USING CIRCULAR DICHROISM TO DETERMINE HOW FLUORINATION AFFECTS PEPTIDE FOLDING

Benjamin J. Levin

April 23rd, 2013

This thesis has been read and approved by Professor Neil Marsh.

Signed: _____ Date: ___/___/___

Faculty advisor e-mail: nmarsh@umich.edu Phone: (734) 763-6096

Abstract:

It has been established that substituting hydrophobic amino acids in proteins with fluorinated analogs leads to an increase in their stability against heat and chemical denaturants. The cause of this effect has been difficult to determine. Some results suggest it is due to fluorine's increased size and that the effect is entropic, while other evidence points towards an enthalpic contribution, stemming from the fluorous effect. To better understand the causes of this enhanced stability, the Marsh Lab synthesized 12 α_4-helix peptides that incorporate different numbers and types of fluorinated residues into the interior of the folded helical bundle. Heat and guanidinium hydrochloride were used to denature the tetramer and this two-state equilibrium was monitored using circular dichroism. An algorithm was developed to describe this equilibrium and a script was written in MATLAB to relate the CD signal to the thermodynamic parameters of interest. The $\Delta H°$, $\Delta S°$, and $\Delta C_p°$ of unfolding were calculated for each peptide and it was found that the change in entropy and heat capacity were highly correlated with the change in apolar solvent accessible surface area, while the enthalpy change had no such correlation. These results indicate that fluorine's stabilizing effect is due to its larger size and the conventional hydrophobic effect and not due to any unusual fluorous interactions.

Introduction:

Proteins are among the most important components of biological organisms (*1*). With water and protein composing 70% and 15% of *E. coli* by weight respectively, it is no wonder that much of the earliest biochemical research was concerned with studying them (*2*). In fact, the term protein was coined in 1838 by Berzelius and it stems from the Greek *proteios*, meaning of first importance (*1*). They are macromolecules and it was realized in the 19th century that they consist of amino acid monomers (*3, 4*). There are 20 standard amino acids and those plus specialized nonstandard amino acids are the principle components of proteins (*1*). Critically, Anfinsen demonstrated in the 1960s that the primary structure or amino acid sequence of a protein is enough to determine its three-dimensional structure and therefore its function (*5*).

The structure of each amino acid plays a key role in determining how proteins fold. Amino acids consist of an α-carbon bonded to an amino group, a carboxyl group, a hydrogen, and a side chain (R group). The α-carbon is a stereocenter and in the vast majority of biological systems the only stereoisomers present are the L-amino acids (using the Fischer convention) (*1, 6*). The only feature that differs between amino acids is their side chains. Any introductory biochemistry textbook will have the standard amino acids illustrated (*1*). Amino acids link together via peptide bonds; a short sequence of amino acids is called a peptide. Given the primary structure of a peptide, the next step is examining its secondary structure.

The secondary structure of a peptide is formally the local conformation of its backbone. It is the three dimensional configuration of a small portion of the molecule. The major determinants of secondary structure are electrostatics, hydrogen bonding, disulfide bonds, and perhaps most importantly the hydrophobic effect (*1*). The favorability of salt bridge formation between oppositely charged residues and of hydrogen bonding between amino acids is a factor in

the overall structure, but they are often not substantially more favorable than interactions with solvent. The hydrophobic effect refers to the tendency of nonpolar groups to minimize their contact with water (*7*). An aqueous solution can maximize its entropy and minimize its free energy by excluding nonpolar side chains found in peptides from the aqueous environment (*8*). An unfolded protein in solution will almost immediately undergo hydrophobic collapse into a molten globule state, indicating just how influential the hydrophobic effect is (*9*).

It may come as a surprise that the vast majority of protein secondary structure can be described by just a few different motifs. Ramachandran rigorously showed that the conformational degrees of freedom for amino acids are relatively restricted and this greatly limits the configuration space for proteins (*10*). The major repeating structures are the α-helix and the β-sheet, while the nonrepetitive structures are the coil or loop conformations (*1*). Of particular interest here are the α-helical and random coil conformations. The α-helix has a number of features that lead to its prevalence. The amide hydrogens are an optimal distance from the carbonyl oxygen, which leads to an extensive hydrogen bond network. The backbone arrangement is very compact, which leads to an entropic gain for water and an enthalpic gain from van der Waals interactions for the protein. There is a net dipole from the orientation of the amide groups, leading to further stabilization. Finally, the side chains are all oriented outward, minimizing steric interactions (*1, 11*). The random coil configuration is one in which the residues are totally disordered and are rapidly fluctuating.

Protein folding can be regarded as an equilibrium process (*12*). Regardless of whether denaturation is performed with a chaotropic agent like guanidinium hydrochloride (GuHCl) or with heat, protein unfolding is cooperative (*1*). As soon as part of the protein becomes unfolded, the rest of the structure will unfold as well. For example, as soon as part of an α-helix loses its

structure, the entire chain will collapse into a random coil. Although relatively stable intermediates may exist on the unfolding pathway, especially for smaller peptides the folded and unfolded conformations reach equilibrium very quickly.

The secondary structure of peptides plays a key role in determining their function, of which there are many. Perhaps the most well-established function of peptides is as hormones. Glucagon (29 residues), insulin (51 res.), parathyroid hormone (84 res.), calcitonin (33 res.), and human growth hormone (191 res.) are just a few well studied examples (*1*). More recently, there has been substantial interest in antimicrobial peptides (*13*, *14*). These are evolutionarily ancient peptides that target infective agents without harming animal or plant cells in physiologically relevant concentrations (*13*). Although their structural diversity is impressive, these peptides all seem to differentiate between bacterial cells and eukaryotic cells via charge interactions on the cell membranes (*15*). The outer surface of bacterial membranes is composed of lipids with negatively charged phospholipid headgroups, while in eukaryotes the outer surface is typically more neutral with the negative groups on the internal side of the membrane. Thus it seems that the peptides interact specifically with the negatively charged membranes and then kill the microbes through a variety of different pathways. How each peptide kills pathogens has been the subject of much debate (*15*). An understanding of these mechanisms could lead to a new class of antibiotics, especially considering bacteria cannot seem to develop any resistance to these peptides in a physiologically relevant setting or timescale (*16*).

In order to better study the structure of peptides, the element fluorine has been utilized extensively (*17*). Fluorine is a fascinating element with a number of unusual properties, particularly in organic compounds. For instance, it exhibits the aptly named “fluorous effect”. A solution of perfluorohexane is not soluble in hexane or in water and a mixture of the three results

in three phases, indicating that the compound is neither truly hydrophobic nor hydrophilic (*18*, *19*). Even more strangely, 1,2-difluoroethane is the canonical example of the Gauche Effect (*20*). Steric considerations can explain why the staggered form of butane is 0.9 kcal/mol more stable than the gauche configuration, but the opposite holds for 1,2-difluoroethane: The gauche form is more stable by about 0.6 kcal/mol (*20*). This is due to hyperconjugation (*21*). That is, the $\sigma_{C\text{-}H}$ bond interacts favorably with the $\sigma^*_{C\text{-}F}$ that is anti to it or equivalently that the HOMO of the carbon-hydrogen bond interacts with the low lying LUMO of the anti carbon-fluorine bond. Although fluorine does have some unusual features, it has several that could make it a powerful tool in the study of proteins. The van der Waals radius of fluorine, about 1.4 Å, is similar to that of hydrogen (~1.2 Å), and so the two are relatively isosteric and can be substituted for each other while only minimally perturbing the structure (*22*). The fluorine-19 nucleus has a spin of ½, is 83% as sensitive as ^{1}H, and is 100% abundant (*23*). This has made ^{19}F-NMR not only a powerful tool *in vitro*, but due to the lack of fluorine in biological compounds (see ref. 24 for exceptions), *in vivo* ^{19}F-NMR has become quite popular as well (*18*).

The incorporation of fluorine into peptides can be done in a number of ways. Solid phase synthesis with fluorine-containing amino acid analogues is the most flexible tool for short peptides of less than 50 residues (*18*). Another method requires the use of an amino acyl-tRNA synthetase that can recognize fluorinated substrates. Although the latter method is less developed currently, it has the potential to allow the site-specific incorporation of fluorinated amino acids into large proteins (*25*). After incorporation, the analysis of fluorine's local chemical environment can be done with NMR, as there is no background signal. Just one example is the determination of how fluorinated analogs of antimicrobial peptides interact with bacterial cell

membranes (*26*). By examining how the chemical shift and width of the peaks change over time, insight into the peptide's mechanism can be gained.

In addition to their usefulness as probes, fluorinated analogs of proteins can have a more stable folded form when packed appropriately. When coiled-coil proteins were analyzed and the interior hydrophobic leucine residues were replaced with trifluoroleucine, the melting temperature and the midpoint concentration for urea increased, indicating increased resistance to thermal and chemical denaturation (*27*). In addition, proteins fluorinated in specific locations are known to become more resistant to proteolysis (*28*). Increased resistance to proteolysis is medically relevant, as it could lead to peptide pharmaceuticals that have a longer half-life, as typically these drugs are metabolized too rapidly to be useful. The basis for this is straightforward. The hydrogen-to-fluorine substitution alters the structure just enough that proteolytic enzymes cannot conform to fit around that segment of the protein, but not enough to dramatically decrease the activity of the peptide (*28*).

The increased resistance to thermal and chemical denaturation has been harder to explain (*29*). Methods to measure this increase in stability most commonly use the difference in melting temperature or free energy of unfolding, but these cannot give a clear structural basis for the cause. Complicating matters further is enthalpy-entropy compensation, which makes the thermodynamic data even tougher to interpret without structural support (*30*). Several different theories have been proposed. Some rely simply on the increased size and hydrophobicity of fluorine *(28)*. Because fluorocarbon chains are even more hydrophobic than hydrocarbons, they will be relatively more stable in a folded protein where they can be put into a water-excluded cavity. However, fluorine's unusual properties suggest the cause might be more complex. The fluorous effect and the unusually low lying $\sigma^*_{C\text{-}F}$ make alternative explanations plausible.

Perhaps fluorine atoms in the hydrophobic cavity are interacting favorably with one another, providing an enthalpic contribution to folding. To fully utilize fluorine in peptides, an understanding of why they increase stability is essential.

A model system to explore the contribution of fluorine to peptide stability was developed by the Marsh Lab in 2004 (*31*). See Figure 1 for its structure. In their design, the folded tetramer has hydrophilic residues placed externally, while the hydrophobic fluorous residues are internal and able to interact with each other while avoiding the aqueous solution. Conversely, in the unfolded state the protein is in a random coil conformation. Verification of the α-helical content was done via circular dichroism, while analytical ultracentrifugation and gel filtration confirmed the monomer-tetramer equilibrium. The fluorinated analog was more stable to chemical denaturation than the non-fluorinated version, demonstrating that fluorine is somehow playing a stabilizing role (*31*).

Additional analogs were synthesized, each containing differing amounts of fluorine in the interior cavity (*32*). These peptides were analyzed using the above techniques and similar results observed. Interestingly, the free energy of folding seemed to increase linearly with the number of fluorine atoms in the peptide. Subsequently, proteolysis experiments revealed that the fluorinated analogs have a substantially increased resistance to degradation (*33*). Thus this sequence of peptides demonstrates the desired properties, but it is still unclear how fluorine confers these effects.

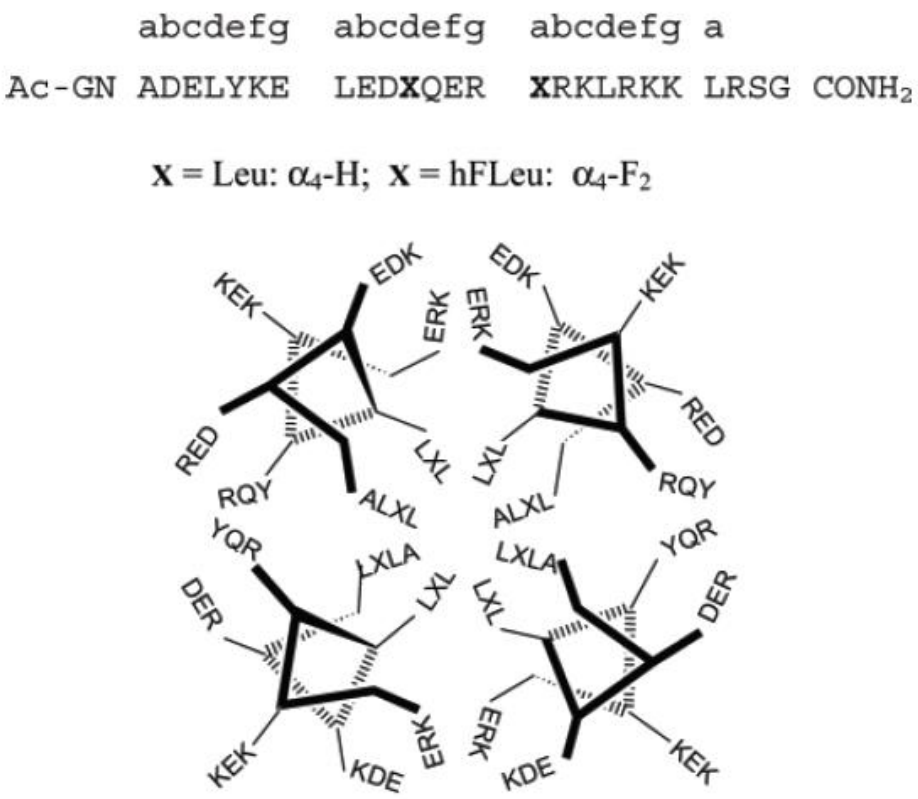


Figure 1: Reprinted with permission from Ref. 31. Copywrite 2004 American Chemical Society. This was the first antiparallel α_4-helix bundle designed by the Marsh Lab to determine how fluorine affects the stability of folded proteins. Note how the hexafluoroleucine residues are packed into the hydrophobic interior of the tetramer.

Further variants of α_4H made it clear that it is not only the amount of fluorine, but also where the fluorine is located that plays a role (*34*). By selectively placing hexafluoroleucine at specific 'a' and 'd' positions in the chain (see Figure 2), the packing structure was shown to influence the free energy change. However, crystal structures of the tetramers were not immediately available and so it was challenging to determine the details of the effect. Notably, enthalpy-entropy compensation could have been taking place, making anything other than free energy calculations impossible to interpret.

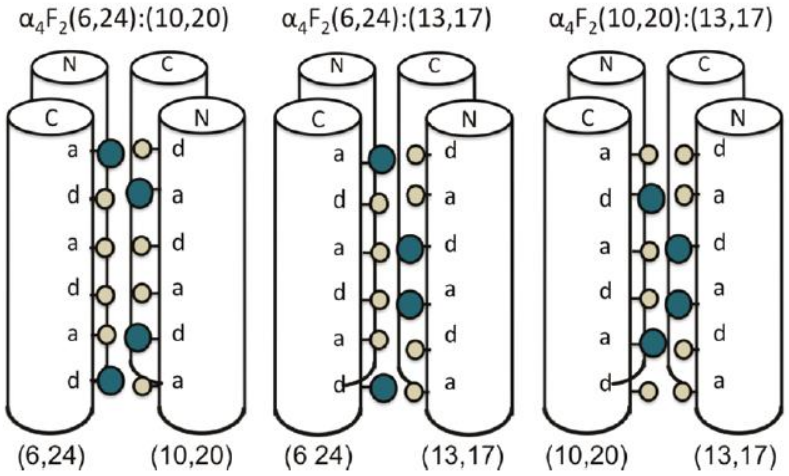


Figure 2: Reprinted with permission from Ref. 34. Copywrite 2009 American Chemical Society. These are several variants of the synthesized α_4H peptide, where the hexafluoroleucine is indicated with darker, larger spheres. The packing differences are intimately related to the stability of the tetramer.

To conclusively determine the source of the stability increase, a detailed thermodynamic analysis was undertaken (*29*). Guanidinium hydrochloride and heat were used to denature the peptides and the folding was monitored using circular dichroism. Circular dichroism is a powerful tool to monitor protein secondary structure, showing characteristic peaks for α-helices, β-sheets, and random coils (*35*). In particular, the ellipticity at 222 nm is strongly correlated with α-helical content and since all of the α_4 analogs above have an α-helical folded state and a random coil unfolded state, this provides a convenient and accurate way to determine the proportion folded under given conditions.

Materials and Methods:

Before peptide synthesis could be performed, the fluorinated amino acids had to be obtained. 4,4,4-Trifluoroethylglycine was purchased from SynQuest Laboratory and enzymatically resolved using porcine kidney acylase I as previously described (*36*). Next, L-

5,5,5,5',5',5'-hexafluoroleucine was synthesized using established methods (*37*). Boc and Fmoc protected β-*tert*-butyl-L-alanine were bought from AnaSpec Inc. Peptide synthesis was performed using standard protocols (*31*, *38*). The peptide sequences are given in Figure 3: α_4H and $\alpha_4$tbA_6 were synthesized via the Fmoc protocol, while the others were done with Boc procedures.

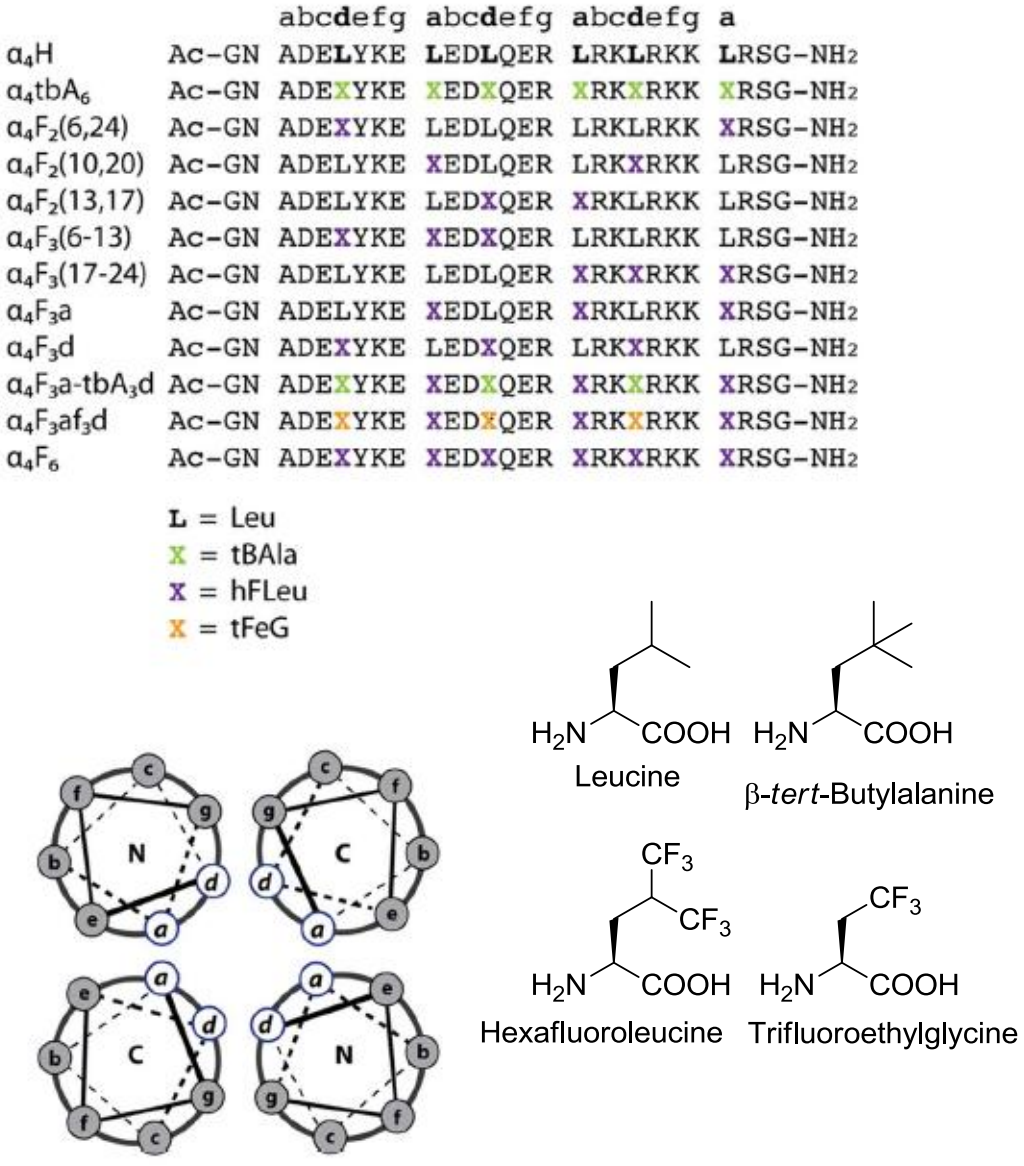


Figure 3: Adapted with permission from Ref. 29. Copywrite 2012 American Chemical Society. This shows the sequences of all of the synthesized peptides, as well as the structures of the folded proteins. The structures of the amino acids are also given.

The circular dichroism experiments were done in an Avis 62DS spectropolarimeter at 222 nm with a 1 mm path length cuvette. Stock solutions were prepared that contained 40 μM peptide (monomer concentration), 10 mM potassium phosphate buffer at pH 7.0, and 9 to12 different concentrations of guanidinium hydrochloride. Each of those solutions had ellipticity measurements averaged over 10 seconds and the temperature for each was varied from 4 °C to 90 °C in 2 °C increments. Between 430 and 512 data points were obtained for each peptide.

The relationship between the CD signal and the thermodynamic parameters was established as follows. It was assumed that the equilibrium was between the monomeric random coil peptide and the folded 4-α-helix bundle. Suppose [P] is the total amount of peptide and let [U] and [F] be the concentrations of monomeric peptide and folded tetramer respectively, so $[\mathrm{P}] = 4[\mathrm{F}] + [\mathrm{U}]$. For all of the experiments presented here [P] = 40 μM. Then at any given set of conditions, the ellipticity θ can be written as the sum of the contributions from the unfolded and folded components, as in Eqn. 1.

$$\theta = \theta_{\mathrm{u}} * \frac{[\mathrm{U}]}{[\mathrm{P}]} + \theta_{\mathrm{f}} * \frac{[\mathrm{P}] - [\mathrm{U}]}{[\mathrm{P}]} \quad (1)$$

This relates the measured ellipticity to the amount of unfolded peptide, and so now a relationship between [U] and the thermodynamic parameters must be found.

The equilibrium constant between F ⇄ 4U is given by the following equation.

$$K = \frac{[\mathrm{U}]^4}{[\mathrm{F}]} = \frac{4[\mathrm{U}]^4}{[\mathrm{P}] - [\mathrm{U}]} \quad (2)$$

For monomer and dimer folding, the above equation can be solved for [U] because the exponent is only equal to 1 or 2. In general though, the above equation cannot be solved for [U] in a simple or convenient way (although an analytic solution does exist) (*39*). It is far simpler to use a numeric solution. With modern computers, the error from this is completely negligible. In Eqn. 2

we are given a function from [U] to *K*, and so an inverse must be shown to exist and to be unique in the domain where [U] lies in the interval (0, [P]). This will be done using real analysis (*40*). *K* is clearly defined for all [U] in the desired interval and as a rational function of [U], it is also differentiable where defined.

$$\frac{dK}{d[U]} = \frac{4[\mathrm{U}]^3 * (4[\mathrm{P}] - 3[\mathrm{U}])}{([\mathrm{P}] - [\mathrm{U}])^2} \quad (3)$$

This function is positive for all [U] in the domain of interest, and so *K* is monotonic and thus there exists an inverse function that maps from *K* to [U] as desired. To prove uniqueness, it suffices to define [U] = 0 when K = 0 and [U] = [P] as *K* tends to ∞. Thus, Eqn. 2 has a unique inverse function. Informally, Eqn. 2 finds the *K* value for each [U], and the previous steps show that Eqn. 2 can also be used to find the [U] that corresponds to each *K*. Although this cannot be done explicitly, this computation can be done numerically by finding the (unique) root of

$$[\mathrm{U}]^4 + \frac{K[\mathrm{U}]}{4} - \frac{K[\mathrm{P}]}{4} = 0 \quad (4)$$

between 0 and [P][1].

Returning to Eqn. 1, this can be rewritten as

$$\theta = \theta_{\mathrm{u}} * \frac{[\mathrm{U}](K)}{[\mathrm{P}]} + \theta_{\mathrm{f}} * \frac{[\mathrm{P}] - [\mathrm{U}](K)}{[\mathrm{P}]} \quad (5)$$

where now the monomer concentration is a function of the equilibrium constant. The equilibrium constant can be related to the Gibbs free energy, and this can then be substituted into the Gibbs-Helmholtz equation (Eqn. 6 below) as follows (*41*).

$$\Delta G = \Delta H° - T\Delta S° + C_p° \left(T - T_0 + T \ln \frac{T_0}{T}\right) - m[\mathrm{GuHCl}] \quad (6)$$

[1] Although the above procedure is rigorous, inverse functions are rather common. For example, the inverse of the squaring function $y = x^2$ is the square root function $x = \sqrt{y}$. Although the fourth order polynomial in Eqn. 4 is more complicated than the squaring function, the mathematics involved is entirely analogous.

$$K = \exp\left(-\frac{\Delta G}{RT}\right) = \exp\left(-\frac{\Delta H° - T\Delta S° + C_p°\left(T - T_0 + T\ln\frac{T_0}{T}\right) - m[\text{GuHCl}]}{RT}\right) \quad (7)$$

In Eqn. 7, $\Delta H°$, $\Delta S°$, and $\Delta C_p°$ represent the change in enthalpy, entropy, and heat capacity upon unfolding at a standard temperature T_0 (typically 298 K). Over this temperature range it cannot be assumed that the enthalpy and entropy changes are constant, but these are accounted for with the $\Delta C_p°$ term. The m term relates the free energy of unfolding to the concentration of denaturant, namely that the free energy of unfolding decreases linearly with [GuHCl]. In the above it is safe to assume that $\Delta C_p°$ and m are constant (*41*, *42*). This gives K as a function of temperature, [GuHCl], and the desired thermodynamic parameters.

Thus, using Eqn. 5, we have the ellipticity in terms of [U], and Eqn. 4 gives [U] as a function of K. From Eqn. 7, K is dependent on the temperature T, [GuHCl], and the constants $\Delta H°$, $\Delta S°$, $\Delta C_p°$, and m. Combining all of this together gives the ellipticity as a function of the experimental conditions and the thermodynamic parameters. The only note remaining is to determine the functional form for θ_u and θ_f. As done in Ref. 41, these were given a linear dependence on the parameters.

$$\theta_u = a + bT + c[\text{GuHCl}]$$

$$\theta_f = d + eT + f[\text{GuHCl}]$$

This is more of an empirical rather than theoretical assumption, but it has been shown to be accurate in most cases and perhaps more importantly setting these to constants rather than planes gave similar values for the peptides studied.

The above theory was used to create a curve fitting algorithm in MATLAB (MathWorks, Inc.). By using the function given by θ vs. T vs. [GuHCl], a nonlinear least squares program could provide estimates and confidence intervals for the thermodynamic parameters.

Results:

The CD spectra were obtained as expected. It had been determined previously that the peptides existed as monomers and tetramers with no significant intermediate structures (*34*). It was confirmed here that the folding was reversible and that no precipitate formed, indicating that the assumptions used in the modeling held for these peptides (*29*). Before the results of the fits are given, the algorithms developed are given below.

```
R = 8.3145; %Ideal Gas Constant in SI units
P = 40e-6; %Free Peptide Concentration in M
T0 = 298.15; %Reference Temperature in K

%1 kcal = 4184 J exactly
%Make sure data is saved in three columns:
    %Temp (Celcius), Theta, [Den] (Molar)

data = uiimport; %Import data; it will be saved as 'data'
data = data.data; %MatLab Glitch; this is the workaround
Temp = data(:,1) + 273.15; %Converts to Kelvin
Theta_Obsd = data(:,2);
Den = data(:,3);
Cond = [Temp Den];
%Cond is the n x 2 matrix of Temp and Den, and Theta_Obsd is output

%Define additional functions K, U, and f
K =  @(b,Cond)exp(-(b(3)-Cond(:,1)*b(4)+b(5)*(Cond(:,1)-T0 ...
    +Cond(:,1).*log(T0./Cond(:,1)))-Cond(:,2)*b(6))./(R*Cond(:,1)));
U = @(b,Cond) arrayfun(@(k) fzero(@(x) x^4+k*x/4-k*P/4, [0 P]), K(b,Cond));
f = @(b,Cond)
1/P*((b(1)*ones(length(Cond(:,1)),1)+b(7).*Cond(:,1)+b(8).*Cond(:,2)).*U(b,Co
nd)...
+(b(2)*ones(length(Cond(:,1)),1)+b(9).*Cond(:,1)+b(10).*(Cond(:,2))).*(P*ones
(length(Cond(:,1)),1)-U(b,Cond)));

%Initial Values Module
beta0 = zeros(10,1);
beta0(1) = max(Theta_Obsd)+1;
beta0(2) = min(Theta_Obsd)-1;
Uest = P*(Theta_Obsd-beta0(2)*ones(length(Theta_Obsd),1))/(beta0(1)-
beta0(2));
Kest = 4*Uest.^4./(P*ones(length(Temp),1)-Uest);
DGest = -R*Temp.*log(Kest);

TestMat = ones(length(Temp),4);
for i = 1:length(Temp)
    TestMat(i,1) = 1;
    TestMat(i,2) = -Temp(i);
    TestMat(i,3) = Temp(i)-T0+Temp(i)*log(T0/Temp(i));
    TestMat(i,4) = -Den(i);
```

```
end

ParaEst = linsolve(TestMat,DGest);

beta0 = [max(Theta_Obsd);min(Theta_Obsd); ParaEst(1);
ParaEst(2);ParaEst(3);ParaEst(4);0;0;0;0];
%Actual Curve Fitting Portion
[beta, r, J, COVB, mse] = nlinfit(Cond, Theta_Obsd, f, beta0);

ci = nlparci(beta,r,'covar', COVB);
```

Because the above algorithm can be generalized to systems beyond studying tetramer folding with circular dichroism, it will be analyzed in depth. Through defining the matrix 'Cond' in the above script, the program is defining constants and putting the experimental data into a format that MATLAB can work with. The equilibrium constant 'K' is defined using Eqn. 7 to be a function of the constant thermodynamic parameters and *m* and the experimental conditions *T* and [GuHCl]. Note that the constants are not defined yet, even though 'K' is a function of them. Next, the free monomer concentration 'U' is defined as a function of those same parameters by using 'K' and Eqn. 4. Finally, the ellipticity 'f' is presented as a function of all of the variables through Eqn. 5. In the above, all of the parameters are given in the vector b = [*a*; *e*; $\Delta H°$; $\Delta S°$; $\Delta C_p°$; *m*; *b*; *c*; *f*; *g*].

After the required definitions were programmed, parameter estimates were computed. When doing nonlinear least squares it is necessary to input a functional form, experimental data, and then initial values for the constants to be determined. Instead of requiring a user to estimate parameters each time, a module was created to do this automatically. To determine estimates, it was first assumed that the base planes were constant (i.e. $b = c = f = g = 0$) and that the contribution of the folded peptide to ellipticity was just above the maximum signal at 222 nm and contribution from the unfolded peptide was just below the minimum signal at 222 nm (*35*).

This allows [U] to be estimated for each experimental condition using Eqn. 1. From that value of [U], Eqn. 2 gives a value for *K* and then

$$-RT \ln K = \Delta G$$

allows ΔG to be estimated at each experimental condition. From Eqn. 6, using the estimated free energy of unfolding as well as the temperature and [GuHCl], which are the known experimental conditions, we have a linear system of equations where the unknowns are the desired parameters. This system can be solved easily and the resulting 'ParaEst' is the collection of those estimates.

It is reasonable to ask if these computed parameters could be used as the results. Just doing that would be inaccurate. In particular, it is rarely appropriate to transform a nonlinear model into a linear one and then perform straightforward linear regression and expect the results to be precise. This is a common problem when analyzing Lineweaver-Burk plots, for example (*43*). More specifically, the procedure gives additional weight to certain points, which causes the errors to be non-normally distributed. The same problem occurs in the 'Initial Values Module' above. Fortunately, these values only need to be approximate enough to be used in the nonlinear regression component, and the results indicate that this method gives good enough values for the nonlinear least squares algorithm to converge.

The last two lines of code perform the nonlinear curve fitting algorithm using the experimental conditions, observed ellipticity, the assumed model, and the estimated parameters. The 'ci' portion computes the confidence intervals from the regression output. This outputs the desired parameters and the confidence intervals in which they lie. To determine the accuracy of these calculations, the experimental data was plotted with the theoretical curve in MATLAB using the algorithm below.

```
X = (min(Temp):(max(Temp)-min(Temp))/99:max(Temp)).';
Y = (min(Den):(max(Den)-min(Den))/99:max(Den)).';
[X,Y] = meshgrid(X,Y);
```

```
Z = zeros(size(X));
for i = 1:100
    for j = 1:100
        Z(i,j) = f(beta, [X(i,j) Y(i,j)]);
    end
end

%plottools is the function to graph stuff; plot the surface X vs. Y vs. Z
%with a scatterplot of the data
```

By plotting the data, it could be made clear if the data was appropriately fitted. That provided a good first test, but the residuals were also analyzed to confirm the fit was valid. The plots are given below.

For five of the peptides (α_4H, $\alpha_4F_3af_3d$, $\alpha_4F_2(6,24)$, $\alpha_4F_2(10,20)$, and $\alpha_4F_2(13,17)$), the above algorithm did not converge. Because they unfolded so easily, the lower base plane could not be accurately estimated. A slight variation of the above algorithm was used that fixed the lower based plane. Running the data for the more stable peptides through this program revealed that the parameters only changed around 10%, and so for comparative purposes was acceptable. The portions with changes are given here for completeness, but there are really only minor computational differences.

```
%Define additional functions K, U, and f
K =  @(b,Cond)exp(-(b(3)-Cond(:,1)*b(4)+b(5)*(Cond(:,1)-T0 ...
    +Cond(:,1).*log(T0./Cond(:,1)))-Cond(:,2)*b(6))./(R*Cond(:,1)));
U = @(b,Cond) arrayfun(@(k) fzero(@(x) x^4+k*x/4-k*P/4, [0 P]), K(b,Cond));
f = @(b,Cond)
1/P*((b(1)*ones(length(Cond(:,1)),1)+b(7).*Cond(:,1)+b(8).*Cond(:,2)).*U(b,Co
nd)...
    +b(2)*(P*ones(length(Cond(:,1)),1)-U(b,Cond)));

%Initial Values Module
beta0 = zeros(8,1);
beta0(1) = max(Theta_Obsd)+1;
beta0(2) = min(Theta_Obsd)-1;
Uest = P*(Theta_Obsd-beta0(2)*ones(length(Theta_Obsd),1))/(beta0(1)-
beta0(2));
Kest = 4*Uest.^4./(P*ones(length(Temp),1)-Uest);
DGest = -R*Temp.*log(Kest);

TestMat = ones(length(Temp),4);
for i = 1:length(Temp)
    TestMat(i,1) = 1;
    TestMat(i,2) = -Temp(i);
```

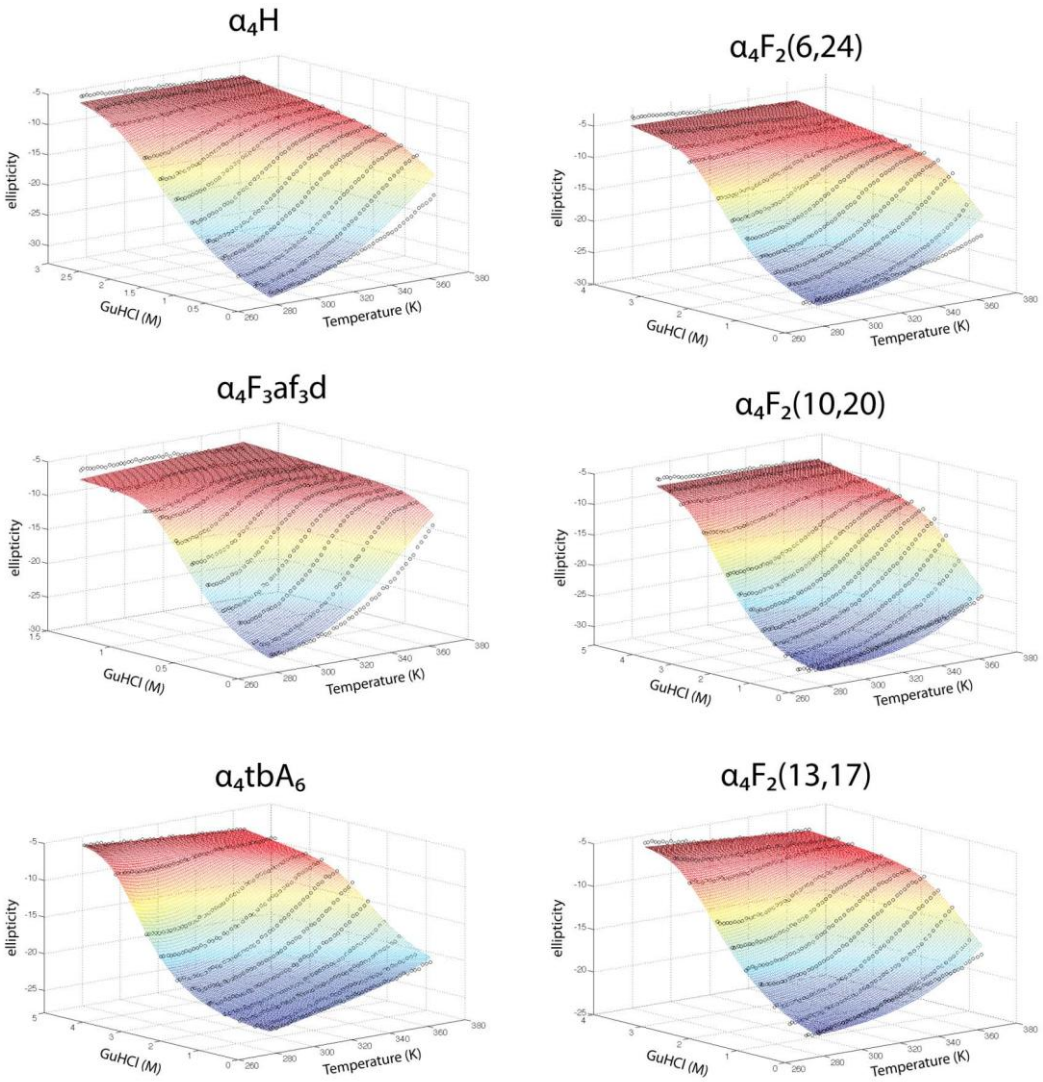
```
    TestMat(i,3) = Temp(i)-T0+Temp(i)*log(T0/Temp(i));
    TestMat(i,4) = -Den(i);
end

ParaEst = linsolve(TestMat,DGest);

beta0 = [max(Theta_Obsd);min(Theta_Obsd); ParaEst(1);
ParaEst(2);ParaEst(3);ParaEst(4);0;0];
%Actual Curve Fitting Portion
[beta, r, J, COVB, mse] = nlinfit(Cond, Theta_Obsd, f, beta0);

ci = nlparci(beta,r,'covar', COVB);
```

The plots of the data (the individual points) and the theoretical fitted results (the surface) are given below in Figure 4. All of the surfaces fit the data very well.



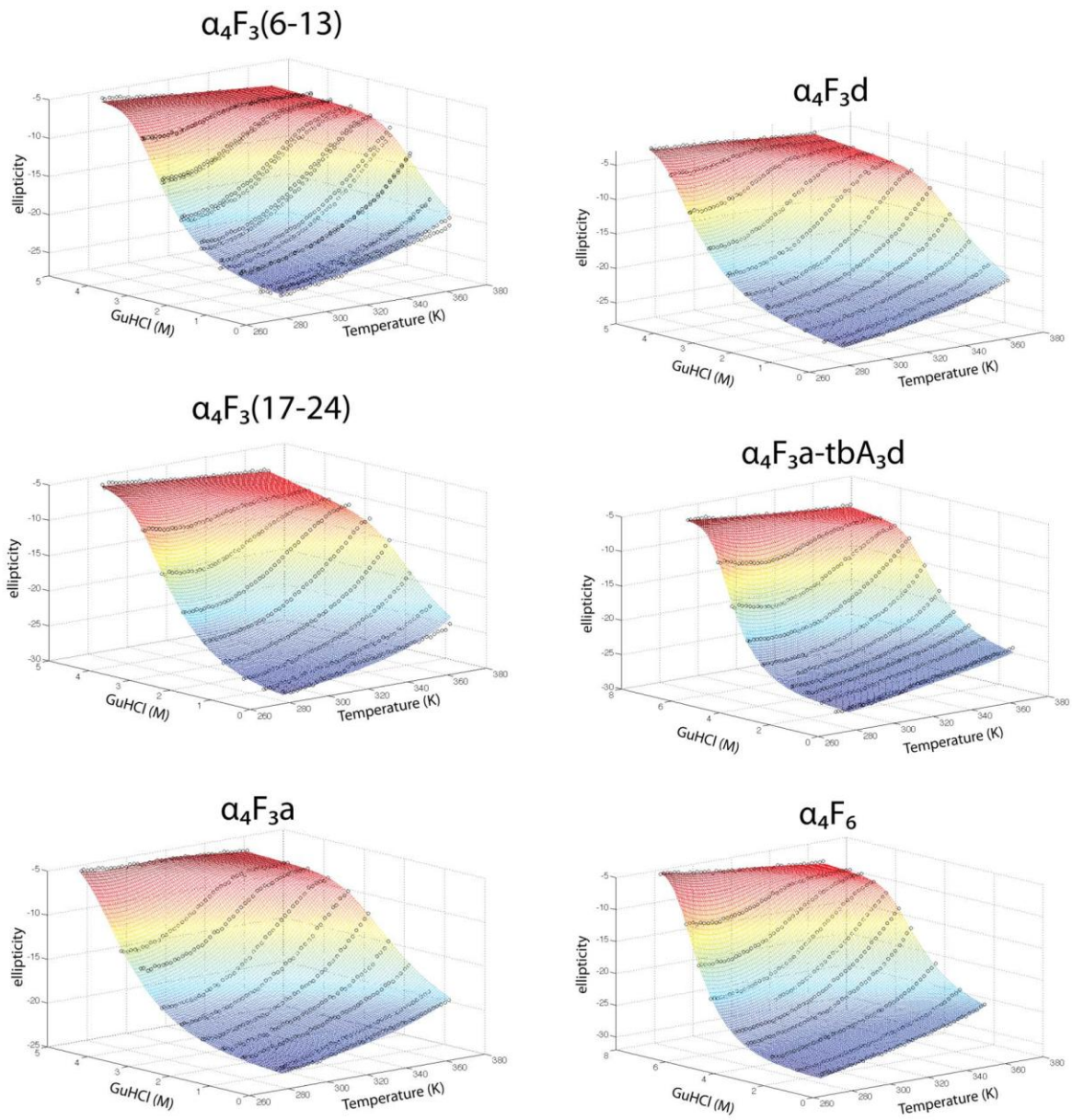


Figure 4: Reprinted with permission from Ref 29. Copywrite 2012 American Chemical Society. These are the plots of the circular dichroism data for each peptide. For each temperature and denaturant concentration the ellipticity was measured at 222 nm and the theoretical surfaces were determined by calculating the thermodynamic parameters from the data.

The surfaces illustrate just how accurate the model is. To compare how the fluorine content affects the peptide stability the actual calculated values must be analyzed. They are given below in Figure 5.

protein	$\Delta H°$ (kcal/mol)	$T\Delta S°$ (kcal/mol)	$\Delta C_p°$ (kcal mol^{-1} K^{-1})	m (kcal mol^{-1} M^{-1})	$\Delta G°_{unfold}$ (kcal/mol)
α_4H	15.87 ± 0.57	−4.69 ± 0.48	0.22 ± 0.02	2.19 ± 0.07	20.56 ± 0.74
$\alpha_4F_3af_3d$	20.70 ± 1.05	−0.13 ± 0.90	0.19 ± 0.03	6.32 ± 0.28	20.83 ± 1.38
$\alpha_4F_2(6,24)$	16.13 ± 0.62	−5.86 ± 0.52	0.29 ± 0.02	2.51 ± 0.18	21.99 ± 0.81
$\alpha_4F_2(10,20)$	14.55 ± 0.48	−8.26 ± 0.41	0.37 ± 0.02	2.15 ± 0.05	22.81 ± 0.63
$\alpha_4F_2(13,17)$	12.54 ± 0.38	−9.44 ± 0.32	0.34 ± 0.02	2.19 ± 0.05	21.98 ± 0.50
α_4tbA_6	13.17 ± 0.28	−12.56 ± 0.21	0.56 ± 0.02	2.35 ± 0.05	25.72 ± 0.35
$\alpha_4F_3(6-13)$	19.56 ± 0.61	−6.43 ± 0.38	0.53 ± 0.03	2.71 ± 0.09	25.99 ± 0.72
$\alpha_4F_3(17-24)$	12.95 ± 0.39	−12.02 ± 0.28	0.60 ± 0.03	2.38 ± 0.08	24.97 ± 0.47
α_4F_3a	19.07 ± 0.60	−7.99 ± 0.38	0.54 ± 0.04	2.29 ± 0.09	27.05 ± 0.71
α_4F_3d	21.37 ± 0.30	−6.99 ± 0.19	0.69 ± 0.02	2.68 ± 0.04	28.36 ± 0.36
$\alpha_4F_3atbA_3d$	12.95 ± 0.32	−17.46 ± 0.27	0.64 ± 0.02	2.45 ± 0.05	30.41 ± 0.41
α_4F_6	18.30 ± 0.28	−13.42 ± 0.17	0.63 ± 0.04	2.21 ± 0.04	31.72 ± 0.33

Figure 5: Reprinted with permission from Ref. 29. Copywrite 2012 American Chemical Society. The calculated thermodynamic parameters for each peptide are presented. Each parameter and 95% confidence interval was computed using the algorithm developed.

Discussion:

The model developed above had generally mild conditions. It required that the peptide exist in equilibrium between two conformations and that the measured ellipticity be the sum of the contribution from each component. Because the peptides fold as α-4-helix bundles, the signal at 222 nm was substantially different and so a precise measure of [U] vs. [F] was possible. In addition, it required that the dependence of the free energy on the denaturant concentration be linear (i.e. $\Delta G = -m * [\mathrm{GuHCl}]$) . This is true for most peptides (*41*, *42*). Because the experiments above were done over such a wide range of temperatures, it could be that the proportionality constant depends on the temperature. Fortunately assuming *m* was constant for each peptide above gave excellent fits, but in general this may not hold. The final major assumption is that the change in heat capacity upon unfolding, $\Delta C_p°$, was constant. This is also typically a valid assumption (*44*). Again, this held for the peptides above as evidenced by the small errors in the fitting (see Figure 5).

Notably, the theory above can be easily generalized beyond tetramer folding. If it is assumed that there are two possible states, a monomer and an *n*-mer with no intermediates, we have the following two equations.

$$\theta = \theta_{\mathrm{u}} * \frac{[\mathrm{U}]}{[\mathrm{P}]} + \theta_{\mathrm{f}} * \frac{[\mathrm{P}] - [\mathrm{U}]}{[\mathrm{P}]} \quad (8)$$

$$K = \frac{[\mathrm{U}]^n}{[F]} = \frac{n[\mathrm{U}]^n}{[\mathrm{P}] - [\mathrm{U}]} \quad (9)$$

The exact same analysis applied to the special case $n = 4$ can be applied here. Taking the derivative of *K* with respect to [U] and observing that it is always positive for $0 < [\mathrm{U}] < [\mathrm{P}]$ proves the existence of an inverse function and so we can find [U] for a given *K* by calculating the unique root of the following polynomial between 0 and [P].

$$0 = n[\mathrm{U}]^n + K[\mathrm{U}] - K[\mathrm{P}] \quad (10)$$

Furthermore, Eqn. 8 can be applied to more than just circular dichroism. Any technique where the measured value differs between the folded and unfolded proteins and is the sum of only those two structures is applicable. Measuring absorbance or using a particular NMR peak, for example, is completely amenable to the procedure above. For large *n* it becomes less likely that there are no intermediate species and this must be verified for Eqn. 8 to be valid however.

The fits were all performed with two independent variables, namely temperature and denaturant concentration. It is entirely possible to do the fitting with just one of the independent variables. This can require less stringent assumptions. In particular, fitting θ vs. [GuHCl] at a constant temperature does not require that *m* be independent of temperature. However, it severely limits the number of data points that can be taken. For the peptides used in these experiments, between 430 and 512 measurements were used in the fitting for each peptide. If we

were restricted to just one temperature, fewer points would be available and so the error from the fit would be substantially higher. In other words, fitting the entire surface gives much smaller errors than just fitting a single curve.

The only remaining concern is a statistical one. Overfitting occurs when regression is done with more terms than necessary or with more complicated relationships than necessary (*45*). In the most general case above 10 different variables are being fitted, so it is a valid concern that there is so much flexibility in the model that regardless of the quality of the data a good fit will always be found. This is because the parameters have too much possible variance. Fortunately, because there were so many data points in each set and the resulting fits had such small confidence intervals for each parameter, it is likely not an issue here. The base planes could have been defined to be constants, higher order polynomials, or any function that empirically gives a good fit. The first was done above for the least stable peptides, while the second has been done in previous work (*41*). Because these did not change the more important thermodynamic parameters by very much, the fits seem very dependent on the thermodynamic data, which is essential for a strong fit. Furthermore, there is strong theoretical backing for the model used. Further simplifications would hurt, not help the quality of the fits.

Circular dichroism was used to study the thermodynamics of these peptides quite deliberately. Differential scanning calorimetry or other arguably more direct techniques were not used for a number of reasons. The most straightforward reason was convenience. A substantial amount of CD data was taken for each peptide in a relatively brief amount of time and the analysis of each data set could be done quickly and automatically using the program developed. In addition to convenience, the tools developed here can be generalized to monomer to *n*-mer folding very easily mathematically and computationally. Plus the experimental protocol used can

be extended to other tools whose output resembles Eqn. 1, including NMR and many absorbance studies. The convenience and generality of the approach made it very appealing to attempt to use here.

The theoretical model described fits the sequence of fluorinated peptides quite well. All 12 peptides in the sequence were well fit by the data and the confidence intervals around each parameter were quite small. As it was previously known that the peptides fit the required assumptions, the quality of the fits is not a surprise. The question remains as to how the thermodynamic parameters of folding were affected by fluorination. The most basic question is what the enthalpic and entropic contributions are to unfolding at room temperature. These are plotted in Figure 6. A glance reveals that the enthalpic component does not follow a significant trend as the free energy of unfolding increases. Thus, it seems that fluorination does not affect the enthalpy nearly as much as it affects the entropy.

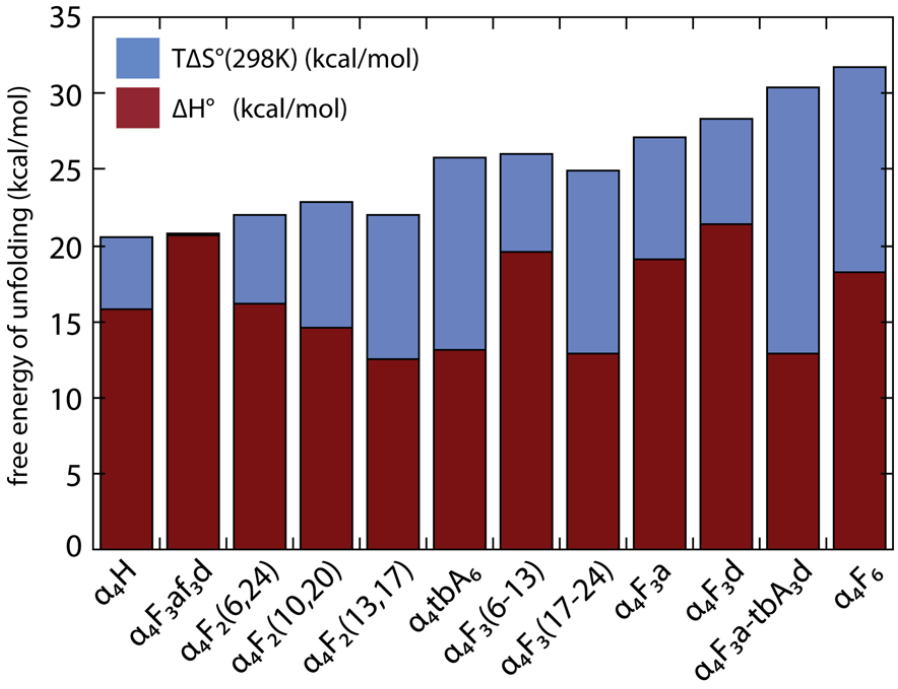


Figure 6: Reprinted with permission from Ref. 29. Copywrite 2012 American Chemical Society. For each peptide, the free energy of unfolding at room temperature is given. This is further broken down into enthalpic and entropic contributions.

Fluorination appears to affect the entropy of unfolding much more so than the enthalpy, but this trend can be made more precise. The change in solvent accessible surface area upon unfolding (ΔASA) will be used to do just that. It is well established that the change in ASA has a substantial effect on protein folding (*46*). If polar side chains are being solvated upon unfolding, then it should be favorable, while if nonpolar groups are leaving a cavity, it should be unfavorable. From examining Figure 1, when the tetrameric bundles unfold there is an increase in the apolar solvent accessible surface area because as the peptides unfold the hydrophobic or fluorinated residues leave the solvent-excluded cavity and enter the aqueous phase. The ΔASA was determined for the above peptides previously (*47, 48, 49*). Informally, a sphere about the size of a water molecule is "rolled" over the structures of the folded and denatured peptides and the total area it can cover is roughly the solvent accessible surface area. The resulting ΔASA was then compared with the thermodynamic parameters (see Figure 7 on the following page).

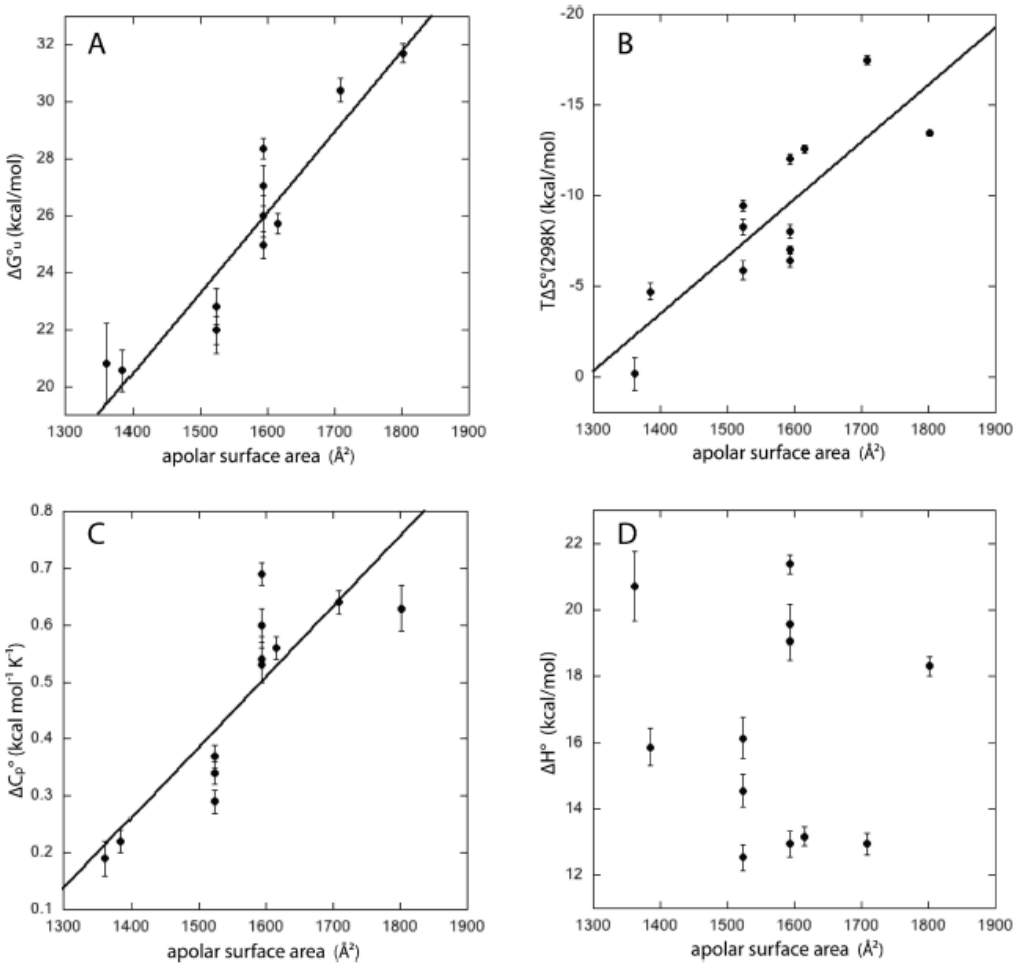


Figure 7: Reprinted with permission from Ref. 29. Copywrite 2012 American Chemical Society. The change in apolar surface area upon unfolding was determined for each of the 12 peptides. These were then compared with the various thermodynamic parameters. The changes in free energy, entropy, and heat capacity are strongly correlated with ΔASA, while the enthalpy change is uncorrelated.

The existence of and lack of correlations illustrated in Figure 7 demonstrate how fluorination affects these peptides. The plot of Δ*G*° vs. ASA was as expected. The greater the number of hydrophobic residues introduced into an aqueous environment upon unfolding the harder it will be to denature the peptide, which is confirmed by the strong correlation in Fig. 7A. Notably the correlation is almost as strong in the plot of Δ*S*° vs. ASA in Fig. 7B, while the graph of Δ*H*° vs. ASA in Fig. 7D has nearly no correlation. Thus the changes between the peptides' stabilities that arise from fluorination are of an entropic nature and not an enthalpic one. The

entropic contribution to protein folding arises due to the hydrophobic effect (*1*). In the folded structure the peptides are very compact and the surrounding water molecules have many possible states in which to maximize their interactions with one another. Upon unfolding, the peptides take up more space and the water molecules must reorient themselves to maximize the interactions with themselves and limit those with the hydrophobic groups. The process is not dependent on anything other than fluorine's size.

The lack of a correlation between the enthalpic contribution and the stability of the folded peptide contrasts with earlier results (*50*). It has been observed that fluorine interacts favorably with other groups due to its high electronegativity and small size, positioning itself in electropositive regions to provide an enthalpic contribution to stability, while moving out of electron dense areas to minimize unfavorable contacts (*50*). On that basis, it might be expected to cause a decrease in stability in the most highly fluorinated peptides, as those would have the greatest amount of fluorine-fluorine interactions. Instead, there seemed to be no correlation of fluorination with enthalpy. This seems to imply that fluorine is simply functioning as a hydrophobic group larger than the hydrogen it replaces. There is no evidence of any "fluorous effect" or any noncovalent interactions, either favorable or unfavorable, in the packed hydrophobic core of the peptides.

The final parameter to be examined is the change in heat capacity. The fact that $\Delta C_p°$ and ΔASA are highly and positively correlated is not surprising (*44*). The change in heat capacity is known to be positive for the solvation of hydrophobic residues and negative for hydrophilic residues (*44*). As the α_4-helix peptides considered have no increased solvated hydrophilic content but a greater amount of solvated hydrophobic content upon unfolding, this agrees with previous results and again illustrates how these fluorous residues are hydrophobic in nature.

Ultimately, in the α_4-helix bundles described fluorine's stabilizing effect arises from it being slightly bigger than the hydrogen it is being substituted for. As a control, the noncanonical amino acid β-*tert*-butylalanine was incorporated into some of the peptides (see Figure 3). That residue is slightly larger than leucine, but peptides containing it fit the exact same trends as the fluorinated ones, despite β-*tert*-butylalanine not containing any fluorine. It seems the ΔASA is the correlating factor and not any properties specific to fluorine. Fluorine is often considered to be isosteric with hydrogen, while in reality it is significantly larger (*18*, *50*). In this case it was just large enough to increase the ΔASA without disrupting the packing structure in the hydrophobic cavity. Using these ideas, fluorine could provide a nice tool in protein design. It can safely be substituted into proteins for hydrogen in hydrophobic environments to increase the stability with the caveat that it cannot be too large and cause steric clashes.

The lack of enthalpic changes due to fluorine in these peptides has significant implications. On one hand it is convenient for the protein designer, as abnormal interactions involving fluorine do not have to be considered. On the other, fluorine is known to have unusual properties and there should be some way to utilize them. The obvious question is why we did not observe any of those effects in the α_4-helix peptides considered. The fluorous effect, illustrated by the three phase solution of hexane, water, and perfluorohexane, occurs when there is a substantial amount of fluorine in the molecules. The peptides considered may not have had enough fluorine to generate this effect. If a peptide could be designed that incorporated even more fluorine, the effect might be more pronounced. The construction of such a peptide has its difficulties though. It may have to have a larger hydrophobic cavity to accommodate all of the fluorine atoms, but if it were too large the fluorous effect would be decreased. It is unknown if it is even possible to get the density of fluorine atoms high enough in the peptide interior to

actually observe this. Furthermore, it would be challenging to compare it to the nonfluorinated version, as both size and fluorine content would have to be controlled for. The other effect discussed above, the Gauche Effect, occurs between fluorine atoms that are three-bond neighbors. Its exploitation would require a much more detailed view of the protein interior, but it is plausible that the favorability of the $\sigma^*_{C\text{-}F}$ and $\sigma_{C\text{-}H}$ interaction leads to alternative packing features. This was not that the case here, as ΔASA accounted for the change in stability, but it could conceivably be incorporated into future designs. Furthermore, as it does not substantially perturb the structure, fluorine's place in protein design may be in catalysis. Fluorine is extremely electronegative and will alter the pK_a and chemical reactivity of nearby groups greatly. In the examples above fluorine's effect on reactivity could obviously not be considered, but incorporating it into or nearby the active site of an enzyme could lead to catalytic differences stemming from its electronegativity.

Speculation aside, the experiments above indicate that the effects of fluorine on stability are strictly due to fluorine's size. A series of fluorinated peptide analogs was synthesized and the thermodynamics of their folding was analyzed using circular dichroism. The peptides were known to form α_4-helix bundles from prior work, and so an algorithm to compute the thermodynamic parameters was designed. It was formally proven to be numerically exact and so a MATLAB script was written to perform these calculations. The results indicated that fluorine's role in peptide stability is primarily due to its size. The correlations between the change in solvent accessible surface area and the changes in free energy, entropy, and heat capacity of unfolding were quite strong, while the lack of a correlation between ΔASA and the enthalpy change led to the conclusion that the effects of fluorine on these peptides stem from size alone and not from any special fluorous interactions. Future work should be performed to determine

what conditions are necessary, if any, for fluorine to exhibit any of the unusual properties it is known for in peptides. Furthermore, the fitting algorithm developed is quite general and can be applied to any *n*-mer folding that can be modeled with a two-state equilibrium.

Acknowledgments:

I wish to thank Professor E. Neil G. Marsh for being my research and concentration advisor over the past three years. I have thoroughly enjoyed working for him and feel his instruction and knowledge have been crucial to this thesis and to my success. I also wish to thank Dr. Benjamin C. Buer for both his work on this project and for his mentorship. Finally, I would like to thank all of the other members of the Marsh Lab for their support and guidance.

References:

1. Voet, D., and Voet, J. G. (2011) *Biochemistry*, 4th ed., John Wiley & Sons, Inc. Hoboken, New Jersey.
2. Watson, J. D. (1976) *Molecular Biology of the Gene*, 3rd ed. W. A. Benjamin, Inc., Menlo Park, California.
3. Vickery, H. B, and Schmidt, C. L. A. (1931) The History of the Discovery of the Amino Acids, *Chem. Rev. 9*, 169-318.
4. Vickery, H. B. (1972) The History of the Discovery of the Amino Acids. A Review of Amino Acids Discovered Since 1931 as Components of Native Proteins, *Adv. Protein Chem. 26*, 81-171.
5. Anfinsen, C. B. (1973) Principles that Govern the Folding of Protein Chains, *Science 181*, 223-230.
6. Fischer, E. (1891) Ueber die Configuration des Traubenzuckers und seiner Isomeren, *Ber. 24*, 1836-1845.
7. Kauzmann, W. (1959) Some Factors in the Interpretation of Protein Denaturation, *Adv. Protein Chem. 14*, 1-63.
8. Franks, H. S., and Evans, M. W. (1945) Free Volume and Entropy in Condensed Systems III. Entropy in Binary Liquid Mixtures; Partial Molal Entropy in Dilute Solutions; Structure and Thermodynamics in Aqueous Electrolytes, *J. Chem. Phys. 13*, 507-532.
9. Baldwin, R. L. (1989) How does protein folding get started? *Trends Biochem. Sci. 14*, 291-294.
10. Ramachandran, G. N., and Sasisekharan, V. (1968) Conformation of Polypeptides and Proteins, *Adv. Protein Chem. 23*, 283-437.
11. Scholtz, J. M., and Baldwin, R. L. (1992) The Mechanism of α-Helix Formation by Peptides, *Annu. Rev. Biophys. Biomol. Struct. 21*, 95-118.
12. Tanford, C. (1968) Protein Denaturation, *Adv. Protein Chem. 23*, 121-282.
13. Zasloff, M. (2002) Antimicrobial peptides of multicellular organisms, *Nature 415*, 389-395.
14. Hancock, R. E. W., and Sahl, H.-G. (2006) Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies, *Nat. Biotechnol. 24*, 1551-1557.
15. Matsuzaki, K. (1999) Why and how are peptide-lipid interactions utilized for self-defense? Magainins and tachyplesins as archetypes, *Biochim. Biophys. Acta 1462*, 1-20.
16. Ge, Y., MacDonald, D. L., Holroyd, K. J., Thornsberry, C., Wexler, H., and Zasloff, M. (1999) In vitro Antimicrobial Properties of Pexiganan, an Analog of Magainin, *Antimicrob. Agents Chemother. 43*, 782-788.
17. Akçay, G., and Kumar, K. (2009) A new paradigm for protein design and biological self-assembly, *J. Fluorine Chem. 130*, 1178-1182.
18. Marsh, E. N. G. (2000) Towards the nonstick egg: designing fluorous proteins, *Chem. Biol. 7*, R153-R157.

19. Scott, R. L. (1948) The Solubility of Fluorocarbons, *J. Am. Chem. Soc. 70*, 4090-4093.

20. Wiberg, K. B., Murcko, M. A., Laidig, K. E., and MacDougall, P. J. (1990) Origin of the "Gauche Effect" in Substituted Ethanes and Ethenes, *J. Phys. Chem. 94*, 6956-6959.

21. O'Hagan, D. (2008) Understanding organofluorine chemistry. An introduction to the C-F bond, *Chem. Soc. Rev. 37*, 308-319.

22. Rowland, R. S., and Taylor, R. (1996) Intermolecular Nonbonded Contact Distances in Organic Crystal Structures: Comparison with Distances Expected from van der Waals Radii, *J. Phys. Chem. 100*, 7384-7391.

23. Gerig, J. T. (1994) Fluorine NMR of Proteins, *Prog. Nucl. Magn. Reson. Spec. 26*, 293-370.

24. O'Hagan, D., and Harper, D. B. (1999) Fluorine-containing natural products, *J. Fluorine Chem.100*, 127-133.

25. Liu, D. R., and Schultz, P. G. (1999) Progress toward the evolution of an organism with an expanded genetic code, *Proc. Natl. Acad. Sci. USA 96*, 4780-4785.

26. Buer, B. C., Levin, B. J., and Marsh, E. N. G. (2013) Perfluoro-*tert*-butyl-homoserine as a sensitive ^{19}F NMR reporter for peptide-membrane interactions in solution, *J. Pept. Sci.19*, 308-314.

27. Tang, Y., Ghirlanda, G., Petka, W. A., Nakajima, T., DeGrado, W. F., and Tirrell, D. A. (2001) Fluorinated Coiled-Coil Proteins Prepared In Vivo Display Enhanced Thermal and Chemical Stability, *Angew. Chem. Int. Ed.40*, 1494-1496.

28. Koksch, B., Sewald, N., Hofmann, H.-J., Burger, K., and Jakubke, H.-D. (1997) *J. Pept. Sci. 3*, 157-167.

29. Buer, B. C., Levin, B. J., and Marsh, E. N. G. (2012) Influence of Fluorination on the Thermodynamics of Protein Folding, *J. Am. Chem. Soc. 134*, 13027-13034.

30. Lumry, R., and Rajendeer, S.(1970) Enthalpy-Entropy Compensation Phenomena in Water Solutions of Proteins and Small Molecules: A Ubiquitous Property of Water, *Biopolymers 9*, 1125-1227.

31. Lee, K.-H., Lee, H.-Y., Slutsky, M. M., Anderson, J. T., and Marsh, E. N. G. (2004) Fluorous Effect in Proteins: *De Novo* Design and Characterization of a Four-α-Helix Bundle Protein Containing Hexafluoroleucine, *Biochemistry 43*, 16277-16284.

32. Lee, K.-H., Lee, H.-Y., Al-Hashimi, H. M., and Marsh, E. N. G. (2006) Modulating Protein Structure with Fluorous Amino Acids: Increased Stability and Native-like Structure Conferred on a 4-Helix Bundle Protein by Hexafluoroleucine, *J. Am. Chem. Soc. 128*, 337-343.

33. Gottler, L. M., Salud-Bea, R., and Marsh, E. N. G. (2008) The Fluorous Effect in Proteins: Properties of α_4F_6, a 4-α-Helix Bundle Protein with a Fluorocarbon Core, *Biochemistry 47*, 4484-4490.

34. Buer, B. C., Salud-Bea, R., Al Hashimi, H. M., and Marsh, E. N. G. (2009) Engineering Protein Stability and Specificity Using Fluorous Amino Acids: The Importance of Packing Effects, *Biochemistry 48*, 10810-10817.

35. Greenfield, N. J. (2006) Using circular dichroism spectra to estimate protein secondary structure, *Nat. Protoc. 1*, 2876-2890.

36. Tsushima, T., Kawada, K., Ishihara, S., Uchida, N., Shiratori, O., Higaki, J., Hirata, M. (1988) Fluorine-Containing Amino Acids and their Derivatives. 7.[1] Synthesis and Antitumor Activity of α- and γ-Substituted Methotrexate Analogs, *Tetrahedron 44*, 5375-5387.

37. Anderson, J. T., Toogood, P. L., and Marsh, E. N. G. (2002) A Short and Efficient Synthesis of L-5,5,5,5',5',5'-hexafluoroleucine from *N*-Cbz-L-Serine, *Org. Lett. 4*, 4281-4283.

38. Schnölzer, M., Alewood, P., Jones, A., Alewood, D., and Kent, S. B. H. (1992) *In situ* neutralization in Boc-chemistry solid phase peptide synthesis, *Int. J. Peptide Protein Res. 40*, 180-193.

39. Artin, M. (1991) *Algebra*, 1st ed., Pearson. Upper Saddle River, New Jersey.

40. Spivak, M. (2006) *Calculus*, 3rd ed., Cambridge University Press. Cambridge, United Kingdom.

41. 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 H_2O and D_2O. Determination of the thermodynamic parameters, $\Delta H°$, $\Delta S°$, and $\Delta C_p°$, and evalution of solvent isotope effects, *Protein Sci. 7*, 2405-2412.

42. Yi, Q., Scalley, M. L., Simons, K. T., Gladwin, S. T., and Baker, D. (1997) Characterization of the free energy spectrum of peptostreptococcal protein L, *Folding Design 2*, 271-280.

43. Wilkinson, G. N. (1961) Statistical Estimations in Enzyme Kinetics, *Biochem. J. 80*, 324-332.

44. Prabhu, N. V., and Sharp, K. A. (2005) Heat Capacity in Proteins, *Annu. Rev. Phys. Chem.56*, 521-548.

45. Hawkins, D. M. (2004) The Problem of Overfitting, *J. Chem. Inf. Comput. Sci. 44*, 1-12.

46. Richards, F. M. (1977) Areas, Volumes, Packing, and Protein Structure, *Annu. Rev. Biophys. Bioeng.6*, 151-176.

47. Buer, B. C., Meagher, J. L., Stuckey, J. A., and Marsh, E. N. G. (2012) Structural basis for the enhanced stability of highly fluorinated proteins, *Proc. Natl. Acad. Sci. 109*, 4810-4815.

48. Buer, B. C., Meagher, J. L., Stuckey, J. A., and Marsh, E. N. G. (2012) Comparison of the structures and stabilities of coiled-coil proteins containing hexafluoroleucine and *t*-butylalanine provides insight into the stabilizing effects of highly fluorinated amino acid side-chains, *Protein Sci. 21*, 1705-1715.

49. Sanner, M. F., Olson, A. J., and Spehner, J.-C. (1996) Reduced Surface: An Efficient Way to Compute Molecular Surfaces, *Biopolymers 38*, 305-320.

50. Müller, K., Faeh, C., and Diederich, F. (2007) Fluorine in Pharmaceuticals: Looking Beyond Intuition, *Science 28*, 1881-1886.