Organic & Biomolecular Chemistry

Cite this: *Org. Biomol. Chem.*, 2012, **10**, 7241

www.rsc.org/obc

EMERGING AREA

Organic fluorine as a polypeptide building element: *in vivo* expression of fluorinated peptides, proteins and proteomes

L. Merkel and N. Budisa*

Received 15th November 2011, Accepted 9th July 2012
DOI: 10.1039/c2ob06922a

Traditionally, the biological fluorination of complex biological systems like proteins is achieved through substitution of canonical amino acids or addition of fluorinated amino acids in the context of the standard genetic code. Ribosomal translation of monofluorinated amino acids into proteins often yields structures with minimal local changes in the interior but, on the same time, results in large global effects on characteristic features of the biopolymers (such as dramatically changed activity profile or folding stability). This is due to the novel and unique local interactions delivered by fluorine atoms such as (i) increase in the covalent radii (ii) changed polarities; (iii) changed hydrogen bond acceptor ability; (iv) altered water solubility as well as water ↔ organic solvent energy transfer. On the other hand, the biological incorporation of tri- or global fluorinated amino acids (such as trifluoroleucine, triflurovaline, and their hexafluoro counterparts, fluoromethionine and trifluoronorleucine *etc.*) represents still a challenge, as the natural structural scaffolds are optimized for hydrocarbon during evolution but not for fluorocarbon cores. Future work will be focused on the re-design of existing or *de novo* design of novel protein scaffolds capable of accommodating such building blocks into functional biologically active proteins and proteomes in the context of the viable cells.

Introduction

Fluorinated organic compounds and polymers are interesting target molecules, in biological and medicinal chemistry, material sciences, agrochemicals and crop sciences. They also possess unique features highly attractive not only for synthetic applications but also for the pharmaceutical industry as fluorination improve the bioavailability of drugs.[1,2] This is due to the fact that the presence of fluorine atoms or fluorinated groups in a particular molecule is associated with the unusual and unique physicochemical properties, which cannot be achieved by any other element.[1,3]

Fluorine is the most abundant halogen in the Earth's crust.[6] It appears in levels comparable to that of nitrogen; however, only 12 organic compounds are known to occur in nature whereas the majority of fluorine is contained within inorganic substances.[1,2,7] Conversely, synthetic organic chemistry provided us with a large number of compounds in which hydrogen atoms are partially or completely replaced by fluorine atoms. High-level or even complete substitution of hydrocarbons, yields fluorocarbons characterized with lower boiling points and water solubility, and an increased molar volume. On the other hand, many synthetic fluorine containing organic compounds of anthropogenic origin are fluorinated to a lesser extent. For example, monofluorinated analogs of aromatic amino acids have often been used as substitutes in protein biosynthesis.

Certainly, the interest in synthesizing fluorinated biological molecules such as proteins can be explained by the potent electronic effects of this halogen, particularly on nearby functional groups. For example, due to fluorine's very high ionization potential and low polarizability – fluorinated molecules exhibit only weak intermolecular interactions, *i.e.* highly fluorinated compounds are characterized by weak surface energies, dielectric constants and refractive indices. In addition, fluorine has the highest electronegativity, and, thus, fluorine–carbon bonds are strongly polarized with the positive partial charge at the carbon atom and the related negative partial charge at the fluorine atom. Therefore, C–F bonds possess stronger energies than the other C–Hal bonds and possess a high bond dissociation energy.[3] On the other hand, the hydrophobic character of a molecule can be increased by the introduction of fluorine atoms.[1,2]

In this review we aim to present the most recent advances in the field of protein biosynthesis using fluorinated amino acids along with brief references to the benchmarks in related peptide studies and the historical developments. Thus, our focus is on *in vivo* synthesized proteins and the interested reader may consult the literature for further details of the results in related fields.[4,5]

TU Berlin, Department of Chemistry, Biocatalysis Group, Franklinstrasse 29 (OE1), D-10587 Berlin, Germany. E-mail: budisa@biocat.tu-berlin.de

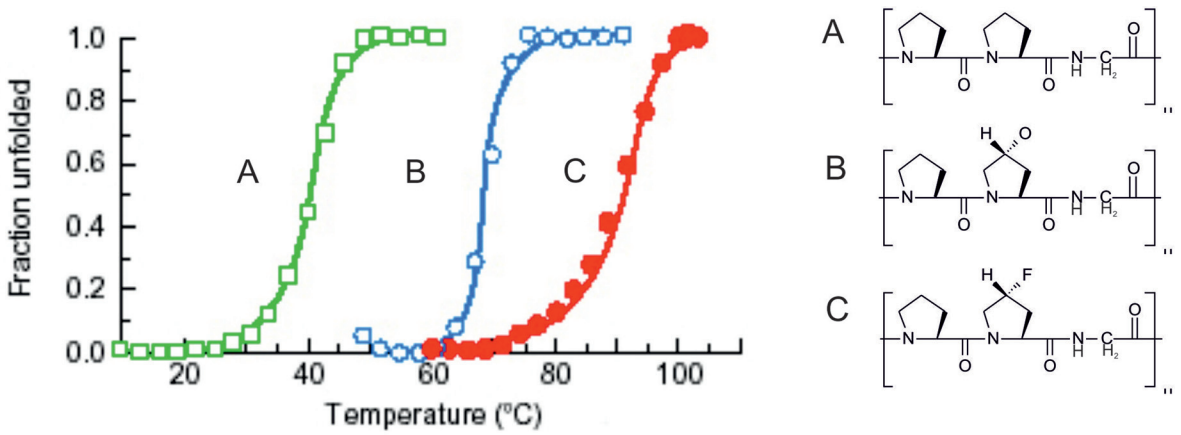


Fig. 1 Unmodified, hydroxylated (natural) and fluorinated (hyperstable) collagen peptides measured in 50 mM acetic acid. Unmodified (Pro-Pro-Gly)$_{10}$ collagen (**A**, Y = Pro) exhibits the lowest melting point (T_m = 41 °C) while the 'natural' (Pro-Hyp-Gly)$_{10}$ sequence (**B**, Y = Hyp) is much more stable (T_m = 69 °C) and, moreover, the fluorine-containing (Pro-(4*R*-F)Pro-Gly)$_{10}$ triple helix (**C**, Y = (4*R*-F)Pro) is hyperstable (T_m = 91 °C). The *trans*-hydroxylation of Pro(Y) leads to an increase in T_m of almost 30 °C whereas its *trans*-fluorination elevates melting temperature for almost 50 °C. This is indeed plausible, since fluorine exerts a strong inductive effect that is sufficient to affect properties such as polarity or binding capacity of functional groups that are even distantly positioned in the structure. The figure is adopted from ref 8.

Effects of fluorine incorporation into peptides and small proteins

As a result of numerous studies showing that fluorinated amino acids can increase the stability of natural polypeptides, fluorine has become of great interest in peptide and protein chemistry. For example, fluorinated derivatives could be used to investigate the factors contributing to the extraordinary stability of collagen.[8] Collagen is the most abundant protein in animals as it is part of connective tissues like bone, cartilage, ligament, skin and teeth. The most characteristic features of this biopolymer are its triple helical structure and its repetitive amino acid sequence which is mostly composed of a Gly-Xaa-Yaa pattern with Xaa and Yaa often either Pro or Hyp ((4*R*)-hydroxyproline, (4*R*-OH) Pro). It was believed that the hydroxyl groups induce the conformational stability of the triple helix by interstrand hydrogen bonds mediated by bridging water molecules. In order to prove this assumption, Hyp was replaced by (4*R*-F)Pro yielding a tremendously increased triple-helical stability of the model peptide compared to its hydroxylated counterpart (T_m of (Pro-Pro-Gly)$_{10}$: 41 °C, (Pro-Hyp-Gly)$_{10}$: 69 °C, (Pro-(4*R*-F)Pro-Gly)$_{10}$: 91 °C) (Fig. 1). This means, bridging water molecules contribute little to collagen's stability and inductive effects are much more important. Additional explanations for the hyperstable triple helix include the assumption of preorganization in terms of *cis*/*trans* isomerization (all peptide bonds in collagen are in *trans* and (4*R*-F)Pro favors *trans* conformation[9]), dipole interactions and the *gauche* effect, as well as effects on the solvent-inaccessible interstrand hydrogen bonds.[6]

The majority of the studies on "fluorous proteins" were conducted using peptides mainly synthesized chemically, such as fluorous bio-active peptides or fluorous antimicrobial peptides.[10] For example, Marsh and associates extensively studied the thermodynamics of fluorinated peptides in comparison with their non-fluorinated counterparts. In particular, by using synthesized peptides containing hexafluoroleucine (Hfl, (5,5′-F$_6$)Leu) they found that fluorination results in an increased thermodynamic stability.[10–12] The authors assumed a higher thermal stability and an increased self-segregation of the peptides into antiparallel 4-helix bundles (Fig. 2) due to the fluorous effect between the fluorinated amino acids based on the results of small fluorocarbon molecules (*vide infra*). However, they found that the effect of Hfl on stability strongly depends on the position of incorporation even if all fluorinated peptides were more stable than the ones without any fluorine. Indeed, the most stable peptides in terms of per-residue stability are these with Hfl in all 'a' or all 'd' positions of the helical peptide (Fig. 2), *e.g.* with an alternating Leu-Hfl arrangement within the core. The sequence with an entirely Hfl packed core had a lower per-residue stability but was the most stable peptide in terms of absolute values. Thus, efficient packing of the fluorous amino acids within the hydrophobic core is more important than the contribution of fluorocarbon-fluorocarbon interactions between fluorinated side-chains. This is an unexpected result, as the fluorous effect would predict that the fluorinated side-chains should be close together in order to interact. There are also examples in the literature which report that fluorination of amino acids leads to a decreased propensity towards helix-formation.[13] However, the arrangement with Hfl in either position 'a' or 'd' provides the maximum dispersion of fluorinated side-chains throughout the structure. In addition, the packing of the peptides was also investigated by studying the generation of homotetramers and heterotetramers. In line with the above mentioned results, the peptides preferred to form homotetramers. Furthermore, the α-helix bundle peptides are more resistant to proteolysis by trypsin and chymotrypsin using Hfl as building block.[14] It is also worth to note that they resist unfolding by organic solvents better than their hydrocarbon parent peptides. Similar results were also obtained for peptides containing amino acids with only one trifluoromethyl group like trifluoroleucine (Tfl, (5-F$_3$)Leu) or trifluorovaline ((3-CF$_3$)Val). Both amino acids were incorporated into peptides similar to parts of the transcription factor GCN4 or NTL9. In all cases a stabilization of the peptides in terms of melting temperatures could be achieved.[15–17]

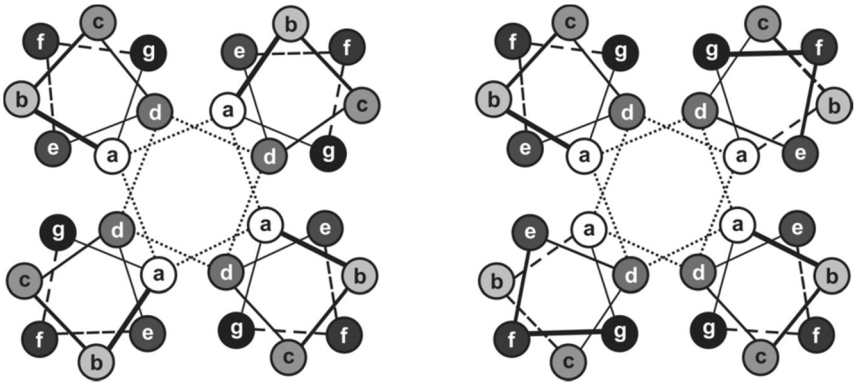
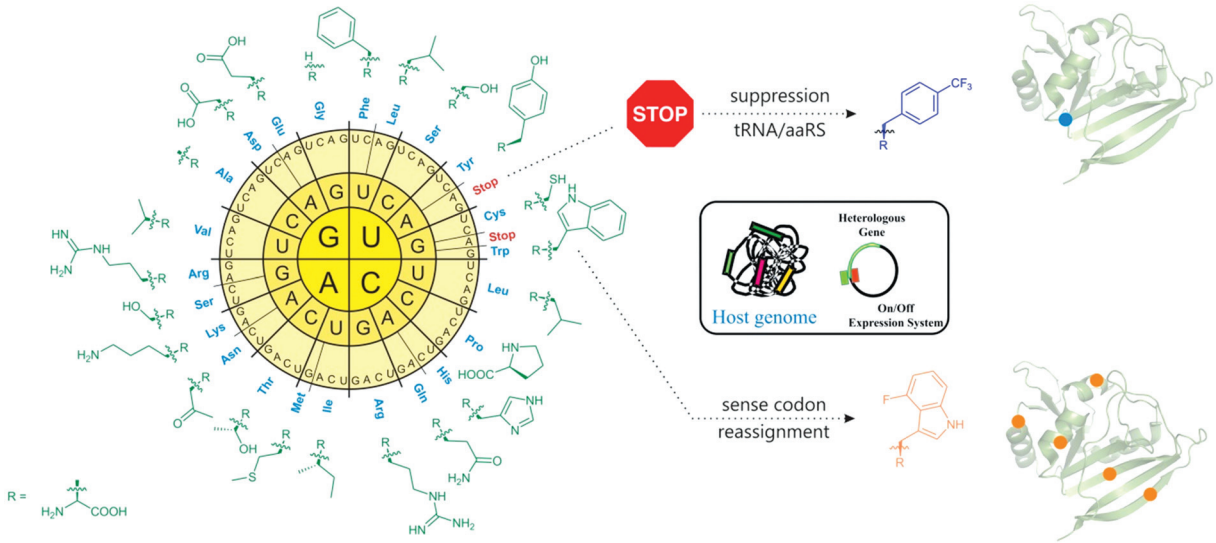


Fig. 2 Helical wheel representation of 4-helix bundles arranged parallel (left panel) and antiparallel (right panel), respectively. Note, that in addition to the differences in 'a' and 'd' interactions also the interactions between the other residues change (from 'g'–'e'-interaction to 'g'–'c'– and 'b'–'e'-interactions).

However, even though Marsh and co-workers revealed that repacking of the hydrophobic core with Hfl in a series of progressively more fluorinated four α-helix bundle peptides results in an increase in free energy of folding, they assume that the six fluorine atoms in Hfl are not sufficient to induce the so-called fluorous effect (*vide infra*).[10] They claim that the stabilizing effects detected originate from modulated hydrophobic interactions due to the use of fluorinated building blocks. However, facing the enhanced stability they ask the question, "How stable would a real fluorous peptide be?" Most recently, the authors solved X-ray structures of these peptides and their analysis supported the former findings that the stability increase could be also explained by an increase in buried hydrophobic surface area that accompanies fluorination rather than by specific fluorous interactions between fluorinated side chains.[171] In fact even extensive fluorination of smaller peptides could preserve the shape of the related side chains which is crucial for the correct packing of side chains within the hydrophobic interior, while at the same time delivering an increased size and hydrophobicity. However, such preservation is difficult in complex proteins and would most probably compromise the functional integrity of the protein.[145]

Other very interesting folding studies examine the effect of fluorine containing amino acids which are not derived from hydrogen → fluorine substitutions of canonical amino acids.[18,19] Koksch and co-workers synthesized a 41 amino acid residue containing α-helical coiled-coil peptide that folds into an antiparallel homodimer. Positions Lys8 and Leu9 were occupied by either (CH_3CH_2)Gly (Abu), (CF_2HCH_2)Gly, (CF_3CH_2)Gly or $(CH_3CF_2CH_2)$Gly. Ala was used as control in position 8 and 9 as well. The fluorinated amino acids were chosen due to the widely accepted assumption that a trifluoromethyl group exhibits approximately the same size as an isopropyl group[20] and, thus, (CF_3CH_2)Gly resembles Leu. Interestingly, all substitutions at position 9 resulted in destabilized peptides as reflected by 15–22 K decreased melting temperatures. However, increasing the side-chain volume by fluorinated analogs caused increased melting temperatures. Whereas the difference in T_m is only 0.4 K for the peptides containing either (CH_3CH_2)Gly or (CF_2HCH_2)-Gly, the difference is more pronounced (5 K) in case of the peptide consisting of (CF_3CH_2)Gly. It is worth mentioning that this shift is induced by a single additional H → F substitution. However, an enlargement of the side-chain with a methyl group as in $(CH_3CF_2CH_2)$Gly resulted in a less stable peptide compared to the (CF_3CH_2)Gly carrying peptide. Thus, the authors conclude that destabilization increases along with the number of alkyl hydrogen atoms polarized by fluorine. In addition, alkyl fluorination induces electrostatic consequences by polarization which affects the hydrophobic interactions to a greater extent than the increased molecular volume perturbs the structural integrity of the peptides. The substitutions of position Lys8 have less pronounced effects on the stability. Here, the T_m is decreased by only 2–5.6 K. This is due to the loss of one salt bridge upon substitution. In this context, it should also be noted, that fluorination of the solvent-exposed residue causes an additional loss in stability. The authors hypothesized that fluorine–fluorine interactions in the unfolded state have an impact on the monomer–dimer equilibrium and performed self-replication studies *via* native chemical ligation in the replicase reaction. The observation that the reaction rate decreases in correlation to an increased number of fluorine atoms in the side-chain might be seen as an inhibition of the reaction cycle by F–F interactions and prove that fluorine prefers to interact with fluorine in nature-like protein environments. The studies on this model system were continued and folding kinetics examined.[21] Here, it could be shown that fluorination has its strongest impact on the association of the coiled-coil peptides.

Another study investigated the stability of peptides by substituting all phenylalanine residues in a subdomain of chicken villin headpiece (VHP, 35 residues, discrete tertiary structure) with pentafluorophenylalanine ((F_5)Phe).[22] This amino acid was chosen as there are several reports regarding the effects of fluorinated saturated carbon atoms but only a few reports of the effects of aromatic amino acids and their impact on stability. In order to assay the conformational stability they performed a backbone thioester exchange because this process can be easily monitored. Model peptides containing three Phe residues were synthesized on solid support in all possible combinations of (F_5)Phe → Phe substitution. Even though every variant adopted a stable tertiary structure, only one out of seven fluorine containing species displayed a greater conformational stability compared to the fluorine free peptide, whereas the other six peptides were less



Scheme 1 The genetic code including the structures of the canonical amino acids (left panel). In the right panel the two approaches for the incorporation of non-canonical amino acids are outlined. A site-specific incorporation of *e.g.* 4-trifluoromethyl-phenylalanine is possible by suppression of an in-frame stop codon by a special tRNA which is charged by a cognate aminoacyl-tRNA synthetase (aaRS). On the other hand, auxotrophic host cell provide a possibility to incorporate non-canonical amino acids residue-specific. Further details, advantages and disadvantages of both methods are described in the text.

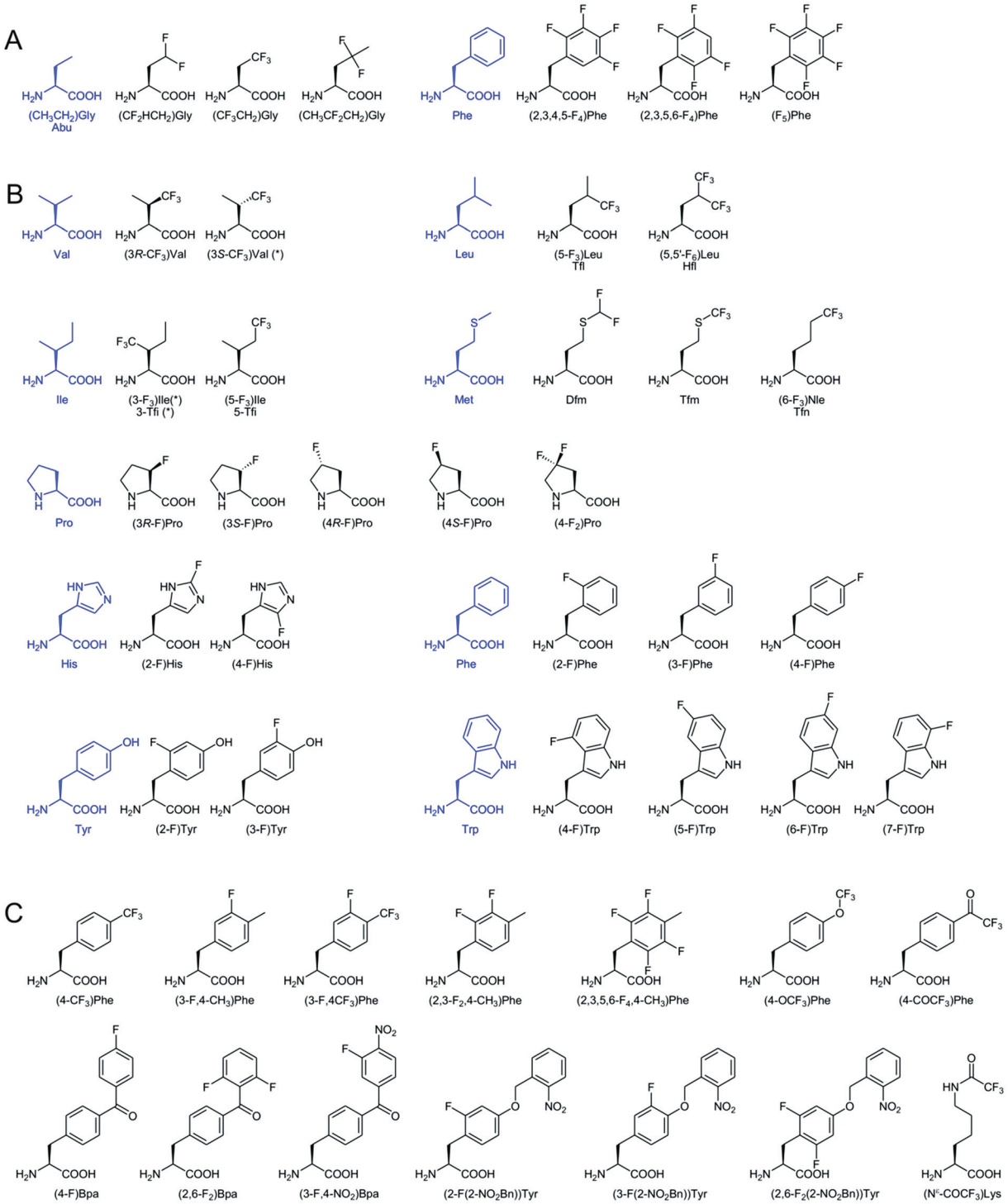
stable. The reason for that might be found in steric repulsion due to a slightly increased side-chain volume, and/or from minor conformational rearrangements. However, the helical propensity of (F_5)Phe is lower compared to Phe.[13] In contrast, Zheng and co-workers could show that one regioisomer of (F_4)Phe, namely (2,3,4,5-F_4)Phe, at one out of three Phe positions stabilizes the VHP peptide because of restored ArH···π interactions.[23] The four fluorine atoms render the remaining ArH on the phenyl ring more electron-deficient in comparison to the unsubstituted phenyl ring. This may leads to stronger ArH···π interactions in case of (2,3,4,5-F_4)Phe. However, for the same reasons ArH···π interactions should be weakened if the π cloud is provided by the fluorinated phenyl ring.

Biological incorporation of fluorinated amino acids into peptides and proteins

The engineering of the genetic code has recently became an important discipline in chemical biology as it provides an efficient platform for the transfer of numerous synthetic chemistries from the laboratory to the biochemistry of living cells. This method also allows us to add fluorinated non-canonical amino acids (NCAAs) to the existing repertoire of the 20 canonical amino acids prescribed by the genetic code. Currently, two approaches exist to ribosomally synthesize proteins using non-canonical amino acids such as fluorinated analogs (Scheme 1 and Scheme 2).[24,25] The first one, genetic code engineering, makes use of amino acid auxotrophic cell strains and can be historically traced back to the experiments of Cowie and Cohen.[26] Whereas the experimental set-up is rather simple and very high protein yields can be isolated following proper protocols, the drawback of this method is that only NCAAs which are sufficiently similar to the original canonical amino acid can be incorporated. However, this approach suits very well for the incorporation of fluorinated amino acids, as the H → F substitution can be considered as non-perturbing, as all endogenous enzymes of the translational apparatus of cells accept almost all monofluorinated amino acids. In addition, this method allows for studying synergistic effects of multiple-site incorporation[27] as well as for the efficient incorporation of different NCAAs simultaneously.[28,29] Due to the advantages of the method, most experiments related to the incorporation of fluorinated amino acids were performed using this approach.[24,25,30–33] The second approach, genetic code expansion, has the major advantage of being site-specific. This is achieved by manipulation of the genetic template in terms of inserting an in-frame stop codon into the coding sequence. This stop codon can be suppressed by specific suppressor tRNAs charged with the desired NCAA either chemically or enzymatically. The drawbacks of this method are the time-consuming screening for effective orthogonal tRNA/aaRS pairs allowing for efficient suppression and incorporation, and the limitation of the number of in-frame stop codons to be effectively suppressed as well as the competition with release factors.[34]

The auxotrophy-based approach

The history of the incorporation of NCAAs into proteins can be traced back to the year 1951 in which Levine and Tarver reported the incorporation of [^{14}C]ethionine into tissue proteins.[35] Six years later, it was Cowie and Cohen who provided in their landmark experiment the first demonstration of a proteome-wide selenomethionine (SeMet) insertion in *E. coli*.[26] To date, SeMet is still the only analog known with such a great substitution capacity. Regarding fluorinated amino acids, it was in 1959 when Munier and Cohen reported the first incorporation of fluorophenylalanine into bacterial proteins *in vivo*.[36] Only two years later it could be shown that mammalian proteins can be



Scheme 2 Survey of fluorinated non-canonical amino acids mentioned in this review. The blue formulas represent the hydrocarbon or canonical counterparts to be substituted and used as controls. (A) Amino acids used only in peptide studies to date, (B) amino acids used in peptide studies and predominantly in genetic code engineering and (C) amino acids used in peptide studies and in genetic code expansion. Amino acids marked with an asterisk could not be incorporated to date *in vivo*.

fluorinated by feeding rabbits with 4-fluorophenylalanine ((4-F)-Phe).[37] This approach was later used in order to perform NMR studies.[38] In 1962, Richmond summarized for the first time the results of NCAA incorporation into proteins in his review entitled 'The Effect of Amino Acid Analogues on Growth and Protein Synthesis in Microorganisms'.[39] In the same year, Chapeville and co-workers reported the first results of amino acid incorporation by using misacylated tRNAs in ribosomal protein synthesis.[40] By supplementing growth medium with Tfl, Rennert and Anker showed one year later that an exponential

growth can be detected and the fluorinated amino acid is incorporated into proteins.[41] At the same time, Conway and co-workers studied the effect of fluorinated Phe analogs on the intracellular accumulation of Phe and its incorporation.[42] In the mid-1970s, the first reports appeared exploring the suitability of fluorinated protein congeners for ^{19}F NMR structural analysis. Sykes and co-workers studied the alkaline-phosphatase equipped with fluorotryptophans,[43] whereas Pratt and Ho examined D-lactate dehydrogenase and lactose permease,[44] also with incorporated fluorotryptophans. Here, up to two doublings could be detected during cell growth but most congeners – non-canonical amino acid bearing proteins – exhibited a reduced activity. However, 4-fluorotryptophan ((4-F)Trp) in D-lactate dehydrogenase showed an enhanced activity which is the first report of a positive effect of fluorinated amino acids on the activity of an enzyme. In the following years, fluorinated amino acids became a widely used tool in structural NMR analysis of proteins.[45–51]

Also in the 1970s, the first experiments using 2-fluorohistidine ((2-F)His) for protein biosynthesis were reported. Its biological effects and its incorporation into (mammalian) proteins were studied.[52–54] (2-F)His is an interesting amino acid as its pK_a is lowered by 4.8 units compared to His and, thus, its incorporation may affect catalysis.[54] Taylor and co-workers published a crystal structure of the semisynthetic ribonuclease-S with incorporated 4-fluorohistidine ((4-F)His) in 1981.[55] However, the enzyme was inactivated due to the altered basicity of the catalytic active His12.

Another benchmark experiment of the 1980s was performed by Wong who showed that by serial dilution experiments, microorganisms can be generated which have a stronger preference to grow on (4-F)Trp than on Trp itself.[56] Similar experiments were attempted by Bacher and Ellington some years later (*vide infra*) but so far without success.[57,58] In 1988, Bronskill and Wong demonstrated that (4-F)Trp could not only be used in ^{19}F NMR analysis but also facilitates the identification of the contribution of chromophores, like Tyr or Phe, to the spectroscopic features of a protein as (4-F)Trp itself is a non-fluorescent Trp analog.[59] In addition, fluorinated Trp analogs have in general been commonly used to perform spectroscopic and unfolding studies (*vide infra* and *e.g.* ref. 60 and 61). In 1997, Duewel and co-workers confirmed the *in vivo* incorporation of trifluoromethionine (Tfm) into a phage lysozyme.[62] They characterized for the first time an enzyme with Tfm as the reports of 1975 only provided indirect evidence for Tfm incorporation into TCA-insoluble protein fractions from *S. cerevisiae*.[63]

Our group also became interested in fluorinated proteins by the end of the 1990s. One of the first experiments was the quantitative incorporation of either (4-F)Trp, (5-F)Trp, or (6-F)Trp into human annexin V (anxA5).[64] As judged from X-ray analysis, the congeners had the same secondary and tertiary structure as the parent protein. However, the congeners differed in stability, folding cooperativity, biological activity and fluorescence properties compared to the parent protein. As expected, the (4-F)-Trp congener exhibited no fluorescence, whereas the (5-F)Trp and (6-F)Trp congener exhibited similar emission profiles as the parent protein but red-shifted by 13 and 16 nm, respectively. Comparison of the T_m values revealed that anxA5[(5-F)Trp] is more stable compared to the parent protein and the other two

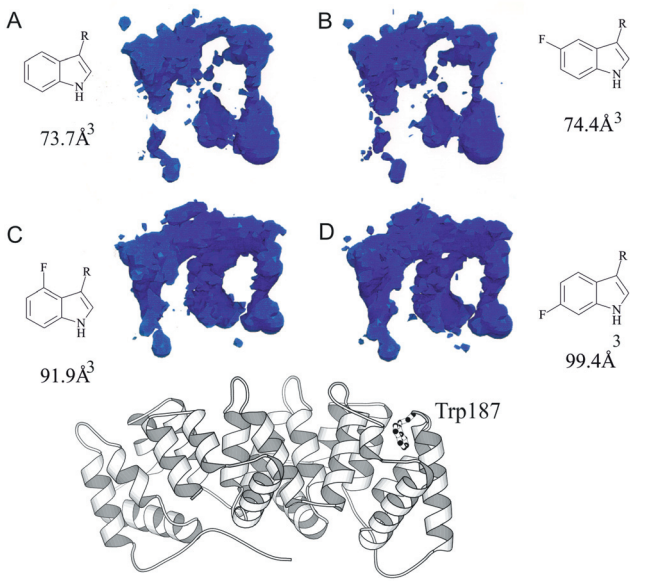


Fig. 3 Packing of fluorinated Trp187 in human annexin V (anxA5) revealed by three-dimensional modeling of the gap regions in Trp's cavity. Ribbon plot of the three-dimensional structure of anxA5 with single Trp187 marked as balls-and-sticks (lower part) along with the calculated gap areas marked in blue (upper part). (A) Native Trp187 occupies 73.7 Å^3 of the space (excluded volume) in its hydrophobic niche; (B) In the anxA5[(5-F)Trp] there is no significant change in the excluded volume in comparison to the native counterpart; (C) (4-F)Trp insertion into anxA5 is accompanied with an increase of the excluded volume of approximately $\frac{1}{4}$ (91.9 Å^3) in comparison to the native protein. D) anxA5[(6-F)Trp] has also a significantly increased excluded volume in comparison to the native form and similar to that of the 4-FTrp substituted protein.[64.]

congeners are less stable (*vide infra*). The biological activity assays showed that the (6-F)Trp congener is threefold more active compared to the parent protein. In contrast, incorporation of (4-F)Trp or (5-F)Trp decreased the activity three- and twofold, respectively. In addition, it is important to mention that anxA5[(5-F)Trp] and anxA5[(6-F)Trp] still crystallized under the same conditions as wild-type anxA5 itself. In contrast, it was impossible for Minks and co-workers to crystallize the congener anxA5[(4-F)Trp]. However, they argued that the overall structure is not affected by incorporation of (4-F)Trp due to the superimposable CD profiles and similar behavior during thermal denaturation.

Interestingly, this study reports the first efforts to produce a data set of molecular packing of fluorine-containing amino acids within a complex protein structure. Molecular modeling was carried out on the basis of the available high-resolution crystal structures. Calculation of the exclusion volumes revealed tighter packing of the parent protein anxA5 and anxA5[(5-F)Trp] whereas the packing of 4-FTrp and 6-FTrp (Fig. 3) into the core of anxA5 seems to be less tight by a factor of $\frac{1}{4}$ compared to the parent protein, and therefore indicating possible unfavorable electrostatic or even steric interactions. Interestingly, these findings correlate well with the thermodynamic stabilities and activity data. Unfortunately these interesting findings were not further studied and elaborated upon using other proteins or model systems.

AnxA5 proved to be a good model protein and was used in further studies in our lab along with azurin from *Pseudomonas aeruginosa* for the incorporation of (3-F)Tyr.[65] Again, there are no dramatic changes in fluorescence and CD profiles upon incorporation of the analog and the thermal stability is comparable to that of the parent proteins with azurin[(3-F)Tyr] being slightly more stable compared to azurin itself. However, anxA5[(3-F)-Tyr] is approximately threefold less active than anxA5 whereas electron transfer processes in azurin seemed not to be altered upon fluorination. Examples of other groups in which fluorination of tyrosine alters enzyme activity include β-galactosidase (fluorination was used to probe whether Tyr is part of the catalytically active portion of the protein),[66,67] ketosteroid isomerase from *Pseudomonas testosterone*[68] and rat glutathione transferase M1-1.[69] These fluorinated proteins are interesting because of their altered pK_a. Remember, the pK_a of the phenolic proton which is 10 in tyrosine varies between 5.2 and 9.0 in its fluorinated analogs depending on the level and position of fluorination.[68,70] Thus, the fluorinated Tyr analogs can be used in order to investigate acid-base catalysis or redox potential properties. In addition, fluorinated tyrosine radicals are characterized by peak reduction potentials between 705 and 968 mV whereas Tyr as a reduction potential of 642 mV.[71]

We could also achieve high incorporation levels of fluorinated Phe derivatives into anxA5 and azurin.[72] Thereby, we observed characteristic absorption maxima in the UV/Vis spectra of the fluorinated amino acid analogs as well as the related protein congeners. These arising maxima can be interpreted as fingerprints of the related (F)Phe and are termed "fluorophenyl-fingers". The protein yields decreased in the order (2-F)Phe ≫ (3-F)Phe > (4-F)Phe for both proteins but the trends are less pronounced in the case of azurin. As before, the structural integrity of the proteins were not affected by NCAA incorporation. The practical importance of these experiments can easily be recognized: The "fluorophenyl-fingers" appear in the UV-spectrum in the range of 260–270 nm and do not overlap with the contributions of tyrosine and tryptophan residues in the protein UV-spectra. Thus, fluorinated phenylalanines, especially those fluorinated at the *ortho*-position of the Phe-ring could be used as markers as they do not affect protein structure or activity and appear exclusively in the chromatographic fractions that contain the target proteins. In other words, the presence of such "fluorophenyl-fingers" ("FF-fingers") represents a unique spectroscopic probe to identify the labeled target protein among other non-labeled cellular proteins in preparative work by simple UV spectroscopy. In the coming era of proteomics such reliable, cheap and easy to reproduce methodology might have a great potential for high-throughput identification and characterization of target molecules in total protein output of a variety of organisms.[72]

The stop codon suppression (SCS) methodology

In 1992, Ellman and co-workers reported an NMR study of T4 lysozyme in which they introduced site-specifically [^{13}C]Ala for the purpose of NMR analysis by *in vitro* synthesis of the protein.[73,74] However, despite this and all related studies affording an enormous amount of useful information, the *in vivo* incorporation of NCAA is still the most prominent goal to achieve. In 1998, Furter performed the first *in vivo* incorporation of fluorinated amino acid using the in-frame stop codon suppression methodology.[33] He introduced an amber stop codon into murine dehydrofolate reductase (mDHFR) and it was translated with (4-F)Trp with approx. 64–75% efficiency (the remainder was translated as Phe or Lys). In addition, he detected also some (4-F)Phe incorporation at Phe positions which can be explained by the substrate tolerance of the endogenous enzymes, which is exploited in genetic code engineering. The expression level of the mDHFR congener was about 2/3 of the parent protein. By using a yeast tRNA^{Phe}/PheRS system in Phe-auxotrophic *E. coli* strains he started what is now dubbed genetic code expansion. However, only in 2007 the first successful incorporation of a fluorinated amino acid in a totally site-specific manner by incorporating 4-trifluoromethyl-phenylalanine (4-CF_3)Phe into either nitroreductase or histidinol dehydrogenase was reported.[75] Both enzymes are functional only as dimers meaning the activity is a measure of structural integrity. The evolution of an orthogonal pair capable for (4-CF_3)Phe incorporation was performed by positive and negative selection starting from a *Methanocaldococcus jannaschii* suppressor tRNA/aminoacyl-tRNA synthetase pair.[76] From the two candidates identified only one was further investigated. In contrast to former systems, the expression yielded protein amounts comparable to the parent proteins. Nitroreductase labeled at different positions with the fluorinated amino acid was used to assay the binding of substrates and inhibition, respectively. Therefore, mutants were generated varying the distance of the active site and the TAG codon/fluorine label from rather long distances to short-distances. However, the introduction of (4-CF_3)Phe had only minor effects on the activity. In addition, the authors measured the chemical shift sensitivity of position 36 and 124 in response to conformational changes by the addition of a competitive inhibitor, and if the cofactor flavin mononucleotide (FMN) was reduced and oxidized. It turned out, that position 36 is not as sensitive as position 124 with regard to conformational changes. The other enzyme studied, histidinol dehydrogenase, is a metallo-enzyme. It was studied to determine whether a conformational change occurs by the addition of Mn^{2+} or histidinol alone. However, no change was detectable. In contrast, by adding both compounds together, a shift of the NMR signals could be detected. Furthermore, ^{19}F NMR could also be performed *in vivo* with the fluorinated congener of nitroreductase. Whereas the first *in vivo* ^{19}F NMR studies examined the mobility of proteins, *e.g.* in *Saccharomyces cerevisiae*,[77–79] (4-CF_3)Phe probes can be used for deeper investigations. This was shown by the incorporation of this amino acid into several globular proteins in the range of 7 to 100 kDa.[80] In a first series of experiments congeners harboring (3-F)Tyr were used and the data were mostly in accordance with the literature.[51] However, to gain a better resolution the fluorinated Phe analog (4-CF_3)Phe was introduced to the proteins leading to more detailed results.

From these studies, it can be deduced that (4-CF_3)Phe could be a useful tool to examine large-sized proteins as well as protein complexes as the independent internal motion of the trifluoromethyl group of the fluorinated amino acid sharpens the resonance and facilitates the detection of larger globular proteins. (4-CF_3)Phe was also used in relaxation analysis. Here, the advantage of the fast rotational motion of the CF_3 group around the C–C bond could be exploited. The ^{19}F T_1 value of free

(4-CF_3)Phe is 200 fold shorter compared to the T_1 of (4-F)Phe (1.47/352 s).[81] In addition, in (4-F)Phe the ^{19}F spin is directly bonded to the aromatic moiety. Thus, (4-CF_3)Phe is better suited for chemical shift and relaxation studies. Examples include the human vinexin SH3 domain which upon (4-CF_3)Phe incorporation revealed different internal motions upon ligand binding *via* relaxation analysis,[82] as well as a nine-transmembrane helices membrane protein trimer of diacyl-glycerol kinase (DAGK) in dodecyl phosphocholine (DPC) micelles.[83]

(4-CF_3)Phe is not the only fluorinated aromatic acid which can be incorporated in response to an in-frame stop codon. For instance, (4-F)Bpa, (2,6-F_2)Bpa and (3-F, 4-NO_2)Bpa could be incorporated[84] with varying yields into a mutant superfolder GFP[85] (V150TAG, His-tagged). By mutations of the BpaRS, they generated various synthetases which differ in permissivity. It is worth to note, that some of these mutants were much more effective for incorporation of particular NCAAs. The authors also identified a mutant of NapRS suitable for 4F-Bpa incorporation. Mehl and co-workers also used their superfolder GFP mutant as model for their attempts to evolve aaRS for (4-CF_3)-Phe and other fluorinated amino acids such as (3-F, 4-CH_3)Phe, (3-F, 4-CF_3)Phe, (2,3-F_2, 4-CH_3)Phe and (2,3,5,6-F_4, 4-CH_3)-Phe.[86] They started their evolution with a modified *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase/$tRNA_{CUA}$ pair[87] and were able to select for enzymes which incorporated the desired amino acids into chloramphenicol acetyltransferase and superfolder GFP. Whereas, for example (3-F, 4-CH_3)Phe gives 50% of the parent protein yield, (4-CF_3)Phe yields only approx. 20%. However, the authors noted that the isolated yields do not match the examined *in vivo* yields which they attribute to problems during the purification procedure. Even though the two mentioned amino acids ((3-F, 4-CH_3)Phe and (4-CF_3)Phe) are very similar in their structure, their related aaRS differ in sequence and their ability to charge tRNAs. In addition, both enzymes were capable of incorporating at least some of the other fluorinated amino acids under investigation, with the size of the *p*-substituent proving to be crucial. The authors also report the generation of a hybrid aaRS bearing elements from both enzymes and, thereby, increasing the ability of incorporation of fluorinated amino acids. For their subsequent NMR studies, they incorporated (4-CF_3)Phe, (3-F, 4-CH_3)Phe, and (3-F, 4-CF_3)Phe into superfolder GFP at positions 134 (loop), 146 (near the fluorophore) and 150 (surface of the β-barrel). Here, they observed that the signals expectedly varied from sharp to broad.

The amino acid (4-OCF_3)Phe was incorporated into the 33 kDa thioesterase domain of the human fatty acid synthase.[88,89] This was combined with the incorporation of ^{13}C and ^{15}N labeled amino acids into 11 different positions around the active site in response to a TAG amber stop codon in order to probe the structure, dynamics and ligand binding of the protein. In addition to the development of a tRNA/aaRS pair specific for (4-OCF_3)Phe, improved yields of protein (2 mg of 50 mL culture in average for all mutants) could be achieved. The advantage of this method is the successful peak assignment in contrast to former studies of large proteins such as the 100 kDa citrate synthase with incorporated (5-F)Trp[77] or the 210 kDa trimeric pyruvate kinase also labeled with (5-F)Trp.[78] However, the first study revealed that the citrate synthase is motionally restricted *in vivo*.

Fluorinated tyrosines are commonly introduced into proteins by genetic code engineering, as the endogenous enzymes accept them as substrate.[24,65] However, recently Wilkins and co-workers could show that these amino acids can also be incorporated by in-frame stop codon suppression. The only prerequisite is a masking before incorporation.[90] The fluorinated Tyr analogs can be masked, for instance, with photoremovable protecting groups.[91–93] In that way, incorporation of *o*-nitrobenzyl-2-fluorotyrosine, -3-fluorotyrosine and -2,6-difluorotyrosine into superfolder GFP in *E. coli* was achieved. They started the evolution of the orthogonal pair from a mutant of the *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase for the incorporation of *o*-nitrobenzyltyrosine.[94] When the three analogs were incorporated into the chromophore (Tyr66), the 2-fluoro and the 2,6-difluoro derivative exhibited a blue shift in the emission profiles of the superfolder GFP (503 and 509 nm compared to 513 of the parent protein). In contrast, fluorine in position 3 leads to a small red shift (515 nm). Thus, the fluorescence properties can be altered in different ways using fluorinated Tyr residues. This is also in line with experimental data gained by genetic code engineering with enhanced green fluorescent protein (EGFP) as model protein (*vide infra*).[95,96] In addition, it is worth to note that even if the protecting group still blocks the tyrosine hydroxyl, the fluorescence maximum of the 2-fluoro derivative has an emission maximum at 503 nm. However, the intensity is decreased by approx 340% and the folding as well as the maturation state with protected fluorotyrosines is not clear yet.

Very recently, Schultz and co-workers reported an evolved aaRS with broad polysubstrate specificity.[97] They screened for the incorporation of (4-F)Phe as well as for *p*-trifluoromethyl-acetylphenylalanine among 17 other aromatic NCAAs. Here, they used a GFP-based screening system for incorporation and validated this by incorporation of the NCAAs into a His-tagged myoglobin mutant which could be expressed in levels up to 45% in the presence of the fluorinated non-canonical amino acids compared to the parent protein.

Aliphatic fluorinated amino acids can also be incorporated *via* the suppression-based approach.[98] Indeed, Hancock and co-workers could demonstrate that N^{ϵ}-trifluoroacetyl-L-lysine ((N^{ϵ}-$COCF_3$)Lys) can be incorporated into proteins of *S. cerevisiae via* a tRNA/aaRS pair which is based upon a pair evolved earlier for N^{ϵ}-acetyl-L-lysine. In this way, they provided an effective system for the creation of orthogonal pairs in *E. coli* which can be used to expand the genetic code of yeast.

Fluorination of enzymes

Already some early studies investigated the effect of fluorinated amino acids on the catalytic activity. For example, Ho and co-workers analyzed the influence of fluorinated aromatic amino acids on the activity of D-lactate dehydrogenase of *E. coli*.[44,99–101] Brooks and co-workers studied the effects of (2-F)Tyr, (3-F)Tyr and (2,3-F_2)Tyr on the catalytic activity of ketosteroid isomerase of *Pseudomonas testosterone*[68] whereas Ring and co-workers used (3-F)Tyr to identify a catalytically active Tyr residue in β-galactosidase.[66] In addition, Xiao and co-workers used these analogs to probe the structure and function of rat glutathione transferase M1-1.[69] More recently, the effect of fluorophenylalanine on the activity of *Pvu*II endonuclease has been studied.[102]

(2-F)Phe, (3-F)Phe or (4-F)Phe were incorporated at the four Phe sites of each subunit. Despite the low incorporation efficiency with respect to fluorination (7–17%), this modification resulted in strong effects on the catalytic activity. Whereas the catalytic specificity was the same for the parent protein and all other congeners, the congener carrying (2-F)Phe was as active as the parent protein but the one with (3-F)Phe was approximately twofold more active than the parent protein. Interestingly, the enzyme with (4-F)Phe in its amino acid sequence had a fourfold decreased activity compared to the Phe containing protein. Also the conformational stability of this congener was decreased as judged from unfolding experiments. A similar trend was observed by the Schultz lab in their *in vitro* experiments using fluorinated Tyr analogs and Staphylococcal nuclease.[103] In addition, a (3-F)Phe bearing congener of galactose binding protein also showed a significant decreased ligand affinity and a decreased stability in comparison with the parent protein.[104]

In 2009, Voloshchuk and co-workers introduced fluorinated phenylalanines into the histone acetyltransferase.[105] Again, the regioisomers differ in their effects on stability and catalytic activity. However, all of them showed decreased values for these properties. Interestingly, the three fluorinated congeners were also more prone to proteolytic digestion than the parent protein, which is in contrast to other examples in the literature.[106] Most recently, Baker and Montclare published their results on the effect of (4-F)Phe on the structural integrity and catalytic activity of S5 phosphotriesterase.[107] The fluorinated congener of this protein showed enhanced refolding properties. It is worth to note that fluorination has different effects on different substrates, as the congeners can exhibit higher and lower k_{cat}/k_M values compared to the parent protein. It should also be mentioned that subtle changes in structure and side-chain interactions/directions of fluorine containing proteins compared to the respective parent protein have been observed in several studies which may account for the decreased stability.[69,95,108,109]

The lab of Montclare also studied the effect of aliphatic amino acid fluorination on the catalytic performance of enzymes. For example, they incorporated Tfl into the chloramphenicol acetyltransferase.[110] This protein was chosen as its 13 Leu residues are not involved in substrate recognition or catalysis and the function can be assayed very easily. They found that fluorination in an efficiency of up to 80% has no effect on the catalytic activity. However, thermo- and chemostability were markedly decreased. This is interesting since they detected the biopolymer being more structured upon fluorination; *i.e.* an enhanced secondary structure does not lead automatically to an enhanced stability (*vide infra*).

A more systematic study was presented by Hoesl and co-workers who investigated the effect of various amino acids on the lipase of *Thermoanaerobacter thermohydrosulfuricus* (TTL).[111] The amino acids studied included fluorinated derivatives such as (3-F)Phe, (4-F)Phe, (4*S*-F)Pro, (4*R*-F)Pro, (2-F)Tyr and (3-F)Tyr. The lipase consists of 6 Pro residues, 16 Phe residues and 7 Tyr residues (Fig. 4). Whereas the fluorinated Pro analogs gave only slightly higher yields compared to the parent protein, the Phe analogs gave much higher yields (up to threefold increase).Tyr analogs, however, yielded lower amounts of protein. The incorporation efficiencies of (4*S*-F)Pro and the Tyr analogs were very good but substitution with fluorinated Phe analogs and (4*R*-F)Pro could be only achieved with a low efficiency. Regarding the catalytic parameters, none of the congeners showed a higher optimal temperature T_{opt} than the parent protein but the (3-F)Phe and (2-F)Tyr congeners retained the 70 °C from the parent protein. Interestingly, the regioisomers have quite different T_{opt} (3-F)Tyr and (4-F)Phe perform best at 50 °C as well as the congeners containing fluorinated Pro residues. With the exception of the (4*R*-F)Pro congener, all congeners show a high thermal stability (75 °C for 180 min). The proteins harboring (4*S*-F)Pro, (4-F)Phe as well as (3-F)Tyr were stable even far above their optimal temperature. A characteristic feature of the lipase is its ability to be activated by heat. For the fluorinated Tyr and Phe residues, activity changes with temperature and is again regioisomer dependent: (3-F)Phe is especially beneficial for the activity whereas (4-F)Phe impairs the activity of the lipase. And whereas the (3-F)Tyr congener is mainly inactive, the one bearing (2-F)Tyr is, after thermal activation, fully active. In the case of (3-F)Phe incorporation, the substrate tolerance of the resulting protein congener was also affected. This congener shows a broader tolerance compared to the parent protein. Interestingly, in case of the proline analogs, no significant heat activation could be detected.

Finally, Wiltschi and co-workers studied the effects of fluorinated CalB expressed in yeast.[106] A glycosylation-deficient mutant of *Candida antarctica* lipase B was expressed in *Pichia pastoris* and the protein segregated to the medium. The analogs (5-F)Trp, (3-F)Tyr or (4-F)Phe were used for incorporation experiments. However, homogenous samples could not be isolated in any case. The mixtures of fluorinated protein species were further analyzed, revealing that the fluorinated congeners were still active, albeit with lower activity. But even with the resistance to proteolytic degradation by proteinase K remaining unchanged, the authors could observe a prolonged shelf life of the lipase activity upon fluorination. The secondary structure analysis revealed some slight changes after fluorine incorporation, especially for (3-F)Tyr. In this variant, the content of random coils was increased. Thus, Tyr might be important for the structural integrity of the protein. This could be also argued from the changes in fluorescence emission spectra upon protein fluorination.

In addition to experimental work, calculations have been performed. For instance, the biotin binding ability of streptavidin in correlation to the fluorination of tryptophan residues was investigated.[112] It turned out, that in this case protein stability and folding are only weakly affected by fluorination of the Trp residues. The binding abilities of the congeners can be predicted to be decreased or increased depending on the position of the fluorine substituent.

Generally speaking, the majority of successful mono-fluorinated amino acid substitutions so far introduced in recombinant proteins are found to be structurally neutral. However, such fluorinations are usually accompanied by the emergence of novel spectral, structural, and enzymatic properties. As already discussed, Ring *et al.*[66,67] probed the role of catalytic Tyr in the active center of β-galactosidase by the global replacement of all Tyr-side chains with (3-F)Tyr, whereas Brooks and co-workers used (2-F)Tyr to dissect the catalytic mechanism of ketosteroid isomerase.[68] More recently, interesting spectral and dynamic properties of green fluorescent protein upon global substitution

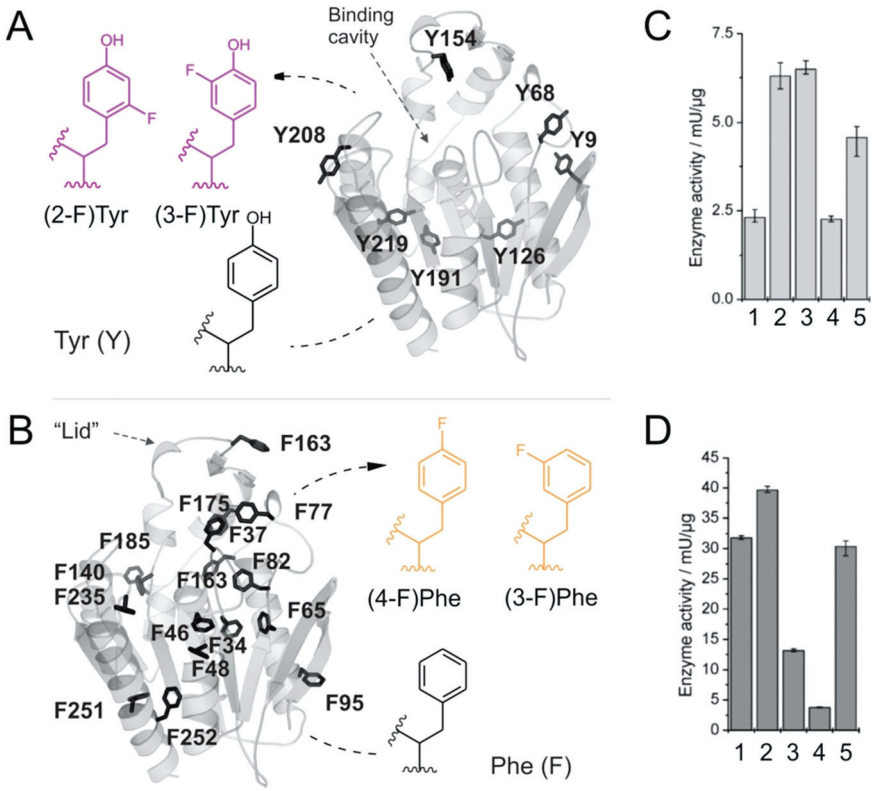


Fig. 4 Influence of the incorporation of monofluorinated Tyr and Phe residues on the enzyme activity of lipase. (*Left*) Three dimensional models of lipase from *Thermoanaerobacter thermohydrosulfuricus* (TTL) with marked Tyr (**A**) and Phe (**B**) residues (one-letter code). Following substitutions have been performed: Tyr → (3-F)Tyr, and (2-F)Tyr; Phe → (3-F)Phe), (4-F)Phe. (*Right*) Results of the activity assays with 1: TTL, 2: TTL[(3-F)Phe], 3: TTL[(4-F)Phe], 4: TTL[(3-F)Tyr] and 5: TTL[(2-F)Tyr]. The lipase is almost not active prior to heat activation (C). Upon activation (D) TTL[(3-F)-Phe] and (TTL[(2-F)Tyr]) along with parent TTL are highly active. The other two fluorinated TTL variants exhibit rater modest activity (TTL[(4-F)-Phe]) (TTL[(3-F)Tyr]). In general, while crystallographic and modeled structures of these proteins are indistinguishable from their native counterparts (*i.e.* they are almost perfectly isomorphous), there are considerable differences in their enzymatic properties.

of its Tyr-residues with (2-F)Tyr and (3-F)Tyr were reported as well (*vide infra*).[95,96] Similarly, the appearance of "fluorophenyl-fingers" discussed above represents a non-invasive fluorinated optical probe designed to introduce dramatic changes in the spectral properties of the protein without compromising its stability and/or biological activity.[72] Finally, these findings and our recent success to generate fluorinated lipase congeners with elevated enzyme activity, enhanced substrate tolerance and dramatic changes in the enzymes' temperature optimum represent a solid basis for the design new classes of fluorine-containing protein-based biocatalysts.

Fluorination of Pro residues in proteins

According to peptide studies, the fluorination of Pro residues can have a strong impact on the stability and folding characteristics of the related peptide. Pro is the only imino acid among the canonical amino acids and is often present in structural import positions in peptides and proteins, for instance in turns, and despite being a common feature of helices, it can also act as a helix breaker. It is hypothesized that the fluorination of position 4 of the pyrrolidine ring affects preorganization principles and that there is a stereochemical bias in folding. This is due to the fact that the stereoelectronic effects of the C4-substituent influence the *cis*/*trans* ratio of the prolyl peptide bond as well as the pucker of the pyrrolidine ring (Fig. 5). Indeed, (4*R*-F)Pro favors the C^{γ}-*exo* puckering whereas (4*S*-F)Pro prefers C^{γ}-*endo* puckering as determined in 1973 for the first time[113] and confirmed in several studies.[9,114–120] These preferences can be explained by the *gauche* effect. In the C^{γ}-*exo* pucker, an $n \rightarrow \pi^*$ interaction between oxygen of the Xaa-Pro peptide bond and the carbon of the Pro-Yaa peptide bond can be observed, which stabilizes *trans* peptide bonds. In contrast, C^{γ}-*endo* favors a *cis* peptide bond. This leads to the assumption that the bias of (4-F)-Pro to adopt either an *exo* or *endo* puckering can have antagonistic effects on peptide/protein stability, depending on the puckering and the structural context in the natural state. For example, incorporation of (4*S*-F)Pro into collagen and its natural precursors was performed in cell culture already in 1974.[121] This incorporation yielded a destabilization as well as a decrease in folding rate of the collagen triple helix. Conversely, in 1999 it was shown that synthetic collagen peptides bearing (4*R*-F)Pro are characterized by an increased stability of the triple helix (*vide supra*).[8]

As the puckering has an impact on the peptidyl prolyl bond conformation, the peptidyl prolyl *cis*/*trans* isomerase (PPIase) reaction was investigated with substrates bearing (4-F)Pro.[122] The 4*S* as well as the 4*R* isomers of the amino acid were used in

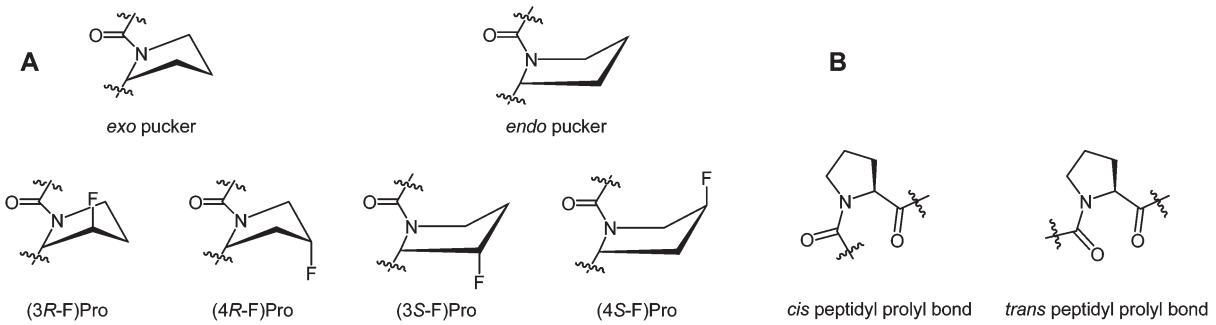


Fig. 5 Schematic representation of the *endo* and *exo* pucker conformation of proline and its fluorinated derivatives (**A**) as well as of the *cis* and *trans* prolyl peptide bond (**B**).

order to study the kinetics and the molecular principles of the enzyme's substrate specificity as the Pro residue at position P1′ of the substrate is assumed to govern the recognition process. In order to assay the effects of the fluorinated derivatives, a tetrapeptide as well as mutants of the nuclease inhibitor barstar (b*) were used. Barstar is a small protein and a well-known folding model. The mutant protein used here had a single Pro residue which exhibits a *trans* conformation in the unfolded state and a *cis* conformation in the native state. Most PPIases tested showed a preferred stereospecificity for the b*[(4*S*-F)Pro] congener, with one exception, (*E. coli* Par10). The authors could also detect a difference between the model peptides and the related proteins in their susceptibility to PPIase catalysis. However, spontaneous isomerization is not affected by fluorination of the C4 of the pyrrolidine ring.

Renner and co-workers also used model peptides and barstar for folding and stability studies.[9] These studies confirmed the preference of (4*S*-F)Pro for a *cis* conformation and the preferred *trans* conformation for (4*R*-F)Pro. They also confirmed the preferred puckers (Pro, (4*S*-F)Pro and (4-F_2)Pro adopt an C^{γ}-*endo* pucker, (4*R*-F)Pro exhibits an C^{γ}-*exo* pucker). By supplementing the medium with (4-F)Pro for the *in vivo* expression of the barstar mutant with a single Pro residue, they could demonstrate that incorporation of (4*S*-F)Pro leads to a more stable and (4*R*-F)-Pro to a less stable congener compared to the parent protein. This is due to the fact that Pro48 of barstar prefers *cis* conformation, which is in accordance with the expectations. Interestingly, (4-F_2)Pro did not have an effect on thermal stability.

However, the situation regarding the effects of (4-F)Pro incorporation is inversed in the Trp cage miniprotein.[123] Here, a *trans* proline bond is present in the model compound and Naduthambi and Zondlo equipped the peptide *in vitro* with different substituents in position four including (4*S*-F)Pro and (4*R*-F)Pro. This allowed them to tune the stability of the peptide in a range of 50 °C. Interestingly, C^{γ} *endo*-favoring residues (such as (4*S*-F)-Pro) at the essential Position Pro12 destabilize the structure, whereas C^{γ} *exo*-favoring residues (such as (4*R*-F)Pro) at that position stabilize the peptide. Other peptide studies evaluated the ring puckering effects on the conformational preferences of human prolyl 4-hydroxylase.[124,125]

Returning to the *in vivo* incorporation of (4-F)Pro into proteins, the results of the Conticello lab have to be considered.[126–128] They attempted to incorporate (4*R*-F)Pro, (4*S*-F)Pro as well as (4-F_2)Pro into a elastin-mimetic polypeptide with 20% Pro residues in the repetitive domain. Whereas the monofluorinated Pro analogs could easily be incorporated by a Pro auxotrophic *E. coli* strain, the difluoro analog (4-F_2)Pro required changes in the expression protocol, *i.e.* the co-expression of the wild-type aminoacyl-tRNA synthetase. It is worth to note that by hyperosmotic concentrations of osmolytes in the media the protein yields could be enhanced in a concentration dependent manner. This is explained by former studies showing that the biosynthesis of endogenous low-affinity proline transporters can be up-regulated by hyperosmotic induction.[129,130] In their study, the authors also confirmed, that the yield of protein expression in the presence of NCAA depends on the host-cell physiology, model protein and experimental conditions by using a variety of analogs. Furthermore, Conticello and co-workers investigated the effects of the incorporated epimers on the self-assembly of the related polypeptides. The transition temperature (T_t) from a random coil conformation to a more ordered one was determined to be 33 °C for the proline bearing parent protein. The congener with (4*S*-F)Pro had a T_t of 41 °C and the one carrying (4*R*-F)Pro had a T_t of 22 °C as determined by differential scanning calorimetry (DSC).[128] Measuring CD spectra, the authors could show that the regioisomers again have different influences on the secondary structure. For example, increasing the temperature leads to CD profiles indicating a higher content of type II β-turn structures for the parent protein as well as the (4*R*-F)Pro congener. In contrast, the profile of the (4*S*-F)Pro congener had a totally different shape, which seems to resemble a type I β-turn. However, for this congener a two-state transition could not be detected by CD spectroscopy, which was the case for the native protein and the (4*R*-F)Pro congener. The related T_t values measured by CD spectroscopy (29 and 23 °C) were in agreement with the DSC experiments for the two congeners. As NMR analysis identified the *trans* isomer as the major conformation in all polypeptides, the *cis*/*trans* isomerization cannot explain the different assembly behavior. It is therefore more likely that the puckering of the pyrrolidine ring has a direct influence on the overall structure. The authors claim that due to stereoelectronic effects, the conformation of the (4*R*-F) Pro might be pre-organized in a way that favors the type II β-turn structure.[128]

The elastin-mimetic polypeptide has also been used for the incorporation of (3*R*-F)Pro and (3*S*-F)Pro in order to study the effect of fluorination in position 3 of the pyrrolidine ring.[126,131] Compared to their 4-fluoro counterparts, these analogs exhibit an

opposed pyrrolidine ring puckering as a consequence of the gauche effect, supported by crystallographic analyses of model compounds. The (3*R*-F)Pro displays an C^{γ}-*exo* puckering whereas the (3*S*-F)Pro displays an C^{γ}-*endo*-puckering. Considerations regarding the *syn*/*anti* orientation are of great interest when comparing (3-F)Pro and (4-F)Pro analogs: (3*R*-F)Pro (*exo*) has a *syn* arrangement in respect to the configuration of the C^{α}, while (4*S*-F)Pro has a *syn* orientation too but shows a C^{γ}-*endo* puckering. The same is true for the *anti*-arrangements: (3*S*-F)Pro exhibits an C^{γ}-*endo* puckering and (4*R*-F)Pro an C^{γ}-*exo*-puckering. The preferences of the analogs are transferred from the small molecules to the related elastin-mimetic.[128,131] The authors succeeded in the expression of (3-F)Pro containing congeners with high substitution efficiencies. Again, it could be shown that an *exo* puckering of C^{γ} stabilizes type II β-turn structures at increased temperatures whereas an *endo* puckering favours more type I β-turn conformations.[131] The transition point for the conformational rearrangement from a random coiled structure to higher ordered β-turn motifs could be determined to be 30 °C and 10 °C for elastin[(3*R*-F)Pro] and elastin[(3*S*-F)Pro], respectively, by CD spectroscopy and differential scanning calorimetry. This means, the *syn*-fluoroprolines raise the transition temperature of the related proteins.[128,131] As for the (4-F)Pros, the *cis*/*trans* isomerization equilibrium can be ruled out as the reason for this observation, as NMR revealed that all 3-fluoroprolines prefer the *trans* conformation.[131,132]

Steiner and co-workers proved that puckering also has a great impact on the folding of the β-sheet-rich green fluorescent protein (GFP).[27] The refolding rate of GFP, and its mutants, is known to be relatively slow.[133] This is often attributed to the *cis*/*trans* isomerization of peptidyl-proline bonds discussed previously. Attempts to incorporate (4*R*-F)Pro into EGFP, a mutant of GFP consisting of 10 Pro residues, led to the formation of insoluble inclusion bodies. In contrast, substitution of all Pro residues with (4*S*-F)Pro yielded in a soluble protein isolation with enhanced (re)folding properties (*e.g.* faster refolding kinetics). In addition, a higher fluorescence recovery after denaturation could be achieved compared to the parent protein. Furthermore, the (4*S*-F)Pro congener is less prone to aggregation. All these findings could be mainly attributed to 12 new stabilizing interactions, as detected by X-ray analysis. These new interactions involve the fluorine atoms and the puckering of the pyrrolidine ring. As nine out of ten (4*S*-F)Pro residues show *endo* puckering, conformational preferences might lead to some form of preorganization.

With regard to the capability of the incorporation of 4-fluoroprolines into the thermophilic KlenTaq DNA polymerase, Holzberger and Marx reported contrary observations compared to the EGFP of Steiner *et al.*[134] The polymerase consists of 32 Pro residues which were exchanged in high efficiency (92%) for (4*R*-F)Pro. However, the authors noted a dramatic decrease in protein yields. On the other hand, the (4*R*-F)Pro congener is as active as the parent protein. The same holds true for fidelity and sensitivity. Conversely, the (4*R*-F)Pro congener displayed a lower thermostability compared to the parent protein. In addition, the authors failed to incorporate (4*S*-F)Pro.

The same holds true for the incorporation of (4-F)Pro residues into human ubiquitin.[135] Again, only the Pro → (4*R*-F)Pro substitution was successful (in yields comparable to the parent protein) whereas (4*S*-F)Pro incorporation could not be detected. This is reasonable, as the Pro residues of the 76 amino acid containing protein favor the C^{γ}-*exo* puckering as (4*R*-F)Pro does. In addition, the Pro residues exhibit *trans* conformation which is also promoted by (4*R*-F)Pro. Fluorination proved again to be beneficial for the stability of the protein which retains its activity (in *in vitro* analysis). The folding kinetics revealed a slightly faster folding and a 2.6-fold slower unfolding (measured at acidic pH). Furthermore, the secondary structure was not affected by fluorination as careful inspection of the CD profiles revealed, but the melting temperature was higher for the congener. The authors also performed computer simulations and found that fluorine is involved in several stabilizing interactions; e. g. with amide groups of the backbone. All these interactions are only possible with C^{γ}-*exo* puckering whereas C^{γ}-*endo* puckering would generate repulsive interactions.

It seems TTL and the elastin-mimetic polypeptides are among the few examples where incorporation of (4*S*-F)Pro or (4*R*-F)Pro into the same protein can be afforded without decreased yields for one of the congeners (*vide supra*).[111,126] This could be explained by the results of peptide studies showing that the biased C4 ring puckering of proline due to stereoelectronic effects cannot selectively stabilize or destabilize the structure when the proline residue is involved in other interactions, e. g. aromatic-proline interactions.[116] On the other hand, the flexibility/structural rigidity of the biopolymer might be important as well. Small peptidic antibiotics like lanthibiotics can be also ribosomally synthesized with either (4*S*-F)Pro or (4*R*-F)Pro in their sequence.[136] In this context it is also worth to note again, that in case of EGFP incorporation of one isomer ((4*S*-F)-Pro) resulted in a soluble congener whereas incorporation of the other one ((4*R*-F)Pro) led to the formation of insoluble protein precipitate.[27] In contrast, in KlenTaq DNA polymerase and human ubiquitin no expression at all could be detected with the disfavored isomer.[134,135] Thus, not only folding but also biosynthesis of proteins might be affected by the different stereochemistry of the fluorinated proline residue.

In summary, these findings clearly demonstrate that with detailed structural information of a protein, the use of fluorinated NCAAs analogs allow for new strategies for the design of tailor-made proteins with folding and stability characteristics superior to those of the native molecules. In addition, stereoselective preferences for the isomers during translation is most probably due to a preorganization effect in the structure of the target protein, although other factors such as ribosomal translational editing and proofreading and co-translational folding cannot be excluded. Therefore, fluorinated Pro-residues might be also seen as a tool to study and understand the basic mechanisms of protein translation.

Fluorinated autofluorescent proteins

Green fluorescent protein and its mutants and variants are attractive targets for studying the effects of non-canonical amino acid incorporation.[137] Of special interest are, for instance, the fluorination of the respective chromophores. For example, the dynamics and spectroscopic characteristics upon fluorotryptophan ((4-F)Trp, (5-F)Trp or (6-F)Trp) incorporation into EGFP

and ECFP (enhanced cyan fluorescent protein) have been studied.[109,138] Whereas EGFP contains only a single Trp residue in position 57, ECFP, in addition, harbors Trp66 as a part of the chromophore. Interestingly, in ^{19}F NMR analyses of the fluorinated EGFP and ECFP congeners, two and four signals were detectable, respectively, reflecting two states of each fluorinated Trp residue. Slow exchange processes are also indicated by broad signals and could be further estimated to be in the time frame of 1.2–1.4 ms. In addition, the population of the two states is temperature dependent and differences in enthalpy and entropy could be measured. Further analyses showed that dimerization and aggregation can be excluded as reasons for the peak widening. Chromophore isomerization can also be ruled out[139,140] as well as different protonation states. As ECFP[(6-F)Trp] exhibits two (6-F)Trp57 signals also in the denatured state, the authors argue that the *cis*/*trans* isomerization of the Pro56-Trp57-Pro58 sequence might be responsible for this observation. In contrast, the signal widening for the fluorinated Trp66 in ECFP might arise from changes in the surrounding of the chromophore. For example, Tyr145 and His148 show an increased flexibility indicating some conformational freedom, and they are located in the vicinity of Trp66.[109,141] To date, the related two conformational states of the proteins could only be detected in solution (^{19}F NMR) but not in crystal structures.[138] As the fluorescence properties of these proteins arise mostly from the chromophore, it was expected that the incorporation of fluorotryptophans into EGFP will have no significant effect and indeed there is no difference in the spectral profiles of the congeners and the parent protein. Conversely, the absorption profile of ECFP is slightly affected by a blue-shift of 6 nm with the exception of ECFP-[(6-F)Trp] where only a 2 nm blue-shift is detectable. The emission maxima are also blue-shifted in the range of 2–10 nm. Furthermore, the relative fluorescence is decreased by 30–60%. This reduction correlates with the position dependence of halogenated porphyrins.[142]

In order to study EGFP in more detail, it was equipped with either (2-F)Tyr or (3-F)Tyr.[95,96] Incorporation of these analogs into EYFP was performed for comparison as well. The eleven Tyr residues were replaced quantitatively and slight changes in the spectral properties could be detected. In EGFP, incorporation of (2-F)Tyr resulted in a 6 nm blue-shift in absorption and the regioisomer (3-F)Tyr showed a blue-shift of only 3 nm whereas the in the emission profiles either a blue-shift of 6 nm or a red-shift of 4 nm was detectable.[95] The relative fluorescence intensities differed markedly when compared to the parent protein. For EYFP, the situation is slightly different. Here, the two isomers exhibited either a blue-shift of 10 nm ((2-F)Tyr) or a red-shift of 4 nm ((3-F)Tyr) in the absorbance spectra. Similar trends were observed in the fluorescence properties. Here, the blue-shift is 7 nm and the red-shift 6 nm. Interestingly, in EYFP the relative fluorescence intensity is either equal to or even slightly higher than for the congeners. The differences between EGFP and EYFP may result from the fact that in the latter protein not only the chromophore itself is affected by fluorination but also Tyr203 in its vicinity (stacking position). Furthermore, it was investigated if any changes can be observed by pK_a shifts upon fluorination as the values differ for the free amino acids (Tyr: pK_a = 10.05, (3-F)Tyr: pK_a = 8.5).[68] Indeed, pH titrations of the chromophore revealed decreased values in the range from 0.1 to 1.2 units (Fig. 6). These experiments also showed that the absorption spectra of EGFP and EYFP are strongly pH dependent. However, this is less pronounced in the fluorinated congeners but EYFP[(2-F)Tyr] behaves similar to the parent protein at low pH values.[95]

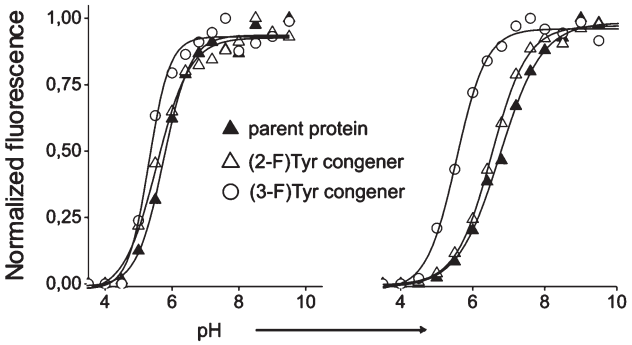


Fig. 6 Influence of chromophore fluorination on fluorescence pH titration profiles in parent and congener EGFPs (left panel) and EYFPs (right panel). Whereas there is almost no change in the pK_a values for most of the congeners EYFP[(3-F)Tyr] has a pK_a value decreased by 1.2 units.

In addition, high-resolution crystallographic analysis was performed on the congeners and revealed only marginal changes of the overall folding. However, careful inspection of the data revealed crystallographic distances between fluorine atoms and hydrogen donors allowing for weak interactions. In contrast to (2-F)Tyr which exhibits only one conformation at all positions and none of the fluorine is involved in special interactions, some of the (3-F)Tyr residues in EGFP[(3-F)Tyr] adopt two conformations (Fig. 7). The authors assumed that in these cases ring flipping is connected to the position of the aromatic residue (buried or solvent-exposed). For instance, (3-F)Tyr145 is involved in several contacts with atoms of different residues in its vicinity as well as water molecules and only one conformer is detectable. In contrast, buried (3-F)Tyr92 and partially exposed (3-F)Tyr143 adopt two conformations, with the latter not involved in particular interactions. This could also be confirmed by the NMR studies of Khan and co-workers.[108] They used a truncated mutant of GFPuv and mutated each Tyr residue one after the other to Phe in order to assign the NMR signals. From their experiments they concluded that ring-flipping must be a slow process and is in a time frame of 160 s^{-1} for (3-F)Tyr143 and 960 s^{-1} for (3-F)Tyr92. As the other residues exhibit no or only marginal ring flipping conformers, the authors argue that packing effects influence the flexibility.

It is even more interesting that (3-F)Tyr in position 66 (as part of the chromophore) exhibits two conformations as well; even though its surrounding is rather rigid as judged from the *B* factors.[96] The phenolate oxygen of the chromophore is in contact with residues in its vicinity and the chromophore itself interacts *via* hydrophobic interactions with Val50 and Phe165.[95] The two conformations are termed major and minor, respectively, as they are occupied to 60 and 40%. In both cases, the hydroxyl group of the (3-F)Tyr residue interacts with Thr203 and His148 through hydrogen bonding (in the major conformation, fluorine interacts with the C$^{\gamma}$ of Thr203 which might be a rather unusual example of a weak CF···CH interactions).[95,96]

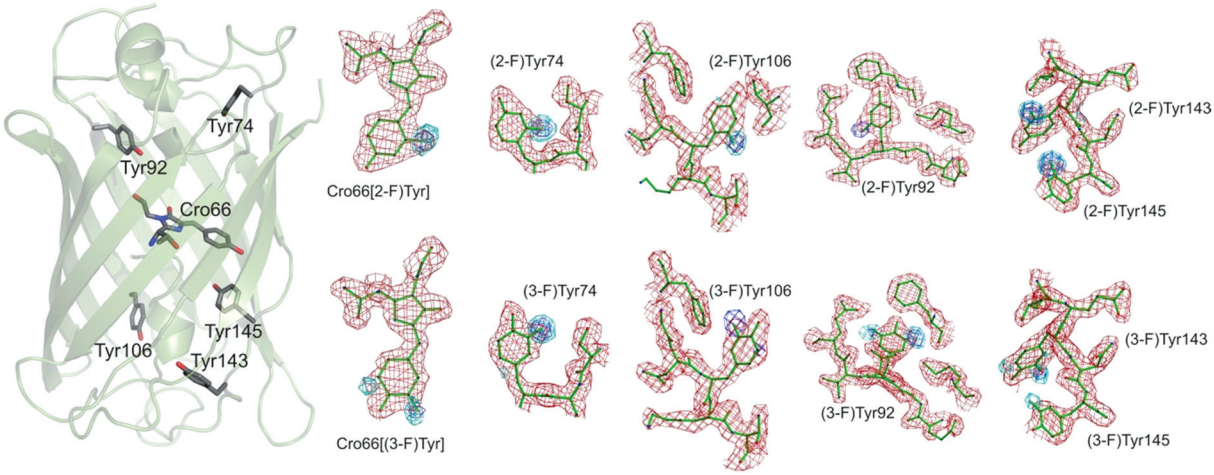


Fig. 7 Survey of the conformation of exemplary Tyr residues in EGFP substituted by either (2-F)Tyr or (3-F)Tyr revealed by X-ray analysis. Note that in the case of EGFP[(2-F)Tyr] all fluorinated residues exhibit only one conformation whereas some of the (3-F)Tyr residues adopt two conformations in the related congener. For further discussion see text.

The authors argued that flipping of the chromophore does not take place. It seems to be more plausible to assume that the two states arise during chromophore maturation. In this context it is worth to note, that the occurrence of only one conformer in EGFP[(2-F)Tyr] must have arisen due to steric reasons, as further analysis revealed that by adaption of the alternative conformation the fluorine atom would directly clash with the ring nitrogen of the chromophore. Additional studies demonstrated that the fluorination of the chromophore by introducing either (2-F)Tyr or (3-F)Tyr has no effect on its photostability in EGFP or EYFP.[143]

(3-F)Tyr was also used to study the effect of fluorination on the chromophore of DsRed.[144] Again, it was argued that even if all Tyrs were exchanged by the analog, only fluorination of Tyr67 will affect the fluorescence properties as shown for GFP mutants.[95,96,141] The fluorescence excitation spectra of the parent protein and the congener showed no differences, except in the intensities. The spectrum of the fluoro-congener is more intense. However, after excitation at 556 nm, the emission maximum of the parent protein can be detected at 603 nm, whereas the maximum of dsRed[(3-F)Tyr] is blue-shifted by 12 nm to 591 nm. This might be attributed to electron-withdrawing effects since comparison with incorporated (3-NH_2)Tyr revealed a red-shift of 12 nm. Whereas the fluoro-congener again showed higher fluorescence intensity when studied under the same conditions as the parent protein, it has a lower extinction coefficient (59 000 compared to 47 000). However, its quantum yield is also higher than the one of the dsRed itself (0.25 compared to 0.04). Careful inspection of the UV/Vis spectra revealed that upon fluorination the spectral shoulder at 480 nm is more pronounced. This absorption maximum is attributed to a green-emitting intermediate of the chromophore. As the 480 nm absorption is even higher as the 556 nm absorption in the fluoro-congener, the question arises whether fluorination hinders a proper chromophore formation.

Besides the efforts to incorporate aromatic non-canonical amino acids, aliphatic analogs have also been tested. For example, Tfl or Tfm have been incorporated into an EGFP

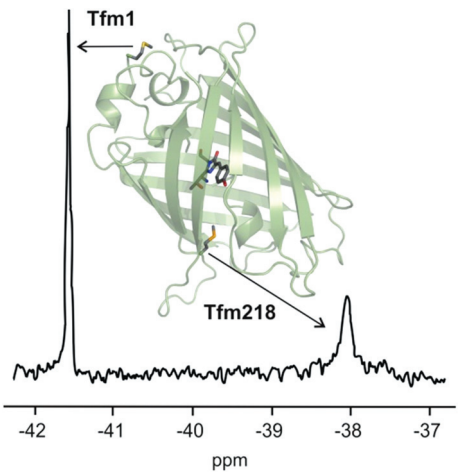


Fig. 8 ^{19}F NMR analysis of the congener EGFP-2M[Tfm]. The difference in chemical shift and shape of the two signals assigned to the two fluorinated residues is due to the different chemical microenvironment of the Tfm residues within the protein structure given rise to the assumption that the congener is correctly folded.

mutant containing only 2 Met residues (Fig. 8).[145] However, these first trials resulted only in a very inefficient substitution and the low expression yields were attributed to the increased size/van der Waals radius [20,146] of the trifluoromethyl group. Despite the low yields, it was possible to analyze the Tfm containing congener by ^{19}F NMR and the detected signals could be assigned. Problems during substitution experiments using Tfl were faced as well by Tirrell and Yoo.[147] Nevertheless, they succeeded in the evolution of a GFP carrying Tfl (*vide infra*).[148]

Taken together, global and site-directed fluorinations of green fluorescent proteins enabled the design of congeners with novel and unique biophysical properties including changes in titration curves as well as variations in the absorbance and fluorescence profiles. The global fluorination of autofluorescent proteins to the level of "atomic mutations" (*i.e.* H → F replacements) offer

unprecedented opportunities to understand and manipulate the relationship between protein structure and its spectroscopic properties.

Towards fluorous proteins[149]–protein design with trifluoromethyl groups

As already mentioned, many studies have been performed using non-canonical amino acids equipped with trifluoromethyl groups. A very interesting result of these studies is the fact that Hfl destabilizes α-helices but stabilizes β-sheets.[13,150] Thus, the structural framework of a model protein should be considered when designing incorporation experiments. However, it is highly desirable to investigate if it is possible to generate "Teflon proteins", regarding their potentials arising from the unique properties of fluorine.[149] Teflon itself is a fluorous polymer and according to Gladysz and Curran the term fluorous means "highly fluorinated or 'rich in fluorine atoms' and based upon sp^3-hybridized carbons".[151] Generally speaking, compounds in which a large potion or even all hydrogen atoms are completely replaced by fluorine atoms, could be defined as fluorous. Furthermore, to generate real fluorous or "Teflon" proteins techniques are needed to synthesize biopolymers with multiple perfluorinated side chains which are sufficiently close in space to allow interaction in the folded state. There are many potential benefits of Teflon proteins, *e.g.* they may find industrial application due to their predicted resistance to unfolding in organic solvents, which has been summarized by Marsh.[149]

A first step on the way to generate Teflon proteins is the global introduction of trifluoromethyl groups into proteins. These works are based on studies involving peptides and chemically synthesized proteins as described in the first part of this review. In 1963, Rennert and Anker were the first ones using Tfl as medium supplement and observed it's *in vivo* incorporation into *E. coli* proteins.[41] In 1975, there was an indirect evidence for Tfm incorporation into TCA-insoluble protein fractions from *S. cerevisiae*.[63]

In 1997, the Honek lab reported the successful incorporation of Tfm into bacteriophage λ lysozyme (LaL) consisting of three Met residues.[62] By careful control of the expression conditions, the efficiency of incorporation could be modulated being either high (70%) or low (30%). The resulting congeners were analyzed by mass spectrometry. Higher Tfm concentrations resulted in lower overall protein yields and *vice versa*, and excessively high concentrations seemed to be toxic for the cells. However, the activity of the congeners was similar to the one of the parent proteins. By ^{19}F NMR analysis, four sharp signals could be detected for the three Tfm residues which means that one residue splits into two signals with their intensities being influenced by the incorporation level. The authors provided a model for the structural reason of the two signals representing one Tfm which includes subtle changes in the protein structure by the incorporation of the fluorinated analog. Later experiments revealed that, in the M14L mutant of LAL, the set of double signals vanishes and only one resonance per residue is detectable.[152] Thus, signal assignment could be performed. Furthermore, not only the size of Tfm is increased compared to Met[62,152] but also the hydrophobicity is altered by the 3 H → F substitutions. For instance, the log *P* value for ethyl trifluoromethyl sulfide was calculated to be 2.73 compared to 0.80 for ethyl methyl sulfide. This is also reflected in the increased retention time of the LaL congener with high substitution level on reversed-phase HPLC columns. Even though the presence of maximal three trifluoromethyl groups is insufficient to make the LAL congener a fluorous molecule, it has to be stated that the hydropathy is already greatly affected. Similar observations have been made for the increased hydrophilicity after the incorporation of a single azatryptophan into a model protein.[153] In addition, CNBr/formic acid reaction of Tfm containing lysozyme yielded only non-digested protein at Tfm positions.[154] In addition to Tfm, difluoromethionine (Dfm) was incorporated into LAL and studied by ^{19}F NMR.[155] In contrast to Tfm, the Dfm congener was expressed in high yields and high substitution efficiency (almost 100% according to ESI MS). Again, the congener retained complete catalytic activity compared to the parent protein. The observed ^{19}F NMR signals could also been assigned to the different Dfm positions.

Both amino acids, Dfm and Tfm, were also used as probes for conformational changes in MetRS upon binding and in the ATP : PP_i exchange assay.[156,157] By doing this, they could identify key residues of *E. coli* MetRS (*e.g.* Y15, W253, H301 and other aromatic amino acids). Dfm binding caused, in general, the same conformational changes as Met although the enzyme affinity was slightly decreased. In case of Tfm, the conformational changes were different compared to the native substrate. With this finding, the authors explained the dramatically lowered Tfm incorporation rate into proteins by the less efficient $tRNA^{Met}$ aminoacylation of Tfm. In another series of experiments, the reduction potential of semisynthetic azurin could be tuned by the incorporation of Dfm or Tfm into the protein at position 121.[158] Following this logic of thinking, Garner and co-workers saw an indication for a linear correlation between the reduction potential and the hydrophobicity. According to them, this is also true for other type 1 copper-containing proteins regardless of their scaffold or the general experimental set-up.

More recent attempts to incorporate Tfm into an EGFP mutant consisting of only two Met residues resulted in only poor substitution efficiencies.[145] During spectral characterization, no shoulders or shifts were detectable in the spectra giving rise to the assumption that Tfm does not affect the spectral features of the protein. The expression yielded enough protein to perform NMR studies and peak assignment (Fig. 5). Furthermore, the experiment indicated an effective blockage of the N-terminal methionine excision by Tfm. In further experiments, the incorporation of Tfl into anxA5, b*, GFP and mini-IGFBP was carried out. Again, only marginal incorporation could be detected in all these model proteins. However, TTL served as a suitable model protein for the incorporation of Dfm which could be incorporated with a good efficiency. Tfm, again, was incorporated only less efficient (unpublished work of the authors). Both congeners had a lower optimal temperature (60 °C compared to 70 °C of the parent protein) and were in general less active. In addition, unpublished data of the authors revealed an incorporation efficiency of approx. 60% of Tfm into anxA5. This was achieved by the co-expression of *E. coli* MetRS.

Very recently, the lab of Marx reported the successful incorporation of Tfm into a DNA polymerase, namely the N-terminally truncated KlenTaq DNA polymerase I of *Thermus*

aquaticus.[159] This mutant consists of 540 amino acids of which 14 are Met residues including the N-terminal one. The efficiency of incorporation was 82%. The congener was an active enzyme with selectivity comparable to the parent protein. However, the activity was only one third of the parent protein and the congener also lost thermal stability. Furthermore, in contrast to a related (4*R*-F)Pro congener (*vide supra*) the Tfm congener was only active in high concentrations or in the presence of high substrate concentrations. Interestingly, in the ^{19}F NMR analysis only 9 signals were detectable but the authors could observe conformational changes after DNA template and subsequent ddCPT addition, respectively, by detecting shifts of the resonances.

Another non-canonical amino acid attracting the attention of researchers is Tfl. For example, Tang and co-workers reported its *in vivo* incorporation into leucine zipper protein A1.[160] The 73 amino acid containing model protein consists of 8 Leu residues which were replaced by Tfl with an efficiency of 92%. The yield of the congener was only 50% of that of the parent protein (whereas lower incorporation levels yielded more isolated protein, but attempts to incorporate Hfl failed completely, *vide infra*). Although the electrophoretic mobility increased, the CD profiles of parent protein and congeners remained almost identical and indicated a high helical content (90%). Substituting the d-positions of the leucine zipper with other canonical amino acids resulted in a reduction of the coiled-coil stability. In contrast, the fluorinated congeners showed a higher resistance to thermal denaturation, reflected in melting temperatures up to 13 K higher compared to the parent protein. The same holds true for the chemical denaturation where the midpoint concentration of urea increased from 2.8 M up to 7 M in case of the high-level fluorinated protein. It should be noted, that already less substituted proteins show the same trends of increased stability. For peptides synthesized *in vitro*, Kumar and co-workers observed similar trends.[15] In contrast, the introduction of Tfl into the chloramphenicol acetyltransferase (CAT) led to a decreased thermostability as well as chemostability (*vide supra*).[110] Since fluorination resulted in a more highly ordered protein as judged by CD spectroscopy, it can be taken as indication that better structured biopolymers do not have to be automatically more stable. In order to tackle the problem of the decreased stability of CAT [Tfl], Montclare and Tirrell combined non-canonical amino acid incorporation and directed evolution.[161] Their evolutionary approach was set up to overcome the 20-fold reduction in $t_{1/2}$ of thermal inactivation at 60 °C. Already after two rounds of random mutagenesis and screening they were able to identify three mutations compensating the loss in thermostability caused by fluorination. A similar approach was used for the incorporation of Tfl into a GFP mutant as mentioned above.[148] Here, the substitution of 78% of the Leu residues led to a vanishing of the protein's fluorescence as the protein was not correctly folded. After 11 rounds of evolution the incorporation efficiency could be increased to 87% and – even more importantly – the fluorescence could be restored. In addition, the refolding rate after acid denaturation was enhanced by the evolution.

The Tfl constitution isomer 5-trifluoroisoleucine (5-Tfi, (5-CF_3)Ile) has been successfully used for incorporation experiments, for instance into murine dihydrofolate reductase (mDHFR) and a cytokine murin interleukin-2 (mIL-2) consisting of 14 and 5 Ile residues, respectively.[162] The level of incorporation was appreciably high (93 and 85%), which is surprising as the specific constant for the activation of 5-Tfi by the *E. coli* isoleucyl-tRNA synthetase is two orders of magnitude lower than for isoleucine. The congener of mIL-2 was further investigated and it turned out that the concentration for 50% of the maximal proliferative response is increased by approx. 30% compared to the parent protein. Any attempt to incorporate 3-trifluoroisoleucine (3-Tfi, (3-CF_3)Ile) failed. In contrast, incorporation of 6-trifluoronorleucine (Tfn, (6-F_3)Nle) could be performed after circumventing some restrictions. Tfn itself does not support protein biosynthesis in *E. coli*. On the other hand, norleucine (Nle) is one of the most effective Met substituents known to date. In order to detect Tfn incorporation Yoo and Tirrell had to engineer *E. coli's* MetRS by saturation mutagenesis with a GFP mutant as reporter protein.[147] By a combined positive and negative screening, they were able to identify a mutant capable of activating Tfn. Subsequently, Tfn was incorporated in very high efficiencies into different model proteins even if the specific rate constant (k_{cat}/k_m) of the mutated MetRS was sevenfold lower for Tfn compared to Met. This approach was also successfully transferred and applied to analogs without fluorine atoms.[163]

The successful incorporation of Hfl into leucine zipper protein A1 could be performed using a modified leucyl-tRNA synthetase of the host (*E. coli*).[164] The yield decreased to a fifth of the parent protein, but in the light of the related *in vitro* activation data this is reasonable. In the related assays it turned out that activation of Hfl is 4000-fold lower compared to leucine, whereas Tfl activation is only 242-fold lower. However, by overexpression of *E. coli* LeuRS 74% of all Leu residues could be replaced by Hfl. The resulting coiled-coil protein proved to be more stable compared to the one with incorporated Tfl in thermal denaturation analysis ($T_{m,Leu}$ = 54 °C, $T_{m,Tfl}$ = 67 °C, $T_{m,Hfl}$ = 76 °C). In addition, even in 8 M urea there is no complete unfolding of the biopolymer. The question arises again if this is due to a kind of fluorous effect or due to the increased hydrophobic character of the trifluoromethyl groups in the protein's core. Furthermore, only a minimal disruption of the folding and assembly of the coiled-coil architecture of the Hfl congener could be detected in comparison to the parent protein. The results obtained in these experiments are in accordance with related peptide studies.[12,14,165]

Interestingly, the first attempts to incorporate the two epimers of 4,4,4-trifluorovaline ((3-CF_3)Val) were unsuccessful.[166] Subsequent *in vitro* activation analyses revealed that the (3*S*-CF_3)Val epimer is not activated by ValRS or IleRS, but (3*R*-CF_3)Val is activated by both enzymes. However, the specificity constant is strongly decreased. The rate of activation is 600- (IleRS) and 2500-fold (ValRS) reduced compared to the natural substrates. Nonetheless, the authors could successfully incorporate (3*R*-CF_3)Val *in vivo* into mDHFR by the co-expression of IleRS or ValRS. In this way, 92% of all Ile residues or 86% of all Val residues could be substituted by (3*R*-CF_3)Val. This is the first striking example of two different canonical amino acids being exchanged by a single fluorinated amino acid.

Whereas the results mentioned above have been gained by supplementation based procedures, stop codon suppression methods can also be used for the incorporation of fluorinated amino acids as described above. Despite the fact that (2,3,5,6-

F_4,4-CH_3)Phe does not fulfill the definition of a fluorous molecule as introduced above (since it is an aromatic amino acid) it should be mentioned here.[86] Furthermore, (4-CF_3)Phe[75] is an aromatic acid bearing an aliphatic trifluoromethyl group, which is now widely used as probe in ^{19}F NMR analysis (*vide supra*).[80] Another example for the successful introduction of a CF_3 group by the SCS approach is the pyrrolysine derivative N^{ε}-trifluoroacetyl-L-lysine, which can be incorporated into target proteins in *S. cerevisiae*.[98]

It is obvious that the major deterrent for trifluoroleucine and hexafluoroleucine incorporation in larger proteins is their steric bulkiness and, thus, their difficult accommodation in the folded protein interior. Therefore, future studies will certainly face quite challenging questions such as: (i) which factors play crucial roles in the translation of trifluorinated amino acids; (ii) to which extent native proteins can be substituted with tri- and hexafluorinated amino acids (trifluoroleucine, trifluorovaline, and their hexafluoro counterparts, trifluoromethionine and trifluoronorleucine *etc.*); (iii) to which extent perfluorinated amino acids (all hydrogens substituted by fluorine) can be introduced to proteins; (iv) what can be learned about protein folding and the possibilities for rational protein design (retaining structure and function of native proteins) with such substances; and finally (v) whether it is possible to perform the *de novo* design of proteins exhibiting a 'fluorous effect'.

Multiple incorporation of different fluorinated amino acids

Whereas the incorporation of non-canonical amino acids with more than one fluorine atom is a first step in the direction of Teflon proteins, another one is the incorporation of different amino acid analogs at the same time.[167] Here, stop codon suppression methods and orthogonal ribosomes are quite restricted, *e.g.* because of the difficulties with more than one in-frame stop codon having to be suppressed and the competition between release factors and suppressor tRNAs. The auxotrophy-based approach offers a much simpler and more effective experimental set-up. In the first proof-of-principle experiment it was demonstrated by Lepthien and co-workers that three amino acids can be substituted *in vivo* in a single expression experiment.[28] In this way, barstar (10 kDa) could be equipped with a reactive handle (homopropargylglycine), a fluorescent tag (4-azatryptophan) and a stabilizing non-canonical amino acid ((4*S*-F)Pro). This approach was further investigated by the triple incorporation of three monofluorinated amino acids into the lipase of *Thermoanaerobacter thermohydrosulfuricus*.[29] The amino acids of choice were (4*S*-F)Pro, (4-F)Phe and (6-F)Trp to substitute Pro, Phe and Trp. The His-tagged protein consists of 267 amino acid of which 6 are Pro, 16 Phe and 2 Trp. This means, 10% of the protein has been fluorinated. The species with all 24 residues substituted was the dominant signal in the ESI MS spectrum but also species with only 23 and 22 fluorine atoms were detectable. The overall yield was approx. one fourth of that of the parent protein. MS/MS analysis of the tryptic fragments revealed a uniform distribution of fluorine through the whole protein. Regarding the spectral features of the protein, the characteristic "fluorophenyl-fingers" of (4-F)Phe[72] could be detected by UV/Vis spectroscopy whereas the fluorescence emission spectrum is only slightly affected by fluorination. This is, again, in accordance with the literature.[64] As revealed by CD spectroscopy, the effects on secondary structure upon fluorination were only marginal. The most prominent differences between parent protein and congener were observed in the catalytic activity which is decreased by approx. 40%. In addition, the congener has a decreased optimal temperature (60 °C instead of 70 °C) which corresponds to the values obtained for Dfm and Tfm incorporation. Also four amino acids could be substituted at the same time. In this experiment, (4*S*-F)Pro, (2-F)Phe or (3-F)Phe, (4-F)-Trp and (2-F)Tyr were incorporated into TTL (Fig. 9). Again, a

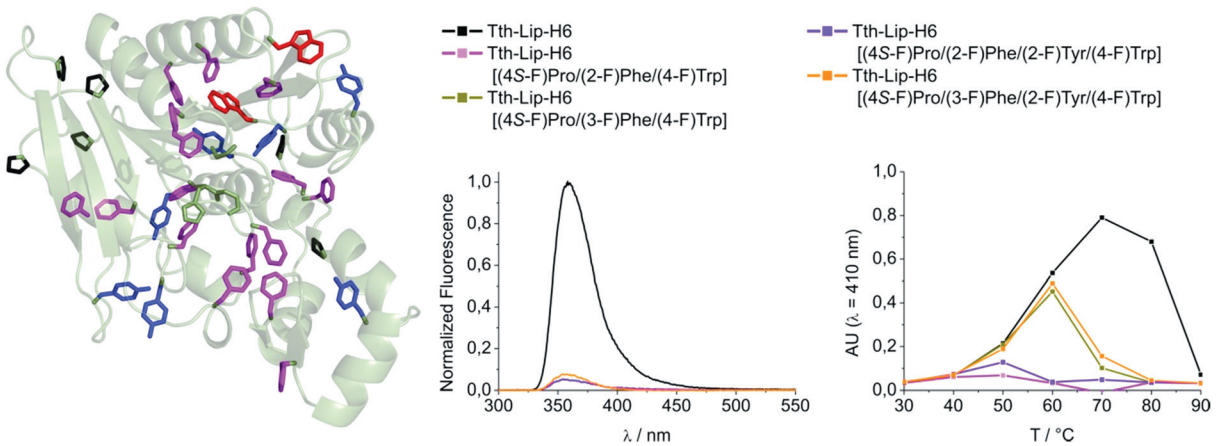


Fig. 9 Parallel incorporation of various fluorine-containing amino acid analogs in a single fermentation experiment. *Left*: Predicted structure of the TTL[111] with highlighted canonical residues substituted by their monofluorinated analogs (black: Pro → (4*S*-F)Pro, red: Trp → (4-F)Trp, magenta: Phe → (2-F)Phe or (3-F)Phe, blue: Tyr → (2-F)Tyr). Congeners with either triple or quadruple residue-specific substitution were generated. *Middle*: Fluorescence emission profiles of the parent protein and the congeners with four different fluorinated amino acids. As expected, the congeners show almost no fluorescence activity due to the incorporation of (4-F)Trp.[59] *Right*: Results of the activity assay studying the parent protein as well as the triple and quadruple congeners. The most striking observation is the correlation of the fluorine position in fluorinated Phe with the activities of the related congeners. Whereas (2-F)Phe bearing proteins are almost not active the congeners harboring (3-F)Phe are approx. 60% as active as the parent protein regardless if three or four types of canonical amino acids are substituted.

very high substitution level could be reached. Most interestingly, the regioisomers of fluorophenylalanine have different effects on the catalytic activity of the enzyme. Triple and quadruple congeners containing (3-F)Phe again have a T_{opt} of 60 °C and an overall activity lowered by approx. 40% compared to the parent protein. In contrast, the (2-F)Phe congeners display very little activity. Thus, we are far away from understanding all the details related to the incorporation of fluorine into enzymes and understanding the resulting effects.

Making fluorine an element of life: what is there to do?

The use of fluorinated amino acids in protein engineering yielded the development of several valuable tools, for example in spectroscopy (NMR analysis of proteins, fluorescence *etc.*) and for the identification of residues important for the catalytic activity of enzymes. But fluorine containing amino acids were also used to study the intracellular accumulation of non-canonical molecules.[42,168] Nonetheless, we are still far away from fundamentally understanding the effects of fluorinated amino acid analogs on the structural context and the catalytic performance of proteins and enzymes. So far, it is not possible to predict the effect of an H → F substitution in a protein. For instance, whereas some proteins are stabilized by the introduction of (4*S*-F)Pro, others are destabilized. On the other hand, the regioisomers of fluorotyrosine and fluorophenylalanine affect different enzymes in a different manner. Thus, more research is indeed needed to obtain novel data and extend existing knowledge in order to gain a much more general view on the effects of carbon-bond fluorine in complex biological systems. There should be no doubt that an investigation of the correlation between structural context, incorporation/translation efficiency and catalytic activity will be beneficial, especially in biocatalyst development. Along these lines, preliminary statements can be made regarding the incorporation of fluoroprolines and the dependence of favored *cis* or *trans* peptidyl prolyl bonds in the parent protein. Our group was among the first reporting a systematic study on the sensitivity of the translation machinery for chemical and chiral isosteric (isostructural), proline analogs that are incorporated into various proteins.[27,111] In particular, we and others [126,131] have discovered the stereochemical bias for the protein folding with fluoroprolines which is directly related to Pro puckering conformation. We are currently performing detailed examinations with various protein sequences in order to elucidate a mechanistic basis for the plasticity of the ribosome for fluorinated substrates.

Finally, the question 'How far can we go?' remains unanswered. How much fluorination would be tolerated in complex systems such as proteins? In order to tackle this problem some fundamental aspects of fluorocarbons should be taken into consideration. First, the source of the enhanced stability (measured as increased resistance against chemical or thermal unfolding) is not necessarily arising from favorable fluorocarbon–fluorocarbon interaction(s) or enhanced hydrophobicity. Marsh and associates argue that it could be "adequately explained by the increases in buried hydrophobic surface area and volume that accompany fluorination".[171] Namely, if we consider fluorous interactions as favorable van der Waals-type interactions between fluorocarbon residues, the question is how strong are these interactions in the context of the complex protein structure? Taking into account that hydrocarbons are more polarizable than fluorocarbons (since the cohesive dispersion forces between two hydrocarbon molecules are greater than between two fluorocarbons) it should not surprising that we still lack solid evidence for this; *e.g.* self-segregating properties of per-fluorocarbon-containing proteins.

Taking all these considerations into account, one can ask: Is it the biosynthesis of "Teflon-proteins" a realistic prospect? Given the case that living beings never adopted fluorine as biogenic element, its accommodation into the protoplasmic chemistry is indeed a formidable challenge. Living organisms would have to be able to survive adaption on fluorine through massive modifications of their enzymes and proteins that are originally evolved on hydrocarbon basis.[24] This certainly requires the rewriting of their entire genomic text by the accumulation of different types of mutations and their combinations. Also evolution of the sequence composition of the proteins of interest will be most probably always required to some extent (as performed by Tirrell and co-workers for the Tfl harboring GFP mutant). Therefore, we believe that the design of artificial cells with fluorine chemistry is a very challenging but achievable goal. On the other hand, if this problem is insurmountable, we need to find out why.

In combination with the works mentioned in this review, the experimental work on microbial cell evolution may lead to cells whose proteome would consist of a novel chemical composition and, thus, being real novel synthetic life forms. In this context it is worth to take a look to the related field of chemical modification of genomes. Only very recently, Marlière and co-workers succeeded in the incorporation of 5-chlorouracil instead of thymine into a genome.[169] The *E. coli* strain whose progeny retained the ability to use 5-chlorouracil is the first *in vivo* proof of a canonical–non-canonical border crossing in the field of nucleic acid research. This success in halogenation of a genome implies the question whether halogenation, *i.e.* fluorination, of a proteome is possible. Would it be possible to design cells that tolerate widely fluorinated proteomes? Would this interfere with the catalytic activity of too many enzymes? What about general metabolic protein turnover and basic metabolism in the presence of fluorine?

There is already solid evidence in the literature that the proteome-wide introduction of fluorine could be possible. Till now, SeMet is the only amino acid which can be incorporated into the whole proteome (as proved by Cowie and Cohen[26]), but several attempts have been reported to increase the number of proteome-wide incorporated amino acids. One of the most important experiments was reported in 1983 by Wong, who selected for microorganisms capable of changing the intrinsic preference for amino acids.[56] He started his experimental evolution with the Trp auxotrophic *Bacillus subtilis* strain QB928 which was serially mutated and cultured for many generations under defined conditions with a medium containing (4-F)Trp. Finally, he could demonstrate that the mutant strain prefers (4-F)Trp over canonical Trp. This microorganism was described as a "first free living organism in the past couple of billion years to have learnt to thrive on a genetic code that departs from the universal code".[56] The ratio behind the choice of (4-F)Trp as non-canonical amino acid was the fact that natural TrpRS cannot distinguish between Trp and the derivative since they have almost the same affinity

for this enzyme in activation and aminoacylation reactions. Later, Ellington and Bacher reported two experiments in Wong's tradition in order to generate "unnatural organisms".[57,58] Their first report deals with the evolution of *E. coli* strains which would be capable for surviving in (4-F)Trp containing medium. They assumed that only a few proteins will be adversely affected, and that mutation can circumvent these problems. Following this line, (4-F)Trp should be inserted into the proteome. However, the resulting strains retained a requirement for Trp, even if the required concentration was very low. Thus, they concluded that further re-optimization of proteins and metabolic pathways are necessary in order to accommodate only (4-F)Trp instead of Trp. For their next experiments, they switched to the bacteriophage Qβ as the model system. Its replication was performed in an *E. coli* strain with Trp auxotrophy. Again, the medium contained different monofluorinated Trp analogs. But as before, Qβ retained an absolute requirement for traces of Trp. Thus, the question arises how to select an organism by nutrient pressure that becomes dependent on a non-canonical amino acid. The very recent work of Wong who extended his examinations on fluorinated Trp analogs gives at least one possible answer.[170]

Finally it should be also mentioned that living beings are indeed capable of developing enzymes for a carbon–fluorine bond formation.[7] This finding indicates that metabolic engineering with organic fluorine should be conceivable as well. In a long term perspective, it might be possible to couple newly designed fluorine-related metabolic pathways with reprogrammed protein translation or other processes (*e.g.* synthesis of biological membranes or signal transduction pathways). The synergy of these attempts and the experimental work on strain evolution and protein engineering may lead to cells whose proteomes consist of a novel chemical composition and, thus, being real novel synthetic life forms.

References

1 I. Ojima, *Fluorine in Medicinal Chemistry and Chemical Biology*, John Wiley & Sons, 2009.
2 J.-P. Bégué and D. Bonnet-Delpon, *Bioorganic and Medicinal Chemistry of Fluorine*, John Wiley & Sons, 2008.
3 J. D. Dunitz, *ChemBioChem*, 2004, **5**, 614–621.
4 N. C. Yoder and K. Kumar, *Chem. Soc. Rev.*, 2002, **31**, 335–341.
5 C. Jäckel and B. Koksch, *Eur. J. Org. Chem.*, 2005, 4483–4503.
6 D. B. Harper and D. O'Hagan, *Nat. Prod. Rep.*, 1994, **11**, 123–123.
7 D. O'Hagan, C. Schaffrath, S. L. Cobb, J. T. G. Hamilton and C. D. Murphy, *Nature*, 2002, **416**, 279.
8 S. K. Holmgren, L. E. Bretscher, K. M. Taylor and R. T. Raines, *Chem. Biol.*, 1999, **6**, 63–70.
9 C. Renner, S. Alefelder, J. H. Bae, N. Budisa, R. Huber and L. Moroder, *Angew. Chem., Int. Ed.*, 2001, **40**, 923–925.
10 K.-H. Lee, H.-Y. Lee, M. M. Slutsky, J. T. Anderson and E. N. G. Marsh, *Biochemistry*, 2004, **43**, 16277–16284.
11 H. Y. Lee, K. H. Lee, H. M. Al-Hashimi and E. N. Marsh, *J. Am. Chem. Soc.*, 2006, **128**, 337–343.
12 B. C. Buer, R. de la Salud-Bea, H. M. Al Hashimi and E. N. G. Marsh, *Biochemistry*, 2009, **48**, 10810–10817.
13 H.-P. Chiu, Y. Suzuki, D. Gullickson, R. Ahmad, B. Kokona, R. Fairman and R. P. Cheng, *J. Am. Chem. Soc.*, 2006, **128**, 15556–15557.
14 L. M. Gottler, R. de la Salud-Bea and E. N. G. Marsh, *Biochemistry*, 2008, **47**, 4484–4490.
15 B. Bilgicer, A. Fichera and K. Kumar, *J. Am. Chem. Soc.*, 2001, **123**, 4393–4399.
16 Y. Tang, G. Ghirlanda, N. Vaidehi, J. Kua, D. T. Mainz, W. A. Goddard, W. F. DeGrado and D. A. Tirrell, *Biochemistry*, 2001, **40**, 2790–2796.
17 J.-C. Horng and D. P. Raleigh, *J. Am. Chem. Soc.*, 2003, **125**, 9286–9287.
18 C. Jäckel, W. Seufert, S. Thust and B. Koksch, *ChemBioChem*, 2004, **5**, 717–720.
19 C. Jäckel, M. Salwiczek and B. Koksch, *Angew. Chem., Int. Ed.*, 2006, **45**, 4198–4203.
20 D. O'Hagan and H. S. Rzepa, *Chem. Commun.*, 1997, 645–652.
21 M. Salwiczek and B. Koksch, *ChemBioChem*, 2009, **10**, 2867–2870.
22 M. G. Woll, E. B. Hadley, S. Mecozzi and S. H. Gellman, *J. Am. Chem. Soc.*, 2006, **128**, 15932–15933.
23 H. Zheng, K. Comeforo and J. Gao, *J. Am. Chem. Soc.*, 2009, **131**, 18–19.
24 N. Budisa, *Angew. Chem., Int. Ed.*, 2004, **43**, 6426–6463.
25 N. Voloshchuk and J. K. Montclare, *Mol. BioSyst.*, 2010, **6**, 65–80.
26 D. B. Cowie and G. N. Cohen, *Biochim. Biophys. Acta*, 1957, **26**, 252–261.
27 T. Steiner, P. Hess, J. H. Bae, B. Wiltschi, L. Moroder and N. Budisa, *PLoS One*, 2008, **3**, e1680.
28 S. Lepthien, L. Merkel and N. Budisa, *Angew. Chem., Int. Ed.*, 2010, **49**, 5446–5450.
29 L. Merkel, M. Schauer, G. Antranikian and N. Budisa, *ChemBioChem*, 2010, **11**, 1505–1507.
30 A. Link, M. L. Mock and D. A. Tirrell, *Curr. Opin. Biotechnol.*, 2003, **14**, 603–609.
31 N. Budisa, C. Minks, S. Alefelder, W. Wenger, F. Dong, L. Moroder and R. Huber, *FASEB J.*, 1999, **13**, 41–51.
32 J. A. Johnson, Y. Y. Lu, J. A. Van Deventer and D. A. Tirrell, *Curr. Opin. Chem. Biol.*, 2010, **14**, 774–780.
33 R. Furter, *Protein Sci.*, 1998, **7**, 419–426.
34 T. S. Young, I. Ahmad, J. A. Yin and P. G. Schultz, *J. Mol. Biol.*, 2010, **395**, 361–374.
35 M. Levine and H. Tarver, *J. Biol. Chem.*, 1951, **192**, 835–850.
36 R. Munier and G. N. Cohen, *Biochim. Biophys. Acta*, 1959, **31**, 378–391.
37 E. W. Westhead and P. D. Boyer, *Biochim. Biophys. Acta*, 1961, **54**, 145–156.
38 M. P. Gamcsik and J. T. Gerig, *FEBS Lett.*, 1986, **196**, 71–74.
39 M. H. Richmond, *Bacteriol. Rev.*, 1962, **26**, 398–420.
40 F. Chapeville, F. Lipmann, G. n. von Ehrenstein, B. Weisblum, W. J. Ray and S. Benzer, *Proc. Natl. Acad. Sci. U. S. A.*, 1962, **48**, 1086–1092.
41 O. M. Rennert and H. S. Anker, *Biochemistry*, 1963, **2**, 471–476.
42 T. W. Conway, E. M. Lansford and W. Shive, *J. Bacteriol.*, 1963, **85**, 141–149.
43 B. D. Sykes, H. I. Weingarten and M. J. Schlesinger, *Proc. Natl. Acad. Sci. U. S. A.*, 1974, **71**, 469–473.
44 E. A. Pratt and C. Ho, *Biochemistry*, 1975, **14**, 3035–3040.
45 F. Evanics, J. Kitevski, I. Bezsonova, J. Formankay and R. Prosser, *Biochim. Biophys. Acta, Gen. Subj.*, 2007, **1770**, 221–230.
46 S. L. Cobb and C. D. Murphy, *J. Fluorine Chem.*, 2009, **130**, 132–143.
47 J. T. Gerig, *Prog. Nucl. Magn. Reson. Spectrosc.*, 1994, **26**, 293–370.
48 M. A. Danielson and J. J. Falke, *Annu. Rev. Biophys. Biomol. Struct.*, 1996, **25**, 163–195.
49 J. Feeney, J. E. McCormick, C. J. Bauer, B. Birdsall, C. M. Moody, B. A. Starkmann, D. W. Young, P. Francis, R. H. Havlin, W. D. Arnold and E. Oldfield, *J. Am. Chem. Soc.*, 1996, **118**, 8700–8706.
50 G. R. Winkler, S. B. Harkins, J. C. Lee and H. B. Gray, *J. Phys. Chem. B*, 2006, **110**, 7058–7061.
51 C. Li, E. A. Lutz, K. M. Slade, R. A. S. Ruf, G.-F. Wang and G. J. Pielak, *Biochemistry*, 2009, **48**, 8578–8584.
52 D. C. Klein and J. L. Weller, *J. Pharmacol. Exp. Ther.*, 1973, **186**, 516–527.
53 D. C. Klein, K. L. Kirk, J. L. Weller, T. Oka, A. Parfitt and I. S. Owens, *Mol. Pharmacol.*, 1976, **12**, 720–730.
54 D. C. Klein, J. L. Weller, K. L. Kirk and R. W. Hartley, *Mol. Pharmacol.*, 1977, **13**, 1105–1110.
55 H. C. Taylor, D. C. Richardson, J. S. Richardson, A. Wlodawer, A. Komoriya and I. M. Chaiken, *J. Mol. Biol.*, 1981, **149**, 313–317.
56 J. T. Wong, *Proc. Natl. Acad. Sci. U. S. A.*, 1983, **80**, 6303–6306.
57 J. M. Bacher and A. D. Ellington, *J. Bacteriol.*, 2001, **183**, 5414–5425.
58 J. M. Bacher, J. J. Bull and A. D. Ellington, *BMC Evol. Biol.*, 2003, **3**, 24.
59 P. M. Bronskill and J. T. Wong, *Biochem. J.*, 1988, **249**, 305–308.
60 C. Y. Wong and M. R. Eftink, *Biochemistry*, 1998, **37**, 8938–8946.
61 C. Y. Wong and M. R. Eftink, *Biochemistry*, 1998, **37**, 8947–8953.

62 H. Duewel, E. Daub, V. Robinson and J. F. Honek, *Biochemistry*, 1997, **36**, 3404–3416.
63 F. Colombani, H. Cherest and H. de Robichon-Szulmajster, *J. Bacteriol.*, 1975, **122**, 375–384.
64 C. Minks, R. Huber, L. Moroder and N. Budisa, *Biochemistry*, 1999, **38**, 10649–10659.
65 C. Minks, S. Alefelder, L. Moroder, R. Huber and N. Budisa, *Tetrahedron*, 2000, **56**, 9431–9442.
66 M. Ring, I. M. Armitage and R. E. Huber, *Biochem. Biophys. Res. Commun.*, 1985, **131**, 675–680.
67 M. Ring and R. E. Huber, *Biochem. Cell Biol.*, 1993, **71**, 127–132.
68 B. Brooks, R. S. Phillips and W. F. Benisek, *Biochemistry*, 1998, **37**, 9738–9742.
69 G. Xiao, J. Parsons, K. Tesh, R. N. Armstrong and G. Gilliand, *J. Mol. Biol.*, 1998, **281**, 323–339.
70 M. R. Seyedsayamdost, C. S. Yee and J. Stubbe, *Nat. Protoc.*, 2007, **2**, 1225–1235.
71 M. R. Seyedsayamdost, S. Y. Reece, D. G. Nocera and J. Stubbe, *J. Am. Chem. Soc.*, 2006, **128**, 1569–1579.
72 C. Minks, R. Huber, L. Moroder and N. Budisa, *Anal. Biochem.*, 2000, **284**, 29–34.
73 J. A. Ellman, D. Mendel and P. G. Schultz, *Science*, 1992, **255**, 197–200.
74 J. A. Ellman, B. F. Volkman, D. Mendel, P. G. Schulz and D. E. Wemmer, *J. Am. Chem. Soc.*, 1992, **114**, 7959–7961.
75 J. C. Jackson, J. T. Hammill and R. A. Mehl, *J. Am. Chem. Soc.*, 2007, **129**, 1160–1166.
76 L. Wang, A. Brock, B. Herberich and P. G. Schultz, *Science*, 2001, **292**, 498–500.
77 P. M. Haggie, *J. Biol. Chem.*, 1999, **274**, 3941–3945.
78 S. P. Williams, P. M. Haggie and K. M. Brindle, *Biophys. J.*, 1997, **72**, 490–498.
79 S. P. Williams, A. M. Fulton and K. M. Brindle, *Biochemistry*, 1993, **32**, 4895–4902.
80 C. Li, G.-F. Wang, Y. Wang, R. Creager-Allen, E. A. Lutz, H. Scronce, K. M. Slade, R. A. S. Ruf, R. A. Mehl and G. J. Pielak, *J. Am. Chem. Soc.*, 2010, **132**, 321–327.
81 U. H. N. Dürr, S. L. Grage, R. Witter and A. S. Ulrich, *J. Magn. Reson.*, 2008, **191**, 7–15.
82 P. Shi, Z. Xi, H. Wang, C. Shi, Y. Xiong and C. Tian, *Biochem. Biophys. Res. Commun.*, 2010, **402**, 461–466.
83 P. Shi, H. Wang, Z. Xi, C. Shi, Y. Xiong and C. Tian, *Protein Sci.*, 2011, **20**, 224–228.
84 A. L. Stokes, S. J. Miyake-Stoner, J. C. Peeler, D. P. Nguyen, R. P. Hammer and R. A. Mehl, *Mol. BioSyst.*, 2009, **5**, 1032–1038.
85 J.-D. Pédelacq, S. p. Cabantous, T. Tran, T. C. Terwilliger and G. S. Waldo, *Nat. Biotechnol.*, 2005, **24**, 79–88.
86 S. J. Miyake-Stoner, C. A. Refakis, J. T. Hammill, H. Lusic, J. L. Hazen, A. Deiters and R. A. Mehl, *Biochemistry*, 2010, **49**, 1667–1677.
87 J. M. Xie and P. G. Schultz, *Nat. Rev. Mol. Cell Biol.*, 2006, **7**, 775–782.
88 S. E. Cellitti, D. H. Jones, L. Lagpacan, X. S. Hao, Q. Zhang, H. Y. Hu, S. M. Brittain, A. Brinker, J. Caldwell, B. Bursulaya, G. Spraggon, A. Brock, Y. Ryu, T. Uno, P. G. Schultz and B. H. Geierstanger, *J. Am. Chem. Soc.*, 2008, **130**, 9268–9281.
89 D. H. Jones, S. E. Cellitti, X. S. Hao, Q. Zhang, M. Jahnz, D. Summerer, P. G. Schultz, T. Uno and B. H. Geierstanger, *J. Biomol. NMR*, 2009, **46**, 89–100.
90 B. J. Wilkins, S. Marionni, D. D. Young, J. Liu, Y. Wang, M. L. Di Salvo, A. Deiters and T. A. Cropp, *Biochemistry*, 2010, **49**, 1557–1559.
91 D. D. Young and A. Deiters, *Org. Biomol. Chem.*, 2007, **5**, 999–1005.
92 G. Mayer and A. Heckel, *Angew. Chem., Int. Ed.*, 2006, **45**, 4900–4921.
93 H.-M. Lee, D. R. Larson and D. S. Lawrence, *ACS Chem. Biol.*, 2009, **4**, 409–427.
94 A. Deiters, D. Groff, Y. H. Ryu, J. M. Xie and P. G. Schultz, *Angew. Chem., Int. Ed.*, 2006, **45**, 2728–2731.
95 P. P. Pal, J. H. Bae, M. K. Azim, P. Hess, R. Friedrich, R. Huber, L. Moroder and N. Budisa, *Biochemistry*, 2005, **44**, 3663–3672.
96 J. H. Bae, P. P. Pal, L. Moroder, R. Huber and N. Budisa, *ChemBioChem*, 2004, **5**, 720–722.
97 D. D. Young, T. S. Young, M. Jahnz, I. Ahmad, G. Spraggon and P. G. Schultz, *Biochemistry*, 2011, **50**, 1894–1900.
98 S. M. Hancock, R. Uprety, A. Deiters and J. W. Chin, *J. Am. Chem. Soc.*, 2010, **132**, 14819–14824.
99 E. A. Pratt, J. A. Jones, P. F. Cottam, S. R. Dowd and C. Ho, *Biochim. Biophys. Acta, Biomembr.*, 1983, **729**, 167–175.
100 O. B. Peersen, E. A. Pratt, T. Nguyen Truong Hoai, C. Ho and G. S. Rule, *Biochemistry*, 1990, **29**, 3256–3262.
101 Z.-Y. Sun, H.-T. N. Truong, E. A. Pratt, D. C. Sutherland, C. E. Kulig, R. J. Homer, S. M. Groetsch, P. Y. Hsue and C. Ho, *Protein Sci.*, 1993, **2**, 1938–1947.
102 M. A. Dominguez, K. C. Thornton, M. G. Melendez and C. M. Dupureur, *Proteins: Struct., Funct., Genet.*, 2001, **45**, 55–61.
103 J. S. Thorson, E. Chapman, E. C. Murphy, P. G. Schultz and J. K. Judice, *J. Am. Chem. Soc.*, 1995, **117**, 1157–1158.
104 L. A. Luck and J. J. Falke, *Biochemistry*, 1991, **30**, 4248–4256.
105 N. Voloshchuk, A. Y. Zhu, D. Snydacker and J. K. Montclare, *Bioorg. Med. Chem. Lett.*, 2009, **19**, 5449–5451.
106 N. Budisa, W. Wenger and B. Wiltschi, *Mol. BioSyst.*, 2010, **6**, 1630–1639.
107 P. J. Baker and J. K. Montclare, *ChemBioChem*, 2011, **12**, 1845–1848.
108 F. Khan, I. Kuprov, T. D. Craggs, P. J. Hore and S. E. Jackson, *J. Am. Chem. Soc.*, 2006, **128**, 10729–10737.
109 M. H. J. Seifert, D. Ksiazek, M. K. Azim, P. Smialowski, N. Budisa and T. A. Holak, *J. Am. Chem. Soc.*, 2002, **124**, 7932–7942.
110 T. Panchenko, W. W. Zhu and J. K. Montclare, *Biotechnol. Bioeng.*, 2006, **94**, 921–930.
111 M. G. Hoesl, C. G. Acevedo-Rocha, S. Nehring, M. Royter, C. Wolschner, B. Wiltschi, N. Budisa and G. Antranikian, *ChemCatChem*, 2011, **3**, 213–221.
112 J. a. J. Panek, T. R. Ward, A. Jezierska and M. Novic, *J. Mol. Model.*, 2008, **15**, 257–266.
113 J. T. Gerig and R. S. McLeod, *J. Am. Chem. Soc.*, 1973, **95**, 5725–5729.
114 E. S. Eberhardt, N. Panasik and R. T. Raines, *J. Am. Chem. Soc.*, 1996, **118**, 12261–12266.
115 M. D. Shoulders and R. T. Raines, *Annu. Rev. Biochem.*, 2009, **78**, 929–958.
116 T.-Y. Zheng, Y.-J. Lin and J.-C. Horng, *Biochemistry*, 2010, **49**, 4255–4263.
117 Y. C. Chiang, Y. J. Lin and J. C. Horng, *Protein Sci.*, 2009, **18**, 1967–1977.
118 C. Boulègue, A. G. Milbradt, C. Renner and L. Moroder, *J. Mol. Biol.*, 2006, **358**, 846–856.
119 J. C. Horng, *Protein Sci.*, 2006, **15**, 74–83.
120 M. D. Shoulders, K. A. Satyshur, K. T. Forest and R. T. Raines, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 559–564.
121 J. Uitto and D. J. Prockop, *Biochim. Biophys. Acta, Protein Struct.*, 1974, **336**, 234–251.
122 R. Golbik, C. Yu, E. Weyher-Stingl, R. Huber, L. Moroder, N. Budisa and C. Schiene-Fischer, *Biochemistry*, 2005, **44**, 16026–16034.
123 D. Naduthambi and N. J. Zondlo, *J. Am. Chem. Soc.*, 2006, **128**, 12430–12431.
124 K. L. Gorres, R. Edupuganti, G. R. Krow and R. T. Raines, *Biochemistry*, 2008, **47**, 9447–9455.
125 M. Tandon, M. Wu, T. P. Begley, J. Myllyharju, A. Pirskanen and K. Kivirikko, *Bioorg. Med. Chem. Lett.*, 1998, **8**, 1139–1144.
126 W. Kim and V. P. Conticello, *Polym. Rev.*, 2007, **47**, 93–119.
127 W. Kim, A. George, M. Evans and V. P. Conticello, *ChemBioChem*, 2004, **5**, 928–936.
128 W. Kim, R. A. McMillan, J. P. Snyder and V. P. Conticello, *J. Am. Chem. Soc.*, 2005, **127**, 18121–18132.
129 S. Grothe, R. L. Krogsrud, D. J. McClellan, J. L. Milner and J. M. Wood, *J. Bacteriol.*, 1986, **166**, 253–259.
130 D. D. Buechter, *J. Biol. Chem.*, 2002, **278**, 645–650.
131 W. Kim, K. I. Hardcastle and V. P. Conticello, *Angew. Chem., Int. Ed.*, 2006, **45**, 8141–8145.
132 C. A. Thomas, E. R. Talaty and J. G. Bann, *Chem. Commun.*, 2009, 3366–3368.
133 G. J. Palm and A. Wlodawer, *Spectral Variants of Green Fluorescent Protein*, Academic Press, 1999.
134 B. Holzberger and A. Marx, *J. Am. Chem. Soc.*, 2010, **132**, 15708–15713.
135 M. D. Crespo and M. Rubini, *PLoS One*, 2011, **6**, e19425.
136 F. Oldach, R. Al Toma, A. Kuthning, T. Caetano, S. Mendo, N. Budisa and R. D. Süßmuth, *Angew. Chem., Int. Ed.*, 2012, **51**, 415–418.
137 M. G. Hoesl, L. Merkel and N. Budisa, in *Fluorescent Proteins* ed. G. Jung, Springer, 2011.
138 N. Budisa, P. P. Pal, S. Alefelder, P. Birle, T. Krywcun, M. Rubini, W. Wenger, J. H. Bae and T. Steiner, *Biol. Chem.*, 2004, **385**, 191–202.

139 W. W. Ward, H. J. Prentice, A. F. Roth, C. W. Cody and S. C. Reeves, *Photochem. Photobiol.*, 1982, **35**, 803–808.
140 M. C. Chen, C. R. Lambert, J. D. Urgitis and M. Zimmer, *Chem. Phys.*, 2001, **270**, 157–164.
141 J. H. Bae, M. Rubini, G. Jung, G. Wiegand, M. H. J. Seifert, M. K. Azim, J.-S. Kim, A. Zumbusch, T. A. Holak, L. Moroder, R. Huber and N. Budisa, *J. Mol. Biol.*, 2003, **328**, 1071–1081.
142 R. Bonnett, A. Harriman and A. N. Kozyrev, *J. Chem. Soc., Faraday Trans.*, 1992, **88**, 763–769.
143 S. Veettil, N. Budisa and G. Jung, *Biophys. Chem.*, 2008, **136**, 38–43.
144 A. Goulding, S. Shrestha, K. Dria, E. Hunt and S. K. Deo, *Protein Eng., Des. Sel.*, 2008, **21**, 101–106.
145 N. Budisa, O. Pipitone, I. Siwanowicz, M. Rubini, P. P. Pal, T. A. Holak and M. L. Gelmi, *Chem. Biodiversity*, 2004, **1**, 1465–1475.
146 M. E. Houston, L. Harvath and J. F. Honek, *Bioorg. Med. Chem. Lett.*, 1997, **7**, 3007–3012.
147 T. H. Yoo and D. A. Tirrell, *Angew. Chem., Int. Ed.*, 2007, **46**, 5340–5343.
148 T. H. Yoo, A. J. Link and D. A. Tirrell, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 13887–13890.
149 E. N. G. Marsh, *Chem. Biol.*, 2000, **7**, R153–R157.
150 H.-P. Chiu, B. Kokona, R. Fairman and R. P. Cheng, *J. Am. Chem. Soc.*, 2009, **131**, 13192–13193.
151 J. A. Gladysz and D. P. Curran, *Tetrahedron*, 2002, **58**, 3823–3825.
152 H. S. Duewel, E. Daub, V. Robinson and J. F. Honek, *Biochemistry*, 2001, **40**, 13167–13176.
153 M. G. Hoesl, M. Larregola, H. Cui and N. Budisa, *J. Pept. Sci.*, 2010, **16**, 589–595.
154 H. S. Duewel and J. F. Honek, *J. Protein Chem.*, 1998, **17**, 337–350.
155 M. D. Vaughan, P. Cleve, V. Robinson, H. S. Duewel and J. F. Honek, *J. Am. Chem. Soc.*, 1999, **121**, 8475–8478.
156 T. Crepin, E. Schmitt, Y. Mechulam, P. B. Sampson, M. D. Vaughan, J. F. Honek and S. Blanquet, *J. Mol. Biol.*, 2003, **332**, 59–72.
157 M. D. Vaughan, P. B. Sampson, E. Daub and J. F. Honek, *Med. Chem.*, 2005, **1**, 227–237.
158 D. K. Garner, M. D. Vaughan, H. J. Hwang, M. G. Savelieff, S. M. Berry, J. F. Honek and Y. Lu, *J. Am. Chem. Soc.*, 2006, **128**, 15608–15617.
159 B. Holzberger, M. Rubini, H. M. Möller and A. Marx, *Angew. Chem., Int. Ed.*, 2010, **49**, 1324–1327.
160 Y. Tang, G. Ghirlanda, W. A. Petka, T. Nakajima, W. F. DeGrado and D. A. Tirrell, *Angew. Chem., Int. Ed.*, 2001, **40**, 1494–1496.
161 J. K. Montclare and D. A. Tirrell, *Angew. Chem., Int. Ed.*, 2006, **45**, 4518–4521.
162 P. Wang, Y. Tang and D. A. Tirrell, *J. Am. Chem. Soc.*, 2003, **125**, 6900–6906.
163 S. Nagasundarapandian, L. Merkel, N. Budisa, R. Govindan, N. Ayyadurai, S. Sriram, H. Yun and S.-G. Lee, *ChemBioChem*, 2010, **11**, 2521–2524.
164 Y. Tang and D. A. Tirrell, *J. Am. Chem. Soc.*, 2001, **123**, 11089–11090.
165 B. Bilgicer and K. Kumar, *Tetrahedron*, 2002, **58**, 4105–4112.
166 P. Wang, A. Fichera, K. Kumar and D. A. Tirrell, *Angew. Chem., Int. Ed.*, 2004, **43**, 3664–3666.
167 M. G. Hoesl and N. Budisa, *Angew. Chem., Int. Ed.*, 2011, **50**, 2896–2902.
168 C. Giese, S. Lepthien, L. Metzner, M. Brandsch, N. Budisa and H. Lilie, *ChemMedChem*, 2008, **3**, 1449–1456.
169 P. Marlière, J. Patrouix, V. Döring, P. Herdewijn, S. Tricot, S. p. Cruveiller, M. Bouzon and R. Mutzel, *Angew. Chem., Int. Ed.*, 2011, **50**, 7109–7114.
170 W.-K. Mat, H. Xue and J. T.-F. Wong, *PLoS One*, 2010, **5**, e12206.
171 B. C. Buer, J. L. Meagher, J. A. Stuckey and E. N. Marsh, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 4810–4815.