPTI	15.0	-0.63	J. Mol. Biol. (1995) 249, 388-397
K46A/C14A/C30A/C38A/C51A-BPTI	39.8	7.03	J. Mol. Biol. (1995) 249, 388-397
S47A/C14A/C30A/C38A/C51A-BPTI	27.3	1.40	J. Mol. Biol. (1995) 249, 388-397
E49A/C14A/C30A/C38A/C51A-BPTI	37.9	6.50	J. Mol. Biol. (1995) 249, 388-397
D50A/C14A/C30A/C38A/C51A-BPTI	36.6	6.13	J. Mol. Biol. (1995) 249, 388-397
M52A/C14A/C30A/C38A/C51A-BPTI	27.1	1.16	J. Mol. Biol. (1995) 249, 388-397
R53A/C14A/C30A/C38A/C51A-BPTI	38.3	7.02	J. Mol. Biol. (1995) 249, 388-397
T54A/C14A/C30A/C38A/C51A-BPTI	38.5	6.74	J. Mol. Biol. (1995) 249, 388-397
G56A/C14A/C30A/C38A/C51A-BPTI	37.8	6.46	J. Mol. Biol. (1995) 249, 388-397

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
G57A/C14A/C30A/C38A/C51A-BPTI	37.7	6.60	J. Mol. Biol. (1995) 249, 388-397
WT-BPTI	104.0	35.15	Biochemistry (1997) 36, 5323-5335
C30A/C51A-BPTI	66.0	23.01	Biochemistry (1997) 36, 5323-5335
C30V/C51A-BPTI	67.0	27.61	Biochemistry (1997) 36, 5323-5335
C30G/C51A-BPTI	53.0	13.39	Biochemistry (1997) 36, 5323-5335
C30T/C51A-BPTI	58.0	18.83	Biochemistry (1997) 36, 5323-5335
C30S/C51A-BPTI	57.0	18.41	Biochemistry (1997) 36, 5323-5335
C30A/C51S-BPTI	52.0	14.23	Biochemistry (1997) 36, 5323-5335
C30S/C51S-BPTI	46.0	11.30	Biochemistry (1997) 36, 5323-5335
C30G/C51M-BPTI	40.0	8.79	Biochemistry (1997) 36, 5323-5335

Table 5.11. Mutants of Chymotrypsin Inhibitor 2.[38, 169, 170]

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
WT-CI2*	64.0	29.55	Biochemistry (1993) 32, 11259-11269
WT-CI2*	73.8	39.69	Biochemistry (1993) 32, 11259-11269
L27A-CI2*	48.5	13.72	Biochemistry (1993) 32, 11259-11269
V38A-CI2*	61.1	26.57	Biochemistry (1993) 32, 11259-11269
I39V-CI2*	57.7	22.69	Biochemistry (1993) 32, 11259-11269
I39V-CI2*	67.3	31.64	Biochemistry (1993) 32, 11259-11269
I48V-CI2*	69.1	34.16	Biochemistry (1993) 32, 11259-11269
I48A-CI2*	52.5	15.80	Biochemistry (1993) 32, 11259-11269
V66A-CI2*	45.3	10.18	Biochemistry (1993) 32, 11259-11269
L68A-CI2*	52.3	15.85	Biochemistry (1993) 32, 11259-11269
V70A-CI2*	51.3	15.90	Biochemistry (1993) 32, 11259-11269
I76V-CI2*	65.4	29.95	Biochemistry (1993) 32, 11259-11269
I76A-CI2*	38.3	5.41	Biochemistry (1993) 32, 11259-11269
I76A-CI2*	50.3	13.20	Biochemistry (1993) 32, 11259-11269
I48A/I76V-CI2*	52.8	12.62	Biochemistry (1993) 32, 11259-11269
WT-CI2	74.3	29.96	Protein Eng. (1994) 7, 103-108
A43E-CI2	73.0	27.28	Protein Eng. (1994) 7, 103-108
A45E-CI2	70.2	27.57	Protein Eng. (1994) 7, 103-108
A43K/A45E-CI2	70.7	25.94	Protein Eng. (1994) 7, 103-108
A44P-CI2	65.2	23.01	Protein Eng. (1994) 7, 103-108
WT-CI2	63.8	31.84	Protein Eng. (1994) 7, 777-782
SA31-CI2	58.0	27.53	Protein Eng. (1994) 7, 777-782

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
EA33EA34-CI2	60.2	31.84	Protein Eng. (1994) 7, 777-782
SG31EA33EA34-CI2	56.6	28.41	Protein Eng. (1994) 7, 777-782
SA31EA33EA34-CI2	56.3	28.70	Protein Eng. (1994) 7, 777-782

Table 5.12. Mutants of Cold shock protein B.[171]

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
WT*	55.3	9.71	J. Mol. Biol. (2007) 366, 842–856
6H-WT*	49.7	7.70	J. Mol. Biol. (2007) 366, 842–856
WT	52.3	9.63	J. Mol. Biol. (2007) 366, 842–856
E3R	70.2	16.74	J. Mol. Biol. (2007) 366, 842–856
E3R*	71.6	16.69	J. Mol. Biol. (2007) 366, 842–856
6H-E3R*	66.9	14.10	J. Mol. Biol. (2007) 366, 842–856
E3Q	62.6	14.18	J. Mol. Biol. (2007) 366, 842–856
K5E*	22.2	-0.62	J. Mol. Biol. (2007) 366, 842–856
K5Q*	38.6	4.40	J. Mol. Biol. (2007) 366, 842–856
6H-K5Q*	37.6	3.80	J. Mol. Biol. (2007) 366, 842–856
N10D	57.7	12.04	J. Mol. Biol. (2007) 366, 842–856
N10K*	41.1	4.42	J. Mol. Biol. (2007) 366, 842–856
E12K*	50.2	8.78	J. Mol. Biol. (2007) 366, 842–856
K13E*	52.7	8.76	J. Mol. Biol. (2007) 366, 842–856
K13Q*	54.0	10.70	J. Mol. Biol. (2007) 366, 842–856
E19K*	50.5	8.68	J. Mol. Biol. (2007) 366, 842–856
E19Q*	53.7	10.12	J. Mol. Biol. (2007) 366, 842–856
V20Q*	39.5	4.56	J. Mol. Biol. (2007) 366, 842–856
6H-V20Q*	30.1	1.26	J. Mol. Biol. (2007) 366, 842–856
V20Q/E3R*	58.2	12.71	J. Mol. Biol. (2007) 366, 842–856
6H-V20Q/E3R*	51.6	8.94	J. Mol. Biol. (2007) 366, 842–856
V20E/E3R*	43.0	5.23	J. Mol. Biol. (2007) 366, 842–856

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
6H-V20E/E3R*	34.7	2.20	J. Mol. Biol. (2007) 366, 842–856
V20K*	37.2	3.40	J. Mol. Biol. (2007) 366, 842–856
6H-V20K*	29.3	1.00	J. Mol. Biol. (2007) 366, 842–856
V20K/E3R*	54.6	9.99	J. Mol. Biol. (2007) 366, 842–856
6H-V20K/E3R*	47.9	7.20	J. Mol. Biol. (2007) 366, 842–856
E21K*	54.3	11.46	J. Mol. Biol. (2007) 366, 842–856
E21Q*	52.5	8.15	J. Mol. Biol. (2007) 366, 842–856
E21Q/E19Q*	52.0	10.16	J. Mol. Biol. (2007) 366, 842–856
D24K*	50.9	7.71	J. Mol. Biol. (2007) 366, 842–856
D24N*	48.4	6.41	J. Mol. Biol. (2007) 366, 842–856
D25K*	35.0	1.24	J. Mol. Biol. (2007) 366, 842–856
D25K/E19Q*	39.4	2.83	J. Mol. Biol. (2007) 366, 842–856
D25Q	44.2	5.31	J. Mol. Biol. (2007) 366, 842–856
D25Q/E19Q*	43.6	4.27	J. Mol. Biol. (2007) 366, 842–856
K39E*	52.8	8.52	J. Mol. Biol. (2007) 366, 842–856
K39Q*	54.9	8.57	J. Mol. Biol. (2007) 366, 842–856
E24K*	55.3	11.46	J. Mol. Biol. (2007) 366, 842–856
E42Q*	54.1	8.58	J. Mol. Biol. (2007) 366, 842–856
E43K*	56.3	12.96	J. Mol. Biol. (2007) 366, 842–856
E43Q*	55.5	11.68	J. Mol. Biol. (2007) 366, 842–856
S48K*	61.6	11.75	J. Mol. Biol. (2007) 366, 842–856
6H-S48K*	56.8	9.97	J. Mol. Biol. (2007) 366, 842–856
S48E	53.0	9.32	J. Mol. Biol. (2007) 366, 842–856

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
E50K*	49.7	6.70	J. Mol. Biol. (2007) 366, 842–856
6H-E50K*	49.7	6.63	J. Mol. Biol. (2007) 366, 842–856
E50Q	42.6	6.22	J. Mol. Biol. (2007) 366, 842–856
E53K*	53.7	10.21	J. Mol. Biol. (2007) 366, 842–856
E53Q*	56.6	12.06	J. Mol. Biol. (2007) 366, 842–856
N55K*	54.3	9.41	J. Mol. Biol. (2007) 366, 842–856
N55D*	59.2	12.66	J. Mol. Biol. (2007) 366, 842–856
R56Q	55.3	11.37	J. Mol. Biol. (2007) 366, 842–856
K65E*	38.2	3.29	J. Mol. Biol. (2007) 366, 842–856
6H−K65E*	33.4	2.00	J. Mol. Biol. (2007) 366, 842–856
K65Q*	48.0	6.43	J. Mol. Biol. (2007) 366, 842–856
6H-K65Q*	44.4	5.27	J. Mol. Biol. (2007) 366, 842–856
E3R/F15A/ D25K/F27A*	44.0	6.05	J. Mol. Biol. (2007) 366, 842–856
E3R/F15A/E19K/F27A*	39.0	4.09	J. Mol. Biol. (2007) 366, 842–856
E3R/F15A/ D25K/F27A*	23.0	-0.31	J. Mol. Biol. (2007) 366, 842–856

Table 5.13. Mutants of RNase Sa.[172-178]

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
WT-RNase Sa	48.4	22.68	Biochemistry (1998) 37, 16192-16200
N39D-RNase Sa	43.2	16.82	Biochemistry (1998) 37, 16192-16200
N39S-RNase Sa	40.4	12.76	Biochemistry (1998) 37, 16192-16200
N39A-RNase Sa	40.8	12.34	Biochemistry (1998) 37, 16192-16200
WT-RNase Sa	47.2	21.51	Protein Sci. (1999) 8, 1843-1849
D25K-RNase Sa	50.2	24.02	Protein Sci. (1999) 8, 1843-1849
E74K-RNase Sa	51.1	24.89	Protein Sci. (1999) 8, 1843-1849
WT-RNase Sa	48.4	23.47	J. Mol. Biol. (2001) 312, 393-404
Y30F-RNase Sa	49.6	20.13	J. Mol. Biol. (2001) 312, 393-404
Y49F-RNase Sa	47.6	19.96	J. Mol. Biol. (2001) 312, 393-404
Y51F-RNase Sa	40.8	12.68	J. Mol. Biol. (2001) 312, 393-404
Y52F-RNase Sa	36.3	8.08	J. Mol. Biol. (2001) 312, 393-404
Y55F-RNase Sa	46.5	21.30	J. Mol. Biol. (2001) 312, 393-404
Y80F-RNase Sa	43.2	13.31	J. Mol. Biol. (2001) 312, 393-404
Y81F-RNase Sa	44.5	17.03	J. Mol. Biol. (2001) 312, 393-404
Y86F-RNase Sa	47.3	18.62	J. Mol. Biol. (2001) 312, 393-404
WT-RNase Sa	48.4	23.47	J. Biol. Chem. (2003) 278, 31790-31795
T18V-RNase Sa	43.7	16.28	J. Biol. Chem. (2003) 278, 31790-31795
T56V-RNase Sa	42.1	16.40	J. Biol. Chem. (2003) 278, 31790-31795
T67V-RNase Sa	48.3	21.30	J. Biol. Chem. (2003) 278, 31790-31795

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
T82V-RNase Sa	42.7	15.90	J. Biol. Chem. (2003) 278, 31790-31795
T5V-RNase Sa	48.4	22.55	J. Biol. Chem. (2003) 278, 31790-31795
T16V-RNase Sa	49.4	21.92	J. Biol. Chem. (2003) 278, 31790-31795
T59V-RNase Sa	42.8	15.94	J. Biol. Chem. (2003) 278, 31790-31795
T72V-RNase Sa	47.6	21.76	J. Biol. Chem. (2003) 278, 31790-31795
V2T-RNase Sa	45.2	18.79	J. Biol. Chem. (2003) 278, 31790-31795
V36T-RNase Sa	43.8	18.33	J. Biol. Chem. (2003) 278, 31790-31795
V43T-RNase Sa	46.8	21.80	J. Biol. Chem. (2003) 278, 31790-31795
V57T-RNase Sa	33.4	7.87	J. Biol. Chem. (2003) 278, 31790-31795
WT-RNase Sa	49.0	26.40	Protein Sci. (2003) 12, 2367-2373
Q38A-RNase Sa	52.5	31.38	Protein Sci. (2003) 12, 2367-2373
E41K-RNase Sa	46.5	23.56	Protein Sci. (2003) 12, 2367-2373
E54Q-RNase Sa	43.1	17.57	Protein Sci. (2003) 12, 2367-2373
R65A-RNase Sa	45.6	20.13	Protein Sci. (2003) 12, 2367-2373
E74K-RNase Sa	52.1	32.13	Protein Sci. (2003) 12, 2367-2373
H85Q-RNase Sa	49.1	22.05	Protein Sci. (2003) 12, 2367-2373
WT-RNase Sa	47.8	21.88	J. Mol. Biol. (2005) 354, 967-978
D79F-RNase Sa	57.8	26.28	J. Mol. Biol. (2005) 354, 967-978
D79Y-RNase Sa	57.4	24.06	J. Mol. Biol. (2005) 354, 967-978
D79A-RNase Sa	57.0	27.99	J. Mol. Biol. (2005) 354, 967-978
D79I-RNase Sa	57.4	26.11	J. Mol. Biol. (2005) 354, 967-978

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
D79R-RNase Sa	56.8	27.11	J. Mol. Biol. (2005) 354, 967-978
D79L-RNase Sa	56.6	20.17	J. Mol. Biol. (2005) 354, 967-978
D79K-RNase Sa	55.4	24.89	J. Mol. Biol. (2005) 354, 967-978
D79W-RNase Sa	55.4	24.89	J. Mol. Biol. (2005) 354, 967-978
D79H-RNase Sa	53.4	26.94	J. Mol. Biol. (2005) 354, 967-978
D79N-RNase Sa	53.4	26.94	J. Mol. Biol. (2005) 354, 967-978
D79E-RNase Sa	47.0	21.05	J. Mol. Biol. (2005) 354, 967-978
Q94K-RNase Sa	48.6	23.93	J. Mol. Biol. (2005) 354, 967-978
D33A-RNase Sa	31.8	4.94	J. Mol. Biol. (2005) 354, 967-978
WT-RNase Sa	49.3	23.47	Protein Sci. (2010) 19, 1044-1052
D25K/E74K-RNase Sa	51.5	24.43	Protein Sci. (2010) 19, 1044-1052
D25K/S31P/S42G/S48P/E74K/T76P/Q77G-RNase Sa	66.6	35.19	Protein Sci. (2010) 19, 1044-1052
D25K/S31P/S42G/S48P/E74K/T76P/Q77G/D79F-RNase Sa	77.2	42.17	Protein Sci. (2010) 19, 1044-1052
D25K/S31P/S42G/S48P/E74K/T76P/Q77G/I92D-RNase Sa	34.0	8.28	Protein Sci. (2010) 19, 1044-1052
D25K/S31P/S42G/S48P/E74K/T76P/Q77G/D79A-RNase Sa	75.3	43.30	Protein Sci. (2010) 19, 1044-1052
D25K/S31P/S42G/S48P/E74K/T76P/Q77G/D79F/I92D-RNase Sa	50.2	21.55	Protein Sci. (2010) 19, 1044-1052
D25K/S31P/S42G/S48P/I70D/E74K/T76P/Q77G/D79F-RNase Sa	41.5	15.23	Protein Sci. (2010) 19, 1044-1052
D25K/S31P/S42G/S48P/E74K/T76P/Q77G/D79F/I92A-RNase Sa	63.9	36.36	Protein Sci. (2010) 19, 1044-1052
D25K/S31P/S42G/S48P/E74K/T76P/Q77G/D79F/Y80A-RNase Sa	58.0	27.15	Protein Sci. (2010) 19, 1044-1052

Table 5.14. Mutants of RNase Sa3.[174]

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
WT-RNase Sa3	46.9	22.18	J. Mol. Biol. (2001) 312, 393-404
Y11F-RNase Sa3	45.0	19.66	J. Mol. Biol. (2001) 312, 393-404
Y33F-RNase Sa3	48.6	24.27	J. Mol. Biol. (2001) 312, 393-404
Y54F-RNase Sa3	38.0	11.30	J. Mol. Biol. (2001) 312, 393-404
Y55F-RNase Sa3	39.8	13.39	J. Mol. Biol. (2001) 312, 393-404
Y58F-RNase Sa3	44.6	19.25	J. Mol. Biol. (2001) 312, 393-404
Y83F-RNase Sa3	41.9	15.90	J. Mol. Biol. (2001) 312, 393-404
Y84F-RNase Sa3	43.6	17.99	J. Mol. Biol. (2001) 312, 393-404
Y89F-RNase Sa3	47.0	22.18	J. Mol. Biol. (2001) 312, 393-404

Table 5.15. Mutants of lambda repressor.[179-184]

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
WT-λRep	55.7	20.08	Biochemistry (1992) 31, 4324-4333
M40V/V47L-λRep	51.3	15.48	Biochemistry (1992) 31, 4324-4333
V36L/M40V/V47I-λRep	53.6	15.48	Biochemistry (1992) 31, 4324-4333
V36I/M40V/V47L-λRep	53.4	16.32	Biochemistry (1992) 31, 4324-4333
V36I/M40V/V47I-λRep	53.7	16.74	Biochemistry (1992) 31, 4324-4333
V36I-λRep	59.1	24.27	Biochemistry (1992) 31, 4324-4333
V36L/M40L/V47I-λRep	59.6	22.18	Biochemistry (1992) 31, 4324-4333
V36F/M40L-λRep	51.6	14.23	Biochemistry (1992) 31, 4324-4333
M40A-λRep	47.1	11.72	Biochemistry (1992) 31, 4324-4333
V36F/M40F/V47I-λRep	47.2	10.88	Biochemistry (1992) 31, 4324-4333
V36F/M40F/V47F-λRep	45.4	7.53	Biochemistry (1992) 31, 4324-4333
L18A/M40A-λRep	23.2	-0.84	Biochemistry (1992) 31, 4324-4333
V36F/M40F/V47I/L65F-λRep	49.1	12.55	Biochemistry (1992) 31, 4324-4333
WT-λRep	53.4	23.96	Proteins (1986) 1, 43-46
G46A-λRep	56.5	28.20	Proteins (1986) 1, 43-46
G48A-λRep	58.1	25.87	Proteins (1986) 1, 43-46
G46A/G48A-λRep	59.6	25.74	Proteins (1986) 1, 43-46
WT-λRep	55	30.75	Proc. Natl. Acad. Sci. USA (1984) 81, 5685-5689
K4Q-λRep	56	19.80	Proc. Natl. Acad. Sci. USA (1984) 81, 5685-5689
Q33S-λRep	55	30.75	Proc. Natl. Acad. Sci. USA (1984) 81, 5685-5689

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
Q33Y-λRep	61	36.21	Proc. Natl. Acad. Sci. USA (1984) 81, 5685-5689
Q44L-λRep	58	29.32	Proc. Natl. Acad. Sci. USA (1984) 81, 5685-5689
Q44Y-λRep	56	23.76	Proc. Natl. Acad. Sci. USA (1984) 81, 5685-5689
S45L-λRep	55	23.07	Proc. Natl. Acad. Sci. USA (1984) 81, 5685-5689
A49V-λRep	49	14.11	Proc. Natl. Acad. Sci. USA (1984) 81, 5685-5689
Y22H-λRep	34	13.09	Proc. Natl. Acad. Sci. USA (1984) 81, 5685-5689
A66T-λRep	42	4.55	Proc. Natl. Acad. Sci. USA (1984) 81, 5685-5689
I84S-λRep	46	15.25	Proc. Natl. Acad. Sci. USA (1984) 81, 5685-5689
WT-λRep	55.0	15.06	Biochemistry (1990) 29, 7563-7571
P78A-λRep	48.0	21.76	Biochemistry (1990) 29, 7563-7571
G46A/G48A/P78A-λRep	55.0	22.59	Biochemistry (1990) 29, 7563-7571
G46A/G48A-λRep	62.0	28.87	Biochemistry (1990) 29, 7563-7571
WT-λRep	53.9	19.85	Biochemistry (1988) 27, 7571-7574
Y88C-λRep	62.7	31.31	Biochemistry (1988) 27, 7571-7574
G46A/G48A-λRep	62.0	41.00	Biochemistry (1988) 27, 7571-7574
G46A/G48A/Y88C-λRep	70.3	40.36	Biochemistry (1988) 27, 7571-7574
WT-λRep	53.9	29.66	Biochemistry (1988) 27, 7571-7574
Y88C-λRep	62.7	44.98	Biochemistry (1988) 27, 7571-7574
G46A/G48A-λRep	62.0	28.94	Biochemistry (1988) 27, 7571-7574
G46A/G48A/Y88C-λRep	70.3	64.56	Biochemistry (1988) 27, 7571-7574
D14A/Y22W/Q33Y/G46A/G48A-λ(6-85)	73.5	38.11	Biochemistry (2004) 43, 13018-13025

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
Y22W/Q33Y/G46A/G48A-λ(6-85)	71.0	50.62	Biochemistry (2004) 43, 13018-13025
Y22W/Q33Y/A37G-λ(6-85)	59.5	21.01	Biochemistry (2004) 43, 13018-13025
Y22W/Q33Y/G46A/S45A/G48A-λ(6-85)	69.5	52.94	Biochemistry (2004) 43, 13018-13025
Y22W/Q33Y/G46A/S45A/G48A/S79A-λ(6-85)	70.5	48.95	Biochemistry (2004) 43, 13018-13025
Y22W-λ(6-85)	61.0	30.73	Biochemistry (2004) 43, 13018-13025
Y22W/G46A/G48A-λ(6-85)	67.5	43.04	Biochemistry (2004) 43, 13018-13025
Y22W/A37G-λ(6-85)	54.5	21.01	Biochemistry (2004) 43, 13018-13025
Y22W/A63V-λ(6-85)	60.0	29.96	Biochemistry (2004) 43, 13018-13025
Y22W/A37G/A49G-λ(6-85)	47.0	12.89	Biochemistry (2004) 43, 13018-13025
Y22W/Q33Y/A37G/A49G/A81G-λ(6-85)	47.5	14.06	Biochemistry (2004) 43, 13018-13025
Y22W/Q33Y/A37G/A49G-λ(6-85)	54.5	20.88	Biochemistry (2004) 43, 13018-13025
Y22W/Q33Y/M42G/S45A/G46A/G48A/S79A-λ(6-85)	59.0	30.44	Biochemistry (2004) 43, 13018-13025

Table 5.16. Mutants of RNase T1.[172, 173, 185-190]

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
WT-RNase T1	49.3	23.01	J. Biol. Chem. (1989) 264, 11621-11625
Q25K-RNase T1	51.7	28.87	J. Biol. Chem. (1989) 264, 11621-11625
E58A-RNase T1	46.0	18.83	J. Biol. Chem. (1989) 264, 11621-11625
Q25K/E58A-RNase T1	48.8	22.59	J. Biol. Chem. (1989) 264, 11621-11625
WT-RNase T1	50.9	30.12	Biochemistry (1992) 31, 725-732
Y11F-RNase T1	44.9	22.59	Biochemistry (1992) 31, 725-732
Y42F-RNase T1	54.3	31.38	Biochemistry (1992) 31, 725-732
Y56F-RNase T1	48.8	25.10	Biochemistry (1992) 31, 725-732
Y57F-RNase T1	49.6	28.03	Biochemistry (1992) 31, 725-732
Y68F-RNase T1	46.9	20.92	Biochemistry (1992) 31, 725-732
S12A-RNase T1	47.7	24.27	Biochemistry (1992) 31, 725-732
S17A-RNase T1	52.6	31.38	Biochemistry (1992) 31, 725-732
S64A-RNase T1	46.3	24.69	Biochemistry (1992) 31, 725-732
N9A-RNase T1	48.8	25.52	Biochemistry (1992) 31, 725-732
N36A-RNase T1	50.9	30.54	Biochemistry (1992) 31, 725-732
N44A-RNase T1	45.4	20.50	Biochemistry (1992) 31, 725-732
N81A-RNase T1	42.3	17.99	Biochemistry (1992) 31, 725-732
WT-RNase T1	57.6	30.54	Biochemistry (1994) 33, 10725-10730
Y45W/W59Y-RNase T1	53.1	24.69	Biochemistry (1994) 33, 10725-10730
W59Y-RNase T1	53.4	24.27	Biochemistry (1994) 33, 10725-10730
Y45W-RNase T1	56.0	25.94	Biochemistry (1994) 33, 10725-10730
WT-RNase T1	50.8	25.52	Biochemistry (1998) 37, 16192-16200

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
N44D-RNase T1	45.3	18.41	Biochemistry (1998) 37, 16192-16200
N44S-RNase T1	45.8	20.08	Biochemistry (1998) 37, 16192-16200
N44A-Rnase T1	45.6	19.25	Biochemistry (1998) 37, 16192-16200
WT-RNase T1	50.8	25.52	Biochemistry (1999) 38, 13379-13384
D76N-RNase T1	37.0	10.88	Biochemistry (1999) 38, 13379-13384
D76S-RNase T1	37.2	11.30	Biochemistry (1999) 38, 13379-13384
D76A-RNase T1	35.6	9.20	Biochemistry (1999) 38, 13379-13384
WT-RNase T1	52.3	29.29	Protein Sci. (1999) 8, 1843-1849
D49A-RNase T1	54.0	32.22	Protein Sci. (1999) 8, 1843-1849
D49Y-RNase T1	52.4	30.54	Protein Sci. (1999) 8, 1843-1849
D49F-RNase T1	52.7	31.80	Protein Sci. (1999) 8, 1843-1849
D49W-RNase T1	51.2	28.45	Protein Sci. (1999) 8, 1843-1849
WT-RNase T1	55.5	33.05	Protein Sci. (1999) 8, 1843-1849
D49H-RNase T1	58.9	35.56	Protein Sci. (1999) 8, 1843-1849
WT-RNase T1	51.6	33.05	Eur. J. Biochem. (1996) 241, 516-524
P73V-RNase T1	45.8	23.43	Eur. J. Biochem. (1996) 241, 516-524
WT-RNase T1	57.2	32.70	Eur. J. Biochem. (1994) 220, 527-534
W59Y-RNase T1	54.3	28.80	Eur. J. Biochem. (1994) 220, 527-534
Y24W-RNase T1	58.8	37.80	Eur. J. Biochem. (1994) 220, 527-534
Y24W/W59Y-RNase T1	56.6	33.90	Eur. J. Biochem. (1994) 220, 527-534
Y42W-RNase T1	56.6	32.10	Eur. J. Biochem. (1994) 220, 527-534
Y42W/W59Y-RNase T1	53.7	28.30	Eur. J. Biochem. (1994) 220, 527-534
Y45W-RNase T1	55.9	35.80	Eur. J. Biochem. (1994) 220, 527-534

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
Y45W/W59Y-RNase T1	52.8	28.90	Eur. J. Biochem. (1994) 220, 527-534
H40T-RNase T1	56.7	33.80	Eur. J. Biochem. (1994) 220, 527-534
H40T/W59Y-RNase T1	53.9	28.80	Eur. J. Biochem. (1994) 220, 527-534
H92A-RNase T1	55.9	30.10	Eur. J. Biochem. (1994) 220, 527-534
W59Y/H92A-RNase T1	53.2	26.20	Eur. J. Biochem. (1994) 220, 527-534

Table 5.17. Mutants of barnase.[191-193]

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
WT-brn	53.9	45.46	Biochemistry (1989) 28, 4914-4922
I96V-brn	51.5	39.95	Biochemistry (1989) 28, 4914-4922
I88V-brn	51.0	39.42	Biochemistry (1989) 28, 4914-4922
I96A-brn	44.9	30.97	Biochemistry (1989) 28, 4914-4922
I88A-brn	42.7	25.82	Biochemistry (1989) 28, 4914-4922
L14A-brn	42.0	24.21	Biochemistry (1989) 28, 4914-4922
WT-brn	54.1	48.70	J. Mol. Biol. (1999) 286, 1471-1485
G53A-brn	45.5	34.48	J. Mol. Biol. (1999) 286, 1471-1485
G52A-brn	40.8	26.53	J. Mol. Biol. (1999) 286, 1471-1485
G53V-brn	34.5	16.07	J. Mol. Biol. (1999) 286, 1471-1485
G52V-brn	33.0	13.56	J. Mol. Biol. (1999) 286, 1471-1485
G53Δ-brn	41.0	26.94	J. Mol. Biol. (1999) 286, 1471-1485
WT-brn	54.5	43.93	Biochemistry (2004) 43, 3346-3356
H102A-brn	52.3	39.75	Biochemistry (2004) 43, 3346-3356
D8G-brn	51.5	39.75	Biochemistry (2004) 43, 3346-3356
D8A-brn	52.5	41.42	Biochemistry (2004) 43, 3346-3356
D12G/H102A-brn	50.0	34.73	Biochemistry (2004) 43, 3346-3356
D12A/H102A-brn	51.3	44.35	Biochemistry (2004) 43, 3346-3356
Y13G/H102A-brn	36.7	12.97	Biochemistry (2004) 43, 3346-3356
Q15G/H102A-brn	50.3	33.05	Biochemistry (2004) 43, 3346-3356
Q15A/H102A-brn	51.4	38.91	Biochemistry (2004) 43, 3346-3356
Y17G/H102A-brn	42.0	23.43	Biochemistry (2004) 43, 3346-3356

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
Y17A-brn	49.4	35.98	Biochemistry (2004) 43, 3346-3356
I55G/H102A-brn	47.0	26.78	Biochemistry (2004) 43, 3346-3356
R72G/H102A-brn	46.6	29.29	Biochemistry (2004) 43, 3346-3356
E73G/H102A-brn	39.6	17.15	Biochemistry (2004) 43, 3346-3356
I88G/H102A-brn	31.5	8.79	Biochemistry (2004) 43, 3346-3356
L89G/H102A-brn	43.6	10.46	Biochemistry (2004) 43, 3346-3356
L95G/H102A-brn	43.2	20.08	Biochemistry (2004) 43, 3346-3356
I96G/H102A-brn	36.6	15.90	Biochemistry (2004) 43, 3346-3356
Y97G/H102A-brn	34.4	12.13	Biochemistry (2004) 43, 3346-3356
T100G/H102A-brn	48.0	28.03	Biochemistry (2004) 43, 3346-3356

Table 5.18. Mutants of RNase A.[194-200]

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
WT-RNase A	63.3	45.44	Protein Sci. (1996) 5, 1697-1703
Y97F-RNase A	53.2	29.46	Protein Sci. (1996) 5, 1697-1703
Y97A-RNase A	29.0	3.51	Protein Sci. (1996) 5, 1697-1703
Y97G-RNase A	30.0	3.97	Protein Sci. (1996) 5, 1697-1703
WT-RNase A	61.3	39.33	Protein Sci. (1997) 6, 1682-1693
N67D-RNase A	61.5	38.91	Protein Sci. (1997) 6, 1682-1693
N67isoD-RNase A	55.0	29.04	Protein Sci. (1997) 6, 1682-1693
WT-RNase A	58.0	34.94	J. Biol. Chem. (2002) 277, 17538-17543
F46V-RNase A	46.5	20.29	J. Biol. Chem. (2002) 277, 17538-17543
F46E-RNase A	26.4	1.30	J. Biol. Chem. (2002) 277, 17538-17543
F46K-RNase A	23.7	-0.79	J. Biol. Chem. (2002) 277, 17538-17543
WT-RNase A	59.7	39.79	Biochemistry (2003) 42, 10651-10658
F46L-RNase A	48.6	25.94	Biochemistry (2003) 42, 10651-10658
F46V-RNase A	47.6	20.21	Biochemistry (2003) 42, 10651-10658
F46A-RNase A	37.3	12.43	Biochemistry (2003) 42, 10651-10658
WT-RNase A	61.1	37.74	Protein Sci. (2007) 16, 1609-1616
S75A-RNase A	53.2	30.59	Protein Sci. (2007) 16, 1609-1616
S75T-RNase A	52.2	30.59	Protein Sci. (2007) 16, 1609-1616
S75C-RNase A	43.2	30.67	Protein Sci. (2007) 16, 1609-1616
S75R-RNase A	38.2	13.47	Protein Sci. (2007) 16, 1609-1616
WT-RNase A	65.6	45.31	Biophys. Chem. (2009) 141, 21-28
A4C/V118C-RNase A	70.5	50.08	Biophys. Chem. (2009) 141, 21-28

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
R10C/R33C-RNase A	56.9	32.09	Biophys. Chem. (2009) 141, 21-28
M30C/N44C-RNase A	47.4	18.03	Biophys. Chem. (2009) 141, 21-28
V43C/R85C-RNase A	67.8	47.74	Biophys. Chem. (2009) 141, 21-28
H105C/V124C-RNase A	64.9	51.84	Biophys. Chem. (2009) 141, 21-28
I107C/A122C-RNase A	57.8	32.97	Biophys. Chem. (2009) 141, 21-28
WT-RNase A	62.8	20.46	Biochemistry (2006) 45, 10795-10806
A4S	77.0	28.28	Biochemistry (2006) 45, 10795-10806
A5S	62.0	18.45	Biochemistry (2006) 45, 10795-10806
S123A	60.0	68.87	Biochemistry (2006) 45, 10795-10806

Table 5.19. Mutants of staphylococcal nuclease.[201-210]

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
wild type (w/o phosphate buffer)	50.6	22.98	Biochemistry (1991) 30, 1193-1199
wild type (w phosphate buffer)	52.1	22.99	Biochemistry (1991) 30, 1193-1199
P I17G	55.6	30.56	Biochemistry (1991) 30, 1193-1199
P I17T	51.2	20.36	Biochemistry (1991) 30, 1193-1199
H124L	56.4	31.18	Biochemistry (1991) 30, 1193-1199
K116G	54.4	27.80	Biochemistry (1991) 30, 1193-1199
nuclease-conA	32.8	4.84	Biochemistry (1991) 30, 1193-1199
nuclease-con A-S28G	30.5	5.02	Biochemistry (1991) 30, 1193-1199
nuclease-conA-S28G	26.0	0.60	Biochemistry (1991) 30, 1193-1199
wild type (w Pi).	53.4	27.79	Biochemistry (1991) 30, 1193-1199
V66L	55.7	22.98	Biochemistry (1991) 30, 1193-1199
V66L + G88V	57.6	17.18	Biochemistry (1991) 30, 1193-1199
G88V	55.9	21.74	Biochemistry (1991) 30, 1193-1199
V66L + G79S + G88V	53.4	11.49	Biochemistry (1991) 30, 1193-1199
A69T	41.2	12.16	Biochemistry (1991) 30, 1193-1199
118M + A90S	41.6	14.41	Biochemistry (1991) 30, 1193-1199
WT	54.1	15.33	Protein Sci. (1994) 3, 952-959
V66L	59.8	15.74	Protein Sci. (1994) 3, 952-959
V66A	42.2	8.31	Protein Sci. (1994) 3, 952-959
V66W	46.3	10.25	Protein Sci. (1994) 3, 952-959
WT	54.1	15.33	Protein Sci. (1994) 3, 952-959

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
WT	53.4	25.52	Biochemistry (1988) 27, 4761-4768
V66L	55.7	23.43	Biochemistry (1988) 27, 4761-4768
G88V	55.9	22.18	Biochemistry (1988) 27, 4761-4768
V66L/G88V	57.6	16.32	Biochemistry (1988) 27, 4761-4768
V66L/G79S/G88V	53.4	13.39	Biochemistry (1988) 27, 4761-4768
A69T	41.2	13.81	Biochemistry (1988) 27, 4761-4768
I18M/A90S	41.6	11.30	Biochemistry (1988) 27, 4761-4768
L7A	50.9	20.99	Biochemistry (1995) 34, 2034-2041
V23A	41.6	12.97	Biochemistry (1995) 34, 2034-2041
K24G	47.9	19.92	Biochemistry (1995) 34, 2034-2041
L137A	45.3	13.78	Biochemistry (1995) 34, 2034-2041
K70W	59.4	20.38	J. Mol. Biol. (1993) 232, 718-724
G88W	49.9	18.12	J. Mol. Biol. (1993) 232, 718-724
H124L	55.9	18.37	Biochemistry (1996) 35, 10328-10338
H124L/Q80C/K116C (oxidized)	57.9	12.55	Biochemistry (1996) 35, 10328-10338
H124L/Q80C/K116C (reduced)	51.1	7.66	Biochemistry (1996) 35, 10328-10338
G79S/H124L/Q80C/K116C (oxidized)	53.8	11.92	Biochemistry (1996) 35, 10328-10338
G79S/H124L/Q80C/K116C (reduced)	43.0	1.34	Biochemistry (1996) 35, 10328-10338
H124L/G79C/N118C (oxidized)	64.3	24.48	Biochemistry (1996) 35, 10328-10338
H124L/G79C/N118C (reduced)	48.9	8.03	Biochemistry (1996) 35, 10328-10338
H124L/D77C/N118C (oxidized)	53.9	8.28	Biochemistry (1996) 35, 10328-10338

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
H124L/D77C/N118C (reduced)	29.9	-5.19	Biochemistry (1996) 35, 10328-10338
G29C	48.4	18.41	Protein Sci. (1995) 4, 2545-2558
G50C	50.4	19.66	Protein Sci. (1995) 4, 2545-2558
E57C	52.1	20.08	Protein Sci. (1995) 4, 2545-2558
A60C	49.4	17.99	Protein Sci. (1995) 4, 2545-2558
K70C	50.2	20.92	Protein Sci. (1995) 4, 2545-2558
K78C	53.4	21.34	Protein Sci. (1995) 4, 2545-2558
R105C	41.9	11.72	Protein Sci. (1995) 4, 2545-2558
A112C	50.2	19.66	Protein Sci. (1995) 4, 2545-2558
K134C	50.1	20.08	Protein Sci. (1995) 4, 2545-2558
D21K	61.2	26.78	J. Mol. Biol. (2000) 303, 125-130
P117G/H124L/S128A	65.9	36.82	J. Mol. Biol. (2000) 303, 125-130
D21N	60.5	28.45	J. Mol. Biol. (2000) 303, 125-130
P117G/H124L/S128A/T41I	69.7	39.75	J. Mol. Biol. (2000) 303, 125-130
T33V	55.3	24.27	J. Mol. Biol. (2000) 303, 125-130
P117G/H124L/S128A/T41I/T33V	70.9	38.49	J. Mol. Biol. (2000) 303, 125-130
T41I	57.2	25.52	J. Mol. Biol. (2000) 303, 125-130
P117G/H124L/S128A/T41I/S59A	71.8	42.68	J. Mol. Biol. (2000) 303, 125-130
S59A	56.2	24.69	J. Mol. Biol. (2000) 303, 125-130
P117G/H124L/S128A/T41I/T33V	72.9	41.42	J. Mol. Biol. (2000) 303, 125-130
P117G	55.6	26.78	J. Mol. Biol. (2000) 303, 125-130
P117G/H124L/S128A/T41I/D21K	76.2	45.19	J. Mol. Biol. (2000) 303, 125-130

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)	Reference
H124L	56.4	28.87	J. Mol. Biol. (2000) 303, 125-130
P117G/H124L/S128A/T41I/D21N	75.0	46.02	J. Mol. Biol. (2000) 303, 125-130
S128A	57.3	25.52	J. Mol. Biol. (2000) 303, 125-130
E73G/D77G	30.0	3.23	J. Biol. Chem. (2001) 276, 46039-46045
E73G/E75G	27.5	1.23	J. Biol. Chem. (2001) 276, 46039-46045
E75G/D77G	32.0	3.82	J. Biol. Chem. (2001) 276, 46039-46045
E73G/E75G/D77G	25.1	0.04	J. Biol. Chem. (2001) 276, 46039-46045
W140H	59.0	23.85	J. Mol. Biol. (2004) 338, 383-400
W140F	56.9	23.01	J. Mol. Biol. (2004) 338, 383-400
W140Y	56.4	20.92	J. Mol. Biol. (2004) 338, 383-400
W140L	43.6	6.69	J. Mol. Biol. (2004) 338, 383-400
F76W/W140H	56.6	21.34	J. Mol. Biol. (2004) 338, 383-400

Table 5.20. Mutants of T4 lysozyme.[211]

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)
WT	67.2	22.18
WT*	65.8	22.18
WT* (SeMet)	78.2	33.89
Circular permutant "PERM1"	75.5	37.11
Extension mutant "PERMEXT"	74.1	34.18
I3A	73.6	28.45
I3C (S-H)	73.3	32.93
I3C (S-S)	72.7	30.54
I3D	72.6	31.88
I3E	72.5	30.54
I3F	72.4	29.50
I3G	72.3	30.54
I3L	71.9	29.71
I3L/S38D/A41V/A82P/V131A/N144D	71.5	29.29
I3L/S38D/A41V/A82P/N116D/V131A/N144D	71.4	29.29
I3L/S38D/A82P/V131A/N144D	71.3	28.87
I3L/S38D/A82P/N144D	71.0	26.36
I3M	71.0	28.83
I3P	70.9	28.03
I3S	70.8	27.49
I3T	70.8	28.03
I3V	70.8	28.45

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)
I3W	70.6	26.11
I3Y	70.6	27.03
F4-[A]/WT*	70.5	27.20
M6A/WT*	70.5	27.61
M6I	70.4	26.36
M6I	70.3	26.78
M6L/WT*	70.3	27.61
L7A/WT*	70.3	26.61
R8-[A]/WT*	70.2	27.20
R8Δ/WT*	70.0	26.15
I9C/L164C/WT* red.	70.0	26.78
I9C/L164C/WT* ox.	69.9	26.36
E11A	69.8	26.78
E11F	69.6	25.10
E11H/WT*	69.5	24.77
E11M	69.4	25.94
E11N/WT*	69.3	25.52
R14A/K16A/I17A/K19A/T21A/E22A/WT*	69.2	25.10
R14K/WT*	69.0	24.69
K16E	68.9	24.27
K16E/R119E	68.9	24.27
K16E/R119E/K135E/K147E	68.8	24.69

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)
K16E/K135E	68.8	24.69
K16E/K135E/K147E	68.8	24.69
K16E/R154E	68.8	25.10
I17A/WT*	68.7	23.89
I17A/K19S/T21S/E22R/Y24I/Y25P/WT* (“R1")	68.7	24.27
I17A...(“R1R2")	68.7	24.69
I17A...(“R1R3")	68.7	24.48
I17M/WT*	68.6	24.27
I17M/I27M/L33M/WT*	68.6	24.56
K19S/WT*	68.6	24.56
D20A/WT*	68.5	24.06
D20N/WT*	68.4	23.85
D20S/WT*	68.4	23.85
D20T/WT*	68.4	23.97
T21C/S38D/L99A/M102E/E108V/S117V/T142C/N114D (“L99A/M102E/St”)	68.4	24.02
T21C/T142C/WT* red.	68.3	24.02
T21C/T142C/WT* ox.	68.3	24.27
T21H/T142H/WT*	68.3	23.81
T21S/WT*	68.2	23.77
E22K/WT*	68.2	23.01
E22R/WT*	68.1	23.81
Y24A/Y25A/T26A/I27A/WT*	68.1	23.56

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)
Y24I/WT*	68.1	23.47
Y25P/WT*	68.1	23.85
T26S/WT*	68.1	23.85
I27A/WT*	68.1	23.85
I27M/WT*	68.0	23.01
I27M/L33M/WT*	68.0	23.01
I27M/L33M/WT*	68.0	23.43
I27-[GIGHLL]/WT* (“L31d”)	68.0	23.43
G28A/I29A/G30A/WT*	68.0	23.30
I29A/WT*	67.9	23.39
I29V/WT*	67.9	23.43
G30-[YTIGIG]/WT* (“L30c”)	67.9	23.43
G30F	67.9	23.18
L32A/L33A/T34A/E108V/WT*	67.9	23.26
L32T/WT*	67.8	22.84
L32T/T34K/K35V/S36D/P37G/S38N/L39S/WT* (R2)	67.8	23.01
L32T...(R2R3)	67.8	23.01
L33A/WT*	67.8	23.01
L33G/WT*	67.8	23.01
L33M/WT*	67.8	23.01
T34A/K35A/S36A/P37A	67.7	22.97
T34A/K35A/S36A/P37A/E128A/V131A/N132A ("7003A")	67.7	23.01

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)
T34K/WT*	67.7	23.01
K35V/WT*	67.7	22.89
S36D/WT*	67.6	22.84
P37G/WT*	67.5	22.68
S38D	67.5	22.18
S38D/A82P/N144D	67.5	22.59
S38D/N144D	67.5	22.59
S38N/WT*	67.5	22.59
L39A/WT*	67.4	22.55
L39I-[NAAKSELNKAI]/WT* ("L20", Crystal form I)	67.4	22.51
L39I-[NAAKSELNKAI]/WT* ("L20", Crystal form II)	67.4	22.59
L39S/WT*	67.4	22.43
N40A/WT*	67.4	22.38
N40A/S44A/E45A/D47A/K48A/WT*	67.3	22.43
N40A/S44A/E45A/D47A/K48A/D127A/E128A/V131A/N132A/WT* ("I001A")	67.3	22.34
N40-[A]/WT*	67.3	22.34
N40-[AA]/WT*	67.3	22.59
N40-[AAA]/WT*	67.3	22.30
N40-[AA]/K48-[LP]/WT*	67.2	22.22
N40-[AAAA]/WT*	67.2	22.18
N40A/S44A/E45A/D47A/K48A/WT*	67.2	22.18
N40A/K43A/S44A/E45A/L46A/D47A/K48A/WT*	67.2	22.18

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)
N40-[ES]/WT*	67.2	21.76
N40-[SLD]/WT*	67.1	22.05
N40-[SLD]/L46A/WT*	67.1	21.76
N40D/WT*	67.1	22.01
N40L-[A]/WT*	67.1	22.18
N40L/K43A/S44-[A]/WT*	67.1	22.18
A41D/WT*	67.1	21.97
A41S/WT*	67.0	21.76
A41V	67.0	21.76
A41V/V131A	67.0	22.18
A42F/WT*	66.9	21.34
A42I/WT*	66.9	21.76
A42K/WT*	66.9	21.71
A42L/WT*	66.9	21.67
A42S/WT*	66.9	21.71
A42V/WT*	66.8	21.34
K43A/WT*	66.8	21.34
S44A/WT*	66.8	21.59
S44-[A]/WT*	66.8	21.46
S44-[AA]/WT*	66.8	21.59
S44-[AAA]/WT*	66.8	21.76
S44-[AAA]/L46A/WT*	66.7	21.30

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)
S44-[AAAA]/WT*	66.7	21.76
S44A-[AA]/WT*	66.7	21.46
S44C/WT*	66.7	21.34
S44D/WT*	66.7	21.46
S44E/WT*	66.7	21.42
S44E/WT*	66.6	21.34
S44F/WT*	66.6	20.50
S44G/WT*	66.6	21.76
S44H/WT*	66.6	21.25
S44I/WT*	66.5	21.30
S44K/WT*	66.5	20.92
S44L/WT*	66.5	21.17
S44M/WT*	66.5	20.96
S44N/WT*	66.4	20.08
S44P/WT*	66.4	20.50
S44Q/WT*	66.4	20.92
S44R/WT*	66.4	20.92
S44T/WT*	66.4	20.92
S44V/WT*	66.4	20.92
S44W/WT*	66.4	21.00
S44Y/WT*	66.3	20.08
S44?/WT*	66.3	20.08

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)
E45A/WT*	66.3	20.84
E45A/K48A/WT*	66.2	20.92
L46A/WT*	66.2	19.66
D47A/WT*	66.2	20.08
K48A/WT*	66.2	20.50
K48-[A]/WT*	66.2	20.50
K48-[AA]/WT*	66.1	20.50
K48-[AAA]/WT*	66.1	20.50
K48-[AAAA]/WT*	66.0	20.21
K48-[HP]/WT*	66.0	20.50
K48-[LP]/WT*	66.0	20.50
A49S/WT*	65.9	20.50
I50A/WT*	65.9	20.08
I50M/WT*	65.8	20.08
R52A/N53A/T54A/N55A/G56A/V57A/I58A ("7004A")	65.8	20.92
R52V/WT*	65.7	19.66
R52V/N53A/T54S/N55G/V57T/WT* (R3)	65.7	20.08
N53A/WT*	65.7	20.08
N53A/T54A/N55A/G56A/V57A/I58A ("6005A")	65.7	20.08
N53A/T54A/N55A/G56A/V57A/I58A/E62A ("7005A")	65.7	19.96
N53A/N55A/V57A	65.6	19.25
N53A/N55A/V57A/E128A/V131A/N132A ("6003A")	65.6	19.25

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)
C54S/C97A	65.6	19.25
N55G	65.6	19.25
G56M/WT*	65.6	19.66
V57T/WT*	65.6	19.66
I58A/WT*	65.5	20.29
I58T/WT*	65.4	19.66
T59A/WT*	65.4	19.25
T59D/WT*	65.4	19.25
T59G/WT*	65.4	19.33
T59N/WT*	65.4	19.92
T59S/WT*	65.3	18.83
T59V/WT*	65.3	19.25
K60H	65.3	19.25
K60H/L13D/WT*	65.3	19.25
K60P	65.2	18.83
E64-[A]/WT*	65.2	18.83
L66A/WT*	65.2	19.25
L66M/WT*	65.2	19.25
F67A/WT*	65.2	19.25
N68-[A]/WT*	65.1	18.83
Q69P/WT*	65.1	18.83
V71A/WT*	64.9	17.15

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)
D72P/WT*	64.9	17.99
A73-[A]/WT*	64.9	18.41
A73-[AA]/WT*	64.9	19.29
A73-[AAA]/WT*	64.9	19.20
A73-[L]/WT*	64.8	18.83
A73-[V]/WT*	64.8	20.08
A73S/WT*	64.7	17.99
A73?/WT*	64.7	18.41
A74P	64.7	19.41
V75-[A]/WT*	64.7	19.08
V75T/WT*	64.6	17.57
G77A	64.6	17.57
I78A/WT*	64.6	17.99
I78M/WT*	64.6	17.99
I78M/L84M/L91M/L99M/I100M/V103M/L118M/L121M/L133M/WT* ("9b-M")	64.6	17.99
I78M/L84M/L91M/L99M/I100M/V103M/L118M/L121M/L133M/WT*, SeMet ("9b-sM")	64.5	17.15
I78M/L84M/L91M/L99M/L118M/L121M/L133M/WT* ("7c-M")	64.5	17.99
I78V/WT*	64.5	17.99
I78V/V87M/L118I/M120Y/L133F/V149I/T152V/WT* ("Core 7")	64.5	18.74
I78V/V87M/M120Y/L133F/V149I/T152V/WT* ("I118L/Core 7")	64.5	19.25
I78V/L118I/M120Y/L133F/V149I/T152V/WT* ("M87V/Core 7")	64.4	17.57
R80K/WT*	64.4	17.57

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)
R80K/R119H/WT*	64.3	17.99
A82P	64.3	17.57
A82S/WT*	64.3	17.57
K83H/WT*	64.3	17.57
K83H/A112D/WT*	64.3	18.16
K83M/T115E	64.2	8.79
L84A/WT*	64.2	17.57
L84M/WT*	64.2	17.57
L84M/V87M/L91M/L99M/I100M/V103M/G110R/V111M/L118M/L121M/L133M/WT* ("10a-M")	64.1	17.15
L84M/V87M/L91M/L99M/I100M/V103M/G110R/V111M/L118M/L121M/L133M/WT*, SeMet ("10a-sM")	64.1	17.15
L84M/V87M/L91M/L99M/G110R/V111M/L118M/L121M/L133M/F153M/WT* ("9a-M")	64.1	17.15
L84M/V87M/L91M/L99M/G110R/V111M/L118M/L121M/L133M/F153M/WT*, SeMet ("9a-sM")	64.0	16.74
L84M/V87M/L91M/L99M/L111M/L118M/L121M/L133M/WT* ("8a-M")	63.8	16.32
L84M/L91M/L99M/WT* ("3-M")	63.8	16.32
L84M/L91M/L99M/WT*, SeMet ("3-sM")	63.8	16.74
L84M/L91M/L99M/I100M/V103M/L118M/L121M/L133M/WT* ("8b-M")	63.7	17.15
L84M/L91M/L99M/V111M/L118M/L121M/L133M/WT* ("7b-M")	63.7	17.61
L84M/L91M/L99M/V111M/L118M/L121M/L133M/WT*, SeMet ("7b-sM")	63.6	16.32
L84M/L91M/L99M/L118M/L121M/WT* ("5-M")	63.5	15.90
L84M/L91M/L99M/L118M/L121M/WT*, SeMet ("5-sM")	63.5	15.90
L84M/L91M/L99M/L118M/L121M/L133M/WT* ("6b-M")	63.5	15.90
L84M/L91M/L99M/L118M/L121M/L133M/V149M/I150M/F153M/WT* ("9a-M")	63.5	16.32

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)
L84M/L91M/L99M/L118M/L121M/L133M/F153M/WT* ("7a-M")	63.5	16.74
L84M/L91M/L99M/L118M/L121M/L133M/F153M/WT*, SeMet ("7a-sM")	63.5	16.74
L84M/L91M/L99M/L118M/L121M/F153M/WT* ("6a-M")	63.5	17.15
L84M/L91M/L99M/L133M/WT* ("4b-M")	63.4	15.48
L84M/L91M/L99M/F153M/WT* ("4a-M")	63.4	16.74
K85A	63.3	15.90
K85A/R96H	63.3	15.90
V87A/WT*	63.2	15.06
V87I/WT*	63.2	15.48
V87I/I100V/M102L/V103I/M106I/V111A/M120Y/L133F/V149I/T152V/WT* ("Core 10")	63.2	15.90
V87I/I100V/M102L/M106I/V111A/M120Y/L133F/V149I/T152V/WT* ("I103V/Core 10")	63.2	15.90
V87I/I100V/M102L/V103I/M106I/M120Y/L133F/V149I/T152V/WT* ("A111V/Core 10")	63.2	15.90
V87I/I100V/V103I/M106I/V111A/M120Y/L133F/V149I/T152V/WT* ("L102M/Core 10")	63.2	15.90
V87M/WT*	63.2	16.32
V87T/WT*	63.2	17.36
Y88-[A]/WT*	63.1	15.48
D89A	63.1	15.48
D89A/R96H	63.1	15.48
S90C/Q122C/WT* red.	63.0	17.57
S90C/Q122C/WT* ox.	63.0	17.57
S90H/WT*	62.8	15.06
S90H/Q122D/WT*	62.7	12.55

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)
L91A/WT*	62.7	15.48
L91M/WT*	62.7	15.48
D92N/WT*	62.6	14.85
A93P	62.6	15.06
A93S/WT*	62.5	14.64
A93T/WT*	62.5	15.06
V94A/WT*	62.5	15.90
R95A/WT*	62.4	16.19
R96-[A]/WT*	62.4	14.64
R96A	62.4	14.64
R96C	62.4	15.48
R96D	62.3	14.23
R96E	62.3	14.64
R96F	62.3	15.06
R96G	62.3	15.90
R96H	62.3	16.74
R96H (100 K)	62.2	9.62
R96I	62.2	14.64
R96K	62.2	16.74
R96L	62.1	12.97
R96M	62.1	12.97
R96N	62.1	16.74

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)
R96P	62.1	13.81
R96Q	62.0	14.23
R96S	62.0	14.23
R96T	62.0	15.06
R96V	61.9	13.81
R96W	61.9	14.23
R96Y	61.8	14.64
A98C/WT*	61.8	16.32
A98F/WT*	61.8	16.74
A98I/WT*	61.7	15.86
A98L/WT*	61.7	13.81
A98M/WT*	61.7	13.81
A98S/WT*	61.6	13.39
A98V	61.5	12.97
A98V/WT*	61.5	13.60
A98V/V149C/T152S	61.5	13.81
A98V/V149I/T152S	61.5	14.23
A98V/T152S	61.5	14.23
A98W/WT*	61.4	13.39
L99A/WT*	61.4	13.81
L99A/E108V/WT*	61.4	15.90
L99A/F153A/WT*	61.4	16.32

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)
L99F/WT*	61.3	12.55
L99F/M102L/WT*	61.3	12.97
L99F/M102L/V111I/WT*	61.3	13.39
L99F/M102L/V111I/F153L/WT*	61.2	12.55
L99F/M102L/F153L/WT*	61.2	12.97
L99F/V111I/WT*	61.2	15.48
L99F/F153L/WT*	61.1	12.55
L99G/WT*	61.1	14.64
L99G/E108V/WT*	60.9	12.55
L99I/WT*	60.9	12.55
L99M/WT*	60.8	12.13
L99V/WT*	60.8	12.13
I100A/WT*	60.8	12.97
I100M/WT*	60.7	15.90
I100V/WT*	60.6	11.30
N101A/WT*	60.6	11.30
M102A/WT*	60.6	12.13
M102A/M106A/WT*	60.6	12.55
M102K/WT*	60.6	12.97
M102L/WT*	60.2	11.30
M102L/V111F/WT*	60.2	11.72
V103A/WT*	60.2	12.13

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)
V103I/WT*	60.1	10.46
V103M/WT*	60.1	10.46
F104A/WT*	60.1	10.88
F104M/WT*	60.1	10.88
Q105A	60.1	10.88
Q105E	60.1	11.30
Q105G	60.1	11.72
Q105M/WT*	60.0	11.30
M106A/WT*	59.9	10.46
M106I/WT*	59.9	11.72
M106K/WT*	59.9	11.72
M106L/WT*	59.8	11.30
E108-[A]/WT*	59.7	11.72
E108V/WT*	59.7	12.55
T109D/WT*	59.7	10.88
T109N/WT*	59.7	11.30
G110R/V111M/WT*	59.6	10.04
V111A/WT*	59.6	10.46
V111F/WT*	59.5	10.04
V111F/F153L/WT*	59.4	10.46
V111I/WT*	59.4	10.46
G113A	59.4	10.88

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)
G113E/WT*	59.3	10.04
T115-[A]/WT*	59.3	10.04
T115A/WT*	59.3	10.88
T115A/N116A/S117A/R119A/M120A/Q122A/Q123A/WT*	59.3	13.39
T115A/S117A/WT*	59.2	9.62
T115A/R119A/WT*	59.2	9.62
T115E	59.2	9.62
N116A/WT*	59.2	9.62
N116D	59.2	10.46
N116D/R119M	59.1	10.46
S117A/WT*	59.0	10.04
S117A/R119A/WT*	59.0	10.88
S117A/N132I/WT*	58.9	9.20
S117A/N132M	58.8	11.21
S117F/WT*	58.8	9.20
S117I	58.8	9.20
S117I/N132I	58.8	10.88
S117I/N132M/WT*	58.7	13.35
S117V	58.7	8.79
L118A/WT*	58.7	9.62
L118I/WT*	58.6	8.79
L118M/WT*	58.6	9.20

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)
L118M/L121M/WT* ("2-M")	58.5	8.79
R119-[A]/WT*	58.4	8.37
R119A/WT*	58.4	8.79
R119A/Q123A/WT*	58.4	9.20
R119E	58.4	9.62
R119E/K135E	58.4	10.04
R119E/K135E/K147E	58.3	9.20
R119H/WT*	58.2	8.79
R119M	58.2	9.20
R119?/WT*	58.2	9.20
M120A/WT*	58.0	8.79
M120K/WT*	58.0	10.88
M120L/WT*	58.0	12.13
M120Y/WT*	57.9	8.37
L121A/WT*	57.9	8.79
L121A/A129L/WT*	57.8	12.55
L121A/A129M/WT*	57.8	34.73
L121M/WT*	57.7	7.53
Q122A/WT*	57.7	8.37
Q123A/WT*	57.7	9.20
Q123E	57.6	7.53
K124G	57.4	7.11

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)
D127A/E128A/V131A/N132A	57.4	7.53
D127A/E128A/V131A/N132A/L133A	57.4	8.37
D127-[A]/WT*	57.3	7.95
D127A/E128A	57.2	6.69
D127A/E128A/V131A/N132A/K135A/S136A ("7002A")	57.2	10.04
D127A/E128A/V131A/N132A/K135A/S136A/R137A/Y139A/N140A/Q141A ("10A01")	57.1	10.46
D127C/R154C/WT* red.	57.1	7.11
D127C/R154C/WT* ox.	57.1	7.53
D127Δ/WT*	57.1	7.95
E128A	57.1	7.95
E128A/V131A	57.1	8.37
E128A/V131A/N132A	57.1	9.20
E128A/V131A/N132A/K135A/S136A/R137A ("6004A")	57.1	9.20
E128A/V131A/N132A/K135A/S136A/R137A/Y139A/N140A/Q141A ("9001A")	57.0	6.28
E128A/V131A/N132A/L133A	56.9	7.32
A129F/WT*	56.9	7.53
A129L/WT*	56.9	7.53
A129M/WT*	56.9	9.20
A129M/F153A/WT*	56.9	9.50
A129V	56.9	11.30
A129W/WT*	56.8	8.79
A130S/WT*	56.8	9.62

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)
V131-[A]/WT*	56.7	6.28
V131A	56.7	6.69
V131A/N132A	56.7	7.95
V131D	56.7	7.11
V131E	56.6	10.46
V131G	56.5	7.95
V131I	56.4	6.69
V131L	56.2	6.69
V131M	56.2	8.79
V131S	56.2	9.20
V131T	56.0	6.69
N132F	56.0	6.69
N132I	56.0	7.11
N132M	55.9	5.86
L133A	55.8	5.86
L133D/WT*	55.7	4.60
L133F	55.5	6.28
L133F/WT*	55.5	9.20
L133G	55.3	4.18
L133M/WT*	55.2	3.77
A134S/WT*	55.2	9.62
K135E	55.1	2.93

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)
K135E/K147E	55.1	3.77
N140-[A]/WT*	55.1	4.18
N144-[A]/WT*	55.0	7.53
N144D	54.9	7.53
N144E/WT*	54.7	4.60
N144E/K147M	54.7	5.44
N144H/WT*	54.5	4.18
A146T	54.4	3.35
K147-[A]/WT*	54.1	3.77
K147E	54.0	2.51
R148-[A]/WT*	54.0	2.51
R148-[AA]/WT*	54.0	3.77
R148-[AAA]/WT*	53.9	2.51
R148-[AAAA]/WT*	53.8	6.28
R148-[D]/WT*	53.5	2.93
R148-[DS]/WT*	53.5	4.60
R148-[S]/WT*	53.3	2.09
R148-[TT]/WT*	53.3	4.18
R148-[VP]/WT*	52.7	1.26
V149A/WT*	52.5	2.09
V149C	52.4	1.67
V149C/WT*	52.0	1.67

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)
V149G/WT*	51.9	1.67
V149I/WT*	51.7	-0.84
V149M/WT*	51.6	1.26
V149S/WT*	51.5	1.26
V149T/WT*	51.4	5.44
I150-[A]/WT*	51.3	1.26
T151S/WT*	51.3	1.26
T152A/WT*	51.2	-0.84
T152C/WT*	51.1	1.67
T152I/WT*	50.9	-1.26
T152S	50.8	5.44
T152S/WT*	50.8	5.44
T152V/WT*	50.2	-0.84
F153A/WT*	50.0	0.42
F153I/WT*	49.7	-0.42
F153L/WT*	49.6	-0.84
F153M/WT*	49.5	-0.42
F153V/WT*	49.3	-1.67
R154E	49.2	-1.67
G156D	48.2	-1.26
T157A	48.1	-1.26
T157C	48.0	-7.95

Mutation	T_M (°C)	$\Delta G°$ (kJ/mol)
T157D	47.9	-1.67
T157E	47.8	-0.42
T157F	47.4	-2.93
T157G	46.9	-6.69
T157H	46.9	-4.18
T157I	46.8	-2.51
T157I/WT*	46.7	-2.51
T157I/W158L/WT*	46.0	-5.02
T157L	45.2	-4.18
T157N	44.0	-1.26
T157R	43.8	-7.11
T157S	42.9	-2.93
T157V	42.8	-7.11
W158L/WT*	42.3	-7.11
N163D/WT*	38.4	-5.44
L164-[AAAA]/WT*	25.4	-12.55

5.8. References

[1] Marcotte, E. M., Pellegrini, M., Ng, H. L., Rice, D. W., Yeates, T. O., and Eisenberg, D. (1999) Detecting protein function and protein-protein interactions from genome sequences, *Science 285*, 751-753.

[2] Martinez Molina, D., Jafari, R., Ignatushchenko, M., Seki, T., Larsson, E. A., Dan, C., Sreekumar, L., Cao, Y., and Nordlund, P. (2013) Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay, *Science 341*, 84-87.

[3] Jafari, R., Almqvist, H., Axelsson, H., Ignatushchenko, M., Lundback, T., Nordlund, P., and Martinez Molina, D. (2014) The cellular thermal shift assay for evaluating drug target interactions in cells, *Nat Protoc 9*, 2100-2122.

[4] Jensen, A. J., Martinez Molina, D., and Lundback, T. (2015) CETSA: a target engagement assay with potential to transform drug discovery, *Future Med Chem 7*, 975-978.

[5] Alshareef, A., Zhang, H. F., Huang, Y. H., Wu, C., Zhang, J. D., Wang, P., El-Sehemy, A., Fares, M., and Lai, R. (2016) The use of cellular thermal shift assay (CETSA) to study Crizotinib resistance in ALK-expressing human cancers, *Sci Rep 6*, 33710.

[6] Thirumalai, D., O'Brien, E. P., Morrison, G., and Hyeon, C. (2010) Theoretical perspectives on protein folding, *Annu Rev Biophys 39*, 159-183.

[7] Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding, *Protein Sci 4*, 2138-2148.

[8] Robertson, A. D., and Murphy, K. P. (1997) Protein structure and the energetics of protein stability, *Chem Rev 97*, 1251-1268.

[9] Dill, K. A., Ghosh, K., and Schmit, J. D. (2011) Physical limits of cells and proteomes, *Proc Natl Acad Sci U S A 108*, 17876-17882.

[10] Sawle, L., and Ghosh, K. (2011) How do thermophilic proteins and proteomes withstand high temperature?, *Biophys J 101*, 217-227.

[11] Leuenberger, P., Ganscha, S., Kahraman, A., Cappelletti, V., Boersema, P. J., von Mering, C., Claassen, M., and Picotti, P. (2017) Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability, *Science 355*, 812-+.

[12] Kumar, M. D., Bava, K. A., Gromiha, M. M., Prabakaran, P., Kitajima, K., Uedaira, H., and Sarai, A. (2006) ProTherm and ProNIT: thermodynamic databases for proteins and protein-nucleic acid interactions, *Nucleic Acids Res 34*, D204-206.

[13] Ghosh, K., and Dill, K. A. (2009) Computing protein stabilities from their chain lengths, *Proc Natl Acad Sci U S A 106*, 10649-10654.

[14] Privalov, P. L. (1979) Stability of proteins: small globular proteins, *Adv Protein Chem 33*, 167-241.

[15] Baldwin, R. L. (1986) Temperature dependence of the hydrophobic interaction in protein folding, *Proc Natl Acad Sci U S A 83*, 8069-8072.

[16] Murphy, K. P., Privalov, P. L., and Gill, S. J. (1990) Common features of protein unfolding and dissolution of hydrophobic compounds, *Science 247*, 559-561.

[17] Murphy, K. P., and Gill, S. J. (1991) Solid model compounds and the thermodynamics of protein unfolding, *J Mol Biol 222*, 699-709.

[18] Zhang, J. (2000) Protein-length distributions for the three domains of life, *Trends Genet 16*, 107-109.

[19] Rees, D. C., and Robertson, A. D. (2001) Some thermodynamic implications for the thermostability of proteins, *Protein Sci 10*, 1187-1194.

[20] Giuliani, S. E., Frank, A. M., and Collart, F. R. (2008) Functional assignment of solute-binding proteins of ABC transporters using a fluorescence-based thermal shift assay, *Biochemistry 47*, 13974-13984.

[21] Krishna, S. N., Luan, C. H., Mishra, R. K., Xu, L., Scheidt, K. A., Anderson, W. F., and Bergan, R. C. (2013) A fluorescence-based thermal shift assay identifies inhibitors of mitogen activated protein kinase kinase 4, *Plos One 8*, e81504.

[22] Ferguson, N., Johnson, C. M., Macias, M., Oschkinat, H., and Fersht, A. (2001) Ultrafast folding of WW domains without structured aromatic clusters in the denatured state, *Proc Natl Acad Sci U S A 98*, 13002-13007.

[23] Hernandez-Arana, A., Rojo-Dominguez, A., Soriano-Garcia, M., and Rodriguez-Romero, A. (1995) The thermal unfolding of hevein, a small disulfide-rich protein, *Eur J Biochem 228*, 649-652.

[24] Wang, T., Zhu, Y. J., and Gai, F. (2004) Folding of a three-helix bundle at the folding speed limit, *J Phys Chem B 108*, 3694-3697.

[25] Hinz, H. J., Cossmann, M., and Beyreuther, K. (1981) lac Repressor headpiece constitutes a reversibly unfolding domain, *FEBS Lett 129*, 246-248.

[26] Thompson, K. S., Vinson, C. R., and Freire, E. (1993) Thermodynamic characterization of the structural stability of the coiled-coil region of the bZIP transcription factor GCN4, *Biochemistry 32*, 5491-5496.

[27] Alexander, P., Fahnestock, S., Lee, T., Orban, J., and Bryan, P. (1992) Thermodynamic analysis of the folding of the streptococcal protein G IgG-binding domains B1 and B2: why small proteins tend to have high denaturation temperatures, *Biochemistry 31*, 3597-3603.

[28] Swint-Kruse, L., and Robertson, A. D. (1995) Hydrogen bonds and the pH dependence of ovomucoid third domain stability, *Biochemistry 34*, 4724-4732.

[29] Swint, L., and Robertson, A. D. (1993) Thermodynamics of unfolding for turkey ovomucoid third domain: thermal and chemical denaturation, *Protein Sci 2*, 2037-2049.

[30] Gushchina, L. V., Gabdulkhakov, A. G., Nikonov, S. V., Mateo, P. L., and Filimonov, V. V. (2009) Structural and thermodynamic studies of Bergerac-SH3 chimeras, *Biophys Chem 139*, 106-115.

[31] Viguera, A. R., Martinez, J. C., Filimonov, V. V., Mateo, P. L., and Serrano, L. (1994) Thermodynamic and kinetic analysis of the SH3 domain of spectrin shows a two-state folding transition, *Biochemistry 33*, 2142-2150.

[32] Makhatadze, G. I., Kim, K. S., Woodward, C., and Privalov, P. L. (1993) Thermodynamics of BPTI folding, *Protein Sci 2*, 2028-2036.

[33] Lim, W. A., Fox, R. O., and Richards, F. M. (1994) Stability and peptide binding affinity of an SH3 domain from the Caenorhabditis elegans signaling protein Sem-5, *Protein Sci 3*, 1261-1266.

[34] Wang, L., Cowley, A. B., and Benson, D. R. (2007) Enhancing the thermal stability of mitochondrial cytochrome b5 by introducing a structural motif characteristic of the less stable microsomal isoform, *Protein Eng Des Sel 20*, 511-520.

[35] Mayor, U., Johnson, C. M., Daggett, V., and Fersht, A. R. (2000) Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation, *Proc Natl Acad Sci U S A 97*, 13518-13522.

[36] Filimonov, V. V., Azuaga, A. I., Viguera, A. R., Serrano, L., and Mateo, P. L. (1999) A thermodynamic analysis of a family of small globular proteins: SH3 domains, *Biophys Chem 77*, 195-208.

[37] Knapp, S., Mattson, P. T., Christova, P., Berndt, K. D., Karshikoff, A., Vihinen, M., Smith, C. I., and Ladenstein, R. (1998) Thermal unfolding of small proteins with SH3 domain folding pattern, *Proteins 31*, 309-319.

[38] Jackson, S. E., Moracci, M., elMasry, N., Johnson, C. M., and Fersht, A. R. (1993) Effect of cavity-creating mutations in the hydrophobic core of chymotrypsin inhibitor 2, *Biochemistry 32*, 11259-11269.

[39] Johnson, C. R., Morin, P. E., Arrowsmith, C. H., and Freire, E. (1995) Thermodynamic analysis of the structural stability of the tetrameric oligomerization domain of p53 tumor suppressor, *Biochemistry 34*, 5309-5316.

[40] Garcia-Arribas, O., Mateo, R., Tomczak, M. M., Davies, P. L., and Mateu, M. G. (2007) Thermodynamic stability of a cold-adapted protein, type III antifreeze protein, and energetic contribution of salt bridges, *Protein Sci 16*, 227-238.

[41] Pfeil, W. (1998) Protein stability and folding.

[42] Schindler, T., and Schmid, F. X. (1996) Thermodynamic properties of an extremely rapid protein folding reaction, *Biochemistry 35*, 16833-16842.

[43] Zavodszky, M., Chen, C. W., Huang, J. K., Zolkiewski, M., Wen, L., and Krishnamoorthi, R. (2001) Disulfide bond effects on protein stability: designed variants of Cucurbita maxima trypsin inhibitor-V, *Protein Sci 10*, 149-160.

[44] Maxwell, K. L., Davidson, A. R., Murialdo, H., and Gold, M. (2000) Thermodynamic and functional characterization of protein W from bacteriophage lambda. The three C-terminal residues are critical for activity, *J Biol Chem 275*, 18879-18886.

[45] Bae, S. J., and Sturtevant, J. M. (1995) Thermodynamics of the thermal unfolding of eglin c in the presence and absence of guanidinium chloride, *Biophys Chem 55*, 247-252.

[46] Petrosian, S. A., and Makhatadze, G. I. (2000) Contribution of proton linkage to the thermodynamic stability of the major cold-shock protein of Escherichia coli CspA, *Protein Sci 9*, 387-394.

[47] Padmanabhan, S., Laurents, D. V., Fernandez, A. M., Elias-Arnanz, M., Ruiz-Sanz, J., Mateo, P. L., Rico, M., and Filimonov, V. V. (1999) Thermodynamic analysis of the structural stability of phage 434 Cro protein, *Biochemistry 38*, 15536-15547.

[48] Renner, M., Hinz, H. J., Scharf, M., and Engels, J. W. (1992) Thermodynamics of unfolding of the alpha-amylase inhibitor tendamistat. Correlations between accessible surface area and heat capacity, *J Mol Biol 223*, 769-779.

[49] Wintrode, P. L., Makhatadze, G. I., and Privalov, P. L. (1994) Thermodynamics of ubiquitin unfolding, *Proteins 18*, 246-253.

[50] Loladze, V. V., and Makhatadze, G. I. (2005) Both helical propensity and side-chain hydrophobicity at a partially exposed site in alpha-helix contribute to the thermodynamic stability of ubiquitin, *Proteins 58*, 1-6.

[51] Horvath, L. A., Sturtevant, J. M., and Prestegard, J. H. (1994) Kinetics and thermodynamics of thermal denaturation in acyl carrier protein, *Protein Sci 3*, 103-108.

[52] Novokhatny, V. V., Kudinov, S. A., and Privalov, P. L. (1984) Domains in human plasminogen, *J Mol Biol 179*, 215-232.

[53] Kumar, S., Tsai, C. J., and Nussinov, R. (2002) Maximal stabilities of reversible two-state proteins, *Biochemistry 41*, 5359-5374.

[54] Hasegawa, J., Uchiyama, S., Tanimoto, Y., Mizutani, M., Kobayashi, Y., Sambongi, Y., and Igarashi, Y. (2000) Selected mutations in a mesophilic cytochrome c confer the stability of a thermophilic counterpart, *J Biol Chem 275*, 37824-37828.

[55] Wenk, M., and Jaenicke, R. (1999) Calorimetric analysis of the Ca(2+)-binding betagamma-crystallin homolog protein S from Myxococcus xanthus: intrinsic stability and mutual stabilization of domains, *J Mol Biol 293*, 117-124.

[56] Nicholson, E. M., and Scholtz, J. M. (1996) Conformational stability of the Escherichia coli HPr protein: test of the linear extrapolation method and a thermodynamic characterization of cold denaturation, *Biochemistry 35*, 11369-11378.

[57] Van Nuland, N. A., Meijberg, W., Warner, J., Forge, V., Scheek, R. M., Robillard, G. T., and Dobson, C. M. (1998) Slow cooperative folding of a small globular protein HPr, *Biochemistry 37*, 622-637.

[58] Scholtz, J. M. (1995) Conformational stability of HPr: the histidine-containing phosphocarrier protein from Bacillus subtilis, *Protein Sci 4*, 35-43.

[59] Razvi, A., and Scholtz, J. M. (2006) A thermodynamic comparison of HPr proteins from extremophilic organisms, *Biochemistry 45*, 4084-4092.

[60] Agashe, V. R., and Udgaonkar, J. B. (1995) Thermodynamics of denaturation of barstar: evidence for cold denaturation and evaluation of the interaction with guanidine hydrochloride, *Biochemistry 34*, 3286-3299.

[61] Pfeil, W., and Bendzko, P. (1980) Thermodynamic investigations of cytochrome b5 unfolding. I. The tryptic fragment of cytochrome b5, *Biochim Biophys Acta 626*, 73-78.

[62] Clarke, J., Hamill, S. J., and Johnson, C. M. (1997) Folding and stability of a fibronectin type III domain of human tenascin, *J Mol Biol 270*, 771-778.

[63] Pfeil, W. (1993) Thermodynamics of apocytochrome b5 unfolding, *Protein Sci 2*, 1497-1501.

[64] Pace, C. N., Hebert, E. J., Shaw, K. L., Schell, D., Both, V., Krajcikova, D., Sevcik, J., Wilson, K. S., Dauter, Z., Hartley, R. W., and Grimsley, G. R. (1998) Conformational stability and thermodynamics of folding of ribonucleases Sa, Sa2 and Sa3, *J Mol Biol 279*, 271-286.

[65] Zerovnik, E., Lohner, K., Jerala, R., Laggner, P., and Turk, V. (1992) Calorimetric measurements of thermal denaturation of stefins A and B. Comparison to predicted thermodynamics of stefin-B unfolding, *Eur J Biochem 210*, 217-221.

[66] Taddei, N., Chiti, F., Paoli, P., Fiaschi, T., Bucciantini, M., Stefani, M., Dobson, C. M., and Ramponi, G. (1999) Thermodynamics and kinetics of folding of common-type acylphosphatase: comparison to the highly homologous muscle isoenzyme, *Biochemistry 38*, 2135-2142.

[67] Notomista, E., Catanzano, F., Graziano, G., Dal Piaz, F., Barone, G., D'Alessio, G., and Di Donato, A. (2000) Onconase: an unusually stable protein, *Biochemistry 39*, 8711-8718.

[68] Yu, Y., Makhatadze, G. I., Pace, C. N., and Privalov, P. L. (1994) Energetics of ribonuclease T1 structure, *Biochemistry 33*, 3312-3319.

[69] Privalov, P. L., and Khechinashvili, N. N. (1974) A thermodynamic approach to the problem of stabilization of globular protein structure: a calorimetric study, *J Mol Biol 86*, 665-684.

[70] Hagihara, Y., Tan, Y., and Goto, Y. (1994) Comparison of the conformational stability of the molten globule and native states of horse cytochrome c. Effects of acetylation, heat, urea and guanidine-hydrochloride, *J Mol Biol 237*, 336-348.

[71] Plaza del Pino, I. M., Pace, C. N., and Freire, E. (1992) Temperature and guanidine hydrochloride dependence of the structural stability of ribonuclease T1, *Biochemistry 31*, 11196-11202.

[72] Lee, C. F., Allen, M. D., Bycroft, M., and Wong, K. B. (2005) Electrostatic interactions contribute to reduced heat capacity change of unfolding in a thermophilic ribosomal protein l30e, *J Mol Biol 348*, 419-431.

[73] Prasad, A., Housley, N. A., and Pedigo, S. (2004) Thermodynamic stability of domain 2 of epithelial cadherin, *Biochemistry 43*, 8055-8066.

[74] Bowie, J. U., and Sauer, R. T. (1989) Equilibrium dissociation and unfolding of the Arc repressor dimer, *Biochemistry 28*, 7139-7143.

[75] Milla, M. E., and Sauer, R. T. (1995) Critical side-chain interactions at a subunit interface in the Arc repressor dimer, *Biochemistry 34*, 3344-3351.

[76] Bae, S. J., Chou, W. Y., Matthews, K., and Sturtevant, J. M. (1988) Tryptophan repressor of Escherichia coli shows unusual thermal stability, *Proc Natl Acad Sci U S A 85*, 6731-6732.

[77] Tamura, A., Kimura, K., Takahara, H., and Akasaka, K. (1991) Cold denaturation and heat denaturation of Streptomyces subtilisin inhibitor. 1. CD and DSC studies, *Biochemistry 30*, 11307-11313.

[78] Griko, Y. V., Makhatadze, G. I., Privalov, P. L., and Hartley, R. W. (1994) Thermodynamics of barnase unfolding, *Protein Sci 3*, 669-676.

[79] Santoro, M. M., and Bolen, D. W. (1992) A test of the linear extrapolation of unfolding free energy changes over an extended denaturant concentration range, *Biochemistry 31*, 4901-4907.

[80] Ladbury, J. E., Wynn, R., Hellinga, H. W., and Sturtevant, J. M. (1993) Stability of oxidized Escherichia coli thioredoxin and its dependence on protonation of the aspartic acid residue in the 26 position, *Biochemistry 32*, 7526-7530.

[81] Cohen, D. S., and Pielak, G. J. (1994) Stability of yeast iso-1-ferricytochrome c as a function of pH and temperature, *Protein Sci 3*, 1253-1260.

[82] Liggins, J. R., Sherman, F., Mathews, A. J., and Nall, B. T. (1994) Differential scanning calorimetric study of the thermal unfolding transitions of yeast iso-1 and iso-2 cytochromes c and three composite isozymes, *Biochemistry 33*, 9209-9219.

[83] Martinez, J. C., el Harrous, M., Filimonov, V. V., Mateo, P. L., and Fersht, A. R. (1994) A calorimetric study of the thermal stability of barnase and its interaction with 3'GMP, *Biochemistry 33*, 3919-3926.

[84] Moulick, R., and Udgaonkar, J. B. (2014) Thermodynamic characterization of the unfolding of the prion protein, *Biophys J 106*, 410-420.

[85] Wallgren, M., Aden, J., Pylypenko, O., Mikaelsson, T., Johansson, L. B., Rak, A., and Wolf-Watz, M. (2008) Extreme temperature tolerance of a hyperthermophilic protein coupled to residual structure in the unfolded state, *J Mol Biol 379*, 845-858.

[86] Liu, C., Chu, D., Wideman, R. D., Houliston, R. S., Wong, H. J., and Meiering, E. M. (2001) Thermodynamics of denaturation of hisactophilin, a beta-trefoil protein, *Biochemistry 40*, 3817-3827.

[87] Deutschman, W. A., and Dahlquist, F. W. (2001) Thermodynamic basis for the increased thermostability of CheY from the hyperthermophile Thermotoga maritima, *Biochemistry 40*, 13107-13113.

[88] Xie, D., Bhakuni, V., and Freire, E. (1991) Calorimetric determination of the energetics of the molten globule intermediate in protein folding: apo-alpha-lactalbumin, *Biochemistry 30*, 10673-10678.

[89] Griko, Y. V., Freire, E., and Privalov, P. L. (1994) Energetics of the alpha-lactalbumin states: a calorimetric and statistical thermodynamic study, *Biochemistry 33*, 1889-1899.

[90] Del Vecchio, P., Catanzano, F., de Paola, B., and Barone, G. (2000) Thermodynamic Stability of Ribonuclease B, *Journal of Thermal Analysis and Calorimetry 61*, 363-368.

[91] Catanzano, F., Giancola, C., Graziano, G., and Barone, G. (1996) Temperature-induced denaturation of ribonuclease S: a thermodynamic study, *Biochemistry 35*, 13378-13385.

[92] Privalov, P. L., Tiktopulo, E. I., and Khechinashvili, N. N. (1973) Calorimetric investigation of ribonuclease thermal denaturation, *Int J Pept Protein Res 5*, 229-237.

[93] Straume, M., and Freire, E. (1992) Two-dimensional differential scanning calorimetry: simultaneous resolution of intrinsic protein structural energetics and ligand binding interactions by global linkage analysis, *Anal Biochem 203*, 259-268.

[94] Steif, C., Hinz, H. J., and Cesareni, G. (1995) Effects of cavity-creating mutations on conformational stability and structure of the dimeric 4-alpha-helical protein ROP: thermal unfolding studies, *Proteins 23*, 83-96.

[95] Karantza, V., Baxevanis, A. D., Freire, E., and Moudrianakis, E. N. (1995) Thermodynamic studies of the core histones: ionic strength and pH dependence of H2A-H2B dimer stability, *Biochemistry 34*, 5988-5996.

[96] Shih, P., and Kirsch, J. F. (1995) Design and structural analysis of an engineered thermostable chicken lysozyme, *Protein Sci 4*, 2063-2072.

[97] Cooper, A., Eyles, S. J., Radford, S. E., and Dobson, C. M. (1992) Thermodynamic consequences of the removal of a disulphide bridge from hen lysozyme, *J Mol Biol 225*, 939-943.

[98] Schwarz, F. P. (1989) Biological Thermodynamic Data for the Calibration of Differential Scanning Calorimeters - Heat-Capacity Data on the Unfolding Transition of Lysozyme in Solution, *Thermochimica Acta 147*, 71-91.

[99] Pfeil, W., and Privalov, P. L. (1976) Thermodynamic investigations of proteins. I. Standard functions for proteins with lysozyme as an example, *Biophys Chem 4*, 23-32.

[100] Herning, T., Yutani, K., Inaka, K., Kuroki, R., Matsushima, M., and Kikuchi, M. (1992) Role of proline residues in human lysozyme stability: a scanning calorimetric study combined with X-ray structure analysis of proline mutants, *Biochemistry 31*, 7077-7085.

[101] Li, W. T., Grayling, R. A., Sandman, K., Edmondson, S., Shriver, J. W., and Reeve, J. N. (1998) Thermodynamic stability of archaeal histones, *Biochemistry 37*, 10563-10572.

[102] Tanaka, A., Flanagan, J., and Sturtevant, J. M. (1993) Thermal unfolding of staphylococcal nuclease and several mutant forms thereof studied by differential scanning calorimetry, *Protein Sci 2*, 567-576.

[103] Burova, T. V., Choiset, Y., Jankowski, C. K., and Haertle, T. (1999) Conformational stability and binding properties of porcine odorant binding protein, *Biochemistry 38*, 15043-15051.

[104] Balsera, M., Menendez, M., Saiz, J. L., de Las Rivas, J., Andreu, J. M., and Arellano, J. B. (2004) Structural stability of the PsbQ protein of higher plant photosystem II, *Biochemistry 43*, 14171-14179.

[105] Gasset, M., Mancheno, J. M., Laynez, J., Lacadena, J., Fernandez-Ballester, G., Martinez del Pozo, A., Onaderra, M., and Gavilanes, J. G. (1995) Thermal unfolding of the cytotoxin alpha-sarcin: phospholipid binding induces destabilization of the protein structure, *Biochim Biophys Acta 1252*, 126-134.

[106] Creagh, A. L., Koska, J., Johnson, P. E., Tomme, P., Joshi, M. D., McIntosh, L. P., Kilburn, D. G., and Haynes, C. A. (1998) Stability and oligosaccharide binding of the N1

cellulose-binding domain of Cellulomonas fimi endoglucanase CenC, *Biochemistry 37*, 3529-3537.

[107] Kelly, L., and Holladay, L. A. (1990) A comparative study of the unfolding thermodynamics of vertebrate metmyoglobins, *Biochemistry 29*, 5062-5069.

[108] Privalov, P. L., Griko Yu, V., Venyaminov, S., and Kutyshenko, V. P. (1986) Cold denaturation of myoglobin, *J Mol Biol 190*, 487-498.

[109] Makhatadze, G. I., Clore, G. M., Gronenborn, A. M., and Privalov, P. L. (1994) Thermodynamics of unfolding of the all beta-sheet protein interleukin-1 beta, *Biochemistry 33*, 9327-9332.

[110] Hollien, J., and Marqusee, S. (1999) A thermodynamic comparison of mesophilic and thermophilic ribonucleases H, *Biochemistry 38*, 3831-3836.

[111] Romero, C. M., Lozano, J. M., Sancho, J., and Giraldo, G. I. (2007) Thermal stability of beta-lactoglobulin in the presence of aqueous solution of alcohols and polyols, *Int J Biol Macromol 40*, 423-428.

[112] Carra, J. H., Murphy, E. C., and Privalov, P. L. (1996) Thermodynamic effects of mutations on the denaturation of T4 lysozyme, *Biophys J 71*, 1994-2001.

[113] Connelly, P., Ghosaini, L., Hu, C. Q., Kitamura, S., Tanaka, A., and Sturtevant, J. M. (1991) A differential scanning calorimetric study of the thermal unfolding of seven mutant forms of phage T4 lysozyme, *Biochemistry 30*, 1887-1891.

[114] Lah, J., Prislan, I., Krzan, B., Salobir, M., Francky, A., and Vesnaver, G. (2005) Erythropoietin unfolding: thermodynamics and its correlation with structural features, *Biochemistry 44*, 13883-13892.

[115] Genzor, C. G., Beldarrain, A., Gomez-Moreno, C., Lopez-Lacomba, J. L., Cortijo, M., and Sancho, J. (1996) Conformational stability of apoflavodoxin, *Protein Sci 5*, 1376-1388.

[116] Shiraki, K., Nishikori, S., Fujiwara, S., Hashimoto, H., Kai, Y., Takagi, M., and Imanaka, T. (2001) Comparative analyses of the conformational stability of a hyperthermophilic protein and its mesophilic counterpart, *Eur J Biochem 268*, 4144-4150.

[117] Teles, R. C., Calderon Lde, A., Medrano, F. J., Barbosa, J. A., Guimaraes, B. G., Santoro, M. M., and de Freitas, S. M. (2005) pH dependence thermal stability of a chymotrypsin inhibitor from Schizolobium parahyba seeds, *Biophys J 88*, 3509-3517.

[118] Fukada, H., Kitamura, S., and Takahashi, K. (1995) Calorimetric study of the thermal unfolding of Kunitz-type soybean trypsin inhibitor at pH 7.0, *Thermochimica Acta 266*, 365-372.

[119] Kodicek, M., Infanzon, A., and Karpenko, V. (1995) Heat denaturation of human orosomucoid in water/methanol mixtures, *Biochim Biophys Acta 1246*, 10-16.

[120] Muccio, D. D., Waterhous, D. V., Fish, F., and Brouillette, C. G. (1992) Analysis of the two-state behavior of the thermal unfolding serum retinol binding protein containing a single retinol ligand, *Biochemistry 31*, 5560-5567.

[121] Ramsay, G., and Freire, E. (1990) Linked thermal and solute perturbation analysis of cooperative domain interactions in proteins. Structural stability of diphtheria toxin, *Biochemistry 29*, 8677-8683.

[122] Moutiez, M., Burova, T. V., Haertle, T., and Quemeneur, E. (1999) On the non-respect of the thermodynamic cycle by DsbA variants, *Protein Sci 8*, 106-112.

[123] Johnson, C. M., Cooper, A., and Stockley, P. G. (1992) Differential scanning calorimetry of thermal unfolding of the methionine repressor protein (MetJ) from Escherichia coli, *Biochemistry 31*, 9717-9724.

[124] Rezaei, H., Choiset, Y., Eghiaian, F., Treguer, E., Mentre, P., Debey, P., Grosclaude, J., and Haertle, T. (2002) Amyloidogenic unfolding intermediates differentiate sheep prion protein variants, *J Mol Biol 322*, 799-814.

[125] Welfle, K., Misselwitz, R., Welfle, H., Politz, O., and Borriss, R. (1995) Influence of Ca2+ on conformation and stability of three bacterial hybrid glucanases, *Eur J Biochem 229*, 726-735.

[126] Arunachalam, U., and Kellis, J. T., Jr. (1996) Folding and stability of endoglucanase III, a single-domain cellulase from Trichoderma reesei, *Biochemistry 35*, 11379-11385.

[127] Spuergin, P., Abele, U., and Schulz, G. E. (1995) Stability, activity and structure of adenylate kinase mutants, *Eur J Biochem 231*, 405-413.

[128] Tishchenko, V. M., Tiktopulo, E. I., and Privalov, P. L. (1974) [Calorimetric studies of conformational transitions in alpha-chymotrypsin], *Biofizika 19*, 400-404.

[129] Tall, A. R., Shipley, G. G., and Small, D. M. (1976) Conformational and thermodynamic properties of apo A-1 of human plasma high density lipoproteins, *J Biol Chem 251*, 3749-3755.

[130] Privalov, P. L., Khechinashvili, N. N., and Atanasov, B. P. (1971) Thermodynamic analysis of thermal transitions in globular proteins. I. Calorimetric study of chymotrypsinogen, ribonuclease and myoglobin, *Biopolymers 10*, 1865-1890.

[131] Ginsburg, A., Szczepanowski, R. H., Ruvinov, S. B., Nosworthy, N. J., Sondej, M., Umland, T. C., and Peterkofsky, A. (2000) Conformational stability changes of the amino terminal domain of enzyme I of the Escherichia coli phosphoenolpyruvate: sugar phosphotransferase system produced by substituting alanine or glutamate for the active-site histidine 189: implications for phosphorylation effects, *Protein Sci 9*, 1085-1094.

[132] Matulis, D., Kranz, J. K., Salemme, F. R., and Todd, M. J. (2005) Thermodynamic stability of carbonic anhydrase: measurements of binding affinity and stoichiometry using ThermoFluor, *Biochemistry 44*, 5258-5266.

[133] Santos, J., Risso, V. A., Sica, M. P., and Ermacora, M. R. (2007) Effects of serine-to-cysteine mutations on beta-lactamase folding, *Biophys J 93*, 1707-1718.

[134] Yutani, K., Hayashi, S., Sugisaki, Y., and Ogasahara, K. (1991) Role of conserved proline residues in stabilizing tryptophan synthase alpha subunit: analysis by mutants with alanine or glycine, *Proteins 9*, 90-98.

[135] Beadle, B. M., Baase, W. A., Wilson, D. B., Gilkes, N. R., and Shoichet, B. K. (1999) Comparing the thermodynamic stabilities of a related thermophilic and mesophilic enzyme, *Biochemistry 38*, 2570-2576.

[136] Pantoliano, M. W., Whitlow, M., Wood, J. F., Dodd, S. W., Hardman, K. D., Rollence, M. L., and Bryan, P. N. (1989) Large Increases in General Stability for Subtilisin Bpn' through Incremental Changes in the Free-Energy of Unfolding, *Biochemistry 28*, 7205-7213.

[137] Fukada, H., Sturtevant, J. M., and Quiocho, F. A. (1983) Thermodynamics of the binding of L-arabinose and of D-galactose to the L-arabinose-binding protein of Escherichia coli, *J Biol Chem 258*, 13193-13198.

[138] Kamal, J. K., Nazeerunnisa, M., and Behere, D. V. (2002) Thermal unfolding of soybean peroxidase. Appropriate high denaturant concentrations induce cooperativity allowing the correct measurement of thermodynamic parameters, *J Biol Chem 277*, 40717-40721.

[139] Ganesh, C., Shah, A. N., Swaminathan, C. P., Surolia, A., and Varadarajan, R. (1997) Thermodynamic characterization of the reversible, two-state unfolding of maltose binding protein, a large two-domain protein, *Biochemistry 36*, 5020-5028.

[140] Moro, F., and Muga, A. (2006) Thermal adaptation of the yeast mitochondrial Hsp70 system is regulated by the reversible unfolding of its nucleotide exchange factor, *J Mol Biol 358*, 1367-1377.

[141] Spector, S., Young, P., and Raleigh, D. P. (1999) Nativelike structure and stability in a truncation mutant of a protein minidomain: the peripheral subunit-binding domain, *Biochemistry 38*, 4128-4136.

[142] Kuhlman, B., and Raleigh, D. P. (1998) Global analysis of the thermal and chemical denaturation of the N-terminal domain of the ribosomal protein L9 in H2O and D2O. Determination of the thermodynamic parameters, deltaH(o), deltaS(o), and deltaC(o)p and evaluation of solvent isotope effects, *Protein Sci 7*, 2405-2412.

[143] Knapp, S., Karshikoff, A., Berndt, K. D., Christova, P., Atanasov, B., and Ladenstein, R. (1996) Thermal unfolding of the DNA-binding protein Sso7d from the hyperthermophile Sulfolobus solfataricus, *J Mol Biol 264*, 1132-1144.

[144] McCrary, B. S., Edmondson, S. P., and Shriver, J. W. (1996) Hyperthermophile protein folding thermodynamics: differential scanning calorimetry and chemical denaturation of Sac7d, *J Mol Biol 264*, 784-805.

[145] Wassenberg, D., Welker, C., and Jaenicke, R. (1999) Thermodynamics of the unfolding of the cold-shock protein from Thermotoga maritima, *J Mol Biol 289*, 187-193.

[146] Perl, D., Mueller, U., Heinemann, U., and Schmid, F. X. (2000) Two exposed amino acid residues confer thermostability on a cold shock protein, *Nat Struct Biol 7*, 380-383.

[147] Clarke, J., Cota, E., Fowler, S. B., and Hamill, S. J. (1999) Folding studies of immunoglobulin-like beta-sandwich proteins suggest that they share a common folding pathway, *Structure 7*, 1145-1153.

[148] Ruiz-Sanz, J., Filimonov, V. V., Christodoulou, E., Vorgias, C. E., and Mateo, P. L. (2004) Thermodynamic analysis of the unfolding and stability of the dimeric DNA-binding protein HU from the hyperthermophilic eubacterium Thermotoga maritima and its E34D mutant, *Eur J Biochem 271*, 1497-1507.

[149] Consalvi, V., Chiaraluce, R., Giangiacomo, L., Scandurra, R., Christova, P., Karshikoff, A., Knapp, S., and Ladenstein, R. (2000) Thermal unfolding and conformational stability of the recombinant domain II of glutamate dehydrogenase from the hyperthermophile Thermotoga maritima, *Protein Eng 13*, 501-507.

[150] Xu, S., Qin, S., and Pan, X. M. (2004) Thermal and conformational stability of Ssh10b protein from archaeon Sulfolobus shibattae, *Biochem J 382*, 433-440.

[151] Byrne, A., Williams, D. V., Barua, B., Hagen, S. J., Kier, B. L., and Andersen, N. H. (2014) Folding dynamics and pathways of the trp-cage miniproteins, *Biochemistry 53*, 6011-6021.

[152] Williams, D. V., Byrne, A., Stewart, J., and Andersen, N. H. (2011) Optimal salt bridge for Trp-cage stabilization, *Biochemistry 50*, 1143-1152.

[153] Barua, B., Lin, J. C., Williams, V. D., Kummler, P., Neidigh, J. W., and Andersen, N. H. (2008) The Trp-cage: optimizing the stability of a globular miniprotein, *Protein Eng Des Sel 21*, 171-185.

[154] Rodriguez-Granillo, A., Annavarapu, S., Zhang, L., Koder, R. L., and Nanda, V. (2011) Computational design of thermostabilizing D-amino acid substitutions, *J Am Chem Soc 133*, 18750-18759.

[155] Williams, D. V., Barua, B., and Andersen, N. H. (2008) Hyperstable miniproteins: additive effects of D- and L-Ala mutations, *Org Biomol Chem 6*, 4287-4289.

[156] Jager, M., Dendle, M., and Kelly, J. W. (2009) Sequence determinants of thermodynamic stability in a WW domain--an all-beta-sheet protein, *Protein Sci 18*, 1806-1813.

[157] McKnight, C. J., Doering, D. S., Matsudaira, P. T., and Kim, P. S. (1996) A thermostable 35-residue subdomain within villin headpiece, *J Mol Biol 260*, 126-134.

[158] Kubelka, J., Eaton, W. A., and Hofrichter, J. (2003) Experimental tests of villin subdomain folding simulations, *J Mol Biol 329*, 625-630.

[159] Kreitler, D. F., Mortenson, D. E., Forest, K. T., and Gellman, S. H. (2016) Effects of single alpha-to-beta residue replacements on structure and stability in a small protein: insights from quasiracemic crystallization, *J Am Chem Soc 138*, 6498-6505.

[160] Hsu, W. L., Shih, T. C., and Horng, J. C. (2015) Folding stability modulation of the villin headpiece helical subdomain by 4-fluorophenylalanine and 4-methylphenylalanine, *Biopolymers 103*, 627-637.

[161] Chiu, T. K., Kubelka, J., Herbst-Irmer, R., Eaton, W. A., Hofrichter, J., and Davies, D. R. (2005) High-resolution x-ray crystal structures of the villin headpiece subdomain, an ultrafast folding protein, *Proc Natl Acad Sci U S A 102*, 7517-7522.

[162] Kubelka, J., Chiu, T. K., Davies, D. R., Eaton, W. A., and Hofrichter, J. (2006) Sub-microsecond protein folding, *J Mol Biol 359*, 546-553.

[163] Xiao, S., Bi, Y., Shan, B., and Raleigh, D. P. (2009) Analysis of core packing in a cooperatively folded miniature protein: the ultrafast folding villin headpiece helical subdomain, *Biochemistry 48*, 4607-4616.

[164] Zheng, T. Y., Lin, Y. J., and Horng, J. C. (2010) Thermodynamic consequences of incorporating 4-substituted proline derivatives into a small helical protein, *Biochemistry 49*, 4255-4263.

[165] Bi, Y., Cho, J. H., Kim, E. Y., Shan, B., Schindelin, H., and Raleigh, D. P. (2007) Rational design, structural and thermodynamic characterization of a hyperstable variant of the villin headpiece helical subdomain, *Biochemistry 46*, 7497-7505.

[166] Merkel, J. S., Sturtevant, J. M., and Regan, L. (1999) Sidechain interactions in parallel beta sheets: the energetics of cross-strand pairings, *Structure 7*, 1333-1343.

[167] Yu, M. H., Weissman, J. S., and Kim, P. S. (1995) Contribution of individual side-chains to the stability of BPTI examined by alanine-scanning mutagenesis, *J Mol Biol 249*, 388-397.

[168] Liu, Y., Breslauer, K., and Anderson, S. (1997) "Designing out" disulfide bonds: thermodynamic properties of 30-51 cystine substitution mutants of bovine pancreatic trypsin inhibitor, *Biochemistry 36*, 5323-5335.

[169] de Prat Gay, G., Johnson, C. M., and Fersht, A. R. (1994) Contribution of a proline residue and a salt bridge to the stability of a type I reverse turn in chymotrypsin inhibitor-2, *Protein Eng 7*, 103-108.

[170] elMasry, N. F., and Fersht, A. R. (1994) Mutational analysis of the N-capping box of the alpha-helix of chymotrypsin inhibitor 2, *Protein Eng 7*, 777-782.

[171] Gribenko, A. V., and Makhatadze, G. I. (2007) Role of the charge-charge interactions in defining stability and halophilicity of the CspB proteins, *J Mol Biol 366*, 842-856.

[172] Hebert, E. J., Giletto, A., Sevcik, J., Urbanikova, L., Wilson, K. S., Dauter, Z., and Pace, C. N. (1998) Contribution of a conserved asparagine to the conformational stability of ribonucleases Sa, Ba, and T1, *Biochemistry 37*, 16192-16200.

[173] Grimsley, G. R., Shaw, K. L., Fee, L. R., Alston, R. W., Huyghues-Despointes, B. M., Thurlkill, R. L., Scholtz, J. M., and Pace, C. N. (1999) Increasing protein stability by altering long-range coulombic interactions, *Protein Sci 8*, 1843-1849.

[174] Pace, C. N., Horn, G., Hebert, E. J., Bechert, J., Shaw, K., Urbanikova, L., Scholtz, J. M., and Sevcik, J. (2001) Tyrosine hydrogen bonds make a large contribution to protein stability, *J Mol Biol 312*, 393-404.

[175] Takano, K., Scholtz, J. M., Sacchettini, J. C., and Pace, C. N. (2003) The contribution of polar group burial to protein stability is strongly context-dependent, *J Biol Chem 278*, 31790-31795.

[176] Yakovlev, G. I., Mitkevich, V. A., Shaw, K. L., Trevino, S., Newsom, S., Pace, C. N., and Makarov, A. A. (2003) Contribution of active site residues to the activity and thermal stability of ribonuclease Sa, *Protein Sci 12*, 2367-2373.

[177] Trevino, S. R., Gokulan, K., Newsom, S., Thurlkill, R. L., Shaw, K. L., Mitkevich, V. A., Makarov, A. A., Sacchettini, J. C., Scholtz, J. M., and Pace, C. N. (2005) Asp79 makes a large, unfavorable contribution to the stability of RNase Sa, *J Mol Biol 354*, 967-978.

[178] Fu, H., Grimsley, G., Scholtz, J. M., and Pace, C. N. (2010) Increasing protein stability: importance of DeltaC(p) and the denatured state, *Protein Sci 19*, 1044-1052.

[179] Lim, W. A., Farruggio, D. C., and Sauer, R. T. (1992) Structural and energetic consequences of disruptive mutations in a protein core, *Biochemistry 31*, 4324-4333.

[180] Hecht, M. H., Sturtevant, J. M., and Sauer, R. T. (1986) Stabilization of lambda repressor against thermal denaturation by site-directed Gly----Ala changes in alpha-helix 3, *Proteins 1*, 43-46.

[181] Hecht, M. H., Sturtevant, J. M., and Sauer, R. T. (1984) Effect of single amino acid replacements on the thermal stability of the NH2-terminal domain of phage lambda repressor, *Proc Natl Acad Sci U S A 81*, 5685-5689.

[182] Reidhaar-Olson, J. F., Parsell, D. A., and Sauer, R. T. (1990) An essential proline in lambda repressor is required for resistance to intracellular proteolysis, *Biochemistry 29*, 7563-7571.

[183] Stearman, R. S., Frankel, A. D., Freire, E., Liu, B. S., and Pabo, C. O. (1988) Combining thermostable mutations increases the stability of lambda repressor, *Biochemistry 27*, 7571-7574.

[184] Yang, W. Y., and Gruebele, M. (2004) Rate-temperature relationships in lambda-repressor fragment lambda 6-85 folding, *Biochemistry 43*, 13018-13025.

[185] Shirley, B. A., Stanssens, P., Steyaert, J., and Pace, C. N. (1989) Conformational stability and activity of ribonuclease T1 and mutants. Gln25----Lys, Glu58----Ala, and the double mutant, *J Biol Chem 264*, 11621-11625.

[186] Shirley, B. A., Stanssens, P., Hahn, U., and Pace, C. N. (1992) Contribution of hydrogen bonding to the conformational stability of ribonuclease T1, *Biochemistry 31*, 725-732.

[187] Fabian, H., Schultz, C., Backmann, J., Hahn, U., Saenger, W., Mantsch, H. H., and Naumann, D. (1994) Impact of point mutations on the structure and thermal stability of ribonuclease T1 in aqueous solution probed by Fourier transform infrared spectroscopy, *Biochemistry 33*, 10725-10730.

[188] Giletto, A., and Pace, C. N. (1999) Buried, charged, non-ion-paired aspartic acid 76 contributes favorably to the conformational stability of ribonuclease T1, *Biochemistry 38*, 13379-13384.

[189] Schindler, T., Mayr, L. M., Landt, O., Hahn, U., and Schmid, F. X. (1996) The role of a trans-proline in the folding mechanism of ribonuclease T1, *Eur J Biochem 241*, 516-524.

[190] Schubert, W. D., Schluckebier, G., Backmann, J., Granzin, J., Kisker, C., Choe, H. W., Hahn, U., Pfeil, W., and Saenger, W. (1994) X-ray crystallographic and calorimetric studies of the effects of the mutation Trp59-->Tyr in ribonuclease T1, *Eur J Biochem 220*, 527-534.

[191] Kellis, J. T., Jr., Nyberg, K., and Fersht, A. R. (1989) Energetics of complementary side-chain packing in a protein hydrophobic core, *Biochemistry 28*, 4914-4922.

[192] Axe, D. D., Foster, N. W., and Fersht, A. R. (1999) An irregular beta-bulge common to a group of bacterial RNases is an important determinant of stability and function in barnase, *J Mol Biol 286*, 1471-1485.

[193] Vu, N. D., Feng, H., and Bai, Y. (2004) The folding pathway of barnase: the rate-limiting transition state and a hidden intermediate under native conditions, *Biochemistry 43*, 3346-3356.

[194] Eberhardt, E. S., Wittmayer, P. K., Templer, B. M., and Raines, R. T. (1996) Contribution of a tyrosine side chain to ribonuclease A catalysis and stability, *Protein Sci 5*, 1697-1703.

[195] Catanzano, F., Graziano, G., Capasso, S., and Barone, G. (1997) Thermodynamic analysis of the effect of selective monodeamidation at asparagine 67 in ribonuclease A, *Protein Sci 6*, 1682-1693.

[196] Salamanca, S., Villegas, V., Vendrell, J., Li, L., Aviles, F. X., and Chang, J. Y. (2002) The unfolding pathway of leech carboxypeptidase inhibitor, *J Biol Chem 277*, 17538-17543.

[197] Kadonosono, T., Chatani, E., Hayashi, R., Moriyama, H., and Ueki, T. (2003) Minimization of cavity size ensures protein stability and folding: structures of Phe46-replaced bovine pancreatic RNase A, *Biochemistry 42*, 10651-10658.

[198] Johnson, R. J., Lin, S. R., and Raines, R. T. (2007) Genetic selection reveals the role of a buried, conserved polar residue, *Protein Sci 16*, 1609-1616.

[199] Pecher, P., and Arnold, U. (2009) The effect of additional disulfide bonds on the stability and folding of ribonuclease A, *Biophys Chem 141*, 21-28.

[200] Gotte, G., Donadelli, M., Laurents, D. V., Vottariello, F., Morbio, M., and Libonati, M. (2006) Increase of RNase a N-terminus polarity or C-terminus apolarity changes the two domains' propensity to swap and form the two dimeric conformers of the protein, *Biochemistry 45*, 10795-10806.

[201] Eftink, M. R., Ghiron, C. A., Kautz, R. A., and Fox, R. O. (1991) Fluorescence and conformational stability studies of Staphylococcus nuclease and its mutants, including the less stable nuclease-concanavalin A hybrids, *Biochemistry 30*, 1193-1199.

[202] Carra, J. H., Anderson, E. A., and Privalov, P. L. (1994) Thermodynamics of staphylococcal nuclease denaturation. II. The A-state, *Protein Sci 3*, 952-959.

[203] Shortle, D., Meeker, A. K., and Freire, E. (1988) Stability mutants of staphylococcal nuclease: large compensating enthalpy-entropy changes for the reversible denaturation reaction, *Biochemistry 27*, 4761-4768.

[204] Carra, J. H., and Privalov, P. L. (1995) Energetics of denaturation and m values of staphylococcal nuclease mutants, *Biochemistry 34*, 2034-2041.

[205] Gittis, A. G., Stites, W. E., and Lattman, E. E. (1993) The phase transition between a compact denatured state and a random coil state in staphylococcal nuclease is first-order, *J Mol Biol 232*, 718-724.

[206] Hinck, A. P., Truckses, D. M., and Markley, J. L. (1996) Engineered disulfide bonds in staphylococcal nuclease: effects on the stability and conformation of the folded protein, *Biochemistry 35*, 10328-10338.

[207] Byrne, M. P., and Stites, W. E. (1995) Chemically crosslinked protein dimers: stability and denaturation effects, *Protein Sci 4*, 2545-2558.

[208] Chen, J., Lu, Z., Sakon, J., and Stites, W. E. (2000) Increasing the thermostability of staphylococcal nuclease: implications for the origin of protein thermostability, *J Mol Biol 303*, 125-130.

[209] Leung, K. W., Liaw, Y. C., Chan, S. C., Lo, H. Y., Musayev, F. N., Chen, J. Z., Fang, H. J., and Chen, H. M. (2001) Significance of local electrostatic interactions in staphylococcal nuclease studied by site-directed mutagenesis, *J Biol Chem 276*, 46039-46045.

[210] Maki, K., Cheng, H., Dolgikh, D. A., Shastry, M. C., and Roder, H. (2004) Early events during folding of wild-type staphylococcal nuclease and a single-tryptophan variant studied by ultrarapid mixing, *J Mol Biol 338*, 383-400.

[211] Baase, W. A., Liu, L. J., Tronrud, D. E., and Matthews, B. W. (2010) Lessons from the lysozyme of phage T4, *Protein Science 19*, 631-641.

6. Future directions

The sensitivity and relative simplicity of fluorescence experiments have already contributed to their importance across many applications in biochemistry. Although new FRET probes and bright dyes such as the Alexa series have spurred recent advances, small fluorophores like F_{CN} and Trp will continue to be essential tools due to their small size and minimally-perturbative nature. These small fluorophores are particularly useful in applications that require a fluorophore to be inserted into a sterically hindered environment, such as the core of a protein, a channel in a transmembrane transport protein or a binding interface. In addition, the small size of Trp and F_{CN} compared to large, bright dyes allows for more site-specific conclusions to be made.

Inherent quantum yield changes can be difficult to predict, particularly in Trp, due to multiple, often competing effects. Consequently, Trp emission frequency shifts are often examined rather than emission intensity. These effects can also be subtle, so the introduction of a quencher is desirable to simplify data interpretation. FRET quenching is a popular and well-established method in the field, but is limited to applications where the distance between the donor and acceptor is close to R_0 or (if the donor and acceptor are very close in one state) the change in distance is very large. In addition, FRET cannot distinguish between direct contact and any other distance closer than ~0.5R_0. Contact quenchers, such as M_{Se} are therefore useful in combination with FRET or other methods to examine specific interactions. M_{Se}, as a single atom substitution of Met, offers one of the least perturbative contact quenchers possible.

Small, minimally perturbative contact quenching pairs are most valuable in small proteins systems that might behave differently with large, polyaromatic dyes attached, or systems in which

a very specific interaction is under examination. Amyloid forming proteins, such as islet amyloid polypeptide (IAPP) are one example of such a system. IAPP forms a regular, cross-β structure, apparent by CD or infrared spectroscopy. However, the details of the side chain packing remain uncertain; F_{CN}-M_{Se} pairs could be used to confirm specific interactions. The structure and interactions of pre-amyloid oligomers are even more poorly understood. There is evidence of transient α-helical structure, which F_{CN}-M_{Se} pairs could be used to more accurately map.

Other amyloid systems, such as amyloid β and α-synuclein, which form amyloid deposits in Alzheimers and Parkinsons patients, respectively, could also be studied using these fluorescence quenching pairs. Whereas IAPP and amyloid β can be prepared by solid phase peptide synthesis, which facilitates the incorporation of multiple unnatural amino acids, α-synuclein must be prepared recombinantly or by native chemical ligation. This highlights an important future direction: development of methods for the simultaneous recombinant incorporation of F_{CN} and M_{Se}. The fluorescence quenching pairs described here could also be applied to studies of membrane proteins. In addition, the ability of M_{Se} to quench multiple fluorophores potentially allows for the simultaneous incorporation of multiple, independently excitable fluorophore-quencher pairs to probe multiple interactions in a single protein construct.

Predictions about the chain-length dependent thermodynamic properties of proteins have potential impacts in many fields. The clear differences observed between mesophilic and thermophilic organisms motivates further examination of the origins of thermal stability in thermophilic organisms. Protein stability is major limitation of protein-based pharmaceuticals; many favorable properties of proteins as drug molecules are undercut by their rapid turnover in the body. The ability to engineer significantly more stable proteins would open up new possibilities in drug design. Sequence data is already widely used to predict protein structure and function, and

this information can be used to model the behavior of the cellular environment. A major missing piece of such models is stability data, as protein turnover is critical to cellular homeostasis. The chain-length dependent stability model could fill this gap, refining systems models. The existing data on thermophilic proteins is clearly more limited than what is available for mesophilic proteins. Further refinement of the model for thermophilic organisms motivates additional studies of proteins from these organisms. In addition, the existing model treats all proteins as single, cooperative folding units. More accurate predictions of proteome properties could be made with a more complete understanding of the distribution of domain sizes. Given sufficient data, the model could also be parametrized for psychrophiles, which could lead to a greater understanding of evolutionary adaptations to resist cold denaturation or methods for the design of cold-stable protein therapeutics.

The chain-length dependent relationship between stability and T_M has clear implications for thermal shift assays. Critically, it provides a reference point for the magnitude of observed thermal shifts, making it clear that a small change in T_M does not necessarily indicate weak drug binding in a large protein. Similarly, the relationship cautions against the use of thermal shift assays in early drug discovery, where potential lead compounds might be missed due to very small changes in T_M. More broadly, it is clear that changes in T_M are strongly correlated to changes in stability, whether the result of drug binding, mutation, or other effects. This could make it possible to rapidly estimate stabilities for a large number of mutants based only on values of T_M.