2 Fluorine in Bioactive Molecules – The Way from Medicinal Chemistry to Peptide Design and Protein Engineering

This chapter will provide a detailed insight into the role of organic fluorine in bioorganic chemistry in respect to its use and applicability for peptide and protein chemistry. It will start with fluorine's unique physical and chemical properties and discuss the most important controversies that are present in the recent scientific literature. The following subchapters will summarize the importance, fluorine evolved for medicinal chemistry and for the analyses of protein structure and function. The benefit that fluorine provides for the modulation of biological activity and stability of peptides and proteins will be described. In this context, the necessity of a systematic investigation of the interaction behavior as well as of the physical and chemical properties of fluoroalkyl-substituted amino acid side chains in a native protein environment will be explained.

2.1 The Properties of Organic Fluorine

Several properties of fluorine as a heteroatom in organic compounds are still controversially discussed in the recent literature. These features comprise physical properties like steric demand and hydrophobicity of functional groups as well as their characteristics of intermolecular interactions, such as partitioning in hydrogen bonds and the formation of fluorine-fluorine interaction.

2.1.1 Fluorine's Capability to Participate in Hydrogen Bonds

Undoubtedly, the ability to act as a hydrogen bond acceptor is the most discussed issue of organic fluorine in literature.[45] Since hydrogen bonds are indispensable features in higher-ordered peptide, protein, and DNA structure and, thus, provide a considerable contribution to the biological function of these macromolecules, this hard-argued aspect increases the value of fluorocarbon-modified amino acids for peptide design and protein engineering. For decades many independent studies on different molecular systems could not clarify this issue.

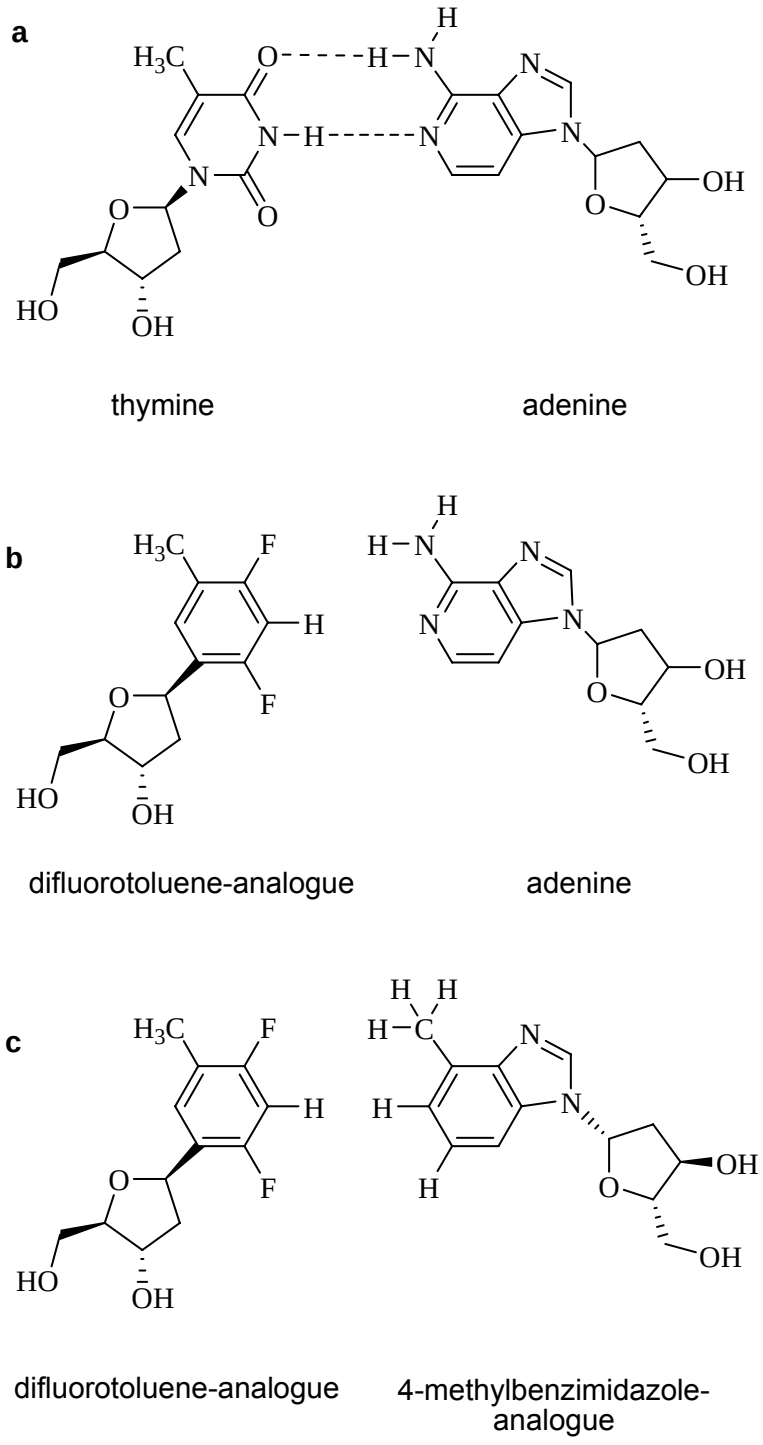
Recent screenings of the Cambridge Crystallographic Structural Database for relevant F...H contacts and *ab initio* energetic calculations by Dunitz and Taylor led to the conclusion that

organic fluorine is at best a very weak hydrogen bond acceptor.[46] According to these studies, C-F...H-C contacts represent the majority of contacts while interactions between carbon-bound fluorine and amine or hydroxy groups are not so common.[47] However, the contacts that were found in these studies were denoted as weak polar interactions rather than hydrogen bonds because of their calculated weak energies. Another, more recent database screening that was performed by Carosat et al. supports the opposite assumption. This exhaustive analysis of the Protein Data Bank (PDB) was done in order to determine the behavior of fluorine as a hydrogen bond acceptor from protein-ligand complexes with cocrystalized fluorine-containing ligands.[48] On the basis of the obtained statistical data, the specific fluorine-hydrogen bonding geometry has been characterized in detail. In this context, a new angular function for fluorine was developed, which could be shown to improve computational methods for the analysis of fluorine-containing ligand-target complexes. Based on these upgraded calculation studies, the authors state that fluorine-containing hydrogen-bonding interactions can be characterized in 18% of the investigated protein-ligand complexes.

Very often, opponent results do not require different experimental strategies or molecular systems that are objects of investigation, but the obtained data can be interpreted in conflictive ways. A descriptive example is provided by DNA polymerase activity studies with thymine analogues, carrying difluorotoluene as nucleoside, which were performed in order to investigate the importance of hydrogen bonds in Watson-Crick base pairings for the fidelity of these enzymes.[49,50] The results of these investigations have led to a new debate about the hydrogen bond acceptor capability of fluorine. Although it was shown that the thymine mimic 2,4-difluorotoluene pairs very poorly with adenine, it appeared to be a surprisingly good substrate for the DNA polymerase. The enzyme efficiency was lowered only by a factor of 40 (Scheme 2.1a,b). Since the authors describe this fluorinated nucleoside analogue to be apolar and unable to form the Watson-Crick like F...H-N hydrogen bond, they consequently devaluated the contribution of hydrogen bonds to polymerase fidelity. Evans and Seddon interpreted these results the opposite way and considered the F...H-N contact to be strong enough to form hydrogen bonds.[51]

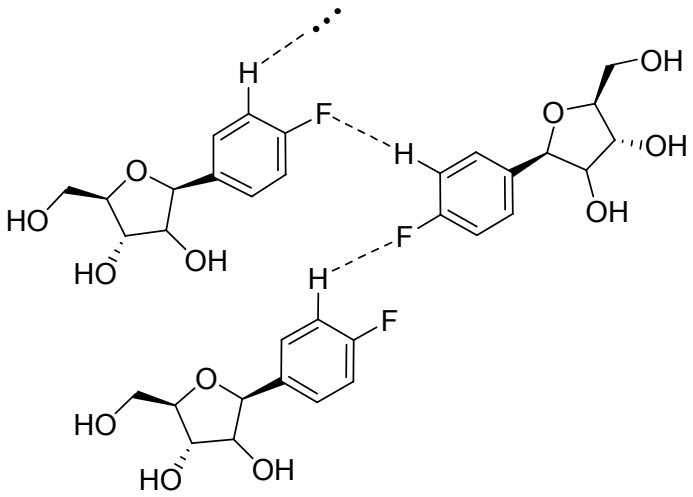
Similar experiments with 2,4-difluorotoluene as a thymine mimic and 4-methyl-benzimidazole (Scheme 2.1c) instead of adenosine as the counterpart for base-pairing could prove that this non-natural base pair is replicated surprisingly well and selectively by DNA polymerase enzymes.[52] In the investigated adenine analogue, the two nitrogen heteroatoms in the six-membered ring and the amino group, which is substituted by methyl, are missing. According

to the theory of Evans and Seddon, C-F...H-C hydrogen bonds must significantly contribute to polymerase fidelity.



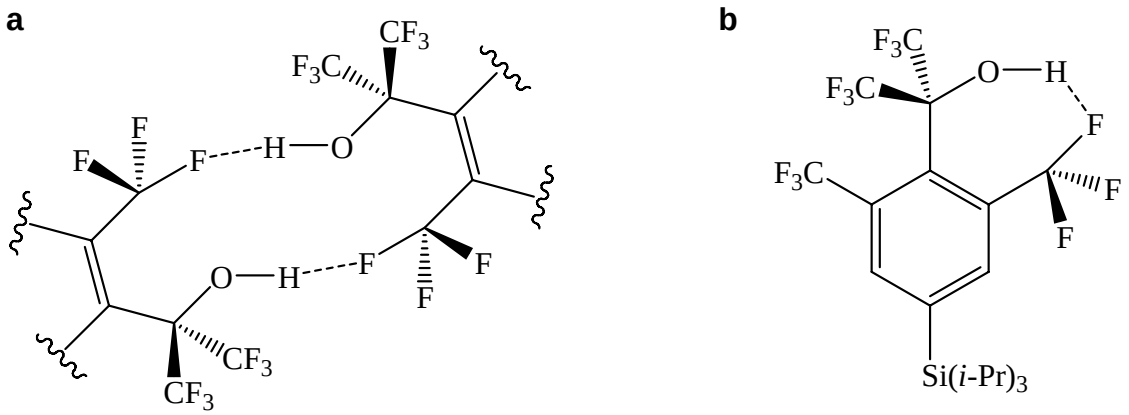
Scheme 2.1: *Watson-Crick pairing of thymine and adenine **(a)**, and base-pair stacking of adenine with a fluorinated thymine analogue **(b)**, according to Diederichsen; **c)** interactions of both fluorinated nucleobase derivatives, according to Guckian et al.*

A few years later, Parsch and Engels reported such C-F...H-C type hydrogen bonds in RNA base pairs of fluorobenzene and fluorobenzimidazole nucleobase analogues, which were proved by crystal structures and calculations from thermodynamic data (Scheme 2.2).[53]



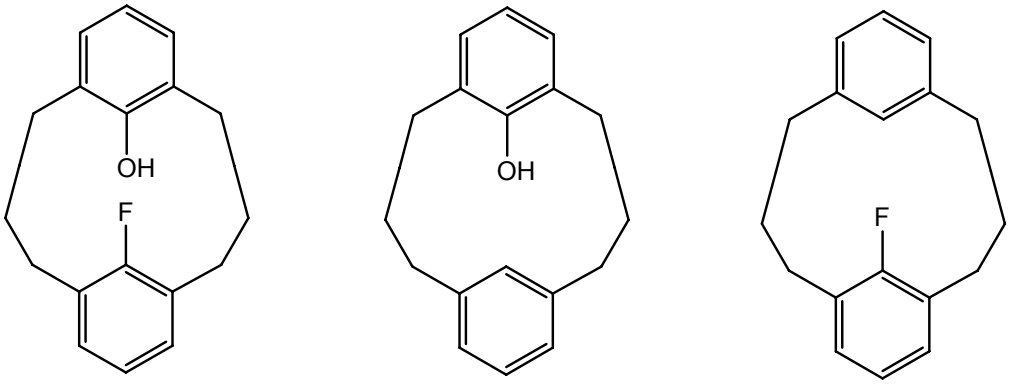
Scheme 2.2: *C-F…H-C type hydrogen bonds in crystals of 1'-deoxy-1'(4-fluorophenyl)-β-D-ribofuranose according to Parsch and Engels.*

Following these studies, weak intramolecular C-F…H-C type hydrogen bonds in polyolefin procatalysts were detected.[54] Although they are described to be even less important than C-F...H-C contacts, their C-F...H-O analoga have been characterized as hydrogen bonds in small molecular model systems. In tetra-trifluoromethyl substituted benzyl alcohols, intra- and intermolecular C-F...H-O type hydrogen bonds have been verified in hexane solution by NMR and in solid-state by x-ray diffraction, respectively (Scheme 2.3).[55]



Scheme 2.3: *Intermolecular C-F…H-C type hydrogen bonds in solid state **(a)** and intramolecular hydrogen bonds of the same type in hexane solution **(b)**, according to Barberich et al.*

The data of this study suggest that the linear fluorine-containing hydrogen bonds in the dimer are significantly stronger than the bent varieties in the dissolved monomer. In general, C-F...H-X type hydrogen bonds can be observed in rigid molecular systems where the optimal atomic distances and angels are realized and any stronger hydrogen bond acceptors are excluded. Model systems applying a cyclophane skeleton (Scheme 2.4) have been used to study F...H-O contacts systematically and have proved the formation of hydrogen bonds by atomic distance measurements in the crystals as well as by ^{1}H NMR spectroscopy.[56]



Scheme 2.4: *Substituted [3,3]metacyclophanes as models for the investigation of C-F...H-O type hydrogen bonds, according to Takemura et al.*

Fluorosubstitution can also increase the hydrogen bond donor capability of alkyl groups, as could be shown for CF_2H. Thereby, the high electron withdrawing effect of the fluorine atoms causes an enhanced acidity of the attached proton.[57] In view of these results, the study from Caminati et al., which suggests that the difluoromethane dimer is stabilized by three C-F...H-C hydrogen bonds, is not surprising.[58]

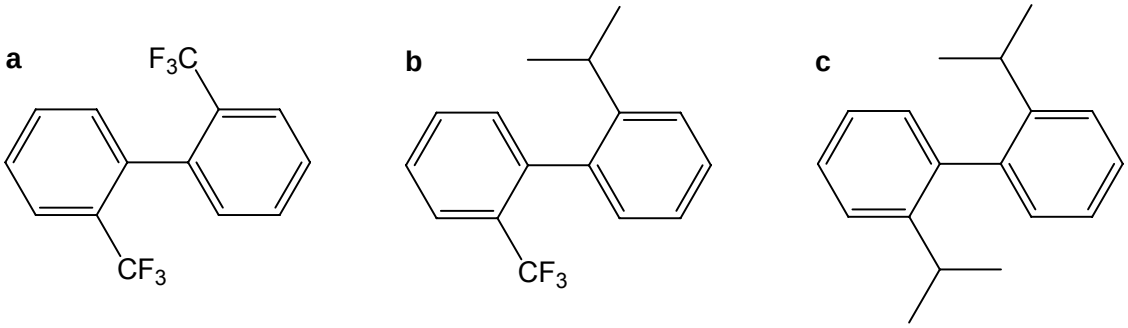
To date, systematic investigations describing the hydrogen bond-forming capability of fluorine in native protein environments in detail are missing. However, such knowledge could be very powerful, e.g., for the rational design of selective receptor binders or enzyme substrate inhibitors.

2.1.2 The Steric Demand of Fluorocarbon Groups

Similar to the hydrogen bond topic, investigations on another contentious issue, the size of fluorine and fluorine-containing organic groups, have not yet resulted in a consistent concept. The van der Waals radius of fluorine (1.47 Å) is close to that of oxygen (1.57 Å) and is, thus, the next smallest atom after hydrogen (1.2 Å).[59] Depending on the application, fluorine is

often used as a hydrogen substitute or a mimic of oxygen and hydroxyl, respectively, in organic chemistry.[60] Very interestingly, the steric demand of fluorocarbon groups seems to increase in an unproportional manner along with the extent of fluorine substitution.[61] For example, the trifluoromethyl group is often described to be at least as large as isopropyl. However, data of its exact size differ within a relatively wide range in literature. Thereby, the results mostly depend on the experimental methods from which they arise as well as on the molecular systems that were applied in the studies. One reason may be the difficulty to distinguish between the steric and electrostatic contribution of the fluorocarbon substituent on a chemical reaction and physical parameters, respectively. A further aspect that has to be considered when comparing such physicochemical data is the discrepancy often found between results that were obtained as "side products" from interpretations of differently motivated experiments and those that were achieved by sytematic investigations using designed molecular models. In the case of the molar volume of a trifluoromethyl group, the values vary from twice the bulk of the methyl group,[62] as calculated from van der Waals radii, to even close to the size of phenyl or *tert*-butyl groups.[63]

As discussed for the issue of fluorine's capability to participate in hydrogen bonds, similar data for the size of a fluorocarbon group that were derived from studies applying the same molecular system can lead to differing interpretations. This is nicely highlighted by investigations about the rotational energy barrier in *ortho* substituted biphenyls (Scheme 2.5).

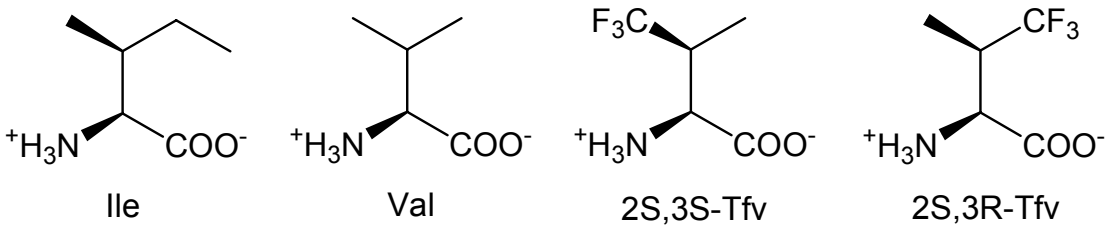


Scheme 2.5: *2,2' substituted biphenyls according to Wolf et al.* ***a)*** *2,2'-bis(trifluoromethyl)biphenyl,* ***b)*** *2-isopropyl-2'-(trifluoromethyl)biphenyl and* ***c)*** *2,2'-diisopropyl-biphenyl.*

The rotational energy barriers of all three types of biphenyls that are substituted with isopropyl and trifluoromethyl in *ortho* position have been determined by polarimetry and dynamic gas chromatography.[64] Almost the same values could be obtained for the bis(trifluoromethyl) and for the isopropyl-trifluoromethyl analoga, while the diisopropylbiphenyl derivative shows a slightly higher stability. The authors of this work declare the electronic influence of the trifluoromethyl groups to be negligible and conclude

that the energy barriers are reasoned by the bulkyness of the *ortho* substituents, which is, therefore, almost equivalent for trifluoromethyl and isopropyl. The interpretation of similar results by Leroux takes into account that the rotationally unsymmetrical isopropyl group can be oriented in a more favorable way, when passing the substituent of the other phenyl ring.[65] Such reorientation is impossible in case of the symmetrical trifluoromethyl group. As a result, the anisotropic effect lets the isopropyl group appear only slightly larger than the trifluoromethyl substituent, although it is supposed to have a much bigger volume.

However, there are several indications that the trifluoromethyl group is a mimic of a larger alkyl moiety rather than of the methyl group. Substitutions of *iso*-butyl groups by trifluoromethyl in pepstatin A, a subnanomolar inhibitor of many aspartyl proteases, resulted in only slightly less potent inhibitors of plasmepsin II, compared to the native peptide.[66,67] Crystal structures of both pepstatin A analogues with plasmepsin II exhibited almost identical conformations of the enzyme inhibitors. The result of this study suggests the trifluoromethyl group is a very effective mimic of isobutyl in peptide/protein interactions. Further studies with peptidomimetic hydroxamate inhibitors of matrix metalloproteinases with a methyl moiety replaced by trifluoromethyl could prove this substitution to provoke a dramatic loss in inhibitory activity towards different enzymes. Obviously, the trifluoromethyl group does not fit into the methyl-binding pocket of the enzymes.
It could be shown in a recent work that the 3*R*-form of 4,4,4-trifluorovaline (2*S*,3*R*-Tfv), which represents the diastereomer with the Tfm-group in the same orientation like that of the ethyl chain in Ile (Scheme 2.6), is accepted as a substrate for both the valyl- and the isoleucyl-tRNA synthetase.[68]



Scheme 2.6: *Amino acids used as substrates for valyl- and isoleucyl-tRNA synthetases, according to Wang et al.*

Thereby, the specificity constant for the valyl-tRNA synthetase was reduced by approximately 2500-fold, while the activation by the isoleucyl-tRNA synthetase was diminished by approximately 600-fold, each with respect to the activation of their native substrates Val and Ile, respectively. In contrast, (2*S*,2*S*)-Tfv, the diastereomeric form with the Tfm-group in the same orientation as the methyl group in Val and Ile was not accepted by

either enzyme. This study demonstrates that the trifluoromethyl group is a better surrogate for ethyl rather than for methyl in respect to interactions with proteins. Because the activity of the isoleucyl-tRNA synthetase was also strongly diminished, it can be assumed that Tfm is larger than ethyl as well.

Although the above-mentioned case studies provide indications about the effective size of fluorocarbon groups in respect to peptide-protein and protein-protein interactions, a systematic investigation is missing on this issue so far as well.

2.1.3 The Unique Electronic Properties of Organic Fluorine

Another valuable contribution to the special role of organic fluorine in bioorganic chemistry is provided by the unique properties of the C-F bond. The specialty is mainly based on the hydrophobic character of this covalent fluorine-carbon assembly, which is associated with a high dipole moment. This effect has recently been described as polar hydrophobicity.[69]

The origin of this unique property is, on the one hand, the electronegativity of fluorine, which is the highest among all elements, combined with a small van der Waals radius, resulting in a short and strong covalent bond. On the other hand, the closeness of the outer electrons to the highly charged nucleus leads to a very low polarizability of the C-F bond.[70] Additionally, the C-F bond causes a reduced overall molecular polarizability throughout the carbon framework of the whole fluorine containing molecule.

In spite of the polar nature of the C-F arrangement, the reduced polarizability caused by fluorination of an organic compound leads to a general increase in lipophilicity. The higher binding affinity of fluorocarbon groups, especially of fluoroalkylated compounds, to hydrophobic regions of, e.g., proteins, compared to their hydrocarbon analogues, is also based on the larger hydrophobic surface areas that are dissolved upon binding.[71] However, it is also stated in the literature that terminal mono-, di-, or trifluorination of an alkane typically decreases the lipophilicity of the molecule, although trifluorination of methyl groups is a widely applied approach in drug design to enhance the affinity of ligands for hydrophobic protein regions. These findings are not necessarily contradictory, as the evaluation of lipophilicity depends on whether it refers to the whole molecule or only to the substitution site that binds to the protein.

However, the dipole character of the C-F bond enables organic fluorine to participate in polar interactions as well. Since carbonyl functionalities are frequently found in active sites of enzymes, the multipolar interactions between the intrinsically polar C-F and C=O units can be effective tools in medicinal chemistry for enhancing ligand affinity and selectivity in lead compound optimization (Figure 2.1).[72] Such C-F...C=O contacts have been found in PDB

screenings for both aromatic C-F residues as well as for $-CF_3$ groups in protein-inhibitor complexes.

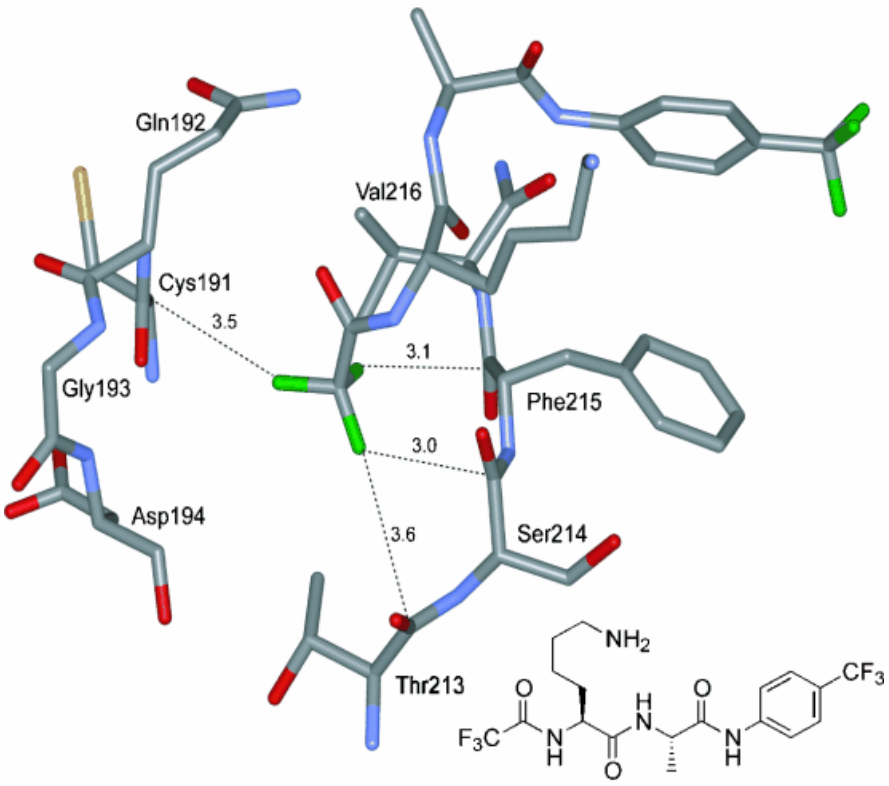


Figure 2.1: *C-F…C=O interactions seen in the crystal structure of the complex between porcine pancreatic elastase and an inhibitor of the trifluoroacetyl peptide class at 2.50 Å resolution (PDB entry 2EST).*[73] *All three F atoms of the CF_3CO group in the inhibitor interact with backbone C=O groups of Cys191, Thr213, Ser214 and Phe215 within the enzyme, according to Olsen et al.*

While the argument that the fluorination of aromatic rings always increases the lipophilicity of the molecule is prevalant in literature, a slightly decreased lipophilicity, caused by the polarity of C-F, has also recently been reported for some cases of aromatic monofluorination.[74] Polarity also strongly influences acidity and basicity of nearby functional groups, shifting the pK_a up to several log units.

All these characteristics of organic fluorine can be utilized by a directed incorporation of this functional group into bioactive compounds to create molecules with improved biochemical, physiological, and pharmacokinetical properties and, thus, to upgrade their values for medicinal chemistry. However, the debate about the electronic properties of different fluorinated alkyl groups, especially about the impact of fluorosubstitution on hydrophobicity, lipophilicity and polarity, demands a systematic investigation of this topic.

2.2 The Advanced Role of Fluorine in Medicinal Chemistry

Organic fluorine became a prominent tool for the design and improvement of pharmacokinetic properties of drug molecules as well as for powerful analytical methods in medical diagnostics.

2.2.1 Fluorine Substitution as a Tool in Drug Research

The outstanding significance of fluorine in the development of bioactive molecules has recently been reviewed by Böhm et al. One eminent effect of the substitution of a carbon-bound hydrogen by fluorine is a tremendous enhancement of metabolic stability caused by the strength and, therefore, the chemical stability of the C-F bond.[75] The introduction of fluorine can, therefore, alter both rate and route of drug metabolism as well as its tissue distribution. This strategy is most important for blocking the oxidative metabolism of a drug by enzymes like cytochrome P450. The rate, route, and extent of metabolic conversion can be influenced. Another very important beneficial aspect of this strategy of drug alteration is the prevention of chemically reactive and toxic metabolites in the body. It could be shown that the *in vivo* hydroxylation of phenyl carbon atoms in drugs by cytochrome P450 leads to metabolites that inhibit the metabolic active enzyme. Such undesirable impacts cause serious drug-drug interactions. The substitution of the effected hydrogen by fluorine blocks this enzymatic conversion owing to the enhanced bond strength. Consequently, fluorination of drugs not only affects the pharmacokinetical properties of the altered molecule, but lowers the toxicity of co-administered agents as well.[76] The big advantage of an aromatic fluorination is the minor impact that it causes on steric demand as well as on the charge of the compound.

The incorporation of fluorine into potent drugs can, however, tremendously affect their bioavailability by enhancing pharmacokinetical properties, such as membrane binding and permeation. This effect can be achieved either by weakening the basicity of a functional group, for example a nitrogen atom,[77] due to the electron-withdrawing capability of fluorine, or by a general increase in lipophilicity of the molecule.[78] The ability to penetrate membranes is indispensible for drugs that are supposed to have cerebral pharmacological effects, since they have to cross the blood-brain barrier. The incorporation of one trifluoromethyl group can thus enhance the membrane permeation level of an agent by almost one order of magnitude. This improvement is most likely caused by the larger hydrophobic surface area of a trifluoromethyl group, compared to its hydrocarbon analogue.

The increase in lipophilicity that is obtained by fluoroalkylation can further be applied in order to increase the binding properties of drug molecules to hydrophobic regions of its target molecule in the body. A lipophilic group attached to the ligand, which does not interfere with the active site of the enzyme to be inhibited, enhances its bioactivity upon binding to hydrophobic residues.

These drug design strategies of improving pharmacochemical properties of potent lead structures by incorporating organic fluorine (Figure 2.2) have led to the development of drugs for a wide variety of biochemical applications,[79] such as inhibitors of several enzymes, anticancer and antiviral agents, antibacterials, antidepressants, antinflammatory, and anorectic agents, hypolipidemic drugs, and antidiabetics.[80]

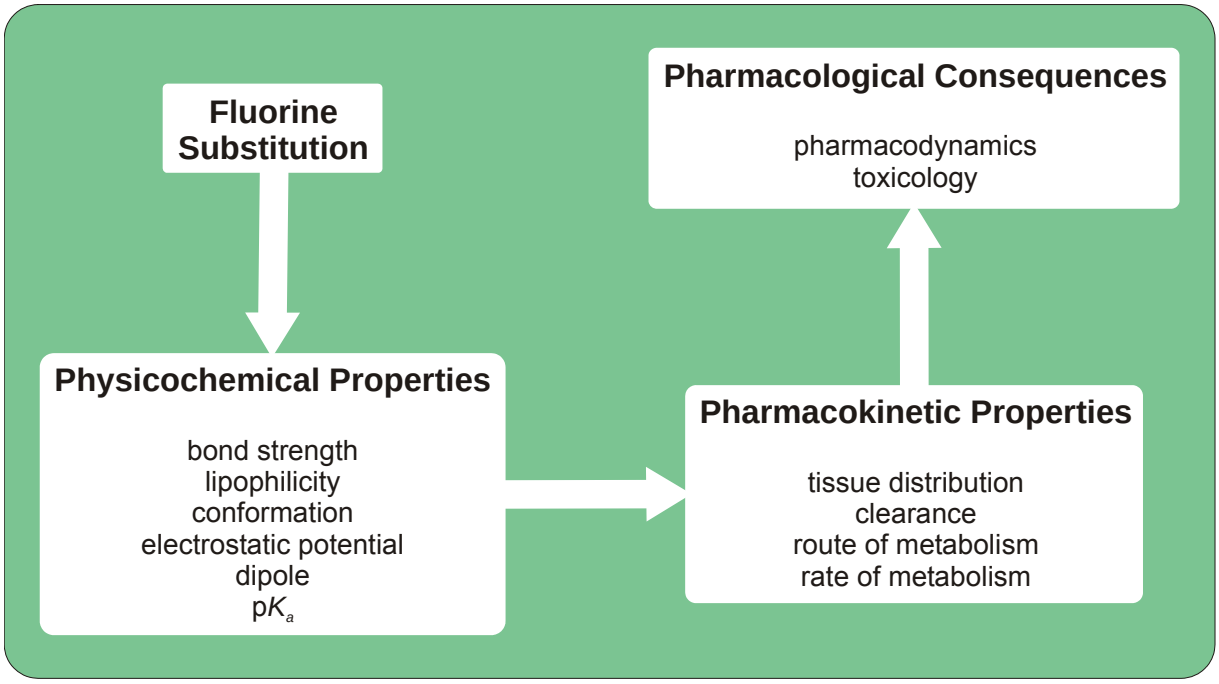


Figure 2.2: *Flow diagram illustrating the effect of fluorine substitution on drug response, according to Park et al.*

At present, more than two hundred fluorinated pharmaceuticals are available and others are appearing.[81] To date there is not much known, however, about the risk of these compounds for human health, e.g., caused by products of *in vivo* biotransformation. When reacting with nucleophilic groups in proteins and amino acids like the amino group in lysine, the sulfhydryl group in cysteine, or the hydroxyl group in serine, defluorination of fluorinated drugs can occur.

2.2.2 Fluorine in Medical Diagnostics – Raman Spectroscopy

A newly developing approach for the qualitative and quantitative analysis of fluoro-organic compounds that is based on Raman spectroscopy has recently been published.[82] Using a copper-vapor laser light source, the excitation of fluorocarbon compounds at 510.8nm results in specific emission bands in the range of 500-800 cm^{-1} for different types of carbon-fluorine bonds (e.g., trifluoromethyl-, aromatic organofluorine-, and difluoromethylene groups).[83] This analytical method can be applied for the investigation of metabolic pathways of administered drugs that contain at least one carbon-fluorine bond. This strategy, which uses novel C-F Raman labels with specific fingerprints, is advantageous because of the chemically easy incorporation of the labels, a directly proportional dependence of the signal on the concentration of the analyte, and the absence of signal-affecting background interferences. Researchers hope that C-F Raman spectroscopy will play an important role in medicinal diagnostics, e.g., for the detection of cancer and pre-cancer conditions using fluorinated bioprobes, such as proteins, nucleic acids, and drug molecules.

2.2.3 Fluorine in Medical Diagnostics – ^{18}F Positron Emission Tomography

Positron Emission Tomography (PET) is a very common and powerful analytical method for medical diagnostics. ^{18}F is the mostly applied radio label in such examinations, based on its low positron energy and the relatively long half-time of the isotope. The low emitting energy results in a marginal risk for the patient due to the short radiation range as well as the limited dose of emission. Fluorinated radiotracers in PET studies for imaging in cancer patients include 2-[^{18}F]fluoro-2-deoxy-D-glucose, analogues of amino acids, peptide derivatives, nucleosides, membrane phospholipids, and steroids.[84]

Possible problems of this analytical technique could be pointed out in studies with three ring-fluorinated isomers of L-DOPA,[85] which is widely used in ^{18}F PET, e.g., to probe the dopaminergic pathway in patients diagnosed with Parkinson's desease.[86] 2-[^{18}F]fluoro-L-DOPA, 5-[^{18}F]fluoro-L-DOPA, and 6-[^{18}F]fluoro-L-DOPA showed differing *in vivo* behavior, which indicates that minor chemical differences like the position of an aryl-fluorine substitution upon labeling, can have a remarkable impact on the performance of an analytical tracer.

2.3 Fluorine's Benefit in Peptide and Protein Chemistry

The described unique properties of organic fluorine, which are established features in drug design and development, attract rising interest in peptide design and protein engineering.[87,88,89] While the incorporation of fluorine into peptides and proteins is a widely applied strategy in protein analysis as well as for the modification of bioactive peptides and peptide mimics, its use for the stabilization of proteins and the design of non-natural peptide-protein interactions is a rather new field that continuously gains in importance.

2.3.1 Fluorine as a Tool in Protein Analysis – ^{19}F NMR Spectroscopy

Due to the fact that fluorine is usually not found in peptides and proteins, ^{19}F NMR spectroscopy has evolved into a sensitive and powerful tool in protein analysis, mainly because of the very low background signals combined with a high sensitivity, which is about 83% of that of ^{1}H NMR spectroscopy. A further advantage of this method is the relatively large range of chemical shift values, which gives access to the detection of even minor differences in the environment of the ^{19}F label. As many proteins and bioactive peptides dramatically change their structural conformation when taken out of their environment or are difficult to crystalize, ^{19}F NMR provides an important method, especially for investigating membrane-associated proteins[90] and peptides.[91,92]

2.3.1.1 Applications of ^{19}F NMR

^{19}F NMR spectroscopy is used to study protein conformations and their changes during folding and unfolding processes, solvent exposure and local mobility of specific regions and residues within the macromolecule, as well as the influence of ligand binding on protein structure. Folding and unfolding kinetics of proteins can be investigated upon substitution of specific amino acid residues by their fluorinated analogues that are located within different elements of the secondary and tertiary structure. The following analysis of urea-induced unfolding can then give insight into sequential denaturation and into the stability of the different structural elements and, therefore, provides a basis for the investigation of folding kinetics of the protein. A demonstrative example is given by the ^{19}F NMR-based evaluation of the unfolding process of a $(\beta/\alpha)_8$ protein by the substitution of four tryptophan residues with 6-^{19}F-Trp (Figure 2.3).[93]

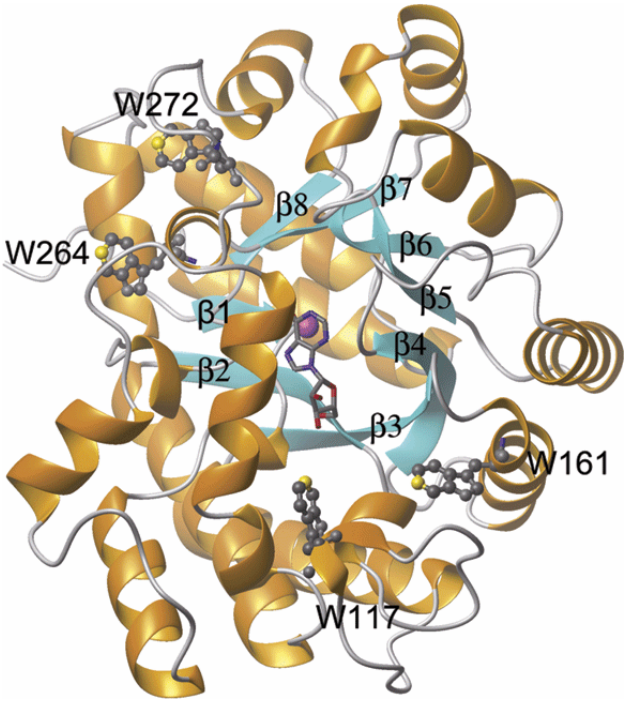


Figure 2.3: *A $(\beta/\alpha)_8$ protein (murine adenosine deaminase) with four 6-^{19}F-Trp residues for ^{19}F NMR spectroscopic studies. The central β-barrel is shown in blue, the α-helices in orange. The four tryptophan analogues are shown in ball-and-stick representations. The fluorinated C-6 atoms of their indole rings are colored in yellow (according to Shu et al.)*

The resulting resonances of the different labels have to be assigned unambiguously by specific site-directed mutagenesis of the labeled residues. This strategy can be also applied to study specific domain-domain interactions within proteins.[94] ^{19}F NMR is also used for studying the properties of amino acid residues such as solvent accessibility and motional environment in so-called molten globule states.[95] In addition to exploring the structural properties of integrated membrane peptides and proteins, the interactions of amino acid side chains with lipid layers, which are involved in many biochemical processes and molecular assemblies, can be studied using ^{19}F NMR.[96] The incorporation of amino acids into proteins that possess two diastereotopic fluorine atoms, e.g. difluoromethionine, can be used to characterize the rotational freedom of the amino acid side chains within solvent-exposed as well as tightly packed protein regions.[97]

2.3.1.2 Incorporation of ^{19}F NMR-Labels into Proteins

One way to label a protein for ^{19}F NMR studies is to incorporate fluorinated amino acids, commonly using fluorotryptophan, fluorophenylalanine, and fluorotyrosine. These

monofluoro-substituted derivatives of aromatic residues are easy to synthesize, exert no enhanced steric demand relative to their native analogues and, thus, do not cause perturbation of the protein structure. These building blocks can be incorporated biosynthetically using auxotrophic bacteria strains. The incorporation rates may differ over a wide range, depending on the fluorinated analogue to be inserted. There is naturally no specificity for any single position in the protein. Thus, the amino acid will be substituted in each position it occurs naturally. Site-directed mutagenesis has to be applied in order to assign the measured resonances to the different labels. An advancement of the biosynthetic incorporation of fluorine-labeled tryptophan into proteins has been recently reported.[98] Instead of using protein expression in auxothrophic strains, which includes extra cloning and bears the risk of low yields and poor incorporation levels, the inhibition of tryptophan biosynthesis with 3-β-indoleacrylic acid can be applied to accomplish the insertion of ^{19}F-Trp into the protein. Fluorinated non-natural aromatic amino acids such as 4-fluorophenylglycine (4F-Phg) are used for structural investigations of bioactive peptides.[99] Monofluorinated aromatic amino acids can be introduced into peptides as ^{19}F NMR lables applying standard peptide synthesis strategies.

In addition to applying fluorine-labeled aromatic amino acids, the biosynthetic incorporation of fluorinated analogues of aliphatic residues is used for ^{19}F NMR spectroscopy.[100] While the fluorination of aromatic amino acids does not significantly affect the steric demand of the altered residue, a trifluoromethyl group has a substantial larger volume than its hydrocarbon analogue. Even one trifluoromethylated amino acid can cause subtle conformational changes in protein structure.[101]

Other methods use site-directed mutagenesis to place cysteine residues in appropriate positions of the protein, followed by chemical loading of these side chains with a fluorine-containing label. In order to incorporate the labels only into the desired positions within the protein, naturally occuring cysteine residues have to be mutated, preferrably to alanine. The labels can then be attached to the cysteine side chains via thioalkylation, e.g., with 3-bromo-1,1,1-trifluoropropanone,[102] or by disulfide linkage, e.g. with 2,2,2-trifluoroethanethiol (TET).[103] Of course, such relatively large labels cannot be incorporated into tightly packed regions of a protein as this would dramatically disturb its overall folding.

Another opportunity to introduce a fluorine label for protein NMR is the labeling of protein binding ligands. Using this methodology, libraries of ligands can be screened for their strength of binding to specific proteins.[104] The strategy of labeling ligands can also be applied in order to characterize specific protein regions. Such investigations can provide very useful information for the design of functional proteins. For example, a metal-assembled

helical peptide trimer with a hydrophobic core has recently been used to evaluate structure and flexibility of hydrophobic binding cavities (Figure 2.4).[105]

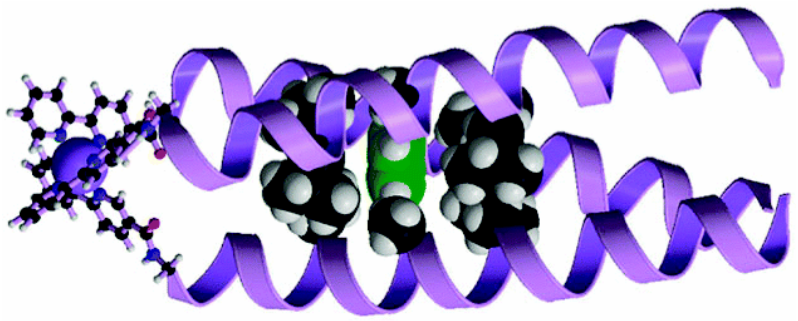


Figure 2.4: *Metal-assembled helical peptide trimer [Fe(bpy)GCN4-N16A$_3$]$^{2+}$. The Benzene-ligand is shown in green, the flanking hydrophobic amino acid residues are shown in black. The N-terminal bipyridine moieties are coordinating an iron atom (accoding to Doerr et al.)*

In this work, competitive titration studies of non-covalently bound hexafluorobenzene with molecules that differ in hydrophobicity and size were followed with ^{19}F diffusion NMR. Binding constants of this library, which derived from these experiments, helped to systematically evaluate the properties of the binding site.

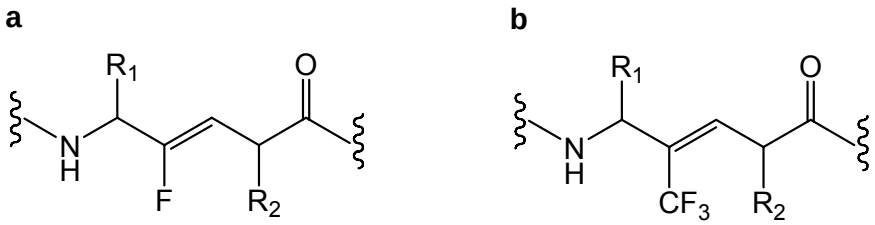
2.3.2 Fluorinated Building Blocks as Modulators of Biological Activity and Stability of Peptides

To date, several peptides that possess a biological relevance have been modified by inserting fluorinated building blocks in order to improve their structural and metabolic stability or to enhance the biological activity of the polymers. Thus allowing different strategies for placing the hetero atom within the peptides to be applied. In general, fluorine can either be incorporated into the amide bond moiety or into the side chains of aliphatic or aromatic amino acids.

2.3.2.1 Fluorinated Amide Bond Isosters

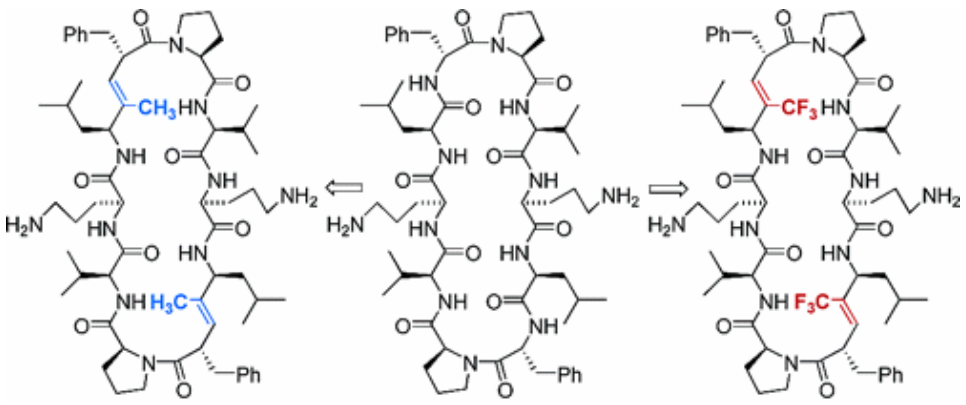
One strategy that exploits the unique steric and electronic features of fluorine to improve the properties of biologically active peptides, is the incorporation of fluorinated amide bond isosters that mimic the peptide linkage. In addition to an improved resistance towards proteolysis, fluorination of such moieties can lead to a better matching of steric demands and

to an alteration of the eletronic properties of the amide group. For example, the replacement of the amide group by a fluoroalkene surrogate in tripeptides (Scheme 2.7a) resulted in inhibitors of the zinc peptidase thermolysin that bind one order of magnitude more tightly than the natural enzyme substrates.[106] Calculations on dipole moments could show that the (trifluoromethyl)alkene isostere (Scheme 2.7b) represents an even better electrostatic mimic of the native amide bond relative to the monofluoro-substituted alkene moiety.[107] Furthermore, x-ray structural analysis revealed an enhancement in backbone rigidity of the CF_3-substituted surrogate relative to the fluorine-free alkene peptide bond isostere.



Scheme 2.7: *Fluoroalkene amide bond isosters according to* ***a)*** *Bartlett et al. and* ***b)*** *Wipf et al.*

The electrostatic similarity of the (trifluoromethyl)alkene isostere to the amide group was recently applied to design a fluorinated analogue of the antibiotic Gramicidin S. The derivatives had two hydrolyzable native peptide bonds substituted but retained the secondary structure that is crucial for its bioactivity.[108] While the analogue containing two methylalkene isosters revealed a much more flexible conformation, a structural stabilization of the cyclic peptide with CF_3-substituted peptide bond surrogates was observed. Thereby, the stabilizing electrostatic CF_3...NH-interactions even overcompensated the steric perturbation caused by the increased volume of the trifluoromethyl group (Scheme 2.8).

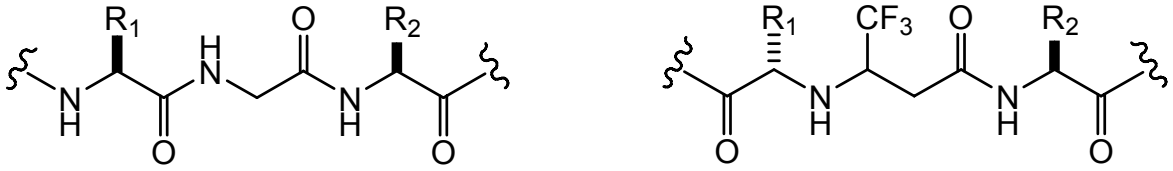


Scheme 2.8: *Gramicidin S* ***(middle)*** *and analogues containing methylalkene* ***(left)*** *and (trifluoromethyl)alkene* ***(right)*** *peptide bond isosters (according to Xiao et al.)*

Furthermore, the fluorine *gauche*-effect, which is a well known phenomenon in organofluorine chemistry,[109] can be used to affect peptide conformation in a directed manner. Since the *gauche* rotamer in a vicinal F-C-C-F moiety possesses a lower electronic energy compared to that of the corresponding *anti* rotamer, this group will induce a specific secondary structure when incorporated into a peptide backbone.[110] Therefore, this special behavior of the C-F bond should be useful in the design of peptide mimics and in influencing the conformation of bioactive molecules.

Further fluorinated peptide bond surrogates that induce the formation of specific peptide folds could be synthesized and characterized recently. Studies on peptidomimetics containing the conformational constraining trifluoromethyl group demonstrated that partially modified retro (PMR) ψ[NHCH(CF_3)]Gly peptides (Scheme 2.9) adopt stable turn-like conformations, mainly owing to the torsional restrictions caused by the stereoelectronic demand of the trifluoromethyl group.

a **b**

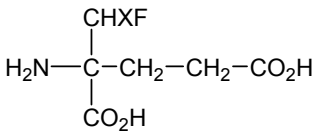


Scheme 2.9: *Standard (-Xaa-Gly-Xaa-) peptide **(a)** and partially modified retro ψ[NHCH(CF_3)]Gly peptide **(b)**, according to Zanda.*

The strong influence of backbone fluorine substitution on the conformational behavior of peptides could also be demonstrated with α-fluorinated β-amino acids. Even a single fluorine substituent in a β-heptapeptide significantly affected the folding of the molecule.[111]

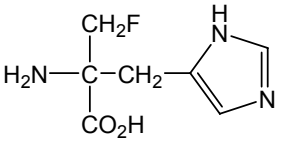
2.3.2.2 Amino Acids with Fluorinated Side Chains

Another approach towards fluorine-containing peptides and proteins uses side chain-fluorinated building blocks. Such fluorinated amino acids, especially α-fluoromethyl-substituted and trifluoromethylated derivatives, possess valuable biological activities as single molecules (Scheme 2.10). Acting as inhibitors for several enzymes, these amino acid derivatives have a significant antibacterial, cancerostatic, or cytotoxic impact and can also be useful in the treatment of disorders of the central nervous system.[112]



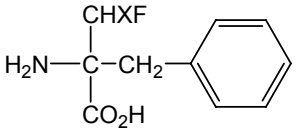
X=H α-(fluoromethyl)glutamic acid
X=F α-(difluoromethyl)glutamic acid

inhibitors of glutamate decarboxylase



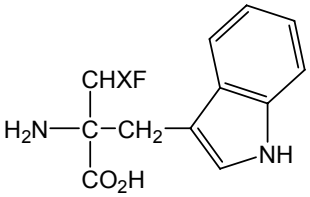
α-(fluoromethyl)histidine

inhibitors of histidine decarboxylase



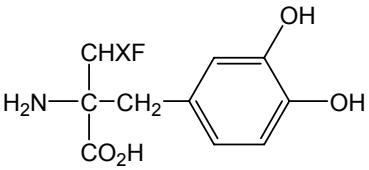
X=H α-(fluoromethyl)phenylalanine
X=F α-(difluoromethyl)phenylalanine

reaction with pyridoxal phosphate



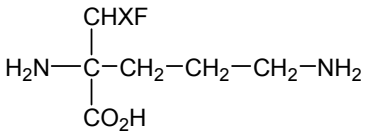
X=H α-(fluoromethyl)tryptophan
X=F α-(difluoromethyl)tryptophan

inhibitors of amino acid decarboxylase (AADC)



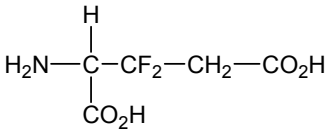
X=H α-(fluoromethyl)-DOPA
X=F α-(difluoromethyl)-DOPA

inhibitors of amino acid decarboxylase (AADC)



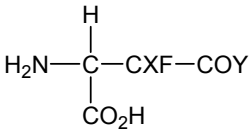
X=Cl α-(chlorofluoromethyl)ornithine
X=F α-(difluoromethyl)ornithine

inhibitors of ornithine decarboxylase



X=F β-difluoroglutamic acid

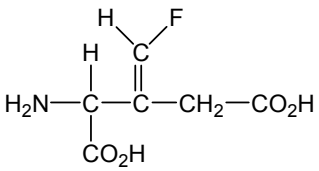
inhibitor of polyglutamylation of folic acid



X=H, Y=OH β-fluoroaspartic acid
X=F, Y=OH β-difluoroaspartic acid
X=H, Y=NH2 β-fluoroasparagine
X=F, Y=NH2 β-difluoroasparagine

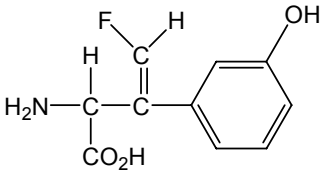
monofluorine analogues:
prodrugs for the inhibition of adenylosuccinate lyase (inhibition of purine synthesis)

asparagine analogues:
inhibition of protein glycosylation



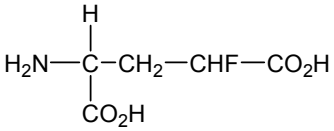
β-(fluoromethylene)glutamic acid

bioprecursor for GABA-transaminase inhibitor



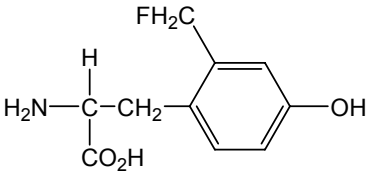
β-(fluoromethylene)-*m*-tyrosine

bioprecursor for monoamine oxidase (MAO) inhibitor



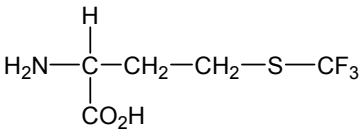
γ-fluoroglutamic acid

inhibitor of polyglutamate synthetase and glutamate mutase



o-(fluoromethyl)tyrosine

inhibitor of tyrosine hydroxylase



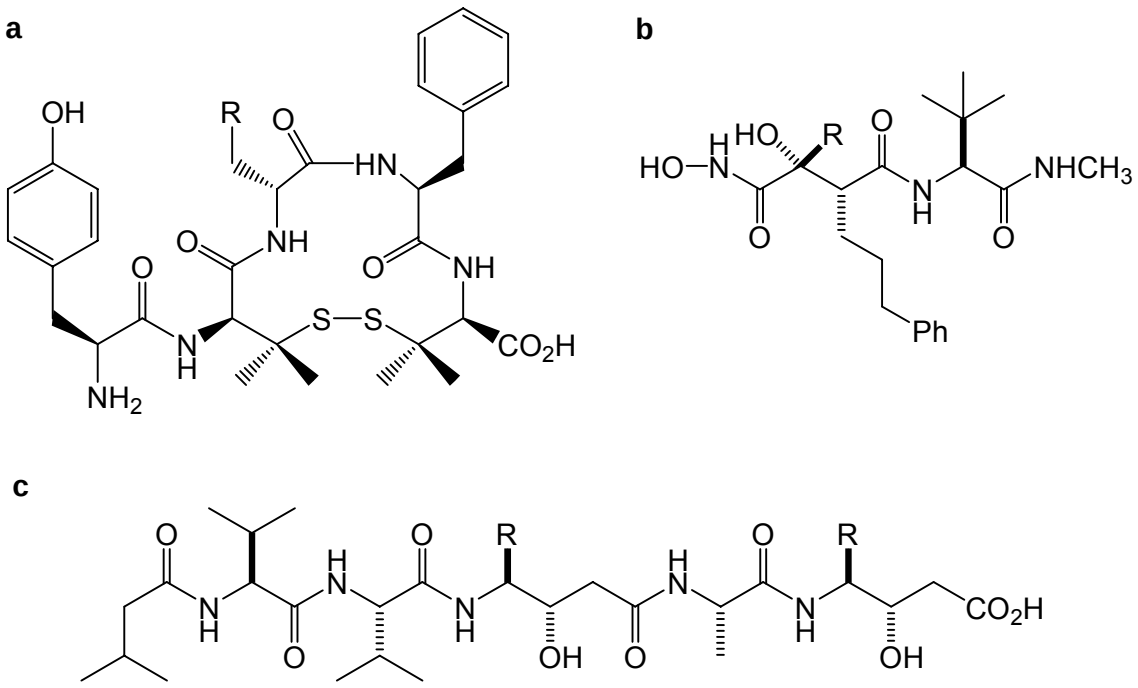
trifluoromethionine

inhibitor of S-adenosylmethionine demethylation

Scheme 2.10: *Fluorinated amino acids with enzyme inhibitory activities (according to J. T. Welch et al.).*

A lot of bioactive peptides, such as Oxytocin, Bradykinin, Physalaemin, Angiotensin II, Thyreotropin-releasing factor (TRH), opioid peptides, Substance P, Tropinin I peptide, and

Folyl-poly-γ-Glu, have been modified using fluorinated analogues of native amino acids, and their impact on the biological activity of the biopolymer was studied.[113] Furthermore, several fluorine-containing peptides have been synthesized that act as active site inhibitors of hydrolytic enzymes like serine-, aspartyl-, and metallo-proteases. The strategies mostly applied for amino acid substitution involve the replacement of a hydrogen atom or a hydroxy group in aromatic amino acids by one fluorine atom or the fluorination of alkyl groups in aliphatic amino acids. In many cases, the directed fluorination of a peptide yielded in significantly increased biological activities, receptor affinities, or inhibitory potential. A demonstrative example is provided by studies on substituted Enkephalin peptides (Scheme 2.11a). It was shown that the analogue containing 4,4-difluoro-2-aminobutyric acid is two orders of magnitude more active *in vivo* than the analogue with the unfluorinated amino acid derivative (2-aminobutyric acid).[114]



Scheme 2.11: *Fluoroalkyl-substitution can alter the biological activity of peptides.* ***a)*** *Enkephalin analogues containing 2-aminobutyric acid (R=CH_3) or 4,4-difluoro-2-aminobutyric acid (R=CF_2H) in amino acid position 3 (according to Winler et al.),* ***b)*** *potent MMP inhibitor (R=CH_3) and its Tfm-analogue (R=CF_3) and* ***c)*** *Pepstatin A (R=isobutyl) and its bis-Tfm-analogue (R=CF_3), both according to Zanda.*

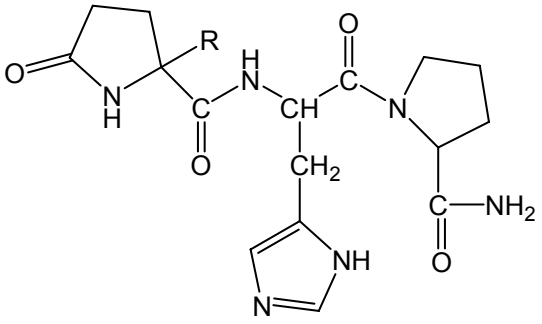
In other cases, the fluorination of amino acid side chains in peptides decreased biological activity. For example, the replacement of CH_3 by CF_3 in the quarternary α-methyl alcohol moiety of a matrix metalloproteinase inhibitor (Scheme 2.11b) led to a dramatic drop in inhibitory activity towards different matrix metalloproteinases (MMP). Here, the enhanced

steric demand of the trifluoromethyl group resulted in an unsatisfactory fit of this moiety into the binding pocket of the enzymes.

The conflictive impacts that fluorination can have on the biological activity of peptides and peptidomimetics can be used to design enzyme inhibitors with an increased selectivity. This concept could be very successfully implemented in studies on pepstatin A, a subnanomolar inhibitor of many aspartyl proteases. The analogue with *iso*-butyl groups in both statine residues substituted with trifluoromethyl groups (Scheme 2.11c) showed an only three times lower potential in inhibition of plasmepsin II, an aspartic protease from a malaria-causing protozoal. In contrast, a dramatic decrease in inhibitory activity towards human cathepsin D by the above-mentioned substitution in pepstatin A was observed. This selectivity, obtained by fluoroalkyl substitution, makes the peptide analogue very attractive for malaria therapy.

In order to systematically investigate the effect that fluoroalkyl substitution has on the bioactivity of peptides, a small library of the chemoattractant For-Met-Leu-Phe-NH_2 (fMLF), in which the leucine residue was substituted by fluorinated amino acids, was studied.[115] The amino acids used for substitution of Leu varied in the content of fluorine, the length of the fluorinated side chain, and the degree of alkylation at the α-carbon atom. Therefore, the impact of fluoroalkyl-substitution of amino acids on the binding of a peptide to a hydrophobic receptor pocket could be studied systematically. The results provide indications of the polarity of difluoromethyl groups as well as of the steric demand of trifluoromethyl groups that will be useful for the design of artificial peptide-based receptor ligands.

Since model studies showed that α-trifluoromethyl-substituted amino acids can increase the proteolytic stability of peptides by 100% when incorporated into the P_1 position (nomenclature according to Schechter and Berger[116]), these building blocks can additionally enhance the peptide's biological potential. Analogues of the Thyreotropin releasing hormone (TRH) that have a pyroglutamic acid substituted by its αTfm-derivative (Scheme 2.12) exhibited a complete resistance to proteolysis.[117]



Scheme 2.12: *Thyreotropin releasing hormone (TRH); native peptide (R=H) and proteolysis-resistant analogue [(αTfm)pGlu^1]-TRH (R=CF_3), according to Koksch et al.*

2.3.3 The Impact of Fluorine on Structural Stability and the Folding of Proteins

While fluorination of potent drug molecules and biologically relevant peptides has been a widely applied strategy for decades to improve their *in vivo* stabilities and activities, a relatively new topic deals with the impact of amino acid fluorination on the structural stability of larger polypeptides and proteins.[118] Since it could be shown that the directed incorporation of fluoroalkyl-substituted amino acids can be used to strongly enhance the thermal stability of proteins and even to program specific protein-protein interactions and folding, fluorine is undoubtedly becoming a valuable tool in protein engineering. However, this new field of application reawakened the controversies about properties of organic fluorine such as hydrophobicity, polarity, space filling of fluoroalkyl groups, and the “fluorous effect”.

2.3.3.1 The Stabilization of Collagen with Fluoroproline Residues

A polypeptide chain of collagen, the most abundant protein in animals, is composed of repeats of the amino acid sequence Xaa-Yaa-Gly and is forming a trimer of left-handed polyproline II-type helices.[119] Thereby, Xaa is often proline and Yaa is often 4*(R)*-hydroxyproline (Hyp). The high content of these conformationally constrained amino acids is indispensable for triple helix formation of collagen molecules. It could be shown that the substitution of the 4*(R)*-hydroxyl group in Hyp by fluorine (Flp) dramatically increases the stuctural stability of the collagen triple helix. While the peptide $[Pro\text{-}Hyp\text{-}Gly]_{10}$ showed a melting temperature of 69°C, the fluorinated analogue $[Pro\text{-}Flp\text{-}Gly]_{10}$ melted at 91°C.[120] Interestingly, substitution of Hyp by 4*(S)*-fluoroproline (flp) strongly destabilizes the protein fold, as could be proven in studies with $[Pro\text{-}Hyp\text{-}Gly]_7$, $[Pro\text{-}Flp\text{-}Gly]_7$, and $[Pro\text{-}flp\text{-}Gly]_7$ peptides.[121] This phenomenon is caused by stereoelectronic effects of fluorine. The electronegative fluorine-amide *gauche* effect leads to a preference for the C_γ-exo ring pucker in Flp (Figure 2.5) as well as for the *trans*- conformation of the amide bond in [ProFlpGly] (Scheme 2.13).[122]

Since the energy difference between the *cis*- and the *trans*- conformation is significantly larger for the exo-pucker than for the endo-pucker,[123] both effects of 4*(R)*-fluorine substitution lead to stabilization of the collagen helix, which contains only *trans* peptide bonds. In contrast, [Pro-flp-Gly] is preferably found with a C_γ-endo pucker of the flp-pyrrolidine ring and a *cis*- conformation of the amide bond. While the exact impact of fluorine substitution in collagen on the formation of secondary and tertiary structure is still controversial,[124] the triplehelix stabilization and destabilization, respectively, by substitution of Hyp with Flp or flp could already be proven to affect biological functions of collagen.[125]

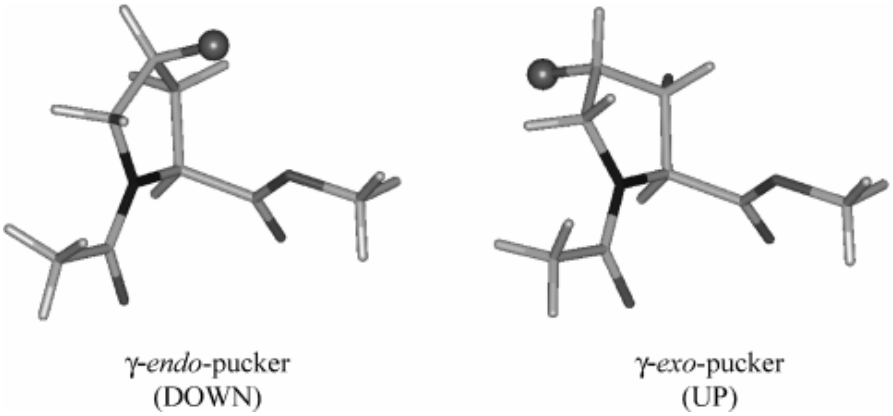
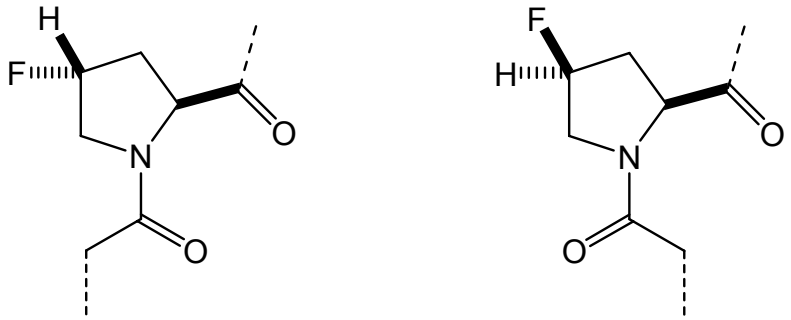


Figure 2.5: *Pucking of the pyrrolidine rings of Ac-flp-OMe* ***(left)*** *and Ac-Flp-Ome* ***(right)****, according to Barth et al.*



Scheme 2.13: *Trans prolyl peptide bond in Xaa-Flp-Xaa* ***(left)*** *and cis prolyl peptide bond in Xaa-flp-Xaa* ***(right)****, according to Bretscher et al.*

2.3.3.2 The Stabilization of Hydrophobic Protein Cores by Perfluorination

In the last five years, a few laboratories have focused their investigations on fluorinated analogues of hydrophobic canonical amino acids such as leucine, isoleucine and valine.[126] They could demonstrate that substitution of the native hydrocarbon building blocks by their fluorocarbon derivatives in hydrophobic protein domains leads to a significant enhancement of the structural stability of the entire protein. However, the authors of these studies explain this effect in different ways.

Substitutions of leucine residues in the hydrophobic interaction domain of helical peptide dimers, that are based on the *coiled coil* region of the yeast transcription factor GCN4, by 5,5,5-trifluoroleucines[127,128] (Figure 2.6) and 5,5,5,5',5',5'-hexafluoroleucines[129] resulted in higher thermal and chemical stabilities of the proteins.

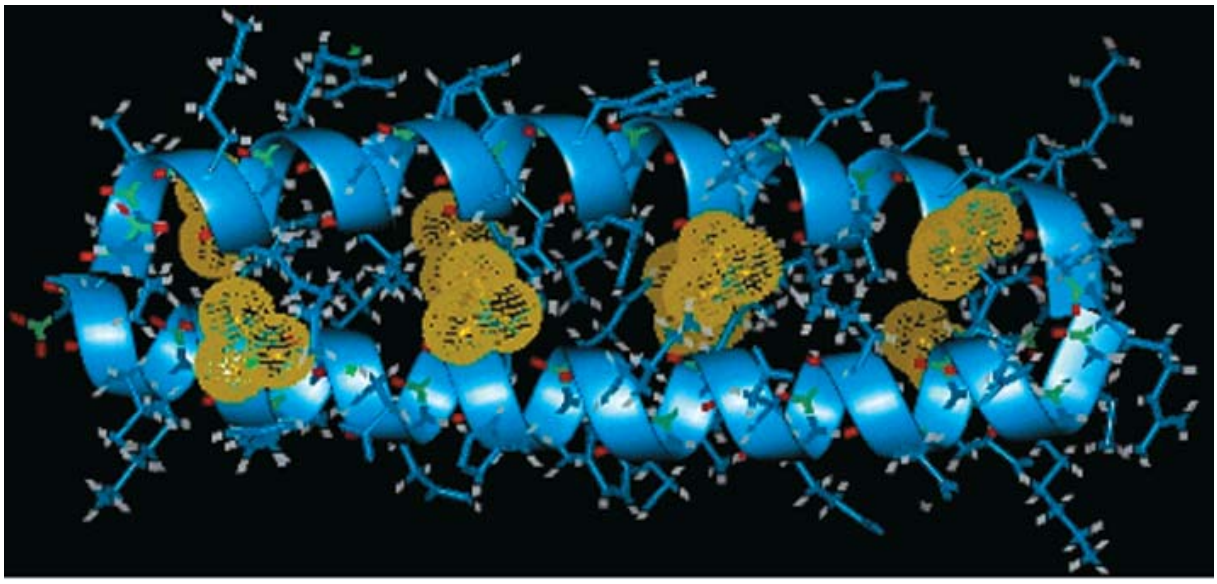


Figure 2.6: *Three-dimensional representation of the coiled coil region of GCN4 substituted with 5,5,5-trifluoroleucine at the four **d** positions in each helix strand. The van der Waals radii of the fluorine atoms are shown as yellow spheres (according to Tang et al.)*

Furthermore, fluorine-substituted full length GCN4 proteins revealed DNA-binding affinities and specificities that were identical to those of the native transcription factor. Further studies with different rates of fluorine substitution in the hydrophobic dimerization domain of GCN4 proved that the stabilizing effect correlates with the content of fluoroalkylated amino acids.[130] Due to the parallel orientation of the peptides, the fluorocarbon side chains generally interact with each other at the dimer interface. Interaction with native Leu residues is prevented, though. To evaluate the nature of the observed effect, self-sorting experiments with parallel oriented *coiled coil* peptides that are formed by one helix strand with a fluoroleucine core and one strand possessing a leucine core (heterodimer) have been performed.[131] The disulfide-bound dimers were allowed to undergo disulfide exchange under appropriate redox conditions (Figure 2.7). Within the equilibrated solution, homodimer formation was preferred over heterodimer formation by at least 26-fold. Furthermore, hydrocarbon-fluorocarbon core heterodimers showed no increased stability compared to pure hydrocarbon homodimers.[132] These results support the hypothesis of the "fluorous effect". This means that the structural stabilization of hydrophobic protein cores upon perfluorination is based on fluorine-fluorine interactions rather than on a general increase in hydrophobicity of the fluorinated amino acid side chains. Such "fluorous effects" are well known in organic chemistry, e.g., for the formation of fluorous phases in phase separation techniques. Perfluorinated tags were recently used in solid phase peptide chemistry for the effective purification of the desired peptide chain applying a highly fluorinated solvent.[133]

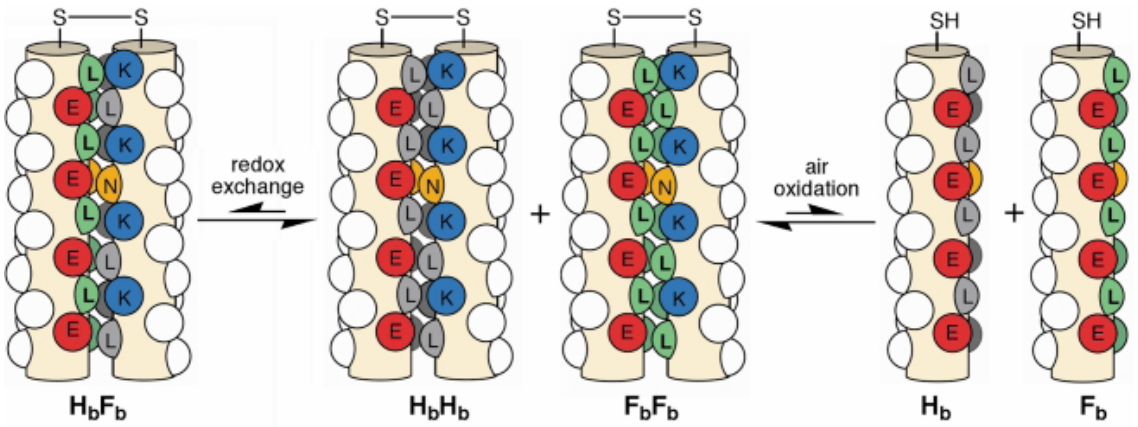


Figure 2.7: *Self-sorting of coiled coil peptides with fluorocarbon cores (F_b) and hydrocarbon cores (H_b), respectively, in disulfide exchange experiments (according to Bilgiçer et al.)*

Studies with fluoroalkyl-substituted peptides in membrane environments also support the theory of "fluorous phase" formation in protein interfaces. Substitutions of leucine residues by 5,5,5-trifluoroleucine in melittin, a 26 amino acid residue amphiphilic peptide from bee venom, resulted in an increased self-association of the peptide in aqueous medium.[134] Furthermore, an enhanced membrane affinity could be observed that was attributed by the authors to fluorocarbon-hydrocarbon separation due to peptide aggregation in the lipid layer. Such simultaneous hydrophobic and lipophobic character of fluorinated peptide surfaces was described by studies about the oligomerization behavior of highly apolar helical peptides containing hexafluoroleucines in membrane-like hydrophobic environments.[135] The solvation of the peptides in micelles initiated helical folding and, thus, the formation of fluoroalkyl helix faces. Fluorous core formation of these surfaces initiated peptide dimerization within the hydrophobic environment (Figure 2.8).

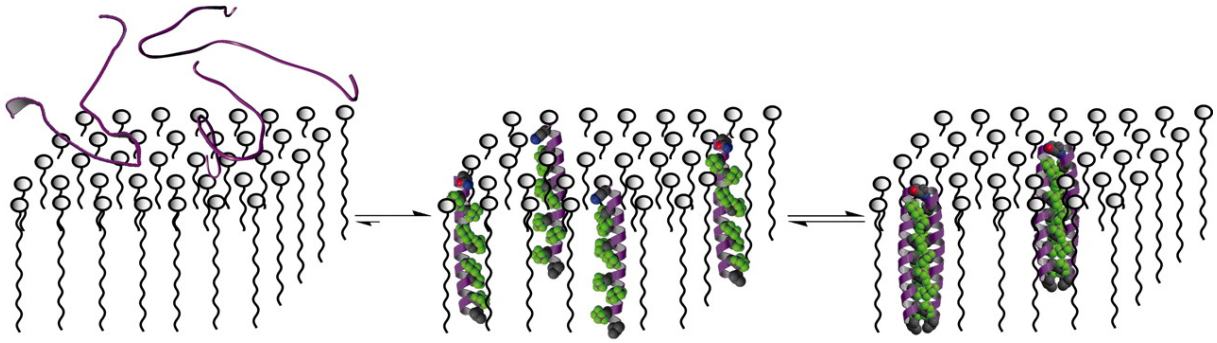


Figure 2.8: *Self-assembly of membrane soluble protein segments in micelles. Dimerization takes place via fluorous core formation of helix faces containing hexafluoroleucine residues (green), according to Bilgiçer et al.*

In addition, it could be shown that 4,4,4-trifluoroethylglycines as hydrophobic residues in amphiphilic helices tend to interact with each other.[136] These results demonstrate that fluorocarbon-hydrocarbon phase separation can direct protein folding.

In contrast to these studies, results of other laboratories suggest that the increase in structural stability of hydrophobic protein domains upon substitution of hydrocarbon groups by their fluorocarbon analogues is caused by an enhancement in hydrophobicity of the substituted side chains. Two variants of an 56-residue α-β globular protein, each modified by a single valine replacement to a 4,4,4-trifluorovaline within largely buried amino acid positions, revealed significant stabilization of the overall structure as well as faster folding and slower unfolding processes compared to the unsubstituted protein.[137] Since no fluorous interactions can occur in either of the variants with single-residue substitutions, the observed structural impact has been ascribed to stabilized hydrophobic interactions.
This interpretation is supported by experiments with antiparallel oriented tetrameric, 27-residue α-helical *coiled coil* peptides that have the central two out of six layers of the leucine-containing hydrophobic interface repacked with 5,5,5,5',5',5'-hexafluoroleucine (figure 2.9). The observed increase in structural stability, relative to the non-fluorinated analogue, has been attributed to the more hydrophobic nature of the hexafluoroleucine rather than to specific fluorous interactions.[138]

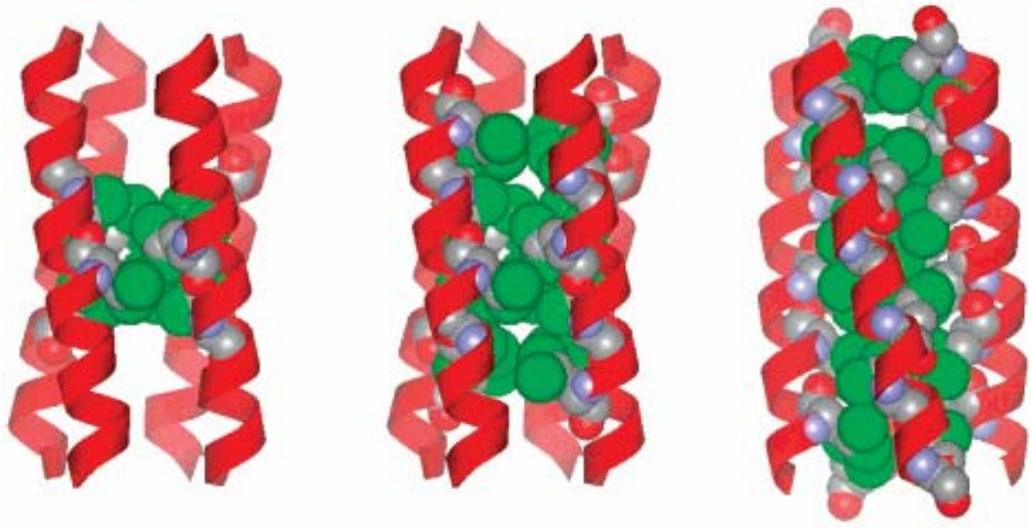


Figure 2.9: *Models of antiparallel α-helical coiled coil peptide tetramers with two **(left)**, four **(middle)**, and all six **(right)** layers of the leucine-containing hydrophobic interface substituted with 5,5,5,5',5',5'-hexafluoroleucine. The trifluoromethyl groups are represented by green spheres (according to Lee et al.)*

The greater hydrophobicity of hexafluoroleucine, compared to the canonical leucine, has been quantified by partitioning experiments from *n*-heptanol into water. The resulting $\Delta\Delta G$ values were in good agreement with the additional stability each fluorinated leucine analogue had contributed to the modified peptide tetramer. In further studies, four and all six hydrophobic layers, respectively, were fluoro-substituted and a structure stabilization that correlates with the extent of the fluorous core was observed.[139] Interestingly, NMR investigations revealed an increase in rigidity of the side chains within the hydrophobic interaction core upon perfluorination, which might be due to the controversially discussed advanced space filling of the trifluoromethyl groups.

Recapitulating, most of the studies with fluoroalkyl-substituted proteins suggested that fluorocarbon-fluorocarbon interactions are much stronger than both fluorocarbon-hydrocarbon and hydrocarbon-hydrocarbon interactions. This “fluorous effect” as well as the unique stereoelectronic properties of organic fluorine could be proven to stabilize protein structure and dictate folding. Thus, fluorinated amino acids can be powerful tools in protein chemistry as well as in the design of bioactive peptides that act as enzyme inhibitors or receptor ligands.

Protein design using fluorinated amino acids is a relatively new but strongly emerging area. The goal is to be able to apply fluorinated amino acids for the rational design of protein structures and peptide-protein interaction sites as well as for controling protein folding. That means, specifically designed fluorinated building blocks that exert favourable and predictable influences on protein structure and folding as well as on binding of peptides to proteins are used in the design of appropriate amino acid side chain interactions.

However, as long as physicochemical properties of fluoroalkyl groups as well as their interaction behavior within native protein environments are controversially discussed and no consistent concepts have been established, this ambition remains a challenge to medicinal and protein chemists.