Max Planck Institut für Biochemie
AG Bioorganische Chemie
BioFuture Forschungsgruppe Molekulare Biotechnologie

EXPANDING THE CODED REPERTOIRE WITH FLUORINATED AMINO ACIDS

Prajna Paramita Pal

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zu Erlangung des akademischen Grades eines

Doktor der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. Stefan Glaser

Prüfer der Dissertation:

1. apl. Prof. Dr. L. Moroder

2. Univ.-Prof. Dr. J. Buchner

Die Dissertation wurde am 27.02.2006 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 22.03.2006 angenommen.

To Dida and Mama

With all my love

Zusammenfassung

Im Rahmen dieser Arbeit wurde der Einfluß von der Substitution aromatischer und aliphatischer Aminosäuren durch fluorinierter Analoge auf Proteine untersucht. Mit Hilfe fluorinierten Varianten von EGFP und β-Galaktosidase in Zellinien wurde das Konzept überprüft, daß so veränderte Proteine als diagnostische Marker dienen könnten, welche in Zellen in inaktiver „prodrug" Form eingeschleust werden und während der Einschleusung in die Zellen noch keine Toxizität ausüben.

Es konnte gezeigt werden, daß die Halogenierung von aromatischen Aminosäuren, speziell der Austausch von Tyrosin durch Fluorotyrosin, dramatische Veränderungen der spektralen Eigenschaften von Proteinen hervorruft, ohne die räumliche Struktur zu verändern. Diese spektralen Eigenschaften dienen als sensible Reporter für Wechselwirkungen innerhalb des Chromophores oder zwischen Chromophor und Proteinmatrix.

Es wurden verschiedene Varianten des Chromophores als auch der umgebenden Proteinmatrix durch klassische Methoden des „Protein-engineering" hergestellt. Mittels ortsspezificher Mutagenese in Kombination mit dem Einbau von monofluorinierten aromatischen Aminosäuren wurden neue Klassen von autofluoreszierenden Proteinen hergestellt. Mit dem Einbau von elektronegativen Fluor-Atomen in Proteine gelang es zum einen den Chromophor kovalent zu modifizieren, zum anderen elektrostatische Wechselwirkungen zwischen Chromophore und umgebender Proteinmatrix zu manipulieren.

In den Proteinen EGFP und EYFP aus der Qualle *Aequorea victoria* wurde Tyrosin durch 2-Fluorotyrosin ((2-F)Tyr oder *o*-FTyr) und 3-Fluorotyrosin ((3-F)Tyr oder *m*-FTyr) ersetzt wodurch der direkte Einfluß dieses Austausches auf die Proteineigenschaften untersucht werden konnte. Es wurde gezeigt, daß sich der pK_a-Werte der Proteine verändert sowie Änderungen im Anion/Kation-Äquilibrium bei Titrationsversuchen und Blau/Rot-Verschiebungen in Absorptions- und Fluoreszenzspektren auftreten. Diese Unterschiede traten verstärkt im fluorinierten EYFP auf und sind zurückzuführen auf die Fluorinierung von Tyr203, welches einzigartig in EYFP ist.

Anhand von Modellingstudien konnte vorgeschlagen werden, das in einem Konformationszustand von EGFP Platz für ein mögliches Chromophore „flipping" vorhanden ist, welches im EYFP nicht der Fall ist, da hier das Tyr203 aufgrund sterischer

Hindernisse durch überlappende Wechselwirkungen beeinträchtigt ist. Tatsächlich konnten in den 3D-Strukturen von (3-F)Tyr EGFP zwei Konformationszustände vom Fluor im Chromophor im Kristall nachgewiesen werden, ebenso wie in der Mehrheit der (3-F)Tyr Seitenketten des Proteins.
Erstaunlicherweise gibt es bei (2-F)TyrEGFP im Chromophore und anderen (2-F)Tyrresten nur einen Konformationszustand. Diese Eigenschaften sind zweifellos wesentliche Merkmale der fluorinierten Chromophore im Zusammenhang mit ihrer starren Proteinumgebung und sind verantwortlich für die einzigartigen Fluoreszenzeigenschaften von fluorinierten autofluoreszierenden Proteine. In einem zusätzlichen Versuch wurde die ausschlaggebende Aminosäure Tyr203 in EYFP durch Phe ersetzt und in *ortho*-, *para*- und *meta*-Position fluoriniert und die spektralen Eigenschaften dieses Proteins untersucht.
Untersuchungen von *ortho*- und *meta* Fluorinierungen des dsRed zeigten keine signifikanten Änderungen in den Absorptions- und Fluoreszenzspektren dieser Varianten, möglicherweise zurückzuführen auf einen nur teilweise erfolgten Ersatz der Analogen.

Ein weiterer Aspekt dieser Arbeit waren Inkorporationsversuche in GFP mit fluorinierten aliphatischen Aminosäuren wie Met und Leu, welche eine Trifluoromethyl-Gruppe besaßen. Die interessanteste Eigenschaft dieser nichtkanonischen Aminosäuren ist deren erhöhte Hydrophobizität (fast zweimal so hydrophob wie das kanonische Gegenstück).
Diese Besonderheit wird oft zur Verbesserung der Leistung bei Medikamenten im Unterdrücken metabolischer Toxizität oder beim Erhöhen der Verfügbarkeit im Beliefern von vielen Pharmazeutika in die Zellen benutzt. Bis zum heutigen Zeitpunkt konnten nur geringfügige Substitutionen von Leucin- und Methioninresten mit deren Trifluoro-Analogen in GFP erreicht werden. Durch Erhöhung des genomischen Level der Methionyl-tRNA-Synthetase, könnte das Substitutionsniveau erhöht werden, mit dem eine fast quantitative Inkorporation von Trifluoromethionin in das Protein Annexin V erreicht werden konnte.

ACKNOWLEDGMENTS

Prof. Luis Moroder is warmly acknowledged for giving me the opportunity to do this Ph D work in his group, and Prof. Dieter Oesterhelt and Prof. Robert Huber, for allowing me free access of the infrastructure. Prof. Padmanabhan Balaram (IISc) is thanked for giving me a foothold in science. Prof. Axel Ullrich and Drs. Pjotr and Tatjana Knayazev for help during work in cell biology. Dr. Habil. Nediljko Budisa is thanked for his help and supervision.

Waltraud Wenger, Petra Birle and Tatjana Krywcun are acknowledged for introducing me to this work, and Elisabeth Weyher for her help in mass spectrometry.

My coworkers in the Departments of Prof. Moroder and Prof. Huber are acknowledged for the collegial atmosphere and for being as much faithful and supportive in front of me, as they were behind me. The “Oesterhelts” are thanked for all the help whenever required and for their warmth when there was none around.

Florian Kurschus and Edward Fellows of the Dept of Neurobiology are remembered for help in experiments and for all the fun times.

My family members are remembered here with love.

Prajna Paramita Pal

Parts of this work are published in the following publications:

[1] Pal, P.P. and Budisa, N. (2006). Engineering green fluorescent proteins using an expanded genetic code. Green Fluorescent Protein (Series**:** Topics in Fluorescence Spectroscopy, Vol. 12) Lakowicz, J.R., Geddes, C.D. (Eds.), Springer Verlag, Berlin, Heidelberg, New York.

[2] Pal, P.P., Bae, J. H., Azim, M. K., Hess, P., Huber, R, Moroder, L. and Budisa, N. (2005). Structural and spectral response of the *Aequorea victoria* Green Fluorescent Proteins to chromophore fluorination. *Biochemistry.* **44**, 3663 -3672.

[3] Budisa, N., Pipitone, O., Slivanowiz, I., Rubini, M., Pal, P. P., Holak, T. A. and Gelmi, M. L. (2004). Efforts toward the design of 'Teflon' proteins: *In vivo* translation with trifluorinated leucine and methionine analogues. *Chem.& Biodiv.* **1**, 1465-1475.

[4] Budisa, N. and Pal, P.P. (2004). Designing novel spectral classes of proteins with tryptophan-expanded genetic code. *Biol. Chem.* **385**, 893 - 904.

[5] Bae, J. H., Pal, P. P. Moroder, L., Huber, R. and Budisa, N. (2004). Crystallographic Evidence for Chromophore Isomerisation in 3-Fluorotyrosyl-Green Fluorescent Protein. *ChemBioChem.* **5**, 720-722.

[6] Budisa, N., Pal, P. P., Alefelder, S., Birle, P., Krywcun, T., Rubini, M., Wenger, W., Bae, J. H. and Steiner, S. (2004). Probing the Role of Tryptophans in *Aequorea victoria* Green Fluorescent Proteins with an Expanded Genetic Code. *Biol. Chem*. **385**, 191-202.

1 CONTENTS

2 DEFINITIONS AND ABBREVIATIONS

2.1 Definitions

Canonical amino acids are defined by a three-letter code: methionine (Met), leucine (Leu), phenylalanine (Phe), tyrosine (Tyr), tryptophan (Trp).

Noncanonical amino acids are termed with comparable names, often denoted by a three letter code a well. The prefixes that denote chemical functionality are marked numerically, e.g. (3-F)Tyr (3-fluorotyrosine) or by using Greek prefixes such as *meta*-fluorotyrosine (*m*FTyr).

The term **analogue** defines the strict isosteric exchange of canonical/noncanonical amino acid (e.g. Tyr/*m*FTyr).

Mutant denotes a protein in which the wild-type sequence is changed by site-directed mutagenesis (codon manipulation at the DNA-level) in the frame of standard amino acid repertoire.

Variant denotes a protein in which single or multiple canonical amino acids from a wild-type or mutant sequence is replaced with noncanonical ones (expanded amino acid repertoire, codon reassignment at the protein translation level).

2.2 Abbreviations

AARS- – aminoacyl-tRNA synthese

AA- – amino acid

AV – annexin V

av – Aequorea victoria

*av*GFP – green fluorescent protein from *Aequorea victoria*

bp – base pairs

β-Gal- beta-galactosidase

CD – circular dichroism

Ct – C-terminal (5') part of sequence

DNA – deoxyribonucleic acid

ds – Discosoma striata

*ds*Red – red fluorescent protein from *Discosoma striata*

ECFP/EGFP/EYFP – enhanced cyan/green/yellow fluorescent proteins

FRET – fluorescence resonance energy transfer

GFP – green fluorescent protein

ε – molar extinction coefficient (in $M^{-1}cm^{-1}$)
kb – kilobase (i.e. bp x 1000)
m – *meta; m*-F – *meta*-fluoro
MIR – *multiple isomorphous replacement*
mRNA – *messenger*-RNA
NMM – new minimal media
NMR – nuclear magnetic resonance
NTPs - nucleotide triphosphates
Nt – N-terminal (3') part of sequence
o – *ortho; o*-F – *ortho*-fluoro
p – *para; p*-F – *para*-fluoro
PAGE – polyacrylamide gel-electrophoresis
PCR – polymerase chain reaction
Phe – phenylalanine
(2-F)Phe - 2-fluorophenylalanine (o-FPhe)
(3-F)Phe - 3-fluorophenylalanine (*m*-FPhe)
(4-F)Phe - 4-fluorophenylalanine (*p*-FPhe)
PheRS – phenylalanyl-tRNA Synthetase
RNA – ribonucleic acid
SDS – sodium dodecyl sulphate
SeMet – selenomethionine
SPI- selective pressure incorporation
TFM – trifluoromethionine
TFL- trifluoroleucine
tRNA – transfer-RNA
Trp – tryptophan
(4-F)Trp - 4-fluorotryptophan
(5-F)Trp - 5-fluorotryptophan
(6-F)Trp - 6-fluorotryptophan
Tyr – tyrosine
TyrRS – tyrosyl-tRNA synthetase
(2-F)Tyr - 2-fluorotyrosine (oFTyr)
(3-F)Tyr - 3-fluoroytrosine (*m*FTyr)
UV - ultraviolet
wt – wild type

3 SUMMARY

Among all the halogens, fluorine is known to produce the most interesting properties when used for substitution/modification in proteins. Monofluorinated analogues of aromatic amino acids were extensively used as markers for biophysical as well as biological and biomedicinal studies. One notable feature of the fluorination of aromatic amino acids is the possibility to dramatically change their spectral properties. For that reason incorporation of monofluorinated amino acids into the chromophore or the surrounding protein matrix of autofluorescent proteins is an attractive approach for a rational design of the spectral properties of such proteins. The primary advantage of this kind of a system is the opportunity to introduce specifically targeted changes since the chromophore of autofluorescent proteins is completely encoded in its amino acid sequence.

Green fluorescent protein (GFP) from *Aequorea victori*a and red fluorescent protein (*ds*Red) from *Discosoma striata* have recently become standard reporters and model systems in molecular and cellular biology. Their chromophores are autocatalytically formed in the post-translational reaction between side chains of residues 65-67 having molecular oxygen as their only external requirement. Their unique spectral properties might be used as sensitive probes and reporters of the interactions within the chromophore and between the chromophore and its protein matrix. For that reason extensive variations of the chromophore itself as well as of its surrounding protein matrix by classical protein engineering were performed. In this study, site-directed mutagenesis is supplemented with methods for incorporation of monofluorinated aromatic amino acids in order to create novel classes of autofluorescent proteins. The rationale of this approach is to perform incorporation of electronegative fluorine atoms in order to (a) covalently modify the chromophore (which can cause a fundamental change of its resonance properties) and (b) to manipulate electrostatic interactions between the chromophore and its surrounding protein environment in order to change its spectral properties.

This was brought about by global replacements of tyrosine residues with 2-fluorotyrosine ((2-F)Tyr or *o*-FTyr) and 3-fluorotyrosine ((3-F)Tyr or *m*-FTyr) in “enhanced green” (EGFP) and “enhanced yellow” (EYFP) green fluorescent proteins (GFPs) from the jellyfish *Aequoria victoria*. Such replacements resulted in shifts in pK_a-values, anion/cation equilibrium upon titration, blue/red shifts in absorbance and fluorescence. These differences are more pronounced in fluorinated EYFPs than in EGFPs due to the fluorination of Tyr203, unique for the EYFPs. Modelling studies indicate

conformational space for possible chromophore flipping in fluorinated EGFPs but not for fluorinated EYFPs due to its involvement in stacking interactions with Tyr203. Indeed, the three dimensional structure of 3-fluorotyrosyl EGFP in the crystalline state revealed two conformational states of fluorine in the chromophore as well as in the majority of (3-F)Tyr-side chains in this protein. Astonishingly, in (2-F)Tyr-EGFP, both the chromophore, and other 2-fluorotyrosyl residues globally represent only one conformeric state. These properties are undoubtedly intrinsic features of the fluorinated chromophores in the context of the rigid surrounding protein matrix, and are responsible for unique fluorescence properties of fluorinated autofluorescent proteins. In addition, the crucial Tyr203 in EYFP was replaced with Phe by site-directed mutagenesis and subsequently fluorinated at *ortho*-, *meta*- and *para*-positions in order to further manipulate the spectral properties of EYFP. Contrary to our expectations, *ortho*- and *meta*-fluorinations of the Tyr-residues in *ds*Red did not induce significant changes in the absorbance and fluorescence profiles of these variants. This was possibly due to partial replacements of the analogues.

Cytotoxic and pharmacological properties of fluorinated aromatic amino acids have still not received enough attention in recent years. The overall protein structure is unchanged upon replacement of native Tyr and Trp residues with their isosteric fluorinated analogues since they resemble their canonical counterparts in shape and size. In this way, pharmacologically active or cytotoxic amino acids that might serve as diagnostic markers, are covalently integrated into the polypeptide in an inactive prodrug form and should exert no toxicity during delivery. In order to test this concept, a preliminary examination in cell lines using fluorinated variants (fluorinated Tyr, Phe, and Trp analogues) of enhanced green fluorescent protein and β-galactosidase are presented in this study as well.

In this context, attempts to incorporate analogues of aliphatic amino acids like Met and Leu containing a trifluoromethyl group into proteins (GFP) are presented as well. The most interesting property of these noncanonical amino acids is their enhanced hydrophobicity (they are almost twice as hydrophobic as their canonical counterparts). This feature is often exploited to improve the performance of small drugs in suppressing metabolic toxicity, increasing bioavailability in the delivery of many pharmaceuticals or enhancing drug activity. Until now attempts to globally substitute Leu and Met residues with their trifluoro-analogues in GFP have resulted in only marginal replacement. By increasing the genomic level of native methionyl-tRNA synthetase, the Met→trifluoromethionine substitution levels can be significantly increased; an almost quantitative incorporation of trifluoromethionine into Annexin V was achieved.

4 INTRODUCTION

4.1 Genetic information and the genetic code

4.1.1 Decoding the code: evolutionary controversies and historical milestones

The genetic code's evolutionary history is a matter of dispute between two schools of thought: the one led by Francis Crick[1] that defines the "frozen accident" theory based solely on random biochemical events, and the other, believing it to be the fortuitous outcome of an evolutionary process directed by the blind forces of natural selection[2]. It was, as a matter of fact, random events that repeatedly went through the mill of the forces of natural selection which endow it with its unique and optimal error-minimization capacity.[3] In other words, the code we know today is a result of interplay between chance and necessity[4].

The fact that amino acids are the basic units that build a protein was already known to protein chemists by the end of the 19th century. Then in 1901 Fisher and Hofmeister[5] proposed the "peptide bond hypothesis" that casts light on the mechanism of the actual process of the amino acid linkages. It took another five decades of advancement of analytical techniques before Sanger[6] made history with the successful sequencing of insulin. After Avery, Hershey and Chase[7] identified DNA as the hereditary molecule, a series of efforts from scientists of all fields led, by the end of the fifties, to the unambiguous conclusion that the genetic code contains the blueprint of intracellular protein synthesis. C. Yanofsky[8] and S. Brenner[9] in the mid sixties were the first to confirm that the linear sequence of nucleotides in DNA specifies the linear sequence of the amino acids.

By then, it was impossible to segregate diverse research directions in elucidating the mysteries of the genetic code. Chemistry, physics, enzymology and genetics needed to merge in order to unravel completely life's encoded information. The most prominent people leading this research were F. Crick, S. Benzer, C. Yanofsky, S. Ochoa, M. Nirenberg, H. Gobind Khorana, and S. Brenner and many others who contributed with equally important discoveries. These concepts are the subject of many recent books, overview articles, and personal accounts (see e.g. recent accounts of B. Wiesblum, M. Nirenberg and C. Yanofsky).[7; 10]

4.1.2 A code that runs like clockwork

The genetic language is composed of five alphabets, A (adenine), G (guanine), T (thymine), C (cytosine) for DNA, and U (uridine), that is specific for RNA, DNA's first cousin. All information regarding the functioning of the molecular machinery is contained in these nucleotides that are strung together in a linear manner and can be millions of molecules long.

Genetic Code

1st base	2nd base				3rd base
	U	C	A	G	
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA Stop	UGA Stop	A
	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
	AUG Met*	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

*Met and initiation.

Scheme 1. The genetic code table: Coding triplets (codons) and their corresponding amino acids are arranged so that twenty amino acids are encoded by 61 triplet coding units or codons ('sense' codons) while the remaining 3 (UAA, UGA UAU) serve as a terminal signals in protein translation ('nonsense' codons). Note that in the coding of amino acids such as Arg, Ser, Leu, as many as six codons is assigned to a single amino acid whereas amino acids such as Met (AUG) or Trp (UGG) are assigned with only single coding triplets.

Physically speaking, the DNA strands twist around each other to form what is known as a double stranded helix[11] that is produced by complementary base pairing of the nucleotides. DNA may well be likened to the reference books in a library from where information can only be copied, not removed. Several things might happen to this information in the DNA: it might be retained in DNA as regulatory DNA, transferred to ribonucleic acid (RNA) and retained in different RNA forms (e.g. tRNA), 'reversed' from

messenger RNA (mRNA) back to DNA (by reverse transcriptase), or further transferred from RNA (through mRNA) to a protein.[12] All three primary nodes of information transfer (replication, transcription, translation) are proof-read, which proves beyond doubt that error correction runs deep. A precise transfer of information is possible only due to some set rules that correlate the linear nucleotide sequences with a linear sequence of the amino acid in proteins. The genetic code that we know today is composed of the triplet bases that form codons, each specifying a particular amino acid (Scheme 1). Along with the concept of the three nucleotides coding for any amino acid, comes the obvious fact that there are far as many codes possible as there are amino acids. This means that there should obviously be synonyms in the coded language and that it is a "degenerate code".[13] Therefore, many alternative nucleic acid sequences could possibly encode a single protein as shown in Scheme 1. Such a coding scheme assigns all the twenty standard amino acids along with three termination signals to 61 triplet combinations. It is worthwhile mentioning another important consideration that shapes the integrity of the code design, viz, the hydrophobicity of the code is conserved, due to the general assignment of uridine to the second position of a triplet that codes for hydrophobic amino acids.[14]

4.1.3 Getting the punctuation marks right

In the process of understanding the functioning of this fundamentally simple yet most sophisticated language that ever was, it is important to gain an insight into its grammatics. Along with the triplet nucleotide synonyms that code for the same amino acid (Met and Trp being the only exceptions), the chain initiation and the stop codons should be noted. Out of the 64 coding triplets, three are termed nonsense codons, as they do not code for any amino acid but signal the ribosome to prevent the polypeptide chain from any further elongation. They are commonly referred to as the amber (UAG), ochre (UAA), and opal (UGA) codons. Similarly, the AUG that normally codes for methionine in positions inside the gene, specifies also the starting point of polypeptide chain synthesis, and is therefore termed the "starting methionine"(Scheme 1). Given the correct starting position, the DNA is reputed to maintain a hundred percent fidelity in translation rate. It is quite apparent that very early on, life became obsessed with error correction as a result of selection processes which implemented the genetic code with a set of constraints to guarantee proof-reading mechanisms in the information transfer. It is worth a mention that neutral mutations (those that do not produce amino acid switches) occur in the third position of the codons. Switching of amino acids in sequences is brought about mainly by varying the second nucleotide in the codon. The next variable position is the first nucleotide in the codon,

followed by the third, whose changes are ignored. This 'intra-codon code' is a subtle strategy used by the translational machinery for a degree of conservation and error minimisation.[15]

4.2 The Flow of the Genetic Information

Before delving any deeper into protein synthesis, it would be worthwhile to dwell briefly on the central dogma of molecular biology (Figure 1) that states the direction of the flow of genetic information to be from the DNA to RNA to protein. Crick stated that "the transfer of information from nucleic acid to nucleic acid or from nucleic acid to protein may be possible, but transfer from protein to protein or from protein to nucleic acid is impossible".[12] DNA replication and gene transcription, the two steps that take place before the actual process of translation are outlined below:

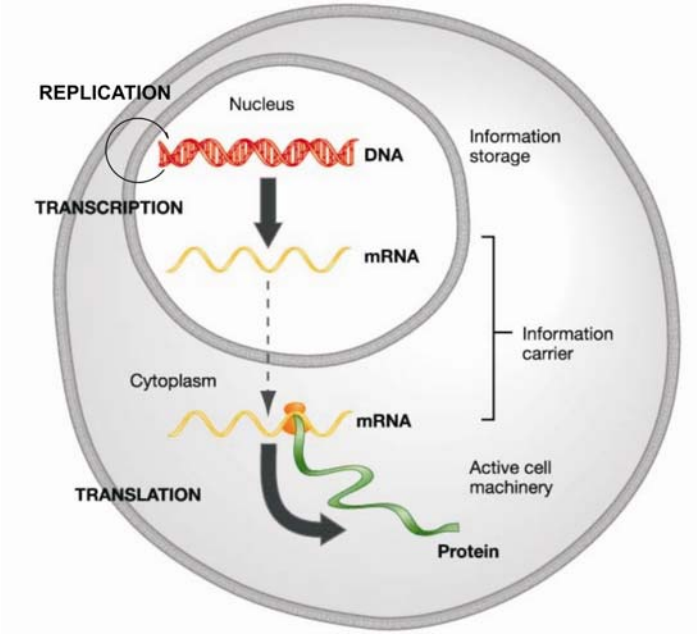


Figure1. General plan for genetic message transmission: The central dogma of molecular biology. Encoded information hardwired into self replicative DNA can be replicated to DNA, or transcribed into various RNAs such as messenger RNA (mRNA); each mRNA string contains the program for synthesis of a particular protein that is executed at the ribosome.[13]

4.2.1 Replication of DNA

This is the first step in the cell division where one DNA essentially produces another identical DNA. During replication, several different proteins collaborate to accomplish the following steps: First, the DNA helicases[11] work to unwind a portion of the double helix. Then some single stranded DNA binding proteins attach to the unwound strands, and

prevent them from winding back together. Next, a DNA polymerase binds to one strand of the DNA and catalyses the elongation of the strands from the 5' to 3' direction, using the initial strand as a template to re-form into a double helix. As the DNA synthesis is possible exclusively in the 5' to 3', an RNA primer is needed to bind to the other template strand upon opening of the double helix, giving rise to discrete segments of polynucleotides known as Okazaki fragments. Finally, the DNA ligase catalyzes the attachment of these fragments to the lagging strand (reviewed in nearly all classical schoolbooks)[11].

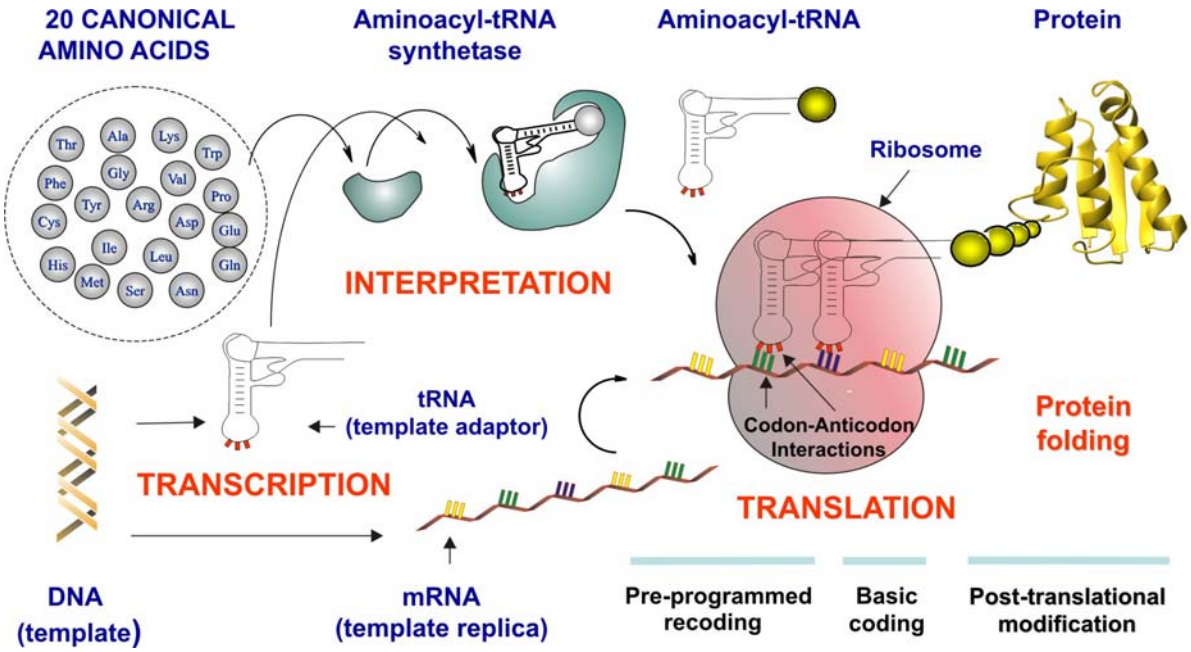


Figure 2. Flow of the genetic information: Transmission of the genetic message is fully completed after the mRNA is translated into protein at the ribosome (protein translation). In this process, a stringent substrate specificity of AARS in the aminoacylation reaction interprets the meaning of a codon (interpretation of the genetic code). On the other hand codon-anticodon interactions (specific base pairing between charged tRNA and ribosome bound mRNA) ensures in sensu strictu translation of the genetic information. The accuracy of both processes contributes significantly to the overall fidelity of protein synthesis (error rates are in the range of about 2 x 10^{-3} to 2 x 10^{-4} in normal *Escherichia coli* cells).[16] Post-translational modifications and pre-programmed recoding (i.e. co-translational modifications) are additional tools that nature uses in order to diversify side chains of proteinogenic amino acids for different purposes.[17]

4.2.2 Gene transcription

Transcription is a process in which one DNA strand is used as a template to synthesize a complementary RNA and works as follows: A cocktail of about fifty different protein transcription factors along with the RNA polymerase bind to the promoter on the 5' end of the gene to be transcribed. This complex, aided by helicase, opens up the double helix in the 5′ → 3′ direction. As the RNA polymerase moves along the unwound DNA strand, it

picks up nucleotide triphosphates (NTPs) to construct a RNA strand (in the 5′ → 3′ direction), following the set rules of base pairing, which inserts, for example, a G in the RNA for each corresponding C in the DNA and vice versa. Since there is no T in RNA, it inserts U instead, every time an A is encountered. Each time an NTP is appended to the 3′ end of the growing strand, the two terminal phosphates are removed. This occurs till the RNA polymerase encounters a termination signal, whereupon both the polymerase and its transcript are released from the DNA.[11]

4.2.3 The highways and byways of protein synthesis

Aminoacyl-tRNA synthetases (AARS) play a cardinal role in maintaining the fidelity of protein biosynthesis (Figure 2), and are believed to be a family of enzymes that evolved and were conserved very early.[18] They are a battery of twenty cellular enzymes (one for each type of amino acid, divided into two classes (class I and class II)) that operate at the interface of nucleic acid and proteins, utilizing ATP as a substrate and Mg^{2+} as a cofactor. Extensive kinetic studies together with the elucidation of the three dimensional structure of AARS provided an insight into the mechanism of their substrate specificity.[19]

Proper functioning of these enzymes requires the presence of the following three domains: first, a catalytic domain (with either a Rossman fold as in the monomeric class I variety, or a seven stranded anti-parallel β-sheet fold as in the dimeric class II variety); second, tRNA binding domains (represented by full motifs or by insertions scattered throughout its structure); and finally oligomerisation domains that are basic parts of each tRNA. The AARS are multi-substrate enzymes, functional in partnering the tRNA, i.e. cognate tRNA binding and charging with single amino acids. This is the Rosetta Stone that translates the nucleotide script of the codons into the amino acid language of the proteins.[20] The secondary (cloverleaf), and tertiary (L-shaped) structures of tRNAs are well-known. The genetic code being degenerate, there needs to be different tRNAs to recognise several codons for any particular amino acid. Different tRNAs chargeable with the same amino acid are called isoaccepting tRNAs. Although the tRNAs are similar in structure and shape, they should have distinct characteristics that enable a precise recognition of cognate AARS to be charged with particular amino acids. In most known organisms, there are more codons than tRNAs and more tRNAs than AARS. The tRNA sequences contain, beside the canonical A, U, G, and C nucleotides, various uncommon nucleotides like pseudouridine, ribothymidine, methylated derivatives of guanosine, etc. Recognition of multiple codons by a single tRNA anticodon is described by “wobble rules”.[13] This can be attributed to

such modifications and tertiary interactions due to non-standard base pairing (i.e. interactions not based on complementary principles).

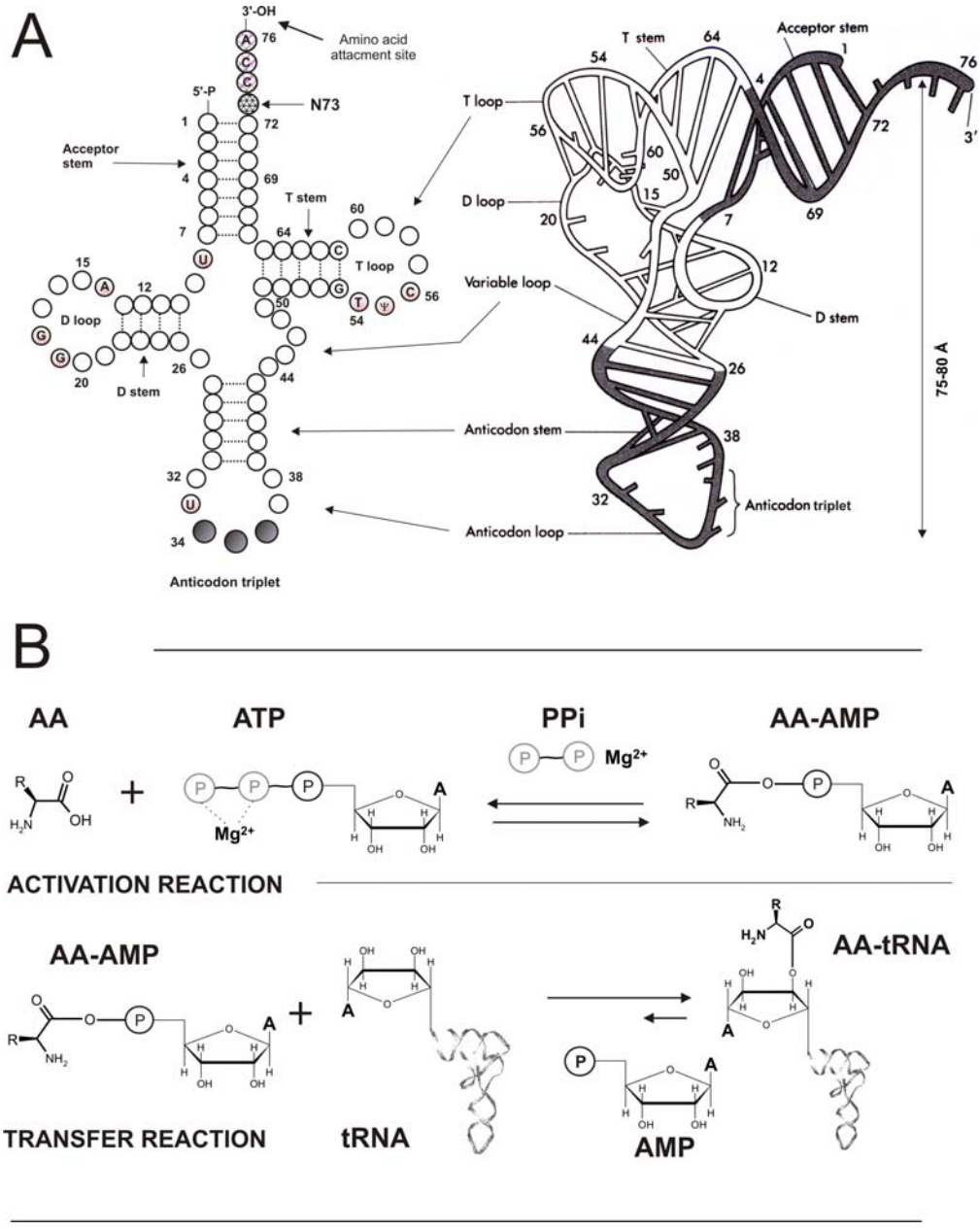


Figure 3: tRNA architecture and generation of aminoacyl-tRNA: (**A**) Two representations of the structure of tRNA. Left: A cartoon showing the cloverleaf-like secondary structure of tRNA. They are composed of 4 stems and three loops. Right: a three dimensional model of $tRNA^{Phe}$ from yeast, showing how the secondary structure folds into tertiary structure.[21] Different regions of tRNA in both representations are interconnected by arrows. (**B**) Generation of aminoacyl tRNAs occurs as a two step chemical reaction, catalysed by AARS. The AARS fishes its pertinent amino acid (AA) from a cellular pool, activates it and charges it onto the tRNA. Following the charging reaction, the aminoacyl tRNA (AA-tRNA) interacts with an elongation factor that is responsible for its delivery to the ribosomal A site (P – phosphate, PPi – pyrophosphate, AMP – adenosine monophosphate, AA-AMP – aminoacyl adenylate).

In the actual aminoacylation reaction, recognition of the correct amino acid by the AARS takes place as the recognition of substrates by enzymes. Each amino acid fits into an active pocket in the AARS where it is bound by electrostatic, hydrophobic, or hydrogen bonds. Activation of the amino acid is brought about in the presence of Mg^{2+} by its attachment to a molecule of ATP at the α-phosphate, resulting in an aminoacyl-adenylate, and inorganic pyrophosphate leaving group (Figure 3B). Then the activated amino acid moiety is transferred to the 2' or 3'-OH of the terminal ribose in the cognate tRNA, producing the specific aminoacyl-tRNA and AMP.[22]

4.2.4 Error correction runs deep

Translational errors appear due either to the incorrect charging of tRNA with non-cognate amino acids or to incorrect decoding of triplets from mRNA.[16] The latter case is due to the fact that the ribosome matches codons with anticodons but never checks strictly whether the tRNAs are charged with cognate or non-cognate amino acids. The job of AARS is to link specific amino acids to their cognate tRNAs which contain the anticodon. This anticodon matches codon (from mRNA) at ribosome. The control of fidelity of these recognition events is achieved by sophisticated proof reading mechanisms at all levels of the genetic message transmission.

Nevertheless, error rates are quite low (error frequency in translation does not exceed on average ≈0.03%)[23] because aminoacylation contributes significantly to the fidelity of ribosome-mediated protein synthesis. This is brought about by positive/negative recognition and regulatory elements of tRNA[24] and error correcting functions of AARS.[25]

Each AARS must be able to discriminate between canonical amino acids as well as between canonical and noncanonical ones in the intracellular pool. Since there are striking resemblances between certain canonical amino acids in that they are isosterically similar in shape (Val-Thr, Cys-Ser), or have the same molecular volume (Ile-Val), proofreading is essential for guaranteeing a correct selection of amino acids. In certain cases like TyrRS, proofreading is unnecessary (in fact it lacks the editing function) due to its uniqueness in recognising tyrosine. In other cases, accuracy in discernment is achieved through the differences in initial binding energies of particular amino acids to AARS. When even such criteria are inadequate, an additional editing step (that enhances AARS specificity by a factor >100) is used as a major determinant of enzyme selectivity. This is because the step that is most susceptible for introduction of errors in the aminoacylation pathway is the formation of aminoacyl adenylate where misacylations by AARS with similar amino acids are frequent. Such non-cognate adenylates are destroyed via two alternative pathways: pre-

transfer, by the hydrolysis of the non-cognate aminoacyl adenylates, or post-transfer, by the hydrolysis of the mischarged tRNA. The generally accepted editing functional model is that either the tRNA accepting end or the non-cognate adenylate shuttles from the catalytic to the editing site.[26]

Aminoacylation, accompanied by some modifications, as the case demands, is followed by translation where AA-tRNAs proceed to the ribosomal elongation cycle. In the ribosome, the initial decoding can distinguish stereochemical correctness of the codon-anticodon duplex that enables discrimination between cognate and non-cognate interactions.[27]

4.3 Reprogramming the coded message

Both eukaryotes and prokaryotes are prone to make errors in translation. On the other hand, there are numerous variations in the translation of the genetic message through all life kingdoms (post-transcriptional modifications, RNA and protein splicing, post translational modification). Some primitive parasitic organisms like retroviruses have even evolved strategies to bend host machinery to favour their replication by hijacking normal translation. In addition, such phenomena are common in a minority of mRNA in many species where sequence-specific information is pre-programmed, both for altering the reading frame, as well as for sense or nonsense codon reassignments (e.g. selenocysteine or pyrolysine incorporation).[17] Biological evolution is remarkable for going through the mill to filter beneficial random mutations in a variety of organisms. Researchers, in turn, have benefited mainly from rapidly evolving organisms like bacteria for work in laboratory conditions. Such organisms prove to be useful for the genetic code expansion as well. They are in fact almost ideal platforms for experimental manipulations with protein translation. In this context, reprogramming of the genetic message can be achieved at two levels, the level of interpretation, and the level of translation.

4.3.1 The weak links in the chain of genetic message transmission: Its interpretation

4.3.1.1 Manipulation of AARS catalytic functions

Reprogramming the genetic message by the reassignment of coding triplets could be possible due to the existence of “weak” links in the interpretation chain. The most obvious weak link is the substrate specificity of the AARS. Indeed as crucial enzymes in the interpretation of the genetic code, the manipulation of their substrate specificity is the most obvious way to reprogramme cellular translational machinery (Figure 4).

1. Catalytic promiscuity of natural enzymes: It is well-documented that natural enzymes, in addition to their main catalytic activity, exhibit an alternative activity with portions situated apart from their active sites. Such a capacity to catalyse secondary reactions in active sites, which are specialized to catalyse a primary reaction, is termed 'catalytic promiscuity'.[28] The catalytic promiscuity of the AARS is at odds with the stringency of the translational standards of the genetic code and gives rise to a "relaxed", or an absence of absolute substrate specificity of these enzymes. Amino acid transport systems reliant on AARS have not evolved editing mechanisms and are incapable of discriminating between chemically and sterically similar amino acids from an endogenous pool, leaving the cells with bitter analogues to swallow.

2. Manipulation with the kinetics of tRNA charging: It is now well documented that the stringent substrate specificity of natural AARS towards cognate amino acids can also be bypassed by sufficient elevation of their intracellular levels (usually by co-expression of particular AARS). This avenue is especially suitable for tRNA-charging with those analogues that are poor substrates in activation reactions. Therefore, the knowledge of kinetics of aminoacylation reactions with unnatural substrates could be used to imperil the fidelity of protein synthesis and increase the turnover of proteins with noncanonical amino acids.[29]

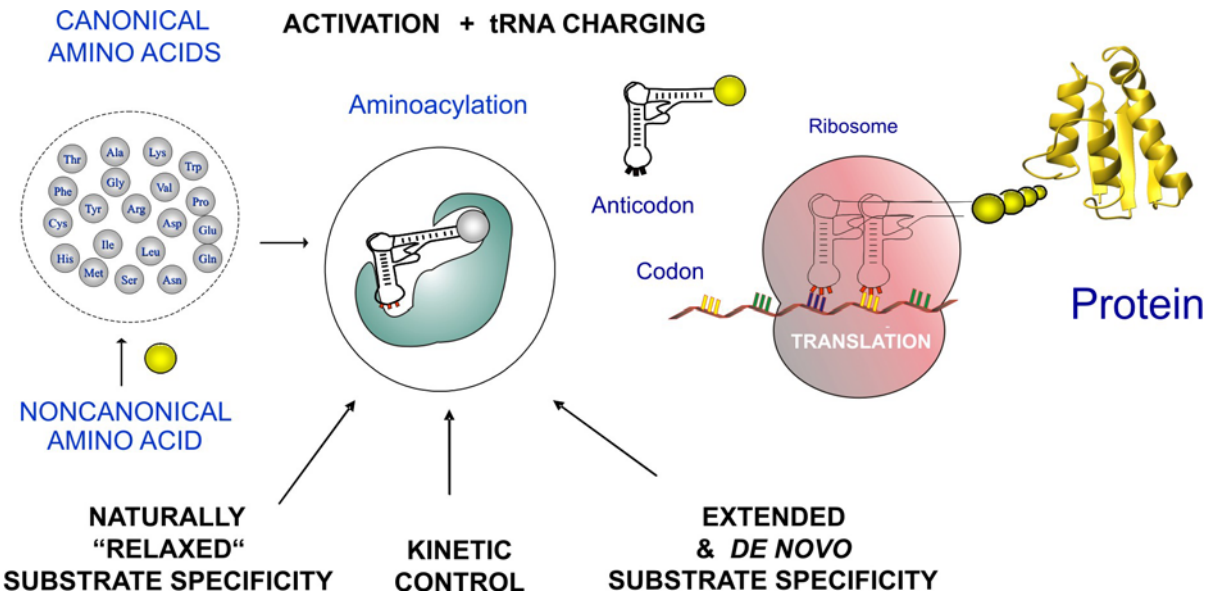


Figure 4. Reprogramming protein translation by AARS manipulation: Protein translation can be manipulated by using different approaches in order to extend the AARS substrate specificity. First, 'catalytic promiscuity' allows the bypass of stringent substrate specificity of natural AARS. Second, by efficient elevation of the intracellular levels of AARS poor substrates in activation reactions can be charged onto cognate tRNAs. The use of designed AARSs with extended substrate specificity, attenuated or changed editing functions, and even novel amino acid substrate specificity, holds a promise greatly to expand the scope of protein synthesis.

3. Engineering active or editing sites of AARS: It is possible, in practice, to extend an amino acid binding pocket of a particular AARS in order to charge tRNAs with bulky amino acids such as brominated or iodinated analogues of aromatic amino acids.[30] It is well known that single catalytic active sites in some of the AARS (LeuRS, IleRS, ValRS) are supplemented with an additional editing (hydrolytic) site whereby it selects amino acids to be charged with a fine-tooth comb ("double sieve" editing or proofreading).[31] A rational mutation of the residues in the editing site of some AARS enables efficient blockage of the editing reaction.[32] This leads to an attenuation of enzyme editing activity and subsequently tRNAs loaded with noncanonical amino acids can enter the ribosome cycle. Finally, engineering AARS that exclusively recognise the desired noncanonical amino acid, would drive the final nail into the coffin of the stringent code we know today. Such customised AARS requires the *de novo* generation of substrate specificity, where it is inclined to favour a non-cognate amino acid along with its corresponding "orthogonal" tRNA over its canonical counterparts. In other words, successful orthogonal engineering would mean the production of a novel adaptor tRNA featuring a strict absence of cross reactivity with other components of the host protein synthesis apparatus.

4.3.2 Other approaches to manipulate protein translation

Other approaches for reprogramming ribosome-mediated protein translation were also developed and are widely used. For example, by mutating the 23S rRNA in the ribosome, it is possible to introduce D-amino acids into proteins.[33] Sisido and co-workers have studied mischarging of tRNA adaptability at the ribosomal A site and have found that many bulky aromatic amino acids could be successfully translated into target proteins.[34] Certain tRNAs with extended anticodons can read frameshifted mRNA sequences (frameshift suppression) – a phenomenon that can be used for noncanonical amino acid incorporation as well.[35] Finally, since the genetic code reserved three codons (UAA, UAG, and UGA) for translational termination, it is possible to reassign at least one of them for noncanonical amino acid incorporation while the other two can serve as stop signals. It is even possible to allow only one of the three stop codons to perform their duty as signal terminators in translation, leaving the remaining two 'blank' codons to be used for encoding specific noncanonical amino acids.[36]

The cells carrying nonsense mutations in the middle of a gene are especially vulnerable to premature chain termination. Possible harmful effects can be reversed by a different

genetic change at different genes of mutant tRNA anticodons and are called suppressor mutations.[37] Suppressor tRNAs are able to insert different canonical amino acids in response to the nonsense codons (when one of the three stop codons appears in the mRNA), such that normal functions of these termination codons are suppressed by tRNAs with altered identity. While nature performs codon reassignment by a switchover of one canonical amino acid with another, it is possible to add chemically or enzymatically to such suppressor tRNAs a wide variety of chemically different noncanonical amino acids that can be successfully translated into proteins.[38]

4.4 Syntax and architecture of the standard and expanded genetic code

4.4.1 Basic features of the universal code

The universal genetic code that we know has at least three basic features such as (a) it has strict codon synonymy; (b) it allows and specifies a finite number of chemically diversified side chain functionalities for protein construction; (c) it exhibits a redundancy in demonstrating encoded physicochemical properties like hydrophobicity.[39]

Encoded amino acids can be divided according to their physicochemical, metabolic and physiological properties. A major underlying property reflected in the code architecture is what determines the polarity of the amino acids. The increased redundancy in the coding of strictly polar/apolar amino acids that exhibits a greater proportion of amino acids like Arg, Ser (XGX) and Leu (XUX) is directly related to their distributions in protein structures. In fact, a major division of amino acids in protein structures follows a simple 'apolar in – polar out' principle that allows them to fold into tight particles that have an internal hydrophobic core shielded from the surrounding solvent. Thus, hydrophobic residues (XUX group of codons) have predominance in protein interiors while hydrophilic residues (XAX, AGX group of codons) are found in abundance at protein surfaces. Such an arrangement encodes hydrophobicity to a great extent which is the main driving force of protein folding, i.e. formation of protein cores and functionally active configurations.[14]

Another remarkable feature of the code is the exclusive repertoire of 'only' 20 amino acids. Many theories are put forward to explain this remarkable feature. For example, Crick, in his 'frozen accident' hypothesis[1] stated that the current code repertoire reached an evolutionary dead end, when cells were no longer capable of assimilating new amino acids in their genetic code. When life had evolved to a certain level of complexity, the proteome of the cell underwent a further diversification by other means like context-dependent pre-programmed reprogramming and post-translational modification.[40] Miller and Weber[41] speculated that a certain class of amino acids was excluded from the code through selection

against adverse effects that they might cause on protein structure, stability and function. The explanations for the exclusion of other amino acids that do not violate these rules are probably historical, which might have been their unavailability or insufficient evolutionary time to produce them metabolically.[42]

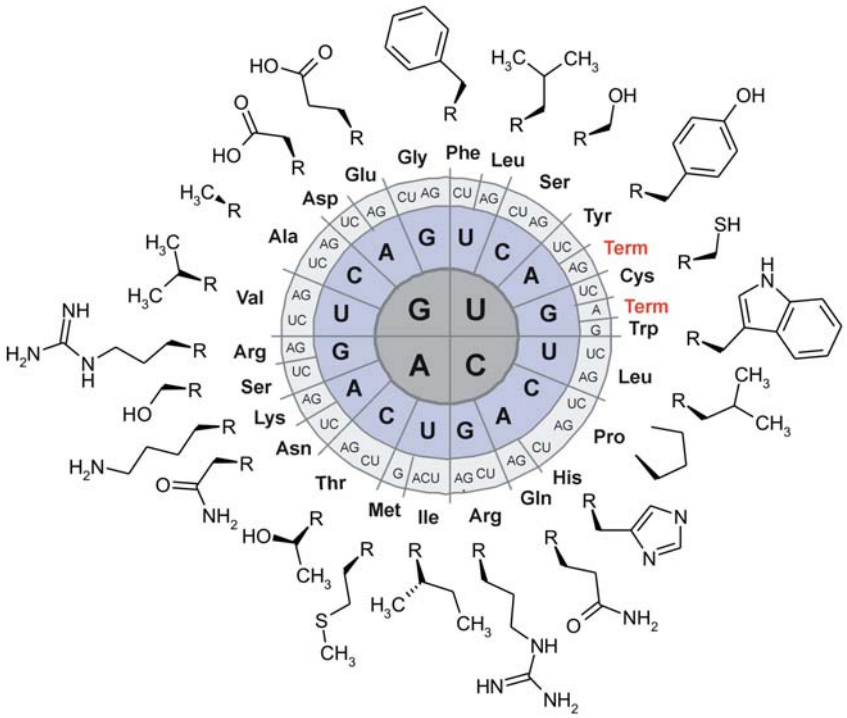


Figure 5. Genetic code (radial representation) and chemical structures of canonical amino acids: The naturally encoded chemical diversity includes a limited number of side-chain functional groups: carboxylic acids (Glu, Asp), amides (Gln, Asn), thiol (Cys) and thioether (Met), alcohols (Ser, Thr), basic amines (Lys), guanidine (Arg), aliphatic (Ala, Val, Leu, Ile) and aromatic side chains (Phe, Tyr, His, Trp), and even the cyclic imino acid Pro. Most of the amino acids are represented by more than one codon - the genetic code is degenerate. It assigns amino acids to related triplets on the basis of polarity/hydrophobicity: especially strictly apolar amino acids like Leu (XUX group of codons) and polar amino acid such as Arg, Ser (XGX group of codons) have larger number of codons assigned to single amino acids.

4.4.2 Strategies for codon reassignment: Flexibility within constraints

There are two currently widely used methodologies for efficient codon reassignments in order to add/substitute canonical with noncanonical amino acids. A new meaning of a particular codon in the genetic code can be achieved either by the reassignment of sense codons or by suppression (read-through) of termination triplets (UGA, UGG, UGU) or non-triplet coding units. These experiments affect the interpretation of the genetic code and lead to the expanded scope of ribosome-mediated protein synthesis.

At the level of living cells, successful experiments include (1) uptake/import of the noncanonical amino acid, (2) its intracellular accumulation at levels high enough for efficient substrate turnover (activation and tRNA acylation) by AARS, (3) metabolic and chemical stability of the imported noncanonical amino acid, (4) tRNA charging (acylation) that must be achieved at a demonstrable rate, and (5) the translation of noncanonical amino acid into nascent polypeptide chain.[43]

4.4.2.1 Nonsense codons suppression – a brief overview

By using chemically or enzymatically aminoacylated suppressor tRNAs, the original meaning of termination (or frameshifted) coding units is 'suppressed'. This leads to suppression-based methodologies that are usually associated with position-specific mode of incorporation. The *in vitro* suppression approaches for an amino acid repertoire expansion are based on the fact that chain termination mutations can be corrected in some bacterial strains by specific suppressor tRNAs that have altered reading capacities. Such tRNAs, capable of carrying canonical amino acids have the termination anticodons as their own anticodons. These tRNAs can suppress nonsense (one of the three terminator codons appears in mRNA) or missense (another sense codon so that different amino acids are determined) codons or frameshift mutations in the parent gene. This is based on the classical discovery that misacylated tRNAs participate in the peptide bond formation at ribosomes[44] in the context of *in vitro* transcription and translation systems.[45]

In this experiment a suppressor tRNA is prepared by chemical aminoacylation such that it recognises a particular stop (nonsense) codon. Conventional site-directed mutagenesis is used to introduce the desired stop codon at the site of interest in the target gene. When the acylated suppressor tRNA and the mutant gene are combined in an *in vitro* transcription/translation system, the non-canonical amino acid is incorporated in response to the termination codon and gives rise to a protein containing that amino acid at the specified position.[46]

Since chemical aminoacylations are expensive, and complicated experiments gave very low yields of mutated proteins (i.e. sub-milligram range), this methodology was developed for *in vivo* expression conditions. In this approach, site-specific incorporation *in vivo* is achieved by cells which can take the desired non-canonical amino acids directly from the growth media. The strategy can be summarised in a three-step procedure. First, a suppressor tRNA should be constructed that would be "orthogonal"; i.e. as a rule, it would not interact with any of the naturally occurring AARS and should be able to deliver non-

canonical amino acid in response to a stop codon (usually UAG) in the mRNA encoding the protein of interest. Second, by engineering of the "orthogonal" AARS (using mutagenesis and directed evolution approaches) it would recognise related orthogonal-tRNAs but not any endogenous tRNAs. Third, by screening a library of mutants of both orthogonal tRNAs and AARS, an"orthogonal" AARS-tRNA pair needs to be found that is capable of activating and transferring the desired non-canonical amino acid to the desired position in the protein sequence in response to a nonsense codon.[47]

4.4.2.2 Sense codon reassignment by using auxotrophic cells

The experiments for sense-codon reassignments are usually performed by using special cell lines known as auxotrophs. This category of mutants (either of bacterial or eukaryotic origin) are incapable of producing its endogenous amino acid and can be bred and selected using methods of classical genetics. Therefore the experiments of sense codon reassignments rely on classical genetics based on breeding, strain construction and selection of intentionally induced genetic variations in intact host expression cells that are forced to make 'mis-incorporations' by environmental control and manipulation, i.e. by an externally imposed selective pressure. The usual terms used for this experimental approach are codon reassignments and multiple-site mode of incorporation. In fact, auxotrophic cells with manipulated translational components, presumably the AARS, offer ideal platforms to change the selection of the amino acids for protein biosynthesis, i.e., to further expand its amino acid repertoire.[43; 48]

The use of auxotrophic strains can be traced back to the 1950s. Cowie and Cohen 1957 reported that "…*E. coli can grow and synthesize active proteins under conditions where an amino acid is totally replaced by a synthetic unnatural analogue*." and "… *that the amino acid composition of proteins may be influenced by environmental changes*".[49] Various pioneering experiments aimed at introducing variations in the amino acid composition in proteins have been made by the fermentation of microorganisms in the presence of structural analogues of amino acids.[50]

4.4.2.3 Selective pressure incorporation (SPI): Basic principles

Nearly all amino acids beside the canonical twenty and their metabolic intermediates are toxic (i.e. they actually prevent cellular growth), selenomethionine being an exception to this rule.[51] Nonetheless it well known that the metabolically toxic analogues such as fluorinated aromatic amino acids, norleucine and canavanine might serve as substrates for protein synthesis. When such toxic analogues are supplied along with their canonical

counterparts in the synthetic growth media, they usually lower the incorporation levels of all cellular proteins. For example, the Met-analogue norleucine leads to only a 38% replacement of the total cellular Met in proteins accompanied by a greatly impaired viability of the *E. coli* ML304d.[52] More recent experiments have shown that it is possible to replace in this manner, about 50% of the total cellular proteins since one more division of bacterial cells takes place after the exhaustion of Met in the intracellular pool and extra-cellular medium.[53]

To achieve a full substitution in target proteins, a basic problem was to resolve the metabolic toxicity from translational capacity. Auxotrophic mutants of bacteria have been exploited as the most straightforward means to circumvent toxic metabolic effects and to achieve preferential incorporation of noncanonical amino acids over canonical ones. This is done by starving host cells of a specific canonical amino acid and by inducing the synthesis of a specific protein, with concomitant addition of the analogue to the culture medium. Even then, complete labelling of a single target protein could be achieved only after the introduction of recombinant DNA techniques. Highly efficient expression systems capable of generating heterologous proteins in large amounts and in a controlled manner are now available. They provide a stringent control of the heterologous gene expression and allow the full exercise of selective pressure on the translation apparatus in order to produce a single substituted target protein in the context of the host cell with an unchanged proteome (Figure 6).

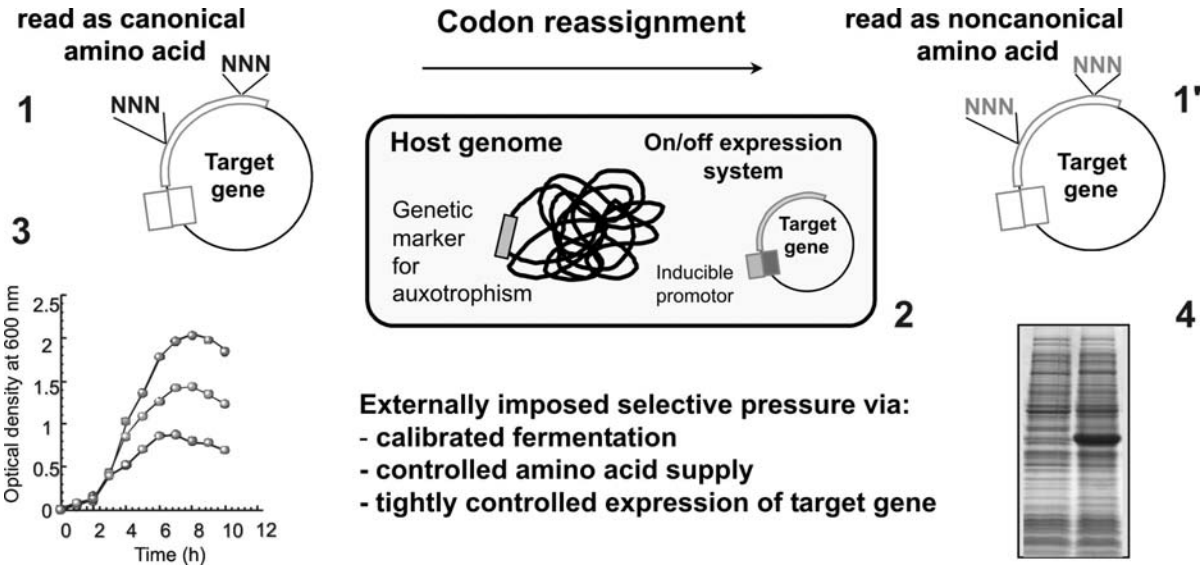


Figure 6. Principles of sense codon reassignments by selective pressure incorporation (SPI): Naturally assigned (1) sense coding triplets (NNN) can be reassigned (1') during properly calibrated fermentation and

controlled expression experiments. This requires intact auxotrophic host cell (2) suitable for calibrated fermentation (3) in the defined minimal media combined with a controllable expression system (4). These experimental conditions when combined result in a single or a set of canonical → noncanonical amino acid substitutions in target protein.[43; 48]

The experimental pressure on a selected expression system has been exercised as follows: Host cells were grown in minimal media having a limited concentration of the specific canonical amino acid, that is consumed thereby holding their growth in NMM in check. After depletion of the canonical amino acid, cells are unable to grow any further and reach a stationary phase. The target gene is transcriptionally silent till now. At this point the amino acid analogue is added, with a simultaneous induction of the target protein synthesis. From this point onwards, the cells stop growing and their only task is to produce the desired modified protein. As a result of such a pressure on the translation apparatus, the target protein containing exclusively the noncanonical amino acids is accumulated. This procedure leads to the arrest of cellular growth, but a successful expression of target protein variants is achieved.[54] Needless to say, the cloned gene should be under the efficient control of the promoter and competitive expression systems should mainly express the target DNA after induction. In the presence of background expression (gene leakage), before the induction of protein synthesis, the resulting protein will be incompletely labelled. Even though such method of incorporation yields substantial amounts of recombinant protein, it is limited to those non canonical amino acids which are in the editing tolerance range of aminoacyl tRNA synthetases.[51]

Most importantly, desired noncanonical amino acid incorporations can be performed on single target proteins without harmful global effects on the expression host. The basic requirements (summarized in Figure 6) include: (i) selection of proper cell and expression system; (ii) control of fermentation conditions (i.e. environment); and (iii) selective pressure for the amino acid replacement (i.e. sense-codon(s) reassignment) at the level of single protein. This approach is termed the selective pressure incorporation (SPI) method since it confirms the principle that the choice of the amino acids for protein synthesis could be conditioned by the control of environmental factors such as amino acid supply and fermentation parameters, i.e. by an experimentally imposed selective pressure.[54]

4.4.2.4 Hierarchy of the expanded code

The basic limits for the successful expansion of the amino acid repertoire are metabolic constraints (almost all amino acids beside the encoded ones and their intermediates are toxic and do not support cellular growth at all), proofreading mechanisms in the genetic message transmission that ensure the fidelity of protein translation, and finally protein

folding rules which are aimed at relating the structure of the genetic code to the physicochemical properties of the amino acids. To challenge these constraints, one has to change or modulate the activities of AARS and to bypass metabolic constraints imposed by intracellular milieu of the expression hosts.

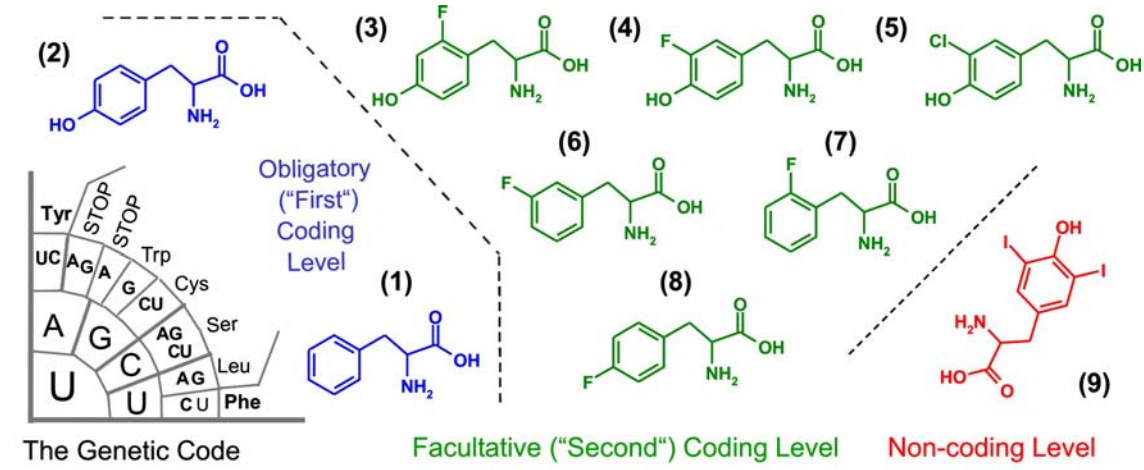
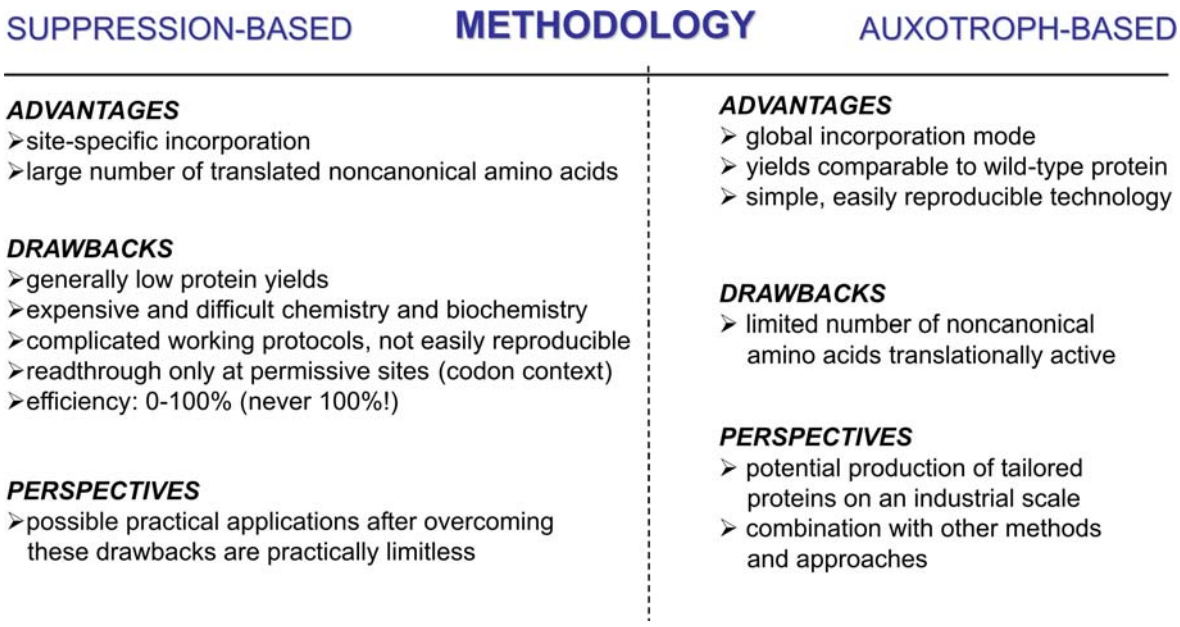


Figure 7. Genetic code of halogenated tyrosine and phenylalanine analogues studied in this work: The 'first' or obligatory coding level marks the evolutionary established assignment of Tyr (UAU, UCU) and Phe (UUU, UUC) coding triplets during protein biosynthesis. The translation of their fluorinated counterparts into target protein sequences enables building of a second or facultative coding level in the genetic code. The non-coding level includes noncanonical amino acids for which experimental attempts for incorporation were unsuccessful. Most of the halogenated noncanonical analogues of Tyr (**1**) and Phe (**2**) presented here are recognized by native TyrRS and charged onto cognate tRNAs.[55; 56] Interestingly, 3-chlorotyrosine ((3-Cl)Tyr, **5**) is translationally active whereas a natural product 3,5-diiodotyrosine (**9**) is not substrate for translation in Tyr-auxotrophic E. coli auxotrophic expression strain AT2471 (used in this study). Full names of other substances: 2-fluorotyrosine, (2-F)Tyr (**3**); 3-fluorotyrosine, (3-F)Tyr, (**4**); 3-chlorotyrosine (3-Cl)Tyr, (**5**); 2-fluorophenylalanine, (2-F)Phe (**7**), 3-fluorophenylalanine, (3-F)Phe (**6**), 4-fluorophenylalanine, (4-F)Phe, (**8**).

The code whose amino acid repertoire is expanded in this way is not limited to its first (restricted) part, but also to its second (relaxed or second code) part which can be experimentally assessed. For that reason three levels in the structure of the genetic code were postulated: (i) the first coding level includes canonical amino acids whose entry in the genetic code is obligatory; (ii) the second coding level includes noncanonical amino acids whose entry is allowed in the code (i.e. entry is facultative); (iii) the third or non-coding level includes all amino acid analogues whose entry in the genetic code by sense-codon reassignments is not allowed (i.e. entry is forbidden) at this level of the method development.[51] This indicates that for each amino acid in the standard genetic code table like Tyr and Phe (as shown in Figure 7) it should be possible to establish a whole library of translationally active substances that are capable of being incorporated in proteins.

4.4.2.5 Comparison of main approaches for noncanonical amino acids incorporation

The basic advantage of the SPI method used in this study is the concept of the use of sense DNA coding triplets and the endogenous AARS:tRNA pair, for residue-specific substitutions.[48] Using this simple, easy and reproducible method, expression and purification of variant proteins are achieved with yields equalling that of their wild type counterparts. For these reasons, this approach is generally considered to be the most promising for large scale synthesis of proteins with noncanonical amino acids for therapeutic, biomaterial or bioengineering applications. On the other hand, position specific incorporation, complimented by site directed mutagenesis can be achieved by a reassignment of codons for those amino acids having a relatively rare abundance in proteins (like Trp, Cys and Met). However, a reassignment of codons of more frequent residues like Gly or Leu might lead to an extensive exchange of residues inside the structure which would undoubtedly be harmful for the protein. Such problems have been circumvented by exploiting the termination (or nonsense) codons used by the genetic code (UAA, UAG, and UGA) to gain a position specific mode of incorporation (see section 4.4.2.1). However the efficiency of this method in many cases is still not high enough to be used routinely in the laboratory. The advantages and disadvantages of both approaches are listed in Scheme 2.



Scheme 2. Comparison of two mainly used methods for noncanonical amino acids introduction into proteins.[57]

4.5 Halogenated aromatic amino acids

Naturally occurring post-translationally halogenated amino acids are usually derivatives of histidine, tryptophan and tyrosine and have proved to be interesting targets for biomedical research. For example, some antibiotics from certain marine algae and some higher plants are in fact chlorinated derivatives of aliphatic or aromatic amino acids. Examples of post-translational iodinations, brominations and chlorinations of tyrosine were first described in vertebrate tyreoglobulins and invertebrate scleroproteins and have lately been found in a variety of marine organisms.[58] Halogenated modifications of amino acids are usually directed towards defence mechanisms against predators and have been isolated from a variety of marine specimens. Of special significance is the post-translationally iodinated, chlorinated, and brominated tyrosine residues.[59] This is not surprising since a marine ambient is characterised by a higher natural abundance of iodine, bromine and chlorine ions. Several marine snails whose natural habitat is in the coral reefs prey upon other marine organisms using a cocktail of post-translationally modified peptides that render toxicity. These toxins have been found to possess pharmacological properties that have proved to be valuable tools for pharmacological research.[60] In the context of protein translation fluorinated and chlorinated amino acids are 'xeno'-compounds and consequently both fluorine and chlorine are 'xeno' elements. It should not therefore be surprising that these different properties will produce different effects upon their integration into cooperatively folded protein structure.[54]

Fluorine in the form of fluoride minerals is one of the most abundant halogen in the earth's crust and occurs only in 12 naturally organofluorine compounds. Conversely, there is a large number and range of commercially generated fluorinated compounds – all of them being of anthropogenic origin. The fluorine's chemistry has not evolved in living cells possibly due to its low availability. It is well known that the available fluoride is mostly insoluble. Sea water, for example contains 1.3 ppm fluorine and 19,000 ppm chloride. Such rarity of the naturally fluorinated products contrasts with the existence of more than 3,500 naturally occurring halogenated compounds.[61]

4.5.1 Why fluorine?

Despite fluorine's natural abundance as fluoride minerals, available fluoride is mostly insoluble, resulting in fluorine chemistry remaining at the unevolved state. Fluorine's low natural abundance among naturally occurring halogenated compounds makes many

scientists keen to study its possible biological effects. Not surprisingly, the discovery of fluorine in teeth enamel and the shell and yolk of eggs prompted scientists like McClure to probe their physiological role as early as 1933.[62] Workers in aluminium factories were hardest hit by osteoarthritis and related symptoms with elevated levels of fluoride in their hair, serum and urine samples. Fluoride in the presence of trace amounts of aluminium forms complex ions that affect all blood elements and circulation, bone cells, fibroblasts, calcium homeostasis etc.[63]

Nonetheless the importance of fluorine in human everyday life is immense. For example, the fact that the presence of fluorine in small amounts in drinking water reduces tooth decay is the main reason why many countries use fluoridated water for dental protection. Since 1956, a class of chloro-fluoro-carbons (CFCs) has been used to assist surgery; (e.g. Fluothane (CF_3CHBr) was used in 70-80% of all anaesthetics but had unpleasant side effects like post operative nausea). Recently Fluothane inhalation anaesthetics have been replaced by fluoro ethers that have no such side effects. Materials known under commercial names such as Gore-Tex, or Teflon (fluorine containing polymer, polytetrafluoroethylene $(CF_2\text{-}CF_2)_n$) are excellent examples of inert and non-toxic material used equally well for cookware or surgery. By virtue of a ready solubility of fluorinated compounds in lipids, the rate of fluorinated drug absorbance and transport can be enhanced. Not surprisingly, fluorine is a component of several drugs including anticancer, antiviral, antiinflammatory, antiarrhythmic heart drugs, etc. Last but not least, modern combat uniforms, astronaut suits and clothing used by mountaineers and sailors are in fact fluorine-based engineered materials. For all these reasons fluorination of proteins should make available wider perspectives. For example, the design of fluorine-containing self-sorting peptides or significantly stabilised fibrils such as collagen or proteins have recently been reported.[64] To examine the possibilities for protein engineering and design using fluorinated amino acids, various proteins (different *av*GFP mutants, *ds*Red, β-Gal, Annexin V) were used as models for incorporation experiments. In particular monofluorinated analogues of aromatic amino acids (Tyr, Phe) and trifluorinated analogues of aliphatic amino acids (Leu, Met) were used to substitute their canonical counterparts in model proteins.

4.5.2 Chemistry and biology of fluorinated amino acids

It is a generally accepted fact that hydrogen to fluorine substitution causes minimal steric demand. In fact, the covalent radius of fluorine is 1.47 Å and lies between that of oxygen (1.57 Å) and hydrogen (1.2 Å) (Table 1). As the most electronegative element in

the periodic table, fluorine can effect a change in the electron distribution of the molecule and is manifested as altered dipole moments, pK_a, and the chemical reactivity and stability of the adjoining functional groups. Such changes are determined by the bonding between the fluorine atom and the functional group. For instance, if the fluorine atom is *ortho*-positioned in the phenolic moiety, the pK_a is reduced by 1.2 units, while its positioning at the *meta*- and *para*-positions cause less significant effects. Fluorination of the amino acid Tyr can be taken as an example.[56] The pK_a of free Tyr in solution is 10.0, whereas the pK_a values of the free fluorinated amino acids *ortho*-FTyr is 9.04 and 8.5 for *meta*-FTyr. For that reason, it is possible to rationalize for pK_a lowering of fluorinated proteins which is in the order of wt > *ortho*-FTyr > *meta*-FTyr-protein as will be discussed later. Obviously, the presence of fluorine closer to a carboxylic acid group of the amino acid seems to have more pronounced effects on the pK_a of the amino acid.

Table 1. Physicochemical properties of the carbon-fluorine bond.[65]

Element (X)	Electronegativity	Bond length C-X (Å)	van der Waals radius (Å)	Bond energy (kcal/mol)
H	2.1	1.09	1.20	99
F	4.0	1.39	1.35	116

Fluorine forms a strong bond with the carbon atom, which is in fact one of the strongest known in organic chemistry (Table 1). The van der Waals radius of fluorine is 1.35 Å, and the C-F bond is 1.39 Å in comparison to the C-H bond which is 1.09 Å. Fluorine generally causes minimal structural perturbations during CH → CF replacements, and such changes can be easily accommodated even in the protein core.[66] Indeed, in most cases the structure of the fluorinated proteins is indistinguishable from that of the wild-type protein form, which is confirmed in numerous crystallographic studies.[67] As the most electronegative element known, fluorine exerts a much stronger electronegativity (4.0) in its local environment than chlorine does (electronegativity of chlorine is 3.16). Both produce strong electron-withdrawing inductive effect (-I effect) and induce a moderate electron-donating resonance effect on the whole aromatic ring (+M effect). They also possess non-shared electron pairs, and carbon-bound fluorine is thought to be a weak hydrogen bond acceptor. Their dipole moments are opposite to that of a C-H bond (C-F: 1.41, C-Cl: 1.46). However, the most dramatic differences between these halogens lie in their steric properties. Chlorine is more bulky than fluorine (van der Waals radius of F is 1.35 Å and Cl 1.8 Å) and is much more polarizable. Thus, chlorine as the bulkier atom (C-H: 1.09 Å, C-F: 1.39 Å, C-Cl: 1.77 Å) should produce more dramatic steric changes in a protein core than fluorine.

In general, monofluorinated analogues of mainly aromatic amino acids were extensively used for various purposes in early incorporation experiments as well as in recent years.[68] For example, fluorination of the catalytic residues such as histidines or tyrosines might lead to changes in enzyme pK_a and subsequently in the novel pH range of catalytic activity.[69] In this context one of the major achievements of the work presented here is the spectral and dynamic properties of green fluorescent protein upon global substitution of its Tyr-residues with *o*-, and *m*-fluorotyrosines and phenylalanines as will be later discussed in more details. Finally, such single atom exchange as H→F provides an "atomic mutation" as an approach for detailed studies of protein folding, activities, dynamics, spectroscopy and stability.[66]

When compared with CH→CF substitution in monofluorinated amino acids (where the fluorine atom is not a sterically demanding substituent), the CH_3→CF_3 substitution increases the sterical bulk by more than 12 Å^3 per replacement. Thus, the volume of the trifluoromethyl group was estimated to be closer to that of the isopropyl group.[70] Therefore, such increased sterical bulk of trifluoromethyl-group relative to the nonfluorinated methyl group would lead to unfavourable interactions with the surrounding residues in a particular protein interior. In general, it is conceivable that the CH_3→CF_3 replacement might be tolerated at single sites, while global substitution of all Leu/Ile/Val side chains in larger proteins would significantly enlarge their fluorous hydrophobic core. It is therefore reasonable to expect that the major deterrent for a full integration of trifluorinated amino acids into protein structures may be their steric bulkiness and subsequent difficulty in accommodation into compact protein cores.[71]

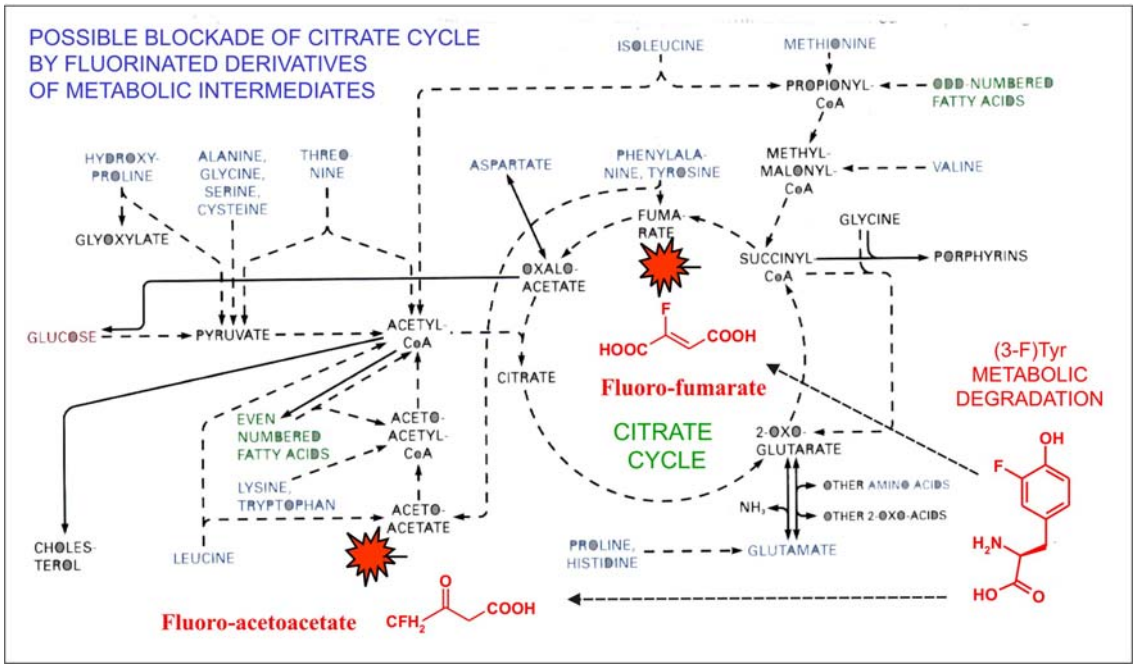
Nonetheless, materials modified by multiple fluorinated carbons possess elevated hydrophobicity and often have profound and unexpected properties. In the context of minimal structural pertubations coupled with characteristics of enhanced hydrophobicity, the perfluorinated polymer Teflon is an excellent example. The trifluoromethyl group is found to be almost twice as hydrophobic as to the methyl group. Perfluorocarbons interact neither with hydrophilic nor hydrophobic solvents, and can best be described as fluorophilic.[64] This would mean that such proteins would have the capacity to fold in aqueous media, yet resist denaturation by efficient exclusion of organic solvents. They perform this by forming a compact core, making the fluorocarbon side chains inaccessible to the surrounding organic solvent molecules. The most conceivable method of achieving such unexpected properties is to replace hydrophobic (core) amino acids with trifluorinated analogues, which offer themselves as rational equipments to shape the interface of proteins, making them simultaneously hydrophobic and lyophobic, thus enhancing their

structural stability. However, natural structural frameworks in respect to hydrocarbon cores have been optimised for billions of years, and as a rule disallow packing, especially with a profusion of fluorine atoms unaccustomed to natural scaffolds. It remains a challenge for *de novo* protein designers to achieve a novel repacking of the protein core to produce "non-stick" proteins that are properly folded to preserve their functional properties.

4.5.3 Canonical amino acid residues that serve as substitution targets

4.5.3.1 Chemistry and metabolism of tyrosine and its fluorinated derivatives

The aromatic amino acid Tyr is normally found buried in the protein hydrophobic core where it is usually involved in various stacking interactions with other aromatic side chains. A reactive hydroxyl group at the *para*-position of the benzene ring makes it suitable for additional interactions like hydrogen bonding or various interactions with non protein atoms. The first ever reported successful attempt of replacing tyrosine with its analogues was the incorporation of (3-F)Tyr in β-galactosidase.[72] Ring and Huber[73] reported a rational manipulation with the role of catalytic Tyr in the active site of β-Gal by the global replacement of all Tyr side chains with 3-fluorotyrosine (Figure 7). More recently, 2-fluorotyrosine and 2,3-difluorotyrosine has been found to be translationally active by Brooks and co-workers,[74] who incorporated them in ketosteroid isomerase from *Pseudomonas testosteroni*.



Scheme 3. Graphic representation of the possible mechanism of metabolic toxicity of (3-F)Tyr in the context of the citrate cycle. The metabolic degradation of canonical amino acid Tyr results in acetoacetate and

fumarate; both are intermediates in the citrate cycle. Similarly, the metabolic degradation of noncanonical (3-F)Tyr results in fluorinated counterparts of these metabolic intermediates, that are highly toxic for the metabolism of living cells. Modified from Michal.[75]

Tyrosine fluorinated at the *meta*-position is known to have particularly strong toxic properties. The metabolism of (3-F)Tyr was examined in detail by Weissman and Koe [76] in an attempt to understand the mechanism of its toxicity. Since the 1940s fluoro-acetoacetate has been recognised as one of the most toxic substances in nature. Administration of (3-F)Tyr in experimental animals has been shown to lead to similar convulsions in mice as those caused by fluoro-acetoacetate poisoning, only the effects are more intense.[77] Toxicity of fluoro-acetoacetate is generally believed to be due to its conversion into fluorocitrate which is a potent inhibitor of the citric acid cycle.[68] Lethality in mice was brought about by an elevation of tissue citrate and fluoro-acetoacetate concentrations by administering higher doses of (3-F)Tyr. These data confirmed that 3-fluorotyrosine is metabolised by enzymes specific for Tyr into fluoro-acetoacetate, with its subsequent conversion to fluorocitrate in the citrate cycle.[76] The concept of metabolic blockage by fluorinated tyrosine analogues is presented in Scheme 3.

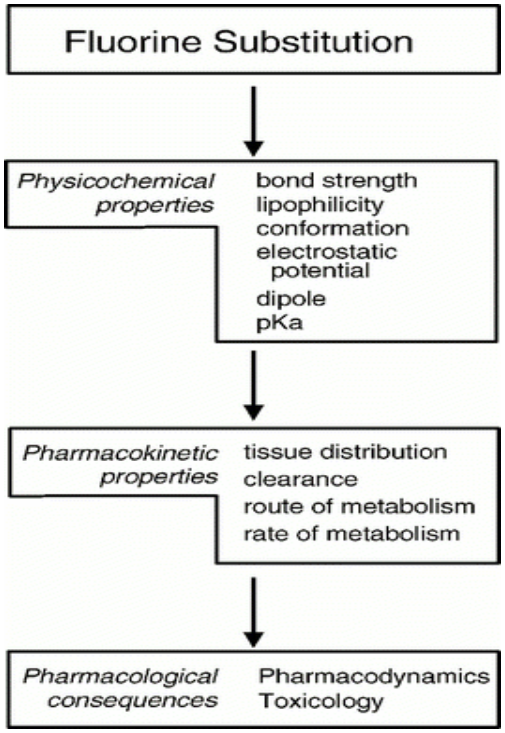


Figure 8. Possible alterations of biological functions induced by substitution of fluorine in drugs.

As already discussed, the presence of fluorine-containing functional groups can enhance the lipophilicity of the molecule, thereby affecting the partitioning of drugs and passive diffusion into membranes. Fluorine substitution can also influence the tissue distribution of a drug, and fluorinated drugs have the distinct advantage that their *in vivo* tissue pharmacokinetics can be monitored by noninvasive ^{19}F-labelled magnetic resonance spectroscopy.[78] Substitution of fluorine in place of hydrogen at the site of oxidative attack can block metabolism, or deflect metabolism along an alternative route. A class of drugs containing fluorine that has earned its place in the pharmacy is the selective serotonin uptake receptors: fluoxetine, fluvoxamine, etc. Fluoroquinolones (for example norfloxacin), on the other hand, are effective against gram positive and negative pathogens, and have applications in respiratory, urinary tract, skin and soft tissue infections. Paracetamol, the most frequently used analgesic, produces a reactive metabolite, N-acetylbenzoquinoeimine, known to cause hepatic necrosis if the drug is taken as an overdose. Introduction of fluorine at the 2- and 6- position decreases hepatotoxicty, but also leads to a substantial loss in the analgesic activity. This illustrates how fluorine substitutions can be used to dissociate the pharmacological and toxicological properties of a drug when a toxic metabolite has been identified and can therefore have complex effects on drug metabolism.[68] Figure 8 outlines the importance of fluorinated subgroups for possible drug dispositions, and the consequent impact on drug efficacy and drug toxicity.

4.5.3.2 Human recombinant annexin V and β galactosidase as carriers of fluorotyrosines

β galactosidase, one of the oldest and most well characterized proteins, was also the oldest model for substitutions with fluorinated tyrosines.[72; 73] β galactosidase is a protein containing 1030 amino acids having a total of 49 Tyr residues and 52 Phe, uniformly distributed throughout the protein. In recent years, Annexin V, a predominantly α-helical human recombinant protein was also extensively used as a model protein for biochemical, biophysical and structural studies upon replacements with fluorinated tyrosines and phenylalanines.[67] It contains only 12 Tyr-residues more or less uniformly distributed through the protein molecule.

Both proteins were substituted by esoteric (and toxic) fluoro-tyrosines in order to serve as candidates for specific carriers ('shuttles') in targeted drug delivery. This means that drug-like amino acids such as fluoro-tyrosine are converted into inactive prodrug forms upon covalent integration into polypeptides. Their delivery into e.g. tumour tissues should induce prodrug conversion into the active drug by its release from the polypeptide chain upon carrier protein internalisation. A similar concept was proposed for peptides as well; a

variety of synthetically substituted small peptides termed "smugglins" or "portable transporters" were proposed as carriers for some antibacterial amino acids, i.e. as vehicles for drug uptake.[79]

4.5.3.3 Biophysics and biomedicine of fluorinated phenylalanines

As a strictly hydrophobic amino acid, Phe is normally non reactive, often found buried in the protein cores where it contributes to stacking interactions. However, it can also be found (seldom enough) at surfaces where it usually contributes to receptor-ligand binding or protein-membrane interactions.[80] Experiments performed as early as the fifties proved fluorinated analogues of Phe (*o*-, *m*- and *p*-FPhe) to be translationally active and capable of incorporation into certain enzymes.[50] These were, in fact, the earliest incorporation experiments that paved the way for more recent systematic incorporation of fluorinated Phe analogues into recombinant proteins, endowing them with interesting optical and thermal properties as well as with altered kinetics. The discovery of a bacterial PheRS variant with relaxed substrate specificity by Kast and Hennecke[30] resulted in a further expansion of translationally active analogues and surrogates of Phe far beyond the fluorinated ones shown in Figure 7.

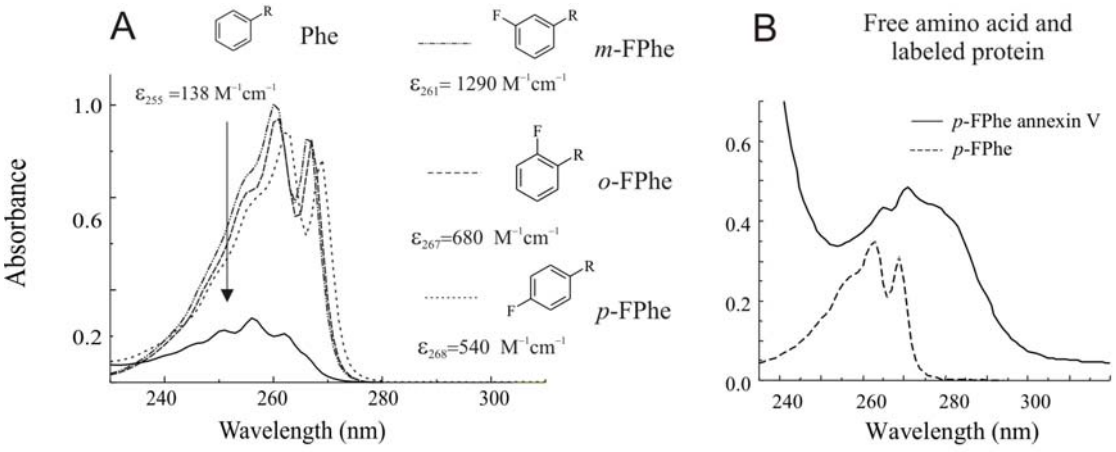


Figure 9. Spectral enhancement of Phe ultraviolet absorbance (UV) spectra by fluorination. (A) UV-spectra of 1mM Phe, o-FPhe, m-FPhe and p-FPhe measured in the range 230 - 320 nm in neutral aqueous buffer. (B) UV-spectra of 50 μM p-FPhe and 50 μM p-FPhe Annexin V are measured under the same conditions. The molar extinction coefficients (εmax) of the absorbance maxima are given in M-1cm-1.[55]

Perhaps the most interesting optical properties of fluorinated Phe analogues were reported recently. Fluorination of the benzene ring of Phe causes a dramatic enhancement of its optical properties having characteristic spectral features termed ‘fluorophenlyalanine

fingers' (FF-Fingers) seen as two distinct peaks in the spectral region between 260-270 nm.[55] In the case of *o*-FPhe, dual absorbance peaks appear at 260 and 267 nm; *m*-FPhe gives rise to peaks at 261 nm and 267 nm, whereas *p*-FPhe exhibits its dual maxima at 262 and 268 nm respectively. It is worthwhile noting that FF-fingers are preserved and integrated in an additive manner into the UV spectra of labelled proteins. The extinction coefficient of fluoro-phenylalanines is almost ten times higher in comparison to Phe by itself.[55] A simple visual inspection of the UV spectra of labelled proteins reveal FF fingers even in the case of proteins having a quantitative predominance of Tyr and Trp chromophores, since FF-fingers appear in a different spectral window. Indeed, global replacement of Phe residues by *o*-, *m*-, and *p*-FPhe analogues results in proteins having unique FF-fingers seen as two separate peaks in the spectral region between 260-270 nm as shown in Figure 9. These unique spectral features enable simple and non cumbersome approaches in distinguishing the labelled target protein among other non labelled cellular proteins.

L-Phe is an essential amino acid that plays a key role in the biosynthesis of other amino acids, including L-Tyr, and is used by the brain to synthesise several important neurotransmitters, principally L-DOPA, dopamine, epinephrine and norepinephrine.[81] Norepinephrine, a chemical that transmits signals between nerve cells and the brain releases signals to trigger alertness, reduces hunger pains, functions as an antidepressant, and helps improve memory. Interestingly, D-Phe is a non-nutrient amino acid that has been shown to inhibit the breakdown of enkephalins, opiate-like substances in the brain and is used for the treatment of a broad spectrum of therapeutic disorders like osteoarthritis, Parkinson's disease, rheumatoid arthritis, and pain from various causes, such as chronic back pain and migraines.[82] In 1951, Armstrong and Lewis, while designing diets for rats meant to exclude Phe, tried using Phe-substituents like *o*- and *p*-FPhe instead of Phe.[83] They found that the growth of rats were inhibited, and the animals could neither recognise nor utilise the structurally similar amino acids. Kaufman, while searching for the enzyme responsible for *in vivo* defluorination of *p*-FPhe found Phe-hydroxylase, the enzyme responsible for converting *p*-FPhe into tyrosine, although one sixth as efficiently as is Phe.[84] Such early experiments marked the crucial place of Phe as a biosynthetic precursor of ring hydroxylated amino acids and thus of phenolic biogenic amines. Both *o*- and *m*-FPhe are also substrates for Phe-hydroxylase, but instead of converting it to native Tyr (like *p*-FPhe), they are transformed into highly toxic tyrosine analogues (3-F)Tyr, and (2-F)Tyr (see previous Section). In addition *m*-FPhe has a medical relevance since it is a weak inhibitor of fungal growth and has been exploited in a prodrug approach to fungal

chemotherapy. Some di and tripeptides containing *m*-FPhe have proved to be active against several fungal organisms.[79]

4.5.3.4 Spectroscopy of Fluorinated Tryptophans

Tryptophan (Trp) as an unique amino acid, is an especially attractive substitution target for protein engineering and design studies for several reasons: (i) the diverse and rich indole chemistry offers numerous analogues/surrogates to be tested for translation activity; (ii) it has relatively low abundance in proteins and a single triplet in the genetic code (UGG); (iii) it possesses special biophysical properties which allow for its participation in numerous interactions in proteins (π–π stacking, hydrogen bonding, cation-π interactions); (iv) it is the main source of UV absorbance and fluorescence of proteins.[85] A very common approach to studying functional roles of Trp-side chains in proteins is to use site-directed mutagenesis for their alteration. In most cases, Trp residues are mutated to phenylalanines or tyrosines in the attempt to minimise structural perturbations by replacing one aromatic, planar moiety with another. There are, however, numerous examples where the scope of this strategy is rather limited. For example, Trp residues may be essential for structural integrity and functionality of the proteins, such as Trp57 in *Aequorea victoria* green fluorescent proteins (*av*GFP) and thus are not replaceable by most of the remaining 19 canonical amino acids.[86] Furthermore, even if such replacement is possible, local structural perturbations may alter the spectral contributions of other chromophoric groups in the target protein. Finally, routine site-directed mutagenesis of Trp with its most similar canonical counterparts Phe or Tyr always brings relatively large alterations either in size, charge, chemical nature or molecular volume.

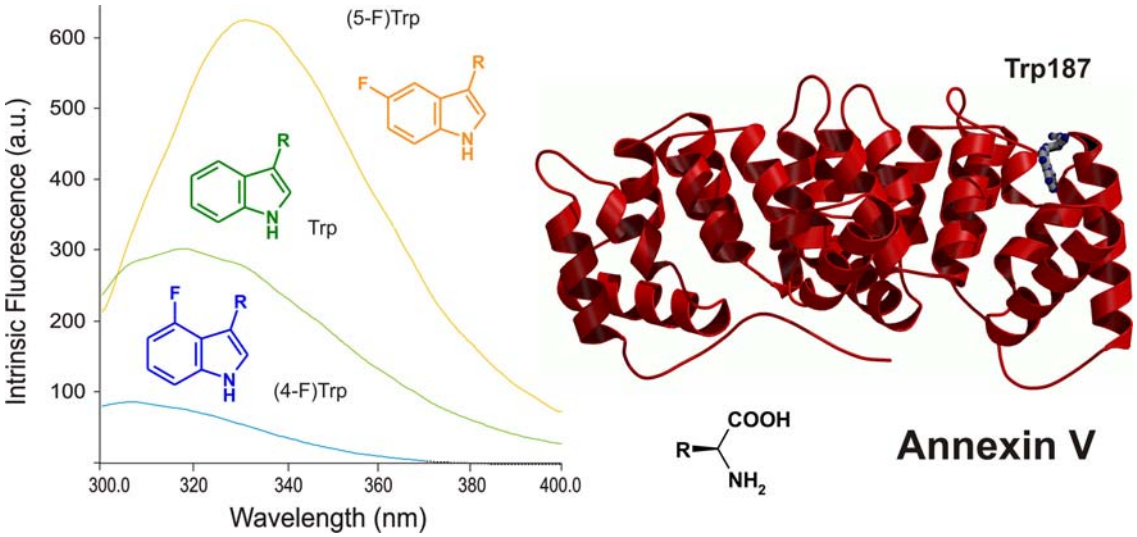


Figure 10. Manipulation with Annexin V fluorescence by fluorinated Trp-analogues (4-F)Trp and (5-F)Trp. The fluorescence emission spectrum of the native human Annexin V (green, excitation at 280 nm, 0.5 μM protein concentration in neutral aqueous buffer (10 mM Tris/ HCl pH 7.5)) could be 'silenced' by replacement of its single Trp157 by (4-F)Trp (blue). In the absence of Trp fluorescence due to its substitution by (4-F)Trp, the fluorescence of Tyr-residues in the spectrum of human recombinant Annexin V is revealed (λ_{max} =307 nm). On the other hand, the fluorescence intensity of this protein can be increased by the integration of (5-F)Trp into the protein structure which also causes a red-shift in its emission maximum by 14 nm.[66; 67] The side view (right) of the three-dimensional structure of human recombinant Annexin V represented as a ribbon plot with the single Trp187 marked as balls-and-sticks.

Fluorinated Trp analogues were often successfully used for ^{19}F-NMR studies on ^{19}F-labeled proteins facilitating chemical shift calculations.[87] On the other hand, it is well established that fluorinated tryptophans can provide unique spectroscopic properties when introduced into proteins. This should not be surprising since the UV-transition of the indole is composed of two almost isoenergetic and overlapping transitions ^1L_a and ^1L_b that are orthogonal in polarisation.[88] In this context, substitutions of atoms or functional groups in the indole moiety might affect the ^1L_a and ^1L_b dipole transition moments to different extents, thus modulating the complex spectral properties of Trp. For example, 4-fluorotryptophan is a non-fluorescent analogue. Indeed, Annexin V containing the non-fluorescent (4-F)Trp exhibits no fluorescence emission upon excitation at 280 nm, as shown in Figure 5. The emission maximum at 307 nm obtained upon excitation at 280 nm identifies unambiguously the contributions of Tyr residues to the emission spectrum. As the energy absorbed by (4-F)Trp is not released as fluorescence, it has to be transferred to other residues. Indeed, Hott and Borkmann demonstrated that (4-F)Trp is significantly more photoactive than the parent Trp, with photolysis quantum yields 7 times larger than that of Trp.[89] Additionally, (4-F)Trp displays fluorescence at 77 K and 295 nm and a phosphorescence comparable to that of Trp at this temperature. Recently it was found that (7-F)Trp exhibits similar properties.[90]

On the other hand, both (5-F)Trp and (6-F)Trp-containing Annexin V exhibit a red shift (13 and 16nm) if compared to the wt-protein (325 nm) and a significantly increased quantum yield upon excitation at 280 nm (Figure 10). Excitation at 295 nm leads to a 3 nm red shift of (5-F)Trp Annexin V, and about 6 nm of (6-F)Trp-Annexin V.[66] There is still no satisfactory explanation of these phenomena.

The most important motivation to use fluorinated Trp-analogues (4-F)Trp and (5-F)Trp in this study is to achieve much subtler alterations in its side chain structure, such as single atom exchanges ("atomic mutations") which should simplify interpretation of the experimental data. In previous studies performed in our laboratory the effects of amino acid substitutions, often to the level of single atom exchange, are immediately observable

in the optical properties without significant changes of the secondary and crystal structure of the proteins.[85] In this study, the influence of fluorination of the single Trp57 in EGFP and fluorination of the additional Trp66 in ECFP, on the spectral properties of these proteins was studied by using the following Trp analogues: (4-F)Trp, (5-F)Trp, (6-F)Trp and (7-F)Trp (Figures 10 and 11).

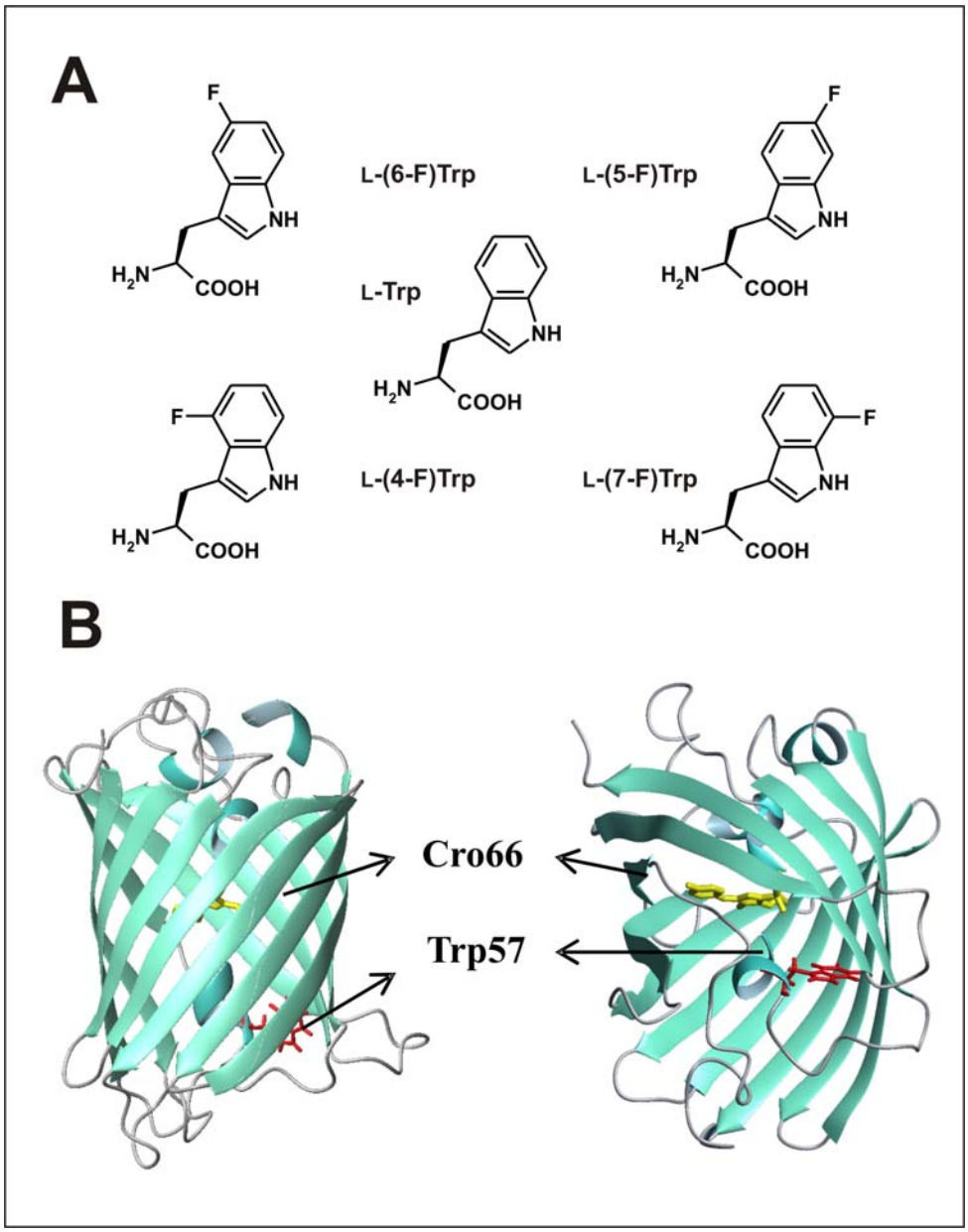


Figure 11. Tryptophan and its fluorinated analogues (**A**) together with *av*GFP as model protein. (**B**) Note that two mutants of *av*GFP have been used in this study: The ECFP with an indole in the chromophore (Cro66) and at position 57 (Trp57) was used for direct chromophore redesign with Fluorinated Trp-analogues. To investigate the influence of a single Trp57 on chromophore spectral properties, EGFP was used as the test system since it contains a tyrosyl-moiety (Tyr66) in chromophore.

4.5.3.5 Trifluoromethionine

Methionine (Met) participates in protein structures with both hydrophobic interactions and hydrogen bonding; it is relatively rare as it accounts for about 1.5 % of all residues in proteins of known structures. Normally it is inaccessible to solvent with only 15 % of all Met residues exposed to the surface. Met residues are significantly more flexible than those of both leucine and isoleucine with their branched and more rigid side chains. Due to the rarity of the Met residues in proteins, their full replacement by TFM would represent an almost site-specific mode of genetically encoded introduction of the trifluoromethyl moiety into large recombinant proteins. The trifluoromethyl group is almost twice as hydrophobic as the methyl group, a property which was widely used to suppress metabolic detoxification or to increase the bioavailability of many pharmaceutically active substances.[64] The most significant difference among these amino acids is the increased steric bulk (about 15%) of noncanonical TFM (84.4 $Å^3$) when compared with that of the canonical Met (71.2 $Å^3$).[70]

Up to now, only Honek and co-workers have provided experimental reports about the possibility of TFM and DFM incorporation into recombinant proteins.[91] They demonstrated that Met-replacements with TFM into phage lysozyme at positions 1, 14 and 107 were possible to levels up to 31 – 70%. In this study it was planned to reproduce these results by using 2M-EGFP and Annexin V as model proteins in the context of standard protocols of the SPI-method, and with genomic levels of MetRS activity (2M-EGFP) as well as with enhanced intracellular amounts of MetRS (Annexin V).

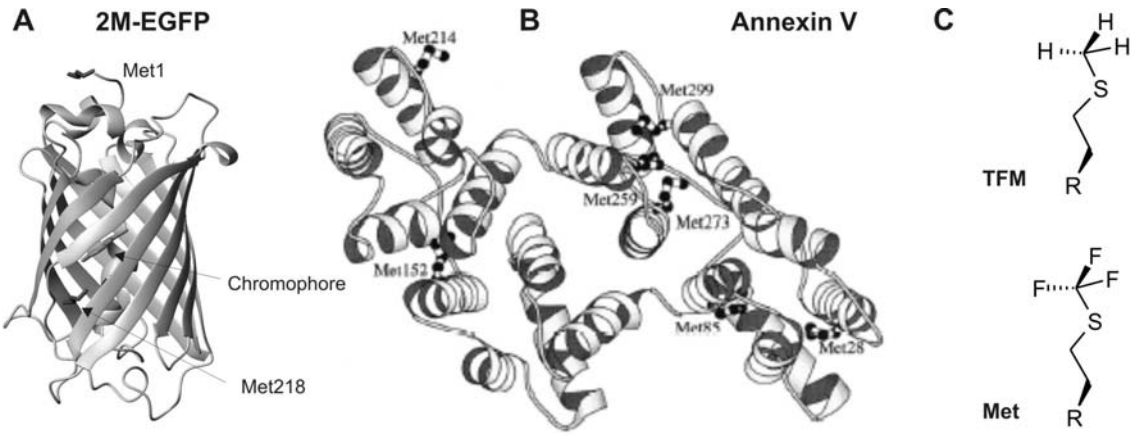


Figure 12. Three-dimensional structure of 2M-EGFP and wt-Annexin V with marked Met residues, and chemical structure of Met and TFM. (**A**) The side view of the three-dimensional structure of 2M-EGFP with marked Met-residues (sticks). (**B**) Ribbon plot of wt-Annexin V (top view) with marked 7 Met-residues

marked as balls-and-sticks. (**C**) Chemical structures of L-Met and L-TFM (R denotes amino-carboxyl moiety).

For Met→TFM substitution studies the human recombinant Annexin V (7 Met, 35810 Da), an exclusive α-helical protein involved in membrane binding processes in a calcium-dependent manner, was chosen. While this protein contains 38 hydrophobic Leu residues (~ 12% of the total amino acid composition) the Met residues constitute only 2.2% of the amino acid composition (7 Met residues with an eighth N-terminal Met excised during expression in *E. coli*).[92] As shown in Figure 12, these are located in different parts of the molecule, *i.e.* in hydrophobic environments (Met28, Met80, Met152 and Met273), in the partially solvent-exposed hydrophobic minicore (Met259 and Met299) and on the outer surface exposed to the bulk solvent (Met214). In previous studies it was demonstrated that Met substitutions by selenomethionine, telluromethionine, or norleucine do not significantly affect Annexin V activity (i.e. binding on membranes in the presence of calcium).[93] Annexin V also has 12 tyrosine residues that make up about 3.8% of the total amino acid composition; they were used as substitution targets for fluorination at *ortho*-, and *meta*-positions of tyrosyl-moieties.

In contrast, the 'enhanced' green fluorescent protein (EGFP, Phe64Leu/Ser65Thr, 15 Leu, 27740 Da), which is a derivative of wt-GFP from *Aequorea victoria*, is predominantly a β-sheet containing protein characterized by a 11-stranded β-barrel that rigidly holds a chromophore anchored at a central helix within the core of the protein structure. In particular, the Met→TFM replacement was attempted by using a mutant of EGFP (2M-EGFP, 27676 Da, Met153Thr/Met233Lys/Met78Leu/Met88Leu)[94] that contains only two Met side chains: Met218 that is buried and the N-terminal Met1 that is solvent-exposed (Figure 12). This mutant shares the general architecture of EGFP and replacement of both its Met residues does not affect the spectral properties of the chromophore.

4.5.4 Fluorination of aromatic residues in autofluorescent proteins

The green fluorescent proteins from *Aequorea victoria* (*av*GFPs) as well as their fluorescent homologs from *Anthozoa* corals and other species such as *Renilla, Discosoma,* and *Anemone* are indisputably some of the most studied and well characterised proteins.[95; 96] The chromophore of these proteins is completely encoded in the amino acid sequence and can fluoresce only after correct folding of the β barrel and in the presence of molecular oxygen.[97] The corollary of such fluorescence behaviour shows that onset of fluorescence could be used as an indicator for completion of the β barrel formation. Since the *Aequorea*

is indigeneous to the cold Northwest Pacific waters, chromophore formation and maturation takes place well below 37 °C.[98] The function of the barrel (diameter ~ 24 Å and height 42 Å)[99] is to protect the chromophore from quenching by oxygen. Several polar residues and water molecules form a hydrogen bonding network around the chromophore. The *av*GFP chromophore (4-(p-hydroxybenzilidene)imidazolid-5-one) is completely encoded in its amino acid sequence.[100] It is autocatalytically formed by the post-translational reaction between the side chains of residues 65-67 having molecular oxygen as its only externally required component.[97; 101] The fluorophore of *av*GFP consists of residues Ser65-*dehydro*Tyr66-Gly67 whose cyclized backbones forms the imidazolinone ring. As mentioned earlier, the unique absorbance and fluorescence properties of GFP are not an inherent property of the isolated fluorophore.

Tsien[102] classified all known GFPs into seven classes depending on their chromophore photophysics. The avGFP, extracted originally from the jellyfish has the most complex spectra among all known GFPs. It has a dual excitation peak, one at 395 nm, and the other at 475 nm. Protonated or neutral chromophores are believed to contribute to the peak at 395 nm while the peak at 475 nm is due to the contribution of the deprotonated or anionic chromophores. The coexistence of the neutral and anionic chromophore has a great many disadvantages for cell biological applications. Classical mutagenesis has yielded several classes of proteins with distinctly improved properties. Comparisons of such mutants are shown in the fluorescence profile in Figure 13. Individual chemistries and photochemistries of such enhanced mutants are extensively discussed below.

Class I is represented by the wt-GFP from *Aequorea victoria,* whose chromophore is in an equilibrium between the phenol and phenolate form. It has two excitation peaks at 395 nm and at 475 nm. In the class II (e.g. EGFP) where Ser65 is replaced by Thr, Ala or Gly, the chromophore contains a phenolate ion and has no excitation peak at 395 nm. Class III (e.g. sapphire; neutral phenol) has a Thr203Ile mutation whereby loss of the alcoholic proton prevents hydrogen bond formation to phenolate and stabilises it (only 399 nm excitation). Class IV (presumably Thr203Tyr mutant, e.g. EYFP) has a stacking π-electron system of the tyrosine side chain on the π-electron system of the chromophore which causes a red-shift due to relative stabilisation of the photoexcited state. The basic characteristic of the class V is the presence of an indole moiety in the chromophore (e.g. ECFP) with other compensatory mutations in the chromophore closer or distant environment. In terms of the spectral properties, ECFP is an intermediate between EGFP and “blue” fluorescent protein (BFP) which contains an imidazole in the chromophore and belongs to the class VI. The shortest wavelengths in the GFP emission are achieved in the

mutant Tyr66Phe due to the presence of the phenyl moiety in the chromophore. The spectra of this class (VII) are not shown in the Figure 13.

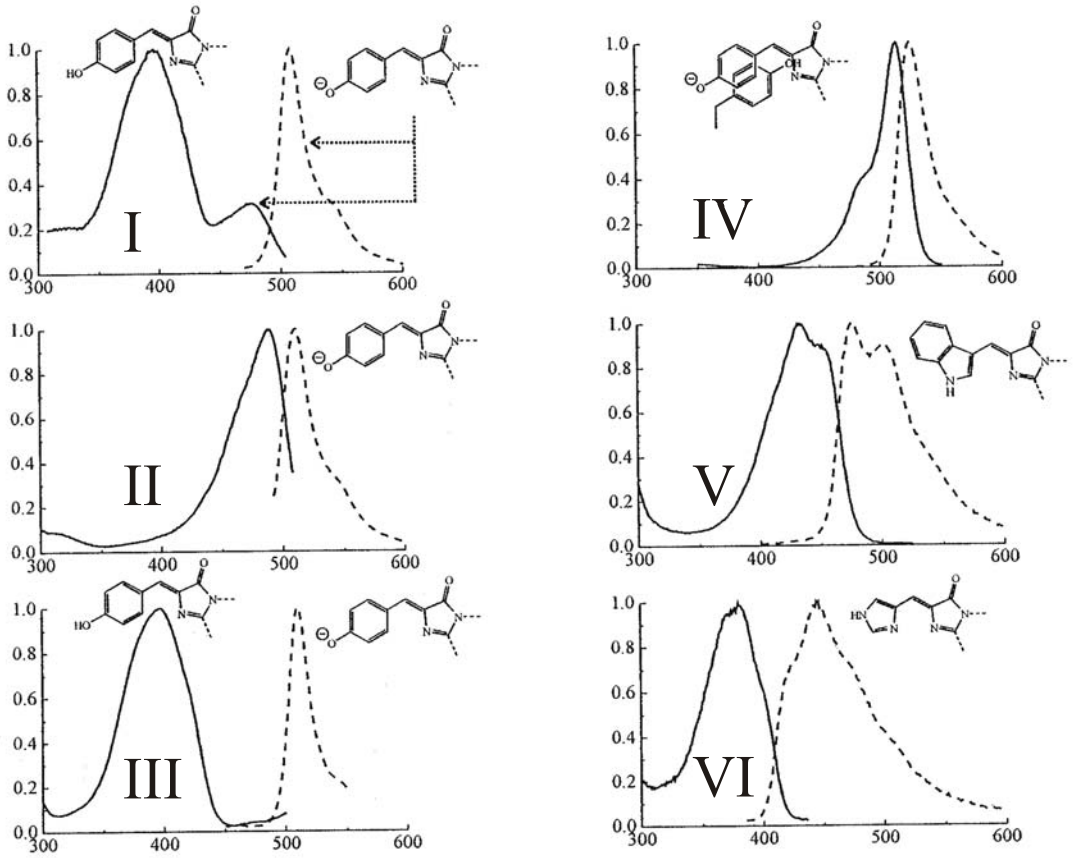


Figure 13. Major classes of GFP proteins derived on the basis of their spectral characteristics.[102]

4.5.4.1 Enhanced Green Fluorescent Protein, EGFP

Enhanced green fluorescent protein, EGFP, (Phe64Leu/Ser65Thr; λ_{max} = 488nm, ε_M = 35570 $M^{-1}cm^{-1}$, λ_{em}= 509 nm), is the most commonly used mutant among *av*GFPs whereby the mutation Ser65Thr is most probably responsible for the existence of the chromophore in the predominantly phenolate form. It contains 11 Tyr and 12 Phe residues, Tyr66 being an integral part of the chromophore (Figure 14). This *av*GFP class contains the phenolate anion in the chromophore and is endowed with a combination of superior brightness along with simple excitation and emission spectra. These properties are brought about by mutation of the residues S65T and F64L, which have wide applications in routine cell biology. It has a single excitation peak at 488 nm with emission ranging from 507-509 nm. On the other hand, having a neutral phenol in the chromophore causes a suppression of the

peak at 475 nm and a concomitant enhancement of the one at 399 nm. This can be brought about by a mutation of Thr203Ile, whereupon the –OH of the threonine is missing, and the chromophore anion cannot be adequately solvated, leaving the chromophore neutral.

4.5.4.2 Enhanced Yellow Fluorescent Protein, EYFP

Enhanced yellow fluorescent protein, EYFP, (Ser65Gly/Val68Leu/Ser72Ala/ Thr203Tyr) contains 12 Tyr residues, Tyr66 being an integral part of the chromophore. It carries an additional Tyr at position 203 (occupied by threonine in wt-*av*GFP) that lies in close proximity (about 3.5 Å) to the chromophore as observed in its X-ray structure.[103] It is, to date, the most red shifted *av*GFP mutant designed by conventional protein engineering methods (λ_{em} = 527 nm) (Figure 13). On this basis a π-π stacking interaction was postulated as being responsible for the observed red-shifted emission. Conversely, dynamic measurements based on studies of a set of single-site aliphatic mutants at position 203 revealed a decrease in stabilization of the deprotonated form as the main origin of the red shift.[104] Therefore, fluorination of Tyr203 in EYFP was expected to offer a good model to study the effects of micro-environment fluorinations on the chromophore spectral properties and, particularly, on the nature of the red shift in its fluorescence.

This is the class of commercially available *av*GFP fluorescent protein mutants that have the longest wavelengths reported so far. Such shifts in wavelength are brought about by a replacement of Thr at position 203 by Tyr. Mutants were rationally designed from the crystal structure of S65T GFP with the supposition that additional polarizability around the chromophore and π-π interactions would increase the excitation and emission wavelengths. Substituting the amino acid at position 203 with any one of the four aromatic residues His, Trp, Phe, and Tyr enhances the excitation and emission up to 20 nm.[105] The crystal structure verified stacking interactions between the aromatic ring and the phenolate anion of the chromophore.[106] The following point mutations were performed on the S65T GFP: T65G, V68L, S72A, T203Y, resulting in an λ_{exc} of 514 nm, and a λ_{em} of 527 nm.[107]

Point mutations on the EYFP yielded a variety of proteins where the tyrosine at position 203 was replaced by phenylalanine. Minks *et al.* showed that the optical properties change significantly upon fluorination of the benzene ring of Phe[55] (Figure 13). Quantum yield of the fluorinated phenylalanine increases about ten times in comparison with the wild type variety. Based on such observations, the mutant EYFP Y203F was designed to be fluorinated with analogues of phenylalanine.

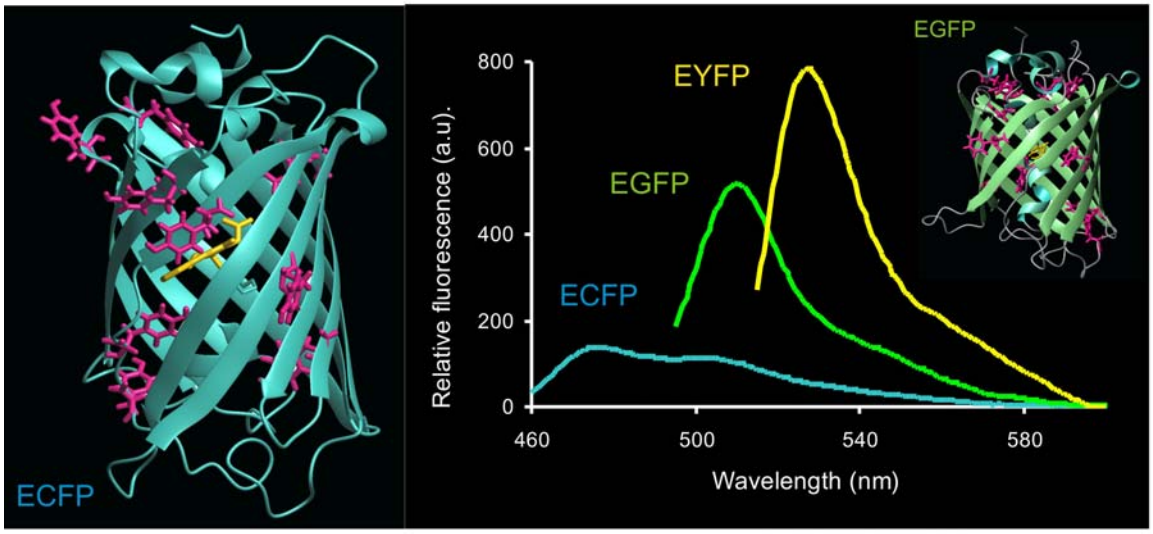


Figure 14. Structures and fluorescence of major avGFP classes: EGFP, EYFP and ECFP. The avGFP mutants EGFP (left) with marked Tyr-residues and EGFP (right above) with marked Phe residues and Annexin V (red) as model proteins for substitutions of Phe and Tyr-resides with their fluorinated counterparts. Note that the fluorescence intensity as well as red shift is most pronounced in EYFP whereas ECFP shows its characteristic double humped peak (emission spectra were measured in sodium phosphate buffer pH 7.0 at 20°C).

4.5.4.3 Enhanced Cyan Fluorescent Protein, ECFP

Enhanced cyan fluorescent protein, ECFP, (Ser65Thr/Tyr66Trp) has an indole ring in the chromophore, which extends toward Val150 and Phe156. In addition, its accommodation is favoured with other compensatory mutations Phe64Leu/Asn146Ile/Met153Thr/Val163Ala.[86] The most curious and not so well understood feature of ECFP is its double-humped excitation and emission peaks (λ_{max1}= 434 nm; ε_{M1}=24700; λ_{max2}=452 nm; ε_{M2}=23600 $M^{-1}cm^{-1}$; λ_{em1}=477 nm, λ_{em2}=504 nm). It was recently suggested that this feature might well derive from the dynamic conformational changes of the neighbouring residues Tyr145 and His148.[87; 108] It contains only 10 Tyr residues while the 11th one is replaced with Trp at position 66 (chromophore) and thus was used as a model protein to study the influence of fluorination of the bulk Tyr-residues outside the chromophore.

The chromophore in this case has neither a phenol nor a phenolate, but an indole, the accommodation of which requires substantial remodelling of the chromophore cavity. Several amino acids were mutated by random mutagenesis that restored fluorescence to fairly satisfactory levels (Gln146Ile/Met153Thr/Val163Ala). Excitation and emission wavelengths are at 436 and 476 nm, respectively, having an intermediate somewhere between neutral phenol and anionic phenolate chromophores. A curious feature of this

class of proteins is their double humped excitation and emission peaks. This feature has so far been unexplained, but has been believed to be due to the vibration levels or other quantum states that equilibrate within the lifetime of the excited state.

4.5.4.4 Red fluorescent protein from *Discosoma striata*, *ds*Red

The 28-kDa red fluorescent protein, *ds*Red is a recently cloned fluorescent protein responsible for the red coloration around the oral disk of the Indo-Pacific reef coral *Discosoma striata*. Like *av*GFP, the tetrameric *ds*Red is an 11-stranded β-can containing a central α-helix with the chromophore, almost identical in topology. The structural rigidity and efficient shielding of the chromophore from bulk solvent are also key factors for the efficient quantum yield of fluorescence observed in *ds*Red. The chemical mechanism of the formation of the chromophore in *ds*Red is similar to that of *av*GFP, viz autocatalytic cyclization of residues Gln 66, dehydroTyr 67 and Gly 68. However, *ds*Red shows an additional oxidation reaction at backbone atoms of the Gln 66 residue to extend the conjugated π-electron system (Figure 15).[109] The extended π-bonding system of the *ds*Red chromophore causes an additional ring resonance stabilisation by greater delocalization of electrons upon photoexcitation. Therefore, *ds*Red with emission maxima of 583 nm is among the longest red-shifted autofluorescent proteins found in nature or generated by classical protein engineering.[96] The absorbance spectra show a major peak at 558 nm and two minor shoulders at 526 and 490 nm. In addition, *ds*Red was found to be stable under harsh conditions of pH and to be extremely resistant to photobleaching. This protein exhibits a bright red fluorescence, and as a potential expression marker and fusion partner, would be an ideal complement to the *Aequorea victoria* GFPs. Extensive work has recently been performed on the biochemical and photophysical properties of *ds*Red. However, the protein forms closely packed tetramers and gives rise to a proportion of curious green protein that is believed to be the immature chromophore.[110] Recently, Campbell *et al.* produced a monomeric red fluorescent protein by directed evolution, although certain refinements are desirable before it can rival GFP as fluorescently tagged fusion proteins.[111]

Potential applications of this class of FPs will be to provide additional colours for multiwavelength monitoring of fusion chimeras or reporters coupled with the enhanced *av*GFP mutants. The red FPs should be a good acceptor of FRET from the green variants produced, and could therefore have wide applications as an expression tracer.

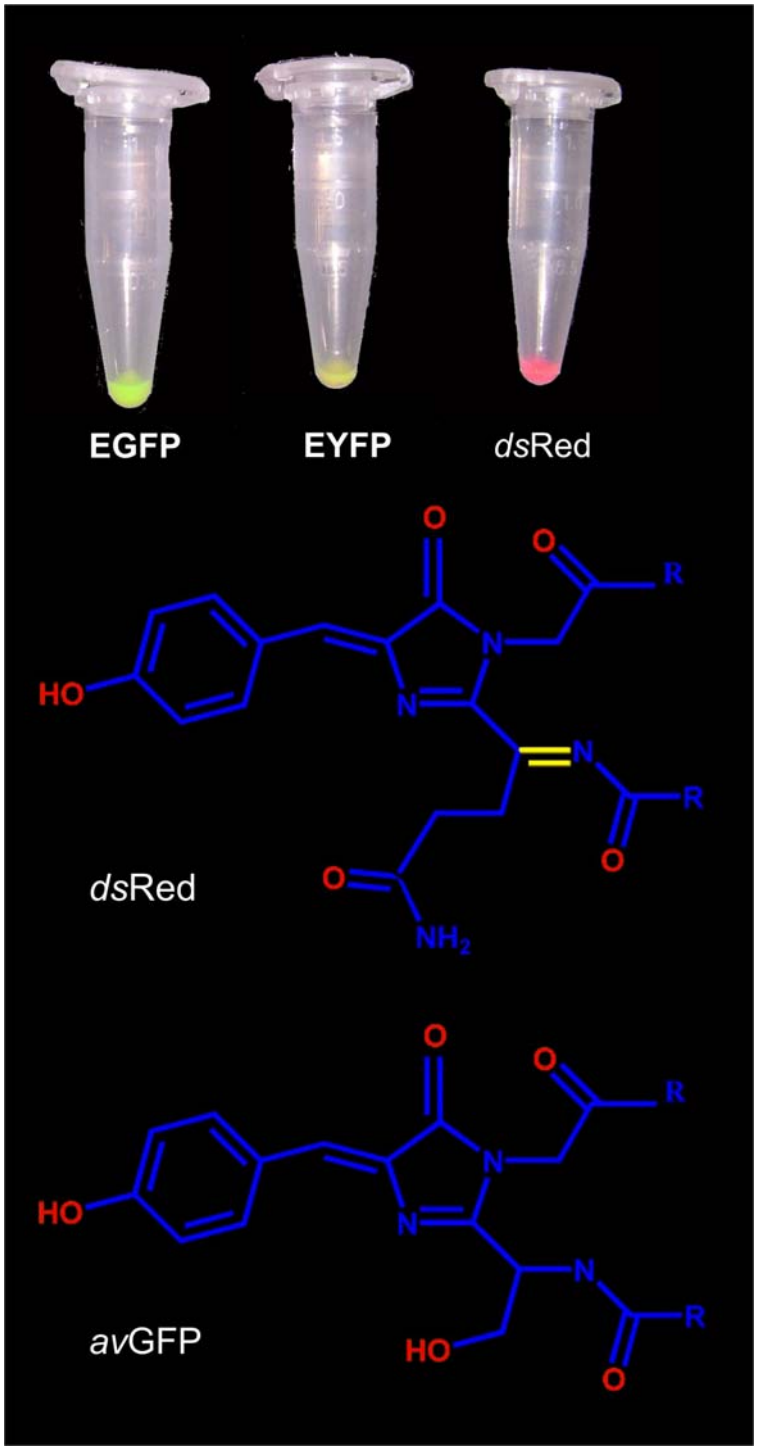


Figure 15. Cell pellets of *av*GFP-derived autofluorescent proteins (EGFP, EYFP) and *ds*Red (upper part). Lower part: comparison between *av*GFP and *ds*Red chromophores. The main difference between them is the chromophore of *ds*Red that contains an extra double bond (yellow) from Gln66 residue which augment the chromophore and in this way causes the red-shift.

5 MATERIALS AND METHODS

5.1 Materials and instruments

5.1.1 Chemicals and accessories

Solvents like acetonitrile (LiChrosolv HPLC-grade), were purchased from Merck (Darmstadt). For SDS polyacrylamide gels, ready-made stock solutions of 30% acrylamide and 0.8% Bis acrylamide were used, that were purchased from National Diagnostics, Atlanta. TEMED (*N,N,N',N'*-tetramethylethylenediamine) was purchased from Serva. Protease inhibitor cocktail tablets (Complete EDTA-free) were obtained from Roche Diagnostics, GmbH. Chelating sepharose and DEAE-Sepharose (Fast Flow) were from Amersham Biosciences and all components for the LB media were from BD Biosciences. Dialysis tubes of various cutoffs were purchased from Chemicon, Invitrogen. Isopropyl-thio-β-D-galactoside (IPTG), DNAse and RNAse were purchased from Roche Diagnostics and Sigma, respectively. BD Falcon polypropylene tubes (15 and 50 ml), and micro-centrifuge tubes (1.5 and 2.0 ml) were from Eppendorf and Cole Parmer, USA, respectively. All natural L-amino acids were purchased from Sigma, as were their analogues, with the exception of (2-F)Tyr which was purchased from SynQuest, Alachua. Centricon and Centriprep concentrators (10, 15, 30 kDa MWCO) were from Amicon Ultra, MILLIPORE. Ampicillin (sodium salt), lysozyme (from hen egg white) (FLUKA, Taufkirchen), Trizma base, guanidinium chloride, urea were purchased from Sigma. All other chemicals were obtained either from Fluka or Merck.

5.1.2 Preparation of (7-F)Trp

Most of the fluorotryptophans are commercially available but not (7-F)Trp which was prepared enzymatically as follows: 200 mg of 7-fluoro-indole (Aldrich) were mixed overnight with 200 mg of L-serine and 0.5 mg of pyridoxal phosphate (PLP) in 60 ml 0.1 M Tris-HCl (pH 7.8) in the presence of $\alpha_2\beta_2$ complex of Trp-synthase (200 units) under argon atmosphere at 37 °C. This reaction mixture was then filtered through a Centriprep-10 to remove the enzyme, concentrated and subjected to reverse-phase HPLC to obtain pure (7-F)Trp.[90]

5.1.3 Liquid Luria-Bertani (LB) Media and LB-Agar plates

Liquid LB-media was prepared by filtering and autoclaving (121 °C, 20 min, 100 bar pressure) a solution containing 10 g/L NaCl, 5 g/L Bacto™ Yeast extract and 10 g/L

Bacto™ Tryptone in water. Solid LB-media used for the plating of bacterial colonies was prepared by adding 6 g of Bacto™ Agar to 400 ml LB, autoclaved, cooled to about 40 °C and poured in plates, 20 ml each. If required, antibiotics were added before pouring (ampicillin: 100 mg/L; kanamycin: 70 mg/L). Antibiotics at appropriate concentrations could also be applied on the LB plate after solidification (i.e. cooling).

5.1.4 New minimal medium (NMM) and limiting concentrations of amino acids of interest

For incorporation experiments, New Minimal Medium (NMM)[92] containing 22 mM KH_2PO_4, 50 mM K_2HPO_4, 8.5 mM NaCl, 7.5 mM $(NH_4)_2SO_4$, 1 mM $MgSO_4$, 20 mM glucose, 1 μg/ml Ca^{2+}, 1 μg/ml Fe^{2+}, 0.001 μg/ml trace elements (Cu^{2+}, Zn^{2+}, Mn^{2+}, MoO_4^{2-}), 10 μg/ml thiamine, 10 μg/ml biotin and the appropriate antibiotics was routinely used. All amino acids except the one whose non-natural analogue needs to be incorporated were weighed individually (~50 mg per liter NMM) and dissolved in water having 2.2 ml/100ml 1 M KH_2PO_4 and 4.8 ml/100ml of 1 M K_2HPO_4.

Stock solutions (50 mM) of sparingly soluble aromatic amino acids like Trp and Tyr were prepared separately by adding a few drops of 25% HCl to facilitate dissolution and added to a final concentration of 0.3 mM in the growth media. Non-canonical amino acid stock solutions (50 mM) were prepared in the same way. Filtered stock solutions of all analogues were kept in the dark at –20 °C, and routinely used for expression and incorporation at a concentration of 0.25 mM without special precautions.

The canonical amino acid whose analogue needs to be incorporated was added in defined amounts, known as its limiting concentrations, since in those quantities they are found to hold the growth in check. These limiting concentrations usually ensure normal culture growth to an appropriate cell density of about 0.5 x 10^9 cells/ml (i.e. optical density at 600 nm (OD_{600}) of about 0.5), before addition of the non-natural analogue and induction of the synthesis of target protein. Using the auxotrophic strains listed in Table 4, the growth of the related culture was efficiently limited by adding 0.015 mM Trp (for Trp-auxotrophic strains), 0.03 mM Tyr (for Tyr-auxotrophic strains), 0.02 mM Phe (for Phe-auxotrophic strains) and 0.05 mM Met (for Met-auxotrophic strains).

5.1.5 Difco yeast nitrogen base without amino acids

In some cases it is difficult to incorporate the desired non-canonical amino acid into a particular target protein. In such cases it was found that supplementation of the NMM with a commercially available powder (Difco yeast nitrogen base without amino acids) containing additional vitamins, cofactors and minerals used normally for culturing yeasts

(BD-Diagnostic systems), vastly improved the level of incorporation. This is supposed to facilitate the incorporation of "difficult" amino acid analogues like trifluoromethionine, since the viability of the host auxotrophic bacterial strain is substantially improved. This powder is added to the media after dissolving in water and filtering (7.4 g in 100 ml H_20 is sufficient for 1 L NMM cultures).

5.1.6 Restriction enzymes, polymerases, ligases and reaction conditions

All restriction enzymes and polymerases used in this work, along with their buffers and accessories were procured from New England Biolabs (Frankfurt am Main). A restriction enzyme recognizes its specific palindromic sequence and cuts DNA only at that particular sequence of nucleotides, producing either blunt or sticky ends in the process. The following restriction enzymes were used in this work (details about their origin, reaction conditions and palindromes are available in the New England Biolabs Catalogues): *BamHI*, *NcoI*, *HindIII* and *EcoRI*.

DNA polymerases were used to extend the primers and synthesize copies of the target DNA. Only polymerases that are able to withstand high temperatures can be used, among which two of the most popular are *Taq* and *Vent* polymerases (Roche, New England Biolabs) extracted from bacteria that grow at extremely high temperatures. Several properties characterize polymerases like thermostability (its half-life at a specific temperature), extension rate (the number of dNTPs polymerized per second per molecule of DNA polymerase), processivity (the average number of nucleotides added by a DNA polymerase in a single binding event), and last but not least, fidelity (an intrinsic property of DNA polymerase that represents the frequency of correct nucleotide insertion per incorrect insertion).

T4 DNA ligase was purchased from Fermentas. This enzyme covalently joins the 5'-phosphate and the 3'-hydroxyl groups of linear double stranded DNA molecules having either blunt or sticky ends. The formation of this phosphodiester bond requires the presence of ATP to a concentration between 0.25 mM to 1 mM. Ligation requires the conversion of the quantity of DNA (in ng) into molar amounts. The ratio of the length of the vector DNA to insert DNA was always calculated; usually the reaction was performed with a 1:5 vector/insert ratio. Prior to ligation, samples were double digested with the relevant restriction enzymes. They were then subjected to the QIAquick PCR purification Kit that ensures cleanup of the enzymatic reactions, followed by DNA quantification on a 1% TBE gel. Ligase buffer containing ATP was added, reaction incubated for 2-3 minutes

at 42°C, followed by 2-3 minutes on ice. DNA ligase was added to the reaction mix at this point and ligation was allowed to proceed overnight at 16°C.

Dpn1, another enzyme purchased from New England Biolabs was used after PCR for point mutation to rid the methylated DNA of the parental strand thereby preserving only the freshly mutated (non-methylated) strand.

5.1.7 Chemicals, media and instruments for handling of tumour cell lines

Media used for cultivation of cell lines: Normal cellular growth for all examined tumor cell lines (see Methods) was achieved either in full or starvation McCOY's 5A medium + 15% FCS (Sigma) with or without foetal bovine serum. Trypsin for cell trypsinisation was purchased form Sigma Aldrich. Cell culture dishes were provided by different companies (CORNING® and Nunclon™); 24, 48 and 96 well plates were from Corning Incorporated costar®. For culture handling, multichannel pipettes (20 – 200 µl) from Pipetman® were routinely used whereas for aspiration of the media, Pipetboy plus IBS Integra Biosciences pipettes were used.

In order to control the composition of amino acids in the media as well as the toxicity of fluorinated amino acids for cultured mammalian cells, a synthetic medium was prepared for cell culture studies. Separate media mixes, each deficient in one of the three amino acids, tryptophan, tyrosine and phenylalanine were prepared from the Dulbecco's minimal media (Sigma Aldrich) comprising of the following ingredients (i) Earle's buffered salt solution (EBSS); (ii) MEM vitamin solution, (iii) MEM non essential amino acids, (iv) extra amount of glucose to a final amount 4.5 g/l (HEK293 cells require extra glucose). Serum used to supplement such minimal medium was dialysed against PBS with a 3000 MWCO membrane.

5.1.8 Reagents for protein delivery in cell lines

Lipofectamine (Invitrogen Gibco®), a liposome formulation comprising a mixture of polycationic lipid and a neutral lipid was used for transfection of adherent cell lines such as MCF10A with wt- and fluorinated green fluorescent proteins.

CHARIOT™ (Active Motif, Vinci Biochem) is a non cytotoxic transfection agent that can transport biologically active proteins, peptides and antibodies into cultured mammalian cells with high (65-95%) efficiency. When added to cells, the complex is rapidly internalised and dissociates, thereby delivering the target compounds directly. While the delivery of β-galactosidase works well, attempts to use it for native and fluorinated GFP delivery were unsuccessful.

BioPORTER (Sigma Aldrich) is another protein delivery reagent that has an unique cationic lipid based formulation (TFA-DODAPL:DOPE), composed of a 2:1 mixture of a cationic lipid, TFA- DODAPL, and a neutral lipid, DOPE. The BioPORTER, when mixed with the protein solution, reacts noncovalently with the protein creating a protective vesicle. This hydrated mixture is then added to cells and the BioPORTER/protein complexes attach to the negatively charged cell surfaces. The BioPORTER can then fuse directly with the plasma membrane and deliver the captured protein into the cells.

5.1.9 Materials and proteins for staining mammalian cell lines

DAPI (4'-6-diamidino-2-phenylindole) for staining cellular nuclei was purchased from Sigma; Annexin V-Cy3 kit (Sigma) allows detection of Annexin V bound to apoptotic cells by fluorescence microscopy, whereas intracellular detection of β-galactosidase was possible by *X-Gal* (5-Bromo-4-chloro-3-indolyl-ß-D-galactoside) (Sigma) according to the principles described in the Introduction.

5.1.10 Instruments

- Amersham Pharmacia Biotech, Freiburg:
CM52 Sepharose column
- Amicon, Beverly, MA, USA:
Microconcentrators Centricon 10 and 5 kDa
Microconcentrators Centriprep 10 and 5 kDa
- Beckmann, Munich:
Centrifuge J2-21 (rotors JA-20 and JA-10), Sorvall RC 5C Plus, DuPont, SLA 3300, Hettich Universal
Ultracentrifuge L7-55 (rotors Ti 45 and Ti 55.2)
UV/VIS-Spectrophotometer DU 7500
- Biometra, Göttingen:
Electrophoresis chambers
- Branson, Dietzenbach:
Sonifier 250
- Eppendorf, Netheler and Hinz GmbH, Hamburg:
Eppendorf tubes (1.5 ml and 2 ml)
Micropipettes 10 µl - 5 ml
- Greiner and Söhne GmbH & CoKG, Frickenhausen:
Petri dishes

- Hellma, Müllheim:

Precision cuvettes Quartzglass SUPRASIL®

- Infors AG, Bottmingen, Switzerland:

Shaker incubator RKF-125 for culture preparation

- Jobin Yvon, Lonjumeau, France:

CD-Spectrophotometer

- Merck Schott, Duran:

Glassware

- Millipore, Billerica:

Bottle top filters

- Perkin Elmer, Weiterstadt:

UV/VIS-Spectrophotometer Lambda 19, Fluorescence Spectrometer, Luminescence Spectrometer (LS50B) with digital software

- Radiometer Copenhagen, Kopenhagen, Denmark:

PHM 82 standard pH-meter

- Roth, Karlsruhe:

Spectrapore dialysis membranes (MWCO 12-16 kDa and 5-8 kDa)

- Sartorius AG, Göttingen:

Laboratory balances LC2200S

5.1.11 Antibiotics, column material and commonly used buffers

Stock concentration of antibiotics ampicillin (100 mg/ml), chloramphenicol (37 mg/ml) and kanamycin (70 mg/ml) were diluted 1000 fold after their addition into the growth media.

Spheroblast-buffer:

200 mM Tris/HCl pH 8.0

0.5 mM EDTA pH 8.0

750 mM saccharose

UZ-buffer:

2 mM Tris/HCl pH 8.0

0.2 mM EDTA pH 8.0

0.5 mM $MgCl_2$

10 mM NaCl

The UZ buffer was supplemented with:
0.1 mg/ml RNase
0.1 mg/ml DNase I
0.1% Triton X100
"Complete" EDTA-free protease inhibitor cocktail tablets (Roche, Penzberg)

Ni-Chelate-Column:
Column material: Ni-chelate Sepharose (Pharmacia)
Washing buffer:
100 mM sodium phosphate buffer pH 8.0
500 mM NaCl
Pre-elution buffer:
5 mM- 20 mM imidazole
100 mM sodium phosphate buffer pH 8.0
500 mM NaCl
Elution buffer
100 mM imidazole
100 mM sodium phosphate buffer pH 8.0
500 mM NaCl

Phenyl sepharose column:
Column Material: Phenyl Sepharose (Pharmacia)
Washing buffer:
20 % ammonium sulfate
20 mM Tris pH 8.0
1 mM EDTA
Elution buffer:
20 mM Tris pH 8.0
1 mM EDTA

GFP denaturing/renaturing buffer:
50 mM sodium phosphate pH 7.4
100 mM NaCl
1.0 mM $MgCl_2$
5.0 mM DTT

8 M urea
(renaturation buffer contains no urea)

5.1.12 SDS-polyacrylamide gel electrophoresis

For the qualitative assessment of protein expression and purification, the method of SDS-PAGE by Laemmli[112] was used. Gels having dimensions of 10.4 cm x 10 cm x 0.7 mm were of the following composition (for 8 gel plates):

Stacking gel (5%):
7.5 ml Protogel (National Diagnostics)
5.65 ml 1 M Tris/HCl pH 6.8
31.2 ml deionized H_2O (dH_2O)
0.45 ml 10 % SDS
0.21 ml 10 % ammonium persulphate (APS)
0.08 ml N, N, N´, N´ - tetramethylethylenediamine (TEMED)

Separating gel (12%):
36 ml Protogel (National Diagnostics)
17 ml 2 M Tris/HCl pH 8.8
36 ml deionized water
0.9 ml 10 % SDS
0.134 ml 10 % APS
0.044 ml TEMED

2 x SDS-loading buffer:
100 mM Tris/HCl pH 6.8
200 mM dithiothreitol
4 % SDS
0.2 % bromophenol blue
20 % glycerol

Running buffer:
0.13 M Tris/HCl pH 8.3
0.19 M glycine
1 % SDS

Staining solution (100 ml):

0.25 g Coomassie-Brilliant-Blue R 250 (Serva)

90 ml methanol/H_2O (1:1)

10 ml concentrated acetic acid (70%)

Destaining solution (5 l):

1.25 l absolute ethanol

0.4 l concentrated acetic acid

5.2 Plasmids and cell lines

5.2.1 Plasmids

To achieve high levels of protein expression, it is required to clone the gene of interest into plasmids which have high copy numbers in the host cell and are replicated under relaxed control. Such plasmids are, e.g. pMB1/ColE1-derived plasmids (15 to 60 copies per cell) or the pUC series of pMB1 derivatives (a few hundred copies per cell).[113] Each plasmid that serves as an expression vector features translational start and stop elements and a multiple cloning site with unique restriction sites to facilitate cloning. Following the induction of a promoter, the target protein is usually produced at a high level. The plasmids used in this study (Table 2) proved to be extremely effective in (a) transformation of auxotrophic strains, (b) excellent expression of heterologous target genes and (c) could be stably maintained during the fermentation in the minimal media. When a co-overexpression of the additional gene of MetRS was desired, a second vector like pQE (*ColE1* derivatives) was usually combined with compatible plasmids bearing a p15A replicon and maintained at about 10–12 copies per cell (pPRO-plasmids from Clontech). Under laboratory conditions, such multi-copy plasmids are randomly distributed during cell division and, in the absence of selective pressure, are lost at low frequency (10^{-5} – 10^{-6} per generation) primarily as a result of multimerization.[113]

Table 2: Basic features of the plasmid used for expression/co-expression experiments used plasmids

Plasmid	Source	Regulatory Element	Insert
pMetRS551 (pPROTetE.133 / pPROLarA.231)	BD Biosciences Clontech	P(*lac*/*ara*) hybrid promoter; *P15A* origin of replication; chloramphenicol resistance gene;	wt-MetRS gene
pQE80-L	Qiagen	T5 promoter; two *lac* operators; *ColE1* origin of replication; cis-repressed high-level expression; ampicillin resistance gene, *lacI*^{q}, N-terminal $(His)_6$-Tag	Annexin V (PP4) gene
pQE30	Qiagen	T5 promoter; two lac operators; *ColE1* origin of replication; ampicillin resistance gene. N-terminal $(His)_6$-Tag	*ds*Red gene
pQE60	Qiagen	T5 promoter; two lac operators; *ColE1* origin of replication; ampicillin resistance gene. C-terminal $(His)_6$-Tag	ECFP/EGFP/EYFP β-Gal gene
pEGFP/pECFP pEYFP	BD Biosciences Clontech	lac promoter; *ColE1* origin of replication; ampicillin resistance gene.	ECFP/EGFP/EYFP gene
pREP4	Qiagen	Constitutive expression of *lac I*^{q}; *P15A* origin of replication; low copy plasmid; kanamycin resistance gene.	*lacI*^{q} gene

5.2.2 Auxotrophic strains of *E. coli* for expression experiments

Among the many systems available for heterologous protein production, the gram-negative bacterium *E. coli* remains to be one of the most attractive due of its ability to grow rapidly and at high density on inexpensive substrates, its well-characterized genetics and the availability of an increasingly large number of cloning vectors and mutant host strains. In addition, the main prerequisite for the successful use of *E. coli* strains for incorporation experiments is the availability of an expression host with a stable auxotrophic genetic marker. In addition, such bacterial strains should also be suitable for transformations with target plasmids. Finally, optimal plasmid-directed protein translation when fermented in minimal medium should be achieved in terms of a well controllable and stringent expression system. The T5 expression system was found to completely suppress basal protein expression before induction ("gene-leakage"). Therefore, the availability of a proper bacterial host and expression system together with the experimentally optimised fermentation parameters was mandatory to achieve high-levels of incorporation. This was

previously demonstrated by replacements of Met,[92] Pro,[114] Trp,[66] Phe[55] and Tyr residues[54] using selective pressure incorporation method.

Table 3. Basic features of *E. coli* auxotrophic strains used for expression and incorporation experiments. Note that *E. coli* DH5α was used for cloning experiments.

Escherichia coli strain	Genotype	Source
DH5α	*F-Φ80 lacZΔM15 Δ(lacZYA-argF) U169 deoR recA1 end A1 hsdR17*	Invitrogen
CAG18491	*λ⁻, rph-1, met⁻, E3079::Tn10* (Met⁻)	Genetic Stock Center at Yale University, New Haven
CAG18492	*WP2 uvrA (Trp⁻)*	ATTC catalogue number 49980
DG30	*(Δ(gpt-proA)62, lacY1, tsx-33, supE44, galK2, λ⁻, aspC13, rac⁻, sbcB15, hisG4, rfbD1, recB21, recC22, rpsL31, kdgK51, xyl-5, mtl-1, ilvE12, argE3, thi-1, tyrB507, hsdS14, (hppT29)) (Phe⁻)*	Genetic Stock Center at Yale University, New Haven
AT2471	*λ⁻*, tyrA4, relA1, spoT1, thi-*(Tyr⁻)*	Genetic Stock Center at Yale University, New Haven

5.2.3 Tumour cell lines

Biological effects of the presence of fluorinated amino acid in growth media as well as their intracellular delivery were studied in suspension and adherent cell lines. These cells were obtained from the American Type Culture Collection (ATCC). The delivery of the substituted fluoro-proteins into cell lines was assisted by different lipid (lipofectamine, BioPorter) or peptide (Chariot) compositions. The following adherent cell lines were used: MCF10a (human breast epithelial tumour cells), SKBR3 (human breast cancer cells), BT 474 (human mammary tumour cells) and HEK293 (human embryonic kidney cells). K562 (hematopoietic malignant cells) was the only suspension tumor cell line used for delivery experiments.

5.3 Methods

5.3.1 Fermentation and expression of the target proteins

Cell culture fermentation and target protein expression were routinely performed in either Luria-Bertani (LB) medium for wild type proteins, or NMM (New Minimal Medium), for expression of proteins for the incorporation of non-canonical amino acids. In incorporation experiments, toxicity of the non-canonical amino acid may present an additional problem. In fact, all amino acids beside the 20 canonical ones do not normally

support cellular growth. The amino acid analogues that are especially close to the native ones or that display reactive functionalities, exhibit the highest toxicity[80] with selenomethionine being an exception from this rule. In this context, high substitution levels or even quantitative substitutions of target proteins by toxic fluorinated analogues can be achieved only by circumventing this problem of toxicity. To combat this, strong auxotrophism of the host cells and controlled amino acid supply in the fermentation media are necessary.[54] At the same time the target gene activity must be kept silent while growing the host cells to an appropriate density. After sufficient "healthy" cells are produced, the synthesis of the target protein can be induced. From this point, the host cells serve only as a "factory" to produce the desired substituted protein and further growth of host cells is of second importance.

5.3.2 Vector multiplication

Multiplication of a vector, a ligation mix, or PCR point mutagenesis products were routinely performed by transformation in a *recA*$^-$ cloning strain, *E. coli* DH5α.

5.3.3 Preparation of electrocompetent cells

Frozen cell stocks at –80°C were regenerated by growing them overnight in LB without antibiotics. Five hundred ml of fresh LB were inoculated with 5 ml of this overnight culture and grown at 37°C with vigorous shaking up to an OD_{600} of 0.5 – 0.7. Cells were harvested by centrifugation (GSA rotor at 5000 rpm at 4°C for 10 min). The pellet was resuspended in 200 ml of 10% ice cold glycerol. The centrifugation step was repeated twice, and the cell pellet was resuspended on ice in whatever little solution was left over in the bottle. Cells were aliquoted in 100 µl samples, shock frozen in liquid nitrogen and stored at -80°C. Freshly thawed aliquots were used for transformation, and the rest discarded after use, ensuring cells not being subjected to repeated freeze-thaw cycles.

5.3.4 Transformation of *E. coli*

Electric fields can induce pore formation and other structural defects in the lipid and cell membranes and make them permeable for DNA fragments and plasmids. Therefore, electroporation is widely used in the laboratory for gene transfection.

Sterile 0.1 cm BioRad *Gene Pulser*™ gap electroporation cuvettes were precooled on ice before adding 40 µl of the desired competent cell strain. Up to 1 µg of plasmid DNA (prepared from Qiagen DNA purification kits), were mixed well with the competent cells. Electroporation was performed in the BioRad Gene Pulser. Plasmids were routinely

transformed into the *E. coli* auxotrophic strains at 1.5 kV, 25 µF, 200 ohms with time constants between 3.5 and 5.5 ms. Transformed cells were washed with 1 ml LB and transferred into a 1.5 ml Eppendorf tube to be incubated at 37°C for 1 h. After incubation, 20 µl and 200 µl of the transformed cells were plated on LB-agar plates (with appropriate antibiotics) and incubated at 37°C overnight.

5.3.5 Gene sequences and expression vectors

5.3.5.1 Expression vectors for Annexin V

The gene for Annexin V (PP4) was cloned from the pRSET-PP4[92] vector as a *NcoI-HindIII* fragment into the pQE80 expression vector (see Figure 16). The advantage of this system is the presence of *lacI*^{q} gene in the construct; this enables an elevated amount of intracellular repressor that efficiently suppresses background expression. Attempts to construct Annexin V with C-terminal $(His)_6$-Tag were also successful. However, the protein of this gene construct does not fold under standard expression and purification conditions. The most probable reason for this is the essential requirement for the C-terminal end to be free to carry out the physiological activity and structural integrity of Annexin V. [93]

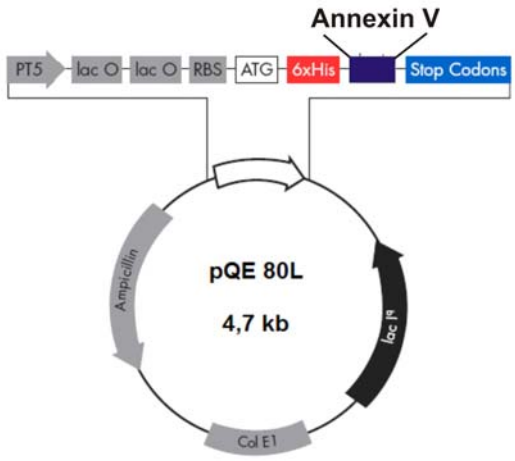


Figure 16: The pQE80-Annexin V expression vector (Table 2 provides description of all necessary regulatory elements).

5.3.5.2 Plasmids for ECFP, EGFP and EYFP expression

Plasmids containing the DNA sequence for EGFP, EYFP and ECFP, respectively, were constructed as follows: *NcoI-HindIII* fragment from commercially available pEGFP/pECFP plasmids (BD Biosciences/Clontech) was inserted into a pQE60 vector.

Resulting constructs have the $(His)_6$-Tag coding sequence at 3'-position respective to the polylinker cloning region (C-terminal end) (Figure 17). The primers were as follows:

5'GFP-His (NcoI): 5' CAT GCA TGC ATG CAT GCC ATG GTG AGC AAG GGC GAG GAG CTG TTC ACC 3'

3'GFP-His (HindIII)

5' CCG TAC CCC AAG CTT AGT GAT GGT GAT GGT GAT GCT TGT ACA GCT CGT CCA TGC CGA G 3'

Hind III 6 X His tag

Since pQE60-based gene constructs do not contain the gene for the *lac*-repressor (*lacI*^{q} gene), the host cells need to be co-transformed with a second plasmid pREP-4 (Figure 16). This plasmid provides elevated intracellular amounts of the *lac*-repressor.

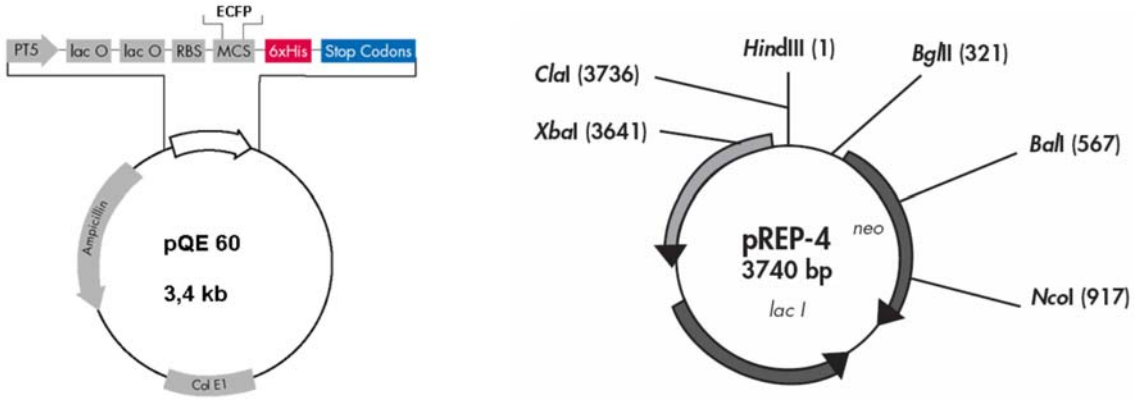


Figure 17. The pQE60-ECFP expression vector (left) and the auxiliary pREP4 plasmid (right) (Reproduced form Qiagen Catalogue). The basic features of these plasmids are listed in Table 2.

5.3.5.3 Red fluorescence protein form *Discosoma striata* (*ds*Red)

The plasmid pQE30 containing the *ds*Red (*BamHI* and *SalI* fragment) with a $(His)_6$-Tag at its N-terminus, was generously donated by K. Lukyanov (Puscino Institute, Russia). This was transformed into Tyr-auxotrophic cells (AT2471) in an attempt to replace the tyrosines with their fluorinated variants. Co-expression with pREP-4 was essential for suppression of baseline expression.

5.3.5.4 β-Galactosidase

The pQE60 vector gene fragment encoding for N-terminal $(His)_6$-Tag β-galactosidase was generously provided by Miss Tatjana Krywcun and Miss Petra Birle. pQE60-β-galactosidase was successfully transformed in electrocompetent Tyr- auxotrophic (AT2471) *E. coli* strain.

5.3.6 Determining plasmid DNA concentrations

DNA samples were measured between 200 and 300 nm on an UV/VIS spectrometer Lambda 19 with the software UV WinLab Version 2.0. The extent of purity of the DNA samples can be judged from the ratio of the absorbance intensity of the signal contributed by the protein aromatic residues at 280 nm, and that contributed by the nucleic acid at 260 nm. Plasmid DNA samples were diluted 50 fold before each measurement and concentration was calculated as follows:

DNA conc (μg/μl) = $\{OD_{260}$ x 50 x (factor of dilution)$\}/1000$.

A basic rule for determining the DNA concentration is the empirical finding that a value of 1 at OD_{260} corresponds to about 50 μg/ml of DNA.

5.3.7 Mutagenesis experiments

Tyrosine203 →Phenylalanine (Y203F) in EYFP. The point mutation Y203F was performed on EYFP that was already cloned in pQE60 C-terminal $(His)_6$-Tag. This protein was used to test the role of the hydroxyl-group in the stacking interaction between Tyr203 and chromophore in EYFP. The mutated proteins were expressed in phenylalanine auxotrophic host cells (*E. coli* DG30); in addition, the incorporation of *meta*-, *para*- and *ortho*-fluorophenylalanine was attempted as well. Primers for classical point mutagenesis were designed with the desired mutation in the middle of the primer with 15 bases of correct sequence flanking the same. Additionally, there is an overhang of seven bases on the 3' end of each primer. Both the mutagenic primers must contain the desired mutation and should anneal to the same sequence on opposite strands of the plasmid. Primer sequences were as follows: Y203F-EYFP forward (5’ CTG AGC TTC CAG TTC GCC CTG A 3’) and Y203F-EYFP backward (5’ GGA CTG GAA GCT CAG GTA GTG G 3’). Routine mutagenesis strategy to introduce point mutations was performed with the *Pfu* polymerase with a template concentration of 0.07 μg/μl and 50 pmol of each of the primers. After the first cycle of 5 min (at 95°C), 25 cycles consisting of denaturing (95°C; 1 min) and renaturing (55°C, 1 min) steps were performed with 10 min at the annealing

temperature (68 °C). After completion of the cyling reactions, an elongation step of 10 min assured ample time for the polymerase to copy the entire plasmid. The sample was then subjected to Dpn1 digestion (for 1 hour before transforming in DH5α competent cells). Distinct, well spaced colonies were observed on the LB agar ampicillin plate after an overnight incubation at 37°C.

Tryptophan57→Phenylalanine (W57F) mutation in ECFP. Using the QuikChange® Site-Directed Mutagenesis Kit from Stratagene (see Figure 18), the point mutation W57F was obtained. The mutagenic primers were designed such that they contain the desired mutation and anneal (68°C, 7 min) to the same sequence on opposite strands of the plasmid. An initial cycle of 5 min was performed at 95°C. Additional 25 cycles consisting of denaturing (95°C; 30 sec) and renaturing (55°C, 1 min) steps were performed. The mutation (TGG – TTC) was in the middle of the primer with 15 bases of correct sequence on both sides. Primer sequences were as follows:

ECFP-W57F forward: 5’ CTG CCC GTG CCC TTC CCC ACC CTC GTG A 3’;

ECFP-W57F backward: 5’ TCA CGA GGG TGG GAA AGG GCA CGG GCA G 3’.

The QuikChange® site-directed mutagenesis was performed by using *PfuTurbo*® DNA polymerase that replicates both plasmid strands with high fidelity and without displacing the mutant oligonucleotide primers. Incorporation of the oligonucleotide primers during cycling generated the mutated plasmid-containing staggered nicks. Following temperature cycling, the product was treated with *Dpn* I that is used to digest the parental DNA template and to select for solely the mutation-containing synthesized DNA. All components for the reaction were provided by the kit, and were mixed as per protocol.

This mutation was performed in order to test the role of Trp57 in ECFP. It is well known from the literature that Trp57 in EGFP is essential. Its replacement with Phe, Leu or Tyr destroys the structural integrity of the protein.[86] The expression experiments with the mutant W57F-ECFP confirmed that the same holds true for ECFP. In this particular experiment, it was not possible to obtain soluble protein under standard expression and purification conditions.

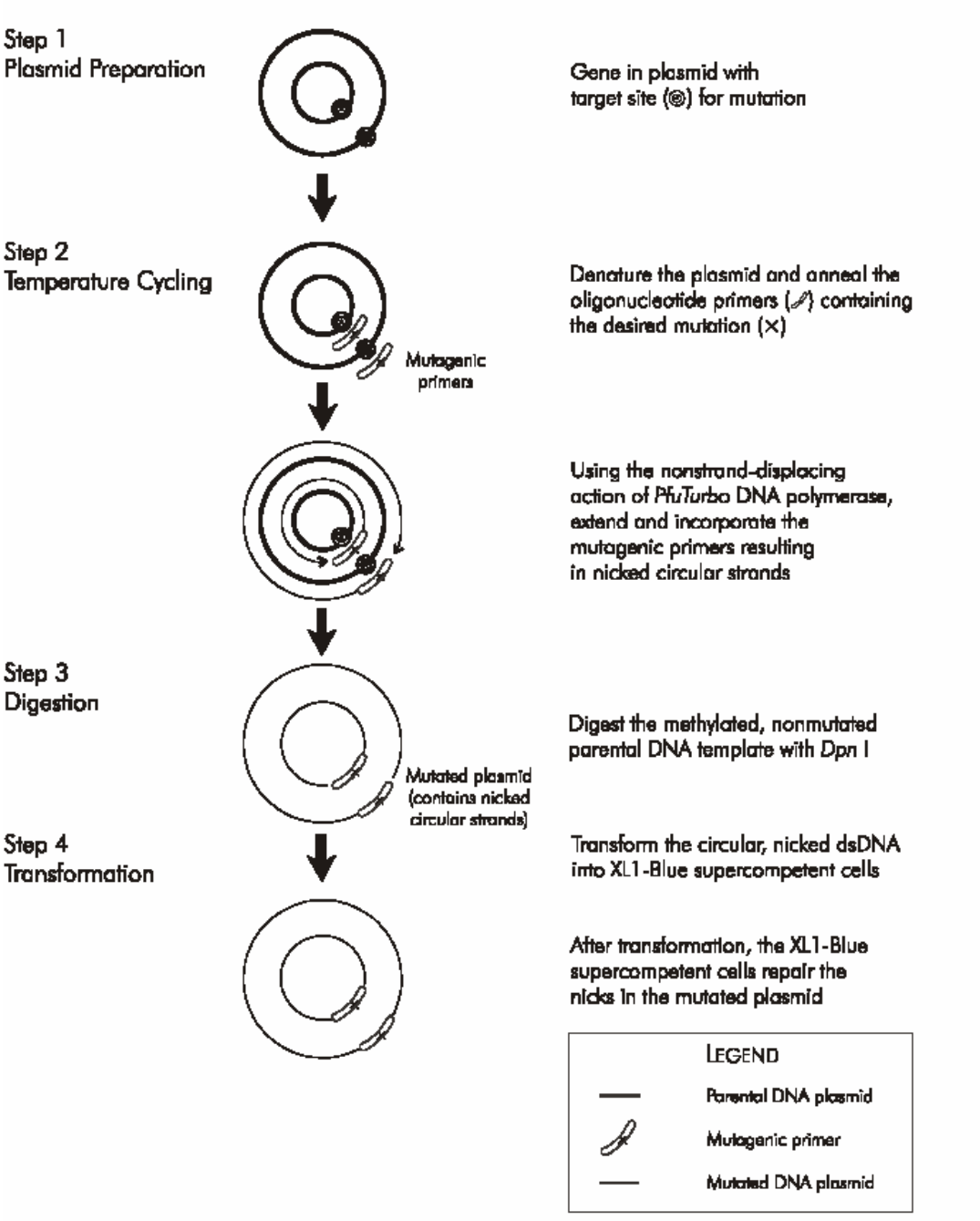


Figure 18. The basic principles of the QuikChange® Site-Directed Mutagenesis approach. Reproduced from Stratagene catalogue.

5.3.8 Expression tests

Before starting with fermentation experiments in larger volumes, an expression test was performed to choose the best colonies for further protein expression. Up to ten colonies were picked at random from the LB- agar plates and were transferred into 5 ml of LB medium with the appropriate antibiotics, and cells were grown at 37°C till an OD_{600} of 0.6 - 0.8. At this point, 1 OD_{600} of cells (i.e. ~ 10^9 cells) were transferred into a 1.5 ml tube, induced with 1 µl of IPTG and allowed to shake overnight (or at least 4-5 hours) at 30°C. After that, the suspension of bacteria was centrifuged for 5 min at 6000 rpm at room temperature and the supernatant was discarded. The cell pellet was resuspended in 50 µl of loading buffer and 50 µl of H_2O and incubated at 95°C for 1 min. Twenty microlitres of the sample was loaded on the 12% SDS gel and was run at 130 V and 30 mA. Staining of the gel was performed by incubation in Coomassie-Brilliant-Blue for 30 min at room temperature. Destaining the gel revealed protein bands and the best expressing cells were chosen for expression in larger cultures.

5.3.9 Protein expression in LB media

Isolated colonies from freshly transformed cells were picked from LB plates containing the appropriate antibiotics. They were grown in 5 ml of LB to an OD_{600} of 0.6 – 0.8 at 37˚C. Three millilitres were set aside from each expression tube, and IPTG was added to a final concentration of 1 mM. Expression was carried out at 30˚C overnight. Both induced and no-induced cultures were centrifuged at 14,000 rpm at room temperature for 10 min, supernatant discarded and pellets either checked for colour in case of GFPs, or processed for SDS expression gel. Molecular weight marker along with non induced cells provided a negative control in contrast to induced cells showing expression.

5.3.10 Protein expression in minimal media

The transformed cells were grown in NMM containing limiting quantities of that amino acid whose non-natural analogue would be added later. Cells were shaken at 200-250 rpm at 37˚C before limitation, and at 30˚C after addition of IPTG. In this way, the cells are made to grow up to a density that could be represented as the mid logarithmic growth phase. This is the optimal state of the dividing cells that are best suited for both expression as well as the replacement/incorporation experiments. This means that by the time the cells deplete the media of the natural amino acid; they have already reached an OD_{600} of 0.7-1.0, and are thenceforth unable to multiply further. At this point, the non-canonical substrate,

along with IPTG was added simultaneously. IPTG switches on the target gene induction for optimal expression of labelled proteins.

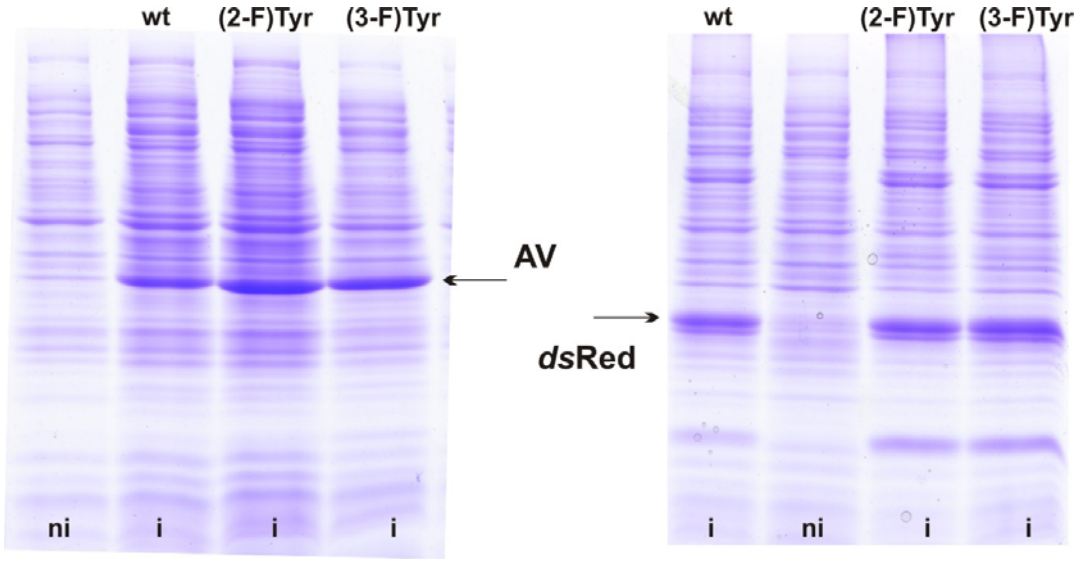


Figure 19. Fluorinations in minimal media. The lanes on the right are the lysates with the expression of wt-Annexin V and its fluorinated variants, whereas on the left side are expression profiles of native and fluorinated *ds*Red proteins. (i- induced cell lysates; ni – non-induced cell lysates; P – pellet, S – supernatant, AV- Annexin V).

5.3.11 Fermentation protocol modification – media shift for TFM

To elevate the pressure on the Met-auxotrophic strains CAG18491 and B834 for higher levels of incorporation of TFM into 2M-EGFP and Annexin V, the fermentation protocol was modified as follows. Cells were grown in LB up to a density of 0.7 at OD_{600} and centrifuged at 6000 rpm for 20 minutes at 4˚C, in sterile containers. The pellet was resuspended in NMM containing Difco Nitrogen base without Met and cells were pre-incubated for 4 hours at 30˚C (in order to exhaust even trace amounts of Met from the culture). After stabilisation of growth, the cells were induced by simultaneous addition of 1 mM TFM and IPTG and expression was performed overnight at 30˚C. In both proteins, 2M-EGFP and Annexin V the level of labelling (as judged form mass analyses) is considerably enhanced. It should be noted however, that this procedure does not increase fermentation yields of labelled proteins significantly, when compared with standard labelling protocols.

5.3.12 Co-expression experiments of Annexin V with MetRS511

The incorporation of non-canonical analogues is dependent on (1) the ability of the host cells to accumulate the analogue in high concentrations and (2) the ability of the analogue to be acylated to cognate tRNA by specific aminoacyl-tRNA synthetase. The first requirement is usually fulfilled in the way that amino acid analogues normally share the same membrane transporters and uptake mechanisms with their canonical counterparts.[80] The second necessity was solved by co-expression of an active fragment (1-551 amino acids) of *E. coli* methionyl-tRNA synthetase (MetRS551) via a low-copy plasmid, pWK1. The increasing concentration of MetRS raises the previously poor acylation level of the TFM analogue to $tRNA^{Met}$ and hence makes it possible to translate the model proteins with this amino acid.

The plasmid pMetRS511 which contains MetRS511 was generously provided by Ms. Petra Birle and Ms. Tatjana Krywcun. This plasmid was constructed by the ligation of the *AveII*/*SprI* fragment (module) of pPROTetE (containing transcriptional and translational control elements, multiple cloning site, and chloramphenicol resistance) to the corresponding fragment of pPROLarA containing the p15A origin of replication and other regulatory elements as listed in Table 2. In this acceptor plasmid MetRS551 (*BamH1* fragment) was cloned within the polylinker region and the resulting expression construct was named pMetRS551. The fermentation and expression experiments in minimal media were performed as described earlier.

5.3.13 Solubility test

Protein expression in 5 ml Falcon tubes was performed as described above. One OD_{600} of cells from the cell culture was taken. The cells were pelleted for 5 min at 6000 rpm at room temperature in a microfuge. To each cell pellet, 100 ml of the solubility buffer (50 mM Tris/HCl pH 7.5, 50 mM NaCl, 5 mM EDTA), were added, cells resuspended and sonified by using a microtip. Lysates were centrifuged for 5 min at 6000 rpm at room temperature in a microfuge. The supernatant (which contains soluble protein) was separated from the pellet (which contains insoluble protein) by transferring the supernatant to a clean tube. The pellet was resuspended in a renaturing buffer (50 mM Tris/HCl pH 7.5, 6 M urea, 50 mM NaCl, 5 mM EDTA), mixed with SDS-gel loading buffer, and 10 µl of each sample was loaded on a SDS-PAGE gel. The supernatant was treated in a similar manner.

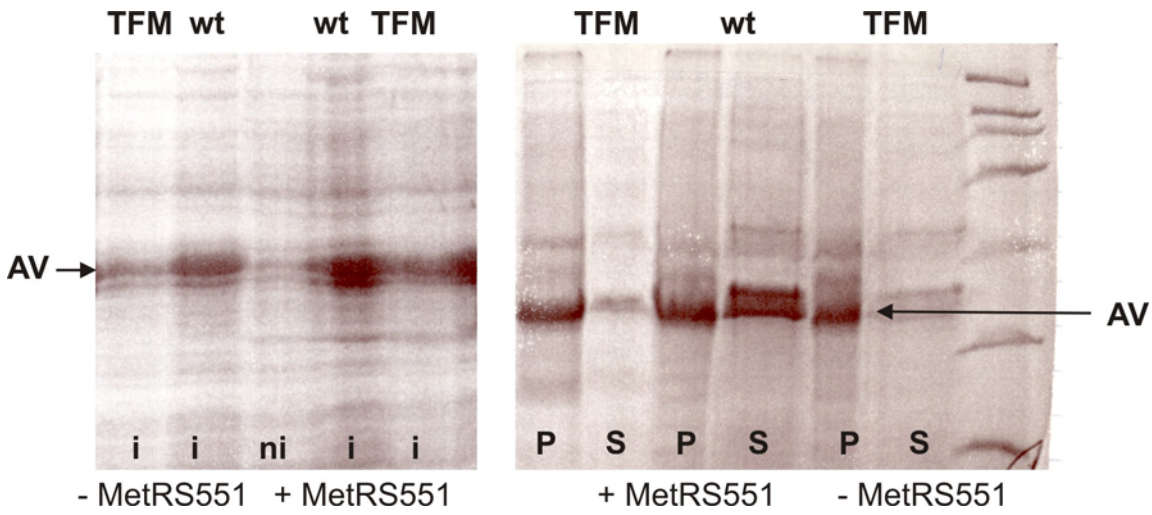


Figure 20. The pQE80-Annexin V/pMetRS551 enhanced expression system. Left: expression test in NMM with media shift in the presence and absence of elevated amounts of MetRS active fragment (MetRS551). Note that the expression level of TFM-Annexin V is considerably increased in the presence of MetRS551. Right: solubility test. Note that the supernatant of the culture with enhanced system contains markedly increased amount of soluble TFM-Annexin V in comparison with those form 'normal' culture. (i- induced cell lysates; ni – non-induced cell lysates; P – pellet, S – supernatant, AV- Annexin V). These gels were kindly provided by Petra Birle and Tatjana Krywcun.

5.3.14 Protein Purification

5.3.14.1 Purification of Annexin V

For purification of the His-tagged Annexin V, the expressing cells were centrifuged at 4°C and 6000 rpm for 15 min. The pellet was then resuspended in Spheroblast buffer (100 ml for 1 l culture) with lysozyme (20 mg per 100 ml buffer) and was kept on ice for 30 min. After another round of centrifugation (6000 rpm for 15 min at 4°C), the pellet was resuspended in 60 ml UZ-buffer and centrifuged at 20 000 rpm for 40 min at 4°C. The supernatant was loaded on a Ni-NTA-column and washed with 40 column volumes (CV) of washing buffer. Then a pre-elution wash was made with 5 CV of 20 mM imidazole in the wash buffer. The protein was eluted with the same buffer containing 200 mM imidazole, and protein fractions were collected and pooled. Annexin V was also purified through the Ni-NTA column using ÄKTA. The washing buffer was 75 mM Tris/HCl, pH 8.0 and 500 mM NaCl, whereas elution buffer contained a gradient between 20 mM – 300 mM imidazole.

Most of the Annexin V purifications were performed at the ÄKTA system from Amersham Pharmacia Biotech. The ÄKTA is an automated chromatographic system utilized in the development of purification protocol and in the large scale purification of proteins.

5.3.14.2 Purification of β-galactosidase

The procedure of breaking the cells and loading the supernatant on the Ni-NTA column was carried out in the same way as for EGFP mutants and variants. Purification was performed on the ÄKTA. The only difference was the composition of the buffer, which is the following. The washing buffer was 75 mM Hepes, pH 7.0 and 500 mM NaCl, whereas elution buffer contained a gradient between 20 mM – 300 mM imidazole in the washing buffer.

5.3.14.3 Purification of *av*GFPs, *ds*Red, and their mutants and variants

The cell suspension was centrifuged for 15 min at 6000 rpm and 4°C, and the pellet was resuspended in 100 mM sodium sulphate buffer, pH 8.0. The harvested cell pellets were resuspended in a minimal volume of washing buffer and sonicated in a Rosetta-flask (10x1 min, with intervals of 2 min between each sonication cycle). Sonication was performed in the Sonifier 250 (Branson; Danbury, CT, USA). The solution is then centrifuged at 30,000 rpm (Beckman Coulter LE-80 K Preparative Ultracentrifuge) for 30 min, 4°C to retrieve the protein in the supernatant. After sonification and centrifugation, the supernatant containing the soluble protein was collected. The protein was loaded on a Ni-NTA-column (Qiagen) and the protein purification was performed similar to the Annexin V purification protocol starting with the NTA wash. In order to obtain a highly concentrated protein solution, a Phenyl-sepharose (Pharmacia) column was used. Ammonium sulphate, 20 % (v/v), was added to the collected protein fractions while the Phenyl-sepharose column was washed with ammonium sulphate buffer. After loading the protein, the column was washed with 20 % ammonium sulphate /Tris/ EDTA and the protein was eluted with a gradient 20 % - 0 % ammonium sulphate. All purified proteins were concentrated with Amicon CentriPrep and Centricon 10 concentrators (MWCO 10 kDa). The purity of the proteins was analysed by SDS-PAGE, HPLC and ESI-mass spectrometry.

5.3.15 Determining protein extinction coefficients and concentrations

Protein concentrations were measured by using molar extinction coefficients previously determined at room temperature (20 °C) for native and fluorinated proteins in the Tyr+Trp UV-absorbance region. The UV-absorbance spectra of the proteins in phosphate buffered saline (PBS: 115 mM NaCl, 8 mM KH_2PO_4, 16 mM Na_2HPO_4, pH 7.3) were recorded on a Perkin-Elmer Lambda 19 UV/VIS spectrophotometer. By using the Beer-Lambert's law, ($\boldsymbol{A = \varepsilon \times c \times d}$) the concentration of the protein solution in the aromatic region

(i.e. 275-280 nm) could be calculated as follows. Protein samples were measured with a slit width of 1.00 nm, a data interval of 0.1 nm and a scan speed of 240 or 480 nm/min. Major contribution to the extinction coefficient (ε_M) of the proteins derives from chromophoric residues like Trp and Tyr at 280 nm and to a lesser extent from the Cys (or disulphide). Since the extinction coefficient of each chromophore in the protein is considered to be additive, the overall molar extinction coefficient of the protein depends on the types and number of these chromophoric residues (n) in the sequence. Therefore, ε_M for proteins at 280 nm (ε_{280}) is a sum of all these residues:

$$\varepsilon_{280} = n_{Trp} \text{ x} \cdot \varepsilon_{Trp} + n_{Tyr} \text{ x } \varepsilon_{Tyr} + n_{Cys} \text{ x } \varepsilon_{Cys}$$

Molar extinction coefficients of chromophoric residues at 280 nm at neutral pH are λ=280 nm (ε_{Trp} = 5690 $M^{-1}cm^{-1}$, ε_{Tyr} =1200 $M^{-1}cm^{-1}$, ε_{Cys} = 120 $M^{-1}cm^{-1}$). Thus, the protein concentration can be calculated as follows.

$$C(\text{mg/ml}) = (A_{280} \text{ x } Mw)/\varepsilon_{280} \text{ x } d$$

where “d” denotes the optical path of a cuvette, (usually 1 cm), “Mw” the molecular weight and A_{280} is the measured absorbance at 280 nm. The quartz cuvettes routinely used for measurements were purchased from Hellma GmbH (Baden, Germany).

The molecular weights and determined molar extinction coefficients of the proteins used in this study are listed in Table 4.

5.4 Analytical quantification of incorporation levels

5.4.1 Amino acid analyses

In the mass spectrometric analyses, ionization of the trifluorinated protein species proved to be difficult. On the other hand, the intensity of these peaks cannot be taken as a quantitative measure of their TFL content. For that reason, the incorporation levels were quantitatively determined only by amino acid analyses. Wild type and mutant-protein sample (10 nM) for amino acid analyses were hydrolysed for 48 h in 6 M HCl containing 2.5% (v/v) thioglycolic acid. The hydrolysates were analysed with a Biotronic LC 6001 amino acid analyser. In comparison to Met (t_R ~ 26.0 min) the t_R value of TFM was ~ 27. 0 min.

Table 4. Molecular masses and molar extinction coefficients (in Tyr+Trp region) of wt-proteins, mutants and fluorinated variants used in this study. Note that fluorination of aromatic residues almost has no effect on extinction coefficients since absorbance properties of fluorinated and parent aromatic residues are almost identical.

Protein	Molecular weight (Da)*	ε_{280} ($M^{-1}cm^{-1}$)
wt-EGFP	27744	21500
(2-F)Tyr/(3-F)Tyr-EGFP	27938	21500
(4-, 5-, 6, 7-F)Trp-EGFP	27762	21500
TFM-2M-EGFP	27800	21500
wt-EYFP	27791	22600
(2-F)Tyr-,/(3-F)Tyr-EYFP	28007	22600
Y203F-EYFP	27776	22600
(2-, 3-, 4-F)Phe-Y203F-EYFP	28028	22600
wt- ECFP	27707	25800
(2-F)Tyr-,/(3-F)Tyr-ECFP	27887	25800
(4-, 5-, 6, 7-F)Trp-ECFP	27743	25800
*ds*Red	27682	21400
(2-F)Tyr-,/(3-F)Tyr-*ds*Red	27942	21400
wt-Annexin V	37684	22800
(2-F)Tyr-,/(3-F)Tyr-Annexin V	37882	22800
wt-β-galactosidase	117714	262550
(2-F)Tyr-,/(3-F)Tyr-β galactosidase	117714	262550

* for fluorinated variants, expected masses for fully substituted proteins are given.

5.4.2 Mass spectrometry

Sample analyses. The replacement of native Trp-residues by its non-canonical analogues was routinely checked by electrospray mass spectrometric analyses (ESI-MS) using an ESI PE SCIEX API 165 instrumental set up (Perkin-Elmer). In addition, these analyses were performed with an ESI PE SCIEX API 165 (Perkin-Elmer) single quadrupole mass-spectrometry system. The spectra for detecting the products were collected with an auto-sampler Series 200 (Perkin-Elmer) at a rate of 10 µl/min, an ion source high voltage of 4900 kV, an orifice voltage of 10 V, a dwell time of 0.4 ms per scan and a step size of 0.2 Da with the scan range of 20 to 50. Injection was regulated with the splitter. The injection volume of samples (dissolved in acetonitrile) was 10 µl. The eluent was 100% acetonitrile (0.05% TFA).

Since β-galactosidase is a relatively large protein, its mass was determined by the more sensitive microTOF-LC (Brucker Daltons) in a positive mode, in a mass range of 50 – 3000 m/z and with a mass accuracy of 5 ppm RMS error. As ion source electrospray, of 1 µl/min – 1 ml/min with CE/MS interface was used; mass resolution was 15.0 (FWHM). About 30,000 spectra were summed for one data point (TOF repetition rate 10-20 kHz) and were calibrated externally by using a sodium formate cluster. The injection volume of the sample was 5 µl (250 ml/min) where the sample was desalted by Aligent 1100 HPLC system (Zorbax C8) with autosampler.

Trp-fluorinations in EGFP and ECFP. The incorporation levels were analysed by mass-spectrometry, since the molecular mass difference between Trp and its analogues is sufficiently large to be determined experimentally. A single Trp57 of native EGFP (expected mass: 27744.5 Da; experimental value: 27743 (±3.0) Da) upon fluorination leads to an increase in the protein's molecular mass of 18 Da (expected mass for all fluorinated EGFPs: 27759.5 Da; experimental value: 27761 (±2.1) Da). In addition, the presence of (4-F)Trp at position 57 of EGFP was unambiguously confirmed by X-ray crystallographic analysis of this protein variant (See Discussion). The ECFP contains two Trp residues (Trp57 and Trp66), with the mass differences upon substitution being much more pronounced. The molecular masses of the expected and calculated weights were in agreement; native ECFP: (found: 27707 (±2.9) Da; calculated: 27707.5 Da), and fluorinated ECFPs were also fully substituted (found: 27743 (±3.5) Da; calculated: 27743.5 Da).

Tyr-fluorinations in EGFP, EYFP, Annexin V and β-galactosidase. The high level replacement of the Tyr residues including Cro66 in all *av*GFPs by (3-F)Tyr and (2-F)Tyr was confirmed by ESI-MS analyses using the same setup as described earlier (Section 4.2.4.1). Deconvoluted mass spectra for EGFP: wt-protein, 27743 (±5.0) Da, (expected mass: 27740.4 Da); (2-F)Tyr-EGFP, 27940 (±4.5) Da, and (3-F)Tyr-EGFP, 27938 (±2.5) Da (expected mass in both cases: 27941.0 Da). All Tyr residues in EYFP were replaced to high level as well: wt-EYFP 27791 (±3.0) Da; (expected mass: 27790.6 Da), (2-F)Tyr-EYFP 28008.0 (±6.5) and (3-F)Tyr-EYFP 27005.5 (±3.5) (expected mass in both cases: 28006.6 Da). Ten Tyr residues in ECFP were also fully replaced with both (2-F)Tyr and (3-F)Tyr. Native ECFP: calcd. 27707.0, found 27710.0 (±4.0) Da); (2-F)Tyr-ECFP and (3-F)Tyr-ECFP: calcd. 27885.15 Da, found 27887.6 (±3.7) Da and 27886.2 (±5.3) Da, respectively.

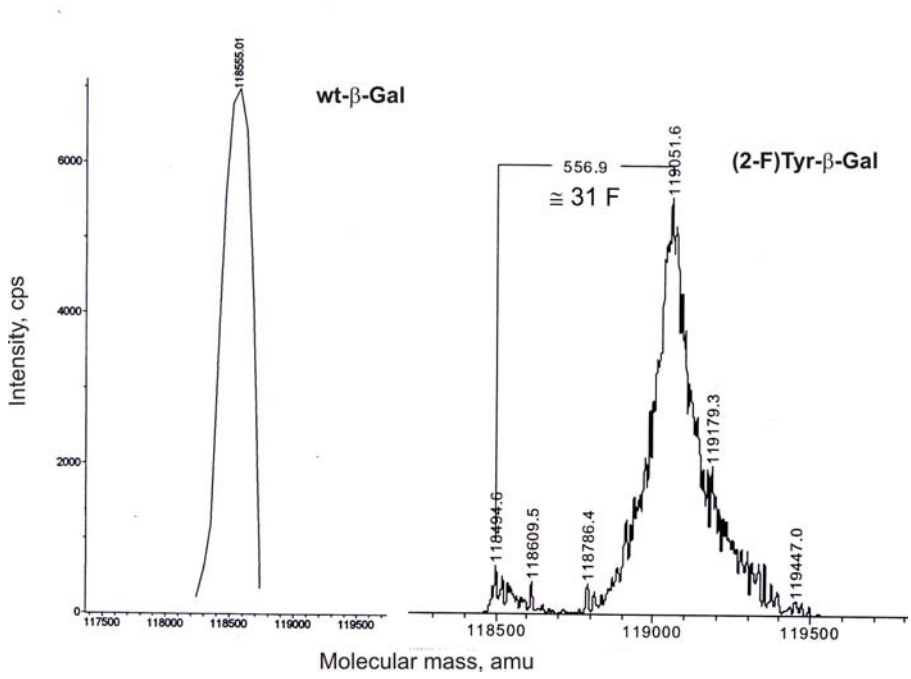


Figure 21. Deconvoluted masses for wt-β-galactosidase (expected mass: 117714.2 Da; found mass 118555) and (2-F)Tyr-β-galactosidase (expected mass: 118258 Da; found mass: 119051.6). Theoretical masses were calculated from the amino acid sequence data. On the other hand, experimental masses show that fluorination was complete as 31 H→F replacements can be detected in the sample. The same holds true for (2-F)Tyr-β-galactosidase. Additional mass peaks are probably sodium adducts of the protein.

Using routine bioexpression protocols, a high-level of incorporation of both fluorinated tyrosines (i.e. *ortho*-, and *meta*-F-Tyr) in N-terminal His-tagged Annexin V (11 Tyr residues) could easily be achieved (wt-protein: 37684 (±5.5) (expected mass: 37620 Da); fluorinated protein: 37882 (±4.5) (expected mass in both cases: 27941.0 Da). The novelty of this experiment was the possibility to incorporate (2-F)Tyr into Annexin V, which was not reported previously. On the other hand, a full labelling of β-galactosidase with both (2-F)Tyr and (3-F)Tyr could be monitored only by using an instrument more sensitive than ESI PE SCIEX API 165. Indeed, as shown in Figure 21 the *microTOF* Brucker instrument provided a direct proof for the complete fluorination of this huge protein.

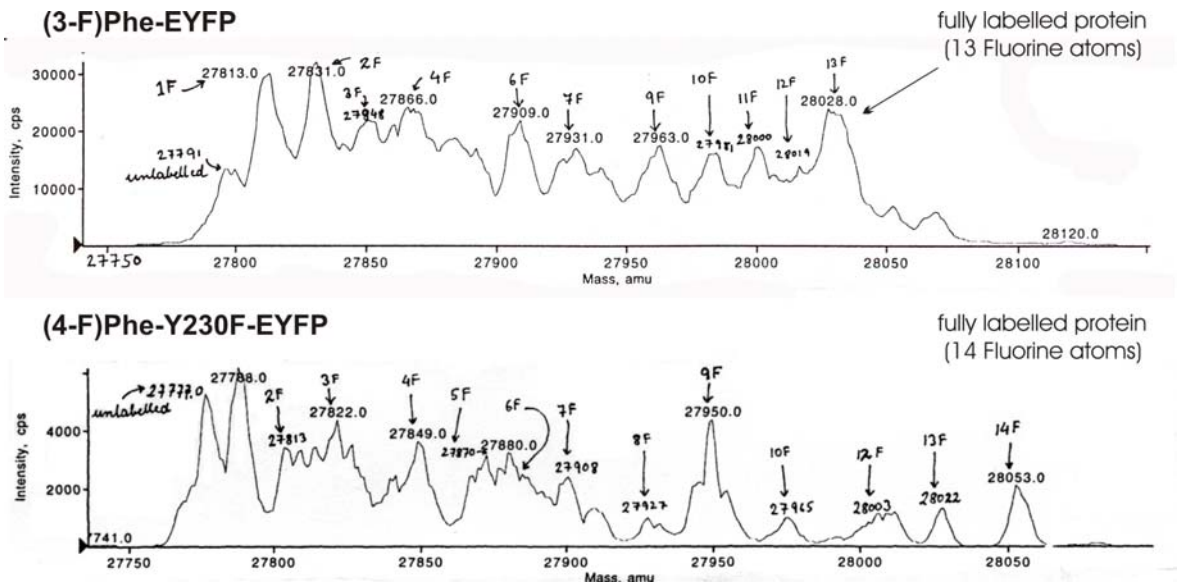


Figure 22. ESI-MS masses for Fluoro-Phe incorporation into EYFP (upper part) and Y203F-EYFP (lower part). Note a distribution of differently labeled species containing 1-13 or 1 -14 fluorine atoms in the range of wt- to fully substituted protein species.

Phe-fluorinations in Y203F-EYFP and Tyr-fluorinations in *ds*Red. Replacement of Tyr203 by Phe in wt-EYFP (containing 13 Phe residues) results in an increment in the number of Phe residues in the mutant. The replacement of Phe-residues of both wt and mutant proteins should reveal the difference of about 18 Da (1 fluorine atom) between these samples as shown in Figure 22. This figure shows in addition, that the replacement was never total, and only a small fraction of the protein sample was fully labelled. Deconvoluted mass spectra for EYFP: wt-protein 27791 (±5.0) Da; (expected mass: 27740.4 Da); EYFP with all fluorinated Phe-residues (fluorine in *ortho*-, *para*-, or *meta*-positions) is expected to have 28028 Da. These species could be found in the analysed protein samples (13 H→F replacements) as shown in Figure 23. Fully labelled species (14 H→F replacements) are also found in Y203F-EYFP samples of fluorinated proteins (fluorine in *ortho*-, *para*-, or *meta*-positions of Phe residues) as well (wt-protein: 27777 Da; fully fluorinated protein: 28053 Da (±4.5)).

A complete replacement of the Tyr-residues (13 residues) including the chromophoric one in *ds*Red by (3-F)Tyr and (2-F)Tyr was also never achieved; the resulting mass Profiles were similar to those shown in Figure 22. It was possible to determine mass species ranging from for wt-*ds*Red (27682 (±3.0) Da) to fully fluorinated protein fractions (determined mass for both fluorinated Tyr-residues was 27940 (±2.5) Da.

5.5 Spectroscopic methods

5.5.1 UV/VIS-spectroscopy.

UV-absorbance spectra of proteins and amino acids in buffer solutions were recorded with a Perkin-Elmer Lambda 19 UV/VIS spectrophotometer. Extinction coefficients (ε_M) for native and substituted proteins in the Tyr+Trp absorbance region (277 – 284 nm) were determined as described in Section 5.3.15, while chromophore ε_M were derived by a normalization to these values. The ratio of the chromophore to Tyr+Trp absorbance ($\varepsilon_{M(Cro66)}/\varepsilon_{M(Tyr+Trp)}$), known to be index of the purity of *av*GFPs is about 1.8 for native EGFP and 1.0 for native ECFP. These ratios may be changed to larger values especially in the cases where the chromophore absorbance is dramatically changed upon replacement with analogues and surrogates. These parameters determined for native EGFP and ECFP as well as their fluorinated variants are presented in Tables 5-8.

For UV-absorbance, protein stock concentrations (10 – 20 mg/ml) were diluted to 0.2 – 0.4 mg/ml in the appropriate buffer (spanning the entire pH range) and allowed to equilibrate for 1 hour at room temperature. The emission spectral profiles of protein samples (0.25 μM) were determined upon excitation (495 – 550 nm; slit 2.5 nm) using parameters and experimental conditions described in Tables 5-8. Instrumental setup for both UV and fluorescence spectroscopy are described in the previous sections.

Molar extinction coefficients (ε_M) in neutral buffers (20 °C) for the Tyr+Trp absorbance region were determined as described in section 4.2.1.5 while chromophore extinction coefficients were derived by normalisation to these values as presented in Table 4. The ratio of the chromophore and Tyr+Trp absorbance ($\varepsilon_{M(Cro66)}/\varepsilon_{M(Tyr+Trp)}$), usually referred in literature [115] as an index of purity of *av*GFPs is about 1.7 for native EGFP and 2.3 for native EYFP. However, it should be noted that these ratios differ among the native proteins and their variants due to the fluorination of the chromophore, and can be considered to be indicators of protein stability. For example protein variants that are more prone to aggregations, such as (2-F)Tyr-EGFP are characterised by lower ($\varepsilon_{M(Cro66)}/\varepsilon_{M(Tyr+Trp)}$) values. Fluorescence spectra were recorded in PBS on a Perkin-Elmer spectrometer (LS50B) equipped with digital software. In fluorescence titration experiments (i.e. determination of pK_a values) samples were excited at different wavelengths (having slit-width 2.5 nm) and the intensities of the emission spectra were recorded at the corresponding emission maximum values as described in Tables 5-8, using the same experimental and instrumental setup as described in a previous section.

5.5.2 Steady state fluorescence

Fluorescence emission and excitation spectra were recorded on a Perkin-Elmer spectrometer (LS50B) equipped with digital software. The samples were present in 50 mM Na-phosphate buffer pH 8.0 and 100 mM NaCl. Fluorescence was measured with proteins (0.5 μM for EGFP and its variants; 1.0 μM for ECFP and its variants) that were excited at 277 nm (Tyr+Trp spectral region), 295 nm (Trp spectral region) or at the chromophore spectral region (slits 5.0 and 2.5 nm, respectively) and the emission spectra were recorded in the range between 460 – 650 nm. The excitation spectra were measured in the range 400 - 550 nm. The determined fluorescence spectral parameters are presented in Tables 6 and 8.

5.5.3 Fluorescence pH titration of chromophore

Degassed and argon-saturated aqueous solutions contained 100 mM NaCl and 50 mM buffering substance: citric acid/Na_2HPO_4 (pH 3.0, 3.5); Na-acetate/acetic acid (pH 4.5, 5.0, 5.5); Na_2HPO_4/ NaH_2PO_4 (pH 6.0, 6.4, 6.8, 7.2, 7.6, 8.0); boric acid/NaOH (pH 9.0); Na-borate/NaOH (pH 8.5, 9.0, 9.5, 11.0). Fluorescence spectra were recorded on a Perkin-Elmer spectrometer (LS50B) equipped with digital software using the "Read-Intensity" option (integration time was 3 seconds). Protein samples (0.25 μM) were examined in the pH range from 1 to 11 in different buffers. Determined pK_a values and emission and excitation values are presented in Tables 6 and 8.

5.5.3.1 Refolding capacity measured by fluorescence recovery

To determine whether the fluorination of Tyr-residues influences the well documented refolding properties of *av*GFPs, fluorescence recovery of the unfolded protein was measured. The spectra were measured after incubating the proteins in a denaturation buffer (8 M urea and by heating at 95 °C for 5 min), followed by 100 x dilution in refolding buffer (10 μl of this solution was diluted into 990 μl of renaturing buffer), and monitoring its refolding over a period of 3 hours. Data were collected with an interval of 5 seconds. The fluorescence recovery was measured at protein emission maximum (EGFP: 505 nm; ECFP: 504 nm and EYFP: 527 nm) by using the option 'Timedrive' of Perkin-Elmer spectrometer (LS50B). The concentration of denatured proteins was typically around 50 μM in denaturation solution.

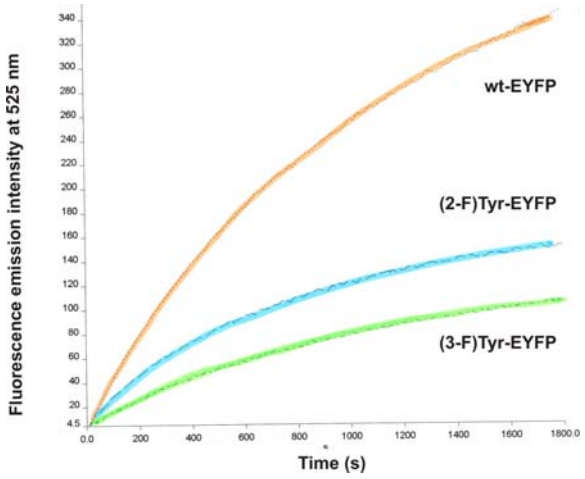


Figure 23. Fluorescence time course of renaturation of wt-EYFP and its fluorinated variants (2-F)Tyr-EYFP, (3-F)Tyr-EYFP.

A time course of fluorescence acquisition of mature fluorescent protein upon dilution from 8 M urea is shown in Figure 23. All tested variants show remarkable renaturing properties (an average of 20% recovery was achieved in both native and fluorinated proteins). These observations show that protein folding/refolding capacity is not considerably influenced by a global fluorination of Tyr-residues.

5.5.4 One-dimensional ^{19}F-NMR of TFM-EGFP

The NMR experiment was performed in the laboratory of Dr. T. Holak by Dr. M. Krajewski. The NMR data (one-dimensional ^{19}F-NMR recorded at room temperature) were collected at 470.4 MHz using a Bruker Avance spectrometer fitted with a 5 mm inverse broad-band probehead with the proton coil tuned to fluorine as described elsewhere.[87]

5.6 Tables with spectroscopic features of autofluorescent proteins

5.6.1 Spectral parameters derived from steady state spectroscopic measurements

Spectral features determined according to the methods described in previous sections are listed in the tables presented below. Accompanying figures are incorporated mainly in Results and Discussion parts.

Table 5. Spectral properties of native EGFP and ECFP and their related mutants brought about by Trp-replacements. Absorbance maxima (λ_{max})[a] and molar extinction coefficients (ε_M)[b] are determined in 50 mM Na-phosphate (pH 7.0) in the presence of 100 mM NaCl at room temperature. Experimental conditions are described in previous sections and related spectra are presented in Figure 26 (Results and Discussion)

	(Trp+Tyr) Absorbance		Chromophore (Cro66) Absorbance			
EGFP						
	ε_M	λ_{max}	ε_M (λ_{max})	λ_{max}		$\varepsilon_{M(Cro66)}/\varepsilon_{M(Tyr+Trp)}$
wt	19887 (± 754)	277	35570 (± 1316)	488	-	1.80
(4-F)Trp	17930 (± 699)	277	31485 (± 1225)	488	-	1.75
(5-F)Trp	22107 (± 996)	277	43630 (± 1670)	488	-	1.97
(6-F)Trp	20710 (± 807)	277	38125 (± 1580)	488	-	1.90
(7-F)Trp	23217 (± 609)	277	38125 (± 1580)	488	-	1.90

ECFP						
	ε_M	λ_{max}	$\varepsilon_{M(\lambda max1)}$	λ_{max1}	λ_{max2}	$\varepsilon_{M(Cro66)}/\varepsilon_{M(Tyr+Trp)}$
wt	25770 (± 1027)	278	24792 (± 1151)	434	452	1.0[c]
(4-F)Trp	21704 (± 686)	276	14810 (± 697)	426	444	0.68
(5-F)Trp	27835 (± 693)	279	19172 (± 730)	426	444	0.69
(6-F)Trp	23291 (± 401)	280	21762 (± 238)	432	450	0.93
(7-F)Trp	23704 (± 987)	276	16650 (± 837)	426	444	0.72

[a]expressed in nm; [b]expressed in $M^{-1}cm^{-1}$ [c]this ratio is between λ_{max1} and (Tyr+Trp) absorbance

Table 6. Fluorescence excitation (λ_{exc})[a] and emission (λ_{em})[a] values and their relative intensities (RF)[b] at 25 °C for native and mutant proteins. Under the same conditions pK_a-values of their chromophores were determined by fluorescence titration[c]. Experimental conditions are described in previous sections and related spectra are presented in Figure 26 (Results and Discussion).

	EGFP				ECFP				
Protein	λ_{exc}	λ_{em}	RF	pK_a	λ_{exc}	λ_{em1}	λ_{em2}	RF	pK_a
wt	488	510	(1.0)	5.85 (± 0.05)	452	476	504	(1.0)	5.28 (± 0.06)
(4-F)Trp	488	510	(1.0)	5.86 (± 0.06)	444	470	498	(0.4)	4.86 (± 0.10)
(5-F)Trp	488	510	(1.0)	5.77 (± 0.10)	444	466	494	(0.7)	5.27 (± 0.08)
(6-F)Trp	488	510	(1.0)	5.87 (± 0.08)	450	474	502	(0.5)	5.10 (± 0.05)
(7-F)Trp	488	510	(1.0)	5.81 (± 0.10)	444	468	496	(0.4)	4.97 (± 0.05)

[a]expressed in nm; [b]RF-values are normalised to 1.0 and represents the fluorescence intensity of the EGFP and ECFP proteins at standard experimental conditions (25°C; 50 mM Na-phosphate pH 7.0, 100 mM NaCl). [c]Each value is an average of at least two measurements.

Table 7. Absorbance parameters of native EGFP,EYFP, *ds*Red and their related mutants produced by Tyr-replacements. Absorbance maxima (λ_{max})[a] and molar extinction coefficients (ε_M)[b] are determined in 50 mM Na-Phosphate (pH 7.6) on the presence of 100 mM NaCl at room temperature. Experimental conditions are available in previous sections; some related spectra are presented in Figure 29.

	(Trp+Tyr) Absorbance		Chromophore (Cro66) Absorbance		
Protein	ε_M (λ_{max})	λ_{max}	ε_M (λ_{max})	λ_{max}	$\varepsilon_{M(Cro66)}/\varepsilon_{M(Tyr+Trp)}$
wt-EGFP	19800 (± 680)	277	34340 (± 1180)	488	1.75
(2-F)Tyr-EGFP	19750 (± 917)	271	22190 (± 1174)	482	1.13
(3-F)Tyr-EGFP	17930 (± 699)	274	30200 (± 1200)	485	1.56
wt-EYFP	20900 (± 475)	277	48770 (± 1250)	514	2.34
(2-F)Tyr-EYFP	20400 (± 730)	273	55460 (± 2180)	504	2.71
(3-F)Tyr-EYFP	21500 (± 590)	274	48790 (± 1440)	518	2.26
wt-*ds*Red	21400	277	75000[c]	514	3.50
(2-F)Tyr-*ds*Red	21400	273	75000[c]	504	3.50
(3-F)Tyr-*ds*Red	21400	274	75000[c]	518	3.50

[a]expressed in nm; [b]expressed in $M^{-1}cm^{-1}$; [c]data provied by K. Lukyanov.

Table 8. Fluorescence excitation (λ_{exc})[a] and emission (λ_{em})[a] values and their relative intensities (RF)[b] at 20 °C for native and mutant proteins. Under the same conditions pK_a-values of their chromophores were determined by fluorescence titration[c]. Experimental conditions are described in previous sections while related fluorescence spectra can be found in Figures 29 and 36.

Protein	λ_{exc}	λ_{em}	RF	pK_a
wt-EGFP	488	510	(1.0)	5.74 (± 0.15)
(2-F)Tyr-EGFP	482	504	(0.6)	5.54 (± 0.17)
(3-F)Tyr-EGFP	485	514	(0.9)	5.30 (± 0.12)
wt-EYFP	514	527	(1.0)	6.62 (± 0.08)
(2-F)Tyr-EYFP	504	520	(1.2)	6.52 (± 0.10)
(3-F)Tyr-EYFP	518	533	(1.0)	5.30 (± 0.05)
wt-*ds*Red	514	527	(1.0)	n.d
(2-F)Tyr-*ds*Red	504	520	(1.0)	n.d
(3-F)Tyr-*ds*Red	518	533	(1.0)	n.d

[a]expressed in nm; [b]RF-values are normalised to 1.0 and represents the fluorescence intensity of the wt EGFP and EYFP proteins at standard experimental conditions (25°C; 50 mM Na-Phosphate pH 7.6, 100 mM NaCl). [c]Each value is an average of at least two measurements. [d]Non-fluorescent proteins; n. d. – not determined.

5.6.2 Absorbance of denatured native and fluorinated *av*GFP at low and high pH values

To determine the extent to which the C-H → C-F exchange in the chromophore and in the rest of the Tyr residues influence the protonated/deprotonated (AH/A^-) state of the chromophore, the "enhanced cyan fluorescent protein" (ECFP, Ser65Thr/Tyr66Trp), was used as a control system for comparison. Since the ECFP contains 10 Tyr residues with the 11th one replaced with Trp at position 66 (chromophore), global fluorination should not affect the chromophore. Indeed, there were no observable differences in the absorbance maxima between the parent protein and either 2-, or 3-fluorotyrosyl-ECFP (Table 9). These results indicate that the observed blue shifts in the variants of EGFP and EYFP principally originate from the fluorine-containing chromophores.

Table 9. The pH-dependence of the absorbance of denatured EGFP, EYFP, ECFP and their fluorinated variants in 100 mM NaCl, 50 mM buffer solutions (KCl/HCl for pH 1.0 and Na-borate/NaOH for pH 11.0) (denaturant 6 M guanidinium chloride)*.

Protein	λ_{max}/nm (pH 1.0)	λ_{max}/nm (pH 11.0)
EGFP	380	447
(2-F)Tyr-EGFP	380	444
(3-F)Tyr-EGFP	380	443
EYFP	378	447
(2-F)Tyr-EYFP	378	444
(3-F)Tyr-EYFP	378	443
ECFP	420	418
(2-F)Tyr-ECFP	420	418
(3-F)Tyr-ECFP	420	418

*There is no change in the absorbance maxima upon addition of 5 mM DTT or use of other denaturant (8 M urea with or without DTT). Proteins with final concentration ~ 0.5 mg/ml were first incubated in denaturing solution for 1 hour and subsequently heated (5-10 min) at 95° C, cooled down and absorbance spectra were measured at room temperature. Each value is an average of at least two measurements.

5.7 Crystallisation experiments, crystal structures and modelling

Both 2- and the 3-fluorotyrosyl-EGFP have been crystallised in 0.05 M NaOAc, 0.1 M Tris/HCl buffer pH 8.5, and 14% (w/v) polyethyleneglycol 4000 within 2 days. For crystallisation, hanging drops were made of 1 µl of protein solution (34 µg/µl) and 1 µl of precipitant solution at 20°C. Drops were equilibrated against 0.2 ml of the precipitant solution. Structure determination of (2-F)Tyr –EGFP and (3-F)Tyr –EGFP was performed by Dr. Jae Hyun Bae whereas modelling studies were performed by Dr. Kamran M. Azim. Detailed accounts on data collection statistics, data processing procedures as well as modelling calculations are available in references.[56; 116] More detailed analyses of the crystallographic micro-environments of fluorinated tyrosines are provided in the "Results and Discussion".

5.8 Delivery of substituted proteins in mammalian cell lines

5.8.1 Protein delivery and processing of transformed cell lines

The delivery of proteins substituted with fluorinated amino acids was assisted by different lipid (lipofectamine, BioPorter) or peptide (Chariot) compositions according to procedures described in sections 5.1.7 and 5.1.8 following the manufacturer's instructions. Adherent cell lines were MCF10a, SKBR3, BT 474, HEK293 and only on suspension

tumor cell line K562 was tested. Protein delivery experiments in these cells were attempted by fluorinated Annexin V, β-galactosidase, and EGFP. While the delivery of Annexin V was not successful at all, the delivery of EGFP/lipofectamine worked well with MCF10a as shown in Figure 24. The delivery of β-galactosidase worked well in all cell lines with both BioPorter and Chariot (which was to be expected since the manufacturer had optimised his delivery reagents with this enzyme).

Adherent cells**:** Gentle aspiration of the media from the plates leave cells attached to the surface. Trypsinisation destroys the extracellular matrix following which the free cells could be centrifuged (4000 rpm, 10 min, room temperature), resuspended in growth media, and counted in a COULTER counter. Cover slides were carefully placed, one in each well, on which 2 x 10^5 cells were seeded. Growth media containing serum was added, and the cells incubated at 37°C overnight to attain ~ 70% confluence. About 4 hours prior to starting the experiment, media was aspirated, starvation media (without serum) added, and protein-transfection agent applied. A further incubation for 1 hour in temperature and humidity controlled environment was required before the addition of media. Visualisation of fixed cells at time points of 0 hrs, 6 hrs, 12 hrs, 24 hrs and 48 hrs were made under the confocal fluorescence microscope (Zeiss Olympus BH series).

Among all adherent cells, MCF10a proved to be the best suited for our studies. These cells were routinely incubated in 24 well plates at 37°C until they reached the concentration about 250,000 units/ml. After a change of growth medium, a mixture of lipofectamine and protein (previously incubated at room temperature for about 30 min) was applied on the cover slips. Cells on the cover slip were provided with fresh medium and incubated at 37°C for 6 to 48 hours. After that, the medium was removed and the cover slips were washed with PBS. Washed cells were fixed with a solution containing PBS with 2% formaldehyde and 100 mM sucrose for about 30 min. The cover slips were then washed again with PBS and finally with H_2O containing DAPI (solution for staining nuclei) and embedded on glass plates with fluoromount (SIGMA).

Suspension cells: Processed in a similar manner as adherent cells, except that trypsin need not be added to aid their unattachment from the surface.

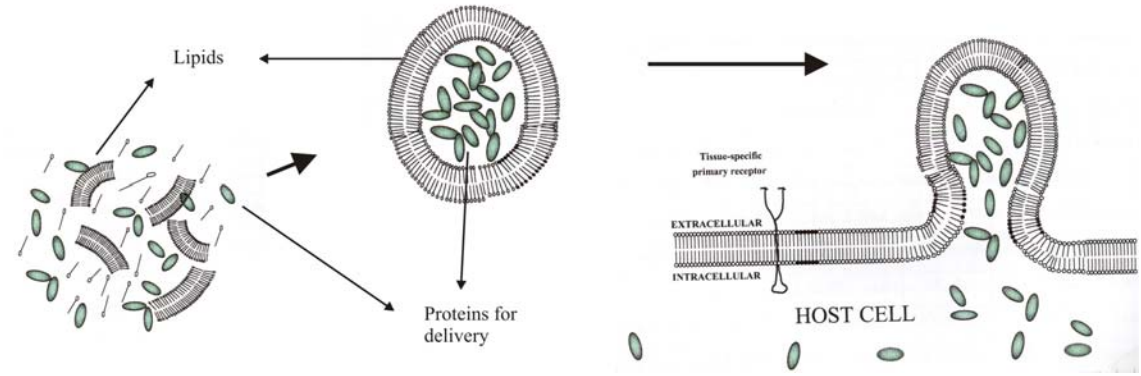


Figure 24a. Principles of lipid encapsulated protein delivery in cells.

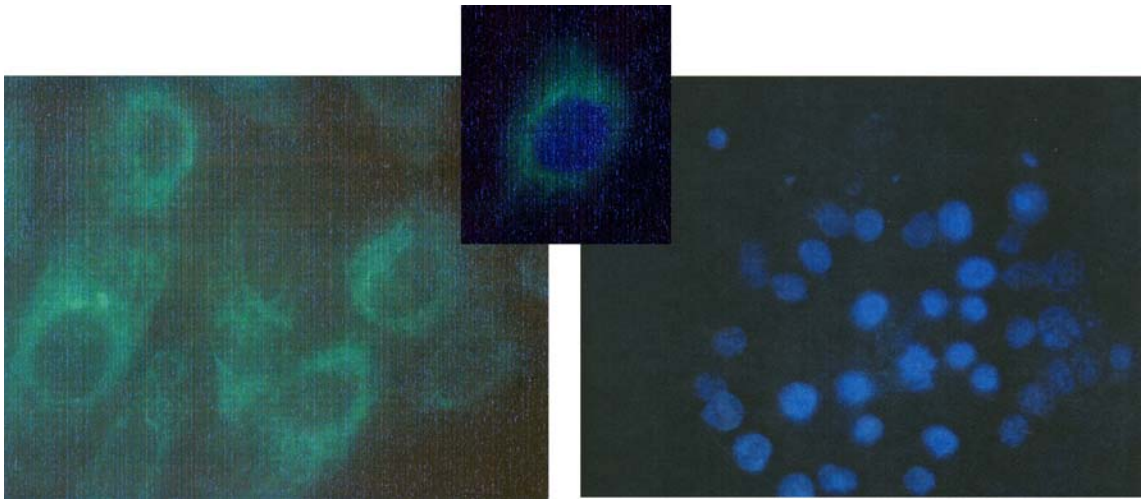


Figure 24b. Confocal microscopic images. Left: Proof of intracellular delivery of (3-F)Tyr-EGFP (excitation at 450 nm) by lipofectamine by its localization of GFP. Right: DAPI-staining (excitation at 350 nm) of cell nuclei in cultured breast epithelial human cells (MCF10). Middle: Superimposition of images of DAPI-stained and GFP-containing single cell. Note clear cytoplasmatic localization of delivered (3-F)Tyr-EGFP.

5.8.2 Tumour cell staining

DAPI is known to form fluorescent complexes with natural, double stranded DNA, showing fluorescence specificity for AT, AU, and IC clusters. The morphology of the nuclei could be observed under the fluorescence microscope at the excitation wavelength of 350 nm. Healthy nuclei having a normal phenotype, show a bright, homogeneous glow, whereas apoptotic nuclei could be identified by the condensed chromatin clusters at the periphery of the nuclear membrane or complete fragmentation of nuclear bodies.

β-galactosidase staining: Media was removed from the cells, washed with PBS, buffer added for fixation and incubated at room temperature for 15 minutes. X-Gal (5-bromo-3-indoyl-β-D-galactopyranoside stock solution) was then added, and cells incubated in a

humidified incubator at 37°C for 12 hrs. Cells were ready to be analysed after washing off the stain and embedment of the cover slips.

5.8.3 Measuring effects of fluorinated proteins on tumour cell lines in culture

The basic intention of performing cell culture studies was to figure out the amounts of canonical amino acids and fluorinated protein required to limit growth, and inhibit any further division of tumor cells. The results obtained in this study are preliminary, and further studies are required to gain a clear insight into the possible effects fluorinated proteins might have on the fate of tumor cells. Synthetic medium, similar to the one used for *E. coli* was prepared for cell culture studies. Separate media mixes, each deficient in one of the three amino acids, tryptophan, tyrosine and phenylalanine were prepared from the Dulbecco's minimal media (Sigma Aldrich) comprising of the following ingredients :

Earle's buffered salt solution (EBSS)

MEM vitamin solution

MEM non essential amino acids

Glucose supplements were made to a final amount of 4.5 g/l (HEK 293 cells require extra glucose) together with serum dialysed against PBS with a 3000 MWCO membrane. Cells were grown in a 96 well optical flat bottomed plate (Nunc). To determine the number of viable cells in culture, CellTiter-Glo® Luminescent Cell Viability Assay (Promega) was used that generates a "glow-type" luminescent signal, produced by the firefly luciferase reaction. Cell lysis of metabolically active cells following the addition of the CellTiter-Glo™ reagent is directly proportional to the amount of ATP present in cells that could be measured in the MICROLUMAT Plus plate luminometer.

6 RESULTS AND DISCUSSION

6.1 GFP design and engineering with an expanded amino acid repertoire

Attempts to redesign the chromophore of GFP by canonical amino acids are limited by the fact that the standard genetic code can offer only four aromatic amino acids in its repertoire (Figure 25). The mutants Tyr66→Phe/His/Trp show a characteristic blue emission since the excited state proton transfer, responsible for the characteristic green fluorescence in native *av*GFP is prohibited.[104] Nature, however, has achieved an extension of the chromophoric π-system by using different principles such as i) in the case of the red fluorescent protein from *Discosoma striata* (*dsRed;* $\lambda_{max(em)}$= 573 nm) integration of an additional amino acid into its chromophore[117] and ii) in the case of the chromoprotein *asFP595* ($\lambda_{max(em)}$= 595 nm) from *Anemone sulcata,* augmentation of its π-system with an imino group.[109] Finally, a new principle of photophysics that causes a red shift in the fluorescence by the incorporation of aminoindole into the chromophore, was developed previously in our laboratory. [108] It is based on protein relaxation as a response to a strong change in dipole moment of the chromophore upon excitation (for more details see Ref. 85). In short, the insertion of an amino group as an electron donor into the chromophore of ECFP leads to a "gold fluorescent protein", (GdFP, $\lambda_{max(abs)}$= 466 nm, ε_M= 23700 $M^{-1}cm^{-1}$, $\lambda_{max(em)}$= 574 nm). GdFP is the most red shifted *av*GFP-variant known to date: 69 nm when compared to the parent ECFP and 47 nm when compared to the emission maximum of "enhanced yellow fluorescent protein", EYFP.

Recently, Budisa[56; 85; 108] Schultz[118] and Sisido[119] reported experiments on the incorporation of mostly aromatic noncanonical amino acids into *av*GFP in order to generate unnatural autofluorescent proteins, either by their introduction directly into the chromophore or in the surrounding protein matrix. These noncanonical amino acids (shown in Figure 25) are mainly analogues or surrogates of Tyr, Phe and Trp. Their abbreviations and full names are: [3,2]Tpa – β-(thieno[3,2-*b*] pyrrolyl)-L-alanine; [3,2]Sep – β-Selenolo [3,2-*b*]pyrrolyl-L-alanine; [(4-F)Trp, L-4-fluorotryptophan; (5-F)Trp, L-5-fluorotryptophan; (6-F)Trp, L-6-fluorotryptophan; (7-F)Trp, L-7-fluorotryptophan; (4-Me)Trp, L-4-methyltryptophan; (4-NH_2)Trp, L-4-aminotryptophan; 2-naphtyl-Ala, 2-naphtylalanine; *p*-Br-Phe, *para*-bromo-phenylalanine; *p*-I-Phe, *para*-iodophenylalanine; *p*-NH_2-Phe, *para*-aminophenylalanine; *p*-NH_2-Phe, *para*-amino-phenylalanine; *O*-Me-Tyr, *O*-methyltyrosine; *m*-F-Tyr, *meta*-fluorotyrosine or (3-F)Tyr, 3-fluorotyrosine; *o*-F-Tyr, *ortho*-fluorotyrosine or (2-F)Tyr), 2-fluorotyrosine.

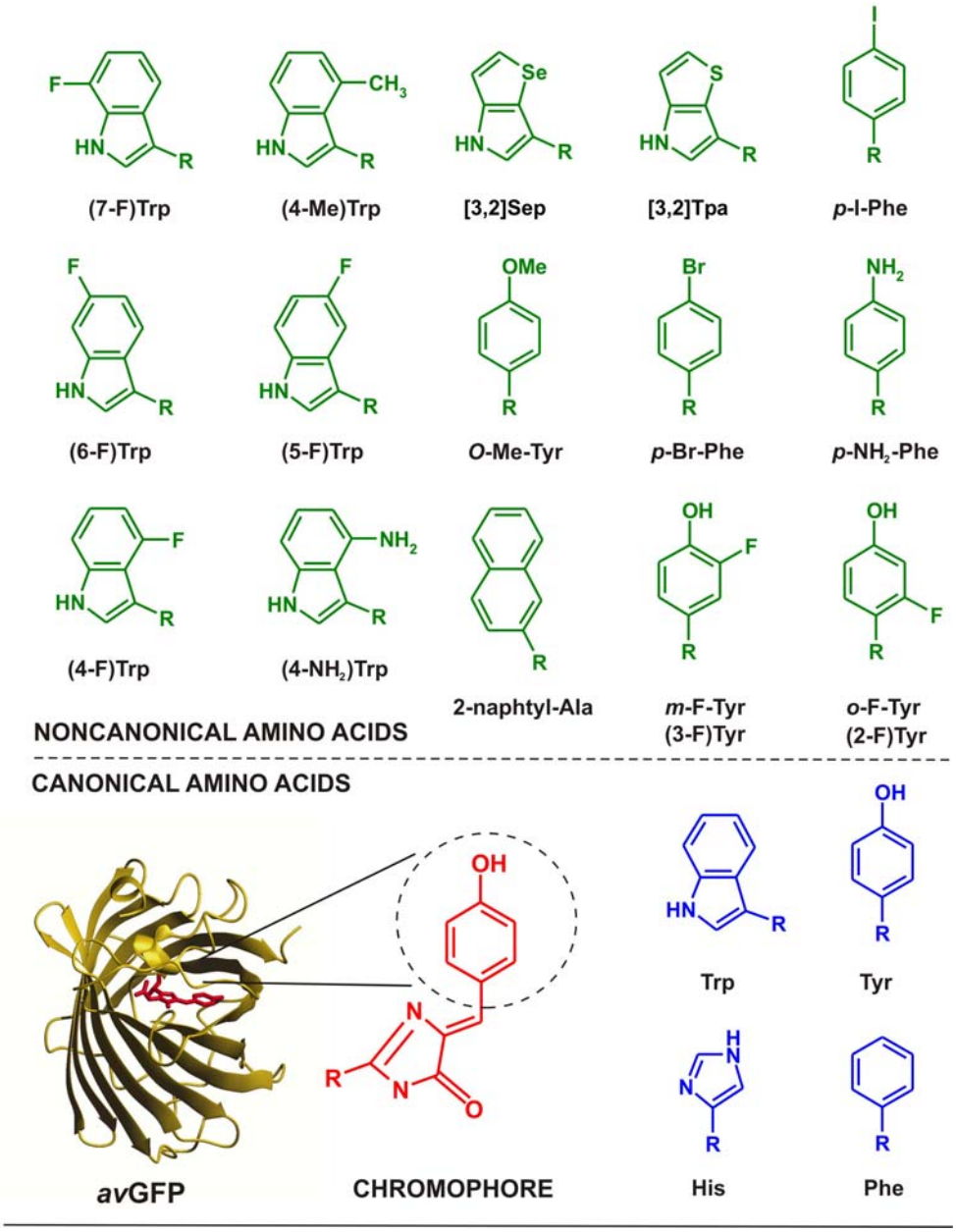


Figure 25. Chromophore engineering of *av*GFP with an expanded amino acid repertoire. One of the basic prerequisites for the formation of the chromophore of *av*GFP is the presence of an aromatic amino acid at position 66 in the protein sequence. The universal genetic code limits the number of possible amino acids at this position to the canonical amino acids tryptophan (Trp), tyrosine (Tyr), histidine (His), and phenylalanine (Phe). A direct redesign the chromophore of *av*GFP was attempted with the whole library of noncanonical Tyr, Phe and Trp analogs and surrogates (full names of these substances are given in the text).

6.2 Fluorination of the Trp-residues in avGFPs

The basic features of *av*GFPs include a visible absorbance and characteristic green fluorescence from their unique chromophores encoded in their gene sequence. Their spatial structures as well as biochemical and biophysical properties are well established. Moreover, the *in vivo* and *in vitro* plasmid-directed expressions of these proteins proved to

be robust, efficient and controllable – a basic prerequisite to take advantage of the expansion of an amino acid repertoire. This makes *av*GFPs almost ideal models to study spectral effects that result either from a direct fluorination of the chromophore or of its microenvironment. These investigations required the replacement of Trp57 of EGFP as well as of the two Trp residues in ECFP by a library of Trp analogues and surrogates (Figure 25) using the Trp-auxotrophic *E. coli* strain ATCC 49980 and the T5-based expression system for applying the SPI-method.

The fluorinated analogues of aromatic canonical amino acids do not support cellular growth of the *E. coli* Trp-auxotrophic strain ATCC49980 in NMM when they serve as a sole source of replacement for L-Trp. In fact, growth of this strain is strongly inhibited from the beginning of fermentation in defined minimal media that contains 0.015 mM of any of these amino acids. Therefore, the cells were grown in minimal medium (NMM) containing 0.015 mM L-Trp as a natural substrate until its exhaustion followed by the simultaneous addition of the Trp fluorinated analogues and target gene induction with IPTG. Under these conditions, optimal expression and production of ECFP and EGFP derivates was achieved in yields comparable to those of the native proteins (10-30 mg/L).

By this method almost quantitative incorporation of fluorinated analogues were readily achieved, although a smaller amount (2-4 %) of wt-protein is always present in all variant samples. It is well known that even the best controllable expression systems are not perfectly "silent" before induction of the target gene expression.[92]

6.2.1.1 Spectral effects of EGFP Trp57 replacements with fluorinated analogues

The wt-*av*GFP contains only one Trp residue. Its inherent fluorescence is quenched due to the energy transfer to the chromophore. In addition, the available tertiary structure reveals a relative rigidity (i.e. immobility) of both Trp57 and the chromophore which are separated by about 15 Å (Figure 11). It is therefore conceivable that the fluorescence of EGFP and its absorbance spectrum should not be affected upon fluorination of Trp57, but that the modified residues should contribute to the fluorescence resonance energy transfer (FRET). Indeed, Figure 26 shows that the UV spectra of all EGFP variants (Table 5) are not affected by the modification of Trp57. The observed differences in the chromophore extinction coefficients among these variants derive from the additive contributions of the Tyr+Trp spectral properties (which are affected to a certain extent by Trp-modifications). Furthermore, $\varepsilon_{M(Cro66)}/\varepsilon_{M(Tyr+Trp)}$, ratios remain more or less constant (Table 5).

It is worth noting that in the structural scaffold of *av*GFP, Trp (and its fluorinated analogues at position 57) are actively involved in FRET (fluorescence resonance energy

transfer) to chromophore. Interestingly, fluorination of the single Trp57 in the micro-environment of the chromophore by the "silent fluorophore" (4-F)Trp, does not suppress its fluorescence, since in both ECFP and EGFP, (4-F)Trp57 (which is about 15 Å distant, but coplanar to the chromophore) participates in FRET with the longer wavelength chromophore. Thus, possible quenching effects of (4-F)Trp57 are masked and not accessible for experimental studies. Detailed studies about the interactions of Trp57 with the chromophore are provided in a recent study of Visser and co-workers.[120] However, other proteins such as human Annexin V or barstar, in which the Trp-residues were globally substituted by “silent fluorophores” like (4-F)Trp and (7-F)Trp, have no Trp-fluorescence.[56]

6.2.1.2 Fluorinating the chromophore of ECFP

The UV-spectra of the ECFP chromophore with fluorinated indole side chains are visibly changed due to the dominant contribution of the indole moiety in Cro66 (Figure 26 and Tables 5 and 6). The electron-withdrawing fluorine substituted at different positions in the indole moiety directly affects absorbance and fluorescence maximum as well as the intensity of related ECFPs. All fluoro-Trp-containing ECFP exhibit a blue shift between 4 nm to 10 nm in emission and absorbance maxima with the exception of (6-F)Trp-ECFP. The $\varepsilon_{M(Cro66)}/\varepsilon_{M(Tyr+Trp)}$ ratio remains more or less unchanged for (6-F)Trp-ECFP, while this ratio is decreased by more than 30% for (4-F)Trp-, (5-F)Trp-, and (7-F)Trp-ECFP (Table 5, Figure 26). There is no doubt that fluorination is responsible for the measured blue shifts and reduced intensities of the ECFP chromophores. Thus, the observed properties can be regarded as an intrinsic feature of the fluorinated chromophore. Curiously, the electron donating methyl-group induces a slight red shift (4 nm) in both the absorbance and fluorescence maximum of the (4-Me)Trp-ECFP without any significant changes in their intensities.[121]

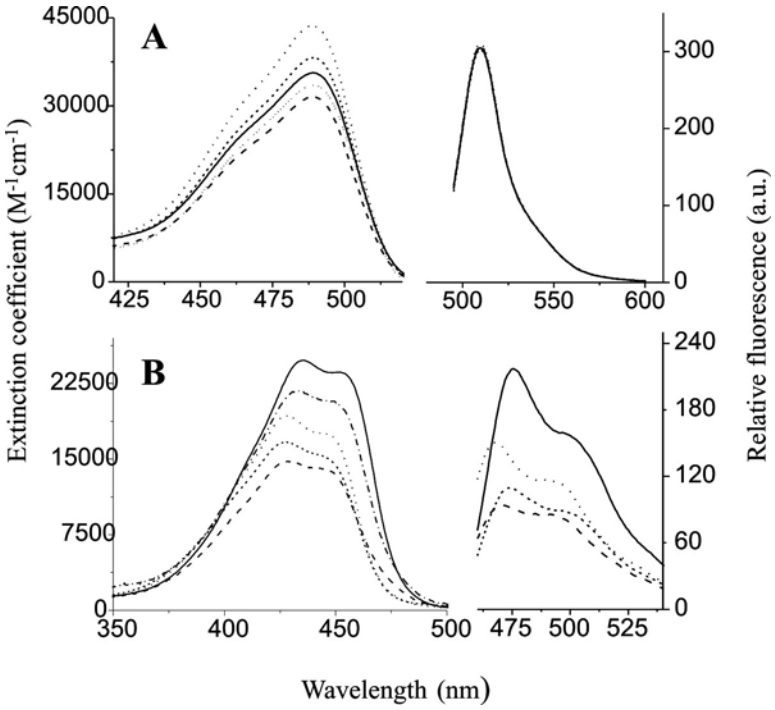


Figure 26. Absorbance and fluorescence spectra of fluorinated *av*GFPs. (**A**) Spectral effects upon substitution of the Trp57 in EGFP with fluorinated indoles. The differences in the chromophore extinction coefficients of EGFP are due to normalisation to the values at 277 nm (Tyr+Trp spectral region), but there are no red or blue shifts in the absorbance or emission maximum. Upon excitation at 488 nm the intensities of the emission spectra are almost identical. (**B**) ECFP variants where both chromophore and Trp57 are replaced with fluorinated Trp analogues exhibit not only differences in the absorbance and emission intensities but also their maxima are blue shifted by 2 – 10 nm. Numerical values are reported in Tables 5 and 6; the experimental conditions are described in Materials and Methods. Full line – wt-protein; longer dashed lines – (4-F)Trp-protein; dotted lines – (5-F)Trp-protein, dot-point lines – (6-F)Trp-protein; shorter dashed lines – (7-F)Trp-protein.

6.2.1.3 Wild-type and fluorinated EGFP/ ECFP: pH titrations

It is well known that the fluorescence and absorbance properties of various *av*GFPs are strongly pH dependent. The π-system of the conjugated chromophore of EGFP contains in principle, two titratable groups, the hydroxyl group at the phenolic end and the imidazolinone ring nitrogen at the heterocyclic end. In this context, replacement of the Trp57 by analogues should not cause large deviations in the pK_a value when compared to native EGFP (5.85). As expected, the pK_a values of fluorinated derivatives are in this range (Figure 27) while the pK_a -values of halogen-derivatives of EGFP are slightly reduced by about 0.15 units (Table 6). The Cro66 in ECFP is not regarded to be neutral, while the protonation state of its indole imino-group has not yet been established. Nevertheless, modifications of the Cro66 in ECFP should directly be visible in the titration curve profile. However, the determined pK_a values of (5-F)Trp-, and (6-F)Trp-ECFP are only slightly reduced. On the other hand, the pK_a values of (4-F)Trp- and (7-F)Trp-ECFP are reduced by 0.3-0.4 pH units (Figure 27). Since no theoretical or model-substance studies of fluorine-

substituted chromophores of ECFP are available, one can only speculate to which extent the lowering of pK_a upon fluorination results from inductive effects exerted by the fluorine atoms themselves or their micro-environments.

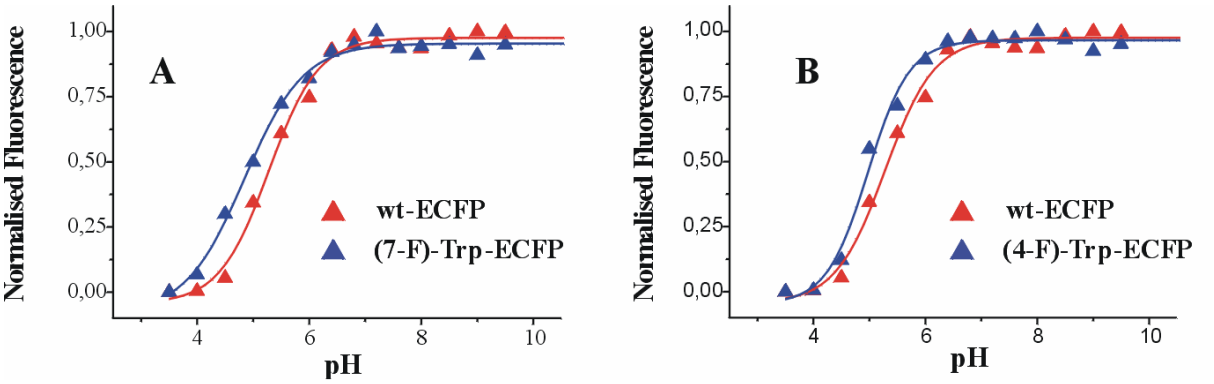


Figure 27. pH titration curves of native ECFP and its fluorinated variants. The fluorescence and absorbance properties of various *av*GFPs are strongly pH dependent. The π-system of the conjugated chromophore of ECFP contains titrable groups. Therefore, its fluorination should affect pK_a values. Lowered pK_a values of about 0.3 - 0.5 pH units upon fluorination (pK_a for wt-ECFP: 5.28; (4-F)Trp-ECFP: 4.86; (4-F)Trp-ECFP:4.97) (see Table 6) are most probably due to an inductive effect exerted by the fluorine atoms or by their micro-environments.

6.2.1.4 Structure of (4-F)Trp-ECFP revealed the importance of Met218/Trp57 pair

It was recently shown the *in vitro* global replacement of all Met-residues with selenomethionine in wt-*av*GFP leads to an increase in the intensity of the chromophoric absorbance by about 30%.[122] This is strikingly similar to the spectral effects of the incorporation of [3,2]Tpa in EGFP as well as the introduction of Met at position 69 in the neighbourhood of the chromophore of "yellow" *av*GFP.[106] Recently, Sisido and co-workers[119] reported an improvement of the rather low quantum yield of *p*-NH_2-Phe-GFP by using a specific approach. They used a combination of random mutagenesis with a cell-free translation system based on four-base codons (i.e. frameshift suppression). In addition, they found that the mutation Tyr145Met increases the fluorescence intensity up to four-fold. The most plausible explanation for these effects is the presence of the electronically rich sulphur/selenium atom which is in a particularly favourable geometric arrangement in the chromophore neighbourhood. This might in turn, facilitate the increase in its absorbance and fluorescence intensities.

Since the crystal structure of (4-F)Trp-ECFP is available, the positions of some indole hydrogens could be "visualised" because they have been replaced by fluorine atoms. The fluorine atom of (4-F)Trp57 was found to be well within the range of hydrophobic contacts (3.5 Å) with the sulphur atom of Met218 (Figure 28). Sulphur atoms are known to lack the

electronegativity of a strongly hydrogen binding atom i.e. they are highly polariziable. It is also known that sulphur-aromatic interactions occur between Met and aromatic side chains like those of Phe, Tyr or Trp. The preferred separation distance between a sulphur atom and the centroid of the aromatic ring is about 6 Å. Interactions occur between the δ^+ ring hydrogen atoms and the δ^- sulphur atom while avoiding the δ^- π-electron cloud of the aromatic group. The crystal structure of (4-F)Trp-ECFP (Figure 28) reveals the configuration of Met218 relative to Trp57 in *av*GFP and is in full agreement with these considerations since the sulphur atom points toward one of the aromatic ring hydrogen. Thus sulphur-aromatic interactions, like oxygen-aromatic interactions in proteins, probably contribute significantly to the overall structural stability in *av*GFP.

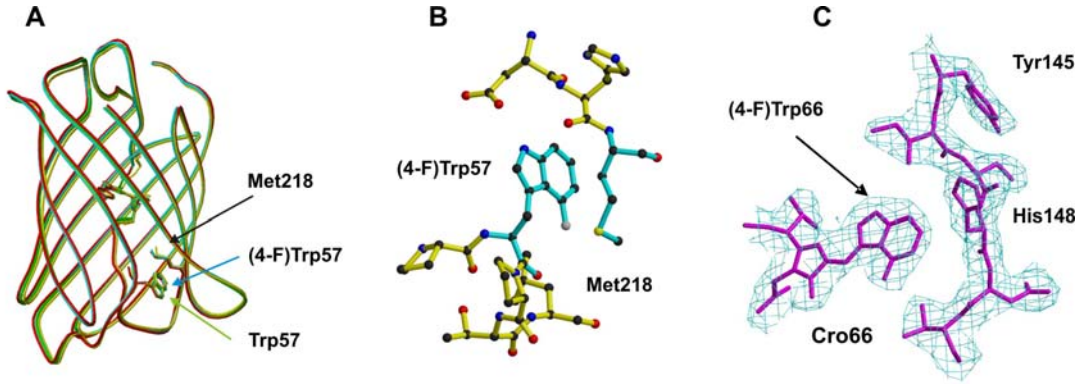


Figure 28. Structural environment of fluorinated chromophore and Trp57 in ECFP. (**A**) Superimposed structures of native ECFP (green) and (4-F)Trp-ECFP (blue) do not show any significant changes in the overall fold at positions 57, 218 or in the Cro66 and its environment upon replacements. (**B**) The electron density of the chromophore of (4-F)Trp-ECFP contoured at 1.7 σ. (**C**) The H $\rightarrow$ F exchange at the indole ring of (4-F)Trp57 of the (4-F)Trp-ECFP. Since fluorine (1.35 Å) and hydrogen (1.20 Å) have a similar van der Waals radii, the distance of 3.5 Å revealed by such "atomic mutation" confirmed a possible sulphur-aromatic interaction between the δ^- sulphur of Met218 and the δ^+ring hydrogen of Trp57 in the native protein.

Systematic mutagenesis experiments showed that out of the five Met-residues in native *av*GFP, all are mutable with other canonical amino acids with the exception of Met218.[94] This cannot be replaced by Leu, Ile, Cys, Ser, Thr and Ala without affecting the biophysical properties and structural integrity. The same holds true for Trp57, whose mutations to other canonical amino acid residues like Leu or Phe (with their relatively large hydrophobic side chains) resulted in an unfolded protein. This highlights the importance of the free energy of folding provided by the complete burial of the large indole ring of Trp57 into *av*GFPs. On the other hand, interactions between the sulfur atom

of Met218 and the aromatic ring of Trp57 seem to be equally important for the structural integrity of *av*GFP and its spectroscopic function.[85]

6.2.1.5 Uptake and influence of fluorotryptophans on the growth of tumour cells in culture

It is known that upon intracellular uptake, noncanonical amino acids may act as (a) inhibitors of biochemical pathways or of specific enzymes and (b) precursors for analogues or surrogates of other critical biomolecules such as acetate, citrate, aminergic neurotransmitters etc. Preliminary experiments were performed in MEM-minimal medium (see Materials and Methods) to see which amounts of fluorinated Trp could inhibit growth in the HEK239 cell culture. The concentrations of (4-F)Trp, (5-F)Trp and (6-F)Trp tested were 5, 50 and 500 μM respectively. While cells grow well in a full medium supplied with L-Trp, all three concentrations of fluorinated analogues effectively inhibit growth. Experiments performed in parallel by H. Lilie and co-workers[123] on MCF-7 cell lines have shown that the presence of fluorinated Trp-analogues extracellularly have no cytotoxic, but rather inhibitory effects on the culture growth. It was demonstrated that the IC_{50} value for (4-F)Trp is about 3 μM and even the presence of L-Trp (25 μM) in the media does not reverse its cytostatic effect. The cytostatic effect was probably due to the blockage of cell cycle. The presence of about 1 μM of (4-F)Trp and (6-F)Trp in the growth media correlated well with the 50 – 120 μM of intracellular concentration of these substances, indicating very efficient intracellular accumulation, most probably via the Trp-uptake system. This indicates that relatively high intracellular amounts of fluorinated analogues should be accumulated in order to achieve inhibitory effects. Therefore, the substituted proteins as a "shuttle" system for specific delivery of these inhibitory substances into tumour cells does not seem too promising for tumour therapy at this stage of our experimentl expertise.

6.3 EGFP and EYFP with globally fluorinated Tyr-residues

6.3.1.1 Spectral properties of fluorinated EGFPs and EYFPs at neutral pH

The basic features of *av*GFPs include the visible absorbance and characteristic green fluorescence from their unique chromophores encoded in the gene sequence. Classical protein engineering based on site-directed mutagenesis yielded several classes of *av*GFP-mutants (Figure 13) with the differences based on their chromophore photophysics.[102] In this system, the wild-type GFP from *Aequoria victoria* (*av*GFP) with the phenolic group of chromophore Tyr66 in equilibrium between protonated (**AH**) or deprotonated ($\mathbf{A}^-$) state, belongs to Class I. Biochemical and optical properties of *av*GFP mutants from Class II

with a mutation Ser65Thr having the chromophore predominantly in $\mathbf{A}^-$ (phenolate) form like EGFP, are especially attractive for numerous practical applications in molecular and cellular biology.[124] The mutation Thr203Tyr not only brings an additional Tyr residue into the structure of *av*GFP but also creates an "enhanced yellow" *av*GFP – EYFP, with the characteristic "yellow" emission, which belongs to Class IV. The available tertiary structure of this mutant shows a relative rigidity (i.e. immobility) of both the Tyr203 side chain and the chromophore which are separated from each other by about 4 Å.[125]

An almost isosteric H→F single atom exchange in the tyrosyl-moiety results in inverted polarities and thus increased charge separation in the chromophore ring. In the present study it was shown that such structurally isomorphous changes introduced by amino acid analogues and surrogates in *av*GFPs usually do not affect the secondary and crystal structure of the protein.[116] But these modifications might well exert dramatic changes in their spectral and thermodynamic properties.[66] In the present work, fluorination of Cro66 of EGFP and ECFP provides not just a direct integration of fluorine into the chromophore, but also global fluorination of all Tyr residues throughout the protein molecule. This enables a study of the effects of micro-environment fluorination on the spectral properties of the protein.

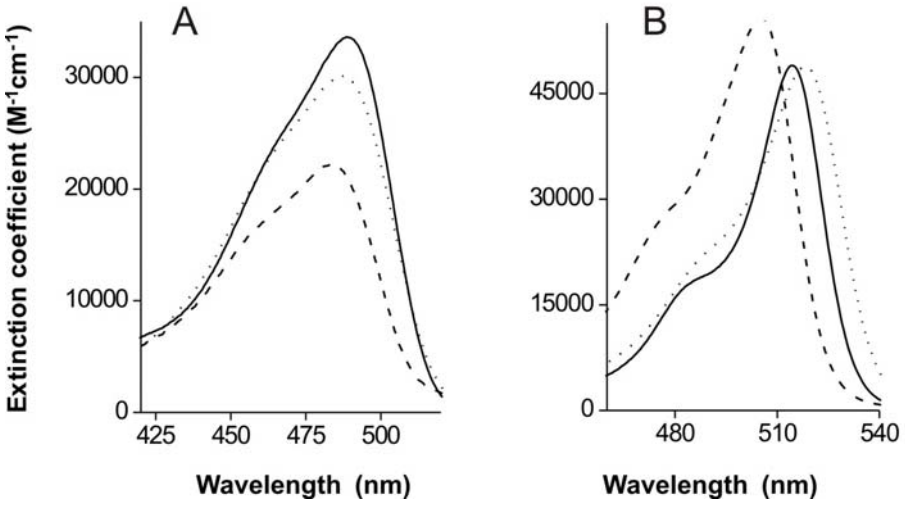


Figure 29. (**A**) Spectral effects upon global fluorination of the Tyr-residues in EGFP. The absorbance maxima of both fluorinated variants are only slightly blue-shifted when compared to that of the parent protein (full line). (**B**) Global replacement of EYFP with (3-F)Tyr (dotted line) does not change the absorbance and emission intensities, while (2-F)Tyr-EYFP (dashed line) exhibits increased spectral intensities when compared with the parent protein (full line). On the other hand, the maxima for the absorbance of (3-F)Tyr-EYFP are red-shifted, and for (2-F)Tyr-EYFP blue-shifted to a much higher extent than those of the EGFP-variants (see Tables 6 and 7). The chromophore extinction coefficients of EGFP are derived by normalization to the values at 277 nm (Tyr+Trp spectral region) as described in the Materials and Methods.

In the absorbance profiles, global replacements of all Tyr side chains in both proteins show blue shifts for the chromophores in (2-F)Tyr-proteins, and red spectral shifts for that of (3-F)Tyr-EYFP (Table 1). In general, the fluorescence emission maxima of (2-F)Tyr-proteins are blue shifted, while those of (3-F)Tyr are red shifted with a significant enhancement in the EYFP variants at neutral pH (Table 2). Relative fluorescence as well as the ε_M of (3-F)Tyr-EYFP when compared to wt-EYFP was almost unchanged, while in case of (2-F)Tyr-EYFP, their intensities increased by almost 15 % (Figure 29). Conversely, there is a remarkable decrease (about 35 %) in the absorbance and the emission intensities of (2-F)Tyr-EGFP when compared to those of the parent protein (Tables 7 and 8). The decreased molar absorbance coefficient in the order wt-EGFP > (3-F)Tyr-EGFP > (2-F)Tyr-EGFP is responsible for the reduced intensities of the fluorescence emission as shown in Figure 29.

To check whether the observed properties in EGFP are due to fluorination of exclusively the chromophore, or if there is additional contribution from the non chromophoric Tyr residues, another "enhanced cyan" *av*GFP mutant (ECFP), [86] was used as a control. It contains the same number of Tyr residues (10) as EGFP while the 11th one is replaced with Trp at position 66 (chromophore). As expected, global fluorination of ECFP with both (3-F)Tyr and (2-F)Tyr does not affect either its absorbance or fluorescence properties (absorbance: λ_{max1}= 434 nm λ_{max2}= 452 nm; fluorescence: λ_{em1} = 476 nm; λ_{em2} = 505 nm). This is clear evidence that the observed spectral shifts in EGFP are not due to a fluorination of the surrounding Tyr residues but are brought about solely by the fluorination of the chromophoric Tyr66. Therefore the measured spectral red- and blue-shifts in the (3-F)Tyr- and (2-F)Tyr-containing EGFPs can undoubtedly be regarded as intrinsic features of the fluorinated chromophores. An additional proof is provided by measuring of absorbance of globally fluorinated ECFP which (unlike EYFP and EGFP) displays no difference in the spectral properties (shifts, intensities) between the parent protein and its fluorotyrosyl-variants also in denatured state (see Table 7). On the other hand, since the main difference between EGFP and EYFP is the presence of the additional Tyr203 in the structure of EYFP, the enhanced blue-, and red shifts can be attributed to the fluorination of this residue which directly affects its stacking interaction with the fluorinated Cro66.

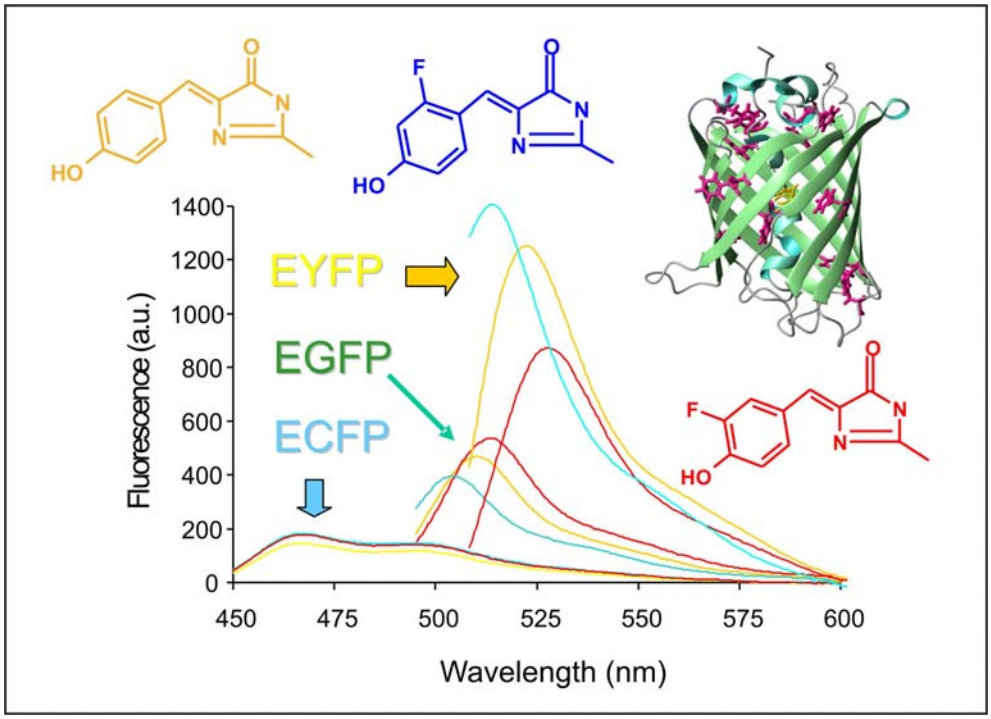


Figure 30. Fluorescence emission spectra of recombinant wild-type proteins (faint orange) EGFP, ECFP and EYFP and their fluorinated variants measured at neutral pH. Tyrosine residues of these proteins were globally substituted with *ortho*-F-Tyr (blue) and *meta*-F-Tyr (red). Proteins (0.5 μM) were excited at their chromophore absorbance maxima as described in Materials and Methods.

6.3.1.2 pH Dependence of absorbance and fluorescence in EGFP and EYFP and their fluorinated variants.

The absorbance of *av*GFPs is characterised by a complicated spectra having a dual maxima at 400 nm and 488 nm for wt-*av*GFP, a feature which has been assigned to the chromophore phenolic oxygen that can exist either in the protonated (neutral, AH) or deprotonated (anionic, A-) state.[124] While the absorbance spectra of native EGFP and EYFP are strongly pH dependent, this tendency is less pronounced in all their fluorinated variants except (2-F)Tyr-EYFP (Figure 31). However, native EGFP, EYFP and (2-F)Tyr-EYFP display a congruence in both absorbance and excitation in the pH range 8.5 –11. Lowering the pH leads to a decrease in absorbance at 488 nm for EGFP and at 515 nm for EYFP respectively. This process is accompanied by a concomitant increase in the absorbance intensity of the non-fluorescent AH species at 398 nm in both proteins (Table 9). The titrations of native proteins and their fluorinated variants are characterised

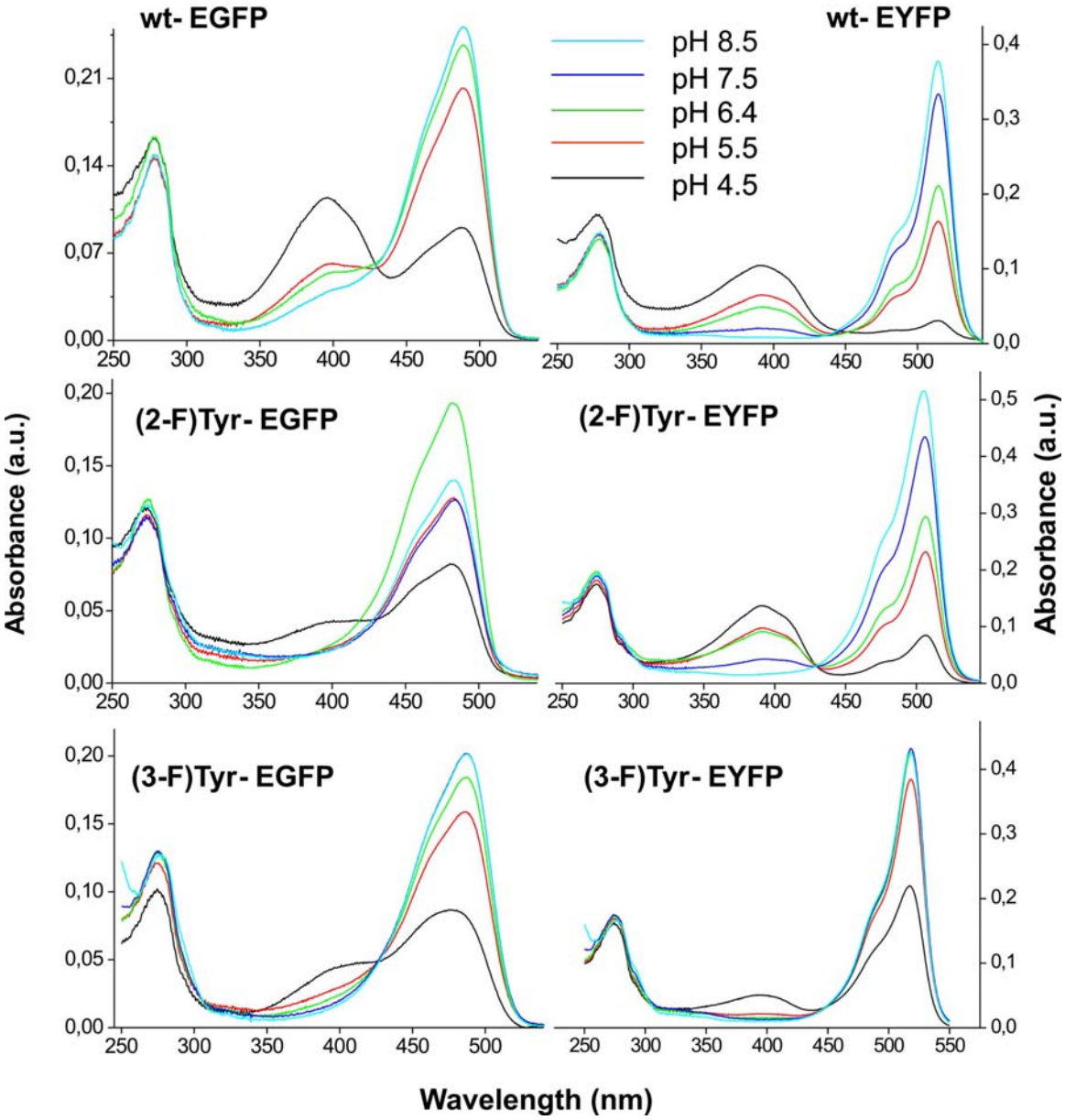


Figure 31. pH-dependence of the absorbance of parent *av*GFPs and their fluorinated variants in 100 mM NaCl and 50 mM buffer solutions. In both wt EYFP as well as its (2-F)Tyr-variant, lowering the pH leads to a decrease in the absorbance at 480 and 515 nm (deprotonated form, A-), and the concomitant increase at 398 nm (protonated form AH) with a clear isosbestic point in parent proteins. Conversely, the increase in absorbance of the protonated form upon a decrease in pH is suppressed in most of the fluorinated protein variants.

by a clear isosbestic point (425 nm), indicating a single protonation/deprotonation process as demonstrated in earlier studies.[126] For example, the fluorinated **AH** chromophores of denatured protein variants at pH 1.0 do not show any shift in their absorbance profiles (λ_{max} = 378 nm) when compared with parent EGFP and EYFP proteins in their folded state. In contrast, at higher pH (11.0) an uniform blue-shift upon fluorination [λ_{max} (parent proteins) = 447 nm, λ_{max} ((2-F)Tyr-variants) = 444 nm; λ_{max}((3-F)Tyr-variants) = 443 nm]

is observed. This suggests that the spectral shifts represent an intrinsic property of the fluorinated $\mathbf{A}^-$ chromophores (Table 9).

The fluorescence emission intensity of *av*GFPs is highly pH sensitive as well, dropping to zero at very low and high pH values. The shapes and maxima of the emission spectra, shown in Figure 31, are pH independent in all analysed protein samples. A fit for one titratable group of the normalised emission intensity versus pH shows that fluorination of EGFP in the *ortho*-position decreases the pK_a-value by about 0.2 pH units, while the fluorine in the *meta*-position causes an additional decrease in the pK_a by about 0.5 pH units. Interestingly, the absorbance of the mutant (2-F)Tyr-EGFP first increases by 35 % starting from pH 6.4 to 6.8, but then decreases by about the same amount at higher pH values, in the range of 7.6 to 8.5 (Figure 31). This mutant is also less stable and more prone to aggregation than wt- and (3-F)Tyr-EGFP. On the other hand, a direct fluorination (either *ortho*- or *meta*-) of the chromophore in EGFP seems to significantly suppress the appearance of the protonated (**AH**) species in these variants with a decrease in pH.

The native EYFP has been characterised by the pK_a value (6.6) that is almost one pH unit higher than wt-EGFP (5.7) (Table 8). One of the most prominent features of the fluorinated variants is the suppression of the appearance of the **AH** species at lower pH values, (2-F)Tyr-EYFP being the exception to this rule. The pH dependence of absorbance and fluorescence in these variants are quite similar to that of their parent protein. In addition, fluorination of both Cro66 and Tyr203 in the *meta*-position is probably the main reason for the considerable decrease in the pK_a-value from 6.6 (native protein) to 5.3 ((2-F)Tyr-EYFP) as well as for the suppression of the appearance of **AH** species at a lower pH values.

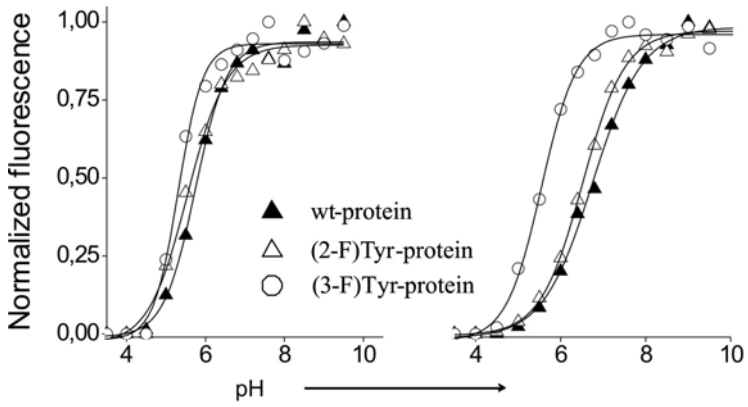


Figure 32. Effects of chromophore fluorination on fluorescence pH-titration profiles in parent and variant EGFPs (left) and EYFPs (right). The data points were fitted to a theoretical titration curve with one titratable group since the absorbance titration revealed only one isosbestic point (see Figure 31) (calculated pK_a values are available in Table 8)

Fluorinated EYFPs having its Cro66 in stacking interaction with the Tyr203 provides a tool to evaluate the extent to which the lowered pK_a upon fluorination was a result of an inductive effect exerted by the fluorine nucleus in Cro66 or by the fluorine atoms in its micro-environments. *Ortho*-fluorination of the Tyr residues in both EGFP and EYFP results in a lowering of the pK_a of the Cro66-fluorescence by about 0.2 pH units and 0.3 pH units respectively. M*eta*-fluorination of the Cro66 in EGFP, on the other hand, is probably the main reason for the decrease of pK*a* by 0.3 pH units (Table 2). Furthermore, the fluorescence pK_a-value in (3-F)Tyr-EYFP is considerably decreased by about 1.2 pH units, which should be at least in part attributed to the stacking interactions with (3-F)Tyr203.

It should be noted that the lowering of pK_a in the fluorinated protein takes place in the following order: parent protein > (2-F)Tyr-protein > (3-F)Tyr-protein. This correlates well with the pK_a values of the free amino acids, Tyr (10.0) > (2-F)Tyr (9.04) > (3-F)Tyr (8.5).[74] In the absence of theoretical or model studies of fluorine-substituted chromophores of *av*GFP it is difficult to rationalise the effects described above.

6.3.1.3 Changes in stability and spectral properties of *av*GFPs upon fluorination.

The chromophore of wt-*av*GFP is formed autocatalytically from Ser65, Tyr66 and Gly67, and exibits two major absorbance maxima at 395 nm and 475 nm, respectively.[98] Chromophore formation has been proposed to proceed through a two-step process consisting of a cyclization of the tripeptide Ser65-Tyr66-Gly67 followed by oxidation.[101] This intramolecular cyclization requires the proper folding of the surrounding matrix, since the chromophore itself is almost non-fluorescent.[127] In the present study, we observed that the tendency for aggregation of protein variants correlates well with the suppression of their AH species at lower pH values (Figure 31). The incorporation of fluorine was found to affect correct post-translational chromophore formation and maturation. For example, the tendency of aggregation of (2-F)Tyr-EGFP in comparison to the parent protein, might at least partly be due to the less efficient chromophore maturation or folding of the surrounding matrix. Indeed, both absorbance and fluorescence intensities of this variant are significantly decreased (about 25%, $\varepsilon_{M(Cro66)}/\varepsilon_{M(Tyr+Trp)}$ = 1.13) when compared to that of either EGFP or (3-F)Tyr-EGFP ($\varepsilon_{M(Cro66)}/\varepsilon_{M(Tyr+Trp)}$ = 1.75 and 1.56, respectively). In contrast, (2-F)Tyr as a part of the chromophore of EYFP was found to considerably enhance the extinction coefficient and relative fluorescence of this variant ($\varepsilon_{M(Cro66)}/\varepsilon_{M(Tyr+Trp)}$ = 2.71) (Figure 29, Tables 7 and 8). In fact, it is significantly less

aggregation-prone and relatively stable over longer periods of time at room temperature as well as in + 4 °C.

6.3.1.4 Effects of the fluorine on the chromophore photophysics

The complex photophysics of the chromophore, including properties like optical absorbance, fluorescence, and ultra-fast fluorescence dynamics of the native *av*GFP can be explained by the three-state photoisomerisation model assuming that the chromophore exists in either a neutral (395 nm, **AH**) or in an anionic (475 nm, $\mathbf{A}^-$) state. The latter exists as a thermodynamically unstable intermediate (493 nm, I-state) and in a stable low-energy form.[124] Structural analysis and quantum mechanical calculations suggest that these states correspond to chromophores with protonated and deprotonated phenolic groups, respectively. [127] Fluorine as the most electronegative element is expected to exert strong inductive electron-withdrawing effects on the fluorinated Cro66 ring. By doing so, it becomes relatively electron deficient and more hydrophilic, which makes its phenolic moiety more acidic. Such nearly isosteric H→F single atom exchanges would certainly affect both the excited and ground states of the phenol oxygen of Tyr66. The electron-withdrawing fluorine atom in the *meta*-position, being positioned in the vicinity of the phenolic group, enhances the acidity of (3-F)Tyr-Cro66 to a greater extent than does (2-F)Tyr-Cro66. This is understandable, as the p*Ka* value of free (3-F)Tyr (8.5) is 1.5 pH units lower than that of Tyr (10.0).[74] It thus follows that the deprotonated anionic $\mathbf{A}^-$ species of the chromophore obviously favours a fluorination at the *meta*-position than its protonated or neutral (**AH**) form. This explains the suppression of the **AH** species in (3-F)Tyr-EGFP and (3-F)Tyr-EYFP with decreasing pH (Figure 31) and thus the observed differences in the pH titration curves (Figure 32).

In addition to a decrease in pK_a, fluorination could bring about other behavorial changes to the chromophores. It causes the differences in the spectral shifts between native and fluorinated *av*GFPs, and therefore the reduced or enhanced intensities in the substituted EGFP and EYFP chromophores (Figures 31, 22, Tables 7 and 8), which can be regarded as intrinsic features of the fluorinated chromophores. The red-shifted spectra of (3-F)Tyr containing proteins can be attributed to their enhanced preference for the anionic $\mathbf{A}^-$ chromophore state. One might assume the same effect for (2-F)Tyr-containing proteins, albeit to a lower extent. Contrary to our expectations, spectra of these proteins are blue-shifted, and there should be other explanations for these experimental evidences. In addition to the inductive electron-withdrawing interactions, the resonance effect induced by fluorine on the phenolic group may well affect the electronic properties of the aromatic

ring in Cro66. Furthermore, the electron-donating resonance effect of the hydroxyl group influences the aromatic ring of Cro66. Such resonance effects of both groups are reflected in the distribution of their lone pairs to the ring carbons. These effects combined, are responsible for the significantly altered biophysical properties of (3-F)Tyr relative to those of Tyr. It could well be possible that the combination or a different balance of these effects in (2-F)Tyr-chromophores is responsible for their intrinsic spectral blue-shifts. Besides, it is likely that inductive electron withdrawing effects in (2-F)Tyr-Cro66 outweighs resonance effects to lower extents than in (3-F)Tyr-Cro66. Thus, it should not be surprising that the biophysical properties of (2-F)Tyr and related chromophores are much more similar to their parent proteins. In general, H→F atom exchange in the Cro66 moiety not only results in inverted polarities and thus increased charge separation, but also might influence electron delocalisations as well as orientations of transition and ground state dipoles in the chromophore ring. [87]

6.3.1.5 Fluorinated tyrosines in the crystal structures of (2-F)Tyr-EGFP, and (3-F)Tyr-EGFP

It is expected that the replacement of *ortho*- or *meta*-hydrogens in Tyr would produce minor steric perturbations in local environments. Not surprisingly, the crystal structures at neutral pH of both proteins have the same overall structure as native EGFP. This was also found in the first crystal structure of a protein containing (3-F)Tyr instead of tyrosine side chains, i.e. in glutathione transferase M1-1 globally substituted with (3-F)Tyr.[128] These studies convincingly demonstrated that the fluorine atoms of (3-F)Tyr residues in proteins might interact with crystallographically well defined water molecules in the vicinity, other neighboring residues or residues from symmetry-related molecules. It also brought to light the crystallographic evidence suggesting that the fluorine atoms are capable to act as hydrogen bond acceptors from both –NH and –OH donors. In addition, fluorine participates in various unusual interactions that can be derived from crystallographic distances. The similar holds true in the crystal structure of (3-F)Tyr-EGFP at a resolution of 2.1 Å.[116] The closer inspection of this structure brought to light libraries of various crystallographically distinct micro-environments where the different types of interactions of fluorine atoms can be mapped. Out of the eleven residues in EGFP, Tyr237 is not crystallographically defined, while Tyr182 is fully exposed and no electron density is observed for the fluorine atom due to the high B-factor (44.20 $Å^2$). Partially exposed Tyr143 has both conformers but has no detectable distances that might indicate its involvement in a particular interaction. Buried residues Tyr74, 106 and 145 are tightly

packed in the protein interior and their fluorine exibits only a single conformation. This is in agreement with the common assumption that ring-flipping of aromatic residues correlates with its burial in the protein molecular cores. In addition, the fluorine of Tyr145 has numerous contacts to neighbouring residues, such as the water molecule S14 (2.99 Å), which is also in the hydrogen bond distance to fluorine as well as a hydroxyl group (2.76 Å) of (3-F)Tyr145.[116] The fluorine of Tyr106 is 2.95 Å distant from the ring carbons of Phe130. In contrast, Tyr92 although buried exhibits both conformations, one of them is at interaction distance of 2.79 Å with an amide oxygen of Phe84, and is probably in contact with the ring carbon of Phe84 through cation-π interaction (3.71 Å) (see Figure 33).

Contrary to logical assumptions, the solvent accessibility of one aromatic residue does not automatically mean that it is prone to flipping. For example, there are surface exposed Tyr residues in glutathione transferase M1-1 where flipping is suppressed through the lattice contacts with neighboring molecules in the crystalline state.[128] A pair of closely placed surface exposed Tyr residues of (3-F)Tyr-EGFP at positions of 151 (two conformers) and 200 (one conformer) are inclined with an angle of about 65°. The fluorine atom of one conformer of the flipping residue Tyr151 is 3.21 and 3.53 Å distant from ring carbon atoms of Tyr200. The single conformer of (3-F)Tyr200 has its fluorine atom quite distant from both the fluorine atoms belonging to the neighbouring Tyr151. Detailed conformational analyses of fluorinated Tyr-side chains are provided in the next section.

On the other hand, our crystal structure of (2-F)Tyr-EGFP at 2.2 Å as the first known protein structure substituted with this analogue, surprisingly shows that all (2-F)Tyr-residues are present in only one conformeric state. While the (3-F)Tyr92 (B-factor: 20.36 $Å^2$) is characterised with two conformers of its fluorine at the *meta*-position, the fluorine atom in the *ortho*-position in (2-F)Tyr92 (B-factor: 16.98 $Å^2$) is present in only one, well defined conformeric state (Figure 33). Surface exposed Tyr-residues at positions 39 and 182 exhibit no electron density for *ortho*-flourine atoms, contrary to a pair of adjacent surface exposed residues (Tyr151 and Tyr200) which might be involved in mutual interactions as well as in interactions with some neighbouring residues. At first glance, it seems plausible that the replacement of an *ortho*-hydrogen by fluorine would induce comparatively less steric perturbation than that in the *meta*-position. Thus, the feature of the exclusive presence of only one conformation in the crystal structure of (2-F)Tyr-EGFP might be a result of the substituted position in the phenol ring (*ortho* or *meta*) or of structural perturbations or a combination of both.

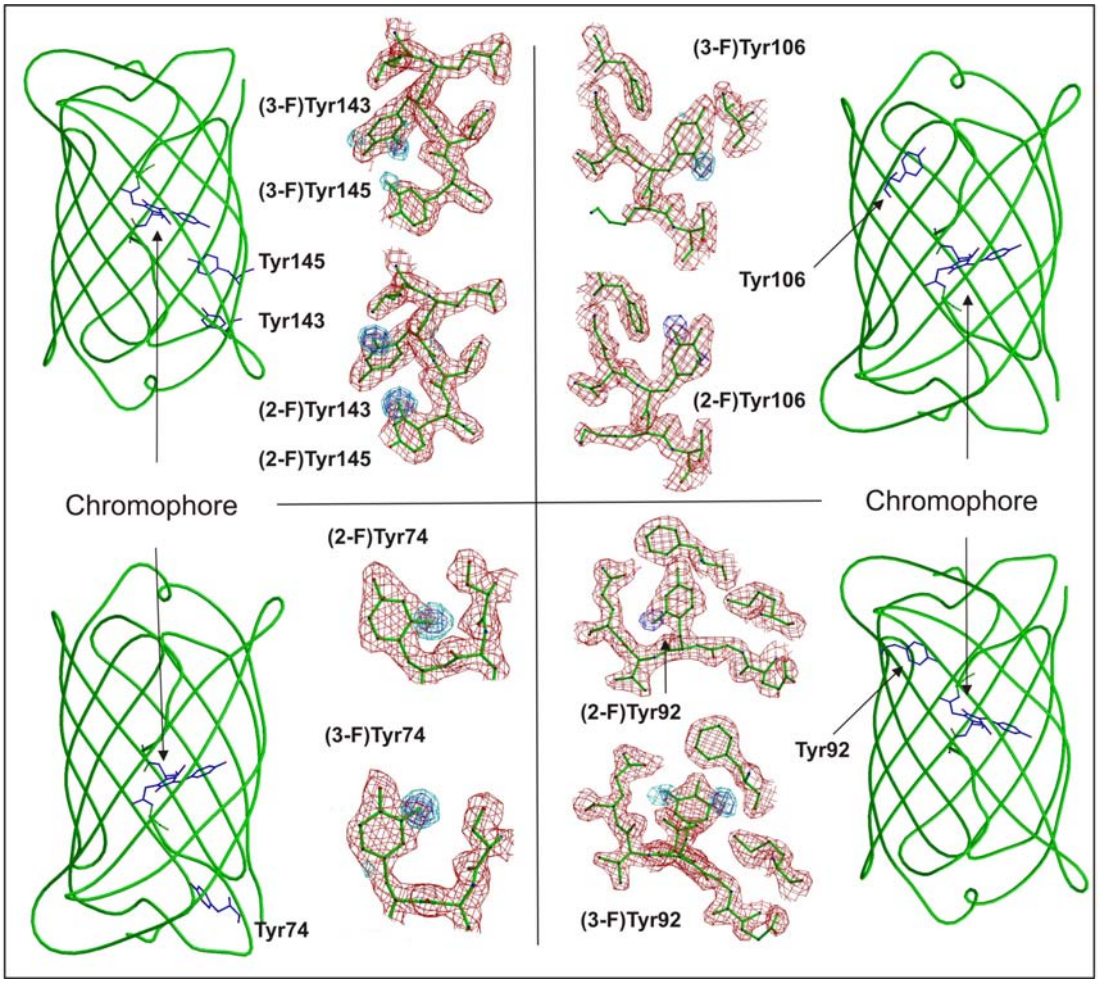


Figure 33. Structural microenviroments of crystallographically well defined Fluoro-Tyr residues in EGFP. Detailed description is provided in the text. The difference Fourier maps (blue, Fo-Fc; contouring levels: 3.0 σ) are superimposed on its continuous electron density (red, 2Fo-Fc; contouring levels 1 σ).

6.3.1.6 Molecular microenviroments of fluorinated Tyr-residues

Out of the eleven residues in EGFP, Tyr237 is not crystallographically defined, while Tyr182 is fully exposed and no electron density is observed for the fluorine atom due to the high B-factor. Diversity of the novel local microenvironments introduced by fluorine atoms either in the *meta*-, or *ortho*- positions in the tyrosyl moieties are demonstrated by the examples described below. For example, partially exposed Tyr143 assumes both conformeric states but no detectable distances that might indicate its involvement in a particular interaction. Buried residues Tyr74, Tyr106 and Tyr145 (Figure 33) are tightly packed in the protein interior and their fluorine assume only one conformation as fluorine electron densities of (2-F)Tyr residues are well defined.

A similar situation is observed in some of the (3-F)Tyr residues. However, most of the electron densities of (3-F)Tyr-residues appear to be distributed (usually in a ratio of 40:60) between the two conformeric states which is obvious from Figure 33. For example, the

(2-F)Tyr39 is solvent exposed and has contact with Thr38. The F atom is forming an H-bond with the main chain oxygen of Thr38. The electron density of the fluorine atom in the surface exposed residue of (3-F)Tyr39 is distributed among two conformeric states. The fluorine atom in the minor conformation is involved in repulsive interaction with the terminal oxygen atoms of the neighboring Asp36 (distances 3.28 Å and 3.27 Å).

The (2-F)Tyr74 is buried inside the molecule and has extensive interactions with the nearby residues. Ala226, Asp82 and His81 form hydrophobic interactions with (2-F)Tyr74, whereas the phenolate is in a H-bond distance with the imidazole group of His199. This in turn forms another H-bond with the main chain of Asp197. The F-atom of (2-F)Tyr74 is in crystallographic H-bond distance with the main chain oxygen of Met78. This interaction is probably repulsive in nature and contributes to a destabilisation of the protein variant (see section 6.3.1.3). The difference electron density map of (3-F)Tyr74 indicates a single conformeric state of the fluorine in crystal structures (Figure 33). Since this residue is buried inside the protein core, it has hydrophobic interactions with the neighboring residues Leu201 and Ala226. Side chain atoms of Asp82, His81, and His199 are also within contact distances. The hydroxyl group forms an H-bond with the imidizole N of His199 which in turn forms a salt-bridge with the carboxylate of Asp197.

The (2-F)Tyr92 is completely buried inside the core of EGFP. It has hydrophobic interactions with Met88, Ile188, Ala87, and Phe84. Nonetheless, its fluorination at the *meta*-position results in an electron density distribution among the two conformeric states. The fluorine of (3-F)Tyr92 present in a 'minor' conformation, is 2.79 Å distant from the amide oxygen of Phe84, and has contact with the ring carbon of Phe84 (3.71 Å) probably through cation-π interactions. Other 'major' (3-F)Tyr92 conformations have extensive interactions with Met88 and Ile188 while its F atom interacts with the main chain of Phe84.

Conversely, all (2-F)Tyr-residues of (2-F)Tyr-EGFP are present exclusively in one conformeric state. The (2-F)Tyr106 side chain is buried and is involved in hydrophobic interactions with Phe100, Phe130 and Leu125. The hydroxyl group of (2-F)Tyr106 forms a hydrogen bond with the side chain of Thr59. On the other hand, the difference electron density for the fluorine atom of the (3-F)Tyr106 is presented in a single conformational state. This residue, buried inside the core, has hydrophobic contacts with Phe130, Phe100, Val22 and Leu125. Its hydroxyl group forms a hydrogen bond with the Thr59 side chain. Interestingly the fluorine atom itself does not exhibit any interaction distances with the surrounding atoms. The fluorine of Tyr106 is 2.95 Å distant from the ring carbons of Phe130.

The (2-F)Tyr143 is partially solvent exposed; however the hydroxyl group is surrounded by Lys209 and Leu207. Its fluorine atom does not show any contact distances with its molecular neighbourhood. Curiously, the (3-F)Tyr143 in spite of its partial solvent exposure assumes both conformeric states in its X-ray structure. In both the 'minor' and 'major' conformation the fluorine atoms show no detectable distances that might indicate its involvement in a particular interaction(s). In the major conformation the fluorine atom is completely solvent exposed.

The side chain of (2-F)Tyr145 is buried inside the molecule with a network of interactions forming Van der Waals contacts with Val61, His148, His169 and Ser205. The fluorine atom is closely placed to the main chain of Asn144 (crystallographic H bond distance). Similarly, the side chain of (3-F)Tyr145 is completely buried and has hydrophobic contacts with Ser205, His169 and Val61. For example, the fluorine of (3-F)Tyr145 is within the contact range of the amide oxygen of Pro58 (2.54 Å). The ring nitrogen of His169 is also quite close to the fluorine (3.27 Å) of (3-F)Tyr145 whereas, the water molecule S14 is close both to the fluorine atom (2.99 Å) as well as to a hydroxyl group (2.76 Å) of (3-F)Tyr145.

6.3.1.7 Fluorination of the Chromophore Environment in EYFP

By virtue of its sequence mutation Thr203Tyr, EYFP is known to be the most red shifted mutant designed by conventional methods ($\lambda_{max(em)}$ = 527 nm). Fluorination of Tyr203 in EYFP offers a fairly good model to study effects caused by fluorinations of its micro-environment on the spectral functions of the chromophore. Fluorinated EYFPs (having the chromophore in stacking interaction with Tyr203) offers a framework to study the nature of lowering of pK_a upon fluorination. It is reasonable to expect that inductive effects exerted by the fluorine nucleus in the chromophore or fluorine atoms from its microenvironments are responsible for such behaviour. Thus, the fluorination of Tyr203 is probably responsible for the enhanced blue and red shifts in the EYFP variants (Figure 30). The Tyr-side chain at position 203 stacks next to the chromophore to induce additional polarization in its micro-environment; π-π interactions are expected to reduce the excited state energy and enhance both excitation and emission wavelengths.[105] Undoubtedly, all these interactions are complicated and are in addition influenced by the introduction of the fluorine nucleus. (3-F)Tyr203 and (2-F)Tyr203 are mutually isosteric, and the differences observed in the related protein mutants may rely not just on different physicochemical properties but also on different stereochemical positions of the fluorine atoms relative to the chromophore protonated/deprotonated phenolic group.

As shown in Figure 34, a closer inspection of the chromophore of (3-F)TyrEGFP in the cavity with Thr203 revealed that there are no steric hindrances in the accommodation of the fluorine atom (purple) in the cavity, and even for the (3-F)Tyr chromophore flipping. The model of (3-F)Tyr-EYFP was constructed by the replacement of a hydrogen by a fluorine at Tyr203 and Tyr66 in wt-EYFP, and by energy minimization using the standard modeling procedures. Although fluorine atoms can easily be accommodated in the chromophore cavity of EYFP, it is obvious that the flipping of the fluorinated chromophore should encounter steric hindrance even in the native EYFP structure. The coplanar position of both Tyr203 and (3-F)Tyr203 that is about 4 Å distant from the centre of chromophore should be the main barrier for possible flipping motions.

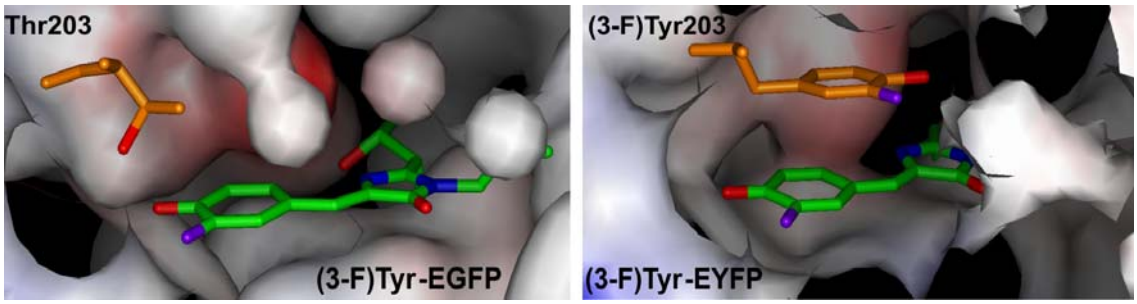


Figure 34. Fluorinated microenvironments in EGFP and EYFP. In the (3-F)Tyr-EGFP chromophore microenvironment with Thr203 (shown as a stick representation) there is no sterical hindrances in accommodating the fluorine atom (purple) in the cavity, and even for possible (3-F)Tyr-chromophore flipping. The (3-F)Tyr-EYFP model constructed by the replacement of a hydrogen to a fluorine at Tyr203 and Tyr66 in EYFP, indicates that the fluorine atoms can easily be accommodated into the EYFP chromophore cavity. The distance between coplanarly positioned Tyr203 and (3-F)Tyr203 in both structures is about 3.5 Å.[56]

6.3.1.8 Y203F-EGFP spectral features in native and fluorinated proteins

No significant shifts in absorbance profiles have been observed in case of the Phe-flourinated variants of EYFP in comparison to the wt protein. However, both the absorbance and emission maxima of Y203F-EYFP are blue shifted (λ_{abs}= 511 nm) λ_{max} = 521 nm). On the other hand, absorbance as well as fluorescence emission of partially fluorinated phenylalanine variants of this mutant exhibit spectral features that are for the most part indistinguishable compared with the parent EYFP (data not shown). Since protein labelling was only partial (full labelling was never achieved although different fermentation strategies were employed), these data should be considered with reserve. Thus, at this stage, it is not possible to use Y203F-EYFP as a model to study the influence of microenvironment fluorinations on the spectral properties of EYFP.

6.3.1.9 Chromophore fluorination.

Global substitution of Tyr side chains in EGFP with both (2-F)Tyr and (3-F)Tyr resulted in the direct incorporation of the fluorine nucleus in the chromophore. The crystal structures of fluorinated chromophores indicate that the same effects observed in Tyr residues can be generalised for them as well. The fluorine in (3-F)Tyr-Cro66 of (3-F)Tyr-EGFP is present in two conformers with slightly different crystallographic occupancy (Figure 34). Conversely, the fluorine in (2-F)Tyr-Cro66 of (2-F)Tyr-EGFP is characterised by a well defined different electron density at the H→F replacement site. A closer inspection of both chromophores and their surrounding residues in terms of crystallographic B-factors indicate a rather rigid internal architecture in the crystalline state. In fact, the average B-factors for atoms in Cro66 in both structures are about 15 $Å^2$ while the B factors of the surrounding residues are from 5 to 25 $Å^2$.[116] In recent studies, it was demonstrated that the chromophore cavity affords some degree of rotational freedom thereby accommodating even some non-planar conformations. Indeed, as shown in Figure 34, there is no steric hindrance to accommodate the fluorine atom in the *meta*-position, and even flipping should be possible. The single conformer of the (2-F)Tyr-Cro66 in (2-F)Tyr-EGFP can be plausibly explained by taking into account sterical considerations. Fluorine, when fitted in the "second" *ortho*-position of (2-F)Tyr-Cro66 induces a direct clash with the hydrogen of the ring nitrogen (i.e. there is ca. 2.1 Å distance between the nitrogen and fluorine). This might explain the single conformer not only in (2-F)Tyr-Cro66, but also in the rest of the (2-F)Tyr-residues in (2-F)Tyr-EGFP. For example, when fluorine from the *ortho*-position in (2-F)Tyr106 is fitted to an alternative *ortho*-position, it causes an electronic repulsion with the amide oxygen of N105. On the other hand, preferred interactions of *ortho* fluorine with the neighboring residues might be the cause for a single conformeric state of (2-F)Ty39 and (2-F)Tyr92, respectively (Figure 33). The X-ray structures presented in this study reflect relatively static properties of the *av*GFP chromophores, and offer a good starting point to study chromophore flipping.

In general, fluorinated residues involved in multiple interactions (with solvent water, neighbouring residues, etc) are "trapped" in only one conformational state. Indeed, Tyr145, as shown in Figure 33 exists only as a single conformer. On the other hand, the (3-F)Tyr-containing chromophore exhibits no additional contacts when compared with those of native EGFP with the exception of its fluorine atom. The hydroxyl groups of native and fluorinated Cro66 conformers interact directly with Thr203 and His148 through hydrogen bonds, and are in hydrophobic contacts with Thr62, Val150 and Phe165. The *meta*-fluorine nucleus creates a set of two crystalographically different micro-environments, in which it

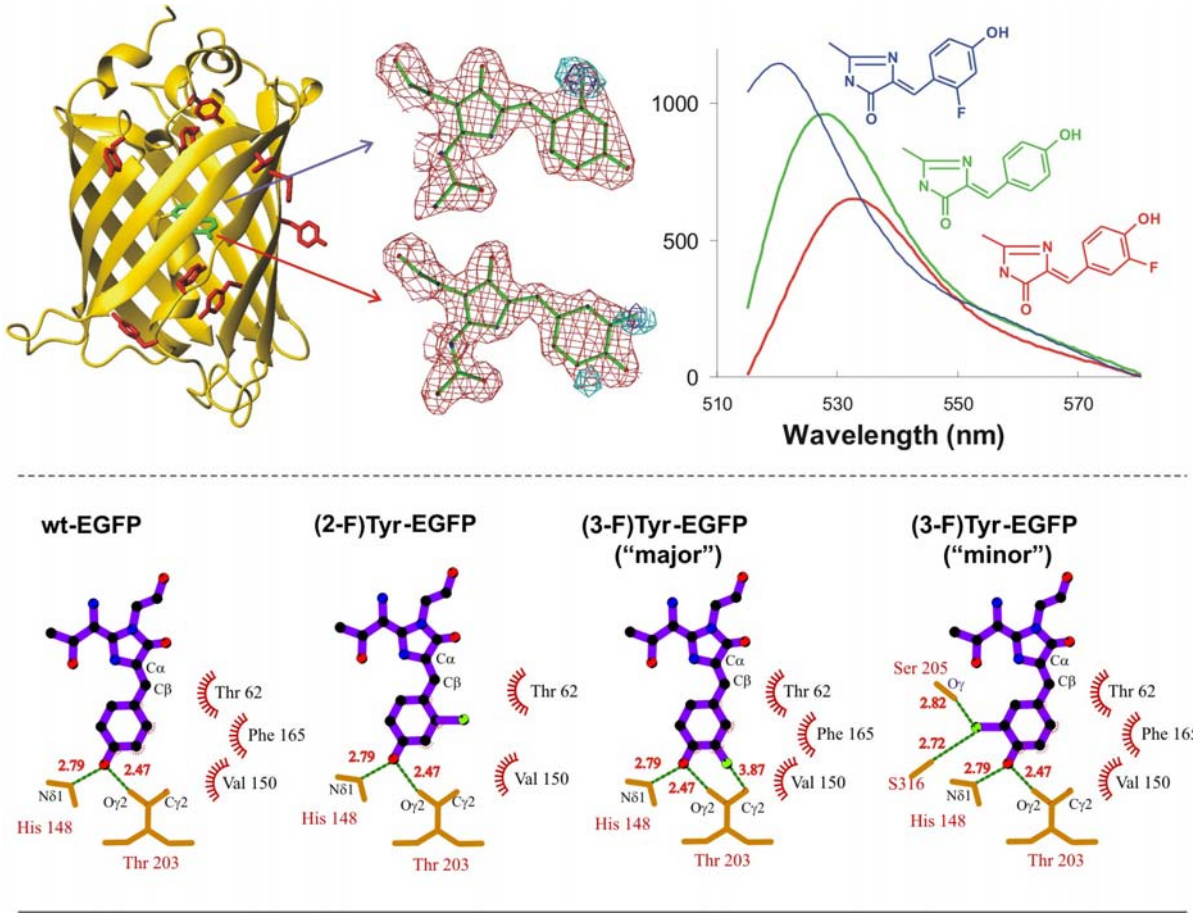


Figure 35. Structural and spectral responses of EGFP brought about be a fluorination of its chromopore in the *ortho*-, and *meta*-positions. Global replacement of Tyr-residues in both proteins by (2-F)Tyr, and (3-F)Tyr respectively, results in concomitant fluorination of the chromophores. Upper part-left: The general architecture of the *av*GFPs is based on the 11 stranded β-barrels that rigidly hold a chromophore anchored at a co-axial helix within the protein core. Upper part-middle: Difference Fourier maps of (3-F)Tyr-chromophore in (3-F)Tyr-EGFP superimposed on its continuous electron density revealed the existence of two conformers ("major" and "minor"). In contrast, X-ray crystallographic structure analysis of (2-F)Tyr-EGFP mapped a single conformer state of (2-F)-Tyr-containing chromophore. Upper part-right: General fluorescence behavior (i.e. red-, and blue-shifting of emission maxima) of fluorinated chromophores. Lower part: Schematic representation of the protein-matrix interactions in native and fluorinated EGFP variants. Note novel interactions in the chromophore microenvironment introduced by the fluorine atoms (detailed description is available in the text).

participates in unusual interactions that can be derived from crystallographic distances such as Cγ2 of Thr203 and the fluorine of conformer 1 (distance: 3.87 Å) in Figure 35. The diversity of the fluorine interactions is also illustrated in the conformer 2 of (3-F)Tyr-Cro66, which is in hydrogen-bond contact (2.82 Å) with the hydroxyl group of Ser205 and a conserved solvent water (S16, 2.74 Å). In contrast, there are no interaction distances of *ortho*-fluorine in the (2-F)Tyr-containing chromophore with its micro-environment, and even the number of its hydrophobic contacts is reduced (Figure 35). Although crystal structures of fluorinated EYFP-variants are not available, it is reasonable to suppose that

most of the fluorotyrosyl-residues behave in the same way as those in fluorinated EGFPs. However, the most interesting question addresses the behaviour of the fluorinated Cro66 and Tyr203. Modelling and energy minimisation procedures show that the planar Cro66 has no rotational freedom for flipping in the same manner as (3-F)Tyr-Cro66 EGFP despite the fairly large central cavity in the protein core. The main barrier for any possible flipping is the nearby positioning of Tyr203 which is in the stacking interaction with the chromophore of EYFP (Figure 34).

6.3.1.10 Fluorination of the Tyrosine residues in *ds*Red

The fluorescence emission spectrum of *ds*Red is shifted to longer wavelengths by almost 75 nanometers (λ_{max} = 585 nm) compared to the enhanced green fluorescent protein (EGFP). It was therefore reasonable to expect that global fluorination (that includes the chromophoric Tyr-residues) would influence spectral properties in a similar manner as EGFP and EYFP. Unfortunately, only partial labelling could be achieved, as shown in Figure 36 with marginal red and blue shifts in fluorescence. Optimal fermentation conditions for complete labelling followed by careful spectral and structural analyses should be done in future as performed with fluorinated EGFPs and EYFPs. Fluorination of *ds*Red might also be useful in understanding important issues like extinction coefficients, quantum yields, photobleaching, oligomerisation and maturation of the fluorinated chromophore.

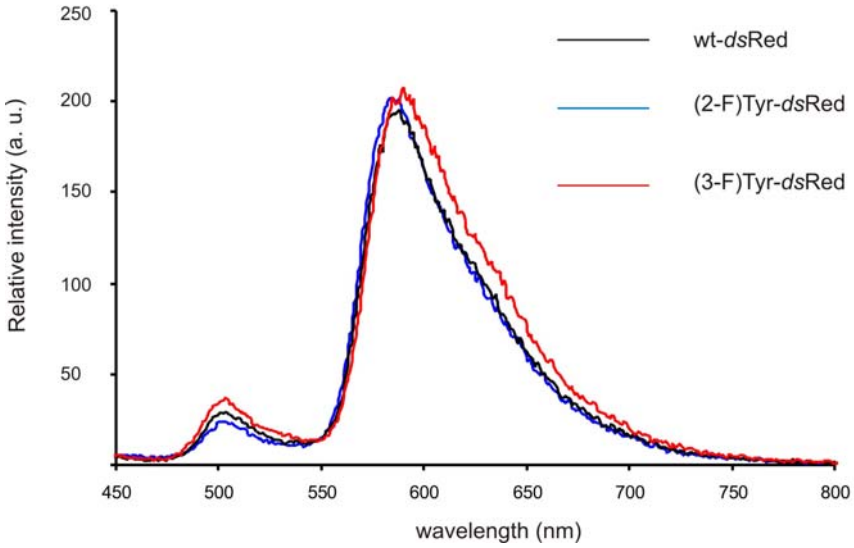


Figure 36. Spectral properties of native and fluorinated *ds*Red. The fluorescene emission maximum of wt-*ds*Red is 585 nm. The protein is partially substituted with (2-F)Tyr and is slightly blue (λ_{max} = 580 nm) whereas partial random incorporation of (3-F)Tyr in the structure of *ds*Red results in a red shift of 7 nm(λ_{max} = 592 nm) (see also Table 8).

6.3.1.11 Electrophoretic behaviour of fluorinated proteins – gel shift

Protein electrophoresis is a technique used to separate and purify proteins that differ in size, charge or conformation. As such, it is one of the most widely-used techniques in biochemistry and molecular biology. When charged molecules are placed in an electric field, they migrate toward either the positive (anode) or negative (cathode) pole according to their charge. Proteins are electrophoresed within a matrix or "gel" which is immersed in an electrophoresis buffer that provides ions to carry a current and buffer to maintain the pH at a relatively constant value. Polyacrylamide is a cross-linked polymer of acrylamide. The length of the polymer chains is dictated by the concentration of acrylamide used, which is typically between 3.5 and 20%. In SDS polyacrylamide gel electrophoresis (SDS-PAGE) separations, migration is determined not by intrinsic electric charge of polypeptides but by molecular weight. Sodium dodecylsulphate (SDS) is an anionic detergent that denatures proteins by wrapping the hydrophobic tail around the polypeptide backbone. For almost all proteins, SDS binds at a ratio of approximately 1.4 g SDS per gram of protein, thus conferring a net negative charge to the polypeptide in proportion to its length. The SDS also disrupts hydrogen bonds, blocks hydrophobic interactions, and substantially unfolds the protein molecules, minimizing differences in molecular form by eliminating tertiary and secondary structures. Disulphide-containing proteins can be totally unfolded when a reducing agent such as dithiothreitol (DTT) is employed. DTT cleaves all disulphide bonds. The SDS-denatured and reduced polypeptides are flexible chains with uniform negative charge per unit length. Because molecular weight is essentially a linear function of peptide chain length, in sieving gels the proteins separate by molecular weight.[112]

In a typical experimental set up a large-pore gel called a stacking gel is layered on top of a separating (resolving) gel. The two gel layers are each made with a different buffer, and the tank buffers differ from the gel buffers. When electrophoresis is started, the ions and the proteins begin migrating into the stacking gel. The denatured proteins concentrate in a very thin zone, called the stack and continue to migrate in the stack until they reach the separating gel which can resolve complex protein mixtures into hundreds of bands. The position of a protein along the separation lane gives a good approximation of its size, and, after staining, the band intensity is a rough indicator of the amount present in the sample. However, it was observed very early on that proteins containing noncanonical amino acids show slightly abnormal electrophoretic behaviour, i.e. their electrophoretic mobility do not correspond strictly to their molecular masses. For example, canavanine-containing viral precursor proteins[129] and more recently trifluoroleucine-containing dihydrofolate reductase[130] as well as annexin V and azurin containing fluorinated tyrosines, show

different electrophoretic mobilities[67]. This feature was observed in EGFP and EYFP with fluorinated tyrosines as well, as shown in the gel of Figure 37. Although there is still no satisfactory explanation for such behaviour, it is obvious that the changed physicochemical properties such as increased/decreased hydrophobicity, suppressed ionisation capacity and altered electrophoretic mobility are due to the presence of the amino acid analogues in the structures of these proteins. In general, the difference in the electrophoretic mobility, i.e. gel-shift, can be observed in both denaturing ad native gels. These phenomena are also well documented for canavanine-, trifluoro-, and hexafluoro-leucine-containing proteins.[130] Such experiments also provide an indirect evidence for the incorporation of an amino acid analogue into proteins.

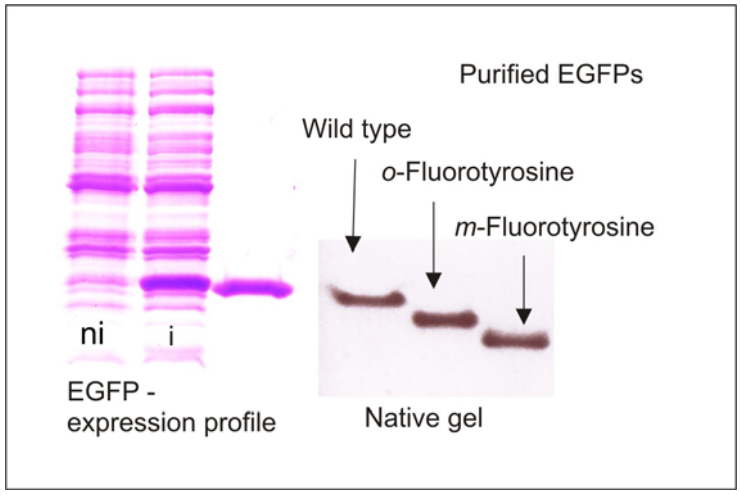


Figure 37. Gel-shift experiment' with the expression profile of native green fluorescent protein (GFP, left) in *E. coli* and the electrophoretic mobility of its 2-fluorotyrosine and 3-fluorotyrosine-containing variants in 12% polyacrylamide gels (right).

6.3.1.12 Affecting the growth of tumor cells by Tyr-fluorinated proteins

The most widely used approach for the intracellular delivery of drugs in proteins is to use lipid-mediated delivery systems. In the initial stages of this study, lipofectamin, a commercially available transfection agent was used (in non toxic concentrations) in order to deliver both (3-F)Tyr-EGFP and EGFP into the cell line MCF10a. These experiments show that the fluorinated protein at an extracellular concentration of 0.05 mM does indeed inhibit cellular growth as seen by confocal microscopy (Figure 38). In order to reproduce these initial experiments other cell lines like HEK293, SKBR3, BT474 and K562 were tested. Proteins as delivery vehicles (β-galactosidase, annexin V) and protein transduction kits (BioPorter, Chariot) were included. In addition (3-F)Tyr-containing β-galactosidase

with a cationic tag was generated (N. Budisa, P. Birle, T. Krywcun; unpublished material); this protein was specifically coupled to an immunotoxin in the laboratory of H. Lilie in order to study delivery in tumour cells (MCF-7). However, these experiments proved that the results of the initial experiments (which were also reported earlier[90]) were not reproducible. Similar findings with the system of immuonotoxin-mediated delivery of fluorinated β-galactosidase were reported as well. In addition there are other technical difficulties in performing these experiments. First, all transduction substances are not suitable for protein delivery. For example, β-galactosidase can be successfully delivered into cells by BioPorter or Chariot, whereas EGFP and annexin V cannot be transduced by using these kits. Second, it is difficult to quantify the amount of intracellular protein. Third, it is even more difficult to determine which amount of protein is intracellulary degraded. There are no recent studies dealing with mechanisms of intracellular degradation of intracellular proteins. On the other hand there are ample data about the processivity of substituted proteins in the context of viral infection or tumour tissues from the sixties and seventies.[43] For example, Jacobson and Baltimore discovered the mechanism of the proteolytic cleavage of a viral protein precursor by studying inhibition effects upon its substitution with amino acid analogues.[131] Other examples include the proteolytic cleavage of the precursor forms of some hormones like prosomatostatin or proinsulin whose cleavage was inhibited by specific analogues of arginine and lysine.[132]

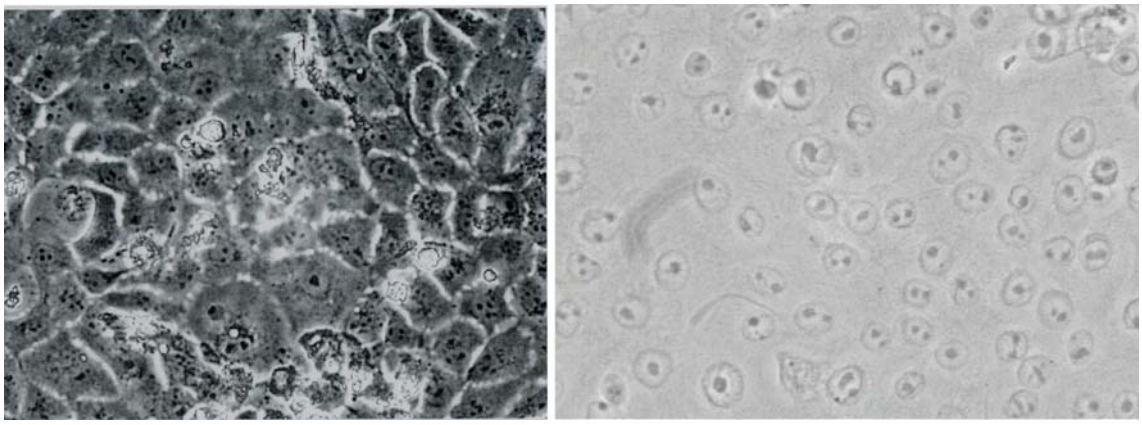


Figure 38. Influence of wt-EGFP (left) and (3-F)Tyr-EGFP (right) after liposome-mediated protein delivery on the growth (for 48 h after delivery) of breast epithelial cell line MCF10A. Note that almost all cells incubated for 48 h with the fluorinated EGFP are apoptotic, whereas cells with wt-EFGP are viable.

Although the fluorinated protein-delivery experiments could not be reproduced, studies of the influence of fluorinated Tyr on cell culture were attempted as well. Surprisingly, it could be demonstrated that large extracellular amounts (over 50 μM) of L-forms of (3-F)Tyr and (2-F)Tyr in minimal media are necessary to efficiently inhibit the growth of

the examined tumour cell lines (MCF10a, HEK293, BT474 and K562). Similar observations were made by the Lilie Group in Halle by measuring the cellular uptake of fluorinated Phe and Tyr-analogues. It was reported that only the L-form of (4-F)Phe has a significant IC50 value of about 10 μM.[123]

Taken together, these data strongly indicate that the experimental proof for the concept of specific intoxication of the tumour cells by using proteins as a carrier of growth inhibiting substances is still missing.

In this context, two main problems remain to be clarified:

First, the carrier protein upon its specific delivery into target issues must be rapidly degraded in order to release the desired noncanonical amino acids in its free form. Second, upon accumulation of the fluorinated amino acids into the cytoplasm (see Section 6.2.1.5) there must be a way to bring them into metabolic circuits in order to generate cytotoxic substances like fluorocitrate and fluoroacetate as shown in Scheme 3 (Section 4.5.3.1) of the Introduction.

6.4 Trifluoromethionine incorporation into 2M-EGFP and annexin V

6.4.1 General considerations

It was generally observed that trifluoro-amino acids are much more toxic for the living cells than their di- or mono-fluoro counterparts.[71] By using lysozyme from bacteriophage λ as a model protein, Vaughan et al. succeeded in obtaining an almost 100% replacement of the Met residues by 5',5'-difluoromethionine (DFM), while with TFM the substitution level was only about 30 %.[70; 91] In contrast, monofluorinated amino acids in most cases are quantitatively incorporated and expressed at identical levels as their parent proteins. For example, *Lactobacillus casei* dihydrofolate reductase expressed in *E. coli* was fully substituted with (2S,4S)-5-fluoroleucine to study conformational changes of the protein core upon ligand binding with ^{19}F-NMR spectroscopy.[133]

In contrast to other hydrophobic residues like Leu, Ile and Val that have a great natural abundance, Met-residues that occur more rarely might be better substitution targets since they are expected to introduce less sterical perturbations in the protein core. Up to now, only Honek and co-workers have provided experimental reports about the possibility of incorporation of TFM and DFM into recombinant proteins.[70; 91] The authors demonstrated that Met-replacements with TFM into phage lysozyme at positions 1, 14 and 107 were possible to levels of 31 – 70%.

6.4.2 2M-EGFP: A mutant designed for TFM incorporation in *av*GFP

The Met→TFM replacement was attempted by using a mutant of EGFP (2M-EGFP, 27676 Da, Met153Thr/Met233Lys/Met78Leu/Met88Leu) that contains only two Met-side chains: Met218 that is buried and the N-terminal Met1 that is solvent-exposed (Figure 39). This mutant shares the general architecture of EGFP and replacement of both its Met-residues does not affect the spectral properties of the chromophore. A controllable, robust and efficient plasmid-directed T5-based expression system along with suitable Met-auxotrophic *E. coli* host strains (see Materials and methods) has been used for protein expression. The method of selective pressure incorporation (SPI) is based on the traditional use of auxotrophic strains and exploits the absence of absolute substrate specificity of aminoacyl-tRNA synthetases, which are crucial enzymes in the genetic code interpretation (see Introduction) It is well known that TFM is a substrate for MetRS (methionyl-tRNA synthetase). Thus it can be efficiently translated during *in vivo* protein synthesis into target protein sequences. As in all other experiments, fermentation procedures were carried out by limiting the concentration of native amino acids (i.e. substitution targets) until their depletion from the defined synthetic growth medium. At this point, the TFM was added and protein synthesis induced. Alternatively, the procedure of media shift was employed during fermentation (see Materials and methods).

Reproduction of these studies was attempted by using 2M-EGFP as a model protein for the standard protocols of SPI. These experiments yielded partially labeled protein (where labeling never exceeded 10-15 %), as shown by amino acid analyses. However, the ESI-MS analysis of the same samples revealed three species of functionally active (i.e. fluorescent) proteins: native 2M-EGFP (2 Met side chains), a species with a molecular mass of 22728 Da which correspond to the presence of 1 Met and 1 TFM in the protein sequence (probably statistically distributed), and 2TFM-EGFP species (27787 Da) having both their Met-residues replaced by TFM (Figure 39). None of the experiments performed yielded a 2M-EGFP protein where their fluorinated species dominates over non-fluorinated ones as analyzed by mass spectrometry. The reason for such behavior is probably the increased steric bulk (about 15%) of noncanonical TFM (84.4 $Å^3$) compared to that of canonical Met (71.2 $Å^3$)[65]

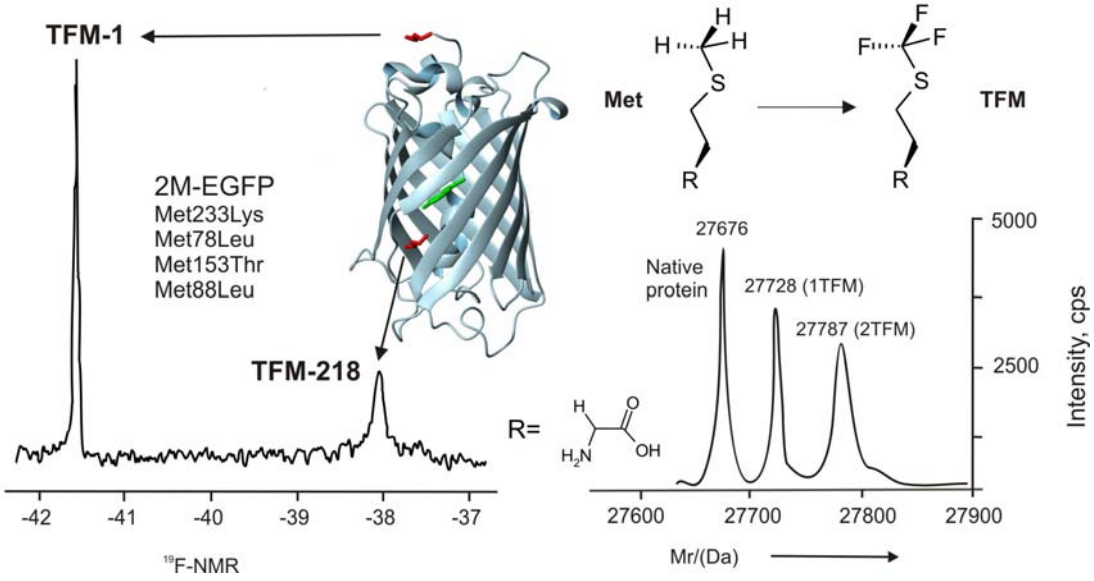


Figure 39. The Met→TFM replacement in the 2M-EGFP. (*Upper Right*) Chemical structures of L-Met and L-TFM. (*Left*) Solvent exposure of TFM-residues in 2M-EGFP protein as determined by one-dimensional ^{19}F-NMR. The profile of ^{19}F-chemical shift reveals different environments for TFM-residues in 2M-EGFP as discussed in the text. (*Right*) ESI-MS analysis of Met→TFM replacement levels. The most prominent feature of native 2M-EGFP mass spectra is the dominance of unprocessed (i.e. N-terminal Met-containing) protein species (27676 Da). Protein species with both TFM residues (27787 Da) and species with one Met and one TFM (27728 Da) appears in the spectra with approximately similar intensities.

It should be kept in mind that the presented mass-spectrometric profiles do not reflect the quantitative incorporation of TFM in 2M-EGFP. Since the content of TFM is less than 30 % in these samples (according to the amino acid analyses), the fluorine atoms in the putative electron densities of the crystals of 2M-EGFP would not be "visible". Such small quantities of fluorinated molecules is easier to trace by using the more sensitive ^{19}F-NMR spectroscopy.[87] In fact, it provides a direct insight into solvent exposure, i.e. dynamic features of both TFM-residues in the 2M-EGFP structure in a similar manner as was demonstrated for the TFM-containing phage lysozyme. The surface exposed and flexible TFM1 is assigned to a narrow signal, while the relatively rigid (and buried) TFM218 is represented by a broad spectral peak in the one-dimensional ^{19}F-NMR spectrum (Figure 39).

Interestingly, a complete absence of the N-terminally processed species of 2M-EGFP in the presence of TFM in the growth media contrasts the observation that at least 10 % of the N-terminal Met-residues of wt-2M-EGFP are processed in the normal growth media (unpublished data). This indicates an efficient blockage of the N-terminal processing of Met by TFM. In many cases this initiator Met is co-translationally removed by a Met-aminopeptidase.[134] It is well known that the metabolism of methionine plays an important role in many aspects of cellular physiology. Protein synthesis initiates with Met in the

cytosol of eukaryotes, and formylmethionine (fMet) in prokaryotes and eukaryotic organelles. Removal of Met is a physiologically critical process: in certain cases the inhibition of Met-aminopeptidases arrests cell growth, thus making these enzymes an interesting target for tumour therapy.[135] In this context, the formylated TFM should be an efficient inhibitor of these enzymes, i.e. an interesting and promising substance for tumour therapy.

6.4.3 Incorporation of TFM in annexin V by co-expression with MetRS551

A full incorporation of TFM was achieved in annexin V only after the expression system was enhanced with elevated intracellular amounts of MetRS (see Figure 20 in Materials and Methods Section). Although the natural MetRS have evolved over billions of years, their rather strict substrate specificity towards cognate Met never encountered synthetic analogues like TFM in its evolutionary paths. On the other hand TFM is obviously quite a poor substrate for the activation reaction mediated by MetRS. The native genomic background activity of MetRS within the auxotrophic host bacterium was sufficient only for trace levels of incorporation, detectable in target proteins only by the extraordinarily sensitive fluorescence-, or radioactivity-based assays. However, this low substrate turnover associated with poor substrate quality can be efficiently circumvented by an increase of the intracellular concentrations of active MetRS. In this particular case, the incorporation of TFM into annexin V is effectively enhanced by a co-expression of an active fragment of MetRS (MetRS551) carried out on a compatible plasmid.

However, it was not possible to estimate the amount of incorporation of TFM in annexin V by ESI-MS, probably because fully fluorinated proteins are poorly ionisable. For that reason, the most reliable method to determine quantitatively the extent of incorporation is amino acid analysis. Indeed, as shown in Figure 40, the amount of TFM incorporation proved to be over 85%. This is the highest level of TFM incorporation into one recombinant protein reported so far. This estimation is based on the disappearance of the intensity of the Met-ninhydrin peak from amino acid composition chromatograms.

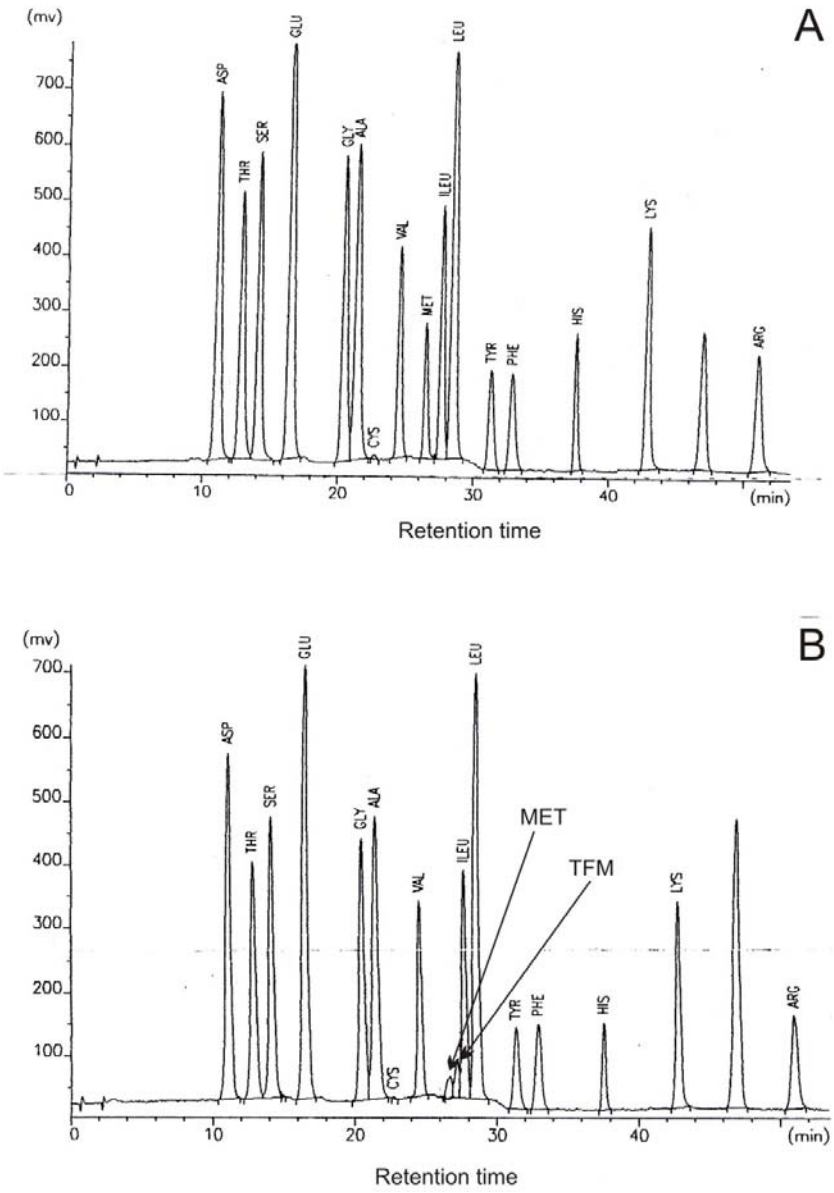


Figure 40. Chromatograms of amino acid composition of wt (left) and TFM-conatining human recombinant annexin V (right). Absorbance peaks of ninhydrine-derivatives in wt-protein, where the total contribution of Met-residues is 2.6 %, whereas in the labeled sample, the reduced Met-peak revealed only about 0.4 %. This indicates that the substitution level is higher than 85%. Note also that an additional signal (in the vicinity of the Met-peak) with slightly shifted retention time appears. This is probably due to the TFM residues that 'survived' the HCl-hydrolysis procedure, or residues that are successfully derivatized by ninhydrine. The team of Stefan Übel (Core facility, MPI) is acknowledged for this chromatogram and Waltraud Wenger for sample preparation.

6.5 Outlook and future directions

The 20 amino acid repertoire of the universal genetic code is the product of evolutionary development and is conserved for several reasons. First, specific amino acids were excluded from the code by selection as they might exhibit adverse effects on the structural integrity of the protein. Second, exclusion of other amino acid types that do not violate these rules is probably due to their unavailability or insufficient evolutionary time for cells to produce them metabolically. Third, the organic compounds containing atoms such as silicon, boron or fluorine are scarce in nature and participate only negligibly in the chemistry of living cells. Finally, the complex proteome organisation of the majority of living cells is the main barrier to an enrichment of the existing amino acid repertoire (i.e. the introduction of novel amino acids into the existing repertoire would be lethal in most cases).

Nonetheless, despite the complexity of interactions between the proteome and other metabolic components in the extant cells, it was possible to introduce a whole variety of noncanonical amino acids into single proteins without lethal effects on the cells themselves. In this context, fluorine-containing amino acids hold great promise for protein design and engineering. Well established and sophisticated fluorine synthetic organic chemistry provides a whole range of fluorinated amino acids as potential substrates for the protein translation machinery. The fluorine atoms endow these amino acids with novel and often unexpected properties which are transmitted to protein structures upon their translation into polypeptide sequences. Finally, along with new and interesting chemistries delivered into proteins, some novel und still unknown features of the translational process may be discovered.

Traditionally, mono-, di-, and tri-fluorinated noncanonical amino acids have been used extensively as markers for structural and bio-medicinal studies and have recently become interesting substrates for designing biomaterials. For example, in the context of aromatic amino acids, fluorine is capable of dramatically changing their spectral properties as shown in this study. Therefore, fluorinated derivatives of canonical amino acids are almost ideal tools for a rational re-design of natural proteins. However it is conceivable that in the near future, these experiments will be scaled from single proteins to whole proteomes. Indeed, the proteome-wide substitutions/insertions of fluorinated amino acids would yield microbes with chemically enriched proteomes. In other words, experiments with residue-specific incorporation of mono-, di-, or tri-fluorinated amino acids in single proteins may represent a solid foundation for more advanced studies and for attempts to generate cells whose proteome would have novel chemical compositions, i.e. synthetic life forms.

7 LITERATURE

1. F. H. C. Crick. Origin of genetic code. *Journal of Molecular Biology* **38**, 367- 379 (1968).

2. C. R. Woese, D. H. Dugre, S. A. Dugre, M. Kondo and W. C. Saxinger. On fundamental nature and evolution of genetic code. *Cold Spring Harbor Symposia on Quantitative Biology* **31**, 723-& (1966).

3. T. M. Sonneborn. *Degeneracy of the genetic code: extent, natureand genetic implications.*, (Academic Press, New York, 1965).

4. J. Monod. *Chance and Necessity*, (Vintage Books, New York, 1971).

5. E. Fischer. Untersuchungen über Aminosäuren, Polypeptide und Proteine. *Berichte der deutschen chemischen Gesellschaft* **39**, 530-610 (1906).

6. F. Sanger and E. O. P. Thompson. The Amino-Acid Sequence in the Glycyl Chain of Insulin. *Biochemical Journal* **52**, R3-R3 (1952).

7. M. W. Nirenberg. Deciphering the genetic code - a personal account. *Trends in Biochemical Sciences* **29**, 46-54 (2004).

8. C. Yanofsky, B. C. Carlton, J. R. Guest and D. R. Helinski. On the Colinearity of Gene Structure and Protein Structure. *Proccedings of the National Academy of Sciences USA* **51**, 226-272 (1964).

9. S. Brenner, F. Jacob and M. Meselson. An Unstable Intermediate Carrying Information from Genes to Ribosomes for Protein Synthesis. *Nature* **190**, 576-581 (1961).

10. C. Yanofsky. Tryptophan synthetase: Its charmed history. *BioEssays* **6**, 133 - 137 (1987).

11. J. D. Watson, N. H. Hopkins, J. W. Roberts, J. A. Steitz and A. M. Weiner. *Molecular Bology of the Gene*, (The Benjamin/Cummings Publishing Company Inc., Amsterdam, 1988).

12. F. H. C. Crick. On protein synthesis. *Symposium of the Society for Experimental Biology* **12**, 138-163 (1957).

13. F. H. C. Crick, L. Barnett, S. Brenner and J. Watts-Tobin. General Nature of the Genetic Code for Proteins. *Nature*, 1227-1232 (1961).

14. G. D. Rose and R. Wolfenden. Hydrogen-bonding, hydrophobicity, packing, and protein-folding. *Annual Review of Biophysics and Biomolecular Structure* **22**, 381-415 (1993).

15. F. J. R. Taylor and D. Coates. The code within codes. *BioSystems* **22**, 177-187 (1989).

16. H. Jakubowski and E. Goldman. Editing of errors in selection of amino acids for protein synthesis. *Microbiological Reviews* **56**, 412-429 (1992).

17. N. Budisa. Basic Features of the Cellular Translation Apparatus. In *Engineering the Genetic Code*, edited.(Viley, Weinheim, 2005) pp. 31-66.

18. L. Ribas de Pouplana and P. Schimmel. Aminoacylations of tRNAs: Record-keepers for the genetic code. In *Protein Synthesis and Ribosome Structure*, edited by.K. NierhausandD. N. Wilson.(Wiley-VCH, Weinheim, 2004) pp. 169-184.

19. R. A. Fersht, J. S. Shindler and W. C. Tsui. Probing the limits of protein-amino acid side chain recognition with the AARS: discrimination against Phe by tyrosyl-tRNA synthetases. *Biochemistry* **19**, 5520-5524 (1980).

20. L. Stryer. *Biochemistry*, New York, 2001).

21. S. H. Kim, F. L. Suddath, G. J. Quigley, McPherso.A, J. L. Sussman, A. H. J. Wang, N. C. Seeman and A. Rich. 3-Dimensional tertiary structure of yeast phenylalanine transfer-RNA. *Science* **185**, 435-440 (1974).

22. M. Ibba, C. Francklyn and S. Cusack. *Aminoacyl-tRNA Synthetases*, (Landes Bioscience, Austin, 2003).

23. C. Francklyn. tRNA synthetase paralogs: Evolutionary links in the transition from tRNA-dependent amino acid biosynthesis to de novo biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 9650-9652 (2003).

24. W. H. McClain. Transfer-RNA identity. *FASEB Journal* **7**, 72-78 (1993).

25. H. Jakubowski. Accuracy of Aminacyl-tRNA Synthetases: Proofreading of amino acids. In *Aminoacyl-tRNA Synthetases*, edited by.M. Ibba, C. FrancklynandS. Cusack.(Landes Bioscience, Austin, 2003).

26. A. C. Dock-Bregeon, B. Rees, A. Torres-Larios, G. Bey, J. Caillet and D. Moras. Achieving error-free translation: The mechanism of proofreading of threonyl-tRNA synthetase at atomic resolution. *Molecular Cell* **16**, 375-386 (2004).

27. J. M. Ogle, A. P. Carter and V. Ramakrishnan. Insights into the decoding mechanismfrom recent ribosome structures. *Trends in Biochemical Sciences* **128**, (2003).

28. N. Budisa. From Reprogrammed Translation and Code Evolution to Artificial Life. In *Engineering the Genetic Code*, edited.(Viley, Weinheim, 2005) pp. 184-213.

29. A. J. Link, M. L. Mock and D. A. Tirrell. Non-canonical amino acids in protein engineering. *Current Opinion in Biotechnology* **14**, 603-609 (2003).

30. P. Kast and H. Hennecke. Amino acid substrate specificity of *Escherichia coli* phenylalanyl-tRNA synthetase altered by distinct mutations. *Journal of Molecular Biology* **222**, 99-124 (1991).

31. O. Nureki, T. Kohno, K. Sakamoto, T. Miyazawa and S. Yokoyama. Chemical modification and mutagenesis studies on zinc-binding of aminoacyl-transfer RNA-synthetases. *Journal of Biological Chemistry* **268**, 15368-15373 (1993).

32. R. S. Mursinna and S. A. Martinis. Rational design to block amino acid editing of a tRNA synthetase. *Journal of the American Chemical Society* **124**, 7286-7287 (2002).

33. L. M. Dedkova, N. E. Fahmi, S. Y. Golovine and S. M. Hecht. Enhanced D-amino acid incorporation into protein by modified ribosomes. *Journal of the American Chemical Society* **125**, 6616-6617 (2003).

34. T. Hohsaka, K. Sato, M. Sisido, K. Takai and S. Yokoyama. Adaptability of nonnatural aromatic amino acids to the active center of the *Escherichia coli* ribosomal A-site. *FEBS Letters* **335**, 47-50 (1993).

35. T. Hohsaka and M. Sisido. Incorporation of non-natural amino acids into proteins. *Current Opinion in Chemical Biology* **6**, 809-815 (2002).

36. L. Wang and P. G. Schultz. Expanding the genetic code. *Chemical Communications*, 1-11 (2002).

37. E. J. Murgola. tRNA, Suppression, and the Code. *Annual Review of Genetics* **19**, 57-80 (1985).

38. D. Mendel, V. W. Cornish and P. G. Schultz. Site-directed mutagenesis with an expanded genetic code. *Annual Review of Biophysics and Biomolecular Structure* **24**, 435-462 (1995).

39. N. Budisa. Amino Acids and Codons- Code Organisation and Protein Structure. In *Engineering the Genetic Code*, edited.(Viley, Weinheim, 2005) pp. 66-90.

40. N. Budisa, L. Moroder and R. Huber. Structure and evolution of the genetic code viewed from the perspective of the experimentally expanded amino acid repertoire in vivo. *Cellular and Molecular Life Sciences* **55**, 1626-1635 (1999).

41. A. L. Weber and S. L. Miller. Reasons for the occurrence of the 20 coded protein amino acids. *Journal of Molecular Evolution* **17**, 273-284 (1981).

42. D. H. Ardell and G. Sella. On the evolution of redundancy in genetic codes. *Journal of Molecular Evolution* **53**, 269-281 (2001).

43. N. Budisa. Prolegomena to future efforts on genetic code engineering by expanding its amino acid repertoire. *Angewandte Chemie-International Edition* **43**, 3387-3428 (2004).

44. F. Chapeville, G. V. Ehrenstein, S. Benzer, B. Weisblum, W. J. Ray and F. Lipmann. On role of soluble ribonucleic acid in coding for amino acids. *Proceedings of the National Academy of Sciences of the United States of America* **48**, 1086-& (1962).

45. Y. Shimizu, A. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, K. Nishikawa and T. Ueda. Cell-free translation reconstituted with purified components. *Nature Biotechnology* **19**, 751-755 (2001).

46. C. J. Noren, S. J. Anthony-Cahill, D. J. Suich, K. A. Noren, M. C. Griffith and P. G. Schultz. In vitro suppression of an amber mutation by a chemically aminoacylated transfer RNA prepared by runoff transcription. *Nucleic Acids Research* **18**, 83-88 (1990).

47. C. J. Noren, S. J. Anthonycahill, M. C. Griffith and P. G. Schultz. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **244**, 182-188 (1989).

48. A. J. Link and D. A. Tirrell. Reassignment of sense codons in vivo. *Methods* **36**, 291-298 (2005).

49. D. B. Cowie and G. N. Cohen. Biosynthesis by Escherichia coli of active altered proteins containing selenium instead of sulfur. *Biochimica et Biophysica Acta* **26**, 252-261 (1957).

50. M. H. Richmond. Effect of amino acid analogues on growth and protein synthesis in microorganisms. *Bacteriological Reviews* **26**, 398-& (1962).

51. N. Budisa, C. Minks, S. Alefelder, W. Wenger, F. M. Dong, L. Moroder and R. Huber. Toward the experimental codon reassignment in vivo: protein building with an expanded amino acid repertoire. *FASEB Journal* **13**, 41-51 (1999).

52. D. B. Cowie, G. N. Cohen, E. T. Bolton and R. H. DeRobichon-Szulmajster. Amino acid analog incorporation into bacterial proteins. *Biochimica et Biophysica Acta* **34**, 39-46 (1959).

53. G. Bogosian, B. N. Violand, E. J. Dorwardking, W. E. Workman, P. E. Jung and J. F. Kane. Biosynthesis and incorporation into protein of norleucine by *Escherichia coli*. *Journal of Biological Chemistry* **264**, 531-539 (1989).

54. C. Minks, S. Alefelder, L. Moroder, R. Huber and N. Budisa. Towards new protein engineering: In vivo building and folding of protein shuttles for drug delivery and targeting by the selective pressure incorporation (SPI) method. *Tetrahedron* **56**, 9431-9442 (2000).

55. C. Minks, R. Huber, L. Moroder and N. Budisa. Noninvasive tracing of recombinant proteins with "fluorophenylalanine-fingers". *Analytical Biochemistry* **284**, 29-34 (2000).

56. P. P. Pal, J. H. Bae, M. K. Azim, P. Hess, R. Friedrich, R. Huber, L. Moroder and N. Budisa. Structural and Spectral Response of Aequorea victoria Green Fluorescent Proteins to Chromophore Fluorination. *Biochemistry* **44**, 3663 -3672 (2005).

57. N. Budisa. Some Practical Potentials of Reprogrammed Cellular Translation. In *Engineering the Genetic Code*, edited.(Viley, Weinheim, 2005) pp. 213-261.

58. N. Fink. Iodotyrosine and tyroglobulin. *Experimentia* **43**, 23-45 (1948).

59. E. C. Jimenez, A. G. Craig, M. Watkins, D. R. Hillyard, W. R. Gray, J. Gulyas, J. E. Rivier, L. J. Cruz and B. M. Olivera. Bromocontryphan: Post-translational bromination of tryptophan. *Journal of Biological Chemistry* **36**, 989-994 (1997).

60. S. R. Patel. Recent advances in systemic therapy of soft tissue sarcomas. *Expert Review of Anticancer Therapy* **3**, 179-184 (2003).

61. D. O'Hagan, C. Schaffrath, S. L. Cobb, J. T. G. Hamilton and C. D. Murphy. Biosynthesis of an organofluorine molecule - A fluorinase enzyme has been discovered that catalyses carbon-fluorine bond formation. *Nature* **416**, 279-279 (2002).

62. F. J. McClure. A review of fluorine and its physiological effects. *Physiological Reviews* **13**, 277-300 (1933).

63. K. Roholm. *Fluorine Intoxication: A Clinical-Hygienic Study*, (HK Lewis, London, 1937).

64. E. N. G. Marsh and E. Neil. Towards the nonstick egg: designing fluorous proteins. *Chemistry & Biology* **7**, R153-R157 (2000).

65. J. D. Dunitz. Organic Fluorine: Odd Man Out. *ChemBioChem* **5**, 614 - 621 (2004).

66. C. Minks, R. Huber, L. Moroder and N. Budisa. Atomic mutations at the single tryptophan residue of human recombinant annexin V: Effects on structure, stability, and activity. *Biochemistry* **38**, 10649-10659 (1999).

67. C. Minks. In vivo Einbau nicht-natürlicher Aminosäuren in rekombinante Proteine. Technische Universität München.(1999).

68. K. L. Kirk. *Biochemistry of halogenated organic compounds*, (Plenum, New York, 1991).

69. M. Ring and R. E. Huber. The properties of beta-galactosidases (*Escherichia coli*) with halogenated tyrosines. *Biochemistry and Cell Biology* **71**, 127-132 (1993).

70. M. D. Vaughan, P. Cleve, V. Robinson, H. S. Duewel and J. F. Honek. Difluoromethionine as a novel F-19 NMR structural probe for internal amino acid packing in proteins. *Journal of the American Chemical Society* **121**, 8475-8478 (1999).

71. N. Budisa, O. Pipitone, I. Slivanowiz, M. Rubini, P. P. Pal, T. A. Holak, R. Huber and M. L. Efforts toward the design of 'Teflon' proteins: In vivo translation with trifluorinated leucine and methionine analogues. *Chemistry & Biodiversity* **1**, 1465-1475 (2004).

72. R. L. Munier and G. Sarrazin. Substitution totale de la 3-fluorotyrosine a la Tyrosine dans les Proteines d'*Escherichia coli*. *Comptes Rendus Hebdomadaires Des Seances De L Academie Des Sciences* **256**, 3376-& (1963).

73. M. Ring, I. M. Armitage and R. E. Huber. Meta-fluorotyrosine substitution in Beta-galactosidase - Evidence for the existence of a catalytically active tyrosine. *Biochemical and Biophysical Research Communications* **131**, 675-680 (1985).

74. B. Brooks, R. S. Phillips and W. F. Benisek. High-efficiency incorporation in vivo of tyrosine analogues with altered hydroxyl acidity in place of the catalytic tyrosine-14 of Delta(5)-3-ketosteroid isomerase of Comamonas (Pseudomonas) testosteroni: Effects of the modifications on isomerase kinetics. *Biochemistry* **37**, 9738-9742 (1998).

75. G. Michal. *Biochemical pathways: An atlas of biochemistry and molecular biology*, (Wiley, Spektrum, New York, Toronto, Heidelberg, Beriln, 1999).

76. A. Weissman and B. K. Koe. Studies on Mechanism of Convulsant Action of Dl-M-Fluorotyrosine. *Pharmacologist* **8**, 215-& (1966).

77. F. L. M. Pattison. Toxic Fluorine Compounds. *Nature* **172**, 1139-1141 (1953).

78. B. D. Sykes, Weingart.Hi and Schlesin.Mj. Fluorotyrosine Alkaline-Phosphatase from Escherichia-Coli - Preparation, Properties, and Fluorine-19 Nuclear Magnetic-Resonance Spectrum. *Proceedings of the National Academy of Sciences of the United States of America* **71**, 469-473 (1974).

79. J. W. Payne. Drug delivery systems: optimising the structure of peptide carriers for synthetic antimicrobial drugs. *Drugs Exp Clin Res* **12**, 585-594 (1986).

80. N. Budisa. Reprogramming the Cellular Translation Machinery. In *Engineering the Genetic Code*, edited.(Viley, Weinheim, 2005) pp. 90-184.

81. G. Rosenthal. *Plant nonprotein amino and imino acids. Biological, Biochemical and Toxicological Properties*, (Academic Press, New York, 1982).

82. C. S. Evans and E. A. Bell. Neuroactive plant amino acids and amines. *Trends in Neurosciences* **3**, 70-72 (1980).

83. M. D. Armstrong and J. D. Lewis. The toxicity of ortho-fluorophenyl-dl-alanine and para-fluorophenyl-dl-alanine for the rat. *Journal of Biological Chemistry* **188**, 91-95 (1951).

84. S. Kaufman. Hydroxylation of 4-fluoro-L-phenylalanine by phenylalanine hydroxylase results in L-tyrosine and fluoride ion. *Biophysica et Biochimica Acta* **19**, 619-627 (1961).

85. P. P. Pal and N. Budisa. Designing novel spectral classes of proteins with tryptophan-expanded genetic code. *Biological Chemistry* **385**, 893-904 (2004).

86. G. J. Palm and A. Wlodawer. Spectral variants of green fluorescent protein. In *Green Fluorescent Protein*, edited, Vol. 302, 1999) pp. 378-394.

87. M. H. Seifert, D. Ksiazek, M. K. Azim, P. Smialowski, N. Budisa and T. A. Holak. Slow exchange in the chromophore of a green fluorescent protein variant. *Journal of the American Chemical Society* **124**, 7932-7942 (2002).

88. J. R. Lakowitz. *Protein fluorescence*. 2nd edit. Principles of fluorescence spectroscopy, (Kluwer Academic/Plenum Publisher, New York, 1999).

89. J. L. Hott and R. F. Borkman. The non-fluorescence of 4-fluorotryptophan. *Biochemistry Journal* **264**, 297–299 (1989).

90. M. Rubini. Noncanonical amino acid in proteins. PhD Thesis, TU Munich.(2004).

91. H. Duewel, E. Daub, V. Robinson and J. F. Honek. Incorporation of trifluoromethionine into a phage lysozyme: Implications and a new marker for use in protein F-19 NMR. *Biochemistry* **36**, 3404-3416 (1997).

92. N. Budisa, B. Steipe, P. Demange, C. Eckerskorn, J. Kellermann and R. Huber. High-level biosynthetic substitution of methionine in proteins by Its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in *Escherichia coli*. *European Journal of Biochemistry* **230**, 788-796 (1995).

93. R. Huber, J. Romisch and E. P. Paques. The Crystal and Molecular-Structure of Human Annexin-V, an Anticoagulant Protein That Binds to Calcium and Membranes. *EMBO Journal* **9**, 3867-3874 (1990).

94. O. Pipitone. Incorporation of trifluoromethionine into green fluorescent protein. Diploma Work, University of Milan.(2004).

95. M. Chalfie, Y. Tu, G. Euskirchen, W. W. Ward and D. C. Prasher. Green fluorescent protein as a marker for gene expression. *Science* **263**, 802-805 (1994).

, V. V. Verkhusha and K. A. Lukyanov. The molecular properties and applications of Anthozoa fluorescent proteins and chromoproteins. *Nature Biotechnology* **22**, 289–296 (2004).

96. N. C. Shaner, P. A. Steinbach and R. Y. Tsien. A guide to choosing fluorescent proteins. *Nature Methods* **2**, 905-909 (2005).

97. B. G. Reid and G. C. Flynn. Chromophore formation in green fluorescent protein. *Biochemistry* **36**, 6786-6791 (1997).

98. O. Shimomura. Structure of the Chromophore of Aequorea Green Fluorescent Protein. *FEBS Letters* **104**, 220-222 (1979).

99. F. Yang, L. G. Moss and G. N. Phillips. The molecular structure of green fluorescent protein. *Nature Biotechnology* **14**, 1246-1251 (1996).

100. C. W. Cody, D. C. Prasher, W. M. Westler, F. G. Prendergast and W. W. Ward. Chemical-Structure of the Hexapeptide Chromophore of the Aequorea Green-Fluorescent Protein. *Biochemistry* **32**, 1212-1218 (1993).

101. H. Niwa, S. Inouye, T. Hirano, T. Matsuno, S. Kojima, M. Kubota, M. Ohashi and F. I. Tsuji. Chemical nature of the light emitter of the Aequorea green fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 13617-13622 (1996).

102. R. Y. Tsien. The green fluorescent protein. *Annual Review of Biochemistry* **67**, 509-544 (1998).

103. R. M. Wachter, M. A. Elsliger, K. Kallio, G. T. Hanson and S. J. Remington. Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein. *Structure* **6**, 1267-1277 (1998).

104. M. Chattoraj, B. A. King, G. U. Bublitz and S. G. Boxer. Ultra-fast excited state dynamics in green fluorescent protein: Multiple states and proton transfer. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 8362-8367 (1996).

105. A. D. Kummer, J. Wiehler, H. Rehaber, C. Kompa, B. Steipe and M. E. Michel-Beyerle. Effects of threonine 203 replacements on excited-state dynamics and fluorescence properties of the green fluorescent protein (GFP). *Journal of Physical Chemistry B* **104**, 4791-4798 (2000).

106. O. Griesbeck, G. S. Baird, R. E. Campbell, D. A. Zacharias and R. Y. Tsien. Reducing the environmental sensitivity of yellow fluorescent protein - Mechanism and applications. *Journal of Biological Chemistry* **276**, 29188-29194 (2001).

107. M. Zimmer. Green fluorescent protein (GFP): Applications, structure, and related photophysical behavior. *Chemical Reviews* **102**, 759-781 (2002).

108. 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. Expansion of the genetic code enables design of a novel "gold" class of green fluorescent proteins. *Journal of Molecular Biology* **328**, 1071-1081 (2003).

109. V. I. Martynov, A. P. Savitsky, N. Y. Martynova, P. A. Savitsky, K. A. Lukyanov and S. A. Lukyanov. Alternative Cyclization in GFP-like Proteins Family. *Journal of Biological Chemistry* **276**, 21012 - 21016 (2001).

110. Y. A. Labas, N. G. Gurskaya, Y. G. Yanushevich, A. F. Fradkov, K. A. Lukyanov, S. A. Lukyanov and M. V. Matz. Diversity and evolution of the green fluorescent protein family. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 4256 - 4261 (2002).

111. R. E. Campbell, O. Tour, A. E. Palmer, P. A. Steinbach, D. A. Baird, D. A. Zacharias and R. Y. R. Y. Tsien. A monomeric red fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 7877 - 7882 (2002).

112. U. K. Laemmli. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **277**, 680–685 (1970).

113. J. Sambrook, E. F. Fritsch and T. Maniatis. *Molecular Cloning: A Laboratory Manual.*, 1,2,3, (Cold Spring Harbor Laboratory Press, New York, 1989).

114. N. Budisa, C. Minks, F. J. Medrano, J. Lutz, R. Huber and L. Moroder. Residue-specific bioincorporation of non-natural, biologically active amino acids into proteins as possible drug carriers: Structure and stability of the per-thiaproline mutant of annexin V. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 455-459 (1998).

115. H. Fukuda, M. Arai and K. Kuwajima. Folding of green fluorescent protein and the cycle3 mutant. *Biochemistry* **39**, 12025-12032 (2000).

116. J. H. Bae, P. P. Pal, L. Moroder, R. Huber and N. Budisa. Crystallographic Evidence for Isomeric Chromophores in 3-Fluorotyrosyl-Green Fluorescent Protein. *ChemBioChem* **5**, 720-722 (2004).

117. D. Yarbrough, R. M. Rebekka M. Wachter, K. Kallio, M. V. Matz and S. J. Remington. Refined crystal structure of DsRed, a red fluorescent protein from coral, at 2.0-Å resolution. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 462-467 (2001).

118. L. Wang, J. M. Xie, A. A. Deniz and P. G. Schultz. Unnatural amino acid mutagenesis of green fluorescent protein. *Journal of Organic Chemistry* **68**, 174-176 (2003).

119. D. Kajihara, T. Hohsaka and M. Sisido. Synthesis and sequence optimization of GFP mutants containing aromatic non-natural amino acids at the Tyr66 position. *Protein Engineering, Design & Selection* **18**, 273-278 (2005).

120. N. V. Visser, J. W. Borst, M. A. Hink, A. van Hoek and A. J. Visser. Direct observation of resonance tryptophan-to-chromophore energy transfer in visible fluorescent proteins. *Biophysical Chemistry* **116**, 207-212 (2005).

121. N. Budisa, P. P. Pal, S. Alefelder, P. Birle, T. Krywcun, M. Rubini, W. Wenger, J. H. Bae and T. Steiner. Probing the role of tryptophans in *Aequorea victoria* Green Fluorescent Proteins with an expanded genetic code. *Biological Chemistry* **385**, 191-202 (2004).

122. N. Budisa. Expression of 'tailor-made' proteins via incorporation of synthetic amino acids by using cell-free protein synthesis. In *Cell Free Protein Expression*, edited by.J. R. Swartz.(Springer Verlag, Berlin, Heidelberg, New York, 2003) pp. 89-98.

123. C. Giese. Einfluss fluorierter aromatischer Aminosäuren auf zelluläres Wachstum. Diploma thesis, Martin-Luther-Universität Halle-Wittenberg.(2005).

124. J. Wiehler, G. Jung, C. Seebacher, A. Zumbusch and B. Steipe. Mutagenic Stabilization of the Photocycle Intermediate of Green Fluorescent Protein. *ChemBioChem* **4**, 1164-1171 (2003).

125. M. Ormo, A. B. Cubitt, K. Kallio, L. A. Gross, R. Y. Tsien and S. J. Remington. Crystal structure of the Aequorea victoria green fluorescent protein. *Science* **273**, 1392-1395 (1996).

126. R. Heim, D. C. Prasher and R. Y. Tsien. Wavelength Mutations and Posttranslational Autoxidation of Green Fluorescent Protein. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 12501-12504 (1994).

127. A. A. Voityuk, A. D. Kummer, M. E. Michel-Beyerle and N. Rosch. Absorption spectra of the GFP chromophore in solution: comparison of theoretical and experimental results. *Chemical Physics* **269**, 83-91 (2001).

128. G. Y. Xiao, J. F. Parsons, K. Tesh, R. N. Armstrong and G. L. Gilliland. Conformational changes in the crystal structure of rat glutathione transferase M1-1 with global substitution of 3-fluorotyrosine for tyrosine. *Journal of Molecular Biology* **281**, 323-339 (1998).

129. M. F. Jacobson and Baltimorore D. Polypeptide cleavages in formation of Poliovirus proteins. *Proceedings of the National Academy of Sciences of the United States of America* **61**, 77-& (1968).

130. Y. Tang and D. A. Tirrell. Attenuation of the editing activity of the Escherichia coli Leucyl-tRNA synthetase allows incorporation of novel amino acids into proteins in vivo. *Biochemistry* **41**, 10635-10645 (2002).

131. G. Hortin and I. Boime. Applications of amino acid analogs for studying co-translational and posttranslational modifications of proteins. *Methods in Enzymology* **96**, 777-784 (1983).

132. G. Hortin and I. Boime. Markers for processiong sites in eukaryotic proteins: characterisation with amino acid analogs. *Trends in Biochemical Sciences* **8**, 320-323 (1983).

133. 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. F-19 nuclear magnetic resonance chemical shifts of fluorine containing aliphatic amino acids in proteins: Studies on Lactobacillus casei dihydrofolate reductase containing (2S,4S)-5-fluoroleucine. *Journal of the American Chemical Society* **118**, 8700-8706 (1996).

134. J. O. Freitas, C. Termignoni, D. R. Borges, C. A. M. Sampaio, J. L. Prado and J. A. Guimaraes. Methionine Aminopeptidase Associated with Liver-Mitochondria and Microsomes. *International Journal of Biochemistry* **13**, 991-997 (1981).

135. R. A. Bradshaw. Methionine aminopeptidase 2 inhibition: antiangiogenesis and tumour therapy. *Expert Opininon on Therapeutic Patents* **14**, 1-6 (2004).