Chapter 4

Structural Basis for the Enhanced Stability of Highly Fluorinated Proteins

4.1 – Introduction

Parts of the work described in this chapter have been published as: Structural Basis for the Enhanced Stability of Highly Fluorinated Proteins. Buer BC, Meagher JL, Stuckey JA & Marsh ENG (2012) *Proceedings of the National Academy of Sciences* 109(13):4810-4815. Co-authors were very helpful in aiding me with X-ray crystallography experiments. Dr. Jennifer Meagher set up 96-well robotic screens of α_4 proteins and trained me to optimize using 24-well grid screens. Training and guidance for data collection and structure refinement came from Prof. Jeanne Stuckey and Dr. Jennifer Meagher.

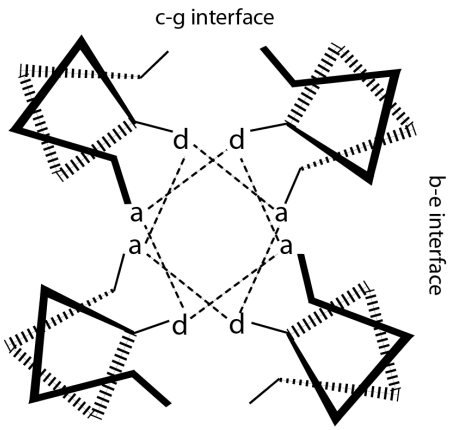
Inspired by the novel properties of fluorocarbons, there have been numerous studies aimed at using extensively fluorinated (or fluorous) amino acids to modulate the properties of proteins, in particular to increase their thermal stability [1,2]. Thus fluorous analogs of hydrophobic amino acids such as leucine, valine and phenylalanine have been incorporated into both natural and *de novo*-designed proteins, either biosynthetically or by chemical synthesis, as detailed in Chapter 1 [3-7]. Proteins with sequences containing up to ~25% fluorous residues have been synthesized without gross structural perturbation. In almost all cases fluorination significantly enhances stability towards thermal

unfolding, chemical denaturation and proteolytic degradation, with minimal impact on the biological activity of the protein or peptide [3,5-19].

Despite the numerous studies on extensively fluorinated proteins and peptides, the origin of their enhanced stability, whether through favorable fluorocarbon-fluorocarbon interactions or simple differences in hydrophobicity, remains a matter of debate. As discussed in Chapter 3, two crystal structures of cVHP have been reported by the Gellman group that contain single Phe to pFPhe substitutions. To date, no structures of proteins containing numerous highly fluorinated amino acids have been reported, which severely hinders our understanding of how interactions between fluorocarbon side-chains within the core of the protein contribute to the dramatic changes in stability observed for many proteins.

		abcdefg	abcdefg	abcdefg	a
α_4H	Ac–GN	ADELYKE	LEDLQER	LRKLRKK	LRSG–NH_2
α_4Ht	Ac–GN	ADE**X**YKE	**X**ED**X**QER	**X**RK**X**RKK	**X**RSG–NH_2
α_4F_3a	Ac–GN	ADELYKE	**X**EDLQER	**X**RKLRKK	**X**RSG–NH_2
α_4F_3d	Ac–GN	ADE**X**YKE	LED**X**QER	LRK**X**RKK	LRSG–NH_2
α_4F_3(6-13)	Ac–GN	ADE**X**YKE	**X**ED**X**QER	LRKLRKK	LRSG–NH_2
$\alpha_4F_3af_3d$	Ac–GN	ADE**X**YKE	**X**ED**X**QER	**X**RK**X**RKK	**X**RSG–NH_2

X = tBAla
X = hFLeu
X = tFeG



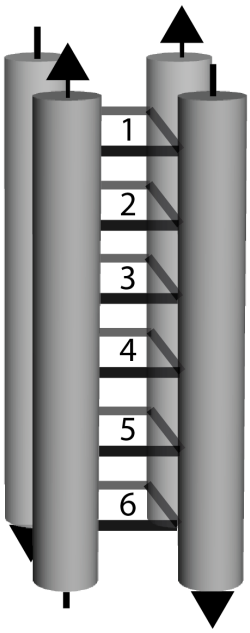


Figure 4.1. The sequences and helical wheel diagram for the α_4 proteins, illustrating positions of the hydrophobic **a** and **d** residues in the anti-parallel 4-helix bundle topology. The hydrophobic core of these proteins comprises 6 layers formed by **a** and **d** residues as illustrated in the diagram on the right.

In this chapter, I describe high resolution X-ray structures for six *de novo*-designed proteins, α_4H [5], α_4Ht, α_4F_3a [20], α_4F_3d [20], $\alpha_4F_3(6\text{-}13)$ and $\alpha_4F_3af_3d$. These proteins are designed to form anti-parallel 4-α-helix bundles in which the hydrophobic core is packed in six layers by residues at the canonical **a** and **d** positions of the helical repeat, as illustrated in Figure 4.1. The design of these proteins has been described previously in Chapters 2 and 3. In α_4H, the hydrophobic core contains leucine at each **a** and **d** position, whereas in α_4F_3a, the leucine residues at the three **a** positions are substituted for hFLeu and in α_4F_3d, the leucine residues at the three **d** positions are substituted for hFLeu, so that 50% of the core is now fluorocarbon. For $\alpha_4F_3(6\text{-}13)$, the 50% fluorocarbon core is retained, but packing order is disrupted with two **d** (6 and 13) positions and one **a** (10) position being hFLeu. $\alpha_4F_3af_3d$ has an all-fluorocarbon core with hFLeu in **a** positions and tFeG in **d** positions. α_4Ht contains tBAla at all **a** and **d** positions.

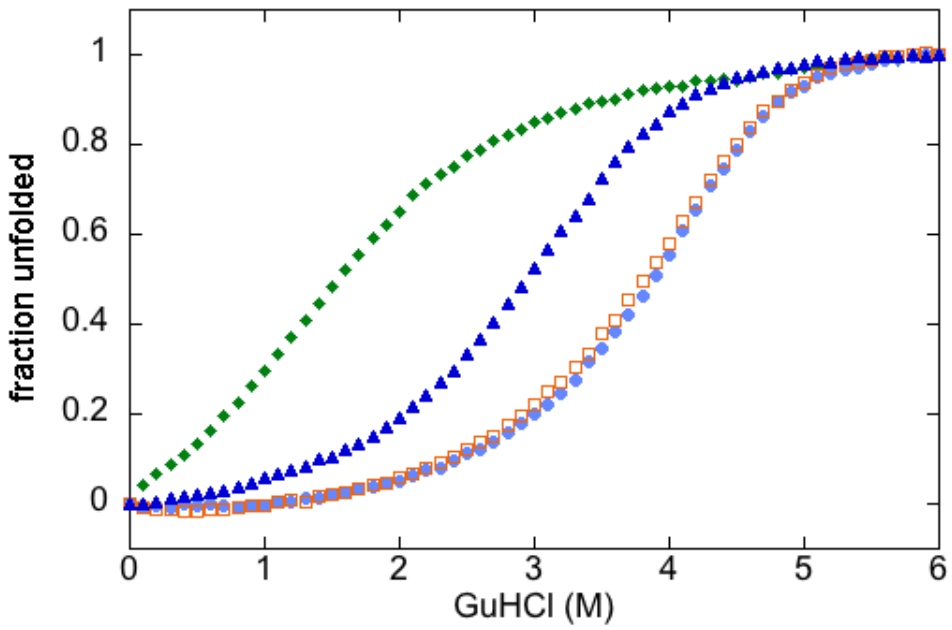


Figure 4.2. Guanidine hydrochloride-induced unfolding curves for α_4H (green), α_4F_3(6-13) (blue), α_4F_3a (light blue) and α_4F_3d (orange). Unfolding was monitored by following changes in ellipticity at 222 nm. Free energies of folding were calculated using these unfolding curves as described in section 2.5.

hFLeu dramatically increases the stability for α_4F_3a (-27.6 ± 0.1 kcal/mol and *m* = -2.47 kcal/mol/M) and α_4F_3d (-26.6 ± 0.1 kcal/mol and *m* = -2.24 kcal/mol/M) compared to α_4H ($\Delta G°_{fold}$ = -18.0 ± 0.2 kcal/mol and *m* = -1.04 kcal/mol/M). This stability increase corresponds to approximately 0.8 kcal/mol/hFLeu residue, with increased hydrophobicity of 24 methyl to trifluoromethyl substitutions being primarily responsible and the efficiency of alternating Leu and hFLeu packing partially responsible for increased stability. To examine the contribution of Leu–hFLeu knobs-into-holes packing on stability, the proteins α_4F_3(6-13) and α_4F_3(17-24) were designed. Although they contain the same number of hFLeu residues, these proteins provide a different packing arrangement by incorporating hFLeu into one **a** and two **d** positions, α_4F_3(6-13), or two **a** and one **d** positions, α_4F_3(17-24). A crystal structure of α_4F_3(6-13) was obtained and it

was determined that differences in packing arrangement lead to loss in stability even though the number of Leu and hFLeu residues remains constant.

A careful comparison of the structures of α_4H and α_4F_3a allowed the design and structural characterization of the protein, $\alpha_4F_3af_3d$, to test the hypothesis that changes in buried hydrophobic surface area, rather than favorable interactions between fluorinated residues, are responsible for the increased stability imparted by fluorination. $\alpha_4F_3af_3d$ contains smaller trifluoroethylglycine residues (tFeG) at the **d** positions, which compensates for the larger hFLeu residues at **a** positions. The crystal structure of this protein was obtained and shows that despite containing 36 trifluoromethyl groups in the core, it is actually slightly less stable than α_4H.

Similarly, synthesis of α_4Ht was motivated by a desire to study a variant of α_4 in which hydrophobic surface area is increased relative to Leu without incorporating fluorine. α_4Ht incorporates β-t-butyl-L-alanine (tBAla) residues at all **a** and **d** positions, so that the entire hydrophobic core increases by 24 methyl groups compared to α_4H. The resultant increase in hydrophobic surface area closely matches that of α_4F_3a, α_4F_3d and $\alpha_4F_3(6\text{-}13)$, this provides a useful protein to investigate whether increasing conventional hydrophobic volume and surface area could contribute to protein stability as effectively as fluorination. The unique shape of the little-studied tBAla residue also presents unique possibilities to protein design as α_4Ht forms a stable, antiparallel, 4-helix bundle with a central hydrophobic void.

4.2 – Experimental Procedures

4.2.1 - Materials and Peptide Synthesis

L-5,5,5,5′,5′,5′-hexafluoroleucine was synthesized as described previously [21] and converted to Boc-protected derivative by procedures described in Chapter 2. 4,4,4-trifluoroethylglycine was purchased from SynQuest Laboratory and enzymatically resolved as described in Chapter 6 [22]. Boc- and Fmoc-protected β-t-butyl-L-alanine were purchased from AnaSpec Inc. Peptides were synthesized by manual Fmoc procedures (α_4H and α_4Ht) or manual Boc procedures (α_4F_3a, α_4F_3d, α_4F_3(6-13) and $\alpha_4F_3af_3$d) as detailed in Chapter 2. All peptides were purified via Waters preparatory RP-HPLC using a linear gradient containing 0.1% TFA with a flow rate of 10 mL/min. Peptide identity was confirmed using MALDI-MS with a matrix of α-cyano-4-hydroxycinnamic acid.

4.2.2 - Crystallization

Peptides were dissolved in 10 mM Tris buffer (pH 7.0) to a concentration of 6 mM as determined by absorbance at 280 nm. Crystals were grown by vapor diffusion at 20 °C in a hanging drop with 2 μl peptide and 2 μl precipitant containing 100 mM CHES buffer (pH 9.0) and 48 % PEG 400 for α_4H, 100 mM Tris buffer (pH 7.8) and 55 % PEG 400 for α_4F_3a, 100 mM Tris buffer (pH 8.5) and 48 % PEG 600 for $\alpha_4F_3af_3$d, 100 mM CHES buffer (pH 9.0) and 48 % PEG 400 for α_4Ht, 100 mM Tris buffer (pH 7.8) and 55 % PEG 400 for α_4F_3d, and 100 mM Tris buffer (pH 8.5) and 48 % PEG 600 for α_4F_3(6-13). Crystals were flash-cooled with liquid N_2 in their mother liquor for data collection.

4.2.3 - Data Collection and Refinement

Data was collected at the Advanced Photon Source (APS) (LS-CAT Beamlines 21-F and 21-G) at the Argonne National Laboratory and were collected on a MarCCD

(Mar USA, Evanston, IL) at wavelengths of 0.97872 Å and 0.97857 Å, respectively, at -180 °C. Data was processed and scaled with HKL2000[23]. The peptides α_4H, α_4F_3a and α_4F_3(6-13) crystallized in space group $I4_1$ while the peptides α_4Ht, α_4F_3d and $\alpha_4F_3af_3$d crystallized in space group $P2_12_12$. All crystals contain a dimer in the asymmetric unit.

Phases were initially determined by molecular replacement using Phaser in the CCP4i suite of programs [24]. The search model for α_4H was a helical monomer of 27 alanine residues based upon the anti-parallel structure of the 4-helix bundle E20S (PDB code: 2CCF) [25] built in Coot. For α_4F_3a, α_4Ht, α_4F_3(6-13) and $\alpha_4F_3af_3$d a monomer of α_4H was used as a starting model with Leu 10, 17 and 24 mutated to hFLeu and all other side-chains truncated to Ala. For α_4F_3d, α_4F_3(6-13) was used as a starting model with hFLeu10 mutated to Leu. The sequence register of α_4H was determined using automated protein model building of the ARP/wARP [26] web service. The PRODRG web server was used to generate coordinates and restraint parameters for hFLeu, tFeG, tBAla and non water solvent molecules [27].

Peptide models were refined by rigid body refinement and restrained refinement using Buster [28]. Side-chains were built using Coot [29] with $2F_o - F_c$ and $F_o - F_c$ electron density maps from Buster. Data refinement and statistics are given in Table 4.1. All residues from the six structures are in the allowed regions of the Ramanchandran plot. Structures were validated with Molprobity [30], Parvarti [31] and whatcheck [32]. Areas of poor electron density were not modeled. These include α_4H residue 27 of chain A and 1 and 27 of chain B; α_4F_3a residues 26 and 27 of chain A and 27 of chain B; $\alpha_4F_3af_3$d residues 1-4 and 27 of chain A and 26 and 27 of chain B; α_4Ht residue 27 of chain A and 1 and 27 of

chain B; α_4F_3d residues 26 and 27 of chain A and 27 of chain B; α_4F_3(6-13) residues 1-4 and 27 of chain A and 26 and 27 of chain B.

4.2.4 - Structure Analysis

Protein models were generated and hydrogens added using PyMOL. Protein volumes and surface areas were analyzed using MSMS [33] in Chimera with a probe radius of 1.4 Å corresponding to a water molecule and a vertex density of 10. The packing arrangement of the hydrophobic core of α_4H was analyzed by SOCKET [34]

Data set	α_4H	α_4F_3a	α_4F_3d
Space group	$I4_1$	$I4_1$	$P2_12_12$
Unit cell	$a = b = 49.04$; $c = 41.23$ $\alpha = \beta = \gamma = 90$	$a = b = 48.35$; $c = 39.75$ $\alpha = \beta = \gamma = 90$	$a = 30.82$; $b = 39.25$; $c = 41.23$ $\alpha = \beta = \gamma = 90$
Wavelength, Å	0.97872	0.97872	0.97872
d_{min}, Å	1.36 (1.36-1.38)	1.54 (1.54-1.57)	1.19 (1.19-1.21)
R_{sym}, %	4.7 (25.8)	5.3 (27.6)	5.7 (15.5)
$\langle I/\sigma_I \rangle$	20 (5)	20 (5)	20 (10)
Completeness, %	99.1 (99.6)	98.8 (100.0)	98.1 (76.2)
Redundancy	7.3 (6.6)	7.0 (6.5)	7.9 (7.6)
Refinement statistics			
Data range, Å	10.0-1.36	34.19-1.54	11.16-1.19
R-factor, %	19.7	18.6	17.8
R_{free}, %	25.5	20.9	18.7
Protein atoms, #	439	457	517
Water molecules, #	41	19	49
Reflections, #	10,467	6,778	16,392
rmsd			
Bonds (Å)	0.010	0.009	0.009
Angles (°)	0.96	1.05	1.08

Data set	α_4F_3(6-13)	$\alpha_4F_3af_3d$	α_4Ht
Space group	$I4_1$	$P2_12_12$	$P2_12_12$
Unit cell	$a = b = 49.58$; $c = 41.57$ $\alpha = \beta = \gamma = 90$	$a = 30.96$; $b = 36.36$; $c = 41.46$ $\alpha = \beta = \gamma = 90$	$a = 31.31$; $b = 37.42$; $c = 40.76$ $\alpha = \beta = \gamma = 90$
Wavelength, Å	0.97856	0.97872	0.97872
d_{min}, Å	1.48 (1.48-1.51)	1.72 (1.72-1.75)	1.54 (1.54-1.57)
R_{sym}, %	4.4 (46.6)	4.6 (52.3)	4.5 (52.2)
$\langle I/\sigma_I \rangle$	20 (3)	20 (3)	20 (3)
Completeness, %	99.9 (100.0)	99.3 (100.0)	98.2 (93.9)
Redundancy	11.0 (10.9)	10.3 (10.7)	10.2 (9.3)
Refinement statistics			
Data range, Å	35.06-1.48	9.08-1.72	27.56-1.54
R-factor, %	25.5	24.1	24.2
R_{free}, %	30.2	29.0	24.5
Protein atoms, #	487	459	431
Water molecules, #	26	19	55
Reflections, #	8,484	5,229	7,348
rmsd			
Bonds (Å)	0.009	0.009	0.008
Angles (°)	1.14	1.17	0.96

Table 4.1. Data collection and refinement statistics

4.2.5 - Circular Dichroism

CD spectra of peptides were recorded with an Aviv 62DS spectropolarimeter at 25 °C. To examine the unfolding of the peptide by GuHCl, stock solutions were prepared containing 40 μM peptide (concentration of monomer) in 10 mM potassium phosphate buffer, pH 7.0, both with and without 8.0 M GuHCl. An auto-titrator was used to mix the two solutions to incrementally increase the concentration of GuHCl in the sample CD cuvette (pathlength 1 cm), after equilibration the ellipticity at 222 nm was measured. The denaturation curves for each peptide are shown in Figure 3.

The denaturation profiles for the peptides were analyzed assuming a two-state equilibrium between unfolded monomeric peptide and folded, tetrameric bundle, assuming no significantly populated intermediates are present, as described previously[5]. Igor Pro software (Wavemetrics, Inc.) was used to fit the denaturation curves. Robust fits were obtained for each protein. For $\alpha_4F_3af_3d$ the absence of a lower base line limited the accuracy with which ΔG_{fold} could be determined, resulting in a larger error in this measurement than for the other two peptides.

4.3 – Results

4.3.1 - Structure of α_4H

As a reference structure against which to compare the effects of fluorination, the structure of α_4H was first determined (Fig. 4.3). The protein crystallized in space group, $I4_1$, and standard molecular replacement methods were used to solve its structure at a resolution of 1.36 Å using a 27 residue poly-alanine helix based on the anti-parallel structure of the 4-helix bundle, E20S [25]. Statistical data for the structure are given in

Table 4.1. The asymmetric unit comprises an anti-parallel dimer of two peptides (A and B chains), with the electron density being well defined for all but the last two residues of chain A and the first residue and last two residues of chain B. The anti-parallel 4-helix bundle structure is generated from the dimer of crystallographically non-equivalent peptide chains by the appropriate crystallographic two-fold operation.

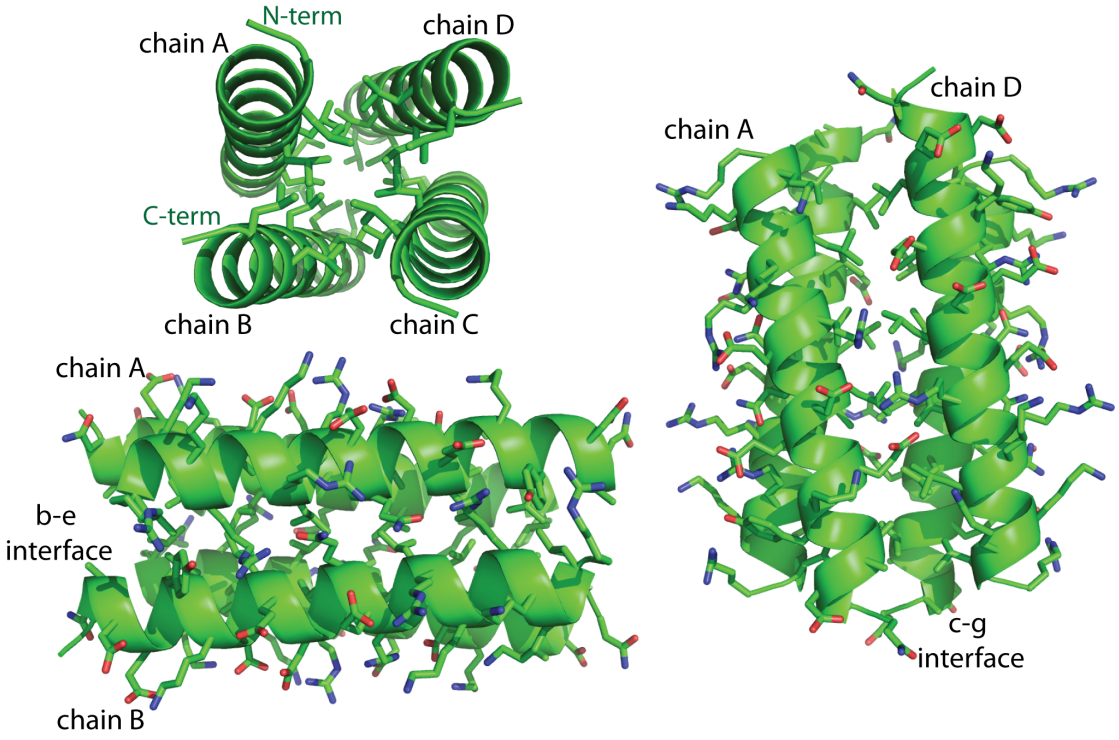


Figure 4.3 Overall structure of α_4H. *Top*: Tetrameric structure of α_4H displaying **a** and **d** position Leu residues as sticks in the coiled-coil core. *Bottom*: Side views of **b**–**e** (*Left*) and **c**–**g** (*Right*) interfaces displaying all side chains as sticks.

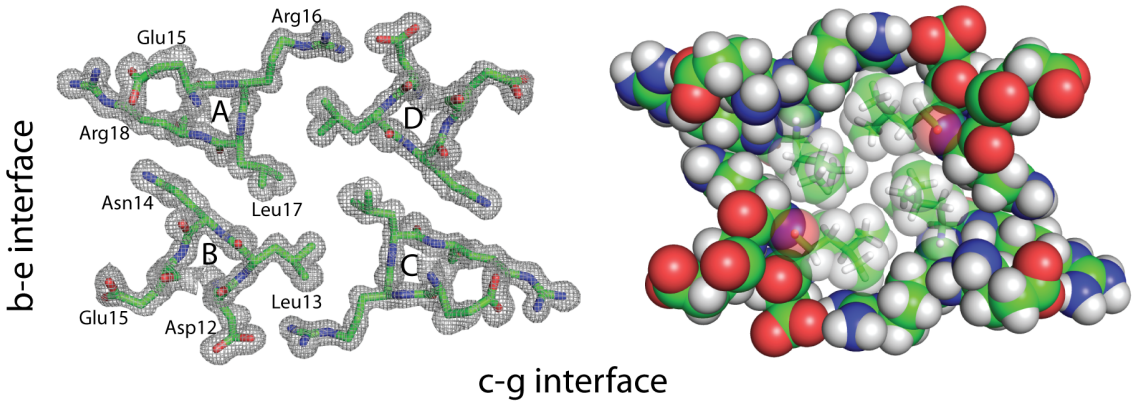


Figure 4.4. *Top*: Representative electron density ($2F_o$-F_c) maps for each protein with residues contoured at 1.0σ. *Bottom*: Space-filling representations of the hydrophobic core illustrating how fluorination conserves the tight packing of side-chains.

The modeled electron density for a cross-section of α_4H is shown in Figure 4.4. In accord with the intended design, the anti-parallel orientation of the α-helices in α_4H is enforced by complementary electrostatic interactions between residues in the **c** and **g** positions (**c-g** interface) and residues in the **b** and **e** positions (**b-e** interface). The two interfaces are non-equivalent, and in the case of α_4H this results in a larger spacing between helices of the **c-g** interface, which is formed by knobs-into-holes packing of the Leu residues at **d** positions (Fig. 4.5), than the **b-e** interface, which is formed by knobs-into-holes packing of the Leu residues at **a** positions (Fig. 4.4).

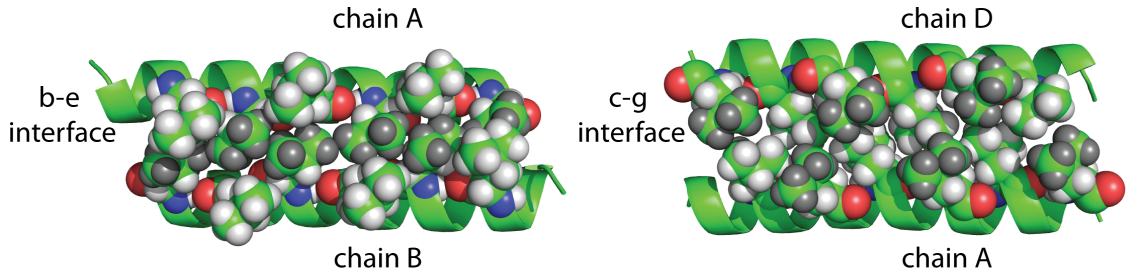
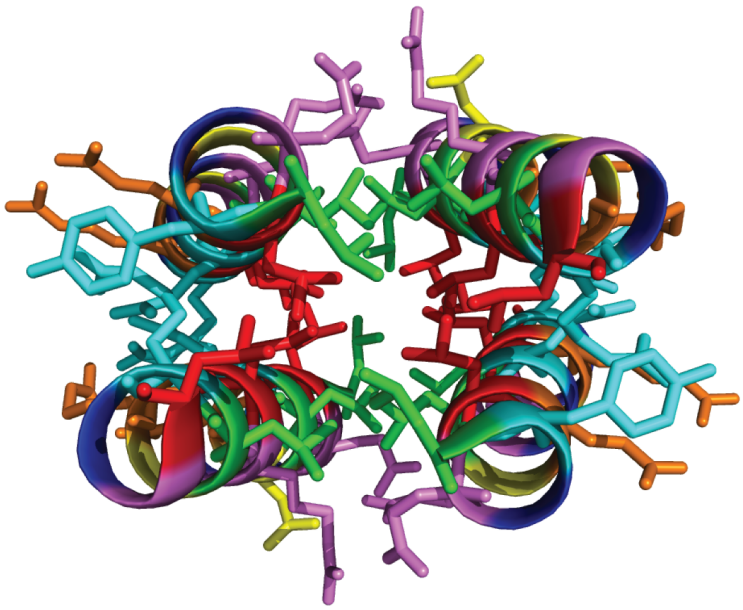


Figure 4.5. Layer 1 of the core is shown to the left; layer 6 to the right. Residues at **a** positions are colored dark gray to distinguish them from residues at **d** positions.

The program SOCKET [34] was used to further analyze the structure of α_4H (Fig. 4.6): the protein adopts a left-handed coiled-coil, with inter-helix angles of 152.36° (**b-e** interface) and 169.15° (**c-g** interface). The program also verified the knobs-into-holes packing arrangement of the leucine residues. Residues in the **a** position have an average packing angle of 63.42° while residues in the **d** position have an average packing angle of 134.34°. Packing angles generated by SOCKET measure the C_α-C_β bond vector of the knob residue relative to the C_α-C_α bond vector of the two residues on the sides of the corresponding hole.



knob residue	side-chain packing angle (degrees)
LeuA10	64.15
LeuA13	132.92
LeuA17	58.08
LeuA20	137.82
LeuB10	73.90
LeuB13	135.50
LeuB17	57.56
LeuB20	131.11

Figure 4.6. *Left*: Structure of the coiled-coil region of α_4H. The colored helices were identified as coiled-coils by SOCKET with knobs shown as sticks. Leucine residues in the **a** and **d** positions of the heptad repeat are colored red and green respectively. The default packing cutoff of 7 Å was used. *Right*: Side-chain packing angles of SOCKET identified type 4 knobs-into-holes (KIH) participating leucine residues.

4.3.2 - Structures of α_4F_3a, α_4F_3d and $\alpha_4F_3(6\text{-}13)$

Of the various fluorinated versions of α_4H studied in the Marsh laboratory, the structures of α_4F_3d and α_4F_3a were of particular interest because they are the most stable on a per-residue basis ($\Delta\Delta G_{fold}$ = -0.72 and -0.8 kcal/mol/hFLeu respectively). The combination of both protein structures provides a complete picture of how hFLeu packs into the hydrophobic core at **a** and **d** positions. α_4F_3a and α_4F_3d crystallized under similar conditions to α_4H with α_4F_3a in the same space group, $I4_1$, minimizing the possibility that altered crystal contacts may be responsible for any changes to the protein structure and α_4F_3d in the $P2_12_12$ space group. The structures were determined at 1.54 Å (α_4F_3a) and 1.19 Å (α_4F_3d) (Table 4.1) and all residues were resolved except the last two residues of the A chain and last residue of the B chain for α_4F_3a and all but the last residue of both the A and B chains for α_4F_3d.

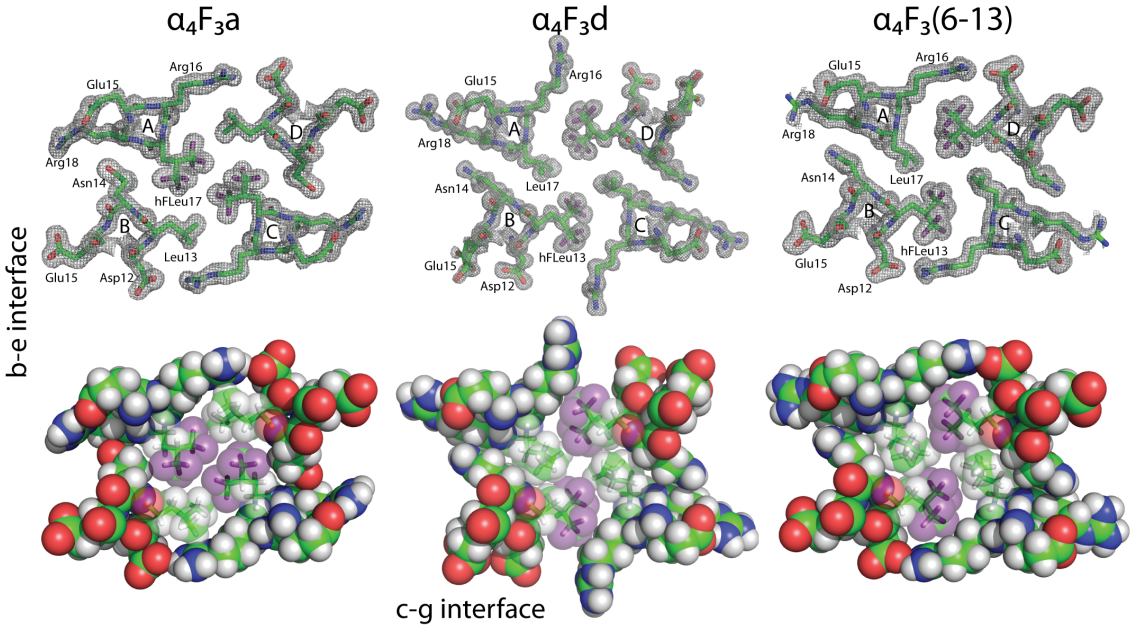


Figure 4.7. *Top*: Representative electron density ($2F_o$-F_c) maps for each protein with residues contoured at 1.0σ. *Bottom*: Space-filling representations of the hydrophobic core illustrating how fluorination conserves the tight packing of side-chains. Fluorine atoms are colored purple.

In particular, the electron density for all the hFLeu residues in α_4F_3a and α_4F_3d are well defined and clearly indicate the shape and orientation of the trifluoromethyl moieties (Fig. 4.7). The trifluoromethyl groups have full occupancy and do not appear to undergo rapid rotation, at least at the cryogenic temperatures at which the data was acquired. In each residue the two trifluoromethyl groups adopt a staggered configuration that minimizes steric repulsion between the trifluoromethyl groups and the β–carbon of hFLeu.

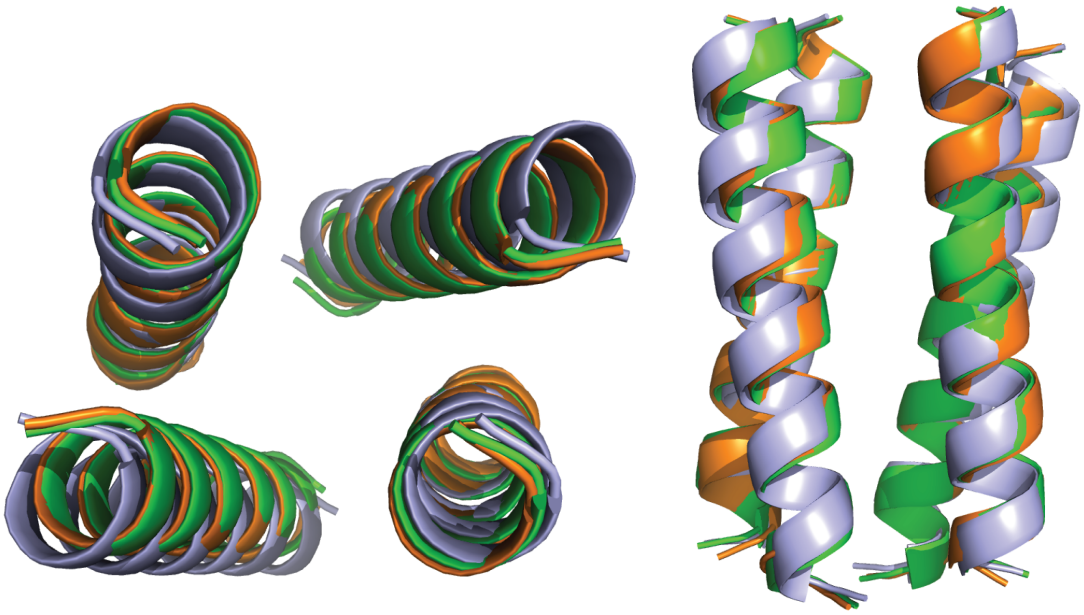


Figure 4.8. End-on and side views of the overlay of backbone atom traces, determined from the crystal structures, of α_4H (green), α_4F_3a (light blue), and α_4F_3d (orange).

Overall, the incorporation of 72 fluorine atoms into either α_4F_3a or α_4F_3d is remarkably non-perturbing to the tetrameric structure of the protein: the C_α positions of the α_4F_3d tetramer are nearly identical to that of α_4H with an rmsd of 0.26 Å, the helices of α_4F_3a move slightly further apart, displacing the C_α atoms of α_4F_3a by an rmsd of 0.95

Å from the coordinates of α_4H (Fig. 4.8). Accordingly, the structures of α_4F_3a and α_4F_3d differ slightly from each other with a C_α rmsd of 0.93 Å. This structural discrepancy between proteins having the same hFLeu content arises from positional differences between the placement of **a** and **d** residues in the heptad repeat.

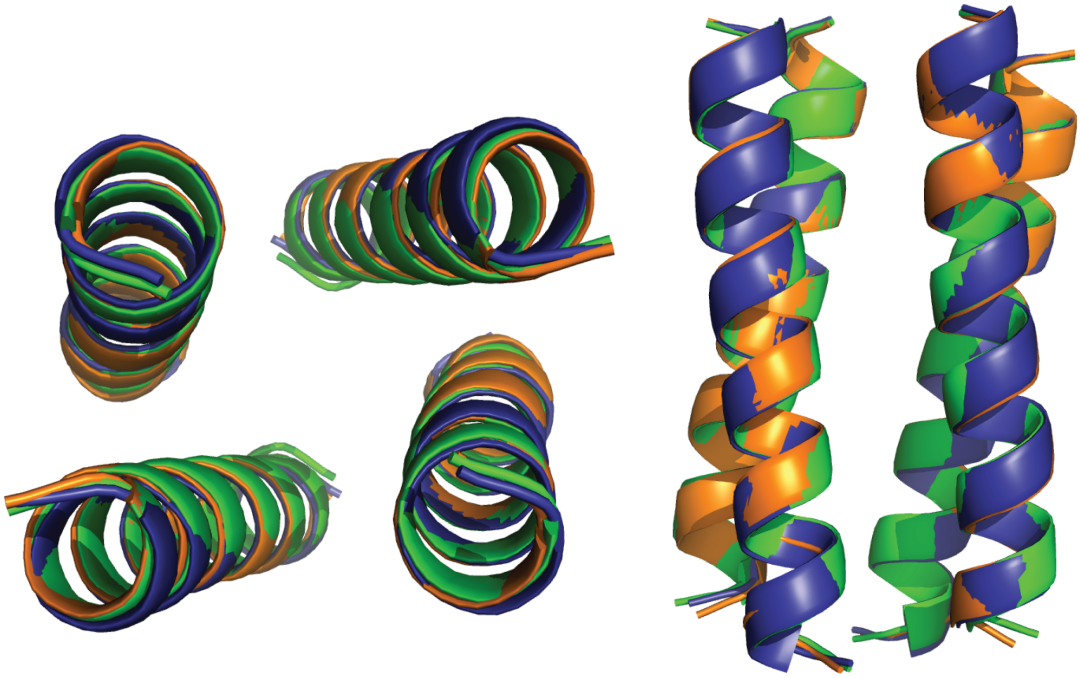


Figure 4.9. End-on and side views of the overlay of backbone atom traces, determined from the crystal structures, of α_4H (green), α_4F_3(6-13) (blue), and α_4F_3d (orange).

The protein α_4F_3(6-13) differs from α_4F_3d by having hFLeu in two **d** positions, 6 and 13, and one **a** position, 10. α_4F_3(6-13) crystallized under similar conditions to other α_4 proteins and in the space group $I4_1$. The structure was determined at 1.48 Å (Table 4.1) and all but the last two residues of both the A and B chains were resolved. The modeled electron density for a cross-section of α_4F_3(6-13) is shown in Figure 4.7. The electron density resembles that of other solved α_4 structures, displaying a tightly packed core and clearly indicating shape and orientation of trifluoromethyl groups. Altering the

packing arrangement of hFLeu in all **a** or all **d** positions by incorporating hFLeu into two **d** and one **a** position, as seen in α_4F_3(6-13), is well accommodated with a C_α rmsd of 0.30 Å, compared to α_4H (Fig. 4.9).

Interactions between hFLeu residues play an important role in forming the **b-e** interface of the α_4F_3a 4-helix bundle. Knobs-into-holes packing of hFLeu in adjacent layers of the core results in a tightly packed fluorinated "stripe" that runs along the entire **b-e** interface, as illustrated in Figure 4.10. The **c-g** interface, in contrast, is formed by knobs-into-holes packing of the Leu residues (Fig. 4.10). For α_4F_3d, hFLeu residues pack as knobs-into-holes into the **c-g** interface and Leu packs into the **b-e** interface – opposite that of α_4F_3a (Fig. 4.10). The **b-e** interface of α_4F_3(6-13) has two separated clusters of hFLeu residues that project into the core, while the **c-g** interface forms a fluorinated "stripe" with residue hFLeu10 projecting into the core.

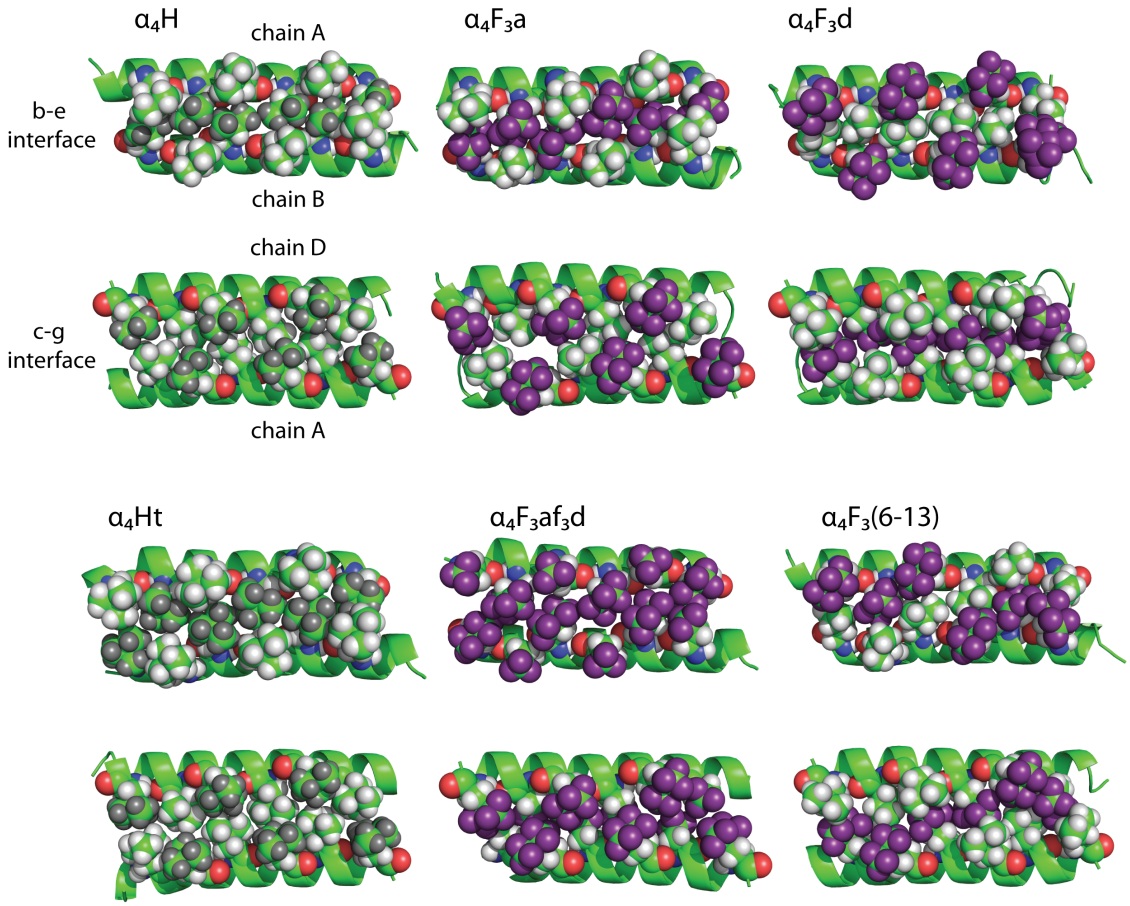


Figure 4.10. Layer 1 of the core is shown to the left; layer 6 to the right. In α_4H and α_4Ht residues at **a** positions are colored dark gray to distinguish them from residues at **d** positions. Fluorine atoms are colored purple.

4.3.3 - Comparison of Core Packing Between α_4H, α_4F_3a, α_4F_3d and α_4F_3(6-13)

Of particular interest, is how fluorination might alter interactions between residues in the hydrophobic cores of α_4F_3a and α_4F_3d in comparison to α_4H. Figures 4.11 and 4.12 compare in detail the interaction of one residue, LeuA17, in α_4H with the corresponding residues, hFLeuA17 in α_4F_3a and LeuA17 in α_4F_3d. In all proteins this residue is tightly packed in the core. The distances between the hydrogen atoms of the α_4H LeuA17 (which were modeled into the structure to facilitate comparison) and

adjacent Leu residues vary between 2.2 – 3.1 Å (Fig. 4.11), whereas the distances between the fluorine atoms of hFLeuA17 and adjacent hFLeu residues range between 2.5 – 3.2 Å (Fig. 4.12). These differences are consistent with the shorter van der Waals radius for hydrogen, 1.2 Å, compared to fluorine, 1.35 Å. The trifluoromethyl groups of hFLeuA17 also form extensive contacts with the methyl groups of adjacent Leu residues in α_4F_3a (Fig. 4.11), with fluorine-hydrogen distances of 2.5 – 3.3 Å. Consistent with α_4F_3a, α_4F_3d shows no preference towards maximizing fluorine-fluorine contacts or minimizing fluorine-hydrogen contacts. The interactions of residue LeuA17 with surrounding residues is shown in Figures 4.11 and 4.12. The distances between the hydrogen atoms of the α_4F_3d LeuA17 and adjacent Leu residues vary between 2.2 – 3.0 Å (Fig. 4.11), whereas the distances between the hydrogen atoms of LeuA17 and adjacent hFleu residues range between 2.7 – 3.1 Å (Fig. 4.12).

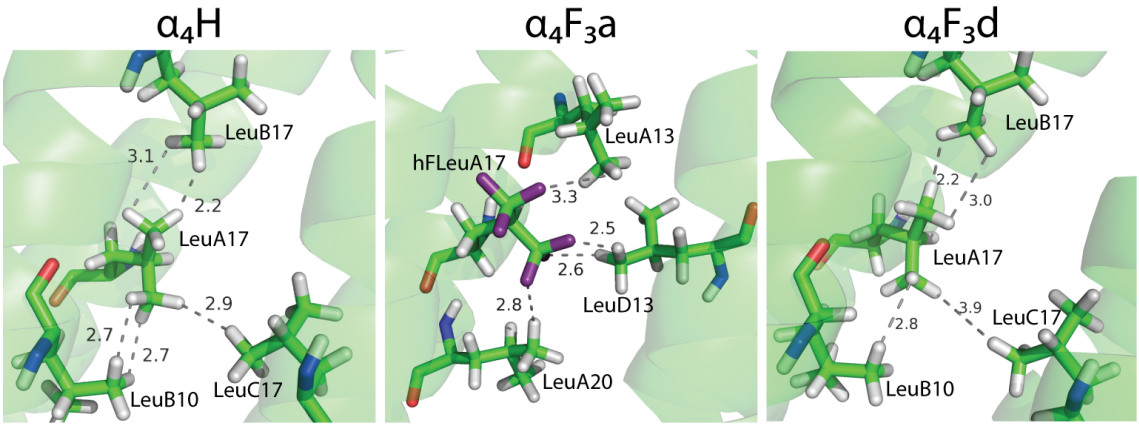


Figure 4.11. In each panel the residue at position A17 is oriented similarly to facilitate comparison. *Left*: Distances between LeuA17 and adjacent Leu residues in α_4H. *Middle*: The equivalent distance measurements between hFLeuA17 and adjacent Leu residues in α_4F_3a. *Right*: Distances between LeuA17 and adjacent Leu residues in α_4F_3d.

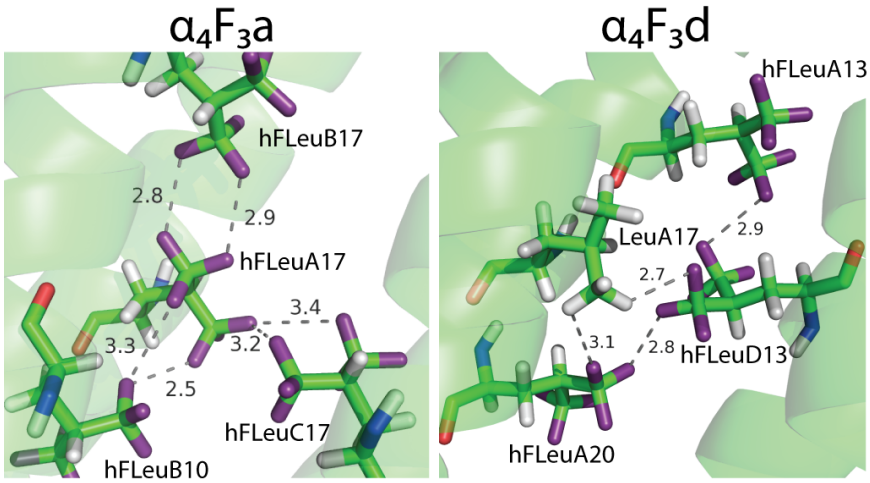


Figure 4.12. In each panel the residue at position A17 is oriented similarly to facilitate comparison. *Left*: Distances between hFLeuA17 and adjacent hFLeu residues in α_4F_3a. *Right*: The equivalent distance measurements between LeuA17 and adjacent hFLeu residues in α_4F_3d.

Overall, there is no evidence from the structures that the hFLeu residues adopt conformations that would either maximize fluorine-fluorine contacts or minimize fluorine-hydrocarbon contacts as would be predicted if favorable fluorous interactions between residues were important. There was also no evidence for dipolar interactions between trifluoromethyl groups and polar groups in the protein (as judged by proximity and alignment of the groups), which have been observed for some fluorinated compounds bound to proteins [35].

Comparison of the packing arrangement of Leu and hFLeu residues in α_4H, α_4F_3a, α_4F_3d and $\alpha_4F_3(6\text{-}13)$ for each of the 6 layers of the hydrophobic core was conducted, as shown in Figure 4.13. In α_4H, the central 4 layers of the core are packed so that the Leu residues at **a** positions extend towards the center of the protein core and make van der Waals contacts with Leu residues at the corresponding **a'** positions. The Leu residues at the **d** positions are less deeply buried and are oriented towards the **c-g** interface. In the

outer layers, 1 and 6, the situation is reversed. Thus Leu residues in **d** positions extend towards the center of the helical bundle, making contact with their counterparts at **d'**, whereas those at **a** and **a'** positions are oriented outwards towards the **b-e** interface. The reason for this change in packing arrangement is not obvious.

In α_4F_3a, the larger hFLeu residues pack the core in an arrangement similar to that seen for leucine in α_4H (Fig. 4.13). Thus, in layers 1 and 6 of the core, the hFLeu residues at the **a** and **a'** positions point towards the **b-e** interface, allowing the Leu residues at the **d** and **d'** positions to extend across the core and make van der Waals contacts with each other. In the central layers, the **a** position hFLeu residues extend into the center to make van der Waals contacts with hFLeu residues at **a'**. The only significant difference between the core packing of α_4F_3a and α_4H involves the packing of layer 3. Here, Leu13 in chains A and C of α_4F_3a extends into the center of core and disrupts the packing of hFLeu17 in chains B and D. This layer appears to be somewhat mobile, as Leu13 in the A and C chains can be modeled in 2 conformations each with ~50% occupancy (Fig. 4.14). Figures 4.10 and 4.13 displaying Leu13 of α_4F_3a only include the side chain conformation that contributes to disruption of hFLeu packing. This change in the packing arrangement is not seen in the chemically equivalent, but crystallographically non-equivalent, layer 4 of α_4F_3a. This may be explained if both central layers of the core have some inherent mobility in solution but crystal packing effects "freeze out" each layer in a different conformation in the crystal.

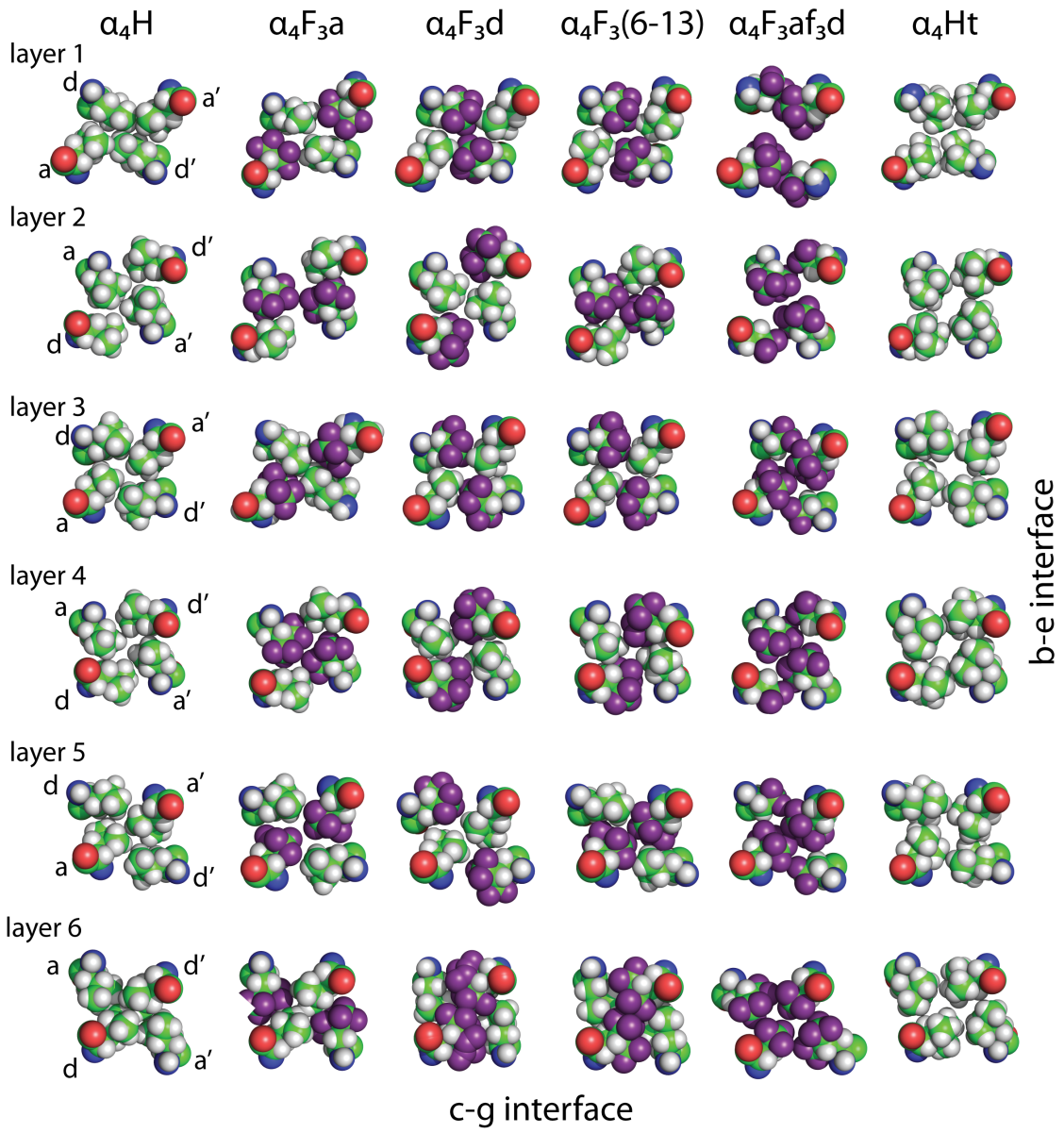


Figure 4.13. The pattern of hydrophobic contacts is generally unchanged by fluorination, despite the hFLeu side-chain being significantly larger: exceptions are layer 3 of α_4F_3a where Leu inserts between hFLeu residues, layer 1 of $\alpha_4F_3af_3d$ where fraying of the core results in a cleft opening across the **b-e** interface and layers 2 – 5 of α_4Ht where absence of side chain contact across the core results in a central cavity.

The core packing of Leu and hFLeu for each of the 6 layers of α_4F_3d and α_4F_3(6-13) was also analyzed (Fig. 4.13). Core packing for both α_4F_3d and α_4F_3(6-13) closely resembles that of α_4H. The central 2 layers of α_4H, α_4F_3d and α_4F_3(6-13) all display

residues in **a** positions extending towards the center of the protein core to make contacts with residues at the corresponding **a'** positions, however, a small cavity between residues in **a** and **a'** positions disrupts direct contact. Layers 2 and 5 of the core are packed so that the Leu residues at **a** positions extend towards the center of the protein core and make van der Waals contacts with Leu residues at the corresponding **a'** positions. The residues at **d** positions in layers 2 through 5 are less deeply buried and are oriented towards the **c-g** interface. In the outer layers, 1 and 6, the packing arrangement is reversed. Residues in **d** positions extend towards the center of the helical bundle, making contact with their counterparts at **d'**, whereas those at **a** and **a'** positions are oriented outwards towards the **b-e** interface.

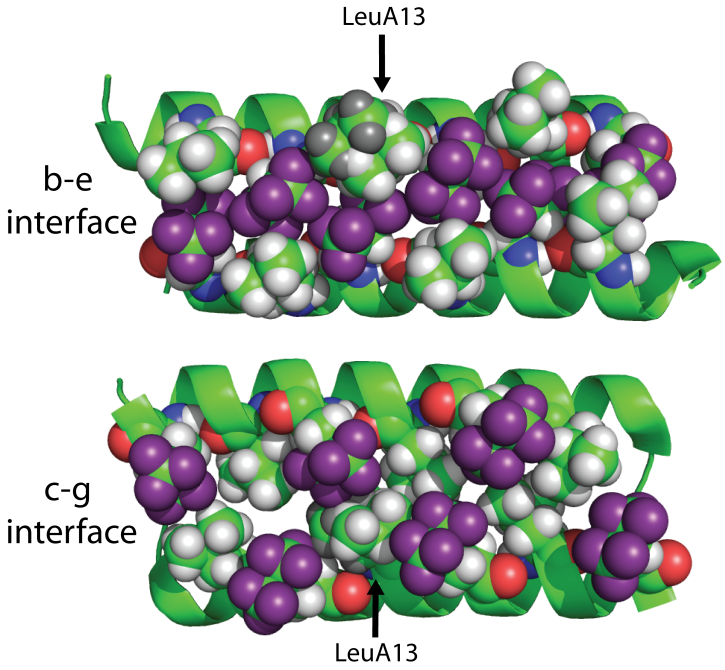


Figure 4.14. Conformational mobility observed for LeuA13 in the structure of α_4F_3a. The electron density for this residue could be modeled in two slightly different conformations as shown, each with ~50% occupancy. The LeuA13 side chain conformation with hydrogen atoms in gray is not displayed in other figures.

4.3.4 - Structural Basis for Enhanced Stability of α_4F_3a, α_4F_3d and $\alpha_4F_3(6\text{-}13)$

The high resolution structures of α_4H, α_4F_3a, α_4F_3d and $\alpha_4F_3(6\text{-}13)$ provide, for the first time, an opportunity to rationalize the enhanced stability imparted by fluorination. First considerations were whether the fluorinated residues are able to pack more efficiently into the hydrophobic core, where packing efficiency is defined as the volume occupied by the peptide chains divided by the total volume of the core. The formation of cavities within proteins is known to be destabilizing [36,37], so a more efficiently packed core should be associated with increased stability. To calculate the packing efficiency of the core truncated structures of α_4H, α_4F_3a and α_4F_3d were used in which the surface-exposed side-chains in the **b, c, e, f** and **g** positions were mutated *in silico* to alanine. This was done to prevent small changes in the conformation of the solvent-exposed side-chains from affecting the calculation. From these structures, the total van der Waals volume of the core and the sum of the van der Waals volumes of the individual peptide chains were calculated.

The tetrameric core of α_4H is calculated to have a total volume of ~8730 $Å^3$, of which the peptide chains occupy ~7820 $Å^3$ resulting in a packing efficiency of ~90%. In α_4F_3a and α_4F_3d substitution of hFLeu for Leu increases the volume of each peptide chain by an average of 96 $Å^3$, or an increase of 5% over the peptide chain volume of α_4H. This represents an increase of 32 $Å^3$/hFLeu residue; a figure that is in good agreement with previous calculations on the volume of hFLeu. The total volume of the α_4F_3a tetramer core expands to ~9220 $Å^3$, an increase of ~6%, therefore, the packing efficiency of α_4F_3a is essentially unchanged at ~89%. By this method, the packing efficiency of α_4F_3d and

α_4F_3(6-13) is similar at ~88% and ~89% respectively. Thus, the additional stability imparted by fluorination does not result from more efficient packing of the protein core, although the density of the core *is* slightly increased because fluorine is 19 times heavier than hydrogen.

This analysis illustrates the ability of the antiparallel, 4-helix bundle to accommodate hFLeu in any hydrophobic **a** or **d** position with little resulting packing deficiencies even though hydrophobic volume increases by 32 $Å^3$/hFLeu residue. However, as seen for α_4F_3a and α_4F_3d the stability imparted by hFLeu incorporation doesn't result from more efficient core packing. As expected, no overt preference for fluorocarbon-fluorocarbon or hydrocarbon-hydrocarbon contact is seen for α_4F_3(6-13), as judged by distance measurements of side chain atoms within van der Waals contact.

Next, "fluorous" interactions, i.e. favorable van der Waals type interactions between fluorocarbon residues may account for the stability of α_4F_3a, were considered, as such interactions have often been hypothesized to account for the high stability of highly fluorinated proteins [3,5,6,8,38]. The structures of the mixed hydrocarbon-fluorocarbon core of α_4F_3a and α_4F_3d provide a unique opportunity to test this hypothesis. If such fluorous interactions were important in α_4F_3a and α_4F_3d, fluorocarbon-fluorocarbon contacts should be maximized at the expense of fluorocarbon-hydrocarbon contacts. However, as discussed above, there is no evidence from the structure that this is the case.

For α_4F_3a, fluorous contacts could be increased by repacking layers 1 and 6 so that the hFLeu residues form contacts across the C_2 axis of the helical bundle. Instead Leu residues at the **d** positions interpose between the hFLeu residues – the same type of packing that is seen in α_4H (Fig. 4.13). Moreover, in layer 3 of α_4F_3a, which is the only

layer that differs in its packing from α_4H, contacts between the 2 hFLeu residues are disrupted by the leucines in the **d** position which would lead to a loss of putative fluorous interactions. For α_4F_3d, fluorous contacts are only seen in layers 1 and 6 where hFLeu residues have contact across the central axis. The central 4 layers retain the same packing as α_4H, with residues in **a** positions making contacts, thereby preventing fluorous contacts between **d** position hFLeu residues. α_4F_3(6,13) is seen to have unique packing due to the 2 **d** and 1 **a** hFLeu residues, resulting in hFLeu–hFLeu contacts in layers 1, 2, 5 and 6: more than either α_4F_3a or α_4F_3d. As discussed in section 4.1, α_4F_3(6-13) is less stable than either α_4F_3a or α_4F_3d even though fluorous contacts within a layer are increased.

Lastly, whether the increased stability of fluorinated proteins could be explained simply by the increase in hydrophobicity of the fluorinated residues was considered. It is well established that changes in the stability of natural proteins correlate with changes in buried hydrophobic surface area, or hydrophobic volume, associated with protein folding [39,40]. And although fluorocarbons are often described as being intrinsically more hydrophobic hydrocarbons, the larger volume and surface area of fluorocarbons is often overlooked in such comparisons; when these factors are accounted for, fluorocarbons and hydrocarbons exhibit similar hydrophobicities [41,42]. The structures of α_4F_3a and α_4F_3d allowed the increase in buried hydrophobic surface area associated with the introduction of hFLeu to be experimentally measured as ~20 $Å^2$/residue. Using the generally accepted value of ~30 cal/mol/$Å^2$ for the hydrophobic effect in proteins [39,40], α_4F_3a or α_4F_3d would be expected to be ~7.2 kcal/mol more stable than α_4H. The experimentally determined stabilization $\Delta\Delta G^o_{fold}$ = 8.6 – 9.6 kcal/mol is somewhat greater, but may be considered to

be in reasonable agreement given the approximate nature of the calculation. This stability value overestimates $\alpha_4F_3(6\text{-}13)$ ($\Delta\Delta G^{o}_{fold}$ = 5.7 kcal/mol and *m* = -1.98 kcal/mol/M) (Fig. 4.2).

The fairly close approximation of protein stability based upon hydrophobic surface area measurements indicates that increased hydrophobic content is likely the main contributor to fluorous protein stability. Discrepancies between the per hFLeu residue stability of α_4F_3a, α_4F_3d and $\alpha_4F_3(6\text{-}13)$ are not due to any distinct van der Waals contact between fluorous side chains or increased packing density. Worth noting again is the number of layers with hFLeu contacts across the central core; α_4F_3d having 2 layers, α_4F_3a having 3 layers and $\alpha_4F_3(6\text{-}13)$ having 4 layers. Differences in stability for proteins of identical hFLeu content likely arise from a preferred knobs-into-holes packing of bulky fluorous side chains. The side chain preference for knobs into holes packing has been thoroughly described by Gellman and Woolfson for canonical hydrophobic residues of the related antiparallel coiled-coil dimer [43-45]. In summary, their investigations conclude that there are distinct side chain packing preferences for both lateral and vertical interactions. Although their studies do not investigate hFLeu, they do generalize how direct packing interactions of non-like hydrophobic amino acids can be more spatially accommodating and thereby more stabilizing.

4.4 – Results: Structures of $\alpha_4F_3af_3d$ and α_4Ht

4.4.1 - Design of a Highly Fluorinated α_4 Protein Lacking Enhanced Stability

With the insights gained from the structure of α_4F_3a, a protein was designed to test whether fluorous interactions, or conventional hydrophobic volume and surface area

changes associated with fluorination, contribute more to protein stability. This peptide, $\alpha_4F_3af_3d$, incorporates smaller trifluoroethylglycine (tFeG) residues at **d** positions and hFLeu at **a** positions, so that the entire core is now packed with fluorocarbon side-chains. The smaller volume and surface area of tFeG with respect to Leu almost exactly compensates for the larger hFLeu side-chain, with result that $\alpha_4F_3af_3d$ has essentially the same volume and surface area as α_4H while containing 50 % more fluorine than α_4F_3a. Therefore, if fluorous interactions contribute significantly to stability, $\alpha_4F_3af_3d$ should be more stable than α_4F_3a; on the other hand if conventional hydrophobic effects dominate, $\alpha_4F_3af_3d$ should have a similar stability to α_4H.

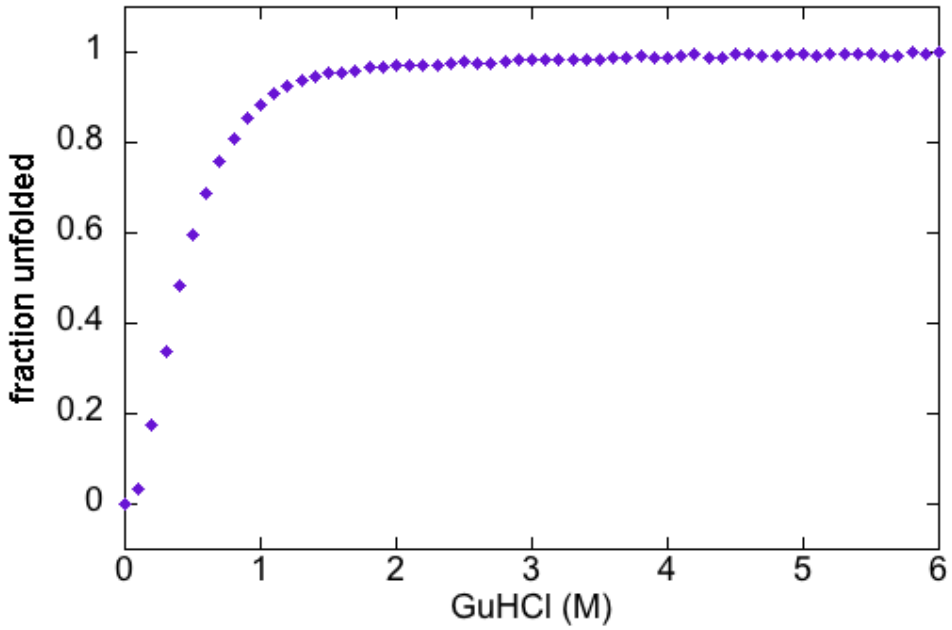


Figure 4.15. Guanidine hydrochloride-induced unfolding curve for $\alpha_4F_3af_3d$. Unfolding was monitored by following changes in ellipticity at 222 nm. Free energies of folding were calculated using these unfolding curves as described in section 2.5.

The free energy of folding for $\alpha_4F_3af_3d$ was determined using guanidinium hydrochloride as the denaturant. The protein unfolds at low concentrations of

guanidinium hydrochloride in a cooperative 2-state transition; fits to the unfolding curve yielded $\Delta G°_{fold}$ = -17.8 ± 1.0 kcal/mol and *m* = -2.78 kcal/mol/M (Fig. 4.15). Therefore, $\alpha_4F_3af_3d$ exhibits very similar stability to α_4H, $\Delta G°_{fold}$ = -18.0 ± 0.2 kcal/mol and *m* = -1.04 kcal/mol/M, consistent with our prediction that stability is primarily affected by changes in hydrophobic surface area and volume.

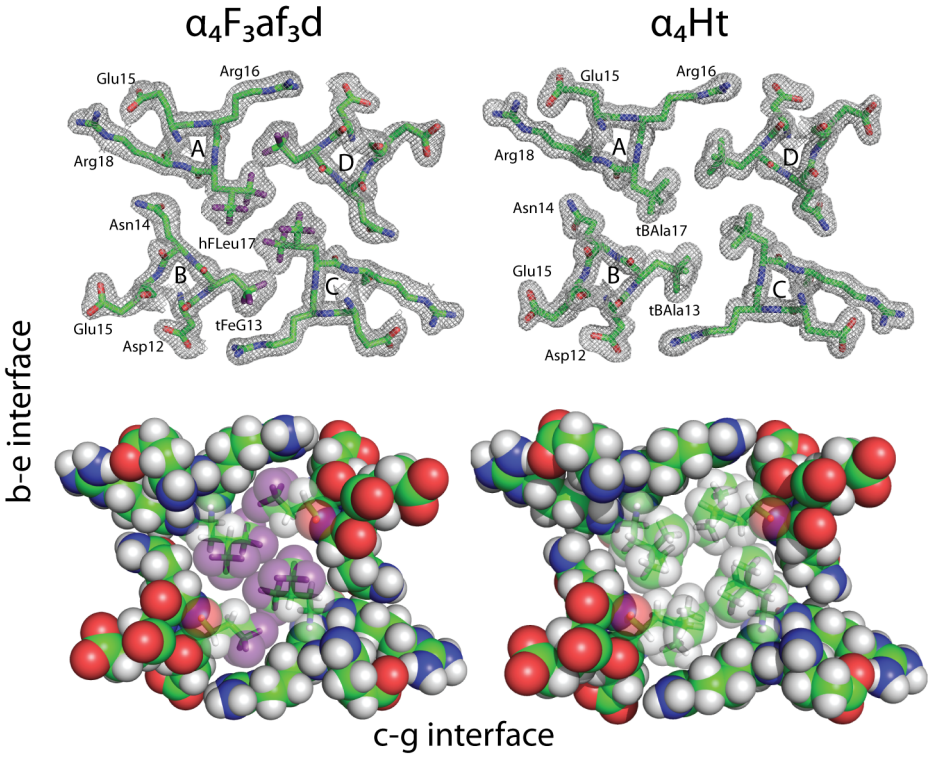


Figure 4.16. *Top*: Representative electron density ($2F_o$-F_c) maps for each protein with residues contoured at 1.0σ. *Bottom*: Space-filling representations of the hydrophobic core illustrating how fluorination conserves the tight packing of side chains and how *t*-butyl side chains form a central void. Fluorine atoms are colored purple.

$\alpha_4F_3af_3d$ crystallized under similar conditions to the other proteins, in space group $P2_12_12$, and its structure was determined at 1.72 Å resolution (Table 4.1). The modeled electron density for a cross-section of $\alpha_4F_3af_3d$ is shown in Figure 4.16. The C_α atoms overlay those of α_4H with an rmsd of 1.02 Å. (Fig. 4.17); however no electron density

was visible for the first 4 residues of the A and C chains, indicating that these residues are unstructured. This disrupts the first layer of the core so that the two hFLeu residues are oriented towards the **c-g** interface and are separated by 4.6 Å, opening up a narrow cleft in this layer (Fig. 4.10).

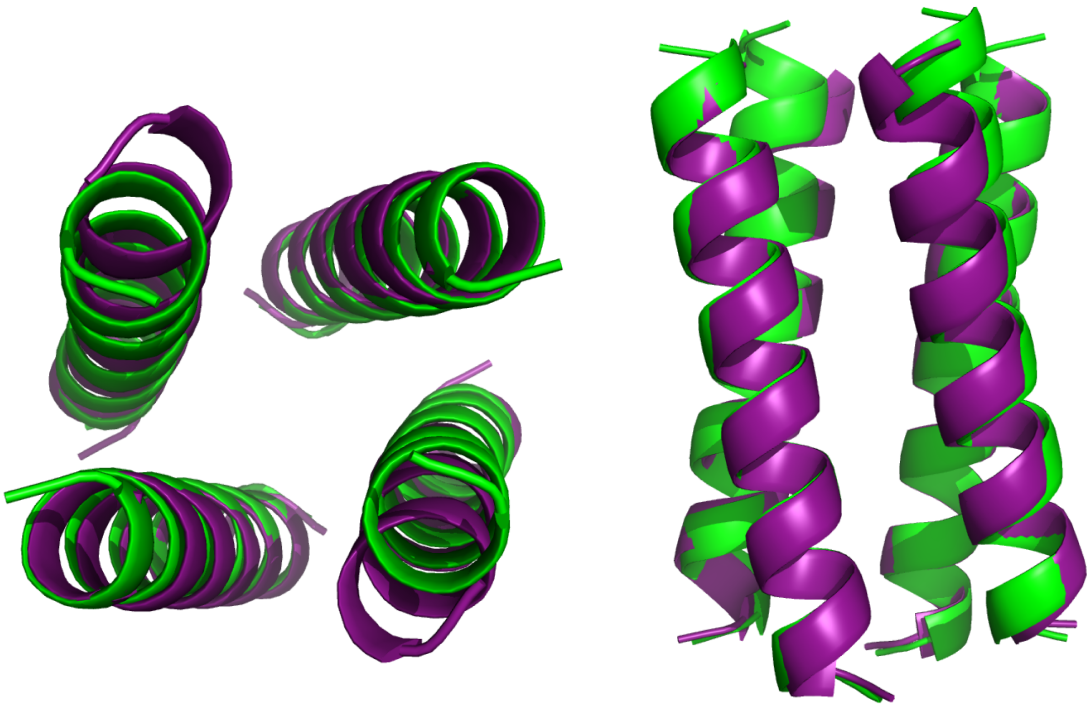


Figure 4.17. End-on and side views of the overlay of backbone atom traces, determined from the crystal structures, of α_4H (green) and $\alpha_4F_3af_3d$ (purple).

Figure 4.18 shows details of the contacts made by one residue, hFLeuA17, with adjacent hFLeu and tFeG side-chains in the core, and may be compared with the equivalent residues in α_4H and α_4F_3a (Fig. 4.7 and Fig. 4.8). Notably, the adjacent tFeGA13 residue points away from the hFLeu residue, resulting in longer fluorine-fluorine distances between neighboring residues than seen in the structure of α_4F_3a.

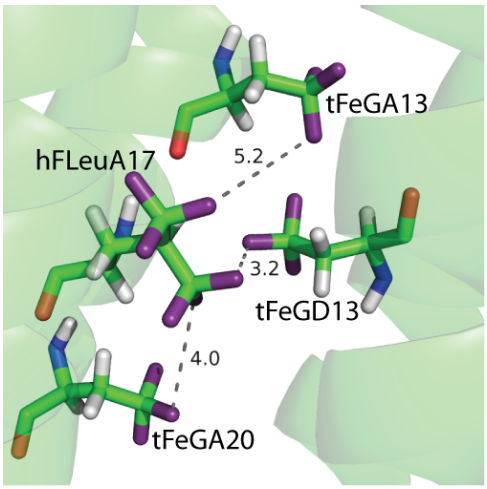


Figure 4.18. Distances between hFLeuA17 and adjacent tFeG residues in $\alpha_4F_3af_3d$; note that the tFeG residues adopt conformations that position the trifluoromethyl group further away from hFLeu than corresponding distances between Leu and hFLeu residues in α_4F_3a.

Examination of the hydrophobic core packing (Fig. 4.13) reveals that the remaining layers adopt an arrangement very similar to α_4H. In layers 2 – 5 hFLeu residues in the **a** positions extend into the center of the core to make contact with their counterparts at **a'** positions. The tFeG residues in the **d** positions are oriented towards the **c-g** interface where they pack in a knobs-into-holes fashion with tFeG residues from the adjacent peptide chain. In layer 6 the tFeG residues point into the center and the hFLeu residues abut them on either side. Using a similar analysis to that described earlier in this chapter, the packing efficiency was calculated for the core of $\alpha_4F_3af_3d$. Fraying of the helices reduces the buried hydrophobic surface area slightly, so the total core volume is 8360 $Å^3$; the volume occupied by the protein chains is 7390 $Å^3$ so the packing efficiency is ~88%, which is very similar to α_4H and α_4F_3a.

4.4.2 - Design of a Hydrocarbon α_4 Protein with Increased Hydrophobic Volume

With insights gained from structures of α_4F_3a, α_4F_3d and α_4F_3(6-13), a protein was designed with increased hydrocarbon volume to test whether increasing conventional hydrophobic volume and surface area could contribute to protein stability as effectively as fluorination. This peptide, α_4Ht, incorporates β-*t*-butyl-L-alanine (tBAla) residues at all **a** and **d** positions, so that the entire core is packed with a Leu analogue containing an additional methyl group at the C_γ. The larger volume and surface area of tBAla compared to Leu compensates for the larger hFLeu side-chain, giving α_4Ht a nearly identical hydrophobic core volume to that of α_4F_3a, α_4F_3d and α_4F_3(6-13). Studies from the Tirrell laboratory have, similarly, used the Leu analogues homoisoleucine (hIle) and trifluoroleucine (tFLeu) to increase hydrophobic volume and stabilize coiled-coil proteins [46]. These nonnatural amino acids have very similar hydrophobic volume as tBAla, with hIle a constitutional isomer. tBAla is hypothesized to stabilize folding to a similar extent as the "half-hFLeu-containing" α_4 proteins and may offer an advantage over hIle of increased packing efficiency due to the more compact *t*-butyl side chain compared to the extended isobutyl side chain of hIle.

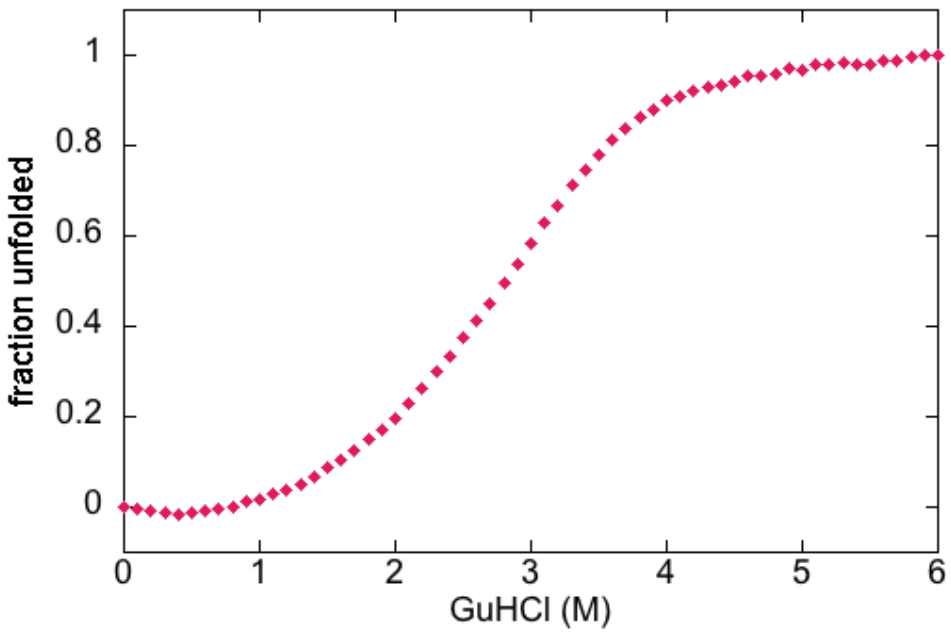


Figure 4.19. Guanidine hydrochloride-induced unfolding curve for α_4Ht. Unfolding was monitored by following changes in ellipticity at 222 nm. Free energies of folding were calculated using these unfolding curves as described in section 2.5.

The free energy of α_4Ht was determined using guanidinium hydrochloride as the denaturant and circular dichroism to follow protein unfolding. The protein unfolds in a cooperative 2-state transition with $\Delta G°_{fold}$ = -22.3 ± 1.0 kcal/mol and m = -1.78 kcal/mol/M (Fig. 4.19), which is 4.3 kcal/mol more stable than α_4H, although lower than that of any α_4F_3 proteins. α_4Ht crystallized under similar conditions to the other proteins, in space group $P2_12_12$, and its structure was determined at 1.54 Å resolution (Table 4.1). All residues but the first two and last two of chain A were resolved. The modeled electron density for a cross-section of α_4Ht is shown in Figure 4.16. Overlaying the tetrameric structure with that of α_4H yields an rmsd of 1.17 Å. The cause for this more significant deviation in backbone alignment is apparent when looking down the helical axis (Fig. 4.20); the helices are pushed apart compared to other 4-helix bundles that have

been analyzed. Volume expansion of the core produces a protein with a more symmetrical cross-section, as the expansion is primarily along the **b-e** interface.

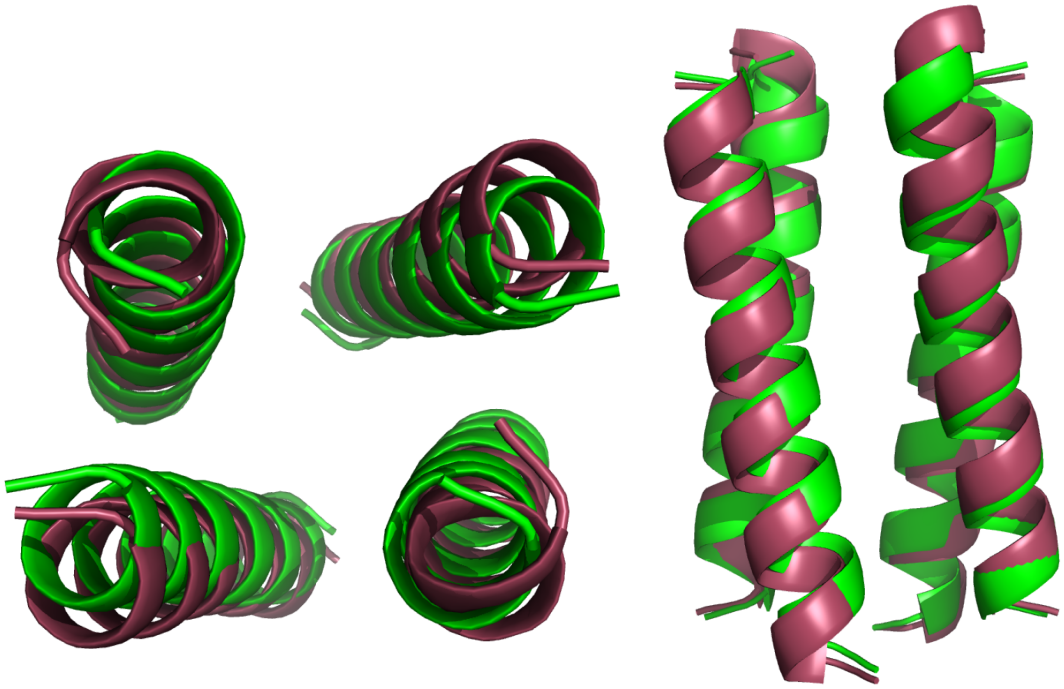


Figure 4.20. End-on and side views of the overlay of backbone atom traces, determined from the crystal structures, of α_4H (green), and α_4Ht (pink).

Interfacial packing of side chains closely resembles that of α_4H (Fig. 4.10), with tight van der Waals packing conservation between residues in **a** positions leading to the expansion of the **b-e** interface. Examination of the hydrophobic core packing (Fig. 4.13) reveals the six layers adopt an arrangement most similar to α_4H, α_4F_3d and α_4F_3(6-13). In layers 1 and 6 tBAla residues in **d** and **d'** positions extend across the center of the core to make contacts. In layers 2-5 the packing substantially deviates from previous structures, with a central void present due to side chains abutting across the **b-e** and **c-g** interfaces, as opposed to contacts across the central core. Creation of a central cavity appears to be a direct result of the unique shape of the *t*-butyl side chain. Few cavities are seen when α_4

proteins are packed with Leu, indicating the coiled-coil motif may be more accommodating to amino acids with side chains containing isopropyl moieties. In the case of hFLeu, the side chain shape of Leu is retained with increased bulk due to fluorine substitution.

The central hydrophobic cavity of α_4Ht is sequestered from water due to the residues in layers 1 and 6 distorting the end of the α-helix to achieve hydrophobic contact. Cavity volume measurements in Chimera using a probe radius of 1.4 Å indicate the presence of 5 discrete cavities with a total volume of ~270 $Å^3$ (Fig. 4.21). The eight tBAla residues of layers 3 and 4 form the central, and largest, cavity of ~100 $Å^3$. The $2F_o - F_c$ electron density map of α_4Ht displays ellipsoid shaped density in the central cavity at ~50% occupancy (Fig. 4.21). The size and hydrophobicity of this cavity could accommodate molecular nitrogen or oxygen and may provide a non-specific binding site for other small hydrophobic molecules.

Other research groups have investigated engineering hydrophobic pockets into coiled-coils. Studies by Johansson and collaborators have designed an anesthetic-binding hydrophobic pocket into the core of a 4-helix bundle protein to aid in understanding the mechanism of anesthetic binding to natural proteins [47,48]. The Ghadiri lab has created several mutants of a GCN4-pLI-based parallel tetramer which have engineered internal cavity volumes ranging from 80 – 370 $Å^3$, by mutating core Ile and Leu residues to Thr, Ser, Ala or Gly [49]. This study is of particular interest due to the wealth of structural data, including 12 distinct structures, 8 of which incorporate hydrophobic cavities, with 2 of these having iodobenzene bound in a hydrophobic pocket. Measurements from the iodobenzene-bound structures indicate that a cavity of ~300 $Å^3$ is needed to bind

iodobenzene which has a volume of ~170 Å^3. This study establishes parameters that would be useful in mutating the hydrophobic residues tBAla13 and tBAla17 of α_4Ht to Leu, Val or Ala in order to accommodate larger substrate molecules and alter substrate affinity.

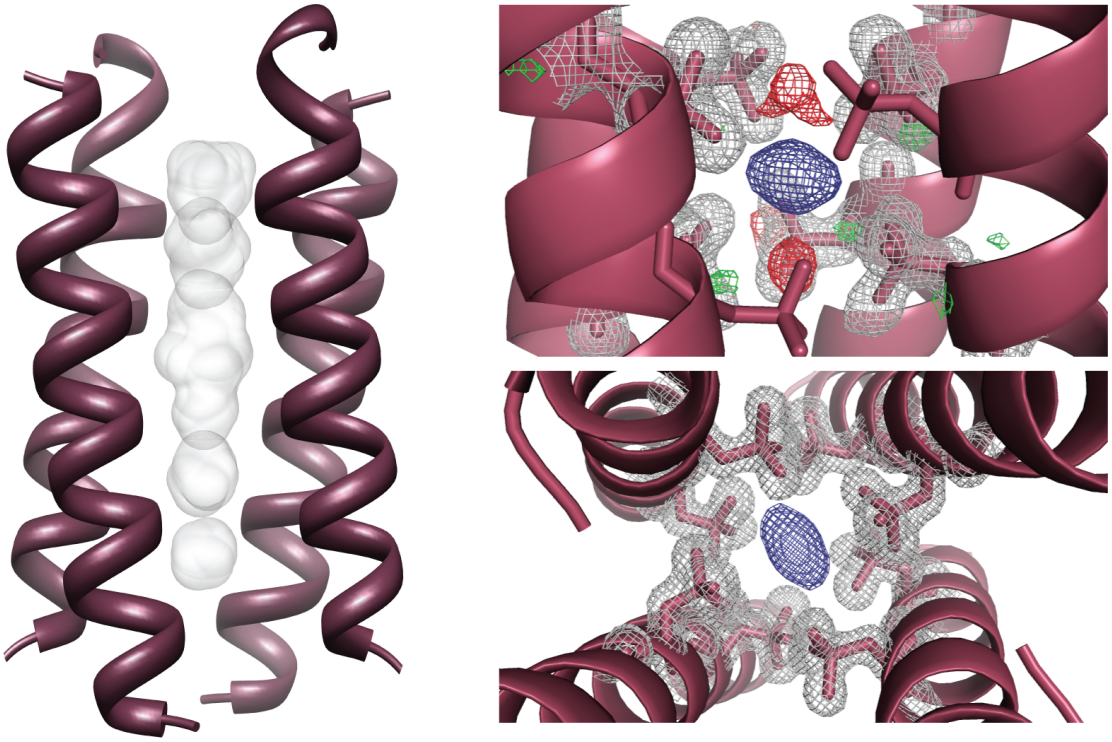


Figure 4.21 *Left*: Side view of α_4Ht displaying cavities in the hydrophobic core. *Right*: $2F_o$-F_c and F_o-F_c electron density maps for layers 3 and 4 of α_4Ht contoured at 1.0σ and 2.5σ respectively. The $2F_o$-F_c density for tBAla residues is shown in gray and unmodeled electron density is shown in blue.

α_4Ht possess similar hydrophobic core volume to the α_4F_3 series of proteins, however it is slightly less stable than α_4F_3(6-13) or α_4F_3(17-24) and considerably less stable than α_4F_3a or α_4F_3d. Analyzing the packing of α_4Ht indicates that it is similarly well-packed with a packing efficiency of ~ 90%. This efficient packing takes into account the ~ 200 Å^3 more internal cavities than other α_4 proteins (α_4 proteins in general

contain ~ 100 Å^3 of internal cavities which are measureable by a 1.4 Å probe). Although this large cavity is present in α_4Ht, similar packing efficiency to other α_4 proteins is seen. This is due to small voids, below the 1.4 Å probe cutoff, which are more prevalent in the α_4F_3 structures than in α_4Ht. Large hydrophobic voids, such as those seen in α_4Ht are known to destabilize the folded state, this is exemplified by GCN4-pLI mutants, where Leu9Ala and Ile12Ala have ~ 200 Å^3 cavities and their melting temperatures decrease by 14 and 6 °C respectively. The stability penalty for α_4Ht cavities can be estimated by using a generally accepted value of 24 cal/mol/Å^3 for the hydrophobic effect in proteins. This estimates a potential loss in stability of ~ 4.8 kcal/mol for the hydrophobic cavity, providing a possible explanation for the decreased stability compared to fluorinated proteins with similar hydrophobic surface area.

4.5 – Discussion

The field of protein design is now advancing to embrace amino acids beyond the 20 canonical residues [50-52]. However, to successfully exploit the potential of new amino acids to augment the functions of natural proteins, it is important to understand both how the novel residues modulate the structures of the proteins that they are incorporated into and how structural changes, in turn, give rise to changes in the physical and biological properties of the protein. In the case of fluorinated amino acids, numerous studies have demonstrated their utility stabilizing proteins against thermal unfolding and chemical denaturation, however the explanation for this enhanced stability has remained a matter of debate. The studies in this chapter provide detailed structural information on how

highly fluorinated amino acids are accommodated within a protein and provide insights into the origin of the stabilizing effect.

The X-ray structures of α_4F_3a, α_4F_3d, $\alpha_4F_3(6\text{-}13)$ and $\alpha_4F_3af_3d$ reveal that large numbers of fluorine atoms (72 and 108 atoms) can be incorporated into proteins with only minimal perturbation of their structure, even though the hFLeu side-chain is some 32 $Å^3$ (~30%) larger than Leu. It has been conjectured that the unusual phase-segregating properties of per-fluorinated molecules, ingeniously exploited in organic synthesis [53,54], could be used to direct protein-protein interactions in a manner orthogonal to the conventional hydrophobic effect [5,6,8,9,38]; however, the structures of α_4F_3a, α_4F_3d, $\alpha_4F_3(6\text{-}13)$ and $\alpha_4F_3af_3d$ reveal no evidence for preferential fluorous interactions between fluorinated residues.

Instead, the increased thermodynamic stability of α_4F_3a, α_4F_3d and $\alpha_4F_3(6\text{-}13)$ can be adequately explained by the increases in buried hydrophobic surface area and volume that accompany fluorination. In other words, the same principles that underpin the stability of natural proteins, efficient packing of side-chains and conventional hydrophobic effects, appear to be responsible for the enhanced stability of fluorinated proteins. Furthermore, although almost all studies have reported increases in protein stability upon incorporation of fluorinated residues, the design of $\alpha_4F_3af_3d$, which has a highly fluorinated core, demonstrates that fluorination *per se* does not stabilize proteins. Thus, if changes in residue size are controlled for, extensively fluorinated proteins can be designed that have very similar structures and stabilities to their natural counterparts.

The above discussion raises the question of why fluorination has proved such a generally successful strategy for increasing protein stability. It appears that this may be

due to the fact that fluorination closely preserves the *shape* of side-chains, which is important for the correct packing of side-chains within the hydrophobic core, whilst increasing size and hydrophobicity. This allows the fluorinated residue to be introduced with minimal adjustment of the surrounding structure, as demonstrated by the structures of α_4F_3a, α_4F_3d and $\alpha_4F_3(6\text{-}13)$. The alternative approach to increasing residue hydrophobicity would be to add extra carbon atoms to the side-chain, e.g. by changing an alanine to a valine. However, such modifications will also change the side-chain's shape significantly so that often it will fit poorly into the existing structure and hence will tend to compromise stability and/or biological activity of the protein.

A further question is whether "fluorous" interactions can really be used to facilitate specific interactions between proteins. To address this, it is necessary to consider the nature of the fluorous effect in more detail. Although the phase separation of fluorocarbon:hydrocarbon solvent mixtures is often ascribed to "fluorophilic" or "fluorous" interactions between fluorocarbon molecules, this is not strictly correct. The phenomenon arises because the cohesive dispersion forces between two hydrocarbon molecules are greater than between two fluorocarbon molecules, or a between a fluorocarbon and a hydrocarbon molecule (this because hydrocarbons are more polarizable than fluorocarbons), and thus fluorocarbons are excluded from the hydrocarbons. More generally, the mutual solubility (or immiscibility) of a mixture of two non-polar solvents is related to the difference in the solubility parameter, δ, which is defined as:

$$\delta = (\Delta E^{V}/V)^{1/2}$$

where ΔE^V is the energy of vaporization and V the molal volume of the pure liquid at a given temperature [55,56]. As the difference in δ between the two solvents increases, the heat of mixing becomes more unfavorable until they are no longer miscible. As discussed by Scott [55], fluorocarbons have low δ values both because they have low boiling points and larger molal volumes than hydrocarbons.

Clearly there are many differences when considering the hydrophobic interface between two proteins and the immiscibility of two liquids, such that the principles discussed above that underlie the segregating tendency of small fluorocarbon molecules cannot be simply applied to protein-protein interactions. Notably, protein interfaces are highly structured and formed by specific interactions between side-chains whereas solvent-solute interactions are transient, non-specific and dynamic. Steric effects appear to play a far more important role in specifying hydrophobic interfaces between proteins than the potential differences in dispersion forces between fluorocarbon and hydrocarbon residues. Moreover, although fluorinated residues are similar in shape, they are not completely isosteric with their hydrocarbon counterparts, so the influence of steric effects can never be entirely ignored.

The peptides described in the literature [3,9] that appear to exhibit "fluorous" segregation were designed to form parallel coiled-coils, whereas our studies use peptides that form anti-parallel coiled-coils. As discussed previously by others [57,58], the oligomerization state of parallel coiled-coils is very sensitive to changes in the volume of the hydrophobic core, whereas the anti-parallel arrangement is far more robust. Significantly, fluorination also induced a change in the oligomerization state of the self-segregating peptides from a dimeric to a tetrameric coiled-coil [3], consistent with the

larger volume of the fluorinated side-chains introduced into the core. Thus the self-segregating behavior of these peptides might be better ascribed to steric effects, rather than a manifestation of fluorous segregation *per se*.

Incorporation of hFLeu increases stability via added hydrophobic volume while minimizing structure perturbation by closely retaining the shape of Leu. α_4Ht demonstrates that although the increased hydrophobic content leads to higher stability, a difference in side chain shape can alter the core packing. Somewhat surprising is the antiparallel 4-helix bundle accommodation of α_4Ht core repacking brought about by tBAla shape, allowing the formation of a unique structure with a large internal hydrophobic cavity.

Further studies into the thermodynamic parameters that contribute to the stability of these and other α_4 proteins is discussed in Chapter 5 with the expectation that a comprehensive structural and thermodynamic study will help guide future endeavors stabilizing proteins through hydrophobic interactions.

4.6 – References

(1) Marsh, E. N. G. *Chem. Biol.* **2000**, *7*, R153.
(2) Yoder, N. C.; Yüksel, D.; Dafik, L.; Kumar, K. *Curr. Opin. Chem. Biol.* **2006**, *10*, 576.
(3) Bilgiçer, B.; Xing, X.; Kumar, K. *J. Am. Chem. Soc.* **2001**, *123*, 11815.
(4) Chiu, H.-P.; Suzuki, Y.; Gullickson, D.; Ahmad, R.; Kokona, B.; Fairman, R.; Cheng, R. P. *J. Am. Chem. Soc.* **2006**, *128*, 15556.
(5) Lee, K.-H.; Lee, H.-Y.; Slutsky, M. M.; Anderson, J. T.; Marsh, E. N. G. *Biochemistry* **2004**, *43*, 16277.
(6) Tang, Y.; Ghirlanda, G.; Petka, W. A.; Nakajima, T.; DeGrado, W. F.; Tirrell, D. A. *Angew. Chem.* **2001**, *40*, 1494.
(7) Tang, Y.; Ghirlanda, G.; Vaidehi, N.; Kua, J.; Mainz, D. T.; Goddard, W. A.; DeGrado, W. F.; Tirrell, D. A. *Biochemistry* **2001**, *40*, 2790.
(8) Niemz, A.; Tirrell, D. A. *J. Am. Chem. Soc.* **2001**, *123*, 7407.

(9) Bilgiçer, B.; Kumar, K. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 15324.
(10) Jäckel, C.; Salwiczek, M.; Koksch, B. *Angew. Chem.* **2006**, *45*, 4198.
(11) Lee, H.-Y.; Lee, K.-H.; Al-Hashimi, H. M.; Marsh, E. N. G. *J. Am. Chem. Soc.* **2006**, *128*, 337.
(12) Khan, F.; Kuprov, I.; Craggs, T. D.; Hore, P. J.; Jackson, S. E. *J. Am. Chem. Soc.* **2006**, *128*, 10729.
(13) Woll, M. G.; Hadley, E. B.; Mecozzi, S.; Gellman, S. H. *J. Am. Chem. Soc.* **2006**, *128*, 15932.
(14) Gottler, L. M.; de la Salud-Bea, R.; Marsh, E. N. G. *Biochemistry* **2008**, *47*, 4484.
(15) Meng, H.; Krishnaji, S. T.; Beinborn, M.; Kumar, K. *J. Med. Chem.* **2008**, *51*, 7303.
(16) Montclare, J. K.; Son, S.; Clark, G. A.; Kumar, K.; Tirrell, D. A. *ChemBioChem* **2009**, *10*, 84.
(17) Cornilescu, G.; Hadley, E. B.; Woll, M. G.; Markley, J. L.; Gellman, S. H.; Cornilescu, C. C. *Protein Science* **2007**, *16*, 14.
(18) Gottler, L. M.; Lee, H.-Y.; Shelburne, C. E.; Ramamoorthy, A.; Marsh, E. N. G. *ChemBioChem* **2008**, *9*, 370.
(19) Son, S.; Tanrikulu, I. C.; Tirrell, D. A. *ChemBioChem* **2006**, *7*, 1251.
(20) Buer, B. C.; de la Salud-Bea, R.; Al Hashimi, H. M.; Marsh, E. N. G. *Biochemistry* **2009**, *48*, 10810.
(21) Anderson, J. T.; Toogood, P. L.; Marsh, E. N. G. *Org. Lett.* **2002**, *4*, 4281.
(22) Buer, B. C.; Chugh, J.; Al-Hashimi, H. M.; Marsh, E. N. G. *Biochemistry* **2010**, *49*, 5760.
(23) Otwinowski, Z.; Minor, W. *Methods Enzymol.* **1997**, *276*, 307.
(24) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *J. Appl. Cryst.* **2007**, *40*, 658.
(25) Yadav, M. K.; Leman, L. J.; Price, D. J.; Brooks, C. L.; Stout, C. D.; Ghadiri, M. R. *Biochemistry* **2006**, *45*, 4463.
(26) Langer, G.; Cohen, S. X.; Lamzin, V. S.; Perrakis, A. *Nat. Protocols* **2008**, *3*, 1171.
(27) Schuttelkopf, A. W.; van Aalten, D. M. F. *Acta Cryst.* **2004**, *D60*, 1355.
(28) Bricogne, G.; Blanc, E.; Brandl, M.; Flensburg, C.; Keller, P.; Paciorek, W.; Roversi, P.; Sharff, A.; Smart, O. S.; Vonrhein, C.; Womack, T. O.; Global Phasing Ltd.: Cambridge, United Kingdom, 2011.
(29) Emsley, P.; Cowtan, K. *Acta Cryst.* **2004**, *D60*, 2126.
(30) Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall, W. B.; Snoeyink, J.; Richardson, J. S.; Richardson, D. C. *Nucleic Acids Res.* **2007**, *35*, 375.
(31) Zucker, F.; Champ, P. C.; Merritt, E. A. *Acta Cryst.* **2010**, *D66*, 889.
(32) Hooft, R. W. W.; Vriend, G.; Sander, C.; Abola, E. E. *Nature* **1996**, *381*, 272.
(33) Sanner, M. F.; Olson, A. J.; Spehner, J.-C. *Biopolymers* **1996**, *38*, 305.
(34) Walshaw, J.; Woolfson, D. N. *J. Mol. Biol.* **2001**, *307*, 1427.
(35) Müller, K.; Faeh, C.; Diederich, F. *Science* **2007**, *317*, 1881.
(36) Eriksson, A.; Baase, W.; Zhang, X.; Heinz, D.; Blaber, M.; Baldwin, E.; Matthews, B. *Science* **1992**, *255*, 178.

(37) Serrano, L.; Kellis Jr, J. T.; Cann, P.; Matouschek, A.; Fersht, A. R. *J. Mol. Biol.* **1992**, *224*, 783.
(38) Bilgiçer, B.; Kumar, K. *Tetrahedron* **2002**, *58*, 4105.
(39) Sharp, K.; Nicholls, A.; Fine, R.; Honig, B. *Science* **1991**, *252*, 106.
(40) Sharp, K. A.; Nicholls, A.; Friedman, R.; Honig, B. *Biochemistry* **1991**, *30*, 9686.
(41) Gao, J.; Qiao, S.; Whitesides, G. M. *J. Med. Chem.* **1995**, *38*, 2292.
(42) Mecinović, J.; Snyder, P. W.; Mirica, K. A.; Bai, S.; Mack, E. T.; Kwant, R. L.; Moustakas, D. T.; Héroux, A.; Whitesides, G. M. *J. Am. Chem. Soc.* **2011**, *133*, 14017.
(43) Hadley, E. B.; Testa, O. D.; Woolfson, D. N.; Gellman, S. H. *Proceedings of the National Academy of Sciences* **2008**, *105*, 530.
(44) Hadley, E. B.; Gellman, S. H. *Journal of the American Chemical Society* **2006**, *128*, 16444.
(45) Steinkruger, J. D.; Bartlett, G. J.; Hadley, E. B.; Fay, L.; Woolfson, D. N.; Gellman, S. H. *Journal of the American Chemical Society* **2012**, *134*, 2626.
(46) Van Deventer, J. A.; Fisk, J. D.; Tirrell, D. A. *ChemBioChem* **2011**, *12*, 700.
(47) Cui, T.; Bondarenko, V.; Ma, D.; Canlas, C.; Brandon, N. R.; Johansson, J. S.; Xu, Y.; Tang, P. *Biophysical journal* **2008**, *94*, 4464.
(48) Johansson, J. S.; Scharf, D.; Davies, L. A.; Reddy, K. S.; Eckenhoff, R. G. *Biophysical journal* **2000**, *78*, 982.
(49) Yadav, M. K.; Redman, J. E.; Leman, L. J.; Alvarez-Gutiérrez, J. M.; Zhang, Y.; Stout, C. D.; Ghadiri, M. R. *Biochemistry* **2005**, *44*, 9723.
(50) Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z.; Schultz, P. G. *Science* **2003**, *301*, 964.
(51) Link, A. J.; Mock, M. L.; Tirrell, D. A. *Curr. Opin. Biotechnol.* **2003**, *14*, 603.
(52) Valiyaveetil, F. I.; Sekedat, M.; Muir, T. W.; MacKinnon, R. *Angew. Chem.* **2004**, *43*, 2504.
(53) Horvath, I. T.; Rabai, J. *Science* **1994**, *266*, 72.
(54) Luo, Z.; Zhang, Q.; Oderaotoshi, Y.; Curran, D. P. *Science* **2001**, *291*, 1766.
(55) Scott, R. L. *J. Am. Chem. Soc.* **1948**, *70*, 4090.
(56) Hildebrand, J. H.; Cochran, D. R. F. *J. Am. Chem. Soc.* **1949**, *71*, 22.
(57) Betz, S. F.; DeGrado, W. F. *Biochemistry* **1996**, *35*, 6955.
(58) Harbury, P. B.; Tidor, B.; Kim, P. S. *Proc. Natl. Acad. Sci.* **1995**, *92*, 8408.

Chapter 5

Thermodynamics of Fluorinated Protein Unfolding

5.1 – Introduction

The thermodynamic parameters described in this chapter were determined by fits to protein denaturation data. The programming of equations describing protein denaturation in Matlab was essential to this work and accomplished thanks to Benjamin J. Levin.

Nature provides many exquisite protein structures and highly efficient enzymes that have evolved to perform specific biological tasks. These natural proteins function optimally in a narrowly defined set of environmental conditions, which are conducive to supporting life. The ability of biologically based proteins to function in harsh environmental conditions (high temperature, organic solvents or extreme pH) can be achieved through engineering greater conformational stability. More stable proteins may also be more resistant to protease degradation.

Using a palette of the 20 canonical amino acids, researchers have been able to produce biologically compatible proteins with increased stability by introducing disufide linkages, salt bridges or repacking of the hydrophobic core. Other strategies involve the global or site-specific incorporation of nonnatural amino acids in place of their natural

counterparts. The incorporation of nonproteinogenic side chains allows a greater sampling of chemical space – including funtionalities orthogonal to those seen in biology.

The introduction of highly fluorinated amino acids into hydrophobic positions of a protein has proved a very effective means for stabilizing a variety of protein folds[1-13]. As discussed in previous chapters the inspiration for using fluorinated amino acids comes from the unique physicochemical properties of perfluorinated small molecules. Fluorous analogs of the hydrophobic amino acids leucine, valine and phenylalanine have been incorporated into both natural and *de novo*-designed proteins to increase the free energy of unfolding ($\Delta G_u°$) by ~0.4 to 1.4 kcal/mol/fluorous residue[2-5,7,9,11]. The dramatic increases in thermodynamic stability are testaments to the potential of fluorous amino acids to stabilize proteins and the biocompatibility of these nonnatural moieties.

Studies of fluorinated proteins have focused on increasing Gibbs free energy ($\Delta G°$) by replacing hydrocarbon residues with more hydrophobic fluorocarbon analogues. $\Delta G°$ is a measurement of thermodynamic stability and consists of two terms describing enthalpic ($\Delta H°$) and entropic ($-T\Delta S°$) contributions. The balance between $\Delta H°$ and $\Delta S°$ determines thermodynamic stability of a protein. Measurements have been made for many natural proteins to assess contributions of $\Delta H°$ and $\Delta S°$ to overall stability and determine how mutations can alter these values. Studies of natural proteins have elucidated common trends showing that changes in $\Delta H°$ are the product of van der Waals interactions, hydrogen bonding and electrostatics whereas changes in $\Delta S°$ result from solvation differences that occur in burying hydrophobic groups.

Hydrocarbon and fluorocarbon amino acids have been shown to stabilize proteins to approximately the same extent when volume is taken into account. In Chapters 2 and

4, the protein compatibility of hFLeu is seen by retention of residue packing and retention of higher order structure for fluorous α_4H analogues. These fluorous analogues have increased $\Delta G°$ values for unfolding, however, values of the thermodynamic terms contributing to $\Delta G°$ are not known. It is thought that, despite being non-biological in origin, fluorinated proteins in general will have similar trends in $\Delta H°$ and $\Delta S°$ to those found in natural, hydrocarbon-containing proteins.

Studies of tFLeu-GCN4-p1d and tFLeu-A1 from the Tirrell laboratory have established the changes in some thermodynamic parameters that stabilize fluorinated proteins[8,9]. Increases in enthalpy, heat capacity and free energy were seen upon incorporation of tFLeu into these model coiled-coil proteins. As with natural proteins, increases in buried hydrophobic content leads to stabilization through concomitant increased free energy. In these studies, the analysis of a single tFLeu mutant protein makes discerning definitive trends for thermodynamic parameters associated with incorporating fluorous residues difficult. To determine how highly fluorinated amino acids influence protein stability it is necessary to define trends by comparing numerous related proteins with similar structural properties.

Thorough analysis of the thermodynamic terms ($\Delta H°$, $\Delta S°$ and $\Delta C_p°$) contributing to fluorous protein stability is currently lacking. The high thermal stability of fluorinated model proteins precludes using calorimetry to measure $\Delta H°$ and $\Delta C_p°$. Measurements of $\Delta G°$ are commonly made by fitting chemical denaturation data to standard models for protein unfolding. This technique monitors protein secondary structure content, as a proxy for the folded proteins, as a function of denaturant concentration assuming $\Delta G°$ is a linear function of denaturant concentration. $\Delta G°$ is an easily comparable value for

overall protein stability, but it doesn't provide information on the thermodynamic parameters contributing to this stability: ΔH°, ΔS°. In this chapter, the influence of fluorous amino acids on the parameters ΔH°, ΔS° and ΔC_p° of a model protein, α_4H is investigated.

		abc***d***efg	***a***bc***d***efg	***a***bc***d***efg	***a***
α_4H	Ac–GN	ADE**L**YKE	**L**ED**L**QER	**L**RK**L**RKK	**L**RSG–NH_2
α_4Ht	Ac–GN	ADE**X**YKE	**X**ED**X**QER	**X**RK**X**RKK	**X**RSG–NH_2
α_4F_2(6,24)	Ac–GN	ADE**X**YKE	LEDLQER	LRKLRKK	**X**RSG–NH_2
α_4F_2(10,20)	Ac–GN	ADELYKE	**X**EDLQER	LRK**X**RKK	LRSG–NH_2
α_4F_2(13,17)	Ac–GN	ADELYKE	LED**X**QER	**X**RKLRKK	LRSG–NH_2
α_4F_3(6-13)	Ac–GN	ADE**X**YKE	**X**ED**X**QER	LRKLRKK	LRSG–NH_2
α_4F_3(17-24)	Ac–GN	ADELYKE	LEDLQER	**X**RK**X**RKK	**X**RSG–NH_2
α_4F_3a	Ac–GN	ADELYKE	**X**EDLQER	**X**RKLRKK	**X**RSG–NH_2
α_4F_3d	Ac–GN	ADE**X**YKE	LED**X**QER	LRK**X**RKK	LRSG–NH_2
$\alpha_4F_3$at	Ac–GN	ADE**X**YKE	**X**ED**X**QER	**X**RK**X**RKK	**X**RSG–NH_2
$\alpha_4F_3af_3d$	Ac–GN	ADE**X**YKE	**X**ED**X**QER	**X**RK**X**RKK	**X**RSG–NH_2
α_4F_6	Ac–GN	ADE**X**YKE	**X**ED**X**QER	**X**RK**X**RKK	**X**RSG–NH_2

L = Leu
X = tBAla
X = hFLeu
X = tFeG

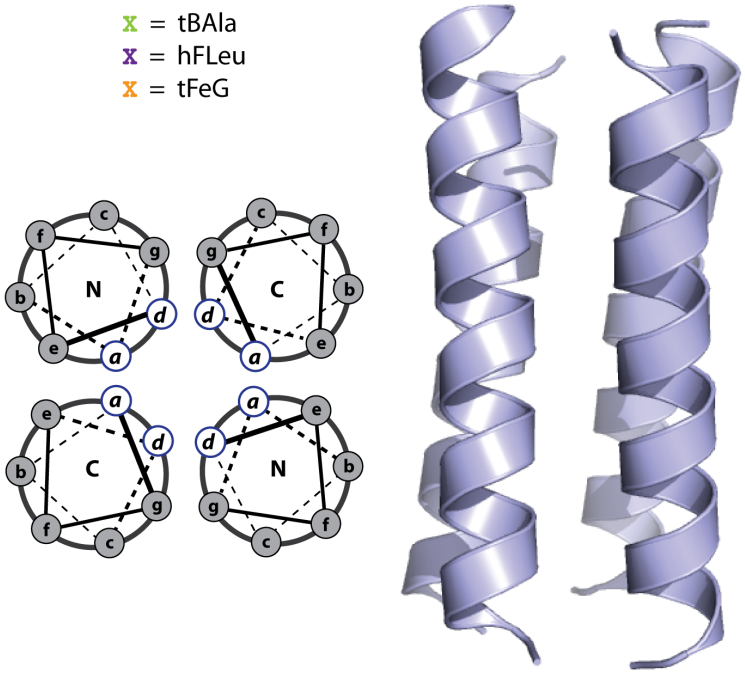


Figure 5.1. *Top*: Sequences of α_4 proteins with hydrocarbon and fluorocarbon analogues of Leu that are substituted at **a** and **d** positions. *Bottom*: Helical wheel diagram of antiparallel 4-helix bundle with hydrophobic residues at **a** and **d** positions and coiled-coil structure.

In this chapter, nonlinear global fitting is used to determine the thermodynamic parameters $\Delta H°$, $\Delta S°$ and $\Delta C_p°$ for 12 different α_4 proteins with varying fluorine content and core packing arrangement. CD was used to monitor protein unfolding as a function of both GuHCl concentration and temperature. Analysis of this data indicate that, much like the conventional hydrophobic effect, there is a linear increase in $\Delta S°$ and $\Delta C_p°$ when hydrophobic content is increased through fluorination. Results from this analysis should help to guide future endeavors towards tailoring the stability of fluorinated proteins.

5.2 – Experimental Procedures

5.2.1 – Materials and Synthesis

L-5,5,5,5′,5′,5′-hexafluoroleucine was synthesized as described in Chapter 2. 4,4,4-trifluoroethylglycine was purchased from SynQuest Laboratory and enzymatically resolved as described in Chapter 6. Boc- and Fmoc-protected β-t-butyl-L-alanine were purchased from AnaSpec Inc. Peptides were synthesized by manual Fmoc procedures (α_4H and α_4Ht) or manual Boc procedures ($\alpha_4F_2(6,24)$, $\alpha_4F_2(10,20)$, $\alpha_4F_2(13,17)$, α_4F_3a, α_4F_3d, $\alpha_4F_3(6\text{-}13)$, $\alpha_4F_3(17\text{-}24)$, $\alpha_4F_3af_3$d, $\alpha_4F_3$at and α_4F_6) as in Chapters 2 and 4.

5.2.2 - Circular Dichroism

CD spectra of protein unfolding were recorded with an Aviv 62DS spectropolarimeter at 222 nm with a 1 mm pathlength cuvette. To examine the unfolding of the peptide by heat, stock solutions were prepared containing 40 μM peptide (concentration of monomer) in 10 mM potassium phosphate buffer, pH 7.0, with varying

amounts of GuHCl. Temperature scans were taken between 4 and 90 °C in steps of 2 °C with a 10 second averaging time.

5.2.3 - Data Fitting

The following describes the two-state equilibrium between the folded tetrameric protein (F) and the unfolded monomeric protein (U), with tetramer dissociation constant, *K*:

$$F \rightleftharpoons 4U \quad (1)$$

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

Rearranging and simplifying the dissociation constant to reduce the quantity of unknown terms allows us to express the concentration of tetramer [F] in terms of total protein [P] and monomer [U]. Noting that [P] = 4[F] + [U]:

$$K(T,[\text{GuHCl}]) = \frac{[U]^4}{[F]} = \frac{4[U]^4}{[P]-[U]} \quad (3)$$

Rearrangement results in the following polynomial:

$$[U]^4 + \frac{K(T,[\text{GuHCl}])}{4}[U] - \frac{K(T,[\text{GuHCl}])}{4}[P] = 0 \quad (4)$$

For fixed [P], it can be shown that given any nonnegative *K*, the above polynomial has a unique solution for [U] between 0 and [P]. This value cannot be expressed in a simple algebraic way, as Equation 4 is a quartic polynomial, but it can be solved numerically. Using Equations 3 and 4, there is a one to one correspondence between *K* and [U].

The Gibbs free energy can then be solved according to the well-known relationship:

$$\Delta G°([\text{GuHCl}]) = \Delta G°(0\ \text{M GuHCl}) - m*[\text{GuHCl}] \quad (5)$$

$$K([\text{GuHCl}]) = \exp\left(\frac{-\Delta G°([\text{GuHCl}])}{RT}\right) = \exp\left(\frac{-\left(\Delta G°(0\ \text{M GuHCl}) - m*[\text{GuHCl}]\right)}{RT}\right) \quad (6)$$

The set of equations listed above describes the determination of folded and unfolded protein concentrations to obtain GuHCl concentration dependent unfolding. In order to fully assess the thermodynamic parameters contributing to protein stability (enthalpy, entropy, and heat capacity), denaturant-dependent unfolding can be combined with temperature-dependent unfolding in a model to globally fit protein unfolding data as a function of two denaturation variables[14]. For these studies of α_4, protein stability as a function of denaturant concentration ([GuHCl]) is modeled by the following relationship:

$$\Delta G°\left(T,[\text{GuHCl}]\right) = \Delta G°(T,0) - m*[\text{GuHCl}] \quad (7)$$

where $\Delta G°(T,0)$ is the protein stability in the absence of GuHCl and the constant of proportionality, m, is the linear dependence of $\Delta G°$ on GuHCl concentration.

Protein stability as function of temperature (T) is given by the Gibbs-Helmholtz equation:

$$\Delta G°(T) = \Delta H° - T\Delta S° + \Delta C_p° * \left(T - T_0 + T\ln\tfrac{T_0}{T}\right) \quad (8)$$

where T is temperature, T_0 is the reference temperature of 25 °C, $\Delta H°$ is the change in enthalpy, $\Delta S°$ is the change in entropy, and $\Delta C_p°$ is the change in heat capacity, all at the reference temperature. Combining these equations gives the protein stability as a function of both GuHCl concentration and temperature:

$$\Delta G°(T,[\text{GuHCl}]) = \Delta H° - T\Delta S° + \Delta C_p° * \left(T - T_0 + T\ln\tfrac{T_0}{T}\right) - m*[\text{GuHCl}] \quad (9)$$

$$K(T,[\text{GuHCl}]) = \exp\left(\frac{-\left(\Delta H^\circ - T\Delta S^\circ + \Delta C_p{}^\circ * \left(T - T_0 + T\ln\frac{T_o}{T}\right) - m*[\text{GuHCl}]\right)}{RT}\right) \quad (10)$$

It has been observed that ΔC_p° and m change little over the measured range of denaturant concentration and temperature and are assumed to be constant[14,15].

Circular dichroism at 222 nm was used to monitor α-helical content as a surrogate for protein stability. Plotting the ellipticity of α_4 proteins as a function of GuHCl concentration and temperature forms a two-dimensional surface with the pre- and post-transition base planes corresponding to the ellipticity of folded protein (θ_f) and unfolded protein (θ_u). The base plane parameters are determined through global fitting analysis with the first term being the ellipticity value of the base plane intercept, the second term describes the linear slope of ellipticity as a function of temperature while the third term describes the linear slope of ellipticity as a function of GuHCl concentration.

$$\theta_u(T,[\text{GuHCl}]) = a + b*T + c*[\text{GuHCl}] \quad (11)$$
$$\theta_f(T,[\text{GuHCl}]) = d + e*T + f*[\text{GuHCl}] \quad (12)$$

It is known that the observed ellipticity will be the sum of the contributions from the unfolded and folded fractions. This is given by Equation (13).

$$\theta_{\text{Obsd}} = \theta_u(T,[\text{GuHCl}])\frac{[\text{U}]}{[\text{P}]} + \theta_f(T,[\text{GuHCl}])\frac{[\text{P}]-[\text{U}]}{[\text{P}]} \quad (13)$$

Both of the base planes are dependent on the temperature and GuHCl concentration, as well as three additional constants each. [P] is constant and by Equation 4, [U] is dependent on K, which by Equation 10 is dependent on temperature, GuHCl concentration, and four constants: ΔH°, ΔS°, ΔC_p°, and m. Thus, θ_{Obsd} is dependent on temperature, [GuHCl], and 10 constants.

To determine the values of those constants, which include the thermodynamic parameters of interest, nonlinear regression was implemented in MatLab. The Levenberg-Marquardt algorithm, a standard way to perform nonlinear least square curve fittings, was utilized. If β is a vector representing all of the desired constants, $f(T,[\text{GuHCl}],\beta)$ is the calculated ellipticity at each point based on the theoretical parameters, and $\theta_{\text{Obsd}}(T,[\text{GuHCl}])$ is the experimental result, then the goal is to find the parameters that minimize the following quantity, where the sum is over all measured data:

$$E(\beta) = \sum \left(f(T,[\text{GuHCl}],\beta) - \theta_{\text{Obsd}}(T,[\text{GuHCl}]) \right)^2 \quad (14)$$

The only input required is the form of the model, given by Equation 13, the experimental data, and initial estimates for the parameters. Base plane estimates were obtained by assuming constant planes and using intercepts just under the minimum and just over the maximum from the experimental data. Using those approximations, the thermodynamic constants were estimated by solving the linear system resulting from Equations 13, 3 and 9.

In addition to finding values for thermodynamic constants, MatLab calculated confidence intervals. By plotting the calculated ellipticity vs. *T* and [GuHCl] as a 2-D surface and comparing with the experimental data, it could be determined if the theoretical model was a good global fit.

5.2.4 – Volume and Surface Area Calculations

Protein surface areas were analyzed using MSMS in Chimera with a probe radius of 1.4 Å corresponding to a water molecule and a vertex density of 10. For nonpolar core

surface area, the surface area of tripeptides of Ala-X-Ala, where X is Leu, tFeG, hFLeu or tBAla, were measured and the difference from Ala-Gly-Ala used. Solvent accessible surface area (ASA) was measured from tetramers of α_4 protein crystal structures. To determine ΔASA upon unfolding, the surface area of four individual α_4 peptides in an extended conformation with β-sheet dihedral angles was measured using MSMS and compared to ASA measurements of the folded tetramers.

5.3 – Results

5.3.1 - Fitting of Denaturation Data

The stability of all 12 α_4 proteins was analyzed by global analysis of thermal and GuHCl denaturation experiments. The signal (θ_{obs}) at 222 nm, corresponding to an α-helix, was recorded for temperatures between 4 °C and 90 °C and 9 – 12 different GuHCl concentrations between 0 M and sufficient concentration to yield complete denaturation at all temperatures studied. The thermal stability of all α_4 proteins has prevented previous determination of thermodynamic parameters by temperature denaturation alone due to an absence of unfolded protein even at high temperature. Global analysis, however, allows data to be fit over a range of both temperature and GuHCl concentration to obtain ellipticity measurements of both pre- and post-transitional baseplanes. Global fits to the temperature and GuHCl dependent denaturation of α_4 proteins are shown in Figures 5.2 and 5.3. Thermodynamic values from fitting data to Equations 3, 9 and 13 are listed in Table 5.1.

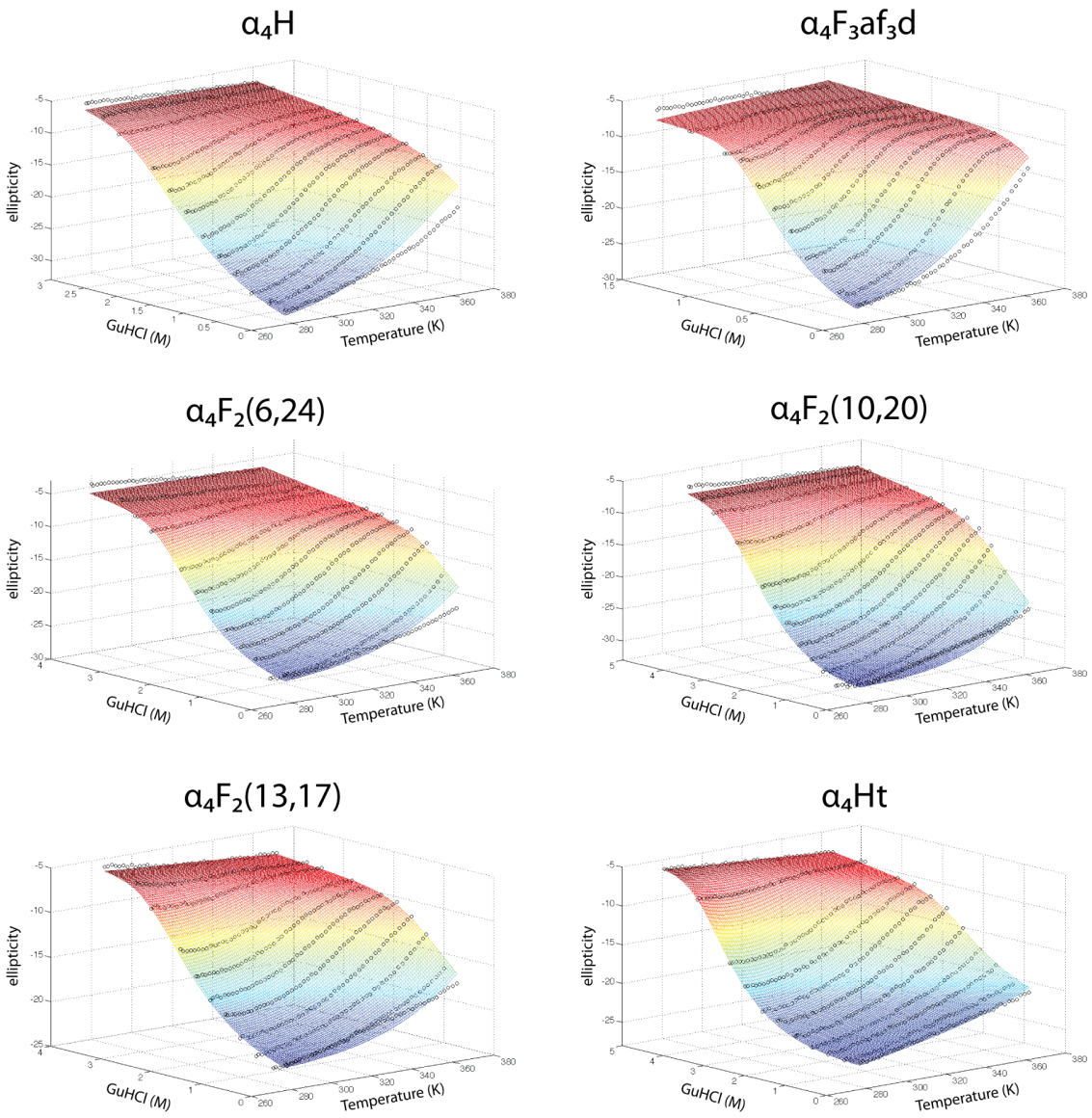


Figure 5.2. Unfolding of proteins with temperature and GuHCl. Fits to the data are represented by a colored surface with blue as the folded base plane and red as the unfolded base plane.

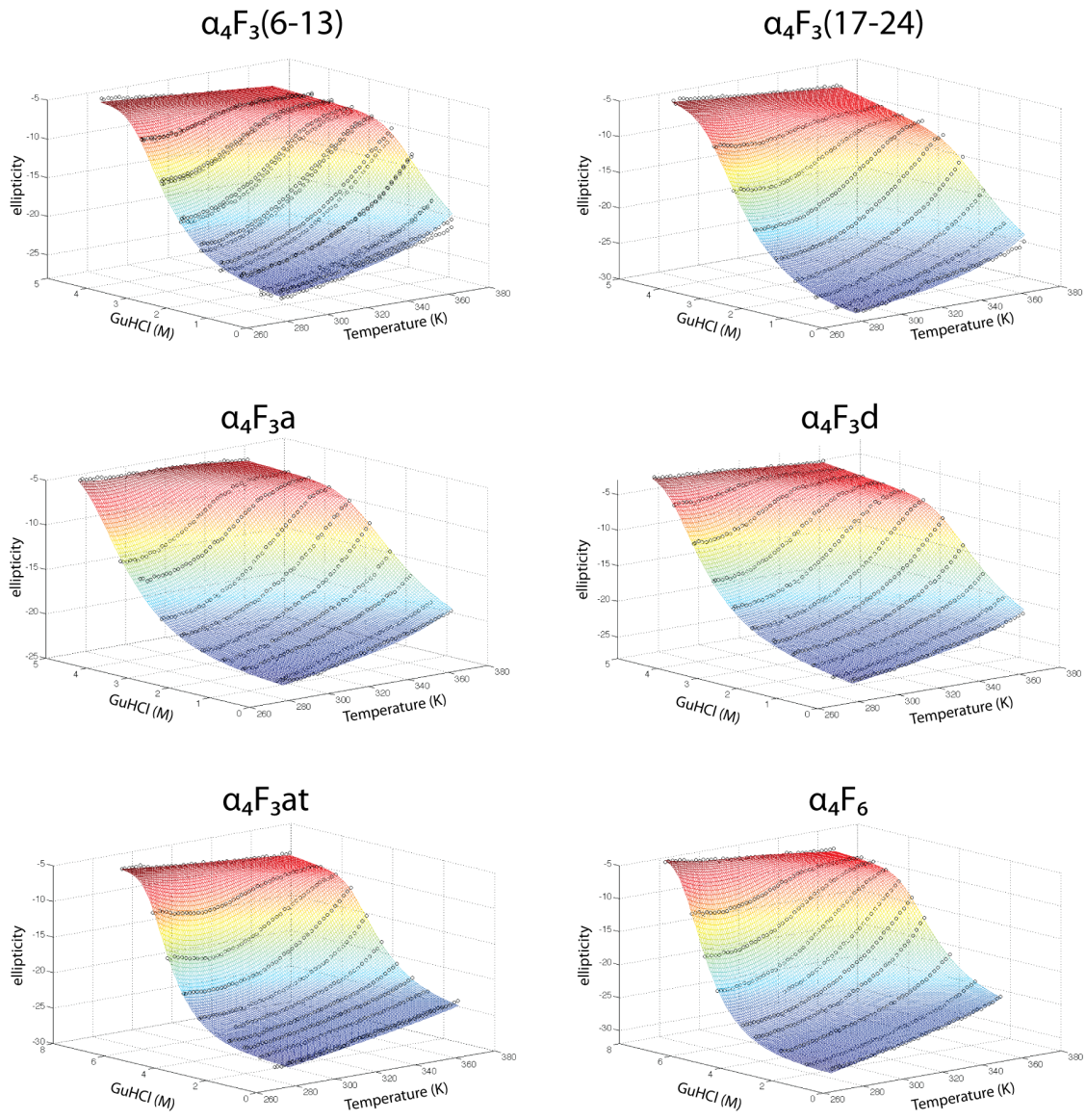


Figure 5.3. Unfolding of proteins with temperature and GuHCl. Fits to the data are represented by a colored surface with blue as the folded base plane and red as the unfolded base plane.

The denaturation surface of the most stable proteins – α_4Ht, α_4F_3(6-13), α_4F_3(17-24), α_4F_3a, α_4F_3d, $\alpha_4F_3$at and α_4F_6 is well-fit by Equation 13. Absence of a sufficient lower base plane, describing signal from folded protein, for the less stable proteins – α_4H, $\alpha_4F_3af_3$d, α_4F_2(6,24), α_4F_2(10,20) and α_4F_2(13,17), necessitated the modification of Equation 12 to estimate the lower base plane (θ_f) as a single point. This approximation of

θ_f results in higher error for fitted parameters, with $\alpha_4F_3af_3d$ showing the greatest error as it denatures at the lowest GuHCl concentration for all proteins and has a very small observable base plane from folded protein.

protein	ΔH° (kcal/mol)	TΔS° (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
α_4F_2(6,24)	16.13±0.62	-5.86±0.52	0.29±0.02	2.51±0.18	21.99±0.81
α_4F_2(10,20)	14.55±0.48	-8.26±0.41	0.37±0.02	2.15±0.05	22.81±0.63
α_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_4Ht	13.17±0.28	-12.56±0.21	0.56±0.02	2.35±0.05	25.72±0.35
α_4F_3(6-13)	19.56±0.61	-6.43±0.38	0.53±0.03	2.71±0.09	25.99±0.72
α_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
α_4F_3at	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

Table 5.1. Summary of the thermodynamic parameters derived from temperature and GuHCl induced unfolding of proteins.

A value for $\Delta G_u°$ is calculated from ΔH° and ΔS° derived by data fitting. Values for $\Delta G_u°$ differ slightly from values previously calculated by monitoring GuHCl induced protein denaturation using CD (Table 5.2). The most noticeable differences are the higher $\Delta G_u°$ values for all proteins except α_4F_3a and α_4F_6. Fitting of the denaturation surface reflects how a protein unfolds in the presence of both heat and chemical

denaturant. Differences in $\Delta G_u°$ between the two methods are likely due to the added temperature dependence on $\Delta G_u°$, which is not taken into account for protein denatured solely by GuHCl at 25 °C. Each protein denaturation data set is comprised of ~500 data points leading me to believe that values resulting from fits of the seven most stable proteins represent accurate measurements of $\Delta H°$, $\Delta S°$ and $\Delta C_p°$ and therefore a more accurate estimate of ΔG_u than that calculated from GuHCl denaturation alone.

The larger *m* values for α_4H, $\alpha_4F_2(6,24)$, $\alpha_4F_2(10,20)$ and $\alpha_4F_2(13,17)$ reported here, as compared to those in Chapter 2, reflect the different method for determining signal attributable to folded and unfolded protein. These proteins have a lower hydrophobic content than the α_4F_3 proteins, and hence, are less stable. Unfolding of these less-stable proteins occurs at comparatively low concentrations of GuHCl, with an accompanying low cooperativity, making determination of the pre-transition (folded protein) signal difficult. For α_4H and $\alpha_4F_3af_3d$, the temperature-dependent denaturation at low GuHCl concentrations displays little cooperativity in unfolding. Thus, fitting a base plane to describe the folded state was not possible and necessitates estimating the signal as a single point with no slope. Approximating the signal due to folded protein for α_4H, $\alpha_4F_3af_3d$, $\alpha_4F_2(6,24)$, $\alpha_4F_2(10,20)$ and $\alpha_4F_2(13,17)$ allows estimation of thermodynamic parameters and although the values for $\Delta G_u°$ and *m* vary somewhat compared to previous results, the same trends in stability are present.

protein	$\Delta G^{\circ}_{unfold}$ (kcal/mol)	m (kcal mol^{-1} M^{-1})
α_4H	18.0 ± 0.2	1.04 ± 0.04
$\alpha_4F_3af_3d$	17.8 ± 1.0	2.78 ± 0.01
$\alpha_4F_2(6,24)$	18.7 ± 0.1	1.18 ± 0.03
$\alpha_4F_2(10,20)$	19.8 ± 0.1	1.22 ± 0.02
$\alpha_4F_2(13,17)$	20.1 ± 0.1	1.53 ± 0.04
α_4Ht	22.3 ± 0.1	1.78 ± 0.03
$\alpha_4F_3(6\text{-}13)$	23.7 ± 0.2	1.98 ± 0.05
$\alpha_4F_3(17\text{-}24)$	23.1 ± 0.1	2.16 ± 0.03
α_4F_3a	27.6 ± 0.1	2.47 ± 0.03
α_4F_3d	26.6 ± 0.1	2.24 ± 0.03
α_4F_3at	28.8 ± 0.1	2.18 ± 0.02
α_4F_6	32.6 ± 0.3	2.38 ± 0.04

Table 5.2 Summary of the thermodynamic parameters derived from GuHCl induced unfolding of proteins as described in Chapters 2 and 4.

5.3.2 - Free Energy

Of greatest interest, is to determine whether the high stability of fluorinated proteins follows the same thermodynamic trends as those of natural, hydrocarbon-containing, proteins or whether different types of stabilizing interactions are present. In Chapter 4, the crystal structures of four fluorinated proteins are described and compared to two non-fluorinated analogues. The structural study shows that there is little preference for distinct fluorous interactions in the folded protein, indicating that highly fluorinated amino acids are compatible with natural, hydrophobic residues, and in fact stabilize a protein in a similar thermodynamic manner to that of natural proteins. The hydrophobic effect has been established as the main factor contributing to the overall stability of a protein. Where the hydrophobic effect in proteins is defined as the favorable transfer of a nonpolar surface away from water into the nonpolar protein core.

The 12 analogues of α_4 provide a unique opportunity to investigate the thermodynamic properties of the hydrophobic effect in fluorinated proteins since all analogues contain identical residues in all except the **a** and **d** positions of the canonical heptad, which provide the major contribution to the hydrophobic effect.

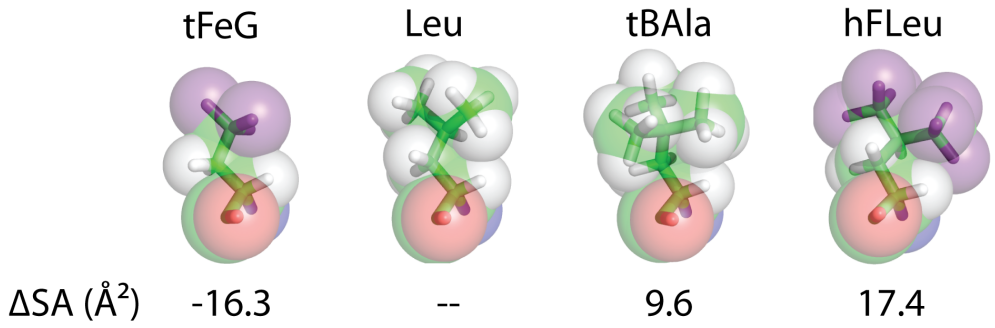


Figure 5.4. Hydrophobic amino acids incorporated into α_4 proteins and resulting changes in surface area as measured from crystal structures using MSMS.

Using the crystal structures of six α_4 proteins, surface area was calculated with MSMS in Chimera. Surface area calculations were made from the crystal structure-based values of Leu, tFeG, tBAla and hFLeu (Fig. 5.4). There are some small differences in the surface area measurements of these proteins compared to literature values as the hydrophobic side chains at **a** and **d** positions include hydrogen atoms to facilitate direct comparison with fluorine substitution.

Plotting $\Delta G_u°$ versus the change in nonpolar surface area contributed by the 24 core residues results in a linear correlation (Fig. 5.5). This trend is similar to what we have observed in previous studies of α_4 proteins. $\Delta G_u°$ correlates well with increasing nonpolar surface area (R = 0.925). Fits of the data yield values of 28.3 cal/mol/Å^2 with no proteins (hydrocarbon or fluorocarbon) significantly deviating. These values lie close to the values determined for energetic contribution of the hydrophobic effect on protein

folding of $\Delta G_u° = 25 – 30$ cal mol^{-1} $Å^{-2}$ determined from numerous studies on hydrophobic small molecules and proteins[16-18].

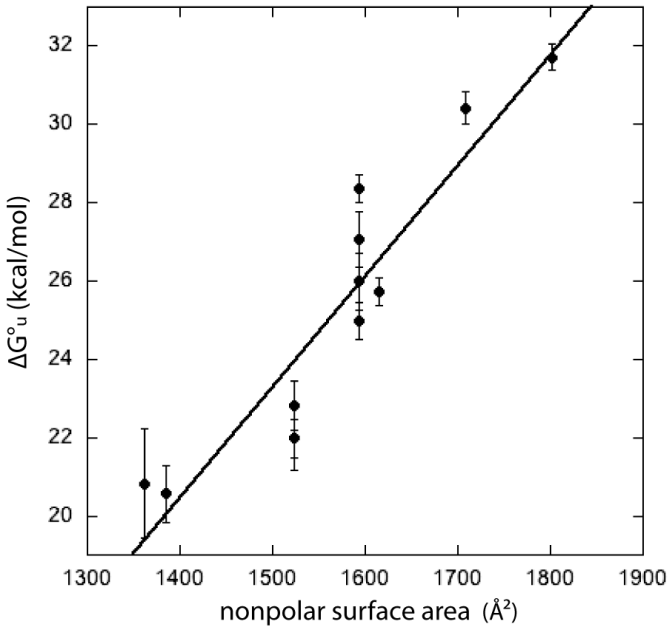


Figure 5.5. Linear correlation of $\Delta G_u°$ from global fitting analysis as a function of increased nonpolar surface area. The slopes are 28.3 cal mol^{-1} $Å^{-2}$ with R = 0.925.

5.3.3 - Enthalpy and Entropy

Analyzing plots of ΔH° versus changes in nonpolar surface area shows that there is no correlation (Fig. 5.6). The linear increase seen for $\Delta G_u°$ with increasing hydrophobic content does not appear to be as a result of increasing ΔH°. There may be a weak association between fluorine content and ΔH° as the five proteins with the highest ΔH° values ($\alpha_4F_3af_3d$, α_4F_3a, α_4F_3d, α_4F_3(6-13) and α_4F_6) contain 72 to 144 fluorine atoms. There are outliers to this trend, as α_4F_3(17-24) and α_4F_3at each have 72 fluorine atoms and relatively low ΔH° values, this lack of a direct correlation makes definitive

conclusions unreliable. These results do indicate that other forces besides hydrophobic interactions are likely responsible for ΔH° values of α_4 proteins.

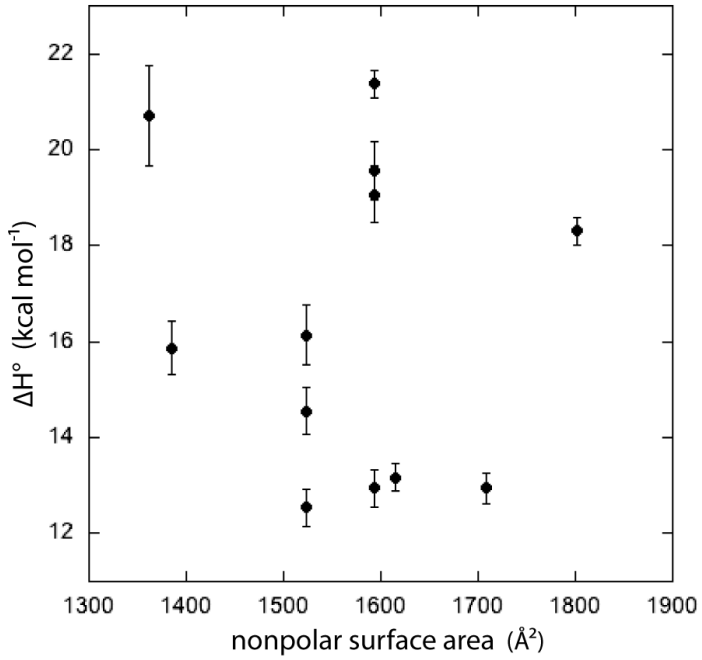


Figure 5.6. Plot of ΔH° from global fitting analysis versus increased nonpolar surface area.

Plots of TΔS° against changes in nonpolar surface area (Fig. 5.7) also reveal a correlation, $R = 0.834$. Notably, the correlation is not quite as strong as that found with $\Delta G_u°$, indicating that a small degree of enthalpy – entropy compensation is present. Since 10 of the 12 proteins contain highly fluorinated residues, this correlation indicates that increasing hydrophobic content, through fluorination, predicatively increases the entropic component of stability. The slope of the linear fit shows an entropy increase due to the increasing hydrophobicity corresponding to 31.6 cal mol^{-1} $Å^{-2}$ (Fig. 5.7) implying that most of the increase in the free energy of unfolding is due to entropic effects. The proteins with the smallest hydrophobic content accordingly have the lowest entropy

contribution, with α_4H having a value of 4.69 kcal/mol and $\alpha_4F_3af_3d$ a value of 0.13 kcal/mol. The nearly zero entropy for $\alpha_4F_3af_3d$ may be explained by structural differences; although this protein is highly α-helical in solution, the crystal structure reveals instability manifested as fraying at the ends of the 4-helix bundle. Fraying of the protein exposes core residues to the solvent, thereby decreasing the difference in entropy between folded and unfolded states. The increase in hydrophobic content stabilizes the folded state due to an entropic penalty of nonpolar compounds interacting with solvent.

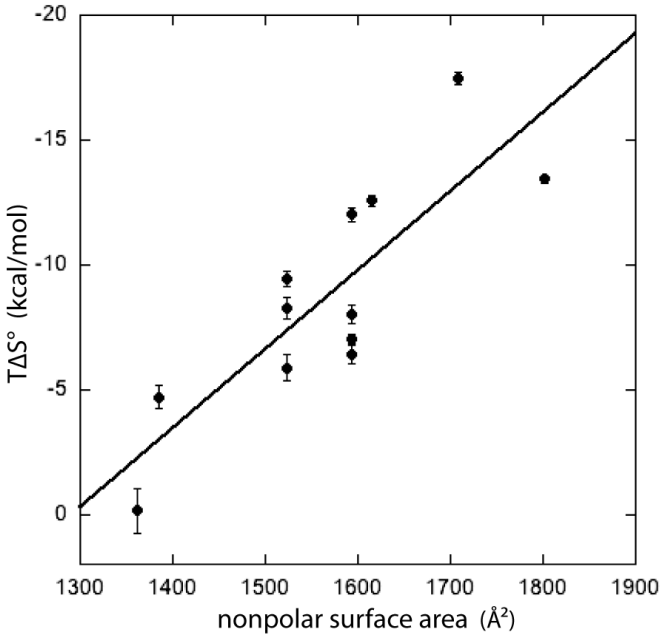


Figure 5.7. Linear correlation of TΔS° from global fitting analysis as a function of increased nonpolar surface area. The slope is 31.6 cal mol^{-1} $Å^{-2}$ with R = 0.834.

Plotting the contributions of ΔH° and TΔS° to overall free energy shows that ΔH° is the main stabilizing factor and remains somewhat constant for most proteins while TΔS° shows a general trend of increasing with hydrophobic content (Fig. 5.8). The contributions from ΔH° and TΔS° to $\Delta G_u°$ appear to compensate. Proteins with similar hydrophobic content but different packing arrangements such as those within the α_4F_2 and

α_4F_3 series of proteins show some degree of compensation between $\Delta H°$ and $\Delta S°$ to maintain a similar $\Delta G°$ value.

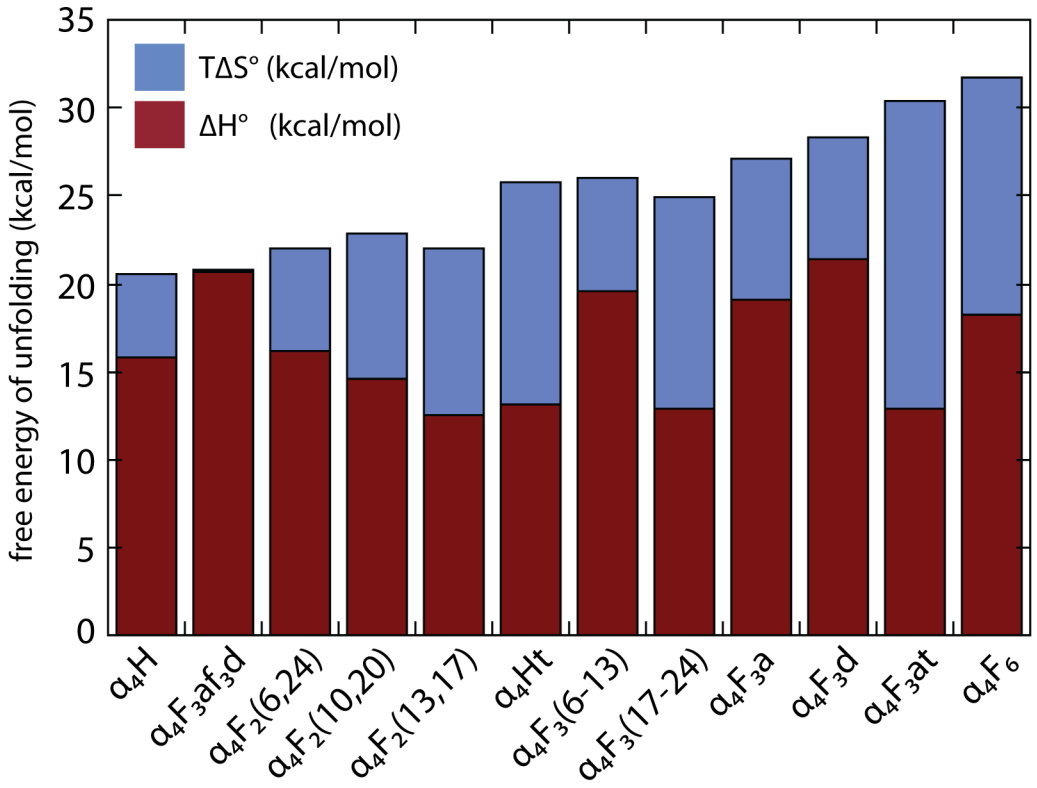


Figure 5.8. Comparison of contributions from entropy ($T\Delta S°$) and enthalpy ($\Delta H°$) to free energy ($\Delta G°$).

5.3.4 - Heat Capacity

The change in heat capacity ($\Delta C_p°$) is particularily informative because changes in both polar and nonpolar surface area upon protein unfolding lead to equivalent changes to $\Delta C_p°$ with opposite signs. A positive value indicates the solvation of hydrophobic side chains while a negative value indicates the solvation of polar side chains. The $\Delta C_p°$ term is largely due to noncovalent interaction differences between the folded and unfolded

states. Noncovalent interactions include interactions between residues within the protein and interactions with the protein and the solvent. The working definition of heat capacity is: the greater amount of heat required to raise the temperature of a solution of unfolded protein than a solution of equal concentration of folded protein. This phenomenon arises from the structuring of water molecules directly interacting with exposed nonpolar groups of unfolded protein.

For natural well-folded proteins, per-residue $\Delta C_p°$ generally lies with the range of 10 – 15 cal mol^{-1} K^{-1} $residue^{-1}$ [19]. However, for the α_4 series of proteins (which all have the same number of residues) the per-residue $\Delta C_p°$ values are rather low and range from 2 – 6 cal mol^{-1} K^{-1} $residue^{-1}$. The low per-residue $\Delta C_p°$ values are one hallmark of a molten globule state[20]. However, the α_4 proteins such as α_4F_6 exhibit well-dispersed amide resonances in ^{15}N-^{1}H-HSQC NMR spectra and exclude hydrophobic dyes such as ANS from their cores – both traits of well folded proteins[5]. Furthermore, with the exception of $\alpha_4F_3af_3d$, the crystal structures of even the least stable proteins show a well-packed hydrophobic core, although the first and last one or two residues are typically nor resolved in the structure implying some fraying of the ends of the 4-helix bundle[21]. However, even if these fraying effects are corrected for, the per-residue $\Delta C_p°$ values are still very low. One possibility is that in solution, even in the folded state, the α_4 proteins are much more fluxional than the crystal structures would suggest.

Calculating $\Delta C_p°$ provides a measure of the amount of nonpolar surface area that becomes exposed upon protein unfolding. Since the folded conformations of α_4 proteins are very similar and the side chains differ, at most, for 6 of the 27 residues, the $\Delta C_p°$ is predominantly a reflection of the hydrophobic content. Examination of $\Delta C_p°$ in Table 5.1

shows a general increase in values going down the column from smallest to largest hydrophobic content. Plotting of $\Delta C_p°$ versus the increase in hydrophobic content shows a correlation of 1.2 cal mol^{-1} K^{-1} $Å^{-2}$ for increasing the core hydrophobic surface area (Fig. 5.9). $\Delta C_p°$ increases fairly linearly (R = 0.856) with increasing nonpolar surface area, which is consistent with the hypothesis that $\Delta C_p°$ is proportional to the change in nonpolar surface area upon unfolding.

To compare this value with those reported in the literature, the change in accessible surface area upon unfolding (ΔASA) was calculated from the six α_4 crystal structures. The $\Delta C_p°$ values corresponding to the change in accessible surface area upon protein unfolding vary greatly between α_4 proteins, which have previously been shown to exhibit similarly folded structures. The area coefficients are 0.04 cal mol^{-1} K^{-1} $Å^{-2}$ for α_4H and $\alpha_4F_3af_3d$, 0.11 cal mol^{-1} K^{-1} $Å^{-2}$ for α_4Ht and 0.10 – 0.13 cal mol^{-1} K^{-1} $Å^{-2}$ for the α_4F_3 protein series. The literature values for area coefficients calculated from other sets of proteins vary quite widely from 0.16 to 0.5 cal mol^{-1} K^{-1} $Å^{-2}$ [19]. It is unclear whether the atypical properties of $\Delta C_p°$ associated with the α_4 proteins arise from fluorination, *per se*, or the de novo-designed 4-helix coiled-coil scaffold. The latter is expected because α_4Ht and $\alpha_4F_3$at which both contain the larger unnatural leucine analog tBAla follow the same trend even though this side-chain contains no fluorine.

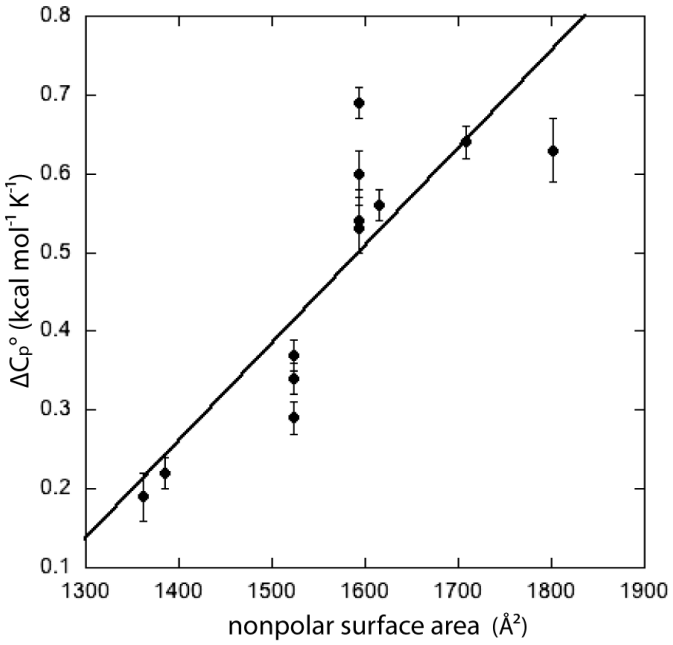


Figure 5.9. Linear correlation of $\Delta C_p°$ from global fitting analysis as a function of increased nonpolar surface area. The slope is 1.2 cal mol^{-1} K^{-1} $Å^{-2}$ with R = 0.856.

5.4 – Discussion

The temperature and GuHCl dependent denaturation profiles of α_4 proteins are well fit to a monomer-tetramer equilibrium. The thermodynamic parameters of $\Delta H°$, $\Delta S°$ and $\Delta C_p°$ have been determined using nonlinear least squares fitting of the global denaturation profiles. This fitting method allows the contribution of enthalpy and entropy to the overall free energy to be determined from a single data set. The increased stability of fluorinated proteins has been discussed in previous chapters, with increase in hydrophobic volume, regardless of fluorine content being the main determinant in fluorous protein stability.

Comparison of the 12 α_4 variants provides a means to determine both how increasing fluorine content and increasing hydrophobic volume influence the

thermodynamics of folding. Studies of α_4 crystal structures described in Chapter 4 indicate that highly fluorinated amino acids have no distinct fluorine-fluorine interaction, which may stabilize the folded state. Also, increases in stability for fluorocarbon amino acids are equivalent to hydrocarbon amino acids when the increased size is taken into account. For this study we determined the thermodynamic parameters that contribute to protein stability in hopes of determining if the physical properties of highly fluorinated amino acids are analogous to those of hydrophobic amino acids found in natural proteins.

Calculating $\Delta G_u°$ from global fitting-derived values of $\Delta H°$ and $\Delta S°$ reconfirmed previous trends discussed in Chapters 2 and 4 of stability increasing as a function of hydrophobic content and not fluorine content. Values of $\Delta H°$ appear to be independent of hydrophobic content or fluorine content. In proteins, $\Delta H°$ is primarily influenced by van der Waals interactions, hydrogen bonding and electrostatic interactions. Differences between $\Delta H°$ values for α_4 proteins are likely from variations in the van der Waals packing of the protein core. Hydrogen bonding and electrostatic interaction is largely between polar residues on the protein exterior and is likely conserved among all analogues. Worth noting is the lack of crystal structures for $\alpha_4F_2(6,24)$, $\alpha_4F_2(10,20)$, $\alpha_4F_2(13,17)$, $\alpha_4F_3(17\text{-}24)$, $\alpha_4F_3$at and α_4F_6 coinciding with their lower values of $\Delta H°$; this may indicate poor packing of the hydrophobic core resulting in loss of van der Waals interactions. α_4Ht also has low $\Delta H°$, this is possibly explained by the large cavity seen in the center of the 4-helix bundle crystal structure in Chapter 4. These measurements indicate that increasing the hydrophobic content, much like in natural proteins, doesn't have a direct influence in $\Delta H°$. The connection between lower $\Delta H°$ values and lack of

crystallographic structural data indicates that a larger enthalpic component may be a requirement for crystal packing of α_4 proteins.

Whether fluorinated proteins stabilize the folded state in accordance with the hydrophobic effect, as seen in conventional proteins, was of utmost interest to this study. Differences in protein hydrophobicity often manifest in thermodynamic terms as differences in $\Delta S°$ and $\Delta C_p°$. The change in entropy of folding showed a general increase when the hydrophobic content of the protein core is increased. Varying the hydrophobic content with fluorous amino acid substitution led to change in both $\Delta S°$ and $\Delta C_p°$ corresponding to the change in hydrophobic surface area. This indicates that for the α_4 proteins and likely for all proteins in general, fluorous amino acid substitutions stabilize the folded state through increasing hydrophobicity thereby increasing entropy.

5.5 – References

(1) Bilgiçer, B.; Fichera, A.; Kumar, K. *J Am Chem Soc* **2001**, *123*, 4393.
(2) Buer, B. C.; de la Salud-Bea, R.; Al Hashimi, H. M.; Marsh, E. N. G. *Biochemistry* **2009**, *48*, 10810.
(3) Chiu, H.-P.; Kokona, B.; Fairman, R.; Cheng, R. P. *J Am Chem Soc* **2009**, *131*, 13192.
(4) Horng, J.-C.; Raleigh, D. P. *J Am Chem Soc* **2003**, *125*, 9286.
(5) Lee, H.-Y.; Lee, K.-H.; Al-Hashimi, H. M.; Marsh, E. N. G. *J Am Chem Soc* **2006**, *128*, 337.
(6) Lee, K.-H.; Lee, H.-Y.; Slutsky, M. M.; Anderson, J. T.; Marsh, E. N. G. *Biochemistry* **2004**, *43*, 16277.
(7) Son, S.; Tanrikulu, I. C.; Tirrell, D. A. *ChemBioChem* **2006**, *7*, 1251.
(8) Tang, Y.; Ghirlanda, G.; Petka, W. A.; Nakajima, T.; DeGrado, W. F.; Tirrell, D. A. *Angew Chem Int Ed* **2001**, *40*, 1494.
(9) Tang, Y.; Ghirlanda, G.; Vaidehi, N.; Kua, J.; Mainz, D. T.; Goddard, W. A.; DeGrado, W. F.; Tirrell, D. A. *Biochemistry* **2001**, *40*, 2790.
(10) Tang, Y.; Tirrell, D. A. *J Am Chem Soc* **2001**, *123*, 11089.
(11) Wang, P.; Fichera, A.; Kumar, K.; Tirrell, D. A. *Angew Chem Int Ed* **2004**, *43*, 3664.
(12) Wang, P.; Tang, Y.; Tirrell, D. A. *J Am Chem Soc* **2003**, *125*, 6900.

(13) Woll, M. G.; Hadley, E. B.; Mecozzi, S.; Gellman, S. H. *J Am Chem Soc* **2006**, *128*, 15932.
(14) Kuhlman, B.; Raleigh, D. P. *Protein Science* **1998**, *7*, 2405.
(15) Yi, Q.; Scalley, M. L.; Simons, K. T.; Gladwin, S. T.; Baker, D. *Folding and Design* **1997**, *2*, 271.
(16) Chothia, C. *Nature* **1974**, *248*, 338.
(17) Eriksson, A.; Baase, W. A.; Zhang, X.; Heinz, D.; Blaber, M.; Baldwin, E. P.; Matthews, B. *Science* **1992**, *255*, 178.
(18) Richards, F. M. *Annual review of biophysics and bioengineering* **1977**, *6*, 151.
(19) Myers, J. K.; Nick Pace, C.; Martin Scholtz, J. *Protein Science* **1995**, *4*, 2138.
(20) Betz, S. F.; DeGrado, W. F. *Biochemistry* **1996**, *35*, 6955.
(21) Buer, B. C.; Meagher, J. L.; Stuckey, J. A.; Marsh, E. N. G. *Proceedings of the National Academy of Sciences* **2012**, *109*, 4810.

Chapter 6

Using Fluorine NMR to Probe the Interaction of Membrane-Active Peptides with the Lipid Bilayer

6.1 – Introduction

Most of work described in this chapter has been published as: Using Fluorine Nuclear Magnetic Resonance To Probe the Interaction of Membrane-Active Peptides with the Lipid Bilayer. Buer BC, Chugh J, Al-Hashimi HM, & Marsh ENG (2010) *Biochemistry* 49(27):5760-5765. Co-authors were very helpful in conducting this research and analyzing the results, Dr. Jeetender Chugh assisted me with NMR experiments and carried out theoretical calculations. Numerous suggestions and guidance on NMR experiments came from Prof. Hashim Al-Hashimi and Dr. Jeetender Chugh.

Antimicrobial peptides (AMPs) comprise a diverse family of membrane-active peptides that are found in essentially all multi-cellular organisms. They are components of the innate immune system and in higher organisms are also implicated in the activation of the adaptive immune response against infection. Although some AMPs have specific intracellular targets [1], most exert their antimicrobial activity by binding directly to the membrane and compromising its integrity [2,3]. AMPs are one class of a growing number of membrane-active peptides that include anti-cancer and anti-viral peptides, cell-penetrating peptides, viral fusion peptides and venom peptides. For all these classes of

peptides, interactions between the membrane lipid bilayer and the peptide are central to their biological functions.

Although highly diverse in sequence and structure, almost all AMPs share the property of being highly amphipathic, with one face of the peptide being hydrophobic and the other face presenting a cluster of positively charged residues [4-6]. The selectivity of AMPs for bacterial membranes arises primarily from electrostatic interactions between the positively charged peptide and the negatively charged phospholipids that predominate in bacterial cell membranes. Eukaryotic membranes, which contain predominantly neutral phospholipids, are usually less susceptible to disruption by AMPs; the presence of cholesterol in eukaryotic membranes also helps prevent membrane disruption by AMPs [7]. Upon association with the bacterial membrane, membrane disruption may proceed through a number of mechanisms, including the formation of pores, membrane thinning and detergent-like action [8,9]. The mechanism of membrane binding and subsequent destabilization is outlined in Figure 6.1.

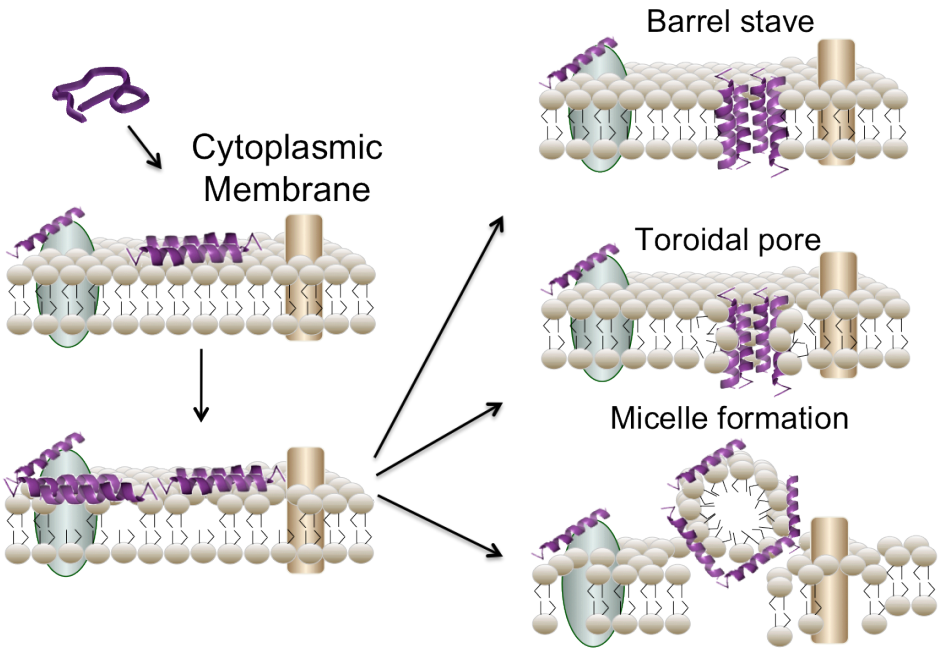


Figure 6.1. Antimicrobial peptides (AMPs) are potent therapeutic agents that act by disrupting bacterial membranes. Membrane disruption by AMPs is initiated by attraction of the positively charged peptide with the negatively charged bacterial membrane lipid headgroups. Loss of membrane integrity may result from three distinct pore forming mechanisms.

Characterizing peptide-membrane interactions is challenging because of the transient nature of these interactions. Furthermore, the peptide may adopt different orientations with respect to the lipid bilayer and different oligomerization states that are concentration-dependent. Detailed structural models derived from solid state NMR data are available for some AMPs bound to lipid membranes [10-15]. These experiments require peptide concentrations that are orders of magnitude higher than their physiologically active range, so it is not clear whether such structures represent biologically active species.

Studies in the Marsh laboratory have previously investigated the effect of incorporating extensively fluorinated amino acids into both α-helical (MSI-78) [16] and β-sheet (protegrin) AMPs [17]. Fluorination can be used to modulate the biological properties and membrane interactions of these AMPs; in particular fluorination of MSI-78 resulted in increased potency towards some bacterial strains and protection against proteolysis when bound to lipid vesicles. Other research groups have also demonstrated that fluorination is an effective strategy to modulate bio-active peptides [18-20], these experiments are discussed in more detail in Chapter 1 section 5.

In this chapter, I discuss using fluorine-containing peptides to probe the interaction of AMPs with membranes by exploiting the sensitive NMR properties of the ^{19}F nucleus. The introduction of fluorine atoms into biological molecules is usually minimally perturbing to structure and function, and fluorine NMR has several advantages

for studying biomolecular interactions [21,22]. These include: 100 % natural abundance and high intrinsic sensitivity - 83 % that of proton NMR; high sensitivity of chemical shift to local environment; and no background signal because fluorine is absent from biological molecules.

In this study, MSI-78 analogs were synthesized that incorporate trifluoromethyl groups within their sequence to probe the interaction of the peptide with lipid bilayers. These studies show that ^{19}F solution NMR can readily be used to detect peptide binding to lipid bilayers and that furthermore the chemical shift change upon binding is sensitive to the position of the probe within the peptide. CPMG experiments were performed and proved crucial in examining the dynamics of the peptide interacting with the membrane. Increasing the signal of the fluorine probe was also investigated by incorporating perfluoro-*t*-butyl groups into the sequence of MSI-78 analogues.

6.2 – Experimental Procedures

6.2.1 - Peptide Preparation

Racemic trifluoroethylglycine (tFeG) was purchased from SynQuest Labs and resolved according to the protocol by Tsushima *et al.* [23]. TfeG was first acylated by slowly adding 2 equivalents acetic anhydride to the amino acid dissolved in 1 M NaOH. The products were extracted by ethyl acetate and dried. L-tFeG was enzymatically resolved from *N*-acyl-D,L-tFeG by porcine kidney acylase I. The enzymatic reaction was buffered with acetic acid/LiOH and allowed to proceed at 37 °C for ~4 hours until deacylation of the L- form was judged complete by ^{19}F NMR. The resulting solution was filtered and L-tFeG separated from *N*-acyl-D-tFeG using an ion-exchange column packed

with ~30 mL Dowex 50X8-200 ion-exchange resin. L-tFeG was identified by ninhydrin staining of column fractions and judged to have >99% ee by derivatization with Mosher's acid and subsequent ^{19}F NMR. The pure amino acid was converted to its Boc- derivative by standard procedures. The sequences of MSI-78 derivatives are shown in Figure 6.1. All peptides were synthesized by manual Boc procedures on MBHA resin as described previously in Chapter 2 [24,25]. The use of Boc SPPS for MSI-78 derivatives was prompted by the observed failure of Fmoc protocols to synthesize hFLeu containing peptides as discussed in previous chapters. However, Fmoc SPPS has been shown to readily couple all residues of tFeG containing MSI-78 analogues, as described in Appendix 1.

All peptides were purified on a Waters preparatory HPLC using a linear gradient of 95% water, 4.9% acetonitrile and 0.1% TFA for solvent A and 9.9% water, 90% acetonitrile and 0.1% TFA for solvent B, with a flow rate of 10 mL/min on a Waters C18 preparatory column. After lypholization, peptides were dissolved in water to 20 mg/mL and residual TFA was removed by passing this solution through a Stratosphere SPE column (Varian) which was conditioned with 5 mL 50:50 water/MeOH. The column was then rinsed with 4 mL methanol and 10 equivalents of formic acid were added to the peptide solution before lyophilization. Stock peptide concentrations were determined using ^{19}F NMR with a known concentration of TFA as an internal reference. Peptide identities were confirmed using MALDI-MS with a matrix of α-cyano-4-hydroxycinnamic acid.

6.2.2 - Lipid Preparation

1-Palmitoyl-2-oleoyl-*sn*-glycero-3-phosphatidylcholine (POPC) and 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phospho-(1'-*rac*-glycerol) (POPG), 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC), 1,2-dimyristoyl-*sn*-glycero-3-phospho-(1'-*rac*-glycerol) (DMPG) and 1,2-dihexanoyl-*sn*-glycero-3-phosphocholine (DHPC) were purchased from Avanti Polar Lipids. Fresh POPC/POPG (3:1) SUVs were prepared in PBS buffer, pH 7.4 with 10% D_2O. To make SUVs, multilamellar liposomes were sonicated to clarity using Fischer Scientific 550 sonic dismembrator, centrifuged to remove insoluble particulate and used immediately. Isotropic bicelles were made in PBS buffer, pH 7.4 with 10% D_2O by adding a solution of 3:1 DMPC/DMPG to a solution of DHPC giving q=0.5 resulting in a clear, non-viscous solution.

6.2.3 - Circular Dichroism

To examine the secondary structure, CD spectra of peptides were recorded with an Aviv 62DS spectropolarimeter at 25 °C. Spectra were taken of 100 μM peptide in buffered solution and 100 μM peptide in the presence of 100mM SDS micelles. Mean residue ellipticities, $[\theta]$, were calculated using Equation 1:

$$[\theta] = \theta_{obsd}/10lcn, \qquad (1)$$

where θ_{obsd} is the ellipticity measured in millidegrees, c is the molar concentration, l is the cell path length in cm and n is the number of residues in the protein.

6.2.4 - MIC Determinations

The peptide minimum inhibitory concentrations (MIC) against *E. coli* K12 were determined by the microdilution antimicrobial assay procedure, using 96-well plates in

replicates of four [26]. A 5 mL overnight culture of *E. coli* was grown in 2xYT media and transferred to 100 mL 2xYT media where it was grown to an OD_{600} of 0.4. Cells were then diluted 1:800 in 2xYT media to an OD_{600} of 0.0005 and 100 μL of the OD_{600} 0.0005 media was added to each well of a sterile 96-well plate. Peptides were dissolved to 400 μg/mL in PBS buffer and 100 μL peptide solution was added to the 100 μL of the OD_{600} 0.0005 media in the first well of each row giving a final peptide concentration of 200 μg/mL. 100 μL of the first well was then transferred to the second well, effectively diluting the peptide concentration to 100 μg/mL repeating this with each subsequent well allowed testing of 10 different peptide concentrations, ranging from 200 μg/mL to 0.39 μg/mL. Plates were incubated overnight at 37 °C and checked for growth by eye.

6.2.5 - ^{19}F NMR

All ^{19}F NMR experiments were performed at 30 °C using a Varian Inova 400 MHz NMR spectrometer equipped with a double-tuned ^{1}H-^{19}F room-temperature probehead. Peptide and lipid samples were prepared with 10% D_2O in PBS, pH 7.4. All experiments were performed at a constant peptide concentration of 400 μM unless indicated otherwise and referenced to trifluoroacetate ion at 0 ppm. ^{19}F CPMG relaxation dispersion experiments were performed for the two peptides in the free state and in the presence of lipid bicelles (200 mM total lipid concentration, q=0.5, long chain lipids 3:1 mol/mol DMPC/DMPG and short chain lipid being DHPC). CPMG delays (τ_{cp}) were varied from 0.2 to 2.0 ms with each data point recorded as a series of standard 1-D transverse relaxation rate measurements with T_2 delays of 0.05, 0.1, 0.2, 0.4, 0.8 and 1.6 ms for free peptide and 0.0125, 0.025, 0.05, 0.1, 0.2 and 0.4 ms for bicelle-bound peptide.

Data sets were recorded with acquisition time of 1 s in T_1 along with a 10 s pre-scan delay and 512 scans for a net acquisition time of 2.3-4.6 h/data point. Data was processed and analyzed in VNMRJ and plotted in Origin 8.0.

6.2.6 - Theoretical R_2 Calculations

Intrinsic R_2 values for free peptide and peptide bound to bicelles were calculated using Equation 2 below:

$$\frac{1}{T_2} = R_2 = \frac{d^2}{8}\left[4J(0) + J(\omega_F - \omega_C) + 3J(\omega_C) + 6J(\omega_F) + 6J(\omega_F + \omega_C)\right] + \frac{c^2}{6}\left[4J(0) + 3J(\omega_C)\right] \quad (2)$$

where $d = \frac{\mu_0}{4\pi}\gamma_F\gamma_C\frac{h}{2\pi}\left(r_{CF}^{-3}\right)$ and $c = \frac{\omega_C\left(\sigma_{\parallel} - \sigma_{\perp}\right)}{\sqrt{3}}$.

$J(\omega)$ in Equation 2 is defined as the spectral density at frequency ω and was calculated using Equation 3.

$$J(\omega) = \frac{2}{5}\left(\frac{S^2\tau_m}{1+\tau_m^2\omega^2} + \frac{\left(1-S^2\right)\tau}{1+\tau^2\omega^2}\right) \quad (3) \quad \text{where } \frac{1}{\tau} = \frac{1}{\tau_m} + \frac{1}{\tau_e}$$

In which μ_0 is the permeability of free space; γ_F and γ_C are gyromagnetic ratios of ^{19}F and ^{13}C, respectively; h is Planck's constant; r_{CF} is the C-F bond length which was taken to be 1.35 Å; ω_C and ω_F are Larmor frequency of ^{13}C and ^{19}F, respectively; $\left(\sigma_{\parallel} - \sigma_{\perp}\right)$ is the CSA of ^{19}F spin and was taken to be 52 ppm [27] (This is the value measured for 5,5,5-

trifluoroleucine, the system closest in structure to tFeG for which CSA has been measured.); τ_m and τ_e are correlation times for global and internal motions, respectively; τ_m is assumed to be 1 ns for free peptide and 20 ns for peptide bound to bicelles [28] and τ_e is assumed to be 1 ps; S^2 is the order parameter which defines the amplitude of the motions and is assumed to be 0.85 for peptides and bicelles.

6.3 – Results

MSI-78 is thought to disrupt bacterial membranes by forming toroidal pores in the lipid bilayer [29]. The peptide has been shown in NMR structural studies to adopt a dimeric α-helical coiled-coil structure in which the dimer interface is formed by contacts between hydrophobic residues, and the positively charged lysine residues face the exterior of the structure and interact with hydrophilic lipid head groups [30]. Using this structural model, shown in Figure 6.2, as a guide, variants of MSI-78 in which Leu-6 (MSI-F6) or Lys-7 (MSI-F7) were substituted with trifluoroethylglycine (tFeG) were sythesized. This introduces the CF_3 reporter group into, respectively, the hydrophobic core of the peptide and the positively charged exterior.

MSI-78 GIGKFLKKAKKFGKAFVKILKK-NH_2
MSI-F6 GIGKFXKKAKKFGKAFVKILKK-NH_2
MSI-F7 GIGKFLXKAKKFGKAFVKILKK-NH_2

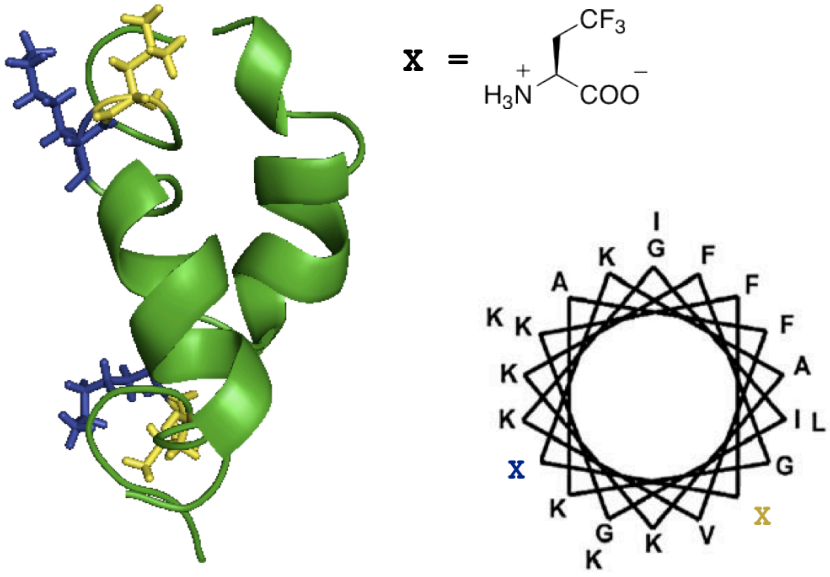


Figure 6.2. *Top*: Primary sequence of MSI-78 with substitution of Leu in MSI-F6 and Lys in MSI-F7 for trifluoroethylglycine (tFeG). *Bottom*: Structure of MSI-78 dimer in DPC micelles showing the position of amino acid substitutions. Helical wheel diagram illustrating the amino acid substitutions on the hydrophobic and hydrophilic face of MSI-78. Leu substitution of MSI-F6 shown in yellow and Lys substitution of MSI-F7 shown in blue.

6.3.1 - Conservation of MSI-78 Structure and Activity

Substitution of Leu or Lys by tFeG was considered to be nonperturbing to the membrane-bound structure of MSI-78 because the tFeG side chain is relatively small. Experiments described in Chapter 4 demonstrated tFeG to be compatible with the protein environment of α_4; the tFeG containing peptide, $\alpha_4F_3af_3d$, had a smaller hydrophobic core and decreased stability compared to α_4F_3a, however, it was highly α-helical under non-denaturing conditions. Substituting one residue for tFeG likely has little influence on helical structure: $\alpha_4F_3af_3d$ incorporated three tFeG residues per chain and was still a well-folded α-helical protein. MSI-78 binds bacterial lipids through charge – charge

interactions, for MSI-F7 the Lys substitution results in a 10% loss of overall positive charge. Decreased positive charge in all likelihood doesn't decrease peptide affinity towards lipid as the parent peptide, magainin 2, has a +5 charge – half that of MSI-78. Incorporation of single tFeG residues into MSI-78 does not appear to cause any gross structural changes to the peptides. Both peptides are unstructured in buffered solution and exhibit extensively α-helical CD spectra in the presence of SDS micelles as shown in Figure 6.3. Both peptides also exhibit MIC values (~4 μg/mL) against *E. coli* K12 strains that are similar to that of MSI-78 [16].

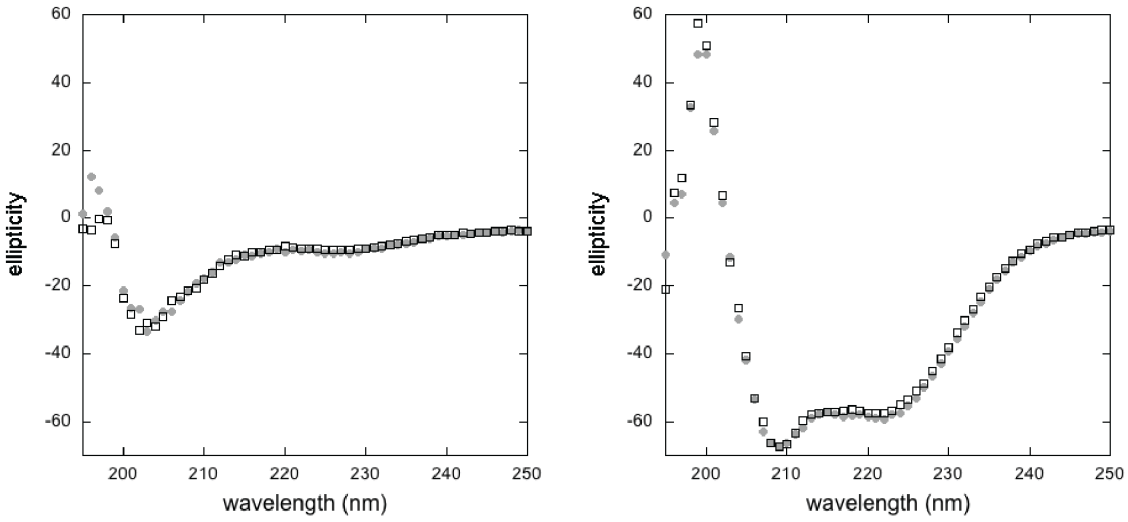


Figure 6.3. CD spectra of 100 μM MSI-F6 (●) and MSI-F7 (□) in buffered solution (*left*) and in the presence of 100 mM SDS micelles (*right*).

6.3.2 – ^{19}F NMR of Peptide Binding Lipid Vesicles

The interaction of the peptides with small unilamellar vesicles (SUVs) were first investigated; SUVs are often used as a model membrane surface. In free solution at pH 7.4 both MSI-F peptides are unstructured and their ^{19}F NMR spectra exhibit a well resolved triplet at 11.64 ppm relative to the TFA internal standard (Fig. 6.4). As the

peptides are titrated with increasing concentrations of SUVs, a new, broadened up field signal is observed due to the bound peptide. Notably, on binding SUVs the chemical shifts of MSI-F7 and MSI-F6, become significantly different, which implies that the CF_3 groups are sampling different chemical environments. The signal of MSI-F7 moves upfield by only 0.2 ppm, indicating that the environment of the CF_3 probe at position 7 does not change much. In contrast, the signal from MSI-F6 moves upfield by more than 1.0 ppm, suggesting that the CF_3 probe at position 6 experiences a more significantly hydrophobic environment. The degree of peak broadening is also different for the two peptides, with the signal due to MSI-F6 becoming more broadened upon binding to the vesicles.

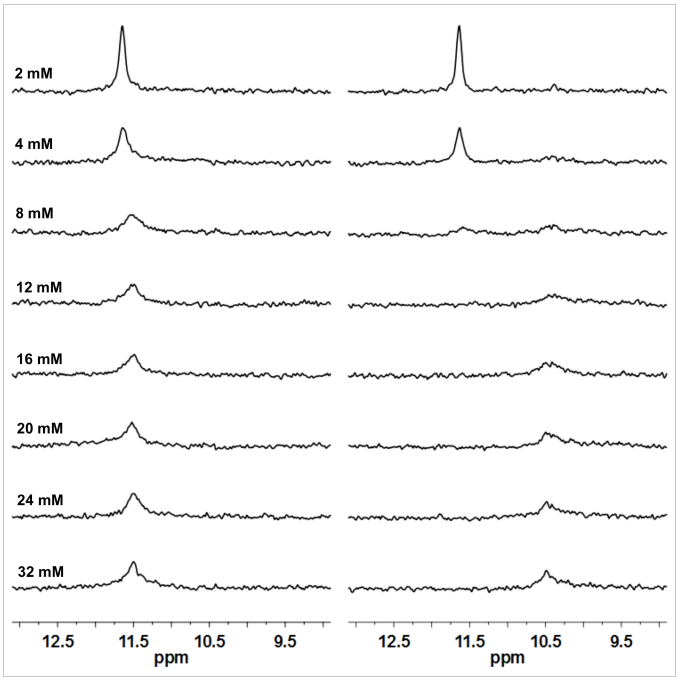


Figure 6.4. ^{19}F NMR spectra of MSI-F7 (*left*) and MSI-F6 (*right*) in the presence of increasing concentrations of SUVs. Spectra recorded at 30 °C at pH 7.4 in PBS buffer with 10% D_2O.

These observations are consistent with the CF_3 group in MSI-F7 occupying a position at the interface of the peptide with the lipid membrane surface where it may be expected to be in a more polar environment and somewhat more mobile. In contrast, the CF_3 group in MSI-F6 should be buried within the hydrophobic interface of the peptide dimer where its mobility would likely be more restricted.

Titration of the peptides with SUVs also provided an estimate for an apparent dissociation constant for the peptide-membrane complex of approximately 5 mM (assuming 60% of the lipid molecules in the SUVs are available for binding on the outer leaflet of the lipid bilayer).

6.3.3 - ^{19}F NMR T_2 Analysis of Peptide Binding

The presence of lipids results in a marked broadening of the peak due to the *unbound* peptide (Fig. 6.4), suggesting that chemical exchange between the bound and free peptides is occurring on the NMR timescale. Therefore to investigate the dynamics of the peptide interacting with the lipid bilayer the transverse relaxation times (T_2) of the CF_3 reporter nuclei were measured as a function of the CPMG pulsing rate ($1/\tau_{cp}$). The relaxation rate, R_{2obs}, as a function of τ_{cp} is given by Equation 4:

$$R_{2obs} = \frac{1}{T_{2obs}} = \frac{X_f}{T_{2f}} + \frac{X_b}{T_{2b} + \tau_b} + \tau_b X_b X_f (\delta\omega)^2 \left[1 - \frac{2\tau_b}{\tau_{cp}} \tanh\left(\frac{\tau_{cp}}{2\tau_b}\right)\right] \qquad (4)$$

Where X_f and X_b are the mole fractions of the free and bound peptide; T_{2f} and T_{2b} are the transverse relaxation times for free and bound peptide and τ_b is the residence time for the

peptide bound to the lipid and $\delta\omega$ is the difference in chemical shifts between the bound and free peptides. For $\tau_{cp} << \tau_b$ the chemical exchange contribution to T_2 is removed and R_{2obs} is independent of τ_{cp}. As τ_{cp} increases to the point that $\tau_{cp} \sim \tau_b$, then $R_{2\text{obs}}$ will increase as chemical exchange begins to contribute to relaxation.

SUVs have a number of drawbacks for NMR experimentation, including broad signal from bound peptide as a result of the large lipid aggregate size and low tumbling rate. SUVs also proved insufficiently stable on the longer timescale needed to perform the CPMG measurements. Therefore, lipid bicelles, another commonly used model membrane system, were used to examine the binding of MSI-F6 and MSI-F7. Bicelles have the advantage of being more stable than SUVs, although higher concentrations of lipid are needed to form them. Interestingly, compared with SUVs the MSI-F6 peptide exhibits a much greater change in chemical shift on binding to bicelles, shifting upfield by ~0.4 ppm. Similarly, the signal due to MSI-F6 shifts upfield by 1.6 ppm on binding to bicelles (Fig. 6.5).

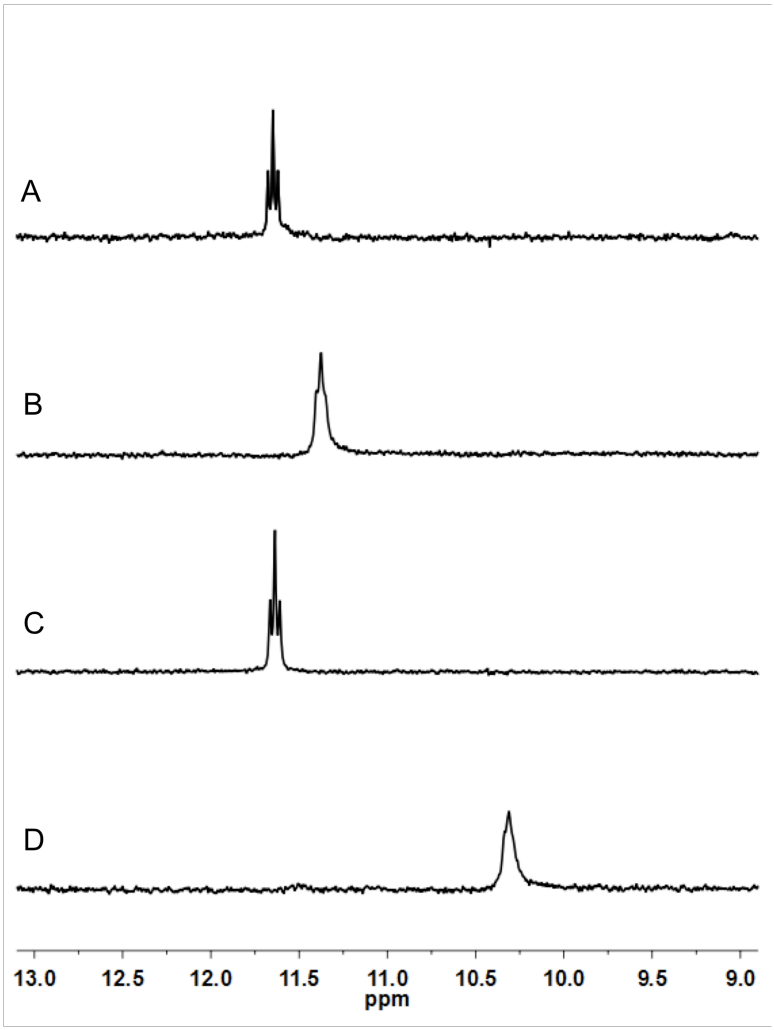


Figure 6.5. ^{19}F NMR spectral changes associated with MSI-F7 and MSI-F6 binding to bicelles. (A) 400 μM MSI-F7, (B) 400 μM MSI-F7 in the presence of 200 mM bicelles, (C) 400 μM MSI-F6, (D) 400 μM MSI-F6 in the presence of 200 mM bicelles. Spectra were recorded at 30 °C at pH 7.4 in PBS buffer with 10% D_2O and referenced to TFA.

The R_2 (= $1/T_2$) values were measured for both peptides, in free solution and bound to bicelles, for $1/\tau_{cp}$ ranging from 100 – 2000 Hz (Fig. 6.6). In free solution both peptides are characterized by R_{2f} ~3 Hz which, as expected, is independent of the CPMG pulsing rate and is consistent with the CF_3 group being highly mobile in the unstructured peptide. The experiment was then performed with MSI-F6 and MSI-F7 bound to high concentrations of lipid bicelles, so that contribution to R_{2obs} due to free peptide is minimal

and thus $R_{2obs} = R_{2b}$. For both MSI-F6 and MSI-F7 R_{2b} remains constant for pulsing rates between 200 and 2000 Hz, and only at longer pulse intervals does R_{2b} appear to increase. However, at these timescales the experiment approaches the limits of sensitivity and measurements are consequently accompanied by large uncertainty in the value of R_{2b}. Although the data cannot be reliably fitted to Equation 6.4 to allow τ_b to be calculated, the experiment does allow an upper limit to be put on the rate at which the peptide dissociates from the membrane ($1/\tau_b$) of ~200 s^{-1}.

At high CPMG pulsing rates R_{2b} for the peptides represents the intrinsic relaxation rate of the ^{19}F nucleus with the chemical exchange component removed. The difference between R_{2b} for MSI-F7, R_{2b} ~18 Hz, and MSI-F6, R_{2b} ~35 Hz, is significant and may be attributed to differences in the dynamics of the peptide at the two positions monitored by the probe. To gain insights into these differences, the R_2 values were calculated, as described in section 2.5, for the peptides in free solution and when bound to bicelles.

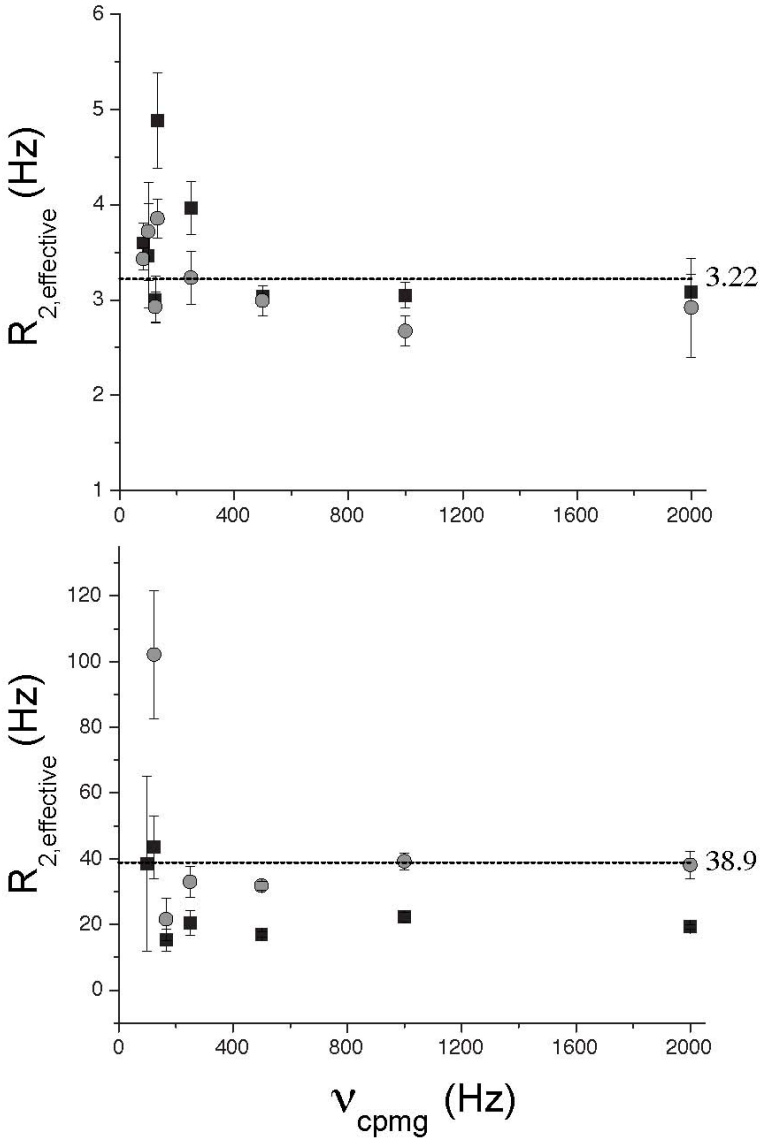


Figure 6.6. Observed transverse relaxation rate, $R_{2,effective}$, plotted as a function of CPMG pulsing rate (ν_{cpmg}). Data are for 400 μM MSI-F7 (■) and MSI-F6 (●) in buffered solution as unstructured peptides (*top*) and with 200 mM bicelles (bottom). The calculated R_2 values for free peptide and peptide bound to bicelles are indicated by the dashed lines.

In free solution the calculated value for R_{2f} = 3.22 Hz, which assumes the peptides are unstructured, agrees very well with the experimentally determined R_{2f} for MSI-F6 and MSI-F7 (Fig. 6.6). Interestingly however, in the bound state whereas the calculated value for R_{2b} = 38.9 Hz is in excellent agreement with that measured for MSI-F6, it is

significantly larger than that measured for MSI-F7 (Fig. 6.6). The lower R_{2b} for MSI-F7 suggests that the CF_3 probe at position 7 is more mobile than if it were tumbling at the correlation time of the bicelle, although not as mobile as in free solution. These observations are in accord with a model that places the CF_3 group in MSI-F7 close to the solvent exposed lipid head groups, where the side-chains may be expected to exhibit greater conformational mobility. In contrast, the model places the CF_3 group in MSI-F6 in the hydrophobic core of the peptide dimer; here the side-chain is immobile and R_{2b} is dominated by the tumbling motion of the bicelle.

6.4 – A Fluorinated Amino Acid with Increased NMR Signal

Using tFeG as an NMR probe has advantages and disadvantages; the side chain is small and nonperturbing to the structure of MSI-78 thusly allowing retention of biological activity, however, having only 3 chemically equivalent fluorine atoms which are coupled to adjacent protons limits the useable concentration realm for some biological studies. As a means to overcome the signal to noise disadvantages of tFeG, the novel amino acid *O*-perfluoro-*t*-butyl-L-homoserine (nFhSer) was synthesized. This amino acid incorporates 9 chemically equivalent fluorine atoms into the side chain, thereby increasing the ^{19}F NMR signal sixfold over that of tFeG.

6.4.1 - Synthesis and Purification

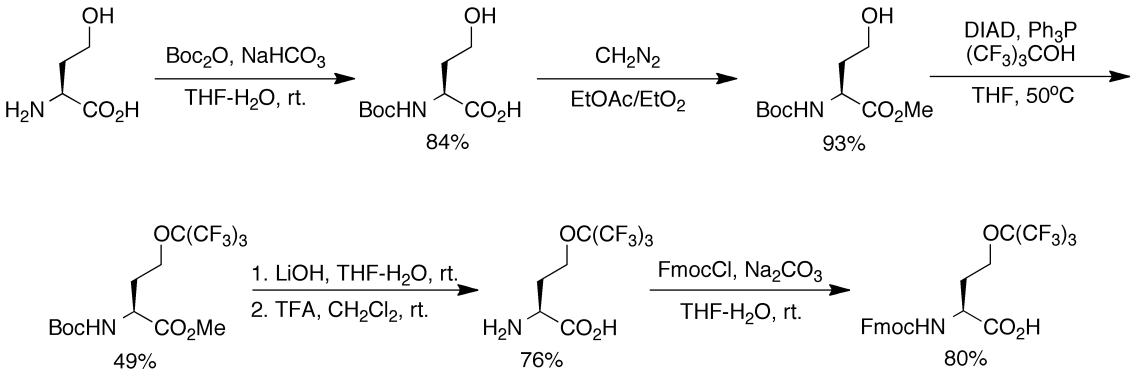


Figure 6.7. Synthetic route of Fmoc-protected *O*-perfluoro-*t*-butyl-L-homoserine (nFhSer) starting with L-homoserine.

Synthesis of *O*-perfluoro-*t*-butyl-L-homoserine takes advantage of the Mitsunobu reaction[31,32] (Fig. 6.7), which is routinely used to convert an alcohol to an ether provided the pKa of the reactant alcohol is sufficiently low (≤ 10). To compare the properties of nFhSer to tFeG, the residue was incorporated into two MSI-78 analogues, MSI9-F6 and MSI9-F7, replacing Leu6 and Lys7 respectively. Fmoc-protected solid phase peptide synthesis (SPPS) on Rink Amide MBHA resin was used to produce both peptides, as described in Chapter 2 section 2.1. The relatively mild resin cleavage with TFA was preferred instead of the HF used in Boc-protected SPPS to prevent cleavage of the perfluoro-*t*-butyl group. During synthesis, no incomplete coupling due to nFhSer was seen, as judged by Kaiser test.

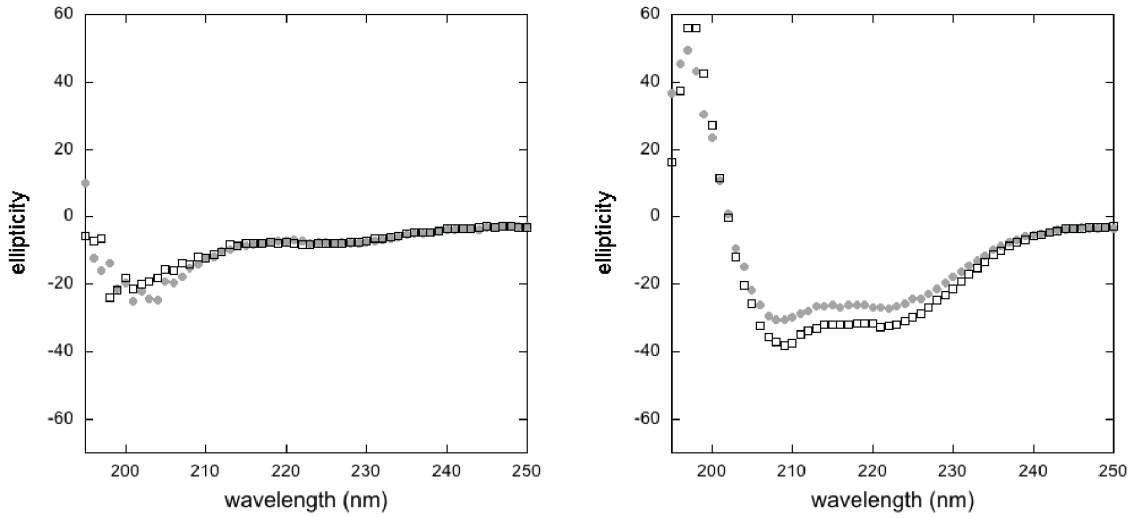


Figure 6.8. CD spectra of 100 μM MSI9-F6 (●) and MSI9-F7 (□) in buffered solution (*left*) and in the presence of 100 mM SDS micelles (*right*).

6.4.2 – Structural Consequences of MSI-78 Containing nFhSer

The ^{19}F NMR signal increases linearly with the number of chemically equivalent fluorine atoms. In using fluorine to probe biological systems, there is a fine line between having adequate signal and altering the natural biological function of the system under study. To assess whether nFhSer incorporation into MSI-78 alters structure thereby changing the mechanism of membrane disruption, CD spectroscopy was performed. Both MSI9-F6 and MSI9-F7 are unstructured in buffered solution and exhibit α-helical CD spectra in the presence of SDS micelles as seen in Figure 6.8. The helical content, however, is decreased by ~40–50% compared to SDS bound MSI-F6 and MSI-F7 (Fig. 6.3). This difference in helical content is likely caused by the bulkiness of the hydrophobic perfluoro-*t*-butyl group disrupting the dimeric hydrophobic core in MSI9-F6 and substantially altering the properties of the Lys containing charged exterior of MSI9-F7.

Although the secondary structure of MSI9-F6 and MSI9-F7 differs from MSI-78 and tFeG containing MSI-78 peptides, biological activity is still retained. Both peptides exhibited MIC values (~4 μg/mL) against *E. coli* K12 strains, similar to values for MSI-78, MSI-F6 and MSI-F7 [16]. It is worth noting that the perfluoro-*t*-butyl group also seems to lower peptide solubility as both MSI9-F6 and MSI9-F7 are insoluble at concentrations greater than ~300 μM in pH 7.4 PBS buffer.

6.4.3 – Lipid Binding of MSI9-F6 and MSI9-F7

Lipid bicelles were used to examine the binding of MSI9-F6 and MSI9-F7. Consistent with tFeG containing peptides, position 6 exhibits a much greater change in chemical shift on binding to bicelles, shifting upfield by ~0.4 ppm (Fig. 6.9). MSI9-F7 displays a more subtle change in chemical shift with a slight upfield shift of ~0.05 ppm upon bicelle binding. The increase in chemically equivalent fluorine atoms is readily apparent with peptides in both free and bound states showing a sharp singlet (compared to the triplet of MSI-F which broadens substantially in the bound state). Although the ratio of signal to noise is greatly enhanced, sensitivity to changes in chemical environment is decreased. MSI-F6 and MSI-F7 displayed a difference in chemical shift of 1.6 and 0.4 ppm respectively between free and bound states while MSI9-F6 and MSI9-F7 chemical shifts only vary by 0.4 and 0.05 ppm.

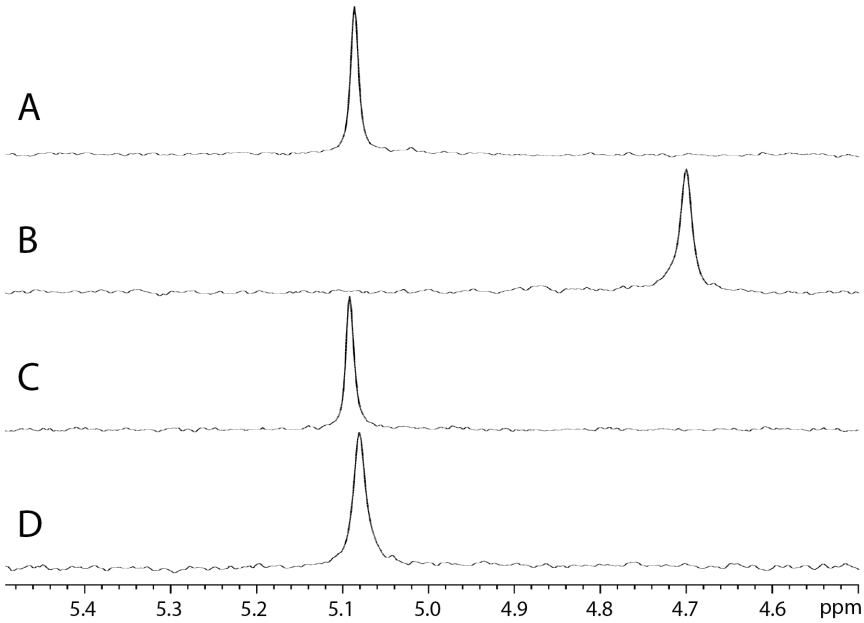


Figure 6.9. ^{19}F NMR spectral changes associated with MSI9-F7 and MSI9-F6 binding to bicelles. (A) 100 μM MSI9-F6, (B) 100 μM MSI9-F6 in the presence of 200 mM bicelles, (C) 100 μM MSI9-F7, (D) 100 μM MSI9-F7 in the presence of 200 mM bicelles. Spectra were recorded at 30 °C at pH 7.4 in PBS buffer with 10% D_2O and referenced to TFA.

The R_2 ($1/T_2$) values were measured for both MSI9-F peptides, in free solution and bound to bicelles, for $1/\tau_{cp}$ ranging from 100 – 2000 Hz (Fig. 6.10). In free solution both peptides are characterized by R_{2f} ~2 Hz which, as expected, is independent of the CPMG pulsing rate and is consistent with the perfluoro-*t*-butyl group being highly mobile in the unstructured peptide – slightly more mobile than that of the CF_3 group of tFeG (R_{2f} ~3 Hz). The experiment was then performed with MSI9-F6 and MSI9-F7 bound to high concentrations of lipid bicelles, so that contribution to R_{2obs} due to free peptide is minimal and thus $R_{2obs} = R_{2b}$. For both MSI9-F6 and MSI9-F7, R_{2b} remains between ~10 and 15 Hz, for pulsing rates between 100 and 2000 Hz.

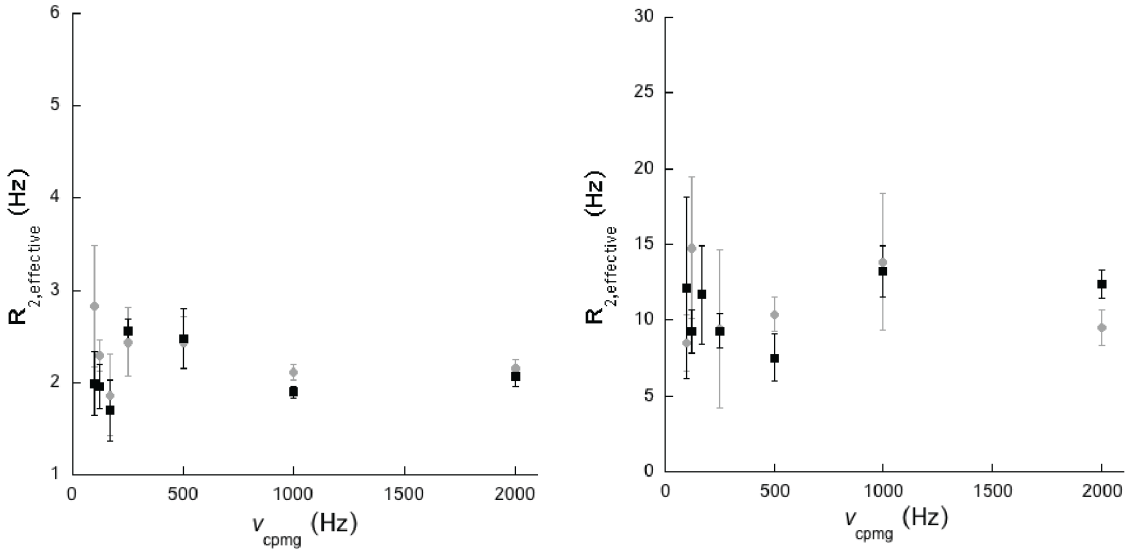


Figure 6.10. Observed transverse relaxation rate, $R_{2,effective}$, plotted as a function of CPMG pulsing rate (ν_{cpmg}). Data are for 100 μM MSI9-F6 (●) and MSI9-F7 (■) in buffered solution as unstructured peptides (*left*) and with 200 mM bicelles (*right*).

The similar values of R_{2obs} and insensitivity of both peptides to varying pulsing rates demonstrates that the perfluoro-*t*-butyl probe is a poor reporter for measuring local peptide dynamics and can't be used to estimate a dissociation rate of peptide from membrane. The inability to discriminate differences in membrane-bound-peptide dynamics at the different residue positions likely stems from the side chain of nFhSer being too long and flexible to adequately couple to the local backbone dynamics. The values of R_{2b} for MSI9-F6 and MSI9-F7 are much lower than the calculated value of 38.9 Hz discussed in section 3.3 of this chapter. This indicates that both peptides are bound to the bicelle since the relaxation rate is higher than that for free peptide, but both perfluoro-*t*-butyl probes are considerably more mobile than either the correlation time of the bicelle or of the trifluoromethyl reporter of bicelle-bound MSI-F7.

6.5 – Discussion

These experiments demonstrate that fluorine NMR can be used as a sensitive probe to investigate the interactions of peptides with membranes. The ^{19}F nucleus is intrinsically very sensitive, 100 % abundant and exhibits wide chemical shift dispersion. Moreover, fluorinated probes can readily be incorporated into peptides site specifically and in a non-disruptive manner as many fluorinated analogs of amino acids are commercially available or easily synthesized. The experiments with tFeG were routinely conducted using peptide concentrations of 400 μM to facilitate R_2 measurements, however these peptides can be detected binding to bicelles at much lower concentrations: 40 μM peptide is readily detected at signal-to-noise of 4:1 at 376 MHz after 512 scans in a spectrum that takes ~15 min to acquire. (With a more sensitive probe operating at higher field strengths sensitivity would be significantly improved.) This potentially allows one to study membrane-peptide interactions using this technique at the concentration range at which they exert their biological activity. Increasing the content of chemically equivalent fluorine atoms also allows a decrease in the peptide concentration needed, however, probes larger than a trifluoromethyl group tend to be fairly hydrophobic and can alter protein structure.

It was demonstrated that the fluorine chemical shift is both sensitive to position of the probe with respect to the membrane surface, and unexpectedly, to the nature of the lipid bilayer – larger changes in chemical shift being observed upon binding to bicelles than binding to SUVs. This suggests that the peptides interact slightly differently with these two systems. One explanation is that the binding interaction is sensitive to membrane curvature: SUVs, being spherical, exhibit positive curvature of the bilayer and

are relatively rigid whereas bicelles, being disc-like, present a flat surface that is more flexible. The difference in membrane topology may well result in subtle changes to the structure of the peptide-membrane complex.

In addition to monitoring peptide binding, NMR also provides information on the dynamics of the peptide-membrane interaction, which would be hard to obtain by other methods. The relaxation measurements of the tFeG probe show that buried and solvent- or lipid head group-exposed positions within the peptide exhibit significantly different dynamics. Whereas the buried position appears to tumble at the frequency of the bicelle, the exposed positions are significantly more dynamic.

In conclusion, studies in this chapter demonstrate the utility of ^{19}F NMR for investigating the interactions of peptides and proteins with their membrane targets. In particular, the high sensitivity and lack of background signal point to the feasibility of using fluorine NMR to study peptide-membrane interactions *in vivo* at physiologically relevant concentrations. A more comprehensive study based on this work, by Suzuki *et al.*, which utilizes seven additional tFeG-containing MSI-78 analogues can be found in Appendix A.

6.5 – References

(1) Brogden, K. A. *Nature Reviews Microbiology* **2005**, *3*, 238.
(2) Huang, H. W.; Chen, F.-Y.; Lee, M.-T. *Phys Rev Lett* **2004**, *92*, 198304.
(3) Oren, Z.; Shai, Y. *Biopolymers* **1998**, *47*, 451.
(4) Hancock, R. E. W.; Lehrer, R. *Trends Biotechnol.* **1998**, *16*, 82.
(5) Shai, Y. *Biochimica Et Biophysica Acta-Biomembranes* **1999**, *1462*, 55.
(6) Wu, M. H.; Maier, E.; Benz, R.; Hancock, R. E. W. *Biochemistry* **1999**, *38*, 7235.
(7) Epand, R. F.; Ramamoorthy, A.; Epand, R. M. *Protein and Peptide Letters* **2006**, *13*, 1.
(8) Oren, Z.; Shai, Y. *Biochemistry* **1997**, *36*, 1826.

(9) Selsted, M. E.; Novotny, M. J.; Morris, W. L.; Tang, Y. Q.; Smith, W.; Cullor, J. S. *Journal of Biological Chemistry* **1992**, *267*, 4292.
(10) Ramamoorthy, A.; Thennarasu, S.; Lee, D. K.; Tan, A. M.; Maloy, L. *Biophysical Journal* **2006**, *91*, 206.
(11) Porcelli, F.; Verardi, R.; Shi, L.; Henzler-Wildman, K. A.; Ramamoorthy, A.; Veglia, G. *Biochemistry* **2008**, *47*, 5565.
(12) Powers, J. P. S.; Tan, A.; Ramamoorthy, A.; Hancock, R. E. W. *Biochemistry* **2005**, *44*, 15504.
(13) Porcelli, F.; Buck, B.; Lee, D. K.; Hallock, K. J.; Ramamoorthy, A.; Veglia, G. *Journal of Biological Chemistry* **2004**, *279*, 45815.
(14) Mani, R.; Cady, S. D.; Tang, M.; Waring, A. J.; Lehrer, R. I.; Hong, M. *Proc. Natl. Acad. Sci. (USA)* **2006**, *103*, 16242.
(15) Wu, X.; Mani, R.; Tang, M.; Buffy, J. J.; Waring, A. J.; Sherman, M. A.; Hong, M. *Biochemistry* **2006**, *45*, 8341.
(16) Gottler, L. M.; Lee, H. Y.; Shelburne, C. E.; Ramamoorthy, A.; Marsh, E. N. G. *Chembiochem* **2008**, *9*, 370.
(17) Gottler, L. M.; De la Salud-Bea, R.; Shelburne, C. E.; Ramamoorthy, A.; Marsh, E. N. G. *Biochemistry* **2008**, *47*, 9243
(18) Hsieh, K. H.; Needleman, P.; Marshall, G. R. *J. Med. Chem.* **1987**, *30*, 1097.
(19) Wang, P.; Tang, Y.; Tirrell, D. A. *J. Am. Chem. Soc.* **2003**, *125*, 6900.
(20) Meng, H.; Kumar, K. *J. Am. Chem. Soc* **2007**, *129*, 15615.
(21) Danielson, M. A.; Falke, J. J. *Annual Review of Biophysics and Biomolecular Structure* **1996**, *25*, 163.
(22) Gakh, Y. G.; Gakh, A. A.; Gronenborn, A. M. *Magnetic Resonance in Chemistry* **2000**, *38*, 551.
(23) Tsushima, T.; Kawada, K.; Ishihara, S.; Uchida, N.; Shiratori, O.; Higaki, J.; Hirata, M. *Tetrahedron* **1988**, *44*, 5375.
(24) Lee, H. Y.; Lee, K. H.; Al-Hashimi, H. M.; Marsh, E. N. G. *J. Am. Chem. Soc.* **2006**, *128*, 337.
(25) Gottler, L. M.; de la Salud-Bea, R.; Marsh, E. N. G. *Biochemistry* **2008**, *47*, 4484.
(26) Shelburne, C. E.; An, F. Y.; Dholpe, V.; Ramamoorthy, A.; Lopatin, D. E.; Lantz, M. S. *Journal of Antimicrobial Chemotherapy* **2007**, *59*, 297.
(27) Grage, S. L.; Durr, U. H. N.; Afonin, S.; Mikhailiuk, P. K.; Komarov, I. V.; Ulrich, A. S. *Journal of Magnetic Resonance* **2008**, *191*, 16.
(28) Andersson, A.; Almqvist, J.; Hagn, F.; Maler, L. *Biochimica Et Biophysica Acta-Biomembranes* **2004**, *1661*, 18.
(29) Hallock, K. J.; Lee, D. K.; Ramamoorthy, A. *Biophysical Journal* **2003**, *84*, 3052.
(30) Porcelli, F.; Buck-Koehntop, B. A.; Thennarasu, S.; Ramamoorthy, A.; Veglia, G. *Biochemistry* **2006**, *45*, 5793.
(31) Jiang, Z. X.; Yu, Y. B. *The Journal of organic chemistry* **2007**, *72*, 1464.
(32) Sebesta, D. P.; O'Rourke, S. S.; Pieken, W. A. *Journal of organic chemistry* **1996**, *61*, 361.

Chapter 7

Conclusions and Future Directions

7.1 – Overview

Enhancing protein stability is an essential component in the design of enzymes and biomaterials that will function under harsh environmental conditions such as, high heat, with chemical denaturants and in the presence of proteases. The incorporation of highly fluorinated amino acids into proteins has emerged as a means to create protein-based materials with novel chemical and physical properties. Incorporating fluorinated amino acids into proteins has proved useful for several purposes including enhancing their chemical and thermal stability and as a noninvasive reporter of protein structure and dynamics [1-29]. Of particular interest to my research is the structural accommodation of the larger and more hydrophobic fluorinated Leu analogue, hFLeu, to create hyperstability [6,7,9,11,14,15,24].

The previously described, de novo designed, 4-helix bundle protein, α_4, was used as a model system to study how incorporation of hFLeu into proteins effects thermodynamic stability [15]. It has been suggested that the self-segregating properties of highly fluorinated small molecules may be manifested in fluorinated proteins [2-4,30]. However, CD and NMR studies of various hFLeu-containing α_4 analogues show no evidence of a protein-based "fluorous effect" [6,11,14]. The optimal packing of protein core residues leads to the greatest stability increase on a per-residue basis, rather than

increasing the amount of fluorine-fluorine contacts [6]. To determine how fluorine is accommodated in a protein environment and how the packing of α_4 proteins leads to enhanced stability, X-ray crystal structures were obtained. Van't Hoff analysis of heat- and GuHCl-dependent protein denaturation demonstrated that the thermodynamic parameters contributing to fluorous proteins stability are similar to those of natural proteins. Lastly, ^{19}F NMR was used to probe the interaction of the AMP MSI-78 with lipid membranes.

7.2 – Self-Segregation and Enhanced Stability of Fluorinated Proteins

My work on investigating the enhanced stability of fluorinated proteins builds on previous research in the Marsh laboratory using the protein, α_4. The incorporation of varying amounts of hFLeu into the antiparallel, 4-helix bundle, α_4 has previously been shown to increase $\Delta G°_{fold}$ linearly on a per hFLeu residue basis. By incorporating hFleu into only all **a** or all **d** positions, the proteins α_4F_3a and α_4F_3d exhibited a greater per-residue stability than the fully fluorinated α_4F_6. This indicates that direct fluorine-fluorine contacts do not appear to contribute to stability, instead, optimal packing similar to the classic knobs-into-holes arrangement where interspersing the larger hFLeu side chains between smaller Leu side chains leads to greatest stability.

A unique fluorous phase proposed by Kumar and others[2-4] to be present in fluorinated proteins has not been seen for α_4 proteins, including the highly fluorinated α_4F_6. Experiments using ^{19}F NMR and CD indicate that α_4H interacts with α_4F_3a, α_4F_3d and α_4F_6. These interactions appear to be transient on the NMR timescale as seen by increased broadening of fluorine signal from protein mixtures [6,11]. The results indicate

that stability and selectivity is dictated by size and shape rather than a distinct fluorous interaction.

7.3 – Structural Consequences of Protein Fluorination

NMR and CD studies indicate that highly fluorinated proteins do not self-associate through exclusive "fluorous" interactions in a bio-orthogonal manner. These highly fluorinated proteins are, however, substantially more stable than their nonfluorinated counterparts. To investigate the structural changes brought on by replacing hydrogen-containing Leu with the larger, fluorine-containing hFLeu, and to determine if there is a preference for fluorine-fluorine contacts, crystals of α_4 analogues were grown to obtain X-ray crystal structures. The structural stability of the antiparallel, 4-helix bundle and plasticity of its core towards hydrophobic substitution proved invaluable in accommodating the increased volume of hFLeu and tBAla substitutions and decreased volume of tFeG as compared to a natural Leu residue. Six high resolution structures were obtained for proteins: α_4H, α_4Ht, α_4F_3a, α_4F_3d, α_4F_3(6-13) and $\alpha_4F_3af_3$d. These structures all preserve the originally designed antiparallel, tetrameric, coiled-coil topology. The close structural similarities of hFLeu-containing proteins to α_4H reveal hFLeu to be a viable amino acid for modifying protein physical properties while retaining native protein structure and function.

The four unique crystal structures of fluorinated proteins reveal that fluorination represents a useful tool for stabilizing proteins by providing the ability to increase hydrophobicity whilst closely preserving side chain shape. Comparing the structure and stability of α_4H with $\alpha_4F_3af_3$d demonstrates that maintaining a constant hydrophobic core

volume while increasing fluorine content to 108 fluorine atoms doesn't lead to increased stability as predicted by the "fluorous effect". Instead, the stability increases seen for Leu to hFLeu substitutions are better explained through increased hydrophobicity as described by the conventional hydrophobic effect.

The distinct knobs-into-holes packing of α_4F_3a and α_4F_3d core residues appears to enhance stability over that expected from just the hydrophobic contribution of hFLeu. The structure of the comparatively less stable α_4F_3(6-13) displays a well packed core that differs only in the vertical knobs-into-holes packing of the **b**-**e** and **c**-**g** interfaces. These observations add to the evidence of fluorous protein stability being dictated by interactions analogous to those of natural proteins. The biocompatibility of hFLeu and indeed all fluorinated amino acids is dependent on closely approximating the shape of the analogous native residue. The unique, cavity-containing structure of α_4Ht is a result of the nonnatural shape of tBAla, which packs the entire protein core. Although the hydrophobic volume and thermodynamic stability of α_4Ht closely mimics that of α_4F_3 proteins, the shape of tBAla, which is unlike that of Leu or hFLeu leads to dissimilar core packing.

It is hoped that the insights gained from these crystallographic studies will aid future efforts to modulate protein stability and protein-ligand interactions using this versatile class of non-canonical amino acids.

7.4 – Thermodynamic Consequences of Protein Fluorination

Structural studies of α_4 proteins provide detailed information to guide future incorporation of fluorinated amino acids. The thermodynamic parameters determined by

global fitting of denaturation data accompany the structural data to provide a thorough analysis of how fluorination may alter the natural properties of a protein. The values of $\Delta H°$, $\Delta S°$ and $\Delta C_p°$ for 12 α_4 proteins indicate that incorporation of hFLeu alters thermodynamic parameters to a similar extent as natural proteins. The increases in $\Delta S°$ and $\Delta C_p°$ correlate with increased hydrophobic content of either fluorocarbon or hydrocarbon nature. $\Delta H°$ doesn't appear to increase linearly as a function of fluorine content, as would be expected if fluorine-fluorine interactions contributed to stability. This investigation confirms the findings from fluorous protein crystal structures that hFLeu is compatible with protein environment due to the conservation of Leu shape. There is also little evidence for the "fluorous effect" in these fluorinated proteins with the increases in $\Delta G°$ better ascribed to general increases in hydrophobic surface area as predicted by the classic hydrophobic effect.

7.5 – Probing Dynamics with ^{19}F NMR

The NMR sensitive properties of the ^{19}F nucleus make the incorporation of fluorinated amino acids useful probes to study dynamics in biological systems. By incorporating tFeG into the potent AMP, MSI-78, it was possible to probe transient interactions with lipid membranes. Fluorine chemical shift is sensitive to positional differences of the trifluoromethyl group, as demonstrated by ^{19}F chemical shift differences between the bound and free states of both MSI-F6 and MSI-F7 interacting with either bicelles or lipid SUVs. These observations indicate that the transition of MSI-78 from unstructured to α-helical, upon lipid binding, places charged and hydrophobic residues in distinctly different chemical environments.

CPMG experiments show that the unbound peptide is unstructured, with a correspondingly low relaxation rate. Upon binding to lipid bicelles, the relaxation rate greatly increases depending upon location of the trifluoromethyl reporter. The tFeG residue of MSI-F6 has a relaxation rate corresponding to the tumbling rate of the bicelle-peptide aggregate, indicating that residue 6 is located in a motionally restrictive environment, such as the interior of a peptide oligomer or lipid interior. MSI-F7, however displays a lower relaxation rate than that of MSI-F6 indicating that residue 7 is in close proximity to the mobile lipid head groups or solvent.

The lack of fluorine in Nature and high sensitivity of the fluorine nucleus make ^{19}F NMR a promising tool to study biological interactions. By increasing the sensitivity of fluorine reporters, the study of less populated molecules in living systems or binding events with small binding constants could be observed. The synthesis and study of nFhSer on binding of MSI-78 to lipids shows promise towards increasing the signal sixfold compared to tFeG, however aqueous solubility remains a limiting factor for future studies.

7.6 – Future Directions

The research presented within this dissertation provides insight into the structural and thermodynamic properties of fluorinated proteins. These investigations should assist in future endeavors to stabilize protein various folds using highly fluorinated or nonfluorinated hydrophobic amino acids. The incorporation of hFLeu into buried positions of α-helical coiled-coil proteins is seen to be greatly stabilizing with minimal structural perturbation. This stabilizing potential of fluorous amino acids has broad

applicability towards modifying the properties of all proteins, natural or synthetic. Further investigation into other, less well-studied, protein structural motifs, such as β-sheets is needed, since fluorous amino acids incorporated into both buried and solvent-exposed positions of β-sheets has resulted in enhanced stability [8,13].

The non-disruptive nature of substituting hydrocarbon containing amino acids for their highly fluorinated analogues signify potential in stabilizing large globular proteins including those with enzymatic activity. Of great interest is the development of stable enzyme catalysts, which can function under the environmental extremes of high heat and in the presence of organic denaturants. Therapeutic proteins and peptides may also benefit from incorporating fluorous amino acids through increased resistance to proteolytic degradation, thereby offering a means to increase bioavailability while retaining a natural-like structure and function.

Through dissecting the structural and thermodynamic consequences of fluorine incorporation into α-helical proteins, I hope to establish renewed interested in this technique as a means to enhance protein stability. Fluorous analogues of natural, hydrocarbon-containing, amino acids are exceptionally biocompatible with the solvent-excluded environment of the protein core. The production of fluorinated proteins currently relies on synthetic or semi-synthetic methods. However, ongoing development of site-specific incorporation of nonnatural amino acids into proteins would allow for nearly limitless options in introducing new chemical functionalities including those of a highly fluorous nature.

7.7 – References

(1) Bilgiçer, B.; Fichera, A.; Kumar, K. *J Am Chem Soc* **2001**, *123*, 4393.
(2) Bilgiçer, B.; Kumar, K. *Tetrahedron* **2002**, *58*, 4105.
(3) Bilgiçer, B.; Kumar, K. *Proc Nat Acad Sci USA* **2004**, *101*, 15324.
(4) Bilgiçer, B.; Xing, X.; Kumar, K. *J Am Chem Soc* **2001**, *123*, 11815.
(5) Buer, B. C.; Chugh, J.; Al-Hashimi, H. M.; Marsh, E. N. G. *Biochemistry* **2010**, *49*, 5760.
(6) Buer, B. C.; de la Salud-Bea, R.; Al Hashimi, H. M.; Marsh, E. N. G. *Biochemistry* **2009**, *48*, 10810.
(7) Buer, B. C.; Meagher, J. L.; Stuckey, J. A.; Marsh, E. N. G. *Proceedings of the National Academy of Sciences* **2012**, *109*, 4810.
(8) Chiu, H.-P.; Kokona, B.; Fairman, R.; Cheng, R. P. *J Am Chem Soc* **2009**, *131*, 13192.
(9) Chiu, H.-P.; Suzuki, Y.; Gullickson, D.; Ahmad, R.; Kokona, B.; Fairman, R.; Cheng, R. P. *J Am Chem Soc* **2006**, *128*, 15556.
(10) Gottler, L. M.; de la Salud Bea, R.; Shelburne, C. E.; Ramamoorthy, A.; Marsh, E. N. G. *Biochemistry* **2008**, *47*, 9243.
(11) Gottler, L. M.; de la Salud-Bea, R.; Marsh, E. N. G. *Biochemistry* **2008**, *47*, 4484.
(12) Gottler, L. M.; Lee, H.-Y.; Shelburne, C. E.; Ramamoorthy, A.; Marsh, E. N. G. *ChemBioChem* **2008**, *9*, 370.
(13) Horng, J.-C.; Raleigh, D. P. *J Am Chem Soc* **2003**, *125*, 9286.
(14) Lee, H.-Y.; Lee, K.-H.; Al-Hashimi, H. M.; Marsh, E. N. G. *J Am Chem Soc* **2006**, *128*, 337.
(15) Lee, K.-H.; Lee, H.-Y.; Slutsky, M. M.; Anderson, J. T.; Marsh, E. N. G. *Biochemistry* **2004**, *43*, 16277.
(16) Meng, H.; Krishnaji, S. T.; Beinborn, M.; Kumar, K. *J Med Chem* **2008**, *51*, 7303.
(17) Meng, H.; Kumar, K. *J Am Chem Soc* **2007**, *129*, 15615.
(18) Montclare, J. K.; Son, S.; Clark, G. A.; Kumar, K.; Tirrell, D. A. *ChemBioChem* **2009**, *10*, 84.
(19) Niemz, A.; Tirrell, D. A. *J Am Chem Soc* **2001**, *123*, 7407.
(20) Son, S.; Tanrikulu, I. C.; Tirrell, D. A. *ChemBioChem* **2006**, *7*, 1251.
(21) Suzuki, Y.; Buer, B. C.; Al-Hashimi, H. M.; Marsh, E. N. G. *Biochemistry* **2011**, *50*, 5979.
(22) Tang, Y.; Ghirlanda, G.; Petka, W. A.; Nakajima, T.; DeGrado, W. F.; Tirrell, D. A. *Angew Chem Int Ed* **2001**, *40*, 1494.
(23) Tang, Y.; Ghirlanda, G.; Vaidehi, N.; Kua, J.; Mainz, D. T.; Goddard, W. A.; DeGrado, W. F.; Tirrell, D. A. *Biochemistry* **2001**, *40*, 2790.
(24) Tang, Y.; Tirrell, D. A. *J Am Chem Soc* **2001**, *123*, 11089.
(25) Wang, P.; Fichera, A.; Kumar, K.; Tirrell, D. A. *Angew Chem Int Ed* **2004**, *43*, 3664.
(26) Wang, P.; Tang, Y.; Tirrell, D. A. *J Am Chem Soc* **2003**, *125*, 6900.
(27) Woll, M. G.; Hadley, E. B.; Mecozzi, S.; Gellman, S. H. *J Am Chem Soc* **2006**, *128*, 15932.
(28) Zheng, H.; Comeforo, K.; Gao, J. *J Am Chem Soc* **2008**, *131*, 18.
(29) Zheng, H.; Gao, J. *Angew Chem Int Ed* **2010**, *49*, 8635.

(30) Marsh, E. N. G. *Chem Biol* **2000**, *7*, R153.

Appendix A

Using Fluorine NMR to Probe Changes in Structure and Dynamics of Membrane-Active Peptides Interacting with Lipid Bilayers

Yuta Suzuki, Benjamin C. Buer, Hashim M. Al-Hashimi and E. Neil G. Marsh

Biochemistry, 2011, 50(27):5979-5987

Introduction

The interaction of small peptides with the lipid bilayer of the cell membrane is important in a variety of biological processes [1-6]. These membrane-active peptides may have protective properties such as antimicrobial peptides[1] (AMPs), anti-cancer and anti-viral peptides, or be involved in pathological processes such as cell-penetrating peptides, viral fusion peptides and venom peptides. For all these classes of peptides, interactions between the membrane lipid bi-layer and the peptide are central to their biological functions.

Characterizing peptide-membrane interactions is often challenging because of the transient and dynamic nature of these interactions. The peptide may adopt different orientations with respect to the lipid bi-layer and different oligomerization states that are concentration-dependent. Solid state NMR experiments have used various NMR-active nuclei to investigate the structures and orientations of AMPs bound to lipid membranes [7-12]. These include studies using fluorine-labeled AMPs to provide information on peptide orientation in membranes [13-15]. However, these experiments require peptide concentrations that are orders of magnitude higher than their physiologically active range, so it is not always clear whether such structures represent biologically active species. Solution-phase fluorine NMR has proved an informative tool for investigating biological interactions; for example in probing the dynamics of soluble proteins [16-18] and the immersion depth of lipophilic molecules in lipid bilayers [19,20].

AMPs are a diverse family of membrane-active peptides found in essentially all multi-cellular organisms. Although some AMPs have specific intracellular targets [1],

most exert their antimicrobial activity by binding directly to the microbial membrane and compromising its integrity [21-23]. Almost all AMPs are highly amphipathic, with one face of the peptide being hydrophobic and the other face presenting a cluster of positively charged residues [24-26]. The selectivity of AMPs for bacterial membranes arises primarily from electrostatic interactions between the positively charged peptide and the negatively charged phospholipids that predominate in bacterial cell membranes. Eukaryotic membranes, which contain predominantly neutral phospholipids, are usually less susceptible to disruption by AMPs; the presence of cholesterol in eukaryotic membranes also helps prevent membrane disruption by AMPs [27]. Upon association with the membrane, disruption of the bacterial membrane may proceed through a number of mechanisms, including the formation of pores, membrane thinning and detergent-like action [23,28,29].

Our studies have focused on the potent, synthetic AMP, MSI-78 (pexiganan), which provides a convenient model system to investigate peptide-membrane interactions. MSI-78 is thought to disrupt bacterial membranes by forming toroidal pores in the lipid bilayer [30], as illustrated in Figure 1. Based on a combination of solution and solid-state NMR experiments the peptide has been shown to adopt a dimeric α-helical coiled-coil structure (Figure 1) in the presence of 3:1 POPC-POPG liposomes [31]. The dimer interface is formed by contacts between hydrophobic residues and the positively charged lysine residues that face the exterior of the structure and interact with hydrophilic lipid head groups.

In previous studies, our group and others [32-36] have demonstrated that the incorporation of fluorinated amino acids into AMPs is an effective strategy to modulate

their biological properties and can be used to study AMP-membrane interactions. For example, we showed that incorporating hexafluoroleucine at four positions in the α-helical AMP MSI-78 resulted in increased potency towards some bacterial strains and protection against proteolysis when bound to lipid vesicles [32].

Recently we turned our attention towards using fluorine-containing peptides to probe the interaction of AMPs with membranes by exploiting the sensitive NMR properties of the ^{19}F nucleus. We demonstrated that binding of MSI-78 to small unilamellar vesicles and bicelles could easily be detected by following changes in the ^{19}F chemical shift of MSI-78 variants containing L-4,4,4-trifluoroethylglycine (TfeG) [37]. We also showed that the local dynamical properties of the membrane-bound peptide in the vicinity of the label could be investigated by measuring the transverse relaxation rate (R_2) of the ^{19}F nucleus using CPMG relaxation dispersion experiments. Here we have extended these measurements to examine the dynamics of the MSI-78:membrane complex by incorporating CF_3- probes at strategic positions throughout the peptide and measuring the associated changes in chemical shifts and R_2 values.

Experimental Procedures

Peptide synthesis Racemic 4,4,4-trifluoroethylglycine (TfeG) was purchased from SynQuest Labs and enzymatically resolved (porcine kidney acylase I) resulting in L-4,4,4-trifluoroethylglycine having >99% ee [38]. The pure amino acid was converted to its t-Boc- derivative by standard procedures. The sequences of MSI-78 derivatives are shown in Figure 2. Peptides were synthesized manually by either standard t-Boc procedures on MBHA resin or by f-moc procedures on PAL-PEG resin, as described

previously [39,40]. Peptides were purified by reverse phase HPLC using a gradient of water/acetonitrile with 0.1% TFA; excess residual TFA was removed by a Stratosphere SPE column (Varian). Stock peptide concentrations were determined using ^{19}F NMR with a known concentration of TFA as an internal reference. Peptide identities were confirmed using MALDI-MS.

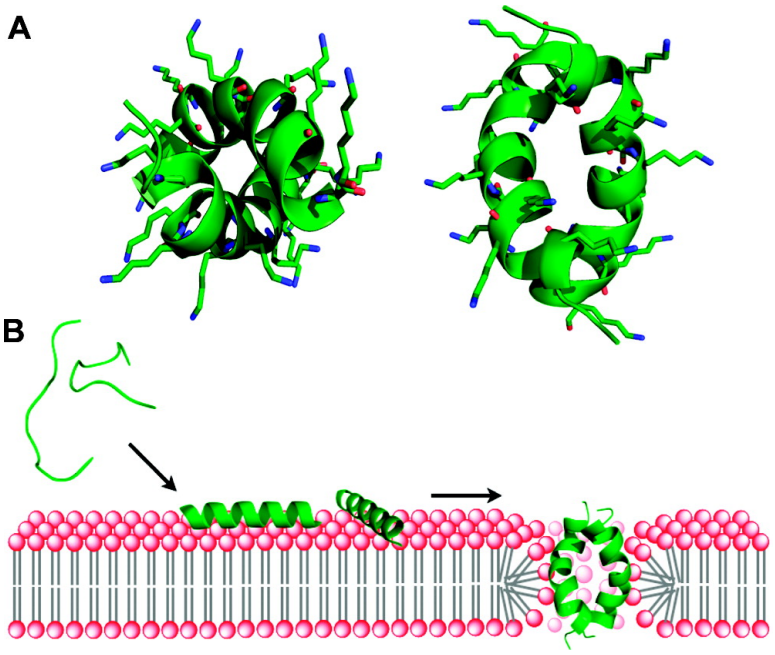


Figure 1. *Left*: Cartoon illustrating the mechanism for MSI-78 insertion into a lipid bilayer to form toroidal pores. *Right*: structure of MSI-78, determined from NMR experiments *(31)*.

Lipid Preparation 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC), 1,2-dimyristoyl-*sn*-glycero-3-phospho-(1'-*rac*-glycerol) (DMPG) and 1,2-dihexanoyl-*sn*-glycero-3-phosphocholine (DHPC) were purchased from Avanti Polar Lipids. Dodecylphosphocholine (DPC) was purchased from Affymetrix. Isotropic bicelles were made in PBS buffer, pH 7.4 with 10% D_2O by adding a solution of 3:1 DMPC/DMPG to a solution of DHPC giving q=0.5 resulting in a clear, non-viscous solution.

MIC determinations The peptide minimum inhibitory concentrations (MIC) against *E. coli* K12 were determined by the microdilution antimicrobial assay procedure, using 96-well plates in replicates of four, as described previously [41].

Proton NMR Samples were prepared in PBS, pH 6.0, containing 10% D_2O, 100 mM DPC and 400 μM peptide. A water suppression pulse sequence (WET) was employed in the acquisition of NMR spectra.

^{19}F NMR All ^{19}F NMR experiments were performed at 30 °C using a Varian Inova 500 MHz NMR spectrometer equipped with a double-tuned ^{1}H-^{19}F room-temperature probehead. Peptide and lipid samples were prepared with 10% D_2O in PBS, pH 7.4. All experiments were performed at a constant peptide concentration of 400 μM unless indicated otherwise and referenced to trifluoroacetate ion at 0 ppm. To measure solvent isotope-induced changes in chemical shift, peptide and lipid samples were first prepared with 10% D_2O in PBS, pH 7.4. The samples were then lyophilized overnight and re-dissolved in either 10% or 90% D_2O. The solvent-induced changes were referenced to trifluoroacetate ion at 0 ppm as an external standard in PBS, pH 7.4, 10% D_2O.

^{19}F CPMG relaxation dispersion experiments were performed for the peptides in the free state and in the presence of lipid bicelles (200 mM total lipid concentration, q=0.5, long chain lipids 3:1 mol/mol DMPC/DMPG and short chain lipid being DHPC). CPMG delays (τ_{cp}) were varied from 0.5 to 10.0 ms with each data point recorded as a series of standard 1-D transverse relaxation rate measurements with T_2 delays of 0.05, 0.1, 0.2, 0.4, 0.8 and 1.6 ms for free peptide and 0.0125, 0.025, 0.05, 0.1, 0.2 and 0.4 ms for bicelle-bound peptide. The pulse width was 7.9 μs. Data sets were recorded with acquisition time of 1 s in T_1 along with a 10 s pre-scan delay. 16 scans, net acquisition time of 17 min/data point, were required to achieve adequate signal-to-noise ratios for

peptides in the free state; 32 scans, net acquisition time of 35 min/data point, were required to achieve adequate signal-to-noise ratios for peptides bound to bicelles. ^{19}F spin-lattice relaxation times (T_1) were acquired by an inversion recovery sequence (180° – τ - 90°) using a total of six τ values of 0.0625, 0.125, 0.25, 0.5, 1.0, 2.0 s. Data were processed and analyzed using VNMRJ software and plotted using the *Kaleidagraph* software package.

All other experimental details have been described previously [37,42].

Results

We previously demonstrated the feasibility of using fluorine-labeled amino acids to probe the chemical environment and dynamics of membrane-bound peptides containing TfeG [37]. TfeG provides a useful probe, as it is relatively small (approximately comparable to valine), is commercially available and readily incorporated into peptides. The C-3 methylene group affords one degree of rotational motion to the trifluoromethyl group, but on the timescale of our experiments backbone dynamics are likely to dominate its NMR behavior. Although substitution of hydrophilic residues, such as lysine, by TfeG obviously cannot be considered as conservative, numerous experiments on AMPs have shown that their biological properties and structure depend primarily on their overall physicochemical properties and are not highly sequence dependent. Changes to a single site are usually minimally perturbing to structure and activity.

In studies of two MSI-78 variants [37] incorporating TfeG at a positively charged exterior position (Lys-7) and a hydrophobic position (Leu-6) we demonstrated position-dependent changes in ^{19}F chemical shift and transverse relaxation times upon peptide

binding to small unilamellar vesicles and lipid bicelles. Whereas these initial studies provided "proof-of-concept", we considered it important to conduct a more comprehensive study to determine, more generally, how sensitive fluorine probes are for distinguishing chemical environment and local dynamics.

In the present work we have extended our investigation by synthesizing a series of 9 MSI-78 variants in which TfeG has been introduced at strategic positions throughout the 22 residue peptide. (We name these peptides as MSI-Fn^2 when n refers to the position in the peptide that has been substituted with the fluorinated amino acid.) Guided by the NMR structure of the peptide bound to lipids [31], we introduced TfeG at 4 hydrophobic positions and at 4 lysine residues spaced along the length of the peptide, including the C-terminal residue Lys-22; we also substituted the N-terminal glycine, which may be expected to show a large change in mobility. The positions that were substituted are shown in Figure 2. For each peptide we have measured the changes in ^{19}F chemical shift, longitudinal and transverse relaxation rates (R_1 and R_2), and D_2O-induced ^{19}F chemical shifts that occur upon peptide binding to lipid bicelles. This has allowed us to obtain a detailed picture of the local changes in chemical environment peptide dynamics that occur when MSI-78 binds to the lipid bilayer.

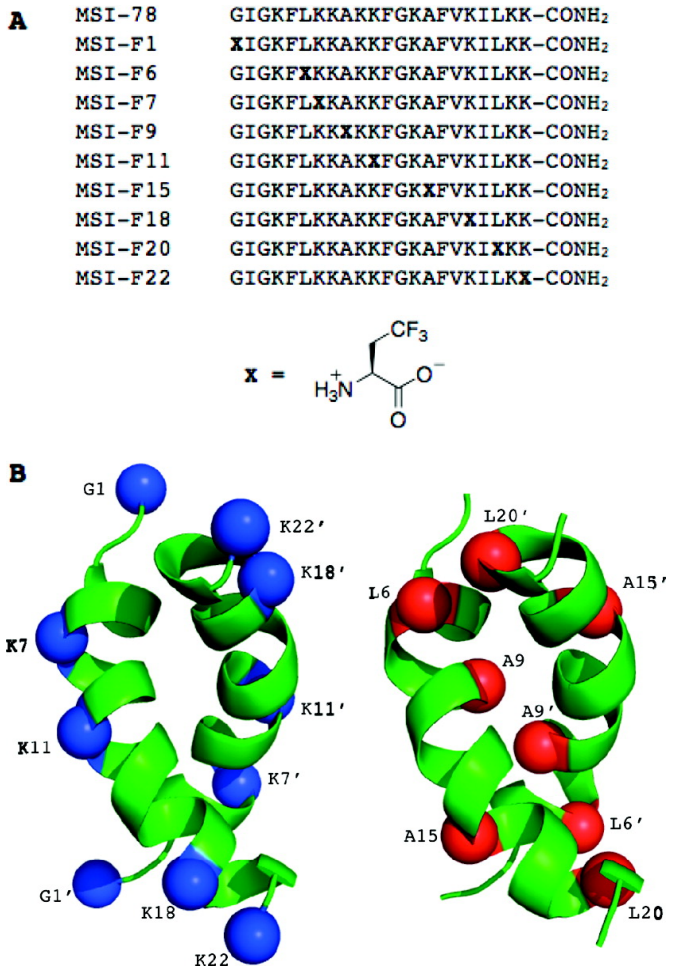


Figure 2. **A:** The primary sequence of MSI-78 and the sequences of tFeG-substituted peptides used in this study. **B:** The positions of amino acid substitutions are mapped on to the structure MSI-78 dimer formed in DPC micelles. For clarity only one peptide in the dimer is shown. Substitutions at hydrophilic positions (blue balls) are shown on the structure to the left; substitutions at hydrophobic positions (red balls) are shown on the structure to the right. Structures are rotated ~ 180° with respect to each other.

Effects of TfeG substitution on secondary structure and biological activity Substitution of TfeG at different positions in these peptides does not appear to cause any gross structural changes or changes in biological activity. The MIC values for all the peptides were within the range of 3 – 6 μg/mL against *E. coli* K12 strains; any differences in MICs between the parent MSI-78 peptide[32] and the TfeG-labeled peptides were not statistically

significant. The peptides all appear to adopt a predominantly random coil structure in plain buffer and exhibit extensively helical CD spectra in the presence of SDS micelles. The mean residue ellipticities, Θ_{222}, of the peptides varied by less than 10 % (-20,200 > Θ_{222} > -21,900 $cm^2dmol^{-1}res^{-1}$) between the different peptides and are similar to that of MSI-78 Θ_{222} = -20,700 $cm^2dmol^{-1}res^{-1}$.

As a further check on the structural integrity of the TfeG-labeled peptides, 1-D proton NMR spectra of the peptides were recorded for the peptides bound to DPC micelles (See supplementary material). In each case the peptides exhibited well-dispersed resonances in the amide region, characteristic of structured peptides. As would be expected, there small changes between the spectra of each TfeG-labeled peptide and MSI-78 were observed, which may reflect small local structural changes and/or slight differences in sample preparation. The resonances from the 3 Phe side-chains in the peptide could clearly be distinguished and were very similar in each case, indicating that the hydrophobic core of the dimeric peptide bundle remains intact. Resonances from the aliphatic side-chains were obscured by signals from the lipid bicelles.

Sensitivity of ^{19}F chemical shift to position of fluorination In the absence of lipids all the peptides, except MSI-F1, exhibited a sharp triplet in the ^{19}F NMR spectrum between 11.45 and 11.75 ppm relative to TFA (Figure 3 and Table 1). The resonance for MSI-F1, in which Gly is substituted by TfeG, is shifted significantly downfield at 12.10 ppm. This is likely due to the influence of the positively charged amino group at the N-terminus.

Peptide	Amino acid substituted	[a]δ_{free} (ppm)	[a]δ_{bound} (ppm)	R_{1free} (Hz)	R_{1bound} (Hz)	[b]R_{2free} (Hz)	[b]R_{2bound} (Hz)
MSI-F1	Gly	12.13	13.41	1.40 ± 0.05	2.55 ± 0.06	1.7 ± 0.1	27 ± 1
MSI-F6	Leu	11.64	10.29	2.18 ± 0.05	3.09 ± 0.04	3.7 ± 0.2	58 ± 9
MSI-F7	Lys	11.65	11.38	2.31 ± 0.03	3.10 ± 0.14	3.7 ± 0.2	22 ± 2
MSI-F9	Ala	11.73	10.57	2.37 ± 0.05	3.10 ± 0.14	4.1 ± 0.2	43 ± 3
MSI-F11	Lys	11.55	11.49	2.30 ± 0.04	3.06 ± 0.05	4.1 ± 0.2	27 ± 3
MSI-F15	Ala	11.47	11.56	2.26 ± 0.05	3.86 ± 0.11	3.9 ± 0.1	49 ± 3
MSI-F18	Lys	11.52	11.46	2.23 ± 0.05	3.24 ± 0.07	3.1 ± 0.1	29 ± 2
MSI-F20	Leu	11.60	10.69	2.05 ± 0.03	3.12 ± 0.09	2.8 ± 0.2	36 ± 2
MSI-F22	Lys	11.40	11.25	1.53 ± 0.06	2.27 ± 0.05	2.1 ± 0.2	20 ± 0.4
TfeG		11.62		0.54 ± 0.03		1.1 ± 0.1	

Table 1. ^{19}F Chemical shifts, R_1 and R_2 values for MSI-F peptides. [a] chemical shift relative to TFA. [b] value at $1/\tau_{cp}$ = 2000 Hz.

We then examined the binding of the fluorinated MSI-78 variants to lipid bicelles, which are commonly used as a model membrane system. [As we have discussed previously[37], for these studies bicelles are preferable to SUVs, which are also commonly used to study AMP -membrane interactions. Bicelles are more stable than SUVs and present a flat, rather than highly curved surface, for peptide binding.] All the peptides exhibited distinct changes in their ^{19}F chemical shifts upon binding to bicelles, which appeared as broadened single peaks (Figure 3).

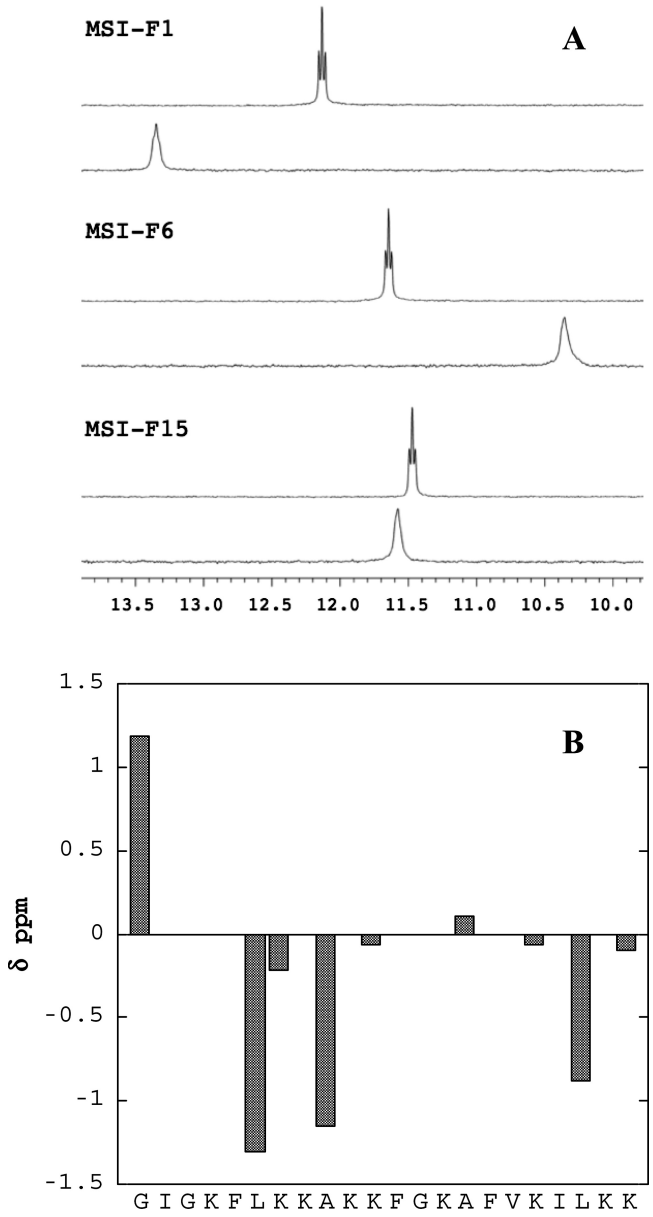


Figure 3. **A:** Representative ^{19}F NMR spectra illustrating changes in chemical shift and peak width that occur when TfeG-substituted peptides bind to lipid bicelles. The upper trace in pair of spectra is for the free peptide and the lower trace for the peptide bound to bicelles. Spectra were recorded at 30 °C at pH 7.4 in PBS buffer with 10% D_2O and referenced to TFA. **B:** Chemical shift changes, Δδ, associated with peptide binding to bicelles plotted as a function of label position.

The change in chemical shift was highly dependent on the position of the TfeG residue (Table1), demonstrating the sensitivity of the ^{19}F nucleus to local chemical

environment. The positions that exhibited the largest upfield shifts, ranging from -0.9 to -1.4 ppm, are those at hydrophobic positions that are deeply buried in the core of the coiled coil (Figure 4). The lysine positions, which project out from the coiled-coil and interact with the lipid head groups, exhibited smaller upfield shifts ranging from -0.05 to -0.26 ppm. Two positions seem to deviate from this general trend: substituting TfeG at the N-terminal glycine position (MSI-F1) results in a large downfield shift of 1.3 ppm; MSI-F15, in which Ala is substituted by TfeG, also showed a downfield shift, although this was quite small, only 0.09 ppm.

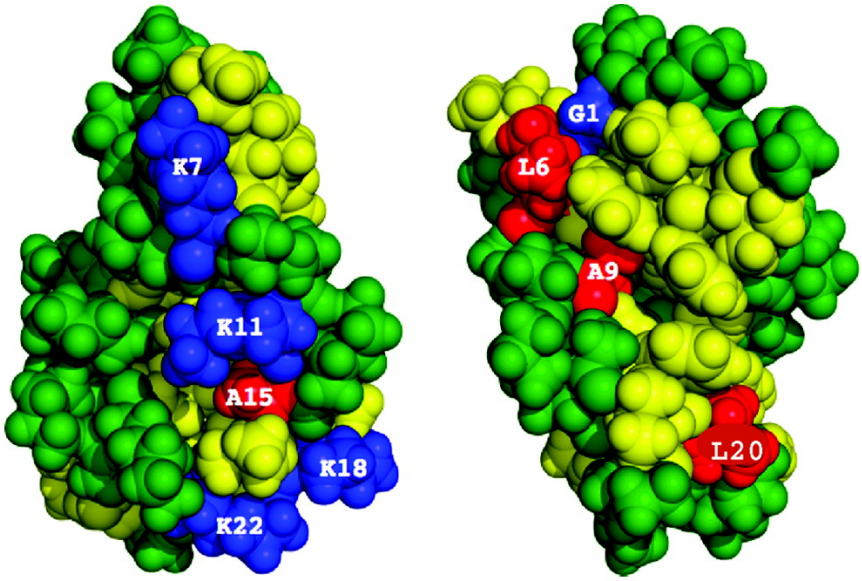


Figure 4. Space filling models of the MSI-78 dimer illustrating the chemical environment of the positions substituted by TfeG. Residues substituted by TfeG are colored blue (hydrophilic) and red (hydrophobic). Other hydrophobic and hydrophilic residues are colored yellow and green respectively. For clarity, the substituted positions are only shown on one peptide of the dimer. The two structures are related to each other by a ~ 180° rotation about the vertical axis.

Although fluorine chemical shifts are influenced by various factors [43], the "anomalous" chemical shift of MSI-F15 might be explained by the fact that the TfeG

side-chain is expected to protrude towards the charged face of the α-helix (Figure 4). Its chemical shift is likely to be influenced by the positively charged lysine residues that are adjacent to it; this could result in deshielding of the nucleus and thereby shift the resonance to lower field. In contrast, substitution of Ala-9 with TfeG (MSI-F9) places the side-chain pointing into the hydrophobic core (Figure 4), presumably resulting in a more shielded environment. The downfield shift observed when TfeG is at the N-terminal position is harder to explain. Possibly in its folded state the N-terminus of the peptide is less well solvated so that the deshielding effect of the positively-charged amino terminus is greater. The N-terminus is also in close proximity to the amino group of the C-terminal lysine of the peptide forming the opposite strand of the coiled-coil, which could also influence the fluorine chemical shift.

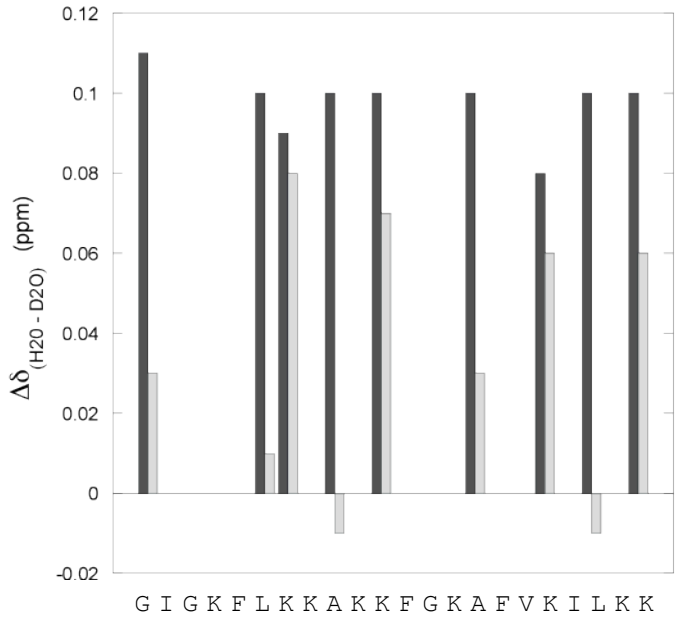


Figure 5. Solvent isotope effects on ^{19}F chemicals shift ($\Delta\delta_{(H_2O - D_2O)}$) plotted as a function of label position. Dark grey bars: peptides in free solution; light grey bars: peptides bound to lipid bicelles.

Changes in solvent exposure on binding to lipid bicelles The ^{19}F chemical shift is sensitive to the isotopic composition of the solvent, e.g. whether H_2O or D_2O, [16,18], and this provides a means to investigate changes in solvent exposure that may occur when MSI-78 binds to lipid bilayers. In the absence of bicelles, changing the solvent composition from 10 % D_2O to 90 % D_2O results in the ^{19}F chemical shift for each of the peptides moving fairly uniformly upfield by 0.08 – 0.11 ppm (see Figure 5). In the presence of bicelles, the chemical shift changes, $\Delta\delta_{(H_2O - D_2O)}$, ($\Delta\delta_{(H_2O - D_2O)} = \delta_{H_2O} - \delta_{D_2O}$) are more variable and range from -0.01 ppm to 0.08 ppm (changes of $\leq$ 0.01 ppm were not considered significant). In theory, positions that are deeply buried should exhibit no changes in chemical shift upon changing the solvent, whereas positions that are completely exposed should exhibit changes similar to those observed for the unbound peptides. Inspection of the data reveals this to be qualitatively true: positions 7, 11, 18 and 22, which are occupied by lysines in MSI-78, exhibit the largest chemical shift changes, $\Delta\delta_{(H_2O - D_2O)}$ = 0.08 – 0.06 ppm. These are somewhat smaller than those observed for the free peptide, indicating that interactions with the lipid head groups may reduce solvent exposure. In contrast, at positions 6, 9, 15 and 20, which are occupied by hydrophobic residues in MSI-78, $\Delta\delta_{(H_2O - D_2O)}$ is very small, -0.01 – 0.03. This indicates a very low degree of solvent exposure in the bound state and provides support for the proposed structural model of MSI-78 in which these positions form part of the hydrophobic core of coiled-coil peptide dimer in the membrane.

Various studies on fluorinated proteins have shown that ^{19}F longitudinal relaxation rates ($R_1 = 1/T_1$) rates are generally faster for buried residues in proteins than for solvent-exposed residues [16,18] and these measurements have been used to probe protein structure and folding. Therefore we also investigated how R_1 changed when the peptides bound to bicelles. The R_1 values for the free peptides range from 1.4 – 2.4 Hz, with the terminal residues exhibiting the lowest R_1 values (Table 1). The R_1 values increase fairly uniformly for the peptides on binding bicelles with R_1 values ranging from 2.3 – 3.2 Hz (Table 1). Only MSI-F15 stands out with R_{1bound} = 3.9 Hz being significantly higher, although the reason for this is unclear. It appears that longitudinal relaxation measurements are not especially sensitive to differences the local chemical environment in this particular system.

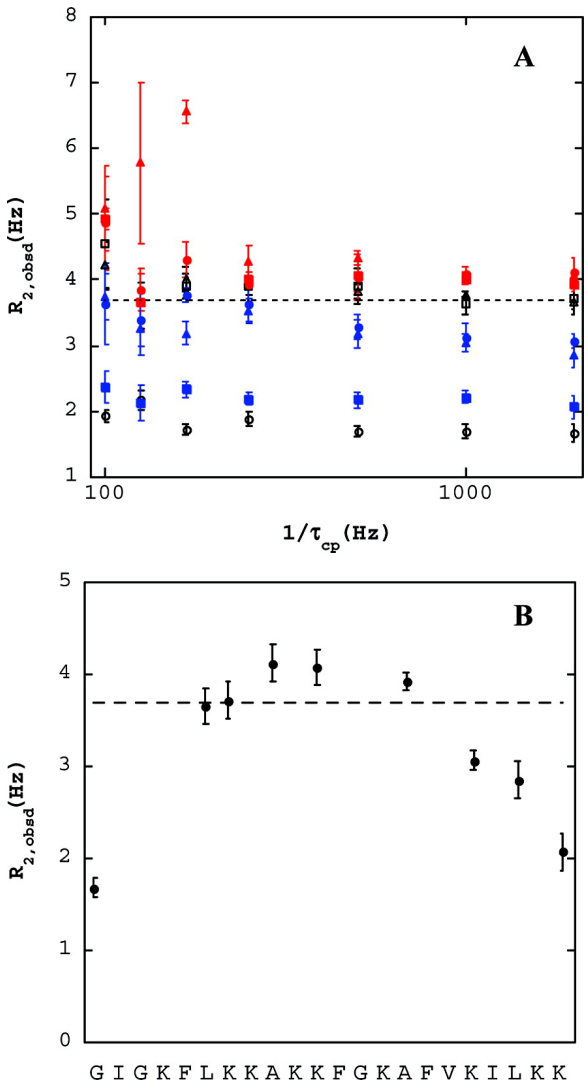


Figure 6. **A:** Observed transverse relaxation rates, $R_{2,observed}$, for MSI-F peptides in free solution plotted as a function of CPMG pulsing rate ($1/\tau_{cp}$). Symbols used are MSI-F1 (○); MSI-F6 (△); MSI-F7 (□); MSI-F9 (●); MSI-F11 (▲); MSI-F15 (■); MSI-F18 (●); MSI-F20 (▲); MSI-F22 (■). **B:** Transverse relaxation rates for peptides as plotted as a function of sequence. The data were obtained using $1/\tau_{cp}$ = 2000 Hz so that the chemical exchange component is removed.. The calculated R_2 = 3.7 Hz for free peptide is indicated by the dashed line.

Changes in peptide dynamics probed by ^{19}F transverse relaxation rates The transverse relaxation rates of the free peptides (R_{2f}) were measured using a CPMG experiment in which $1/\tau_{cp}$, was varied from 100 - 2000 Hz. For all the peptides, R_2 values are similar and are independent of τ_{cp} (Figure 5A), as expected for an unstructured peptide in free

solution. The experimental R_2 values (Figure 5B) generally agree well with that calculated for a random coil peptide in free solution [37]; however, residues at the N- and C-termini exhibit slightly slower R_{2f} values, indicating that the ends of the peptide are mobile than center (Figure 6B and Table 1)

The transverse relaxation rates were measured for each peptide bound to bicelles, R_{2b}, using the same set of CPMG pulse sequences. Under the conditions of the experiment nearly all of the peptide was bound to the lipid bilayer (As shown in Figure 3, no signal for the free peptide could be detected). As discussed previously [37,44,45], when τ_{cp} is short relative to the residence time, τ_b, of the peptide in the lipid, the observed R_{2b} values reflect the intrinsic relaxation rates of the ^{19}F nuclei unencumbered by chemical exchange. Under these conditions differences in R_{2b} may be attributed to changes in the local dynamics of the peptide.

The observed R_{2b} values for all the peptides were essentially invariant for values of $1/\tau_{cp}$ greater than 250 Hz, consistent with our previous results [37]. At longer pulse intervals R_{2b} appears to increase, indicative of chemical exchange (Figure 7A). However, at these timescales the experiment approaches the limits of sensitivity due to inter-conversion between in-phase and anti-phase magnetization during the spin-echo period, as well as loss of signal intensity due to chemical exchange. Therefore measurements are accompanied by large uncertainty in the value of R_{2b} and τ_b cannot be reliably determined from the data. However, the experiment does allow us to put an upper limit on the rate at which the peptide dissociates from the membrane ($1/\tau_b$) of ~ 200 s^{-1}, consistent with our previous data.

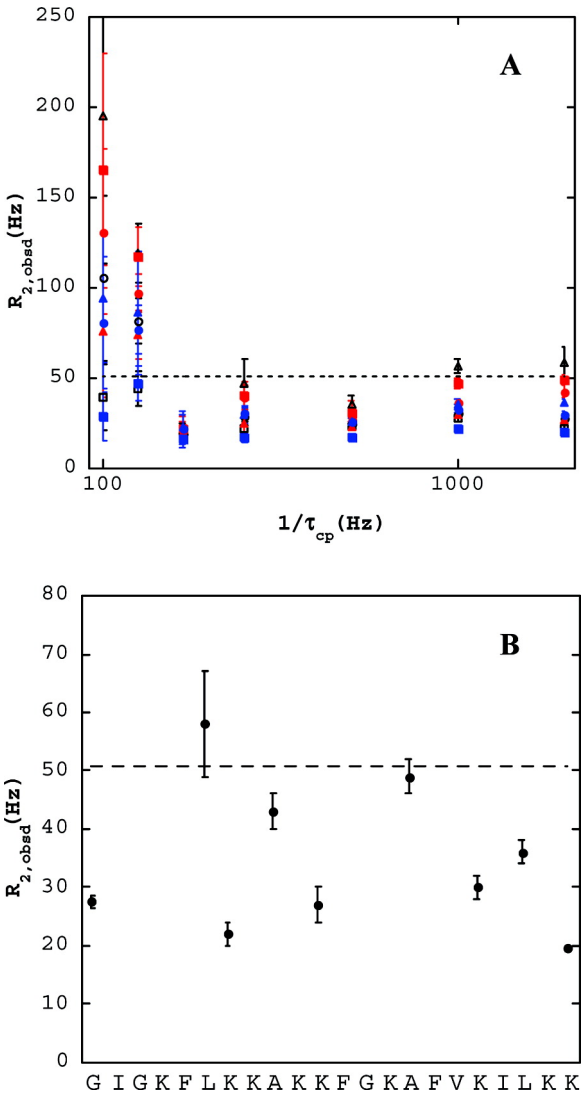


Figure 7. **A:** Observed transverse relaxation rates, $R_{2,\text{observed}}$, for MSI-F peptides bound to lipid bicelles plotted as a function of CPMG pulsing rate ($1/\tau_{cp}$). Symbols used are MSI-F1 (○); MSI-F6 (△); MSI-F7 (□); MSI-F9 (●); MSI-F11 (▲); MSI-F15 (■); MSI-F18 (●); MSI-F20 (▲); MSI-F22 (■). **B:** Transverse relaxation rates for peptides as plotted as a function of sequence. The data were obtained using $1/\tau_{cp}$ = 2000 Hz so that the chemical exchange component is removed. The calculated R_2 = 50.8 Hz for the peptide bound to lipid bicelles (assuming relaxation is only due tumbling of bicelles) is indicated by the dashed line.

The transverse relaxation rates measured for the CF_3- reporter group varied significantly depending upon the location of the reporter nuclei (Figure 7B). The R_{2b} values for each of the peptides determined at $1/\tau_{cp}$ = 2000 Hz are given in Table 1. In

general, hydrophobic positions exhibit the fastest relaxation rates (e.g. Leu-7, R_{2b} = 73 Hz and Ala-15, R_{2b} = 89 Hz) indicating that these residues, which are predicted to be in the hydrophobic core of the coiled-coil, are relatively immobile. Hydrophilic positions, which are predicted to be on the surface of the coiled-coil and interact with lipid head groups, exhibit slower relaxation rates, indicating that these positions have greater mobility. Moreover, positions towards the center of the peptide, whether hydrophobic or hydrophilic, tend to relax faster than those at either end. Thus the slowest R_{2b} values are for terminal residues Gly-1 (33 Hz) and Lys-22 (27 Hz) whereas the lysine residues at the center of the peptide have significantly faster relaxation rates (Lys-11 = 43 Hz and Lys 18 = 46 Hz). This suggests that within the toroidal membrane pore formed by the peptide the central portion of the peptide is more restricted in its motion than are the ends.

The changes in chemical shift and transverse relaxation rates observed upon the peptides binding to bicelles appear to be independent of each other. Thus although the positions occupied by Ala-9 and Ala-15 both appear to be quite immobile (R_{2b} = 66 and 89 Hz respectively), Ala-9 shifts downfield by 1.16 ppm whereas Ala-15 shifts slightly upfield by -0.09 ppm. The chemical shifts of N- and C-terminal positions change in opposite directions (+1.28 and -0.15 ppm respectively) even though both positions exhibit similar mobility as judged by their R_{2b} values.

Discussion

Solution phase ^{19}F NMR has been used to study the dynamics of integral membrane proteins and the interactions of other proteins with the membranes [46-49]; these studies use proteins labeled with fluorinated tryptophan, phenylalanine and tyrosine

analogs into which can readily be incorporated biosynthetically. Our studies extend the use of ^{19}F NMR to examine the more transient and dynamic interactions of peptides with membranes and provide position-specific information on how the chemical environment and dynamics of residues in MSI-78 change upon binding to the lipid bilayer. The results show that ^{19}F NMR is a sensitive and generally useful method for interrogating peptide membrane-interactions in free solution.

Changes in ^{19}F chemical shift provide a simple way of detecting binding at concentrations that are close to those at which the peptide is biologically active. For MSI-78, spectral shifts were apparent for all the TfeG-labeled peptides, regardless of the position of the label. In these studies spectra were acquired on a 500 MHz spectrometer at peptide concentration of 400 μM to facilitate CPMG experiments, however 40 μM of TfeG-labeled peptide is readily detected at signal-to-noise of 4:1 after 64 scans in a spectrum that takes only 2 min to acquire. Another advantage is that the wide range of chemical shift changes observed potentially allows multiple peptides to be studied in one experiment (or multiple fluorine probes to be introduced into one peptide), so that more complex multi-component interactions can be studied.

Solvent, H_2O/D_2O, isotope effects on the fluorine chemical shift have been employed to study the structure and dynamics of large proteins [16,18]. Our results demonstrate that this technique also provides a simple but effective probe of peptide-membrane interactions. In this case, it is evident that, even when embedded in the membrane, the lysine positions remain pre-dominantly solvated whereas the hydrophobic positions are extensively shielded from the solvent. This supports the model for MSI-78

forming toroidal pores in the lipid bilayer [30], in which the lysine side-chains interact with hydrophilic lipid head groups.

From the measurements of transverse relaxation rates along the peptide backbone we have obtained detailed information on the local dynamics of MSI-78 bound to the lipid bilayer.. The hydrophobic positions of the amphipathic peptide that form the core of the coiled-coil are the least dynamic positions. This suggests a well-packed core in which side-chain rotations are restricted. The positively charged face of the coiled coil, which interacts with the hydrophilic lipid head groups, is more dynamic, although far less so than in the unbound peptide. Moreover, positions towards the center of the peptide appear less dynamic than the ends of the peptide. This would be consistent with the peptides forming toroidal pores in the lipid bilayer, as has been deduced for MSI-78. Assuming the peptide sits centrally in the pore, constriction of the peptide by the lipid bilayer would by greatest at the center of the pore, where the opening is narrowest.

The ^{19}F transverse relaxation data provide new and more detailed information on the local dynamics of a peptide changes upon binding to a lipid membrane. The changes in ^{19}F chemical shift can only be interpreted qualitatively using with the structural model of MSI-78 determined by NMR [31]. However, DFT methods have been used to calculate ^{19}F chemical shifts for a large number of fluorinated small molecules [50,51] with a reasonable degree of accuracy. Particularly pertinent to our experiments, it was found that hindered rotation of CF_3 groups led to an upfield shift of the ^{19}F signal [51]. Future advances in computational methods will likely allow ^{19}F chemical shifts in fluorinated peptides and proteins to be calculated. This would permit the relationship between

structure, dynamics and chemical shift to be quantitatively understood, providing a further tool for analyzing peptide-membrane interactions.

In conclusion, these studies demonstrate that ^{19}F NMR provides a relatively sensitive and general technique for investigating the interactions of peptides and proteins with their membrane targets. In particular, the chemical shift and transverse relaxation rates are highly sensitive to the position of the fluorine label in the peptide and inform on changes in local peptide motions. Future work will aim towards using fluorine NMR to study peptide-membrane interactions *in vivo* at physiologically relevant concentrations.

Acknowledgements

We thank Dr. Jeetender Chugh for helpful discussions and advice on CPMG experiments.

References

(1) Brogden, K. A. *Nature Reviews Microbiology* **2005**, *3*, 238.
(2) Dennison, S. R.; Whittaker, M.; Harris, F.; Phoenix, D. A. *Current Protein & Peptide Science* **2006**, *7*, 487.
(3) Dhople, V.; Krukemeyer, A.; Ramamoorthy, A. ***BBA-Biomembranes*** **2006**, *1758*, 1499.
(4) Gennaro, R.; Zanetti, M. *Biopolymers* **2000**, *55*, 31.
(5) Koyama, Y.; Motobu, M.; Hikosaka, K.; Yamada, M.; Nakamura, K.; Saido-Sakanaka, H.; Asaoka, A.; Yamakawa, M.; Sekikawa, K.; Kitani, H.; Shimura, K.; Nakai, Y.; Hirota, Y. *International Immunopharmacology* **2006**, *6*, 234.
(6) Nomura, K.; Corzo, G. *Biochimica Et Biophysica Acta-Biomembranes* **2006**, *1758*, 1475.
(7) Ramamoorthy, A.; Thennarasu, S.; Lee, D. K.; Tan, A. M.; Maloy, L. *Biophysical Journal* **2006**, *91*, 206.
(8) Porcelli, F.; Verardi, R.; Shi, L.; Henzler-Wildman, K. A.; Ramamoorthy, A.; Veglia, G. *Biochemistry* **2008**, *47*, 5565.
(9) Powers, J. P. S.; Tan, A.; Ramamoorthy, A.; Hancock, R. E. W. *Biochemistry* **2005**, *44*, 15504.
(10) Porcelli, F.; Buck, B.; Lee, D. K.; Hallock, K. J.; Ramamoorthy, A.; Veglia, G. *Journal of Biological Chemistry* **2004**, *279*, 45815.

(11) Mani, R.; Cady, S. D.; Tang, M.; Waring, A. J.; Lehrer, R. I.; Hong, M. *Proc. Natl. Acad. Sci. (USA)* **2006**, *103*, 16242.
(12) Wu, X.; Mani, R.; Tang, M.; Buffy, J. J.; Waring, A. J.; Sherman, M. A.; Hong, M. *Biochemistry* **2006**, *45*, 8341.
(13) Afonin, S.; Grage, S. L.; Ieronimo, M.; Wadhwani, P.; Ulrich, A. S. *Journal of the American Chemical Society* **2008**, *130*, 16512.
(14) Buffy, J. J.; Waring, A. J.; Hong, M. *Journal of the American Chemical Society* **2005**, *127*, 4477.
(15) Ieronimo, M.; Afonin, S.; Koch, K.; Berditsch, M.; Wadhwani, P.; Ulrich, A. S. *Journal of the American Chemical Society* **2010**, *132*, 8822.
(16) Evanics, F.; Bezsonova, I.; Marsh, J.; Kitevski, J. L.; Forman-Kay, J. D.; Prosser, R. S. *Biochemistry* **2006**, *45*, 14120.
(17) Hull, W. E.; Sykes, B. D. *Biochemistry* **1974**, *13*, 3431.
(18) Hull, W. E.; Sykes, B. D. *Biochemistry* **1976**, *15*, 1535.
(19) Kitevski-LeBlanc, J. L.; Evanics, F.; Prosser, R. S. *Journal of Biomolecular Nmr* **2009**, *45*, 255.
(20) Prosser, R. S.; Luchette, P. A.; Westerman, P. W.; Rozek, A.; Hancock, R. E. W. *Biophysical Journal* **2001**, *80*, 1406.
(21) Huang, H. W.; Chen, F.-Y.; Lee, M.-T. *Phys Rev Lett* **2004**, *92*, 198304.
(22) Oren, Z.; Shai, Y. *Biopolymers* **1998**, *47*, 451.
(23) Wimley, W. C. *Acs Chemical Biology* **2010**, *5*, 905.
(24) Hancock, R. E. W.; Lehrer, R. *Trends Biotechnol.* **1998**, *16*, 82.
(25) Shai, Y. *Biochimica Et Biophysica Acta-Biomembranes* **1999**, *1462*, 55.
(26) Wu, M. H.; Maier, E.; Benz, R.; Hancock, R. E. W. *Biochemistry* **1999**, *38*, 7235.
(27) Epand, R. F.; Ramamoorthy, A.; Epand, R. M. *Protein and Peptide Letters* **2006**, *13*, 1.
(28) Oren, Z.; Shai, Y. *Biochemistry* **1997**, *36*, 1826.
(29) Selsted, M. E.; Novotny, M. J.; Morris, W. L.; Tang, Y. Q.; Smith, W.; Cullor, J. S. *Journal of Biological Chemistry* **1992**, *267*, 4292.
(30) Hallock, K. J.; Lee, D. K.; Ramamoorthy, A. *Biophysical Journal* **2003**, *84*, 3052.
(31) Porcelli, F.; Buck-Koehntop, B. A.; Thennarasu, S.; Ramamoorthy, A.; Veglia, G. *Biochemistry* **2006**, *45*, 5793.
(32) Gottler, L. M.; Lee, H. Y.; Shelburne, C. E.; Ramamoorthy, A.; Marsh, E. N. G. *Chembiochem* **2008**, *9*, 370.
(33) Gottler, L. M.; De la Salud-Bea, R.; Shelburne, C. E.; Ramamoorthy, A.; Marsh, E. N. G. *Biochemistry* **2008**, *47*, 9243
(34) Hsieh, K. H.; Needleman, P.; Marshall, G. R. *Journal of Medicinal Chemistry* **1987**, *30*, 1097.
(35) Wang, P.; Tang, Y.; Tirrell, D. A. *Journal of the American Chemical Society* **2003**, *125*, 6900.
(36) Meng, H.; Kumar, K. *Journal of the American Chemical Society* **2007**, *129*, 15615.
(37) Buer, B. C.; Chugg, J.; Al-Hashimi, H. M.; Marsh, E. N. G. *Biochemistry* **2010**, *49*, 5760
(38) Tsushima, T.; Kawada, K.; Ishihara, S.; Uchida, N.; Shiratori, O.; Higaki, J.; Hirata, M. *Tetrahedron* **1988**, *44*, 5375.

(39) Lee, H. Y.; Lee, K. H.; Al-Hashimi, H. M.; Marsh, E. N. G. *J. Am. Chem. Soc.* **2006**, *128*, 337.
(40) Gottler, L. M.; de la Salud-Bea, R.; Marsh, E. N. G. *Biochemistry* **2008**, *47*, 4484.
(41) Shelburne, C. E.; An, F. Y.; Dholpe, V.; Ramamoorthy, A.; Lopatin, D. E.; Lantz, M. S. *Journal of Antimicrobial Chemotherapy* **2007**, *59*, 297.
(42) Huhta, M. S.; Chen, H.-P.; Hemann, C.; Hille, C. R.; Marsh, E. N. G. *Biochem. J.* **2001**, *355*, 131.
(43) Lau, E. Y.; Gerig, J. T. *Journal of the American Chemical Society* **2000**, *122*, 4408.
(44) Dubois, B. W.; Evers, A. S. *Biochemistry* **1992**, *31*, 7069.
(45) Luz, Z.; Meiboom, S. *J. Chem. Phys.* **1963**, *39*, 366.
(46) Ahmed, A. H.; Loh, A. P.; Jane, D. E.; Oswald, R. E. *Journal of Biological Chemistry* **2007**, *282*, 12773.
(47) Anderluh, G.; Razpotnik, A.; Podlesek, Z.; Macek, P.; Separovic, F.; Norton, R. S. *Journal of Molecular Biology* **2005**, *347*, 27.
(48) Li, C. G.; Lutz, E. A.; Slade, K. M.; Ruf, R. A. S.; Wang, G. F.; Pielak, G. J. *Biochemistry* **2009**, *48*, 8578.
(49) Wang, G. F.; Li, C. G.; Pielak, G. J. *ChemBioChem* **2010**, *11*, 1993.
(50) Fukaya, T.; Ono, T. *J. Comput. Chem.* **2004**, *25*, 51.
(51) Liu, Z.; Goddard, J. D. *J. Phys. Chem. A* **2009**, *113*, 13921.