Comparison of the structures and stabilities of coiled-coil proteins containing hexafluoroleucine and *t*-butylalanine provides insight into the stabilizing effects of highly fluorinated amino acid side-chains

Benjamin C. Buer,[1] Jennifer L. Meagher,[2] Jeanne A. Stuckey,[2,3] and E. Neil G. Marsh[1,3]*

[1]Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109
[2]Life Sciences Institute, University of Michigan, Ann Arbor, Michigan 48109
[3]Department of Biological Chemistry, University of Michigan Medical School, Ann Arbor, Michigan 48109

Abstract: Highly fluorinated analogs of hydrophobic amino acids are well known to increase the stability of proteins toward thermal unfolding and chemical denaturation, but there is very little data on the structural consequences of fluorination. We have determined the structures and folding energies of three variants of a *de novo* designed 4-helix bundle protein whose hydrophobic cores contain either hexafluoroleucine (hFLeu) or *t*-butylalanine (tBAla). Although the buried hydrophobic surface area is the same for all three proteins, the incorporation of tBAla causes a rearrangement of the core packing, resulting in the formation of a destabilizing hydrophobic cavity at the center of the protein. In contrast, incorporation of hFLeu, causes no changes in core packing with respect to the structure of the nonfluorinated parent protein which contains only leucine in the core. These results support the idea that fluorinated residues are especially effective at stabilizing proteins because they closely mimic the shape of the natural residues they replace while increasing buried hydrophobic surface area.

Keywords: coiled-coil proteins; *de novo* designed proteins; fluorinated proteins; hexafluoroleucine; unnatural amino acids; protein structure

Introduction

The introduction of "unnatural" or noncanonical amino acid side-chains into proteins has proved an effective and versatile approach to designing proteins with novel properties. Applications include enhancing the chemical and metabolic stability of proteins, and introducing chemical functionality capable of acting either catalytically, or as a reporter or sensor into proteins.[1–4] The development of methods to incorporate noncanonical amino acids into proteins biosynthetically, or semisynthetically through native ligation has greatly increased the utility of this approach to protein design.[4–7]

However, to fully utilize the power of these techniques one needs to understand the structural consequences of incorporating noncanonical side-chains into a protein. Whereas there are now many thousands of high-resolution structures for "natural" proteins, there are very few structures for proteins containing noncanonical residues so that the effects of these residues on protein structure remain unclear.

Fluorinated amino acids represent an important class of noncanonical amino acids and a wide variety of fluorinated analogs have been incorporated into proteins.[8–11] The substitution of fluorine for hydrogen is generally considered sterically nonperturbing and many "lightly" fluorinated amino acids can be processed by endogenous tRNA synthetases.[12–14] However, our laboratory and others have focused on introducing highly fluorinated analogs of residues such as valine, leucine, isoleucine, and phenylalanine into proteins as a means of enhancing stability.[12,15–23] In most cases, these modified proteins display significantly increased stability toward unfolding by chemical denaturants and organic solvents, as well as increased resistance to proteolytic degradation.[20,24–26]

Currently, our understanding of how fluorination stabilizes protein folding is impeded by the lack of structures for proteins containing highly fluorinated amino acids. Gellman and coworkers recently determined both NMR and X-ray structures of chicken villin headpiece subdomain (cVHP) in which pentafluorophenylalanine (pFPhe) replaced one of three Phe residues in the core.[27,28] Crystal structures for pFPhe variants at positions 10 and 17 show the pFPhe ring adopts a conformation very similar to Phe in the wild-type protein, however only the pFPhe10 substitution is actually stabilizing. Our laboratory has focused on the *de novo* designed, antiparallel 4-helix bundle protein, $\alpha_4 H$, as a model system to understand the effects of fluorination on protein structure and stability. Toward this goal, we recently solved and compared the crystal structures of $\alpha_4 H$, which contains Leu at the core "a" and "d" positions, and $\alpha_4 F_3 a$, a derivative of $\alpha_4 H$ which contains hFLeu at the "a" and Leu "d" positions. Incorporation of hFLeu stabilizes the structure of $\alpha_4 F_3 a$ by 0.8 kcal mol^{-1} hFLeu-residue^{-1} but is remarkably nonperturbing to the structure, even though hFLeu is ~32 Å^3 larger than Leu. We attributed this to the fact that although side-chain volume is increased, fluorination closely preserves the shape of the side-chain so that it can still be accommodated within the tightly packed hydrophobic core.

In this article, we expand our investigation into the effects of fluorination on protein structure by analyzing high-resolution X-ray crystal structures of three further variants of $\alpha_4 H$. In $\alpha_4 F_3 d$ the Leu residues at the three "d" positions are now substituted by hFLeu, allowing the structural effects of introducing hFLeu at the "a" and "d" positions to be compared. $\alpha_4 F_3 (6–13)$ contains the same number of hFLeu residues as $\alpha_4 F_3 a$ and $\alpha_4 F_3 d$, but the packing is now altered so that hFleu is introduced at one "a" position (10) adjacent to two "d" positions (6 and 13); this allows interaction between hFLeu residues on adjacent "a" and "d" positions to be evaluated. As a nonfluorinated comparison we have determined the structure of $\alpha_4 tbA_6$, which contains β-*t*-butylalanine (tBAla) at all "a" and "d" positions. This side-chain has a surface area intermediate between that of Leu and hFLeu, so that introducing 24 tBAla residues into the core of $\alpha_4 tbA_6$ should result in a buried hydrophobic surface area nearly identical to that of either $\alpha_4 F_3 d$ or $\alpha_4 F_3 (6–13)$. Therefore, any differences in structure and stability are expected to arise primarily from the difference in the shapes of the side-chains.

Results

Effect of hFLeu and tBAla on stability of α_4 proteins

We have previously described the synthesis $\alpha_4 F_3 d$ and determined its free energy of folding.[21] The introduction of three hFLeu residues stabilizes the folding of $\alpha_4 F_3 d$ by -8.6 kcal mol^{-1} relative to $\alpha_4 H$. For the present studies we initially synthesized three new α_4 variants: $\alpha_4 F_3 (6–13)$, which contains hFLeu in place of Leu at positions 6, 10, and 13; $\alpha_4 F_3 (17–24)$, which contains hFLeu in place of Leu at positions 17, 20, and 24; and $\alpha_4 tbA_6$, which contains the leucine analog β-*t*-butylalanine in place of Leu at all 6 "a" and "d" positions.

The free energies of folding for each protein were determined by titration with guanidinium hydrochloride and following ellipticity changes at 222 nm (Supporting Information Fig. S1). Each protein unfolded in a cooperative manner that was well-fitted by assuming a two-state transition between unfolded monomer and folded tetramer. The results are summarized in Table I. Although all these variants are, as expected, more stable than $\alpha_4 H$, it is evident that incorporating hFLeu at mixed "a" and "d" positions is not as stabilizing as incorporating hFLeu at either all "a" or all "d" positions. Similarly, tBAla does not stabilize the structure of α_4 as effectively as hFLeu.

Crystallization of $\alpha_4 F_3 d$, $\alpha_4 F_3 (6$-$13)$ and $\alpha_4 tbA_6$

To investigate how the changes in thermodynamic stability are related to changes in structure we crystallized three of these proteins, $\alpha_4 F_3 d$, $\alpha_4 F_3 (6–13)$ and $\alpha_4 tbA_6$ and determined their structures by X-ray crystallography. We were unable to obtain well-diffracting crystals of $\alpha_4 F_3 (17–24)$, precluding this protein from structural comparison.

Table I. *Free Energies of Folding for the Proteins Described in This Study*

Protein(10)	$\Delta G^{\mathrm{o}}_{\mathrm{fold}}$ (kcal mol^{-1})	m (kcal mol^{-1} M_{GuHCl}^{-1})	$\Delta\Delta G^{\mathrm{o}}_{\mathrm{fold}}$ from α_4H (kcal mol^{-1})	$\Delta\Delta G^{\mathrm{o}}_{\mathrm{fold}}$ (kcal mol^{-1}hFLeu residue^{-1})
α_4H	-18.0 ± 0.2	-1.04 ± 0.04	NA	NA
$\alpha_4$tbA$_6$	-22.3 ± 0.1	-1.78 ± 0.03	-4.3 ± 0.2	NA
α_4F$_3$(6–13)	-23.7 ± 0.2	-1.98 ± 0.05	-5.7 ± 0.3	-0.48 ± 0.03
α_4F$_3$(17–24)	-23.1 ± 0.1	-2.16 ± 0.03	-5.1 ± 0.2	-0.43 ± 0.03
α_4F$_3$d	-26.6 ± 0.1	-2.24 ± 0.03	-8.6 ± 0.2	-0.72 ± 0.02

α_4F$_3$d, α_4F$_3$(6–13) and $\alpha_4$tbA$_6$ crystallized under similar buffer and precipitant conditions to other α_4 proteins (see Materials and Methods section). In each case the asymmetric unit comprises two peptide chains, A and B; the tetrameric bundle was generated from the appropriate symmetry operations. α_4F$_3$d crystallized in the $P2_12_12$ space group and its structure determined at 1.19 Å resolution (Table II), with all residues being well resolved except the last residue of both the A and B chains. α_4F$_3$(6–13) crystallized in the space group $I4_1$ and its structure was determined at 1.48 Å resolution (Table II); in this case all but the last two residues of both the A and B chains were well resolved. $\alpha_4$tbA$_6$ crystallized in space group $P2_12_12$, and its structure was determined at 1.54 Å resolution (Table II); all but the first two and last two residues of chain A were resolved. The structures were solved by molecular replacement using the structure of α_4H as a model.

As expected based on the structure of α_4H, each protein adopted an anti-parallel 4-helix bundle structure in which complementary electrostatic interactions between interfacial residues at the "b" and "e" positions and "c" and "g" positions enforce the antiparallel topology. Overall the helices of α_4F$_3$d, α_4F$_3$(6–13) and $\alpha_4$tbA$_6$ superimpose on the structure

of α_4H with side-chain C$_\alpha$, C$_\beta$, and C$_\gamma$ atom rmsd values of $\sim$1 Å or smaller (Table III).

Crystal packing effects

Interestingly, all the α_4 proteins whose structures we have now solved crystallize in lattices with unusually low solvent contents (three proteins crystallized in the $P2_12_12$ space group and three in the $I4_1$ space group). Whereas the average fractional solvent volume for protein crystal structures in the Protein Data Bank is $\sim$50%, the α_4 proteins crystallize with only $\sim$30% solvent volume. To achieve this, the α_4 proteins pack in the crystal end-to-end with narrow solvent channels between tetramers, as shown in Figure 1. In some cases this results in the charged residues at interfacial positions forming salt-bridges between adjacent tetramers, for example, ArgA16 in α_4F$_3$d (Fig. 2), rather than between the α-helices of the 4-helix bundle as they were designed to do.

Structures of α_4F$_3$d and α_4F$_3$(6-13)

The Leu/hFLeu residues in the α_4 proteins pack the hydrophobic core in a regular manner forming six layers (arbitrarily numbered from the N-terminus of chain A), with each layer comprising side-chains

Table II. *Data Collection and Refinement Statistics for α_4F$_3$d, α_4F$_3$(6–13) and $\alpha_4$tbA$_6$*

Data set	α_4F$_3$d	α_4F$_3$(6–13)	$\alpha_4$tbA$_6$
Space group	$P2_12_12$	$I4_1$	$P2_12_12$
Unit cell	$a = 30.82$; $b = 39.25$;	$a = b = 49.58$;	$a = 31.31$; $b = 37.42$;
	$c = 41.23$	$c = 41.57$	$c = 40.76$
	$\alpha = \beta = \gamma = 90$	$\alpha = \beta = \gamma = 90$	$\alpha = \beta = \gamma = 90$
Wavelength, Å	0.97872	0.97856	0.97872
d_{min}, Å	1.19 (1.19–1.21)	1.48 (1.48–1.51)	1.54 (1.54–1.57)
R_{sym}, %	5.7 (15.5)	4.4 (46.6)	4.5 (52.2)
$<I/\sigma_I>$	20 (10)	20 (3)	20 (3)
Completeness, %	98.1 (76.2)	99.9 (100.0)	98.2 (93.9)
Redundancy	7.9 (7.6)	11.0 (10.9)	10.2 (9.3)
Refinement statistics			
Data range, Å	11.16–1.19	10.00–1.48	27.56–1.54
R factor, %	17.8	25.2	24.2
R_{free}, %	18.7	31.8	24.5
Protein atoms, #	517	487	431
Water molecules, #	49	26	55
Reflections, #	16,392	8484	7348
rmsd			
Bonds (Å)	0.009	0.010	0.008
Angles (°)	1.08	1.19	0.96

Table III. *RMSD Values (Å) of C_α, C_β, and C_γ Atom Coordinates Between all Pairs of Structures from This Study and Previously Reported Structures of α_4H and α_4F_3a*[29]

| Protein	α_4H	α_4F_3a	α_4F_3d	$\alpha_4F_3(6-13)$
α_4H	—			
α_4F_3a	1.01	—		
α_4F_3d	0.55	0.92	—	
$\alpha_4F_3(6-13)$	0.67	0.88	0.49	—
α_4tbA_6	1.11	1.08	1.13	0.97

from two "a" positions and two "d" positions. The modeled electron density for a cross-section of layer 4 of α_4F_3d and $\alpha_4F_3(6-13)$ is shown in Figure 2. The electron density for the hFLeu trifluoromethyl groups is well defined with the individual fluorine atoms clearly resolved. The space-filling representation of layer 4 (Fig. 2) shows that the hydrophobic cores are tightly packed in a similar manner to the parent α_4H protein. However, some charged residues in the external positions of the tetramers exhibit marked conformational differences. These differences may be attributed to crystal packing effects, as noted above.

For antiparallel coiled-coil bundles, there are two different helical interfaces formed by residues at "b" and "e" and "c" and "g" positions. In α_4H this results in Leu at "a" and "d" positions occupying distinctly different environments, with "a" residues packing toward the "b–e" interface and "d" residues packing toward the "c–g" interface. The inter-helical spacing is also different, with the "b–e" interface being narrower than the "c–g" interface; this is a consequence of the heptad repeat placing Leu residues in "a" positions toward the central core axis while Leu residues in "d" positions are oriented away from the central axis, toward adjacent helices. In α_4F_3d this packing arrangement is preserved, with hFLeu packing in a knobs-into-holes manner along

the "c–g" interface and Leu packing along the "b–e" interface to form a fluorinated "stripe" (Fig. 3)—an arrangement that exactly mirrors that found for α_4F_3a. For $\alpha_4F_3(6-13)$ the positions of the hFLeu residues results in a more complex pattern. The "b–e" interface of has two clusters of hFLeu residues that project into the core, which are separated by LeuA17 and LeuB17. The "c–g" interface forms a chain of fluorinated residues that crosses from one helix to the other through contacts between hFLeuA13 and hFLeuD13.

The hydrophobic cores of both α_4F_3d and $\alpha_4F_3(6-13)$ appear to be efficiently packed, with each core residue making extensive van der Waals contacts with other residues, as illustrated for one residue, LeuA17 in α_4F_3d, in Figure 4. The packing efficiency, defined as the volume occupied by the peptide chains divided by the total volume of the core, is $\sim88\%$ for α_4F_3d and $\sim89\%$ for $\alpha_4F_3(6-13)$. These values are very similar to those previously calculated for α_4H, and α_4F_3a, which have packing efficiencies of $\sim90\%$ (29). Thus the additional stability imparted by hFLeu does not result from more efficient core packing in α_4F_3d and $\alpha_4F_3(6-13)$, that is, fewer void spaces in the core. Instead, it appears better attributed to the increase in the hydrophobic buried surface area due to hFLeu.

We compared the core packing of Leu and hFLeu residues for each of the six layers of α_4F_3d and $\alpha_4F_3(6-13)$ with the parent protein α_4H (Fig. 5). The central 4 layers (layers 2–5) of α_4H, α_4F_3d and $\alpha_4F_3(6-13)$ are packed such that the residues at "a" positions extend into the center of the protein core towards the corresponding residue at the opposite "a" position, whereas the residues at the "d" positions are less deeply buried and are oriented towards the "c–g" interface. In the outer layers (layers 1 and 6) the packing arrangement is reversed, with residues in "d" positions extending into the center of the
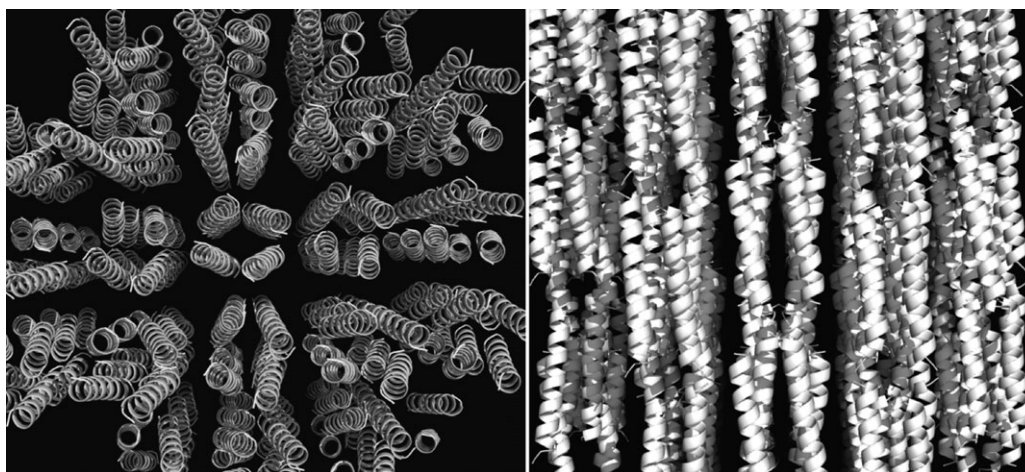
Figure 1. Crystal packing of $\alpha_4F_3(6-13)$ in space group $I4_1$. The proteins pack unusually tightly resulting in a solvent content of only 30%.

Stabilizing Effects of Fluorous Amino Acids
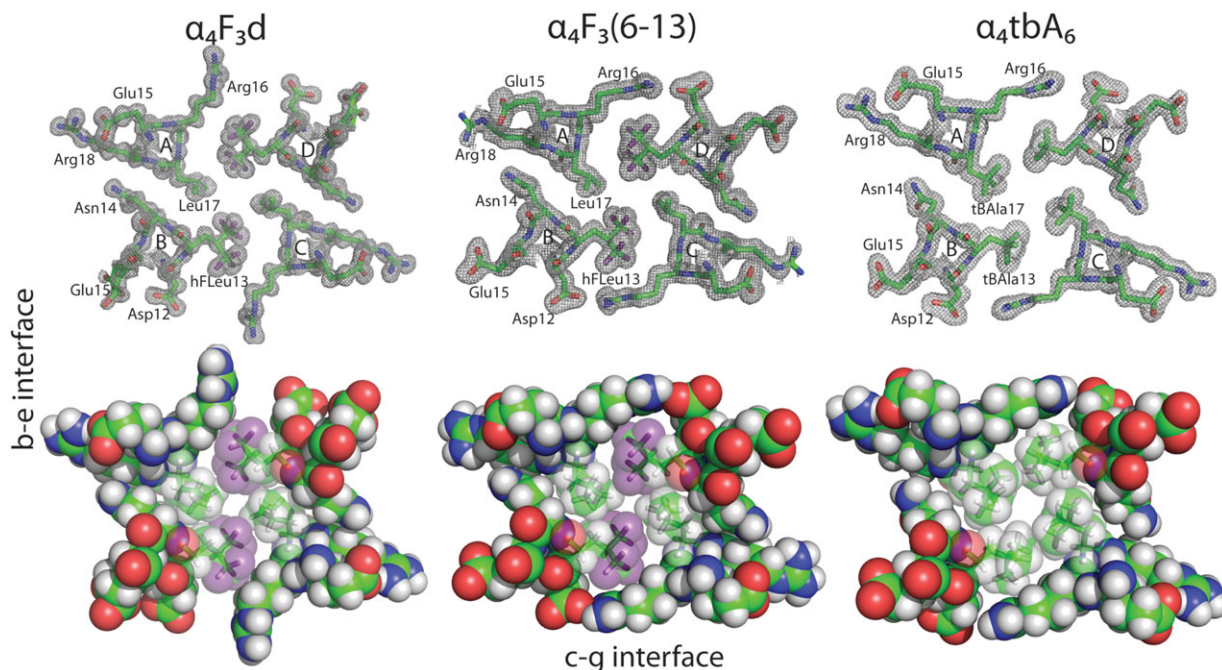
Figure 2. Top: Representative electron density ($2F_o - F_c$) maps for each protein with residues contoured at 1.0σ. Bottom: Space-filling representations of the hydrophobic core illustrating how fluorination conserves the tight packing of side-chains. Fluorine atoms are colored purple. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

helical bundle to make contact with their counterparts a "d," and residues at the "a" and "a'" positions are oriented towards the "b–e" interface.

The introduction of the significantly larger hFLeu residues at either the "a" or "d" position is surprisingly benign. It has essentially no effect on the packing of the layer in which it is introduced nor on adjacent layers (some minor changes in rotomer conformations of adjacent residues occur). Notably, there is no evidence that fluorinated residues form preferential contacts with each other, as has

been predicted based on the tendency of perfluorinated solvents to self-segregate,[15,30,31] also known as the fluorous effect.

Assessing the influence of side-chain shape on core packing and stability

We previously hypothesized that the ability of fluorinated amino acids such as hFLeu to stabilize protein structures resulted from the fact that fluorination closely preserves the shape of the side-chain whilst increasing the overall hydrophobic volume. Indeed
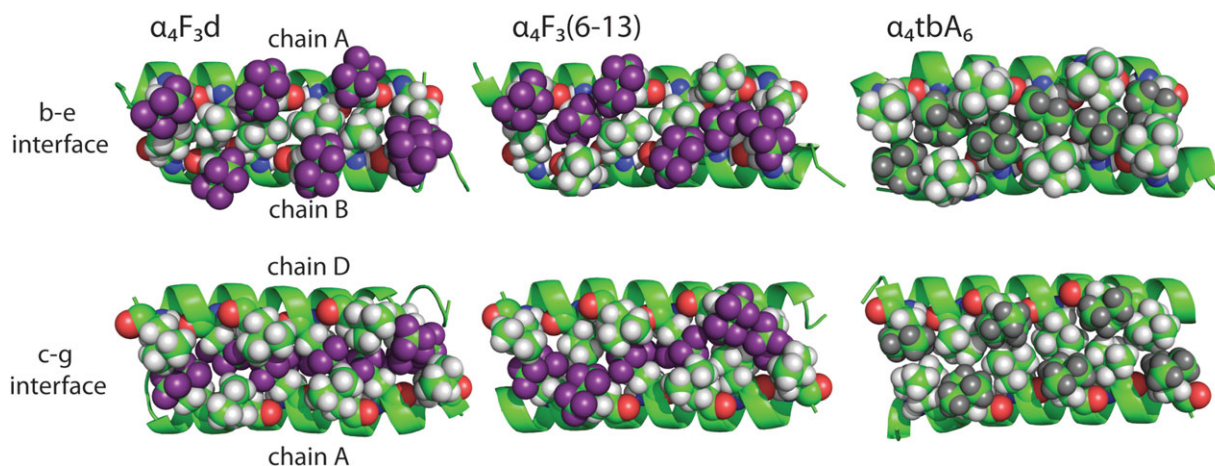
Figure 3. Layer 1 of the core is shown to the left; layer 6 to the right. In α_4tbA_6 residues at "a" positions are colored dark gray to distinguish them from residues at "d" positions. Fluorine atoms are colored purple. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
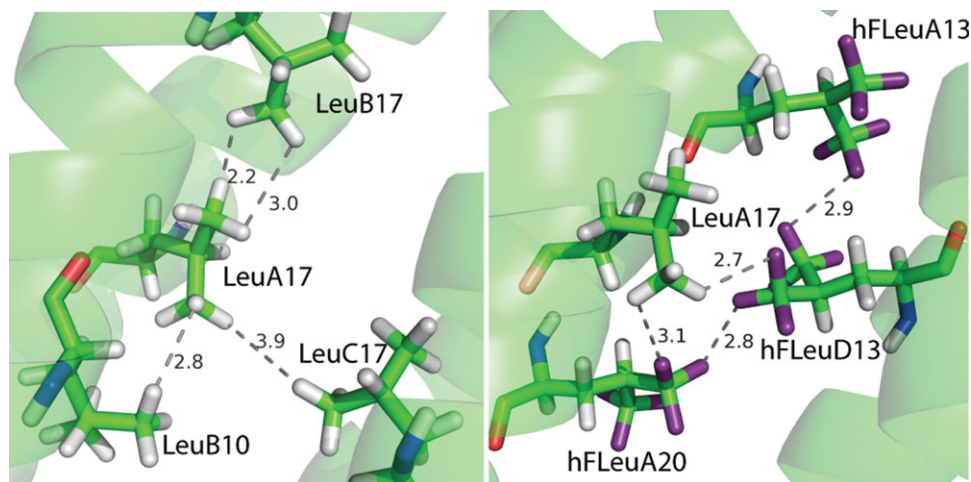
Figure 4. Left: Distances between LeuA17 and adjacent Leu residues in α_4F_3d. Right: The equivalent distance measurements between LeuA17 and adjacent hFLeu residues in α_4F_3d. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

the structures of three highly fluorinated proteins, α_4F_3a, α_4F_3d and $\alpha_4F_3(6{-}13)$, demonstrate the ability of hFLeu to integrate seamlessly into the 4-helix bundle. However, to test this hypothesis more rigorously we designed a nonfluorinated α_4 variant with a core volume very close to that of the α_4F_3 series of proteins. This protein, α_4tbA_6, incorporates β-t-butylalanine (tBAla) residues at all "a" and "d" positions. In essence, this introduces an additional methyl group at the C_γ of Leu and represents one of the least intrusive perturbations to the side-chain that can be accomplished exclusive of fluorination. The volume and surface area of tBAla is intermediate between Leu and hFleu, and thus when incorporated at both "a" and "d" positions the core volume and buried surface area of α_4tbA_6 almost exactly matches that of α_4F_3a, α_4F_3d and $\alpha_4F_3(6{-}13)$.

The packing of tBAla side-chains in α_4tbA_6 overall resembles that of α_4H (Figs. 3 and 5), however, the introduction of the tBAla group leads to an expansion of the "b–e" interface to accommodate this bulkier side-chain. This leads to a larger rmsd of the backbone atoms from the structure of α_4H than is seen for the fluorinated proteins. The packing of the hydrophobic core is significantly different from that of α_4H, α_4F_3d and $\alpha_4F_3(6{-}13)$, as shown in Figure 5. In layers 1 and 6 tBAla residues in "d" and "d'" positions extend into the center of the core to make contact, similar to the other structures. However, in layers 2–5 the side-chains abut each other across the "b–e" and "c–g" interfaces, creating a cavity that runs through the center of the protein that is closed at each end by residues in layers 1 and 6. This alteration in structure demonstrates its sensitivity to the change in shape of the t-butyl side-chain.

Measurements of the cavity volume using Chimera with a probe radius of 1.4 Å indicate the presence of five discrete cavities with a total volume of ~270 $Å^3$ (Fig. 6). 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_4tbA_6 displays additional electron density in the central cavity at ~50% occupancy (Fig. 6). The identity of the guest molecule is unknown, but a hydrophobic cavity of this size could accommodate molecular nitrogen or oxygen and may provide a nonspecific binding site for other small hydrophobic molecules.

Even though α_4tbA_6 contains a large hydrophobic cavity, the overall packing efficiency of the hydrophobic core is still ~90%, which is very similar to the other α_4 proteins. This somewhat surprising result arises because the α_4H and α_4F_3 proteins have more small void spaces within their structures that are not registered by the standard 1.4 Å radius probe used to identify cavities large enough to accommodate water or other small molecules. In α_4tbA_6, the t-butyl side-chains pack together in such a way that there are fewer small voids in the structure and this compensates for the large central cavity. If this cavity is excluded from the analysis, the packing efficiency of α_4tbA_6 increases to ~92%.

The presence of this cavity provides an explanation for the decreased stability of α_4tbA_6 with respect to $\alpha_4F_3(6{-}13)$ α_4F_3a and α_4F_3d, despite the fact that each contains very similar amounts of buried hydrophobic surface area. Large hydrophobic cavities are known from various studies to destabilize the folded state.[32] The energetic penalty for introducing a cavity into α_4tbA_6 can be estimated using the generally accepted value of 24 cal mol^{-1} Å^{-3} for the hydrophobic effect in proteins.[33] The calculated reduction in stability is ~4.8 kcal mol^{-1}, which is in reasonable agreement with the observed decrease in stability compared to fluorinated proteins with similar hydrophobic surface areas.
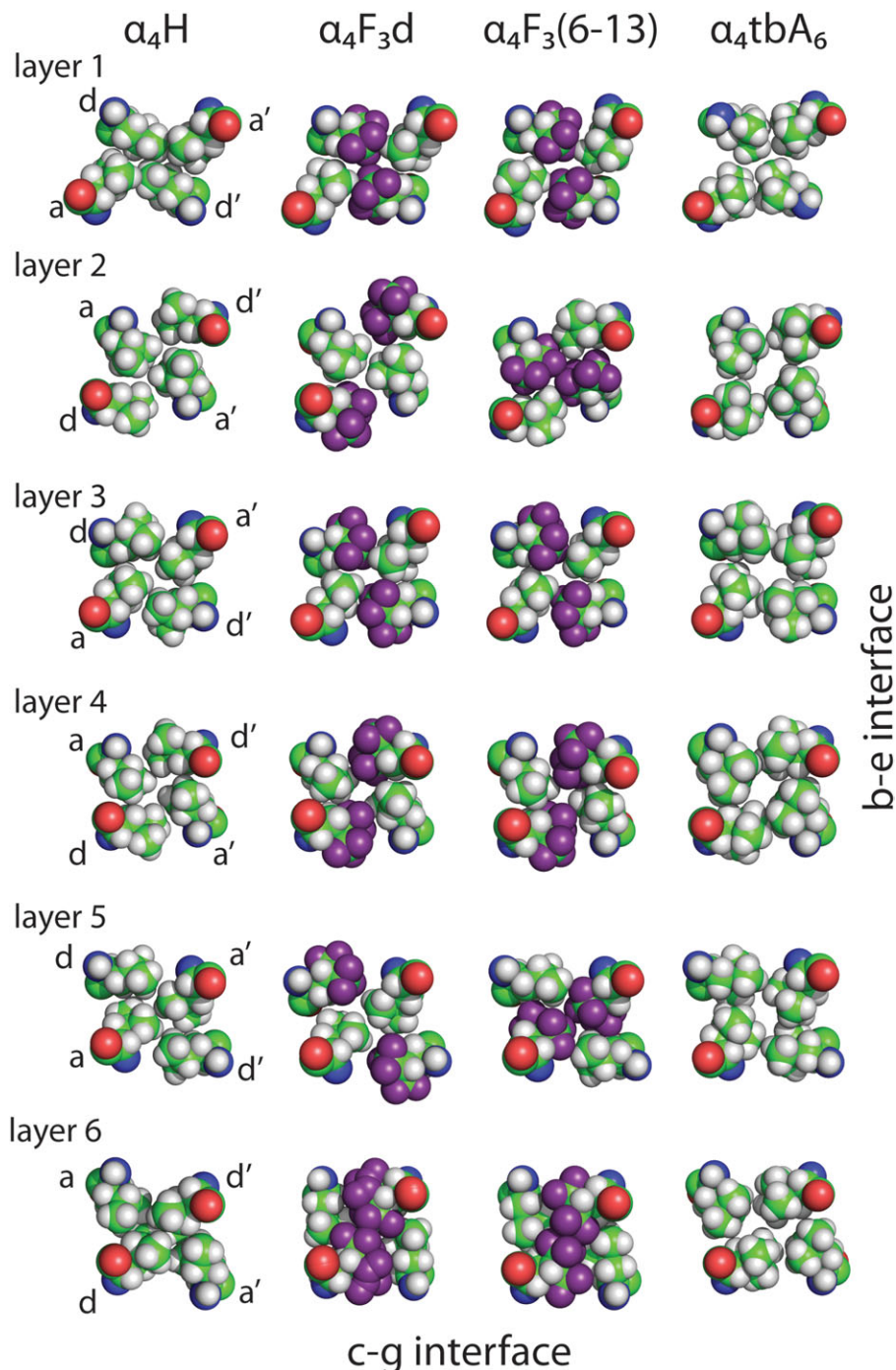
Figure 5. Comparison of core packing in fluorinated and nonfluorinated α_4 proteins. The pattern of hydrophobic contacts is generally unchanged by fluorination, despite the hFLeu side-chain being significantly larger, whereas the introduction of the tBAla side-chain results in the formation of a central cavity in the protein. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Discussion

By determining the structures of these highly fluorinated model proteins we aim to better understand the remarkable and seemingly quite general ability of fluorinated amino acids to stabilize protein structure. In particular, we wanted to test whether preferential interactions between fluorinated residues,[15,16,34,35] sometimes referred to as the "fluorous effect," were responsible for the stabilizing effects of fluorinated amino acids. This idea is predicated upon the unusual phase-segregating properties of perfluorocarbon solvents, which selectively extract highly fluorinated small molecules from organic solvents into the perfluorocarbon solvent.[36,37] Our previous analysis of the structure of α_4F_3a revealed no evidence that contacts between fluorocarbon residues are more stabilizing than contacts between hydrocarbon residues, once differences in buried surface area are accounted for.
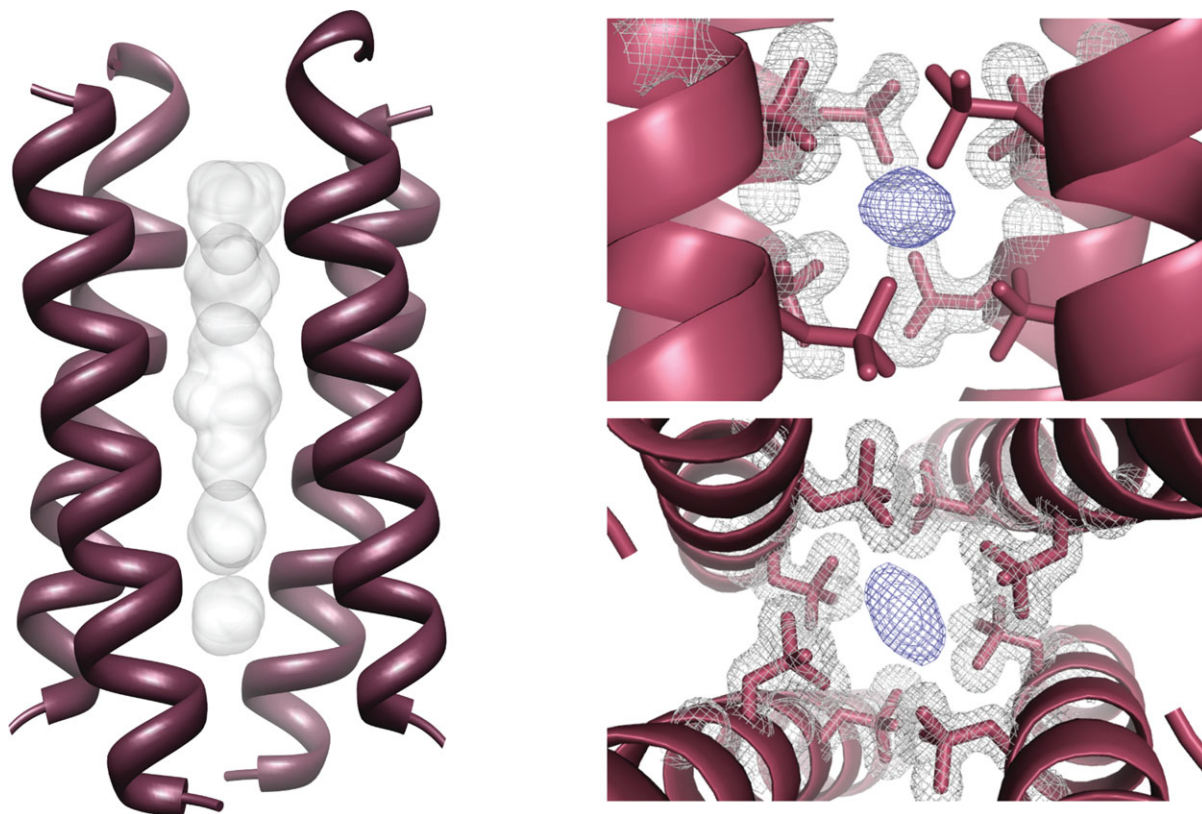
Figure 6. Left: Side view of $\alpha_4$tbA$_6$ displaying cavities in the hydrophobic core. Right: Electron density ($2F_o$ - F_c) maps for layers 3 and 4 of $\alpha_4$tbA$_6$ contoured at 1.0σ. Electron density for tBAla residues shown in gray and unmodeled electron density shown in blue. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

In the present study we have extended our investigation to examine the interactions of hFLeu in two different packing arrangements, one that introduces hFLeu in only "d" positions (α_4F$_3$d) and another that introduces hFLeu in alternating "a" and "d" positions. Here again we find no evidence that fluorous interactions contribute to the stability of these proteins. The core residues in α_4F$_3$d and α_4F$_3$(6–13) show no preference towards maximizing fluorine–fluorine contacts or minimizing fluorine–hydrogen contacts, as might have been expected if segregation of fluorous residues was important. Thus the distances between the hydrogen atoms of the α_4F$_3$d LeuA17 and adjacent Leu residues vary between 2.2 and 3.0 Å, whereas the distances between the hydrogen atoms of LeuA17 and adjacent hFleu residues range between 2.7 and 3.1 Å (Fig. 4). Furthermore, although α_4F$_3$d is more stable than α_4F$_3$(6–13), this does not correlate with an increase in the number of fluorine–fluorine contacts. As illustrated in Figures 3 and 5, hFleu residues in α_4F$_3$(6–13) actually form a more extensive set of fluorine–fluorine contacts than α_4F$_3$d, both across layers of the hydrophobic core (Fig. 5) and between layers of the core (Fig. 3).

Although they each contain the same number of hFLeu residues, α_4F$_3$(6–13) and α_4F$_3$(17–24) are sig-

nificantly less stable than α_4F$_3$a and α_4F$_3$d. These differences may be explained by vertical packing interactions of the "b–e" and "c–g" interfaces, the importance of which has been highlighted in studies on other antiparallel coiled-coil proteins.[38–40] The knobs-into-holes packing of "b–e" and "c–g" interfaces for α_4F$_3$a and α_4F$_3$d contain only like residues, in which Leu residues pack exclusively into the vertical hole created by two Leu residues of the adjacent helix and hFLeu residues follow a similar packing arrangement at the opposite interface. This creates a very regular packing arrangement. However, for α_4F$_3$(6–13) the pattern of hydrocarbon- and fluorocarbon-containing amino acids is altered, so that the hydrocarbon "stripe" that runs through the "b–e" interface is interrupted by hFLeu in position 10 and, similarly, the fluorocarbon "stripe" that runs through the "c–g" interface is interrupted by Leu in position 20 (Fig. 3). The result is that the interface displays a less regular packing arrangement because of the mixed Leu:hFLeu knobs-into-holes packing (Fig. 3), which may be less energetically stable.

The hFLeu side-chain is significantly larger than Leu, with the additional volume occupied by the two CF$_3$ groups being approximately equivalent to adding two further methyl groups to the Leu side-chain. Yet a striking feature of all the α_4F$_3$

fluorinated proteins is how little the introduction of 24 CF_3 groups into the core perturbs the structure of the protein. The rmsd of the C_α, C_β, and C_γ atoms from the structure of α_4H (Table III) provides one measure to quantify the degree of structural similarity. By this measure, the introduction of hFLeu at "d" positions is less perturbing than at the "a" positions. This is because the "b–e" interface, into which "a" position residues pack, is narrower than the "c–g" interface, which accommodates the "d" position residues. The helices in α_4F_3a (rmsd 1.01 Å) are consequently displaced further apart than the helices of α_4F_3d (rmsd 0.55 Å), with the structure of $\alpha_4F_3(6$–$13)$ (rmsd 0.67 Å) deviating by an intermediate amount.

We consider the most likely reason for the general compatibility of hFLeu to be that it closely matches the shape of the Leu side-chain it replaces. The structure of α_4tbA_6 supports this idea. Here, addition of a third methyl group to the Leu side-chain increases the hydrophobic surface area, but also subtly changes the shape of the side-chain. This substitution stabilizes the structure of α_4tbA_6 relative to α_4H, although less than we predicted, but also causes the packing of the core to be significantly altered. As a consequence its structure deviates by the greatest amount from α_4H, with a rmsd of 1.11 Å. The formation of a hydrophobic channel in the center of α_4tbA_6 was unanticipated and illustrates how a very modest change in volume and shape can alter the structure of the generally robust antiparallel 4-helix bundle motif. The channel presumably contributes to the lower stability of this protein compared to the hFLeu-containing proteins that have comparable buried hydrophobic surface areas.

In conclusion, these studies expand the number of crystal structures available for extensively fluorinated proteins and reveal how these larger, but similarly shaped side-chains can be accommodated within the structure of an existing protein core. The structure of α_4tbA_6 is, to our knowledge, the first structure of a protein containing the leucine analog tBAla. The change in core packing imposed by the tBAla side-chain underscores the uniquely non-perturbing nature of fluorinated leucine analogs. Our results suggest that hFLeu is likely to be generally compatible as a substitute for Leu in structures that include knobs-into-holes packing. We hope that these studies may provide a useful guide for future protein design efforts using highly fluorinated amino acids.

Materials and Methods

Materials and peptide synthesis

L-5,5,5,5',5',5'-hexafluoroleucine was synthesized as described previously[41] and converted to the Boc-pro-

tected derivative by standard procedures. F_{moc}-protected β-t-butyl-L-alanine was purchased from AnaSpec. hFLeu containing peptides were synthesized by manual Boc procedures and α_4tbA_6 was synthesized by manual F_{moc} procedures according to established protocols. 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^{-1}. Peptide identity was confirmed using MALDI-MS with a matrix of α-cyano-4-hydroxycinnamic acid.

Free energies of unfolding

The free energies of unfolding were determined by titration with guanidium hydrochloride at 25°C. Stock solutions were prepared containing 40 μM peptide (concentration of monomer) in 10 mM potassium phosphate buffer, pH 7.0, both with and without 8.0M 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 for several minutes the ellipticity at 222 nm was measured. The denaturation curves were fitted assuming a two-state equilibrium between unfolded monomer and folded tetramer as described previously.[21,35]

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_4tbA_6, 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 $\alpha_4F_3(6$–$13)$. Crystals were frozen with liquid N_2 in their mother liquor for data collection.

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 and 0.97857 Å, respectively, at -180°C. Data was processed and scaled with HKL2000.[42] The peptide $\alpha_4F_3(6$–$13)$ crystallized in space group $I4_1$ with unit cell parameters $a = b = 49.58$, $c = 41.57$, $\alpha = \beta = \gamma = 90°$. The peptides α_4tbA_6 and α_4F_3d crystallized in space group $P2_12_12$ with α_4tbA_6 unit cell parameters $a = 31.31$, $b = 37.42$, $c = 40.76$, $\alpha = \beta = \gamma = 90°$ and α_4F_3d unit cell parameters $a = 30.82$, $b = 39.25$, $c = 41.23$, $\alpha = \beta = \gamma = 90°$. All crystals contain a dimer in the asymmetric unit.

Phases were initially determined by molecular replacement using Phaser in the CCP4i suite of programs.[43] For α_4tbA_6 and $\alpha_4F_3(6$–$13)$ 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, $\alpha_4F_3(6\text{–}13)$ was used as a starting model with hFLeu10 mutated to Leu. The PRODRG web server was used to generate coordinates and restraint parameters for hFLeu, tBAla, and nonwater solvent molecules.[44] Peptide models were refined by rigid body refinement and restrained refinement using Buster.[45] Side-chains were built using Coot[46] with $2F_o - F_c$ and $F_o - F_c$ electron density maps from Buster.

The refinement of α_4tbA_6 to 1.54 Å resulted in $R_{work} = 24.2\%$ and $R_{free} = 24.5\%$. The refinement of α_4F_3d to 1.19 Å resulted in $R_{work} = 17.8\%$ and $R_{free} = 18.7\%$. The refinement of $\alpha_4F_3(6\text{–}13)$ to 1.48 Å resulted in $R_{work} = 25.2\%$ and $R_{free} = 31.8\%$. All residues from the three structures are in the allowed regions of the Ramachandran plot. Structures were validated with Molprobity,[47] Parvarti,[48] and whatcheck.[49] Areas of poor electron density were not modeled, these include: α_4tbA_6 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; $\alpha_4F_3(6\text{–}13)$ residues 1–4 and 27 of chain A and 26 and 27 of chain B. Data refinement statistics are given in Table II.

Volume and surface area calculations

Protein volumes and 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. Surface area and volume were measured from monomers of chains A and B with surface-exposed side-chains in the "b," "c," "e," "f," and "g" positions mutated *in silico* to alanine. Rmsd values of protein tetramers were determined using PyMOL.

Acknowledgments

References

1. Kiick KL, Saxon E, Tirrell DA, Bertozzi CR (2002) Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. Proc Natl Acad Sci USA 99:19.
2. Link AJ, Tirrell DA (2003) Cell surface labeling of *Escherichia coli* via copper (I)-catalyzed [3+ 2] cycloaddition. J Am Chem Soc 125:11164–11165.
3. Langer R, Tirrell DA (2004) Designing materials for biology and medicine. Nature 428:487–492.
4. Wang L, Xie J, Schultz PG (2006) Expanding the genetic code. Annu Rev Biophys Biomol Struct 35: 225–249.
5. Muir TW, Sondhi D, Cole PA (1998) Expressed protein ligation: a general method for protein engineering. Proc Natl Acad Sci USA 95:6705.
6. Wang L, Brock A, Herberich B, Schultz PG (2001) Expanding the genetic code of *Escherichia coli*. Science 292:498–500.
7. Muir TW (2003) Semisynthesis of proteins by expressed protein ligation. Ann Rev Biochem 72:249–289.
8. Yoder NC, Kumar K (2002) Fluorinated amino acids in protein design and engineering. Chem Soc Rev 31: 335–341.
9. Marsh ENG, Buer BC, Ramamoorthy A (2009) Fluorine—a new element in the design of membrane-active peptides. Mol BioSyst 5:1143–1147.
10. Buer BC, Marsh ENG (2012) Fluorine: a new element in protein design. Prot Sci 21:453–462.
11. Salwiczek M, Nyakatura EK, Gerling UIM, Ye S, Koksch B (2012) Fluorinated amino acids: compatibility with native protein structures and effects on protein–protein interactions. Chem Soc Rev 41:2135–2171.
12. Tang Y, Ghirlanda G, Petka WA, Nakajima T, DeGrado WF, Tirrell DA (2001) Fluorinated coiled-coil proteins prepared in vivo display enhanced thermal and chemical stability. Angew Chem Int Ed 40: 1494–1496.
13. Wang P, Tang Y, Tirrell DA (2003) Incorporation of trifluoroisoleucine into proteins in vivo. J Am Chem Soc 125:6900–6906.
14. Wang P, Fichera A, Kumar K, Tirrell DA (2004) Alternative translations of a single RNA message: an identity switch of (2S,3R)-4,4,4-trifluorovaline between valine and isoleucine codons. Angew Chem Int Ed 43: 3664–3666.
15. Bilgiçer B, Xing X, Kumar K (2001) Programmed self-sorting of coiled coils with leucine and hexafluoroleucine cores. J Am Chem Soc 123:11815–11816.
16. Bilgiçer B, Kumar K (2004) De novo design of defined helical bundles in membrane environments. Proc Natl Acad Sci USA 101:15324–15329.
17. Chiu H-P, Suzuki Y, Gullickson D, Ahmad R, Kokona B, Fairman R, Cheng RP (2006) Helix propensity of highly fluorinated amino acids. J Am Chem Soc 128: 15556–15557.
18. Jäckel C, Salwiczek M, Koksch B (2006) Fluorine in a native protein environment—how the spatial demand and polarity of fluoroalkyl groups affect protein folding. Angew Chem Intl Ed 45:4198–4203.
19. Lee H-Y, Lee K-H, Al-Hashimi HM, Marsh ENG (2006) Modulating protein structure with fluorous amino acids: increased stability and native-like structure conferred on a 4-helix bundle protein by hexafluoroleucine. J Am Chem Soc 128:337–343.
20. Gottler LM, de la Salud-Bea R, Marsh ENG (2008) The fluorous effect in proteins: properties of α4F6, a 4-α-helix bundle protein with a fluorocarbon core. Biochemistry 47:4484–4490.
21. Buer BC, de la Salud-Bea R, Al Hashimi HM, Marsh ENG (2009) Engineering protein stability and specificity using fluorous amino acids: the importance of packing effects. Biochemistry 48:10810–10817.
22. Chiu H-P, Kokona B, Fairman R, Cheng RP (2009) Effect of highly fluorinated amino acids on protein stability at a solvent-exposed position on an internal strand of protein G B1 domain. J Am Chem Soc 131: 13192–13193.
23. Salwiczek M, Samsonov S, Vagt T, Nyakatura E, Fleige E, Numata J, Cölfen H, Pisabarro MT, Koksch B (2009) Position dependent effects of fluorinated amino acids on the hydrophobic core formation of a heterodimeric coiled coil. Chem Eur J 15:7628–7636.
24. Meng H, Kumar K (2007) Antimicrobial activity and protease stability of peptides containing fluorinated amino acids. J Am Chem Soc 129:15615–15622.
25. Gottler LM, Lee H-Y, Shelburne CE, Ramamoorthy A, Marsh ENG (2008) Using fluorous amino acids to modulate the biological activity of an antimicrobial peptide. ChemBioChem 9:370–373.

26. Meng H, Krishnaji ST, Beinborn M, Kumar K (2008) Influence of selective fluorination on the biological activity and proteolytic stability of glucagon-like peptide-1. J Med Chem 51:7303–7307.

27. Cornilescu G, Hadley EB, Woll MG, Markley JL, Gellman SH, Cornilescu CC (2007) Solution structure of a small protein containing a fluorinated side chain in the core. Protein Sci 16:14–19.

28. Mortenson DE, Satyshur KA, Guzei IA, Forest KT, Gellman SH (2012) Quasiracemic crystallization as a tool to assess the accommodation of noncanonical residues in nativelike protein conformations. J Am Chem Soc 134:2473–2476.

29. Buer BC, Meagher JL, Stuckey JA, Marsh ENG (2012) Structural basis for the enhanced stability of highly fluorinated proteins. Proc Natl Acad Sci USA 109: 4810–4815.

30. Neil E, Marsh G (2000) Towards the nonstick egg: designing fluorous proteins. Chem Biol 7:R153–R157.

31. Yoder NC, Yüksel D, Dafik L, Kumar K (2006) Bioorthogonal noncovalent chemistry: fluorous phases in chemical biology. Curr Opin Chem Biol 10:576–583.

32. Yadav MK, Redman JE, Leman LJ, Alvarez-Gutiérrez JM, Zhang Y, Stout CD, Ghadiri MR (2005) Structure-based engineering of internal cavities in coiled-coil peptides. Biochemistry 44:9723–9732.

33. Eriksson A, Baase WA, Zhang X, Heinz D, Blaber M, Baldwin EP, Matthews B (1992) Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. Science 255:178–183.

34. Bilgiçer B, Kumar K (2002) Synthesis and thermodynamic characterization of self-sorting coiled coils. Tetrahedron 58:4105–4112.

35. Lee K-H, Lee H-Y, Slutsky MM, Anderson JT, Marsh ENG (2004) Fluorous effect in proteins: de novo design and characterization of a four-α-helix bundle protein containing hexafluoroleucine. Biochemistry 43: 16277–16284.

36. Horváth IT, Rábai J (1994) Facile catalyst separation without water: fluorous biphase hydroformylation of olefins. Science 266:72–75.

37. Luo Z, Zhang Q, Oderaotoshi Y, Curran DP (2001) Fluorous mixture synthesis: a fluorous-tagging strategy for the synthesis and separation of mixtures of organic compounds. Science 291:1766–1769.

38. Hadley EB, Gellman SH (2006) An antiparallel α-helical coiled-coil model system for rapid assessment of side-chain recognition at the hydrophobic interface. J Am Chem Soc 128:16444–16445.

39. Hadley EB, Testa OD, Woolfson DN, Gellman SH (2008) Preferred side-chain constellations at antiparallel coiled-coil interfaces. Proc Natl Acad Sci USA 105: 530.

40. Steinkruger JD, Bartlett GJ, Hadley EB, Fay L, Woolfson DN, Gellman SH (2012) The d'--d--d' vertical triad is less discriminating than the a'--a--a' vertical triad in the antiparallel coiled-coil dimer motif. J Am Chem Soc 134:2626–2633.

41. Anderson JT, Toogood PL, Marsh ENG (2002) A short and efficient synthesis of L-5,5,5,5',5',5'-hexafluoroleucine from N-Cbz-L-serine. Org Lett 4:4281–4283.

42. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 276:307–326.

43. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. J Appl Cryst 40:658–674.

44. Schuttelkopf AW, van Aalten DMF (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Cryst D60:1355–1363.

45. Bricogne G, Blanc E, Brandl M, Flensburg C, Keller P, Paciorek W, Roversi P, Sharff A, Smart OS, Vonrhein C, Womack TO (2011) BUSTER version 1.6.0. Cambridge, United Kingdom: Global Phasing Ltd.

46. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Cryst D60:2126–2132.

47. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, Snoeyink J, Richardson JS, Richardson DC (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucl Acids Res 35:375–383.

48. Zucker F, Champ PC, Merritt EA (2010) Validation of crystallographic models containing TLS or other descriptions of anisotropy. Acta Cryst D66:889–900.

49. Hooft RWW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. Nature 381:272–272.