Towards the nonstick egg: designing fluorous proteins

E Neil G Marsh

Anyone who has made scrambled eggs will have had cause to praise the properties of Teflon. Teflon's highly chemically inert and nonstick nature derives from the perfluorinated polymer polytetrafluoroethylene. Perfluorocarbons have unique and valuable physical properties not found in nature. By incorporating fluorine into proteins, it might be possible to produce biological molecules with novel and useful properties.

Address: Department of Chemistry and Division of Biophysics, University of Michigan, Ann Arbor, MI 48109-1055, USA.

E-mail: nmarsh@umich.edu

Chemistry & Biology 2000, **7**:R153–R157

1074-5521/00/$ – see front matter

Fluorine occurs extremely rarely in biological molecules. Only a handful of naturally occurring organofluorine compounds are known, the majority being ω-fluorinated carboxylic acids biosynthesized by the plant *Dychapetalum toxicarium* [1]. In contrast, chemists have synthesized many thousands of fluorinated molecules; indeed, there is a whole journal devoted to the chemistry of fluorine. The carbon–fluorine bond is extremely strong (about 14 kcal/mol stronger than a C–H bond), and perfluorinated carbons are inert to substitution reactions. For this reason, fluorocarbons have found important industrial and medical uses as plastics, refrigerants and fire retardants, and as anesthetics. Fluorinated molecules have also found many uses in investigations of biochemical problems, proving to be valuable probes of enzyme mechanisms, protein structure and metabolic pathways.

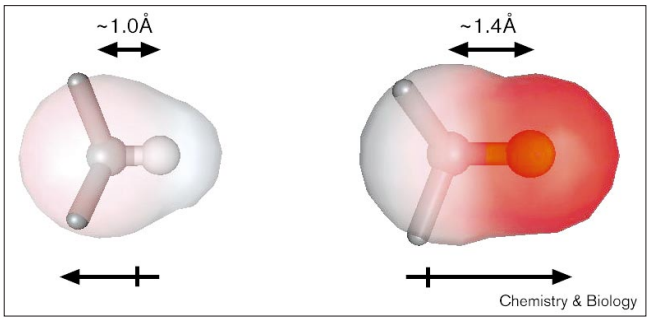
Several physical properties of fluorine make it attractive to chemists wishing to investigate biochemical problems. Firstly, fluorine is extremely small, and for this reason it is often thought of as isosteric with hydrogen (Figure 1). In fact, although the van der Waals radius of fluorine is 1.35 Å, which is only 0.15 Å larger than hydrogen, a C–F bond is significantly longer (~1.4 Å) than a C–H bond (~1.0 Å), and fluorine is better considered as isosteric with oxygen. Nevertheless, the substitution of fluorine for hydrogen is very often sufficiently conservative that fluorinated analogs of natural compounds are still recognized by the target enzymes or receptors. For example, fluorinated amino acids such as trifluoromethylmethionine and fluorine-substituted analogs of tryptophan, phenylalanine and tyrosine are recognized by their cognate amino acyl-tRNA synthetases and incorporated into proteins [2].

Secondly, fluorine is the most electronegative element, and therefore the electronic properties of the C–F bond are quite different from those of a C–H bond. The dipole moment of a C–H bond is relatively small and points towards the carbon, whereas that for a C–F bond is much larger and points towards the fluorine. Similarly, fluorine exerts a strong inductive effect that is sufficient to perturb the reactivity of atoms several bonds removed from the fluorine. For example, introducing a fluorine atom onto carbon-4 of proline decreases the pK_a of the amino group by 1.6 pH units, even though the fluorine is three bonds away from the nitrogen [3].

The naturally occurring isotope of fluorine, ^{19}F, is spin fi and has excellent nuclear magnetic resonance (NMR) properties. It is the second most sensitive nucleus after hydrogen and exhibits a very wide range of chemical shifts that are sensitive to the environment of the fluorine [2,4]. This, combined with the fact that there is no naturally occurring background signal, has made fluorine NMR a useful tool for investigating protein structure and dynamics, and for examining binding of fluorinated ligands to their targets [2].

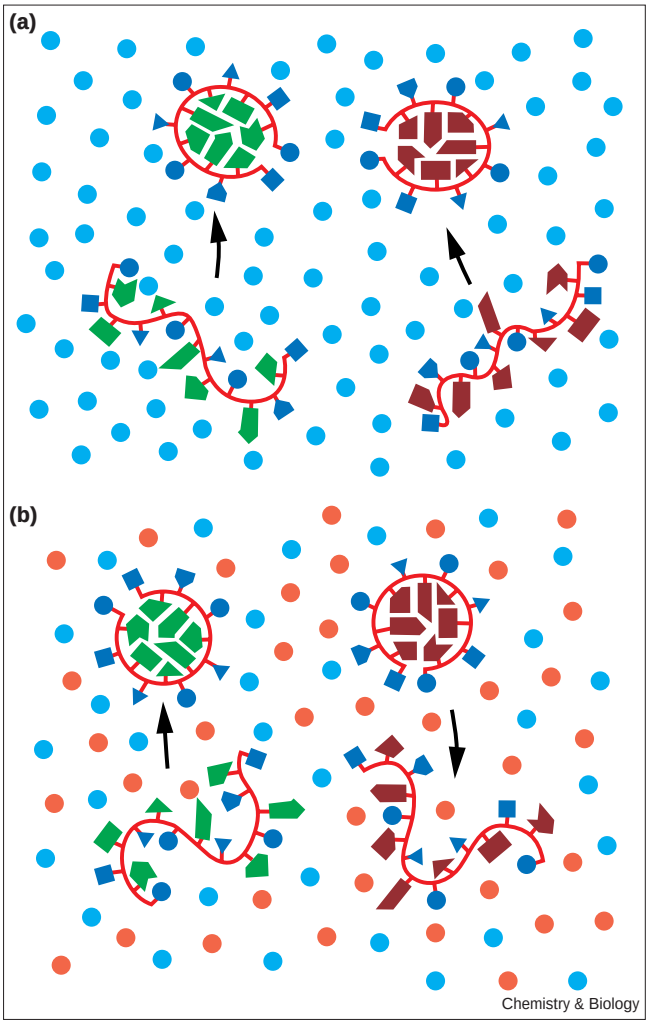
Perfluorinated carbon compounds exhibit interesting solubility properties. Partitioning measurements of fluorocarbon molecules reveals them to be much more hydrophobic than their hydrocarbon counterparts. The partition constant, Π, for the trifluoromethyl group is 1.07, making it over twice as hydrophobic as a methyl group, Π = 0.5 [5]. This property of fluorocarbons has been known for sometime, and has been exploited to increase the lipophilicity, and hence the bioavailability, of various drugs. A high-profile example is the weight-loss drug fenfluramine that contains a trifluoromethyl group.

Figure 1



Comparison of the properties of the carbon–hydrogen bond (left) with the carbon–fluorine bond (right). The van der Waals surface is shaded according to the electrostatic potential.

Figure 2



(a) A schematic illustrating how the hydrophobic effect would drive the folding of both fluorous proteins and natural proteins in aqueous solvents. Fluorous proteins should, however, be resistant to denaturation by organic solvents (b) because the fluorocarbon sidechains will partition away from the organic phase. Hydrophilic sidechains are blue, hydrophobic sidechains brown and fluorous sidechains green. Water molecules are represented by light blue circles and organic solvent molecules by orange circles.

Most interestingly, although fluorocarbons are hydrophobic, perfluorinated molecules tend also to be poorly soluble in hydrocarbon solvents. Thus a mixture of water, hexane and perfluorohexane separates into three layers, each compound being mutually immiscible with the other two. With this in mind, perfluorocarbons are better described as fluorophilic, rather than hydrophobic or lipophilic. This 'fluorous effect' is responsible for the nonstick properties of Teflon because the perfluorinated polymer interacts neither with hydrophilic molecules nor with lipophilic molecules. The fluorous effect has recently been exploited in organic synthesis to facilitate the purification of compounds [6]. By attaching a perfluorocarbon tail to a suitable functional group (such as a hydroxyl or amino group) of one of the starting materials, the products of the reaction can be extracted from the reaction mixture with perfluorohexane, the partitioning into perfluorohexane being driven by the fluorophilicity of the perfluorocarbon tail.

Could the novel phase partitioning properties of fluorocarbons be harnessed in the design of 'Teflon' proteins that might combine biological activity with some of the interesting and valuable properties of abiological fluorocarbons described above? The simplest model for a folded, globular protein is that of a hydrophobic core of 'greasy' amino acids surrounded by a shell of hydrophilic amino acids that provide solvation, as illustrated in Figure 2, with folding being largely driven by the partitioning of hydrophobic residues away from the solvent. But suppose that the hydrocarbon core of the protein were replaced by perfluorinated amino acid sidechains; what properties might this hypothetical protein possess?

It seems reasonable that a hypothetical fluorous protein might fold into a structure similar to that of the wild-type protein. The fluorous residues would still be hydrophobic, thereby maintaining the driving force for folding. Fluorine is bigger than hydrogen, so a fluorous protein would have a larger hydrophobic core than its hydrocarbon counterpart, but many proteins seem able to accommodate moderate changes in sidechain volumes through local reorganization of sidechain packing. Whereas not all natural protein folds are likely to be compatible with a fluorous core, there is a good chance that some of the simpler folding motifs, at least, would translate into stably folded fluorous proteins.

In practice, rather than perfluorinating every buried residue, one would probably design such a fluorous protein using a limited set of hydrophobic amino acids, fluorinated at the extremities of the residue — which is where the majority of inter-residue contacts occur. Indeed, amino acids fluorinated at the β carbon are probably best avoided because the strong electron-withdrawing effect of fluorine would make the amino group much less nucleophilic, which could in turn create problems in the synthesis of the peptide (whether by chemical or biological strategies). A starting set of fluorous amino acids might include hexafluoroleucine, hexafluorovaline, trifluoromethionine and pentafluorophenylalanine (Figure 3), all of which are relatively straightforward to prepare [7–9].

The most straightforward way of synthesizing small fluorous peptides is to use automated peptide synthesis methods. This approach allows almost any amino acid to be incorporated into a peptide, thereby providing enormous flexibility for the design of a protein. A major disadvantage has been that for practical purposes synthesis is restricted to relatively short peptides of less than 50 residues. The recent development of native chemical

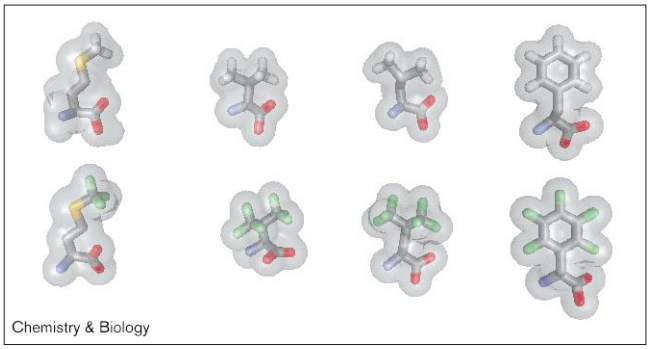
ligation strategies, however, by Kent, Muir and coworkers [10,11] provides a very elegant method of joining several short peptides together under mild conditions, making the total chemical synthesis of proteins comprising 100 to 200 residues a reasonable proposition. Moreover, this technique can be adapted to produce semisynthetic proteins by ligating the peptide containing unnatural amino acids to a larger protein fragment produced by overexpression in *Escherichia coli* [12].

The other approach is to use the cell's own protein-synthesizing machinery to incorporate the desired fluorinated amino acids into proteins. This approach relies on the appropriate amino acyl-tRNA synthetase enzyme recognizing the fluorinated amino acid analog with similar efficiency to the natural amino acid. It has been known for a long time that various mono-fluorinated aromatic amino acids are readily incorporated into proteins, but more extensively fluorinated amino acids may prove harder to introduce. Although methyl-trifluoromethionine and 5,5,5-trifluoroleucine are substrates for amino acyl-tRNA synthetases, they are also quite toxic to cells and inhibit their growth [13]. However, the inducible, high-efficiency expression systems that are widely used today to over-express proteins in *E. coli* effectively subvert most cellular functions towards protein synthesis. It is therefore possible to incorporate some unnatural amino acids in high yields into proteins even though they do not support cell growth. The feasibility of this approach has been demonstrated by Tirrell and coworkers [14,15], who have introduced various unnatural amino acids, including fluorinated amino acids, into small proteins in their experiments to produce novel polymeric materials.

A more significant problem is that extensively fluorinated amino acids might not be recognized by the intended amino acyl-tRNA synthetase, so the specificity of the synthetase would need to be re-engineered. Although this is not a trivial problem, the crystal structures of many amino acyl-tRNA synthetases are now known, and significant efforts are being directed towards re-engineering these enzymes with the ultimate goal of expanding the genetic code [16]. We may be optimistic that these efforts will be successful, and that, as our ability to design proteins continues to improve, it will eventually be possible to incorporate almost any amino acid site-specifically into a protein through an appropriately designed tRNA–aminoacyl-tRNA synthetase pair.

Having established that the synthesis of fluorous proteins is, in principle, feasible, what properties might such proteins possess? As noted above, fluorocarbons are intrinsically more hydrophobic than hydrocarbons, and as partitioning of hydrophobic residues out of the aqueous phase is a major driving force in protein folding, we might expect that fluorous proteins would be more thermodynamically stable than

Figure 3



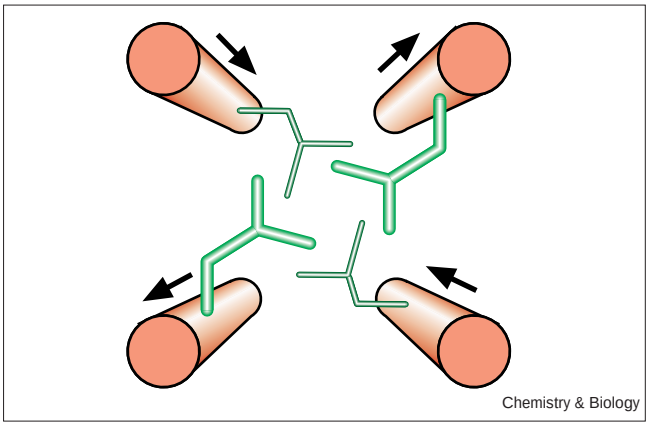
Hydrophobic amino acids (methionine, valine, leucine and phenylalanine, top row left to right) and fluorous analogs (below) that might be useful building blocks in the design of fluorous proteins. Van der Waal's surfaces are superimposed over each structure.

natural proteins. This property might be exploited to produce super-stable proteins that retain structure and activity at high temperatures or under other denaturing conditions. Alternately, it may be possible to design very small peptides with defined structure using fluorous amino acids, because the number of hydrophobic residues necessary to form a stable, folded core should be fewer.

Whereas the increased hydrophobicity associated with fluorous proteins and peptides may prove valuable in protein design, many interesting properties of such proteins may arise from their fluorous nature. The fluorous effect predicts that fluorous proteins should be resistant to denaturation by normal organic solvents such as ethanol, because the fluorocarbon sidechains are fluorophilic rather than lipophilic (i.e. partitioning into hydrocarbon solvents is disfavored; Figure 2b). Solvents such as trifluoroethanol, which are often used to promote structure formation in small peptides, should, however, be effective denaturants as the fluorous effect would result in the fluorinated amino acid residues partitioning into the fluorinated solvent.

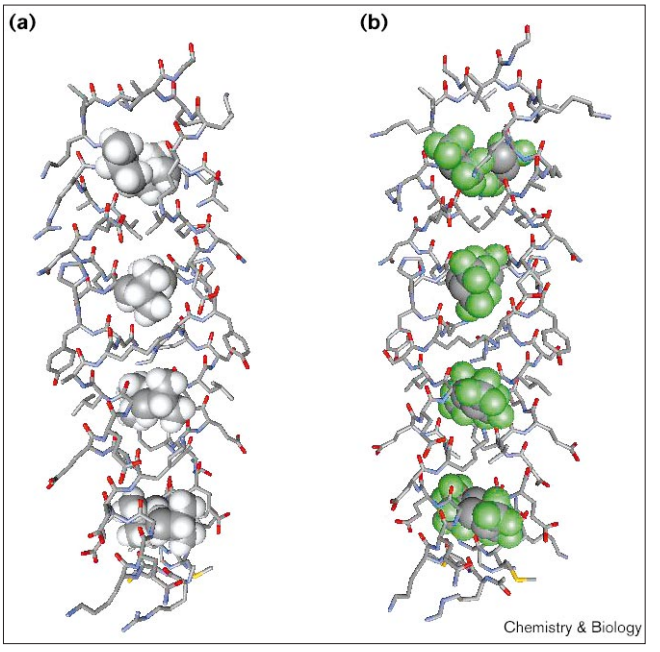
A good starting point for the design of fluorous proteins may be 4-helix bundle proteins, such as cytochrome c and the *E. coli* repressor of primer (ROP) protein. These are small proteins and have been the subject of extensive *de novo* design efforts so that the packing of the hydrophobic core is well understood [17]. In particular, studies by Regan and coworkers [18] have shown that the hydrophobic core of a 4-helix bundle protein can constructed from a very simple packing scheme involving only alanine and leucine at the 'a' and 'd' positions of the helical heptad repeat (Figure 4). Thus 4-helix bundles incorporating hexafluoroleucine in place of leucine should form extensive and well-ordered intermolecular fluorocarbon–fluorocarbon contacts, allowing the effect of these contacts in stabilizing protein folding to be evaluated.

Figure 4



Schematic illustrating how a hydrophobic protein core can be constructed in an antiparallel 4-helix bundle using a simple packing scheme employing leucine sidechains [18]; a fluorous protein core might be created using hexafluoroleucine residues.

Figure 5



(a) The leucine zipper domain of yeast GCN4 protein [20] showing the bridging leucine residues in space-filling representation. (b) Model of the GCN4 domain, in which the bridging leucine residues have been replaced by hexafluoroleucine and the structure energy minimized using the SYBYL force field. Modeling was performed using the PC Spartan Plus software package (Wavefunction Inc., CA, USA). The model suggests that fluorous contacts can be engineered into a protein with only very minor perturbation of the protein structure.

In addition to its role in protein folding, hydrophobicity is important in protein–protein recognition. A paradigm of this type of interaction is the 'leucine zipper' coiled-coil motif that is responsible for the dimerisation of many transcription factors. The most extensively studied example is the dimerisation domain from yeast GCN4 protein, and there have also been a number of *de novo* designed coiled-coil domains [17]. Molecular modeling suggests that the leucine residues in the GCN4 zipper can be replaced by hexafluoroleucines without greatly perturbing the structure, as shown in Figure 5. Furthermore, the phase partitioning properties of fluorous molecules predicts that dimerisation should occur specifically between two fluorous GCN4 peptide analogs (because a hydrocarbon–fluorocarbon interaction would be relatively unfavorable) resulting in a Teflon zipper(!). (Interestingly, close inspection of the model in Figure 5 reveals that only the pro-R trifluoromethyl groups on each pair of hexafluoroleucines make intermolecular fluorous contacts, suggesting that trifluoroleucine might be equally effective.)

The dimerisation of proteins plays an important role in many signal transduction pathways. Thus, if proteins with fluorinated dimerisation domains could be introduced into the cell (see the discussion below), the fluorous effect potentially provides a highly specific way of associating one protein with another. Fluoroprotein–fluoroprotein interactions should be orthogonal to all other protein–protein interactions, and as such might prove to be useful probes of cellular function.

Apart from the intellectual curiosity associated with their novel properties, what practical applications might such fluorous proteins have? Several known properties of fluoropeptides may lend them to pharmaceutical uses. Natural peptides and proteins are generally considered poor drug candidates because, in general, they do not easily cross membranes and are prone to degradation by proteases. It has been found that fluorination increases the bioavailability of many drugs, probably because increased lipophilicity helps them to cross the cell membrane [5]. Small, extensively fluorinated peptides may therefore have much better membrane-crossing properties that would make them useful probes of cellular function and potential drug candidates. Fluorinated peptides may also be resistant to proteases; for example, analogs of the peptide hormone angiotensin II containing hexafluorovaline retain activity but are resistant to a variety of proteases [19].

The excellent NMR sensitivity of the ^{19}F nucleus means that fluoroproteins and peptides should be excellent imaging agents. One can envisage it being possible to design peptides that use the fluorous effect to achieve a stable folded structure with a minimal number of residues, and that display on their surfaces binding sites for cell-specific receptor proteins. Such proteins might be targeted, for example, to receptors that are over-expressed in cancer cells, and this could allow very small tumors to be viewed using magnetic resonance imaging (MRI) because

the fluorinated core of the protein would provide a strong ^{19}F signal with almost no background.

In conclusion, molecules that combine the intricate structures of natural proteins with the unusual physicochemical properties of man-made fluorocarbons, should have novel and potentially useful characteristics. This class of molecules remains essentially unknown at present, but the technology now exists to synthesize small proteins containing completely unnatural amino acids. Their properties will undoubtedly differ in unexpected ways from those that I have predicted here (for example, the above discussion omits any consideration of effects resulting from the change in dipole moment — or, most interestingly, for pentafluorophenylalanine quadrupole moment — associated with the incorporation of fluorine). Nevertheless, 'Teflon' proteins could provide insights into protein folding and enzyme catalysis; they may prove useful tools to probe cellular pathways and could perhaps be developed as therapeutic agents against disease; eventually, they may even provide us with nonstick eggs!

Acknowledgements

Research in my laboratory is currently supported by the National Institutes of Health, the Petroleum Research Fund and Parke Davis Pharmaceutical Research. I thank Mark Banaszak Holl for stimulating discussions on the chemistry of fluorine.

References

1. Harper, D.B. & O'hagan, D. (1994). The fluorinated natural-products. Natl Prod. Rep. 11, 123-133.
2. Gerig, J.T. (1994). Fluorine NMR of proteins. Prog. NMR Spectrosc. 26, 293-370.
3. Eberhardt, E.S., Panasik, N. & Raines, R.T. (1996). Inductive effects on the energetics of prolyl peptide bond isomerization: implications for collagen folding and stability. J. Am. Chem. Soc. 118, 12261-12266.
4. Duewel, H., Daub, E., Robinson, V. & Honek, J.F. (1997). Incorporation of trifluoromethionine into a phage lysozyme: implications and a new marker for use in protein F-19 NMR. Biochemistry 36, 3404-3416.
5. Resnati, G. (1993). Synthesis of chiral and bioactive fluoroorganic compounds. Tetrahedron 49, 9385-9445.
6. Studer, A., et al., & Curran, D.P. (1997). Fluorous synthesis: a fluorous-phase strategy for improving separation efficiency in organic synthesis. Science 275, 823-826.
7. Zhang, C., Ludin, C., Eberle, M.K., Stoeckli-Evans, H. & Keese, R. (1998). Asymmetric synthesis of (S)-5,5,5,5′,5′,5′-hexafluoroleucine. Helv. Chim. Acta 81, 174-181.
8. Eberle, M.K., Keese, R. & Stoeckli-Evans, H. (1998). New synthesis and chirality of (-)-4,4,4,4′,4′,4′- hexafluorovaline. Helv. Chim. Acta 81, 182-186.
9. Houston, M.E., Harvath, L. & Honek, J.F. (1997). Synthesis of and chemotactic responses elicited by fMet-Leu-Phe analogs containing difluoro- and trifluoromethionine. Bioorg. Med. Chem. Lett. 7, 3007-3012.
10. Wilken, J. & Kent, S.B.H. (1998). Chemical protein synthesis. Curr. Opin. Biotechnol. 9, 412-426.
11. Cotton, G.J. & Muir, T.W. (1999). Peptide ligation and its application to protein engineering. Chem. Biol. 6, R247-R256.
12. Cotton, G.J., Ayers, B., Xu, R. & Muir, T.W. (1999). Insertion of a synthetic peptide into a recombinant protein framework: a protein biosensor. J. Am. Chem. Soc. 121, 1100-1101.
13. Chianelli, M.S., Stella, C.A., Saenz, D.A., Ramos, E.H., Kotliar, N. & Mattoon, J.R. (1996). Isolation of a trifluoroleucine-resistant mutant of Saccharomyces cerevisiae deficient in both high- and low- affinity L-leucine transport. Cell. Mol. Biol. 42, 847-857.
14. Kothakota, S., Dougherty, M.J., Fournier, M.J., Mason, T.L., Yoshikawa, E. & Tirrell, D.A. (1995). Templated biological synthesis of polymers of abiological monomers. Macromol. Symp. 98, 573-583.
15. Tang, Y. & Tirrell, D.A. (1999). Stabilization of leucine zipper coiled coils by introduction of trifluoroleucine. Abstr. Pap. Am. Chem. Soc. 218, 416-ORGN.
16. Liu, D.R. & Schultz, P.G. (1999). Progress toward the evolution of an organism with an expanded genetic code. Proc. Natl Acad. Sci. USA 96, 4780-4785.
17. DeGrado, W.F., Summa, C.M., Pavone, V., Nastri, F. & Lombardi, A. (1999). De novo design and structural characterization of proteins and metalloproteins. Ann. Rev. Biochem. 68, 779-819.
18. Munson, M., et al., & Regan, L. (1996). What makes a protein a protein? Hydrophobic core designs that specify stability and structural properties. Prot. Sci. 5, 1584-1593.
19. Hsieh, K.H., Needleman, P. & Marshall, G.R. (1987). Long-acting angiotensin-II inhibitors containing hexafluorovaline in position-8. J. Med. Chem. 30, 1097-1100.
20. O'shea, E.K., Klemm, J.D., Kim, P.S. & Alber, T. (1991). X-Ray structure of the Gcn4 Leucine zipper, a 2-stranded, parallel coiled coil. Science 254, 539-544.