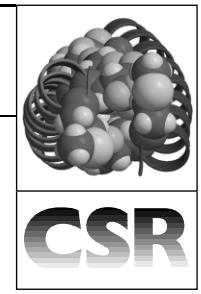
Fluorinated amino acids in protein design and engineering



Nicholas C. Yoder and Krishna Kumar*

Department of Chemistry, Tufts University, Medford, Massachusetts 02155, USA

Received 23rd May 2002
First published as an Advance Article on the web 13th September 2002

Selective incorporation of unnatural amino acids into proteins is a powerful tool for illuminating the principles of protein design. In particular, fluorinated amino acids have recently emerged as valuable building blocks for designing hyperstable protein folds, as well as directing highly specific protein–protein interactions. We review the collagen mimetic and coiled coil peptide systems that exemplify generalizable paradigms for future design. The unique electronic and phase properties of fluorocarbons are discussed, and protein synthesis using unnatural amino acids is briefly reviewed.

1 Introduction

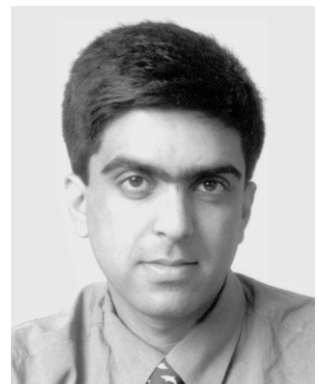
The accelerated sequencing of whole organism genomes and their analysis poses a tremendous challenge to biochemists—the attendant elucidation of the structure and function of the corresponding proteins that are encoded. These efforts will no doubt reveal the functions of individual protein molecules, many of which are likely to be novel and unprecedented.

For twenty years protein scientists have been studying the relationship between protein sequence, structure and function by constructing mutants of known proteins. By comparison, organic chemists have spent two centuries developing reactions, of which many are biologically important but many more are unknown in living organisms. Many of the important technological innovations have come from reactions that are not currently known to exist in nature's arsenal. Inspired by such success, protein scientists have begun to ask 'Rather than limit our vision to the set of extant proteins, can we rationally design and construct novel structures and endow them with predetermined and unusual functions?'

Natural proteins have evolved to accomplish a staggering variety of exquisite tasks, from tissue formation to enzymatic catalysis, in almost all cases using the same twenty amino acid building blocks. Enzymes are remarkable in the precision with which their tertiary structures dictate the arrangement of chemical functionalities to achieve high specificity and catalytic turnover. Site-directed and random mutagenesis, in which specific amino acids in a protein are replaced by one of the other nineteen natural amino acids, have been indispensable tools to understand and manipulate the structure and function of enzymes. Protein engineers do not as yet have such subtle control over structure to direct display of functionality. They do, however, have access to a vast assortment of chemical functionalities which are not found in the side chains of the natural amino acids. The last decade and a half has witnessed a revolution in the ability of chemists and biochemists to site-specifically incorporate amino acids different from the canonical twenty. This research has shown great promise not only for expanding mutagenesis studies of natural proteins, but for making proteins with novel and catalytic properties, as well as unusual structural and ligand-binding characteristics.

Fluorine is one of the most abundant elements on earth, yet it occurs extremely rarely in biological compounds. Most of earth's fluorine is in the form of insoluble fluoride minerals, and consequently the concentration of free fluoride in sea and surface water is very low. Interestingly, the replacement of hydrogen with fluorine in organic compounds is often accompanied by profound and unexpected changes in biological activity. For example, a single fluorine atom on the methyl group of acetic acid turns it into a poisonous substance. Fluoroacetic acid is naturally produced by the gifblaar plant found in the South African veldt, and is deadly for cattle that feed on it. On the other hand, extensive fluorination has become a paradigm for fashioning materials with unusual and useful properties. For example, the DuPont polymer Teflon (polytetrafluoroethylene) is used in waterproofing clothes and shoes, and in nonstick cookware. Fluorinated synthetic oils are now routinely used in lubrication where oxidation is a severe problem. Nafion, a sulfonated perfluorocarbon, is used in the phosphoric acid fuel cell. These materials are useful for the purposes described above because of the high thermal stability and inertness of C–F bonds and low propensity of highly fluorinated compounds to interact with other materials. Although these properties have been known to, and have been

Nicholas Yoder was born in Boston in 1975. He graduated from Brown University in 1997 with his BSc in chemistry after working with Professor David E. Cane on sesquiterpene biosynthesis. He is currently a doctoral student in chemistry at Tufts University and is working with Professor Krishna Kumar on protein design and artificial enzymes.



Krishna Kumar received his doctorate in Chemistry from Brown University in 1996. After two years as a Skaggs Research Fellow at The Scripps Research Institute, he joined Tufts University in 1998 where he is Assistant Professor in the Department of Chemistry and an Associate Member of The Cancer Center at the Tufts-New England Medical Center (NEMC).

exploited by chemists, they have only recently received attention in the design and control of biological macromolecular structure.

Recent studies with fluorinated amino acids suggest new opportunities for the construction of hyperstable protein folds and directing highly specific protein–protein interfaces, and are some of the most successful and potentially general of the studies with unnatural amino acids. These considerations make fluorinated amino acids a valuable addition to the protein engineer's toolbox. Historically, fluorinated amino acids have also proved useful as reporter groups for dynamic NMR studies on proteins, and recent advances in spectroscopic techniques suggest this strategy will remain productive.

Both chemical and biosynthesis of proteins incorporating unnatural amino acids is generally challenging, and has been the limiting factor in the design of novel proteins using this approach. This review will catalogue the preliminary explorations and present new challenges and opportunities in the use of unnatural amino acids containing fluorinated side chains in protein design.

2 Incorporating fluorinated amino acids in proteins

A review of strategies for the synthesis of unnatural amino acids is beyond the scope of this paper, and has been reviewed extensively elsewhere.[1] In most cases, enantiomerically pure fluorinated amino acids can be synthesized by well-established routes.[2] Site-specific incorporation of unnatural amino acids into proteins can be achieved by both synthetic and biosynthetic methods. A few of the techniques currently available are discussed below.

The earliest incorporation of fluorinated amino acids into proteins was accomplished by employing specially bred strains of bacteria called auxotrophs.[3] Auxotrophs lack the ability to biosynthesize one of the naturally occurring amino acids. A structural analog of the required amino acid is introduced into the medium, and if it is recognized by the bacterial aminoacyl-tRNA synthetase, it will be incorporated into all proteins produced by the bacterium. This method has been successfully employed to incorporate *o*-, *m*-, and *p*-fluorophenylalanine, trifluoromethionine, 4-fluoroproline and trifluoroleucine into proteins. While this method is straightforward and produces relatively high yields of protein using analogs of natural amino acids, it has several important drawbacks. The range of amino acids is limited to structurally close analogs of natural ones. Furthermore, there is no control over which locations on the peptide chain are substituted, and if there is more than one occurrence of the desired amino acid in the target protein, in theory they will all be substituted. In practice, however, this does not happen, because the bacteria must be first propagated in the presence of all amino acids, so the naturally occurring amino acid is always present in the culture. This means that substitution of the amino acid is never 100% effective at each position (95% is considered very good) and heterogeneous protein samples are invariably recovered.[4] Nevertheless, auxotrophic synthesis has proven to be effective for many protein engineering applications. Recent advances in auxotroph based synthesis have allowed the incorporation of amino acids with side chain azide[5] and aromatic ketone groups, both of which may be chemically modified to other functionalities, potentially including fluorinated groups, under physiological conditions.[6]

Total control over protein sequence and amino acid content, as well as very high purity, is afforded by solid phase peptide synthesis, but this technique is effectively limited to peptides of up to 50–100 amino acids. Peptide segments can be stitched together by the thioester ligation method of Dawson and Kent which has allowed for the synthesis of 180 residue proteins, though ligation must occur at a cysteine residue.[7] The desired protein is typically synthesized by solid phase protocols in multiple pieces. The N-terminal segment of the target protein is synthesized with the C-terminus functionalized as a thioester. The C-terminal segment is equipped with an N-terminal cysteine. The two fragments are ligated in neutral aqueous solution. First, the C-terminal thioester undergoes a trans-thioesterification with the N-terminal cysteine sulfhydryl group of the other fragment. The resulting thioester rapidly rearranges *via* an S to N acyl transfer reaction to yield a native amide bond. New variations on this strategy have relaxed the requirement for a cysteine at the ligation site either by temporarily linking a thiol to the α-amino group of a glycine residue or by post-ligation reduction of cysteine to alanine using Raney nickel.[8,9]

This strategy can be made significantly more versatile by using recombinant proteins as the C-terminal thioesters.[10] Inteins are intervening peptide segments that are excised from a protein precursor with the concomitant ligation of the flanking protein fragments (exteins) to form a mature extein host protein and the free intein. The mechanism of this splicing reaction involves thioester formation, and this has been exploited as a convenient means to produce protein thioester fragments for use in the so-called expressed protein ligation. Protein domains prepared in this manner could in principle be combined with synthetic fragments incorporating unnatural amino acids.[11] The recombinant protein thioester is limited to the naturally occurring amino acids in this approach.

Another approach to unnatural amino acid incorporation into proteins, known as nonsense suppression, subverts a protein expression system into inserting an unnatural amino acid in response to one of the stop (or nonsense) codons in the genetic code. A 'suppressor' tRNA is modified to provide an anticodon loop complementary to the mRNA stop codon and is chemically charged with the desired unnatural amino acid. A gene corresponding to the desired protein sequence, with the nonsense codon at the desired location of the unnatural amino acid, can easily be synthesized and introduced to a cell-free expression system for translation. Nonsense suppression effectively expands the repertoire of available amino acids, and allows proteins of any size and sequence to be expressed, though not without limitations. The overall yield of protein is entirely dependent on the amount of suppressor aminoacyl-tRNA, which is consumed as a stoichiometric reagent. The synthesis of this component has proved tedious and costly.[12] Some amino acids are simply not added to the nascent polypeptide chain by the ribosomal machinery. In sum, total yields of more than one milligram of protein are almost never observed.

A significant advance to the nonsense suppression technique has been recently reported, in which an enzyme is evolved *in vivo* to charge the suppressor tRNA with the unnatural amino acid, eliminating the need to synthesize this component (Fig. 1).[13,14] Selection of such an enzyme is challenging, since efficient translation is contingent upon the specificity with which the enzyme recognizes the unnatural tRNA and unnatural amino acid in the presence of all the other naturally occurring ones. In addition, enzyme selection must be repeated for every new unnatural amino acid or tRNA used. This technique is also still limited to one nonsense codon, though efforts are under way to use four- and five- base codon–anticodon pairs in order to synthesize proteins incorporating multiple unnatural amino acids *in vivo*.[15,16]

3 Biophysical probes based on ^{19}F NMR

Incorporation of fluorinated amino acids into proteins allows monitoring of the chemical environment of the fluorine-containing residues by ^{19}F NMR. The chemical shift of the ^{19}F

nucleus is extremely sensitive to the surrounding environment; in protein systems, δ may increase or decrease by as much as 8 ppm when the protein is denatured. The ^{19}F nucleus can be observed with high precision quite easily, as it is 83% as sensitive as the proton, occurs in 100% natural abundance, and no background fluorine signal is observed from natural proteins.[17] Typically the change in ^{19}F chemical shift is monitored as some change in protein conformation occurs, such as that induced by substrate binding. Line broadening induced by paramagnetic ion reagents like Gd·EDTA may be used to determine the degree to which a fluorinated side chain is exposed to solvent.

Many important structural questions in protein folding cannot be addressed using X-ray crystallography. Under specific conditions, some proteins occupy stable, partially folded states called molten globules. Characterization of these states may help advance our understanding of protein folding pathways. Recently the molten globule state of α-lactalbumin has been studied with ^{19}F NMR using solvent-exposure measurements on various fluorine-labeled derivatives.[18]

Structure elucidation of proteins embedded in membranes is also challenging, as most of these proteins are insoluble in water and therefore difficult to crystallize. Prosser and Sanders recently described experiments on a membrane-spanning peptide domain which determined the depth of membrane immersion of each amino acid.[19] In this experiment, instead of using a water soluble paramagnetic shift reagent to determine fluorinated side chain exposure to solvent, O_2 was used as a shift reagent to determine fluorinated side chain exposure to the membrane phase. Because O_2 dissolves in membrane phases with a well-defined concentration gradient, side chains near the center of the membrane are exposed to a higher mean concentration of O_2 and a correspondingly higher chemical shift perturbation is observed. While this technique does not deliver the global structure of the protein within the membrane, it adds to the arsenal of biophysical techniques which can be employed to obtain increasingly detailed information.

Solid state NMR experiments using rotational resonance (REDOR) offer some of the best prospects for structure determination in non-crystalline protein aggregates,[20] as demonstrated by experiments with the β-amyloid peptide.[21,22] The REDOR experiment uses rotor-synchronized I-spin π pulses to dephase the transverse magnetization of the S-spin, where I and S are two dipolar-coupled rare spins. A third radio frequency channel is used to remove coupling to protons. The data obtained from REDOR experiments directly gives the strength of the heteronuclear I–S dipolar coupling and therefore I–S internuclear distances. Structure determination based on distance information obtained from REDOR using ^{19}F as one of the nuclei is becoming increasingly popular for several biological systems.[23–25]

Use of fluorine as an NMR probe is not without potential complication. In general, because the fluorine atom is quite compact and does not introduce much steric bulk beyond hydrogen, introduction of two or three fluorine atoms into a protein is not expected to significantly perturb its tertiary structure. However, there have been substantial changes in protein structure observed in at least one ^{19}F NMR study, so introduction of fluorine must be carried out judiciously.[26] As we shall discuss below, the effects of fluorinated amino acids on protein structure are a subject of great interest in their own right.

4 Protein design and engineering

Recently, protein chemists have exploited fluorinated amino acids to introduce novel properties in *de novo* designed peptides. In an innovative series of experiments, Raines and coworkers exploited the highly polarized nature of the C–F bond to probe the origins of stability of collagen, the ubiquitous mammalian structural protein, and to make hyperstable analogues. It has long been known that the triple helix of collagen

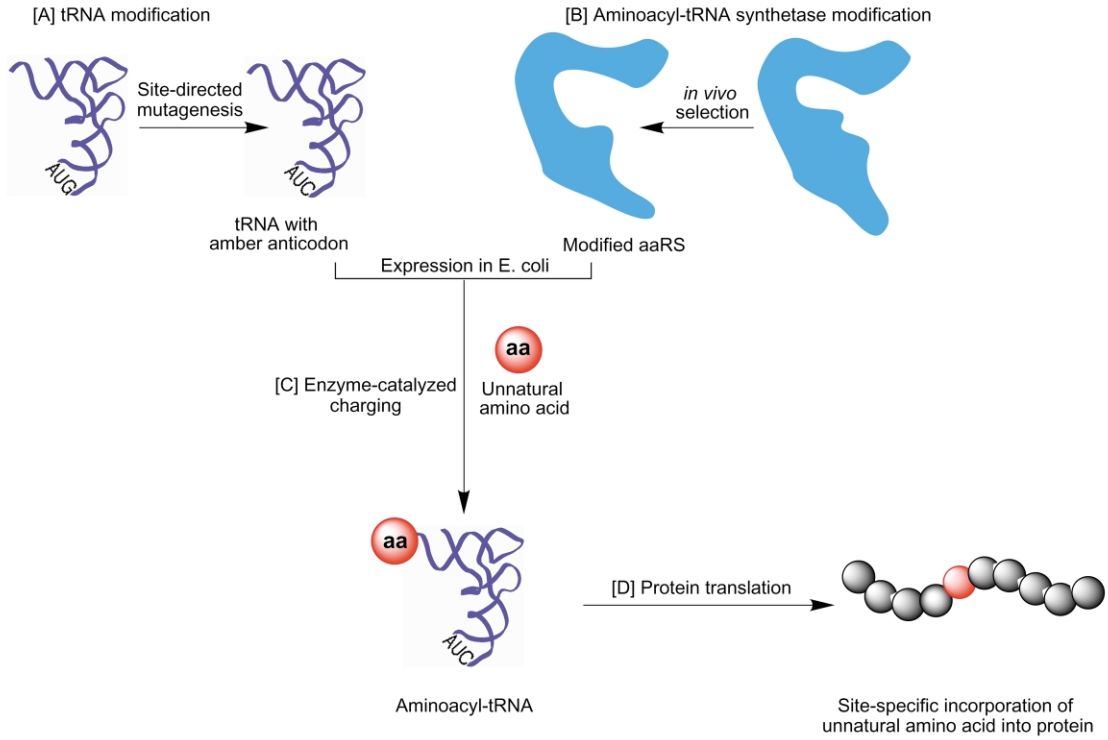


Fig. 1 Unnatural amino acid incorporation using orthogonal tRNA and synthetase pair specific for the unnatural amino acid. A tRNA not recognized by the *E. coli* ribosomal machinery is utilized, along with the corresponding aminoacyl-tRNA synthetase (aaRS). [A] The tRNA is modified so that the anticodon loop corresponds to the amber codon. [B] The synthetase is subjected to extensive *in vivo* selection so that it specifically charges the novel tRNA with the given amino acid. [C] The aaRS charges the suppressor tRNA with the unnatural amino acid present in the growth medium. [D] The unnatural amino acid is incorporated into the target protein in response to the amber codon.

is strengthened by the presence of repeating trimers containing glycine, proline, and the non-canonical amino acid 4(*R*)-hydroxyproline [4(*R*)-Hyp]. These trimers make up a significant fraction of the total collagen content, and it was long thought that hydroxyproline strengthened the collagen helix by allowing increased interstrand hydrogen bonding. Using solid phase synthesis, collagen-like model peptides of the form [ProXaaGly]$_{10}$, with Xaa representing either 4(*R*)-hydroxy- or 4(*R*)-fluoroproline residues were prepared. Both peptides were helical, but the fluoroproline peptide was much more stable, as it showed a cooperative melting temperature of 91 °C as compared to 69 °C for the hydroxyproline peptide.[27] Furthermore, while peptides of the form [ProXaaGly]$_7$ showed melting temperatures of 45 °C and 36 °C with Xaa = 4(*R*)-fluoroproline and 4(*R*)-hydroxyproline respectively, with Xaa = 4(*S*)-fluoroproline, no helical structure was detected above 2 °C.[28] Collagen stability is evidently quite sensitive to a stereoelectronic effect. NMR and computational studies of proline model compounds suggest that electronegative *gauche* effects stabilize both the *exo*- conformation of the proline ring and the *trans*- conformation of the prolyl amide bond (Fig. 2).[29, 30] The predisposition of the substituted proline toward this

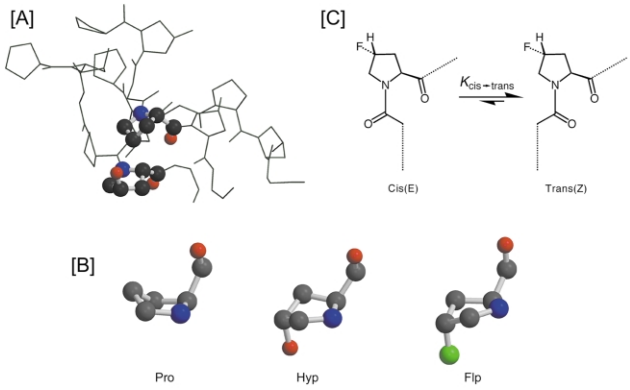


Fig. 2 Stereoelectronic effect of fluorine substitution on proline rings and on collagen mimic peptides. [A] Section of the structure of a collagen mimic peptide (PDB code: 1CAG), with one proline and one hydroxyproline residue shown in ball-in-stick form, showing the preferred conformation of each residue. [B] Alternate views of proline (Pro) and hydroxyproline (Hyp) residues from the structure in [A], as well as of a 4(*R*)-fluoroproline (Flp) derivative (CCDC code: RISDEC) showing the similar *exo* conformation adopted by the 4(*R*)-substituted proline rings. [C] *cis–trans* Isomerism about the prolyl amide bond in a 4(*R*)-fluoroproline model compound. The equilibrium lies farther to the right for the substituted model compound than for the proline model.

conformation leads to stabilization of the collagen helix, a suggestion supported by crystallographic studies of collagen mimics and of polyproline type II helices.[31] Studies with 4(*R*)-aminoproline collagen analogs by Babu and Ganesh support the general model of inductive stabilization of the collagen triple helix.[32] The unique properties of fluorinated amino acids have proved invaluable in elucidating this unprecedented stereoelectronic effect in dictating protein structure.

The ability to construct collagen-like triple helices of varying thermal stabilities has already been exploited to study the relationship between triple helix formation and the signaling effect of certain regions of collagen. It is known that a 15 residue sequence commonly found in collagen known as IV-H1 can increase adhesion and migration of melanoma cells when inserted into triple helical collagen mimetic sequences. A peptide of the form [Gly-Pro-4(*R*)-Hyp]$_4$-[IV-H1]-[Gly-Pro-4(*R*)-Hyp]$_4$ was synthesized as well as variants with one of the hydroxyprolines substituted with either 4(*R*)- or 4(*S*)-fluoroproline. This single substitution had measurable effects on helix stability, with the 4(*R*)-substituent contributing thermal stability and the 4(*S*)- reducing the stability relative to the parent peptide. Using these variants, it was demonstrated that cellular response positively correlated with helical content of peptides containing IV-H1.[33]

In addition to helping characterize the conformational effects of electronegative substituents on proline rings, Moroder and co-workers also demonstrated their broad applicability to protein engineering. Crystal structures of the *E. coli* protein barstar shows the Pro48 prolyl amide bond in the *cis*- and Pro27 in the *trans*-conformation. A barstar mutant gene was prepared with Pro27 mutated to alanine, which does not affect protein function, and was expressed in proline auxotrophs to incorporate 4(*R*)- and 4(*S*)-fluoroproline at Pro48. As expected, 4(*R*)-fluoroproline destabilizes native barstar structure, and 4(*S*)-fluoroproline has the opposite effect, presumably by increasing the preference for the *cis*-prolyl bond conformation.[30]

Other properties of fluorocarbons suggest more widespread uses in protein engineering.[34] It is well known that perfluorinated organic compounds are more hydrophobic than hydrocarbons, as well as insoluble in hydrocarbons at room temperature (forming a third 'fluorous' phase). Many popular pharmaceuticals also incorporate fluorinated functions in order to increase hydrophobicity and suppress metabolic detoxification to increase bioavailability. The fluorous phase has also been exploited in synthetic organic chemistry to recycle catalysts, accelerate organic reactions[35] and provide innovative and practical purification strategies. For instance, metal catalysts bound to fluorinated ligands, are miscible with organic phases at high temperatures, but can easily be recovered by cooling the reaction mixture so the phases separate.[36–38] This effect also interests protein engineers, since the removal of non polar surface area from water provides the dominant driving force in protein folding. While this phenomenon is more complex than simple liquid phase separation, the model is still a useful first approximation.

Among the first reported studies of peptides incorporating highly fluorinated amino acid residues in a site-specific manner was an investigation of the blood pressure regulating peptide hormone angiotensin with ^{19}F NMR. Hexafluorovaline was incorporated into the peptide at selected positions, and *in vivo* control experiments were carried out to verify that the modified peptide retained biological activity. Interestingly, the modified peptide was more active than the wild type, a result which was attributed in part to increased resistance to degradative enzymes but also to increased hydrophobicity and accompanying higher bioavailability from the highly fluorinated amino acid present.[39]

In the last three years research directed toward incorporating the phase properties of fluorocarbons into engineered peptides has begun in earnest in our laboratory and in others. The goal of these studies is to incorporate new structural motifs into proteins using the principles of fluorous phase separation. As such they are distinct from the use of fluorine atoms as spectroscopic reporter groups within a protein framework.

A common starting point for *de novo* peptide design based on hydrophobic binding is the coiled coil. Coiled coils consist of two or more amphipathic α-helices wrapped around each other to form a superhelical structure stabilized by hydrophobic residues spaced in a repeating 3–4 sequence pattern. This characteristic heptad repeat of seven amino acids, denoted (*abcdefg*)$_n$, typically contains hydrophobic residues at the first and fourth (*a* and *d*) positions, and charged residues at the fifth and seventh (*e* and *g*) positions. Supercoiled helix formation positions all the hydrophobic residues in the sequence along one face of the helix surface, with 3.5 residues per turn. The hydrophobic side chains pack against each other in a tight pattern known as 'knobs-into-holes', with a residue from one helix surrounded by four from another. The charged residues provide a secondary recognition motif.[40]

One of the best-characterized coiled coil peptides is GCN4-p1, the dimerization domain of the yeast transcriptional activator protein bZip. This peptide is 33 residues long, with

leucine in every *d* position, and valine in every *a* position. A single *a* position contains asparagine instead of valine. The asparagine decreases the energetic payoff of oligomerization, but leads to exclusively parallel, dimeric coiled coils.[41]

We hypothesized that fluorination of the core residues in GCN4-p1 might increase the driving force for oligomerization. Using solid phase synthesis, trifluoroleucine was substituted for leucine at all *d* positions, and trifluorovaline for valine at all *a* positions, with the exception of a single asparagine maintained for parallel dimer specificity (Fig. 3).[42]The trifluoroleucine and

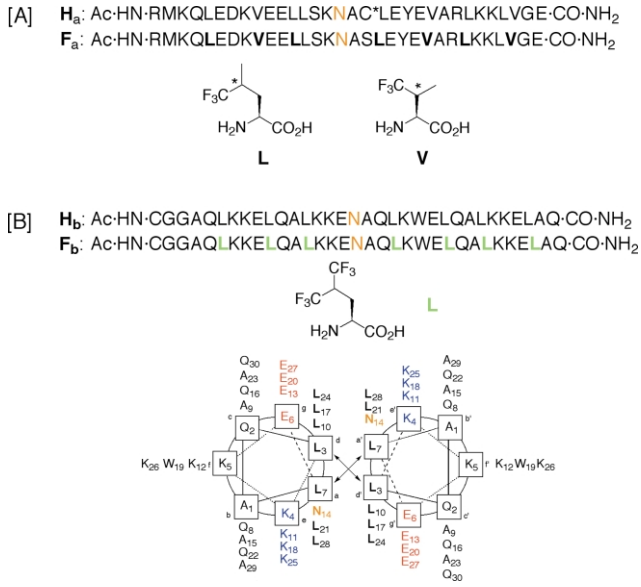


Fig. 3 Design of fluorinated coiled coil peptides. [A] Peptide sequences showing the heptad repeat; $\mathbf{F_a}$ contains trifluorovaline and trifluoroleucine in the core. Asterisks in the amino acid structures indicate unresolved stereochemistry; C* = acetamidocysteine. [B] Peptide sequence and helical wheel diagram of second-generation coiled coil peptide and fluorinated analog. All leucines in $\mathbf{H_b}$ are replaced by hexafluoroleucine, shown, in $\mathbf{F_b}$.

trifluorovaline residues were incorporated as diastereomeric mixtures, stereochemically resolved at the α-carbon but not at the β (trifluorovaline) or γ (trifluoroleucine) carbons. Despite this inhomogeneity, peptides obtained in this manner showed remarkable cooperativity and were amenable to detailed biophysical characterization. The fluorinated peptide $\mathbf{F_a}$ (Fig. 4) showed striking differences in physical properties from the hydrocarbon peptide $\mathbf{H_a}$. Circular dichroism and equilibrium sedimentation experiments established that both were α-helical dimers in solution, but the fluorocarbon core peptide had a melting temperature of 62 °C compared to 47 °C for the control peptide. Denaturation with guanidinium hydrochloride (Gdn·HCl) indicated that the free energy of unfolding was slightly higher for the fluorocarbon peptide by approximately 1.0 kcal mol^{-1} at 5 °C. The higher free energy cost of denaturing the coiled coil dimers was attributable to the larger hydrophobic surface area in the core of the protein.

Similar experiments were performed independently by Tirrell and co-workers using GCN4-p1 analogs with trifluoroleucine at all *d* positions. Interestingly, these peptides showed similar increases in stability to the trifluoroleucine/trifluorovaline peptides discussed above, with the wild-type melting at 48 °C and the trifluoroleucine variant melting at 61 °C.[43] In addition, the incorporation of trifluoroleucine into the entire bZip transcriptional activation factor did not interfere with sequence-specific DNA binding pointing to similarity of the structures. Auxotrophic biosynthesis was also used to produce a larger leucine zipper protein with eight trifluoroleucines in the core with corresponding increases in stability

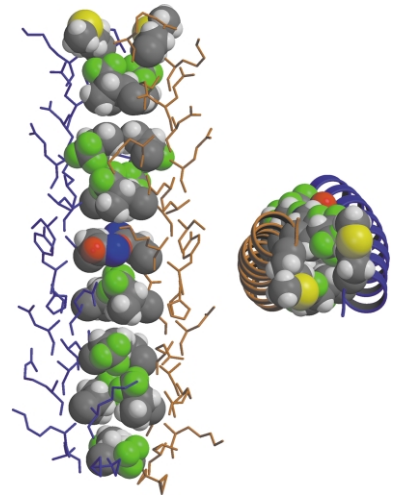


Fig. 4 Three-dimensional rendering of the $\mathbf{F_a}$ dimer with hydrophobic core residues shown in space filling depiction. Color scheme: C = grey, H = white, S = yellow, N = blue, O = red, F = green. Left panel: side view, right panel: top view.

measured by heat and chemical denaturation.[44] In a more recent study, the more highly fluorinated amino acid hexafluoroleucine was introduced into this protein with 74% efficiency by modification of the leucyl-tRNA synthetase activity of the host. In spite of the heterogeneity of the resulting sample, it showed an increase in melting temperature over the hydrocarbon peptide from 54 °C to 76 °C.[4] These studies represent important examples of auxotrophic incorporation of unnatural amino acids with significant and observable differences in stability.

We have recently extended this design further to ascertain the specificity of interaction in the coiled coil assemblies. The peptide system now consisted of 30-residue peptides with lysine residues at all the *e* positions and glutamic acid residues at *g* positions. These charged side chains ensure that the assembly is parallel due to unfavorable interhelical electrostatic interactions in the antiparallel mode. Furthermore, a single asparagine was introduced in the core at a position where it could only hydrogen bond in the parallel arrangement. The sequences were identical except that all seven of the core leucine residues in $\mathbf{H_b}$ were replaced by hexafluoroleucine in $\mathbf{F_b}$. This not only increased the overall fluorocarbon density but resulted in a single diastereomeric peptide product.[2,45] The peptides used were also equipped with N-terminal cysteine residues on a flexible Gly–Gly linker (Fig. 3). A disulfide bound dimer of the hexafluoroleucine core peptide and a leucine core peptide ($\mathbf{H_bF_b}$) was prepared, and allowed to undergo disulfide exchange under appropriate redox conditions. Apparent equilibrium was established within 30 minutes, and at that time only 3% of the initial heterodimer remained in solution, with the remaining peptides present as exclusively leucine or hexafluoroleucine core disulfide bridged dimers, denoted $\mathbf{H_bH_b}$ and $\mathbf{F_bF_b}$, with a small amount of adducts with redox buffer (Fig. 5). Homodimer was preferred over heterodimer by at least 26-fold, and equilibrium thermodynamic analysis suggests that this preference was largely due to the remarkable stability of the hexafluoroleucine core tetramer. This assembly, which incorporates two covalently bound peptide helices into a putative four-helix bundle, resists even minimal denaturation at 6M Gdn·HCl at room temperature, and melts at 45 °C in 7M Gdn·HCl. It is also remarkable that the melting temperature of the covalently bound heterodimer was comparable to the leucine core peptide dimer, suggesting that the hydrocarbon–fluorocarbon interaction contributes little stability.[46] This self-sorting behavior, verified by control experiments ruling out kinetic barriers to heterodimer formation, is the best demonstration so far of the potential for

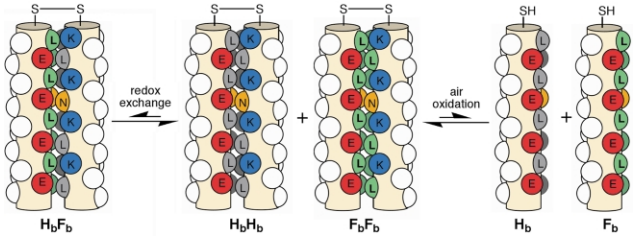


Fig. 5 Thermodynamic self-sorting behavior of hydrocarbon- and fluorocarbon-core coiled coil peptides. Covalently bound heterodimer $\mathbf{H_bF_b}$ undergoes disulfide exchange *via* monomeric peptides to yield covalently bound homodimers $\mathbf{H_bH_b}$ and $\mathbf{F_bF_b}$ in >90% yield. Thiol monomers $\mathbf{H_b}$ and $\mathbf{F_b}$ undergo oxidation to produce the same mixture.

specific protein–protein interactions based on this novel simultaneously hydrophobic and lipophobic behavior.

These experiments suggest that the fluorocarbon orthogonal phase paradigm may be may be applicable to the design of proteins of widely divergent functionalities. The thermal stability of almost any globular protein in solution could, in theory, be increased by replacing its hydrocarbon core with a fluorous one. Specific protein–protein binding is important in signal transduction, and fluorous phase associations could provide a highly specific means of studying cellular signal pathways and other protein–protein interactions. Finally, our laboratory is currently investigating the use of fluorous cores to direct self-association in lipid membrane environments. Unpublished results suggest that fluorinated peptides do in fact form well-defined oligomers within membranes and may provide an unprecedented means to control helix packing within membranes.

5 Conclusions and prospects for the future

Over the last several years, a number of factors have conspired to make the difficult task of site-specific introduction of unnatural amino acids into proteins a matter of routine practice. Coupled with progress in the chemical synthesis of enantiomerically pure amino acids containing side chains with fluorine atoms, these advances have paved the way for the *de novo* design of proteins with unprecedented, and extra-biological properties. Furthermore, creation of such protein molecules has also shed light on the origins of stability for natural biopolymers.

Even though protein design using fluorinated amino acids is a relatively new area of investigation, already two important paradigms have emerged. First, as exemplified by the innovative work of Raines and Moroder, the ability of fluorine to exert a strong inductive effect by virtue of it being the most electronegative element in the periodic table, can play a capital role in protein design. Stereoelectronic effects dictating the structural stability of protein assemblies is a novel concept and may eventually find parallels in natural protein structures other than collagen.

Second, while both the higher hydrophobicity and immiscibility with aqueous and organic solvents of fluorocarbons have been exploited in materials chemistry and in catalysis, recent studies originating from our and other laboratories have established that fluorinated amino acids can be used to stabilize protein folds. Furthermore, appropriately fluorinated protein scaffolds can be used to program selective protein–protein interactions. These experiments suggest widespread and divergent uses for fluorinated proteins. Introduction of a large number of carbon–fluorine bonds in biological macromolecules represents the union of two fertile areas of research: materials science and protein design. Given the unusual and extra-biological properties of fluorine, fluorinated amino acids represent unique building blocks for the design and fabrication of biologically active proteins with novel structure and function.

6 Acknowledgments

This work was supported by the National Institutes of Health (GM65500-01). We thank Basar Bilgiçer for helpful and stimulating discussions.

7 References

1 V. P. Kukhar' and V. A. Soloshonok, *Fluorine-containing Amino Acids: Synthesis and Properties*, John Wiley & Sons Ltd., Chichester, New York, 1995.
2 X. Xing, A. Fichera and K. Kumar, *Org. Lett.*, 2001, **3**, 1285.
3 O. M. Rennert and H. S. Anker, *Biochemistry*, 1963, **2**, 471.
4 Y. Tang and D. A. Tirrell, *J. Am. Chem. Soc.*, 2001, **123**, 11089.
5 K. L. Kiick, E. Saxon, D. A. Tirrell and C. R. Bertozzi, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 19.
6 D. Datta, P. Wang, I. S. Carrico, S. L. Mayo and D. A. Tirrell, *J. Am. Chem. Soc.*, 2002, **124**, 5652.
7 P. E. Dawson and S. B. H. Kent, *Ann. Rev. Biochem.*, 2000, **69**, 923.
8 C. Marinzi, S. J. Bark, J. Offer and P. E. Dawson, *Bioorg. Med. Chem.*, 2001, **9**, 2323.
9 L. Z. Yan and P. E. Dawson, *J. Am. Chem. Soc.*, 2001, **123**, 526.
10 G. J. Cotton and T. W. Muir, *Chem. Biol.*, 1999, **6**, R247.
11 R. S. Roy, O. Allen and C. T. Walsh, *Chem. Biol.*, 1999, **6**, 789.
12 M. A. Gilmore, L. E. Steward and A. R. Chamberlin, *Top. Curr. Chem.*, 1999, **202**, 77.
13 L. Wang, A. Brock, B. Herberich and P. G. Schultz, *Science*, 2001, **292**, 498.
14 L. Wang, A. Brock and P. G. Schultz, *J. Am. Chem. Soc.*, 2002, **124**, 1836.
15 T. J. Magliery, J. C. Anderson and P. G. Schultz, *J. Mol. Biol.*, 2001, **307**, 755.
16 T. Hohsaka, Y. Ashizuka, H. Murakami and M. Sisido, *Nucleic Acids Res.*, 2001, **29**, 3646.
17 J. T. Gerig, *Prog. Nucl. Magn. Reson. Spectrosc.*, 1994, **26**, 293.
18 P. Bai, L. Luo and Z. Y. Peng, *Biochemistry*, 2000, **39**, 372.
19 P. A. Luchette, R. S. Prosser and C. R. Sanders, *J. Am. Chem. Soc.*, 2002, **124**, 1778.
20 S. Dusold and A. Sebald, *Ann. Rep. NMR Spectosc.*, 2000, **41**, 185–264.
21 D. M. Gregory, T. L. S. Benzinger, T. S. Burkoth, H. Miller-Auer, D. G. Lynn, S. C. Meredith and R. E. Botto, *Solid State Nucl. Magn. Reson.*, 1998, **13**, 149.
22 T. S. Burkoth, T. L. S. Benzinger, V. Urban, D. M. Morgan, D. M. Gregory, P. Thiyagarajan, R. E. Botto, S. C. Meredith and D. G. Lynn, *J. Am. Chem. Soc.*, 2000, **122**, 7883.
23 J. Salgado, S. L. Grage, L. H. Kondejewski, R. S. Hodges, R. N. McElhaney and A. S. Ulrich, *J. Biomol. NMR*, 2001, **21**, 191.
24 J. M. Goetz, B. Poliks, D. R. Studelska, M. Fischer, K. Kugelbrey, A. Bacher, M. Cushman and J. Schaefer, *J. Am. Chem. Soc.*, 1999, **121**, 7500.
25 E. A. Louie, P. Chirakul, S. T. Sigurdsson and G. P. Drobny, *Biophys. J.*, 2002, **82**, 621.
26 G. Y. Xiao, J. F. Parsons, K. Tesh, R. N. Armstrong and G. L. Gilliland, *J. Mol. Biol.*, 1998, **281**, 323.
27 S. K. Holmgren, K. M. Taylor, L. E. Bretscher and R. T. Raines, *Nature*, 1998, **392**, 666.
28 L. E. Bretscher, C. L. Jenkins, K. M. Taylor, M. L. DeRider and R. T. Raines, *J. Am. Chem. Soc.*, 2001, **123**, 777.
29 M. L. DeRider, S. J. Wilkens, M. J. Waddell, L. E. Bretscher, F. Weinhold, R. T. Raines and J. L. Markley, *J. Am. Chem. Soc.*, 2002, **124**, 2497.
30 C. Renner, S. Alefelder, J. H. Bae, N. Budisa, R. Huber and L. Moroder, *Angew. Chem., Int. Ed.*, 2001, **40**, 923.
31 C. L. Jenkins and R. T. Raines, *Nat. Prod. Rep.*, 2002, **19**, 49.
32 I. R. Babu and K. N. Ganesh, *J. Am. Chem. Soc.*, 2001, **123**, 2079.
33 N. B. Malkar, J. L. Lauer-Fields, J. A. Borgia and G. B. Fields, *Biochemistry*, 2002, **41**, 6054.
34 E. N. G. Marsh, *Chem. Biol.*, 2000, **7**, R153.
35 K. E. Myers and K. Kumar, *J. Am. Chem. Soc.*, 2000, **122**, 12025.
36 E. de Wolf, G. van Koten and B. J. Deelman, *Chem. Soc. Rev.*, 1999, **28**, 37.
37 I. T. Horvath and J. Rabai, *Science*, 1994, **266**, 72.
38 A. Studer, S. Hadida, R. Ferritto, S. Y. Kim, P. Jeger, P. Wipf and D. P. Curran, *Science*, 1997, **275**, 823.

39 K. H. Hsieh, P. Needleman and G. R. Marshall, *J. Med. Chem.*, 1987, **30**, 1097.
40 A. Lupas, *Trends Biochem. Sci.*, 1996, **21**, 375.
41 P. B. Harbury, T. Zhang, P. S. Kim and T. Alber, *Science*, 1993, **262**, 1401.
42 B. Bilgiçer, A. Fichera and K. Kumar, *J. Am. Chem. Soc.*, 2001, **123**, 4393.
43 Y. Tang, G. Ghirlanda, N. Vaidehi, J. Kua, D. T. Mainz, W. A. Goddard, W. F. DeGrado and D. A. Tirrell, *Biochemistry*, 2001, **40**, 2790.
44 Y. Tang, G. Ghirlanda, W. A. Petka, T. Nakajima, W. F. DeGrado and D. A. Tirrell, *Angew. Chem., Int. Ed.*, 2001, **40**, 1494.
45 B. Bilgiçer, X. Xing and K. Kumar, *J. Am. Chem. Soc.*, 2001, **123**, 11815.
46 B. Bilgiçer and K. Kumar, *Tetrahedron*, 2002, **58**, 4105.