Max-Planck-Institut für Biochemie
Abteilung Strukturforschung

Protein Engineering and Design with Non Canonical Amino Acids

Marina Rubini

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

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr. Dr. Adalbert Bacher
Prüfer der Dissertation:	1. apl. Prof. Dr. Dr. h.c. Robert Huber
	2. Univ.-Prof. Dr. Johannes Buchner

Die Dissertation wurde am 22.09.2004 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 26.11.2004 angenommen.

To the memory of my beloved Granny, Mary

Parts of this work were published or presented at congresses as listed below:

1) Budisa, N., Rubini, M., Bae, J.H., Weyher, E., Wenger, W., Golbik, R., Huber, R. and Moroder, L. (2002) Global replacement of tryptophan with aminotryptophans generates non-invasive protein-based optical pH sensors. *Angewandte Chemie-International Edition*, **41**, 4066-4069.

2) Bae, J.H., Rubini, M., Jung, G., Wiegand, G., Seifert, M.H.J., Azim, M.K., Kim, J.S., Zumbusch, A., Holak, T.A., Moroder, L., Huber, R. and Budisa, N. (2003) Expansion of the genetic code enables design of a novel "gold" class of green fluorescent proteins. *Journal of Molecular Biology*, **328**, 1071-1081.

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

4) Budisa N., Pipitone O., Siwanowicz I., Rubini M., Pal P.P., Holak T. A. and Maria Luisa Gelmi (2004) Efforts toward the Design of 'Teflon' Proteins *In vivo* Translation with Trifluorinated Leucine and Methionine Analogues. *Chemistry & Biodiversity* (in press)

5) Rubini, M., Lepthien, S., Pal, P.P., Huber, R., Moroder, L. and Budisa, N. (2004) Thermodynamics of the expanded genetic code: Structure and stability of Barstar as pH sensor. Dynamics of Proteins, Symposium of the SFB533, Freising, 9-11 July

ABBREVIATIONS AND DEFINITIONS

Canonical amino acids are defined in this work by the three letter code: Tryptophan (Trp), Leucine (Leu), Phenylalanine (Phe), Tyrosine (Tyr), Proline (Pro), etc.

Non canonical amino acids are denoted with the functional groups that characterize them, followed by the three letter code, e.g. (4-NH_2)Trp (4-aminotryptophan); (7-F)tryptophan (7-fluorotryptophan).

The term "**analogue**" refers to strict isosteric exchanges of canonical/non canonical amino acids (e.g.Methionine/Seleno-methionine)

The term "**surrogate**" refers to non isosteric changes of canonical/non canonical amino acids (e.g.Methionine/Ethionine)

The abbreviation **SPI** represents selective pressure incorporation method.

The term "**aaRS**" is generally used to represent aminoacyl-tRNA synthetases.

Mutant denotes proteins in which the wild-type sequence is changed by site-directed mutagenesis in the pool of the 20 canonical amino acids.

Variant denotes proteins in which one or more canonical amino acids from a wild-type or mutant sequences are replaced with non canonical ones.

Ax V is an abbreviation for human recombinant Annexin V.

Barstar, wt-Barstar, and **P27A** define the mutant (Cys40Ala/Cys82Ala/Pro27Ala) of the inhibitor of the extracelluar RNase barnase. **W44F**, **W38F**, and **W3844F**, define the Barstar mutants (Cys40Ala/Cys82Ala/Pro27Ala/Trp44Phe), (Cys40Ala/Cys82Ala/Pro27Ala/Trp38Phe), and (Cys40Ala/Cys82Ala/Pro27Ala/Trp38Phe/ Trp44Phe), respectively.

`***av*GFP**´ defines the wild type Green Fluorescent Protein from *Aequorea Victoria*.

ACKNOWLEDGMENTS

This PhD work was done in the Abteilung Strukturforschung and AG Bioorganische Chemie at Max-Planck-Institut für Biochemie in Martinsried bei München from March 2001 to September 2004.

First, I would like to express my deep gratititude to Prof. Dr. Luis Moroder who gave me the possibility to begin my PhD at Max-Planck-Institut, and to my "Doktorvater" Prof. Dr. Robert Huber.

My special thanks go to my supervisor Dr. Nediljko Budisa for his continuous generous technical, scientific, and human support. Our fruitful discussions have been always been an enrichment for me.

I would like to thank also Mrs Waltraud Wenger, Ms Tatjana Krywcun and Ms Petra Birle for their excellent technical help, their comprehensive explanations and their patient support during my "discovery" of the world of Molecular Biology. I am also thankful to Mrs Elisabeth Weyher for her technical help.

My thanks go to all my colleagues both in the Moroder´s and in the Huber´s Department. Among them I am glad to thank especially my colleague and friend Prajna Paramita Pal (Tabby). A big thank you to you Tabby, for never leaving me alone! I would also like to thank my colleague and friend Thomas Steiner. Merci, Thomas!

Thanks to Dr. Jae Hyun Bae, Dr. Kamran Azim, Olga Pipitone, Dr. Petra Hess, Bojana Bolic, Sandra Lepthien, Dr. Peter Göttig and Dr. Rainer Friedrich. Thanks also to Dr. Piotr Knyazev from the Ullrich´s Department for the interesting collaboration.

I am also in debt to Mrs Renate Rüller, Mrs Monika Bumann, Mrs Monika Schneider, Mrs Marion Heinze, Mr Karsten-Peter Kaerlein, Mr Werner Dersch, and Mr Paul Ottmar for their help in bureaucratic and logistic matters.

Last, I would like to express my heartfelt gratitude to my "Diplomvater" Prof. Fernando Filira, who has taught me never to give up. He is like a second father for me.

SUMMARY

In the last decades, experimental results from our and other laboratories, showed how the translation machinery of cells can efficiently be exploited for the incorporation of various non-canonical amino acids into proteins. The incorporation of noncanonical amino acids in combination with site-directed mutagenesis was used to probe spectroscopic and structural roles of tryptophan (Trp) residues in barstar and Annexin V. Different fluoro-, amino-, and methyl-containing Trp-isosteric analogues were incorporated into model proteins by the use of selective pressure incorporation (SPI) method. Such isosteric replacements introduced minimal local geometry changes in indole moieties, often to the level of single atomic exchange ("atomic mutation") and normally do not affect three-dimensional structures of substituted proteins but induce significant changes in spectral and folding properties. For example, mutants of Barstar can be stabilised by incorporation of fluorinated Trp-analogues, while Trp-fluorescence can be abolished by 4-, and 7-fluorotryprophan.

A novel class of proteins with a fluorous core can be envisaged only if a full replacement of the core-building hydrophobic and aliphatic amino acids such as leucine with the related analogue trifluoroleucine is possible. However, attempts to quantitatively introduce trifluoroleucine in annexin V and green fluorescent protein were not successful. The reasons are high toxicity of these substances and difficulties to accommodate them into the compact cores of natural proteins without adverse effects on their structural integrity.

The replacement of tryptophan residues in barstar with its analogues 4-aminotryptophan and 5-aminotryptophan yielded related protein variants with fluorescence pH-sensitivity. The crystal structure of 4-aminotryptophan-barstar as a pH sensor is almost identical to those of the parent protein, while its thermodynamic behavior in solution proved to be dramatically different. In the native states of both Barstar variants, almost 10% protein is already unfolded and prone to cold-denaturation below the temperature range of 17–22 °C with lowered T_m value (- 20 °C) and substantially reduced unfolding cooperativity. Reasons for these changes are that the 4-aminotryptophan and 5-aminotryptophan are more hydrophilic than Trp itself and their presence in the hydrophobic barstar interior violated the basic rules of protein folding (polar-out; apolar-in). This unambiguously demonstrated that the thermodynamic penalty for amino acid repertoire expansion for protein building might be too high, in order to gain new, or maximize the efficiency of a single function.

Most of the non canonical amino acids are toxic; this toxicity is the result of their conversion into toxic substance by a relatively complex metabolic route. In this way, tumour cells might be cured or selectively killed, if these cytotoxic amino acids could be specifically delivered to them. Proteins substituted by toxic analogues can be specifically intracellulary delivered and after their recycling and the release of an analogue target, cells would be directed to apoptosis or necrosis. Initial experiments done by using liposome-mediated delivery of 3-fluorotyrosine-green fluorescent protein indicate that this scenario is principally possible.

CONTENTS

1. INTRODUCTION

1.1 The structure and properties of the genetic code

For all living organisms, the genetic information is saved in string-like nucleic acids forms (ribonucleic acid, RNA; deoxyribonucleic acid, DNA). DNA is not the direct template for protein synthesis, as it is transcribed into mRNA in the nucleus of eukaryotic cells and mRNA is subsequently translated in the cytoplasm at the ribosome site. Transfer RNA (tRNA) is a molecule that acts as an adaptor between mRNA codons and corresponding amino acids as it was formulated by F. Crick in the "adaptor hypothesis" (Crick, 1958)

The flow of the genetic information takes place only in one definite direction, as the information stored in DNA can be transferred to RNA and from RNA can be reversed back to DNA by reverse transcriptase (present in some viruses) or further transferred to a protein as a final point. This assumption represents the so called "central dogma of molecular biology" originally proposed by Francis Crick (Crick, 1958).

The early studies on the "meaning" of the genetic code showed that the genetic information is stored in the form of nucleotides triplets and in 1961, Nirenberg and co-workers succeeded in assigning the first codon meaning and related the UUU triplet to the amino acid phenylalanine, as he discovered that polyurydilic acid leads to the assembling of phenylalanine molecules (i.e. poly-Phe peptide). After the complete "decoding" of the triplet genetic code it has been clearly explained how the information for linear amino acid sequences is to be found in collinear nucleic acid sequences and how the twenty canonical amino acids can be assigned to the 61 coding triplets (codons UGA, UAA, and UAG induce the termination of the translation). As each amino acid is assigned to more than one coding triplet with Met and Trp as exception to this rule (i.e. amino acids such as Leu, Ser and Arg are encoded by six codons) it can be stated that the code is degenerated.

The presence of high synonymous quotas helps to control the effects of mutations. Actually, it is more likely that a random mutation in a single base does not lead to an amino acid variation or at least to the encoding of an amino acid similar to the original one. For example, the mutation of CUC to CUG has no effect, as both codons represent the amino acid Leu and the mutation of CUU to AUU causes the substitution of Leu with Ile which is a very similar amino acid (i.e. hydrophobic)

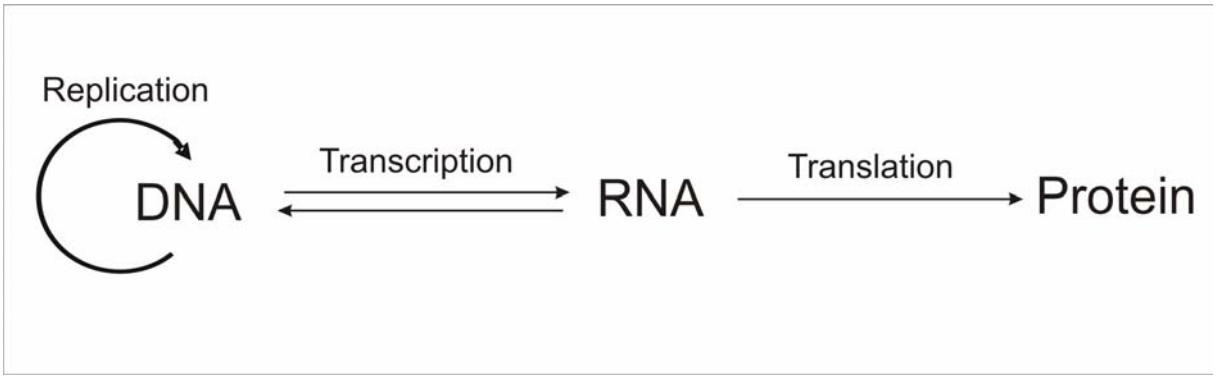


Figure 1.1. The "central dogma of molecular biology". The flow of the genetic information is allowed only from nucleic acids to nucleic acids and from nucleic acids to proteins. In the original proposal Crick (1958) wrote "the transfer of information from nucleic acid to nucleic acid or from nucleic acid to protein may be possible, but transfer from protein to protein or from protein to nucleic acid is impossible".

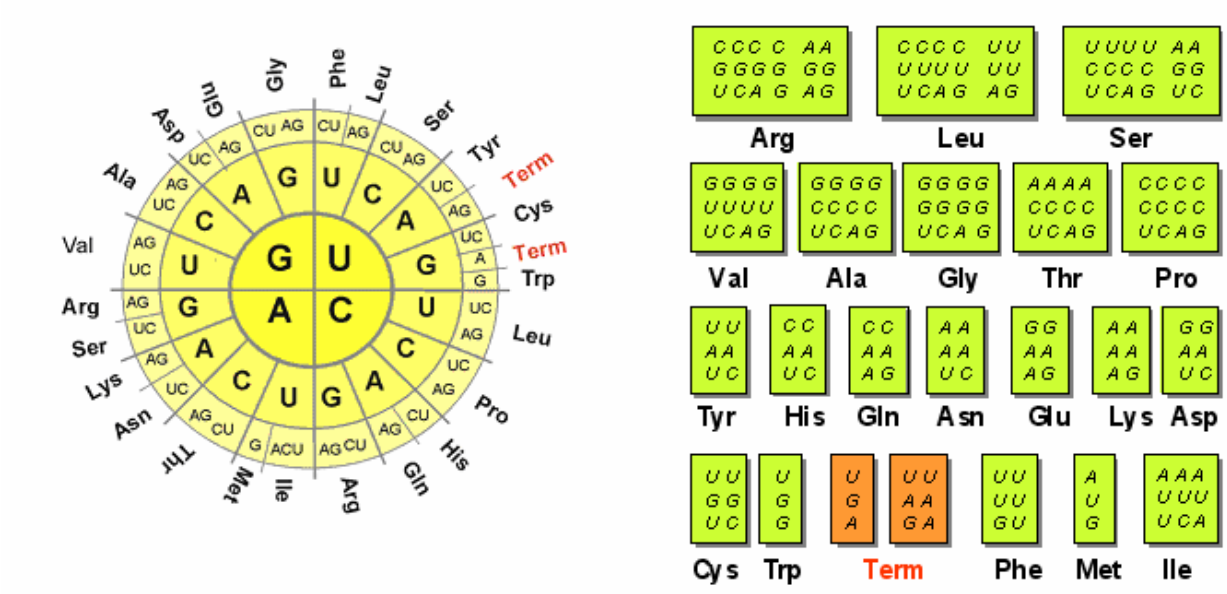


Figure 1.2. The structure of the genetic code represented in radial form (left). On the right side synonymous quotas for particular amino acids are shown. Note that some amino acids like Arg, Leu Ser are characterized by high synonymous quotas (i.e. assigned to larger number of codons) (Taken from Budisa, 2004)

Accordingly, a mutation leading to the introduction of a hydrophilic amino acid in a hydrophobic environment such the protein interior, is not likely to happen since the codons of hydrophobic amino acids are grouped together (the XUX group). The acceptance of amino acid changes critically

depends on the biological role of the respective amino acids. This role, in organisms, refers not only to the known function of a protein, such as biological activity, but to all interactions of a protein with other parts of the organism over the proteins entire lifespan from activation of the corresponding gene to polypeptide degradation.

Is the genetic code the same for all the living organisms? To answer this question it has at first to be established which code has to be considered as the "standard" one, as many changes of the codon meaning have been found both in prokaryotes and eukaryotes (Osawa, *et al*., 1992). For example, in a *Mycoplasma* the codon UGA (STOP codon) codes for the amino acid Trp and in certain species of Cyliata like *Tetrahymena* and *Paramecium*, UAA and UAG (also STOP codons) code for the amino acid Gln. Similar alterations of the genetic code are also observed in mythocondrial DNA (Anderson *et al*., 1981; Borst and Grivell, 1981).

If it is assumed the codon reassignments of E. *coli* or of eukaryotic nuclear genes as representative of the "standard" genetic code, than it should be concluded that there are many "parallel" codes, but, changing perspective, if the attention is focused on the amino acid repertoire it is evident that there have always been only 20 canonical amino acids as substrates for protein synthesis. This means that in every living being all coding triplets do not allow the introduction of any other amino acid outside the conserved pool. Following this logic at reasoning it can be assumed that the genetic code is universal and conserved.

The genetic code is likely to have been established in a very early phase of the evolution. It was speculated that at the very beginning, the relationship between amino acids and corresponding codons was of stereochemical nature at the beginning of the cell evolution. Then, up-to date system for the protein synthesis could have evolved through a selection aiming to a better efficiency or a greater precision.

There are hypothesis that postulate the evolution of the genetic code from a primitive form, where few codons represented few amino acids, or furthermore, where one codon could have represented any amino acid of a certain group (Weberndorfer *et al*., 2003; Alberti, 1997). It is also possible that at the beginning only the first two bases of a codon were used to determine the amino acid and that the discrimination at the level of the third base has evolved only in a later phase (Di Giulio, 1997).

The evolution of the genetic code might have been "frozen" when the system had become so complex that any changes in the meaning of the codons would have led to a structural "upsetting" of the existing proteins, causing substitutions with non suitable amino acids (Crick, 1968).

1.2 Basic features of ribosomal protein synthesis

There are two crucial molecular recognition events in protein translation:

- Amino acids aminoacylation performed by specific aminoacyl-tRNA synthetases (aaRS)
- Codon-anticodon interactions between tRNAs and mRNAs on the ribosome

The ribosome consists in two unequal subunits made of proteins and RNA which can associate with each other in a labile manner. It can be considered as a three-functional entity. It performs a genetic function, as it is responsible for the arrangement of amino acids in accordance with the correspondent nucleotide sequence; it possesses an enzymatic function as it assumes the role of a peptidyl transferase, catalysing the transpeptidation reaction and is also a conveying molecular machine moving along mRNA chain and accommodating tRNA molecules through itself during elongation (Stryer *et al*., 2001; Spirin *et al*., 2002).

At the beginning amino acids are activated by a specific aminoacyl-tRNA synthetase and then transferred to a cognate tRNA (there is at least one specific tRNA per each amino acid) to form an aminoacyl-tRNA, which is sent to the ribosome to take part in the protein synthesis. All tRNAs are able to occupy both sites A (acceptor site) and P (donor site) on the ribosome. When the tRNA is inside one of these sites one of its ends is associated with the mRNA codon, while at the other end of the molecule takes place the transfer of the growing polypeptide. The protein synthesis goes on through three steps:

1. Initial phase: To begin translation, the large (top) and small (bottom) ribosomal units must be bound together to the strand of mRNA. The ribosome positions itself so that the codon sequence AUG (the "start" codon) on the mRNA is exposed. A tRNA molecule with the anticodon sequence UAC binds to the exposed "start" codon. This first tRNA only carries the amino acid methionine in eukaryotes or formyl-methionine in bacteria, which is now set in place.

2. Elongation phase: in this phase take place all reactions between the synthesis of the first peptide bond and the addition of the last amino acid. As the large ribosomal unit sets in place, the second codon on mRNA is exposed. This exposed site is known as the A site. At this point, a special protein unit known as the elongation factor assists the second

tRNA to bond to this newly exposed codon and the newly arrived amino acid is lined up next to the first amino acid. An enzyme binds both amino acids via dehydration synthesis and subsequent amino acids are added one by one to the growing polypeptide chain.

3. Termination**:** All necessary reactions for the release of the polypeptide chain take place. When a codon with the sequence UAA, UAG or UGA is exposed it is a signal that translocation is to stop. The stop codon is not bonded to a complementary anticodon sequence on a tRNA. Rather, a protein known as a *release factor* (RF) binds at the A site. The release factor ultimately will help to release the finished polypeptide chain.

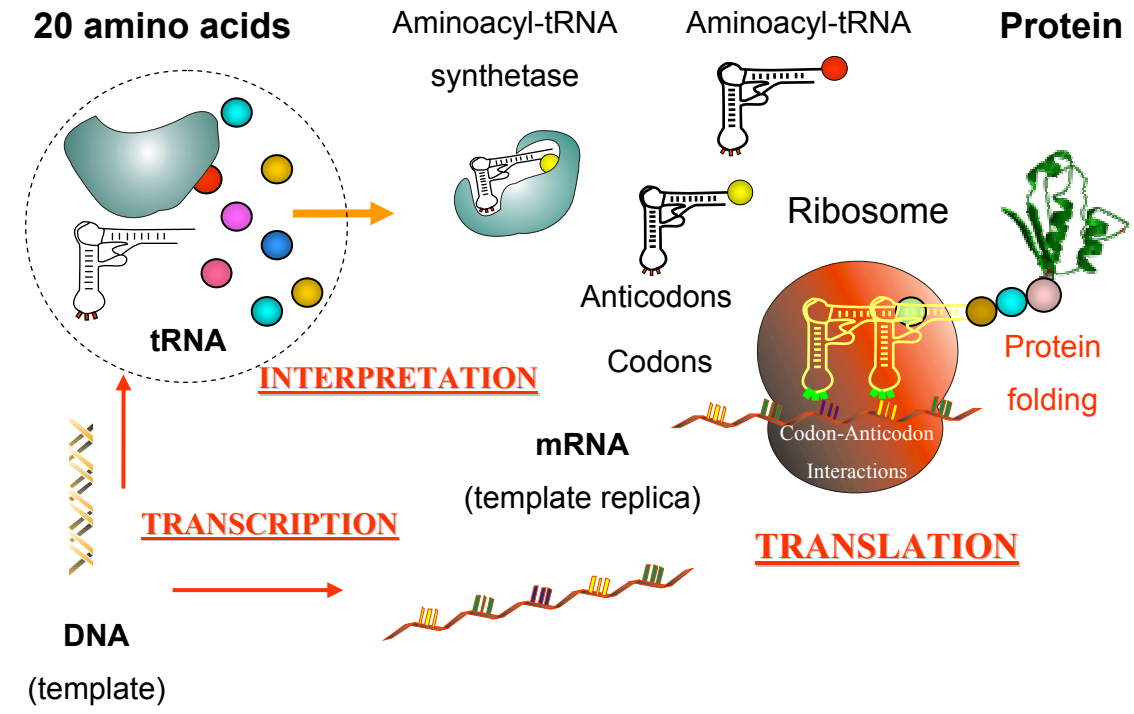


Figure 1.3. Ribosome-mediated protein synthesis in the context of the flow of the genetic information. Amino acids are specifically activated and loaded on cognate tRNAs by aminoacyl-tRNA-synthetases (interpretation of the genetic code). On the ribosome the direct translation of mRNA takes place by base pairing between tRNA-anticodons and mRNA codons. The resulting protein upon folding is capable of performing specific biological functions.

1.3 Amino acid selection, tRNA charging and editing

Aminoacyl-tRNA synthetases (aaRS) are crucial enzymes in the translation of the genetic message. They perform aminoacylation reaction where the specific amino acid is recognized, selected, activated and subsequently transferred onto the cognate tRNA (see Fig.1.4).

aaRS are divided into two different unrelated classes that differ in the topologies of the active site (Eriani *et al*., 1990), but show the same conserved mechanism of catalysis reaction. In the first step ATP and the amino acid bind at the active site, so that the α-carboxylate of the amino acid can attack the α-phosphate of ATP leading to the formation of an aminoacyl-adenylate and an inorganic pyrophosphate leaving group. In the second step of the reaction, the 2′- or 3′-OH of the terminal adenosine of tRNA attacks the α-carbonyl of the aminoacyl-adenylate, so that the tRNA is esterified with the amino acid moiety and AMP is generated as leaving group (Ibba *et al*., 2000).

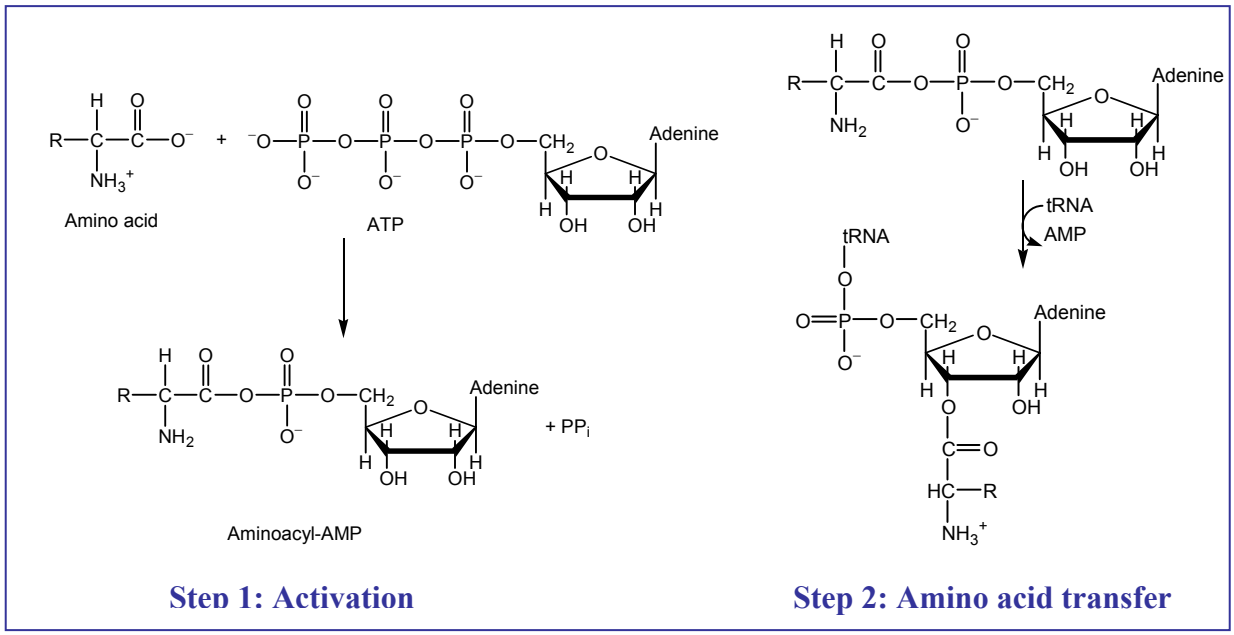


Figure 1.4. Aminoacylation reaction catalysed by aminoacyl-tRNA synthetase can be generally divided into two steps. In the first step (activation) the cognate amino acid is recognized and in the presence of ATP activated. Resulting aminoacyl-adenilate is then dissolved and amino acid is transferred to the cognate tRNA (step 2).

An aminoacyl-tRNA synthetase recognizes the correct tRNA through different interactions between its active site and specific regions on the tRNA itself. Isoacceptor tRNAs are endowed with a greater affinity for the binding site on the aaRS. If the tRNA that has bound is correct, the binding becomes

stabilized by a conformational change on the aaRS molecule, followed by a quick aminoacylation. If the tRNA turns out to be incorrect (i.e. non cognate), the conformational change does not take place, thus the aminoacylation process slow down and it is more likely that the tRNA dissociates from the enzyme. This control modality represents an example of kinetical proofreading. On the other hand, the specificity for the amino acid is preserved by a chemical proofreading or editing and this operation is performed through two different mechanisms depending on the aaRS type (Jakubowsky and Goldman, 1992):

a) pre-transfer editing
b) post-transfer editing

In the pre-transfer editing mechanism, the incorrect aminoacyl-adenylate can be hydrolyzed when the correct tRNA is bound. The best studied example of this editing type is represented by Homocystein mischarge at Met-tRNA synthetase (MetRS). After the proofreading, homocystein is released in the form of thiolactone, a by-product of the "loading" operated by MetRS.

Some other aaRS perform the proofreading in a later step, when the incorrect amino acid has already been transferred to the tRNA and only later is hydrolyzed and released.

Error frequencies during protein synthesis (substitution of one amino acid with another) are in the range from 10^{-3}-10^{-4}(Jakubowsky and Goldman, 1992). In this context, during the selection of amino acids by aaRS error in tRNA charging is in the range 10^{-4} to 10^{-5}. It must be kept in mind that the accuracy of protein synthesis depends on other factors as well. For example, selectivity of ribosomal A site for aminoacyl-tRNA or errors during DNA replication and mRNA transcription.

1.4 Methods for protein modifications

1.4.1 Selective Chemical Modifications

Selective chemical modifications (also known as chemical mutations) of enzymes require that an amino acid be uniquely reactive. Some examples of such uniquely reactive residues are the nucleophile serine in the catalytic triad of serine proteases and the active-site cystein in papain, because of its uniquely reactive thiol group (Kaiser *et al*., 1984).

Chemical mutation has been utilized also to construct semi-synthetic DNA-cleaving proteins by coupling small synthetic molecules to proteins that bind to DNA. One conjugate between CAP protein and copper-phenantroline provided a selective cleavage of megabase DNA (Pendergrast *et*

al., 1994), as phenantroline is accessible to DNA at the cognate binding site but is prevented by the protein to associate random with other DNA within the genome.
Even if nowadays genetic manipulations represent a route to more diverse variants, chemical mutations offer significant advantages for introducing novel chemical functionality into already-folded protein.

1.4.2 Chemical Synthesis

Peptide synthesis methods (both in solution and on solid phase) are very useful tools to incorporate non canonical amino acids.
The solid phase peptide synthesis is the most commonly used method to introduce modified amino acids into relatively short peptides. This method (Merrifield, 1963) makes the peptide chain grow on a solid insoluble resin and after the desired sequence is obtained the peptide is cleaved from the solid support. However, the rate of error during the peptide synthesis increases with the length of the sequence, therefore it is recommended to synthesise separately peptide blocks shorter than 50 residues and to subsequently couple them either enzymatic or chemical to give the entire protein sequence. Kent and co-workers have developed a synthetic method that employs the chemical coupling of unprotected peptide fragments (Dawson *et al*., 1992; Baca *et al*., 1993) by the use of thiol/thioester exchange reaction. Specifically, the carboxy termini of each peptide are activated via a N-terminal α-bromoacyl moiety, while the C-terminus contains an α-thiocarboxylic acid (Baca *et al*., 1993), leading to the formation of a thioester linkage. This technique is called "native chemical ligation", since native amide bonds are created after reaction. In Kent laboratories was developed another technique to provide native amide linkages (Dawson *et al*., 1992) and following this methodology it was possible to synthesize HIV-1 protease (Baca *et al*., 1993) and to examine the insertion of a β-turn mimetic (Baca *et al*., 1993).
Further possibilities were opened by the expressed protein ligation (EPL) technique, a methodology that has become possible after the discovery of protein splicing (Gogarten *et al*., 2002; Muir *et al*., 1998). In principle, it is very similar to the native chemical ligation but exploits the advantages of the recombinant DNA technology to generate protein fragments via ribosomal synthesis, so that large proteins become accessible for chemoligation (Budisa, 2004).

1.4.3 Semi-synthetic Approaches

By linking functionally autonomous units, existing properties of the different molecules are retained and the combination of existing function let new hybrid properties to come out. The fusion of domains by chemical cross-linking in vitro allows the attachment of non protein moieties and attachment of domains to any exposed amino acid residue.

A successful example of the engineering of semisynthetic proteins is offered by the synthesis of hybrid proteins that can target sequence-specific DNA cleavage to selected sites within DNA and RNA. Schultz and coworkers succeeded in conjugating staphylococcal nuclease to either oligonucleotides (Corey and Schultz, 1987) or DNA binding proteins (Pei and Schultz, 1990) to make it suitable for sequence-specific DNA cleavage. In both experiments was exploited the strategy of introducing disulfide exchange as a point of unique reactivity for the linkage.

1.4.4 Site-directed mutagenesis

Through classical site-directed mutagenesis it is possible by codon manipulation to exchange one amino acid at a specific site with one another among the twenty of the basic pool. For example, it is possible to mutate AUG (Met) codon to UGG (Trp) codon using properly designed oligonucleotides in PCR reaction. This method was first used at the end of the 70´s (Hutchinson, *et al*., 1978) to study the relationship between the structure, sequence and function of proteins and to create proteins with enhanced favorable properties, like stability, catalytic function and binding specificity. The limits of this method consist in the narrow choice of amino acids that is limited to the twenty natural ones, thus the capability to incorporate non canonical amino acids into proteins has greatly increased the chances to create proteins with new properties.

By introducing a slight modification of the classical synthetic procedure is also possible to have at disposal a technique based on automatic synthesis of the target sequence in a way that yields to a randomly mutagenized preparation.

1.4.5 DNA Shuffling

By mimicking how genetic variations are performed by nature, it has become possible to develop the technique of DNA shuffling that mimics the process of homologous recombination. The target gene is randomly cleaved by DNase I in large fragments, the resulting duplexes are denatured and the single DNA strands are cooled down to the annealing temperature. The fragments are still able to

reform the original coding sequence and adding Taq DNA polymerase fills in the remaining gaps. In this process, the error rate of Taq DNA polymerase for small sequences is about 0.7%, so that the final resulting sequence differs from the original one in several point mutations. The manipulated gene is then inserted into a vector and a second round of digestion, reconstitution and ligation is performed, combining various mutations. After that, the pool of recombinant plasmids is introduced into an organism to select the clones that show altered properties. This technique has been employed to improve, among others, the function of β-lactamase (Stemmer, 1994) and Green Fluorescent Protein (GFP) (Crameri *et al*., 1996)

1.5 An expanded amino acids repertoire

1.5.1 Terminology

The 20 standard amino acids that are incorporated into proteins due to the natural assignment of amino acids to triplet codons can be defined as canonical amino acids. Other residues like selenocysteine and cystine belong also to the natural proteinogenic amino acids, as they derive from posttranslational or cotraslational modifications of the "basic" residues. There is also another category that includes those amino acids outside the standard set that can be incorporated in a codon-dependent manner. These are called non canonical amino acids and their entry in the genetic code can be achieved experimentally ("expanded" or "second" genetic code). Other terms, like non standard, non natural, unnatural that are often used in the literature to indicate these amino acids are not used in this study.

1.5.2 Selective Pressure Incorporation

The simplest way to incorporate amino acid analogues into proteins is based on the use of auxotrophic bacterial strains. These bacteria are deprived of the capability to produce on their own one or more amino acids.

One of the first problems encountered for the substitution of a canonical amino acid with an analogue counterpart is that almost all amino acidic substances outside the natural pool and metabolic intermediates hamper the cellular growth, with SeMet being an exception to this rule (Budisa *et al*., 1997). This problem could be circumvented by using auxotrophic mutant bacterial strains starved for a specific canonical amino acid and later "fed" with the chosen analogue with concomitant induction of the protein synthesis (Sykes *et al*., 1974, Hortin and Boime, 1983).

Cowie and Cohen (Cowie and Cohen, 1957) reported for the first time the successful incorporation of selenomethionine (SeMet) in all cellular proteins, using an *E. coli* Met-auxotroph mutant strain. As the Met supply for the cellular growth was experimentally defined, it was possible to build up a new "scheme" for the genetic code expansion, based on synthetic media, where the non canonical analogue takes the place of the canonical amino acid. The general principle that comes out from all the experiments that have been performed in this field is that the amino acids composition of proteins can be influenced by imposing an external pressure (i.e. with environmental changes).

The labeling of a single target protein could be achieved only after the advent of the recombinant DNA techniques that allowed extrachromosomal expression of target proteins (Cohen *et al*., 1973). As a consequence, the incorporation of the amino acid surrogate can be performed at single target proteins avoiding harmful global effects on the expression host.

For a successful experiment using the SPI methodology, some criteria have to be fulfilled. The first is the choice of the proper cells in combination with an efficient and controllable expression system like T7 or T5 promoter/polymerase; the second is the possibility to control the fermentation conditions in the defined minimal media and the third concerns the capability to exercise an external selective pressure, i.e. to manipulate amino acid supply or fermentation conditions, for the amino acid replacement (i.e. sense-codon(s) reassignment) at the level of a single protein (Budisa *et al*., 1999; Kiick *et al*., 2000, Minks *et al*., 2000, Ross *et al*., 1997)

The fermentation is usually performed in minimal media supplied with limiting concentrations of the native amino acid that has to be substituted. During the cellular growth, the native substrate is consumed till its complete depletion. At this point, which corresponds to a plateau in the growth curve representing the cellular optical density, the non canonical amino acid is added and the target gene, kept silent until now, is induced. From this point onwards, the cells stop growing and their only task is to produce the desired modified protein.

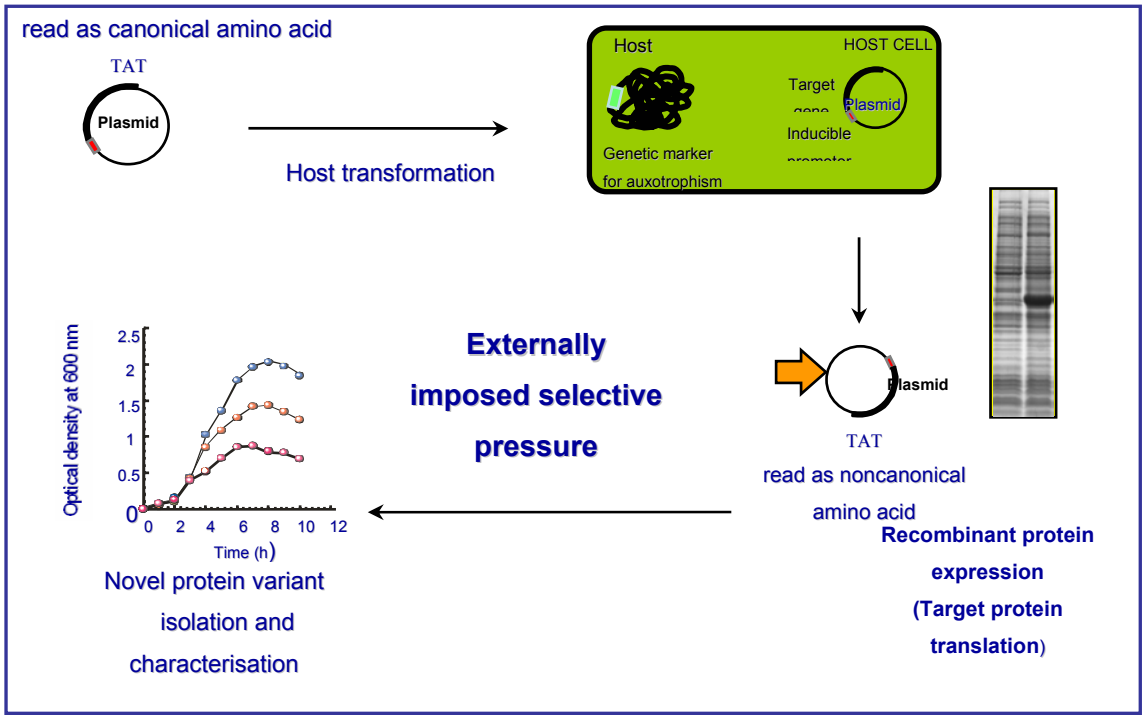


Figure 1.5. The principles of the SPI-method.: i) choice of the proper cells in combination with an efficient and controllable expression system ii) the possibility to control the fermentation conditions in the defined minimal media iii) the capability to exercise an external selective pressure, i.e. to manipulate amino acid supply or fermentation conditions, for the amino acid replacement. (Modified from Minks *et al.*, 1999)

However, not all amino acid analogues can be incorporated by using this *in vivo* methodology. Besides the problems due to toxicity, another feature has to be kept in mind: that SPI method exploits the lack of absolute substrate specificity of aminoacyl-tRNA synthetases (aaRS). The latter are enzymes involved in the activation of the cognate amino acids and in carrying them to the tRNAs, therefore aaRS play a crucial role in the protein synthesis. The only way for the non canonical amino acids to enter the protein synthesis is to exploit the same routes of their canonical counterparts, which are among others, activation, undergoing proofreading and editing mechanisms. This explains why it is so important that the amino acid surrogates show a similar shape and size to natural ones, so that they are not "rejected" from the accurate mechanisms that are active during the protein synthesis.

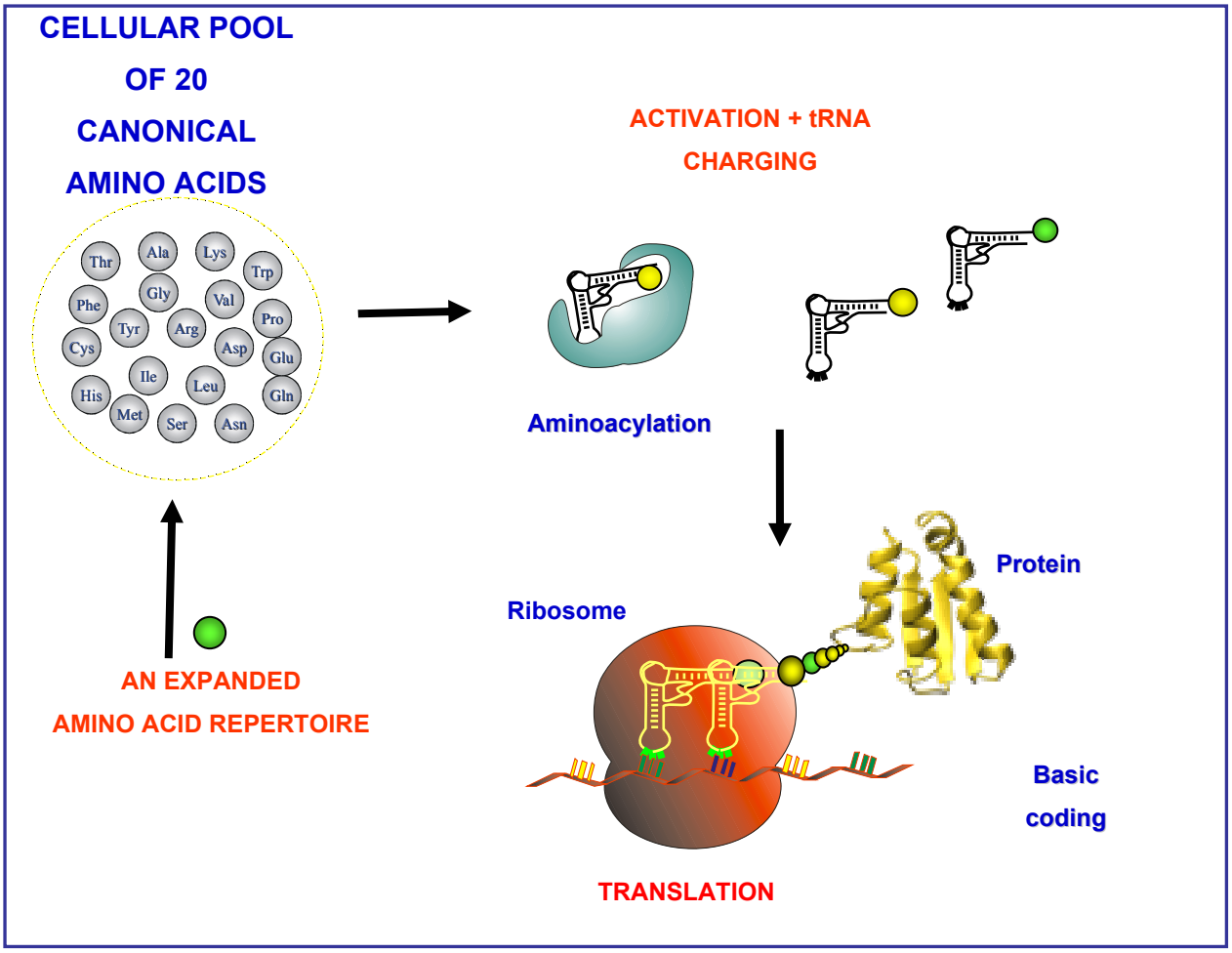


Figure 1.6. Entry of non canonical amino acids into protein synthesis. The SPI method exploits the lack of absolute substrate specificity of aminoacyl-tRNA synthetase to discriminate between chemically and sterically similar amino acids. In this way the standard amino acid repertoire can be enriched by non canonical substrates.

The main advantages of the SPI method are the possibilities to design residue–specific substitutions, to have at disposal a relatively simple, reproducible methodology and to reach great levels of protein expression (comparable to wild type proteins). This last property is very important to perform further structural biological analyses such as NMR-spectroscopy or X-ray crystallography, where a considerable amount of protein is needed or to plan the production of therapeutic proteins on an industrial scale.

1.5.3 How Nature introduces novel amino acids: lessons from selenocysteine

Selenocysteine is a proteinogenic amino acid that is cotraslationally incorporated into proteins in the presence of the codon UGA assisted by a distinct elongation factor and a structural signal in the mRNA (Bock *et al*., 1991; Commans *et al*., 1999). The modification from serine to selenocysteine involves the participation of a special tRNA that possesses the anticodon UCA capable of recognizing the codon UGA. This tRNA becomes acylated by a synthetase specific for serine, and it is then enzymatically converted into selenocysteinyl-$tRNA^{Sec}$ in a two step reaction through the formation of aminoacrylyl-$tRNA^{Sec}$ and the subsequent addition of hydrogen selenide to the aminoacryl residue with release of Sec-$tRNA^{Sec}$. This reaction is catalyzed by the *E.coli* SelD gene product and the selenium donor is selenophosphate. Sec-$tRNA^{Sec}$ is finally bound to the selenocysteine-specific translation factor SELB, which recognizes an mRNA hairpin structure next to the UGA selenocysteine codon on the ribosome (Ibba and Söll, 2000).

1.5.4 Suppressor based methodologies

It has been demonstrated very early by Chapeville and coworkers (Chapeville *et al*., 1962) that tRNAs charged with non-cognate amino acids are able to perform their role of adaptors in the protein synthesis. This discovery has been later exploited for the generation of suppressor based methodologies in order to expand the number of analogues that can be introduced into proteins.

In 1989, Schultz and coworkers developed the amber suppression method for introducing non canonical amino acids into proteins at a specific position. Amber suppressor tRNAs are aminoacylated with non canonical amino acids through chemical aminoacylation and the resulting mischarged aminoacyl-tRNAs are added to an in vitro translation system or in vivo i.e. to *Xenopus* oocytes (Novak *et al*., 1995) together with a mRNA or DNA of interest containing an amber stop codon (UAG) at a desired position. The amber codon acts no more as stop codon, but is suppressed by the misacylated aa-tRNA leading to the incorporation of the non canonical amino acid into the directed position of the target protein.

This method suffers from drawbacks as the UAG stop codons that should “code” for the non canonical amino acid are in competition with the release factors that terminate the protein elongation. Second, among three stop codon only amber can be used for the suppression that means that a multiple incorporation of different non canonical amino acids at the same time is not to be achieved. Moreover, once an aa-tRNA is used for translation, the deacylated tRNA will not be

aminoacylated again either in the cell-free translation system or *Xenopus* oocytes, decreasing the yield of non-natural proteins.

On the way to establish a universal strategy for expanding the genetic code in living cells, the major bottleneck is found at the step of tRNA amino-acylation, as natural aminoacyl-tRNA synthetases have only a limited capability to activate non cognate amino acids and to attach them to tRNAs (Jakubowski & Goldman, 1992). In addition, the concept of "orthogonality" has to be fulfilled. Orthogonality was first defined for $tRNA^{Gln}$ (Liu *et al*., 1997) and refers to the lack of cross-aminoacylation between the engineered synthetase-tRNA pair and the natural host's own tRNAs and synthetases (Giege´, 2003). If this is not achieved, then the suppressor tRNA, once it has delivered its unnatural amino acid to a protein, will be charged with a canonical amino acid and return to the protein synthesis cycle.

Several methods to generate orthogonal suppressor tRNA-synthetase pairs have been developed in *E. coli*. One strategy consists in changing an existing *E. coli* tRNA´s affinity toward its cognate synthetase by mutating nucleotides at the tRNA-synthetase interface, while a mutant synthetase that uniquely recognizes the orthogonal tRNA is evolved (Liu et *al*, 1997). A second method consists in importing a tRNA-synthetase pair from another organism (i.e. *Saccharomices cerevisiae*) into *E. coli*, if cross-species aminoacylation is inefficient (Pastrnak *et al*., 2000).

A third strategy utilizes a heterologous synthetase as the orthogonal synthetase but a mutant initiator tRNA of the same organism or a related organism as the orthogonal tRNA (Kowal *et al*., 2001)

Recently, the attention has been focused also on "sense codon suppression", using mRNA display to identify sense codons that could be efficiently suppressed. The ability to suppress sense codons would enable the facile expansion and potential rewriting of the genetic code with non canonical amino acids and would allow the creation of original polymers and chemically diverse display libraries (Frankel & Roberts, 2003)

Site-specific introduction of non canonical amino acids into proteins in mammalian cell has also been reported (Sakamoto *et al*., 2002) for the incorporation of 3-iodo-L-Tyrosine at amber position. A specific *E. coli* Tyr-tRNAsynthetase variant was expressed in mammalian cells, together with the *Bacillus stearothermophilus* suppressor $tRNA^{Tyr}$, while 3-iodo-L-Tyrosine was supplied in the growth medium.

1.5.5 Extension of codon-anticodon pairs

Alternatively to the amber stop codon, another method based on four and five-base codon suppressors has been introduced (Hohsaka *et al*., 2001; Sisido and Hohsaka, 2001). The tRNA containing the complementary four-base anticodon and the non canonical amino acid decodes the four-base codon that is translated into the non canonical amino acid, while the correct reading frame is maintained. In the case that quadruplet is not decoded, that means if only the first three bases are read leading to the incorporation of a canonical amino acid, the translation will stop at a subsequent stop codon.

The frame shift suppression strategy shows two advantages comparative to amber suppression. First, competition between the suppressor tRNA and the endogenous release factors is efficiently decreased. Second, multiple non canonical amino acids incorporation is possible in the same protein, by using two independent four base codons (Hohsaka *et al*., 2001).

Recently, Hirao and coworkers, succeeded in incorporating *in vitro m*-chlorotyrosine in a human RAS protein, using non natural codon-anticodon pairs (Hirao *et al*., 2002). A pyridine-2-one was introduced in mRNA and a 2-amino-6-(2-thienyl) purine was introduced as matching anticodon in the tRNA charged with *m*-chlorotyrosine. This codon-anticodon non natural pair combines the concepts of hydrogen-bonding patterns and shape complementarities.

1.5.6 Use of missense suppression: breaking the degeneracy of the genetic code

"Breaking the degeneration of the genetic code" was presented as a possible solution of the problem encountered with other methods that is that the non canonical amino acid has to share codons with one amino acid of the basic pool (Kwon *et al*., 2003). In earlier experiments, the protein biosynthetic machinery of Phe was chosen as target, as the amino acid Phe is encoded by two codon triplets (UUC and UUU) which are read by a unique tRNA that bears the anticodon sequence GAA. Therefore, a non natural $tRNA^{Phe}_{AAA}$ was synthesized and misacylated with the analogue L-3-(2-naphtyl) alanine, as it was thought that $tRNA^{Phe}_{AAA}$ could read UUU codon faster than wild-type $tRNA^{Phe}_{GAA.}$ This fact could be exploited to allow codon-biased incorporation of the analogue at multiple sites in recombinant proteins.

Murine dihydrofolate reductase, which contains nine Phe residues, was chosen as target protein and the ratio of mass spectrometric analyses indicates that the incorporation of the non canonical amino acid L-3-(2-naphtyl)-Alanine is dominant, even if not unique, at the UUU codon.

1.6 Model proteins

Proteins used for replacement studies share some common features, as known three dimensional structures and well-known biochemical, biophysical and kinetic properties. Moreover, they can be easily and efficiently expressed in auxotrophic host strains and subsequently recovered in soluble form or refolded from inclusion bodies. Model proteins used in this study are human recombinant Annexin V, Barstar and EGFP and they all fulfill the criteria mentioned above.

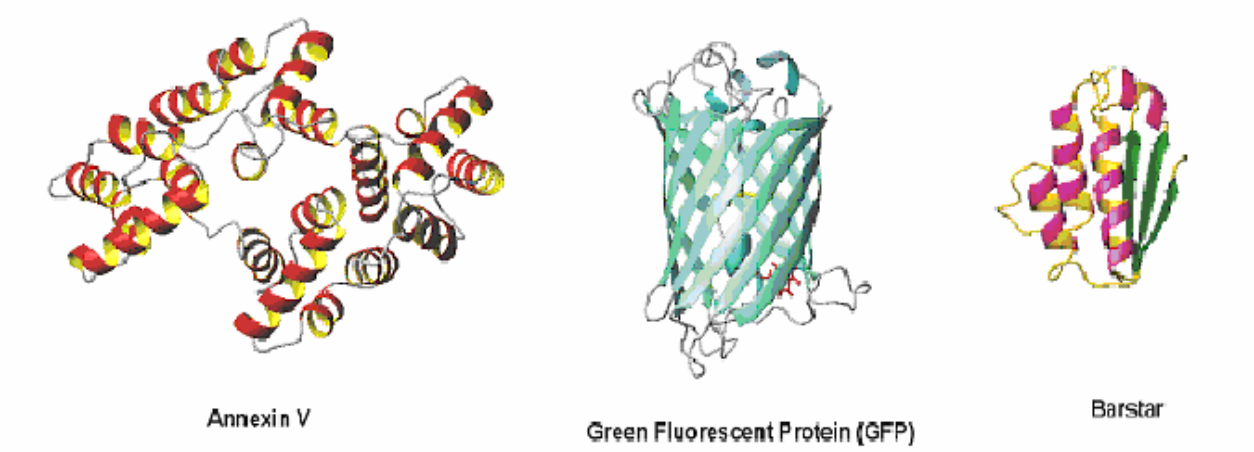


Figure 1.6. Three-dimensional structures of human of human recombinant Annexin V (left), "Enhanced" Green Fluorescent Protein (EGFP, middle) and wt-Barstar (right).

1.6.1 Annexin V

Annexin V is a water soluble protein, which binds to negatively charged phospholipids in a calcium-dependent manner. This protein turned out to be an useful model for incorporation of various amino acid analogues. In this PhD work, the attention has been particularly focused on the amino acids Tryptophan and Leucine.

Annexin V contains one single Tryptophan residue at position 187, which is also involved in the Ca^{2+}-dependent membrane binding and related channel activity, (Huber *et al*., 1990, Budisa *et al*., 1998). The presence of only one Tryptophan residue offers the opportunity to perform substitutions that are at the same time residue-specific and position-specific.

4- and 7-fluorotryptophan were chosen as target analogues in order to proof the changes induced by a single-atom-exchange within the rest of the macromolecule. Fluorine offers isosteric substitution of hydrogen, but has different chemical and physical properties (i.e. weak hydrogen bond acceptor, strong electro negativity) that might induce changes in the entire protein, upon its incorporation.

In other incorporation experiments, amino acid analogues 4- and 5-aminotryptophan have been used. The amino group on the indole ring presents interesting features, such as intra-molecular charge transfer due to the lone electron pair on the nitrogen atom that may deliver new properties to the whole protein. Moreover, from the introduction into a protein of a strong polar group, may arise new interactions that can either enhance or impair the stability of the protein itself, giving a chance to study stability upon mutations.

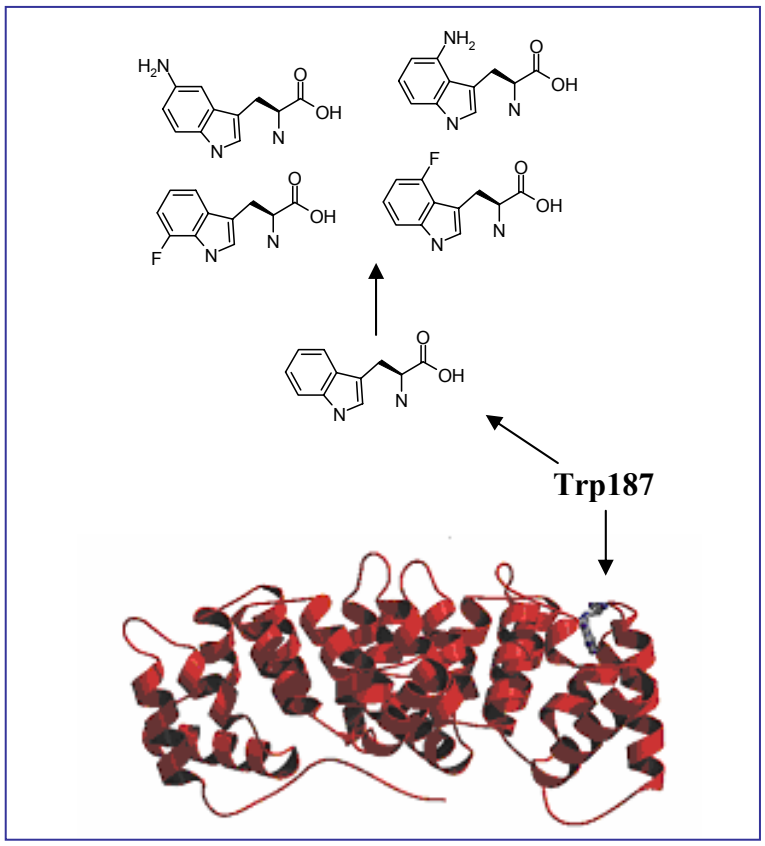


Figure 1.7. 3D structure of model protein Annexin V and Trp analogues used in this study for incorporation experiments are shown.

Annexin V contains in its sequence 35 Leucine residues, uniformly distributed in the protein structure, which offers a possibility to globally fluorinate the protein. Fluorinated derivatives of the canonical amino acids might turn out to be very suitable tools for a rational re-design of natural structural scaffolds in order to obtain structures endowed with predetermined and unusual functions (Budisa *et al*., 2004). This fluorination attempt has been performed with the non canonical amino acid 5´,5´,5´,trifluoroleucine (TFL).

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 (Kukhar,.*et al*., 1991)

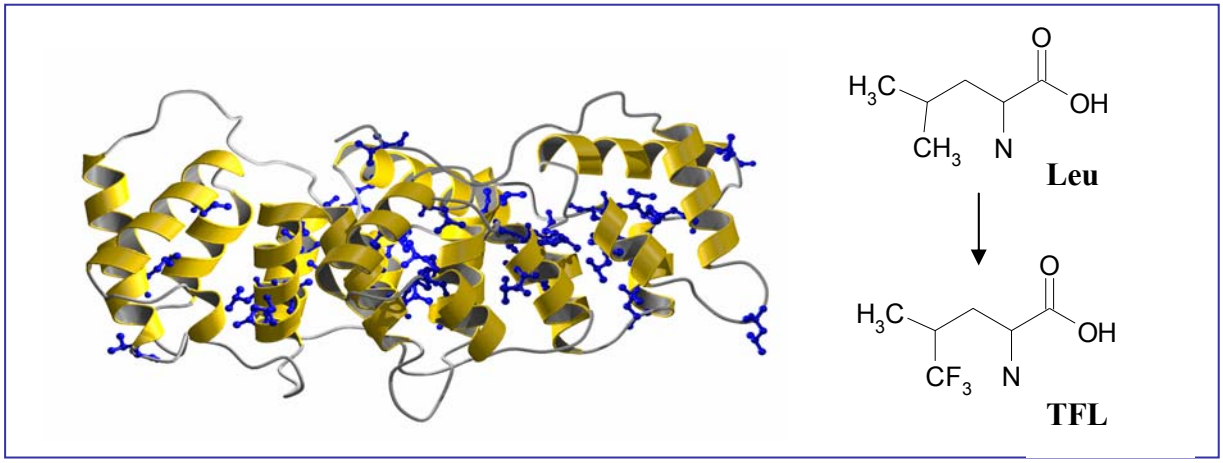


Figure 1.8. TFL was used to replace Leucine residues in Annexin V in the attempt to globally fluorinate the protein core. Leucine are here coloured in blue to show their distribution in the protein structure.

1.6.2 Enhanced Green Fluorescent Protein

Enhanced Green Fluorescent Protein (EGFP) is a derivative of GFP from the jelly-fish *Aequorea Victoria* (*av*GFP) and consists in eleven β-strands forming a barrel with the chromophore buried in the protein interior. The chromophore is completely encoded in the amino acid sequence and is autocatalitically formed during a post-translational reaction between the side chains of residues 65-67 with the participation of molecular oxygen (Shimomura, 1979). Therefore, the effects of amino acids substitutions at position 65, 66 or 67 can be immediately detected as they directly affect the chromophore.

Moreover, another great advantage of GFP as a reporter of gene expression is its high level of stability. This extreme stability seems to be the consequence of the unique three dimensional structure of the protein itself, as the chromophore is positioned near the geometric center of the β-barrel and is protected by the tight strands. Short loops region and distorted α-helices cap both ends of the barrel leading to a complete segregation.

In our laboratories, Enhanced Cyan Fluorescent Protein (ECFP) and EGFP have been used for different replacement experiments. ECFP contains two Trp residues, at position 66 (in the chromophore) and at position 57, while in EGFP sequence position 66 is occupied by Tyrosine.
Tryptophan residues were replaced in both proteins with fluorinated, methylated, hydroxylated and aminated analogues, so that the contribution of the two different Trp residues to spectroscopic and thermodynamic changes could be studied.
As EGFP contains also eleven Tyrosine residues in its sequence (Tyr 66 in the chromophore), the replacement of Tyrosine with fluorinated analogues has also been performed to study the effects of fluorination in chromophore and its environment (Budisa *et al*., 2004). Moreover, EGFP substituted with (3-F)-Tyr was also used to probe its potential use for cytotoxic experiments towards target cells (Data not published). Indeed, 3-Fluor-Tyr is known to be a strong inhibitor of the Krebs cycle, as Tyrosine is a precursor of fumarate, which is a natural substrate in the Krebs cycle. Therefore, the fluorinated compound blocks the final steps of the cycle, causing cell death.
Another interesting target to attempt global fluorination of EGFP is offered by Leucine, as the protein contains 15 Leu residues, uniformly distributed in its structure. As in the case of Annexin, the experiment was performed with TFL, to compare the capability of both proteins to accommodate the bulky fluorous substituents in their interior (see Fig.1.9).

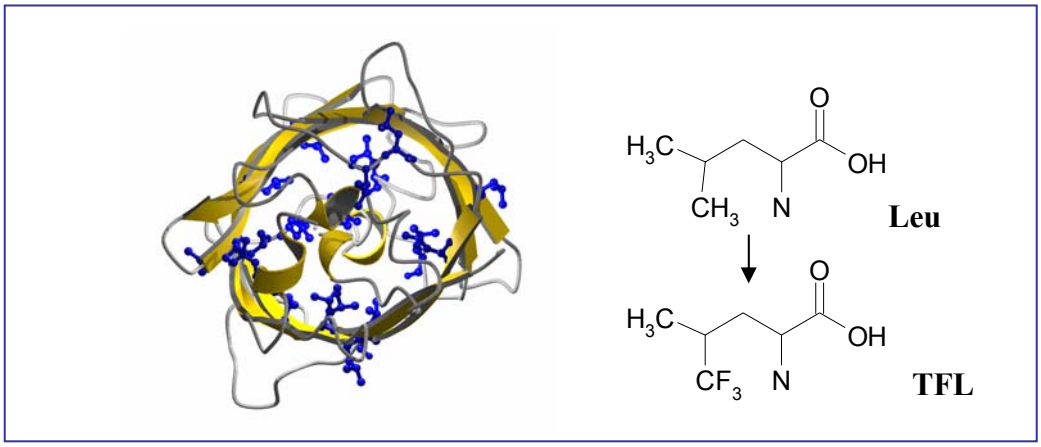


Figure 1.9. The canonical amino acid Leucine was replaced with TFL for incorporation studies in EGFP. The target protein is here proposed in its axial perspective, to show the location of Leucine residues (blue) all over the protein.

1.6.3 Barstar

Barstar is a single domain intracellular inhibitor of ribonuclease barnase from *Bacillus amyloliquefaciens.* The triple mutant C40A/C82A/P27A, obtained through classic site-directed mutagenesis, resemble the native protein but lacks of complications due to oxidation of cysteins and *cis-trans* isomerisations of the peptidyl-Pro27 bond (Golbik, 1999). This mutant is particularly suitable for folding and stability studies. As Barstar contains three Tryptophan residues at positions 38, 44 and 53 (see Fig.1.10), the suitability of many different Trp analogues for incorporation has been tested in this PhD thesis. Most emphasis has been devoted to probe the effects of fluorinated and aminated Tryptophan analogues. Trp53 is buried in the protein core, plays a crucial role in determining the initial phases of the protein folding and cannot be replaced by any of the standard amino acids using the classical site directed mutagenesis (Nath *et al*., 1997).

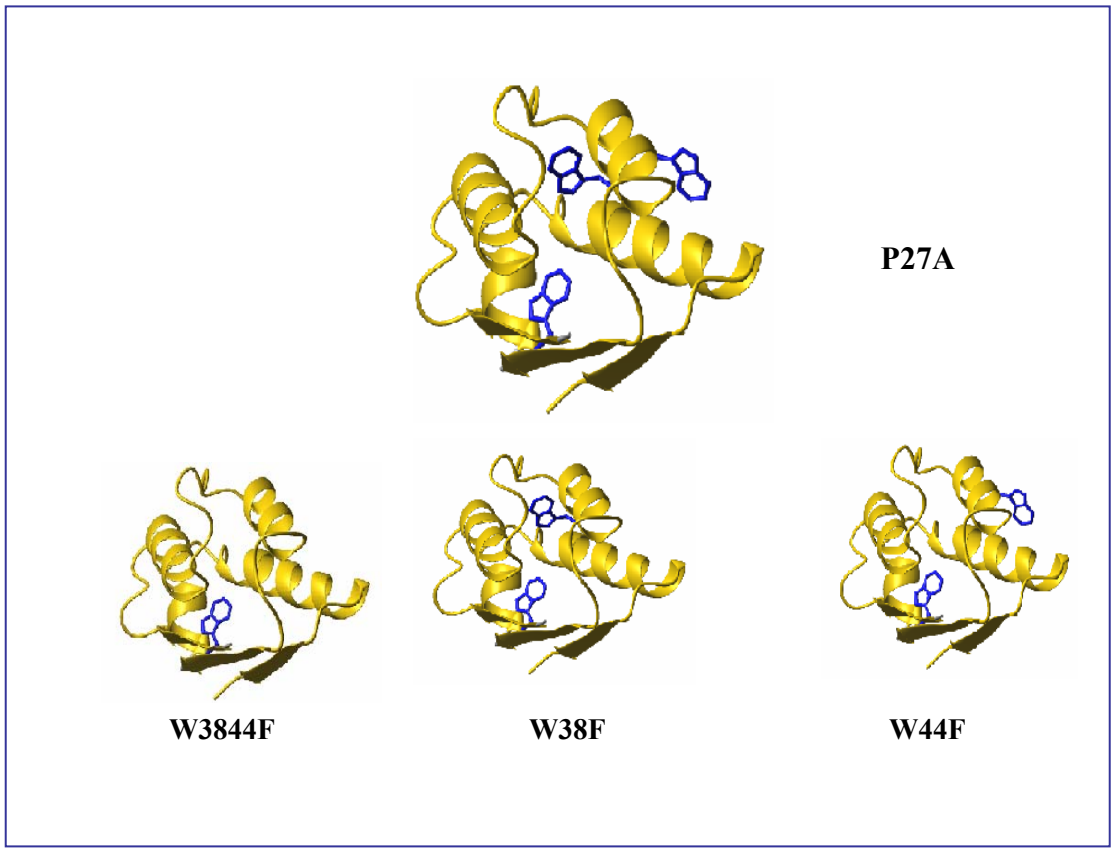


Figure 1.10. Three-dimensional structure of wt-Barstar and its mutants with related Trp residues represented as blue sticks. Trp 38 is partially exposed to the solvent, Trp 44 is fully exposed to the solvent and Trp 53 is packed in the hydrophobic protein interior

1.7 Tryptophan as target for protein engineering

Tryptophan is an attractive amino acid for replacement studies (see Fig. 1.11) in various proteins, as it offers many advantages:

- It is relative low abundant in proteins (1.2%) and is encoded by a single coding triplet, so that its replacement is in many cases almost site-specific
- It is represented by a single coding triplet (UUG)
- The indole chemistry (natural and synthetic) is rich and diverse, offering in this way a large number of related amino acid analogues and surrogates as potential substitute for protein synthesis
- It shows biophysical properties that promote its participation in different interactions in and between proteins (π-π stacking, hydrogen bonding, cation-π interactions)
- It is the principal source of UV absorbance and fluorescence emission
- It plays crucial roles in maintaining protein stability and folding, since it is often part of stabilizing hydrophobic clusters in many proteins
- It is involved in many biological activities (i.e. antigen-antibody binding, membrane binding, voltage gating, substrate recognition, etc.)

The first attempts to incorporate Trp analogues into proteins were performed by Pardee (Pardee *et al*., 1956) and Brawerman and Ycas (Brawerman & Ycas, 1957) who introduced 7-azatryptophan and 2-azatryptophan into *E. coli* proteins. Recently, selenophene-and thienyl-containing Trp analogues have showed interesting properties as pharmacologically active substances and markers for X-ray crystallography of proteins (Bae *et al*., 2001; Budisa *et al*., 2001). As it will be discussed later in this study, the introduction of aminotryptophan analogues into proteins has also opened a further possibility, which enables to design protein-based pH sensors (Budisa *et al*., 2002).
Tryptophan plays also a crucial role in the biosynthetic pathway of hormones such as serotonin and melatonin, therefore the employment of suitable analogues might turn out to be an interesting tool for the design of antagonists, drugs and antibiotics.

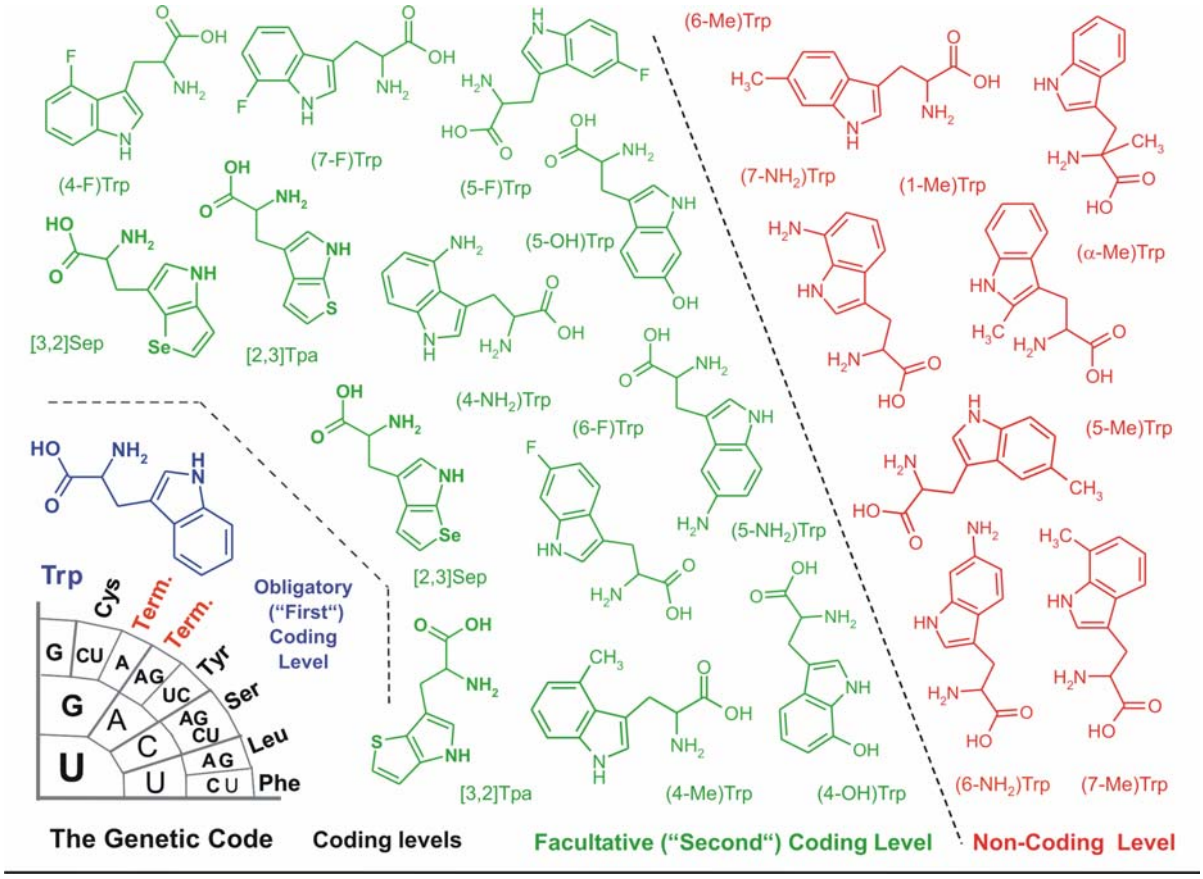


Figure 1.11. An expanded genetic code for Tryptophan. The "first" or obligatory coding level (blue) refers to the established assignment of Tryptophan for UGG coding triplet. The facultative ("second") coding level (green) includes all those Trp analogues that can be incorporated into target proteins as a response to UGG triplet. The non coding level (red) is represented by Trp analogues for which incorporation attempts were not successful. (Taken from Budisa, 2004)

2. MATERIALS AND METHODS

2.1. Materials and instruments

Most of the chemicals that were used in this study were purchased by the following companies: Sigma-Aldrich, Lancaster, BD Biosciences, GIBCO, Molecular Probes, Roche, Invitrogen, NEB Ferment Böhringer-Mannheim, Biomol, Difco Laboratories, Roth, Serva, Qiagen, National Diagnostics and Fluka.

- Amicon, Beverly, MA, USA:

  Microconcentrators Centricon 10 and 5

  Microconcentrators Centriprep 10 and 5

- Beckmann, Munich:

  Centrifuge J2-21 (rotors JA-20 and JA-10)

  Ultracentrifuge L7-55 (rotors Ti 45 and Ti 55.2)

  UV/VIS-Spectrophotometer DU 7500

- Biometra:

  Electrophoreses chambers

- Biosynth AG, Basel, Switzerland

  5- and 6-Chloro-Tryptophane

- Branson:

  Ultrasound instrument (Sonifier 250)

- Charles Supper Company

  CrysChem Plates

- Eppendorf, Netheler and Hinz GmbH, Hamburg:

  Eppendorf tubes (1.5 ml and 2 ml)

  Micropipettes 10 µl -5 ml

  Tablecentrifuge 3200

- Greiner and Söhne GmbH & CoKG, Frickenhausen:

Petri dishes

- Hellma:

  Precision cuvettes Quartzglass SUPRASIL®

- Infors AG, Bottmingen, Switzerland:

  Shaker incubator RKF-125 for culture preparation

- Jobin Yvon, Lonjumeau, France, JASCO:

  CD-Spectrophotometer

- MarResearch Hamburg:

  Image Plate System MAR300

- Merck, Schott (Duran):

  Glassware

- Millipore:

  Bottle top filters

- New England Biolabs, USA:

  Broad Molecular Weight Marker

- Perkin Elmer Weiterstadt:

  UV/VIS-Spectrophotometer Lambda 17, Fluorescence Spectrometer, Luminescence Spectrometer (LS50B) with digital Software

- Pharmacia Biotech, Freiburg:

  CM52 Sepharose column

- Qiagen, Hilden:

  Qiaprep Mini- and Midikits

- Radiometer Copenhagen, Kopenhagen, Denmark:

  PHM 82 standard pH-meter

- Roth, Karlsruhe:

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

- Sartorius AG, Göttingen:

  Laboratory balances LC2200S

- Waters

  Analytical HPLC: system assembly composed of two 515 HPLC pumps, a pump control Module, a 717 plus Autosampler, a 966 Photodiode array detector. The software connected to the system was the Program Millennium version 3.00.

  *Column X-Luna-TM- C_{18} 5µm 150*46mm*

  *Column X-Terra-TM-C_8 5µm 3.9*150mm*

2.2 Enzymatic preparation of Tryptophan analogues

2.2.1 Plasmids, host strains and expression conditions

The *E. coli* strain *CB149* for Tryptophan-synthase was provided by generous help of Prof. E. Wilson-Milles, and was transformed either with the vectors pSTB7 or pEBA10. The cells were allowed to grow and express in Vogel Bonner medium at 37 °C overnight.

2.2.2 Purification of Trp synthase

The cells were harvested by centrifugation (4200 rpm, 4 °C, 30min) and the pellet of collected cells was washed once by resuspension in 30 ml of 0.85 % NaCl followed by centrifugation (5000 rpm, 4 °C, 15 min). After that, cells (wet weight around 12 g) was resuspended in 60 ml of buffer T (0.05 M Tris-HCl pH 7.8, 5 mM EDTA, 10 mM 2-mercaptoethanol and 0.1 mM pyridoxal phosphate) and disrupted by treatment with ultrasonic (macro tip, output 4, duty 20 %, seven times for 1 min with 4 min break in between, on ice). The extract (cell lysate) produced in this way was centrifuged for 20 min at 27000 rpm 4 °C. The supernatant solution (around 60 ml) was saved, and the precipitate discarded. Afterwards, 0.5 M spermine in 30 % (w/v) polyethylene glycol 8000 in buffer T was added to the supernatant (with rapid stirring) to give a final concentration of 6 % PEG 8000 and 5 mM spermine (the spermine precipitate some of the other proteins). The suspension was immediately centrifuged for 5 min at 27000 rpm 4 °C, and the supernatant solution was stored for 4 days at 4 °C.

Yellow crystals precipitated during this period and were collected by centrifugation for 15 min at 27000 rpm 4 °C, and the supernatant solution was decanted. The precipitate was resuspended in 15 ml of buffer B (0.05 M sodium N, N-bis (2-hydroxyethyl) glycine pH 7.8, 1.0 mM EDTA,

0.02 mM pyridoxal phosphate, and 10 mM 2-mercaptoethanol) by stirring for 10 min at 37 °C and finally dialyzed overnight against buffer B.

The enzyme activity was checked with 50µl of 0.5 M L-serine, 25 µl of 0.1 M indole and 50 µl 20 mM PLP in 735 µl of 100mM Tris pH 7.8. The reaction was performed in a quartz cuvette and monitored for 2 min at 295 nm.

2.2.3 Reaction of Tryptophan synthase

The enzyme Tryptophan synthase catalyzes the condensation of indole (or indole derivatives) with L-Serine to L-Trp in a two-step reaction. The α and β subunities that form the tetramer are involved in different reactions (Miles, 1977). The catalyses begins with the binding of 3-indolyl-D-glycerol-3`-phosphate (IGP) to the α–site, and the reaction of L-Ser with Pyridoxalphosphate (PLP) to form the α-aminoacrylate Schiff base, at the β-site. The indole produced by cleavage of IGP diffuses to the β-site through the tunnel that connects the two subunities, to react with the α-aminoacrylate Schiff base. This reaction lead to the formation of a quinoide intermediate, which is then protonated to form the almidine of the product L-tryptophan.

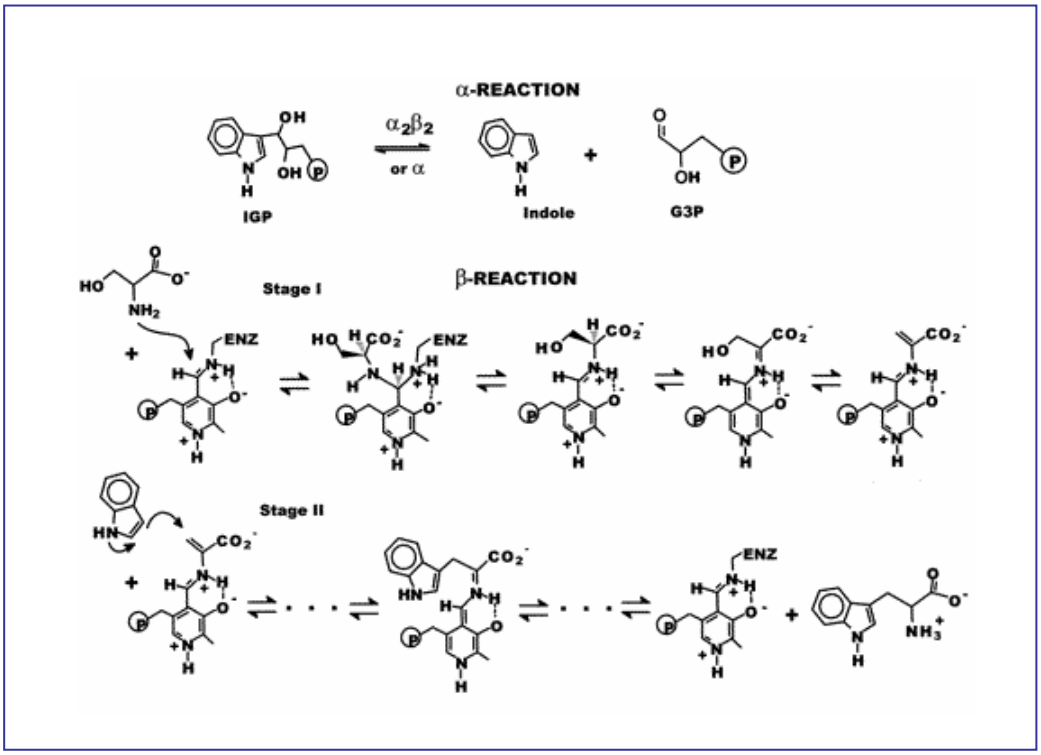


Figure 2.1. The synthesis of Tryptophan performed by the enzyme Tryptophan synthase. The reaction proceeds through a series of PLP intermediates. (Modified from Weber-Ban *et al*., 2001)

The synthesys of 4-, 5-, 6-, and 7-aminotryptophan, 4- and 5-hydroxytryptophan, 7-fluorotryptophan, and 7-bromotryptophan, was performed according to the following protocol:
35 ml of argon saturated distilled water, 5 ml 1M TrisCl pH 7.8, 2 ml 5M NaCl, 5 ml 20mM Pyridoxalphosphate, 2 ml 0.5 M L-Serine and 1 ml 100μM enzyme were mixed up in a round flask and the reaction was started by addition of 100 mg of specific indole anologue.The reaction mix was incubated overnight by gentle shaking in dark at 37 °C.
The completeness of the reaction was checked by analytical HPLC and the enzyme was subsequently separated from the solution by centrifugation at 3500 rpm by using Centricon membrane with molecular weight cut off of 10 KDa.
The Trp analogues and surrogates prepared in this way, were stored under argon at -20 °C.

2.2.4 Isolation of pure aminotryptophans

In order to determine the spectral properties of new synthetized Tryptophan analogues and to perform further kynetical experiments, some efforts were done to isolate these substances in pure solid form. Indeed, after separation of the enzyme from the reaction mixture, a certain amount of reagents and cofactors are still present in the solution. For isolation of (5-NH_2)- and (6-NH_2)Trp was adopted the following protocol. After separating the enzyme, the excess of aminoindole was removed by three extractions with ethyl acetate. As confirmed by Thin Layer Chromatography (TLC) and HPLC spectra, aminotryptophan remains in the water solution while aminoindole enters into the organic phase. The water phase was then evaporated up to 5 ml volume, put on an ice bath and the pH was adjusted with NaOH to pH values between 8.0 and 8.4. When the isoelectric point (pI) is reached, aminotryptophan precipitates as at the pI, the amino acid displays a minimal water solubility. The solution containing the solid aminotryptophan was then centrifuged, the powder has been washed with cold water and subsequently dissolved in water (at room temperature) and lyophilized. The identity and the purity of the compound was confirmed by NMR analyses.

2.3 Attempts to chemically synthesize 4-aminotryptophan

(4-NH_2)Trp could not be precipitate in pure form from the solution as it was done for (5-NH_2)Trp and (6-NH_2)Trp, perhaps because of its much higher solubility. For this reason, some attempts were made in order to synthesize (5-NH_2)Trp using classical synthesis protocols This was done by modification of the synthetic protocol used to generate β-(1-Azulenyl)-L-Alanine (Loidl *et al*., 2000).

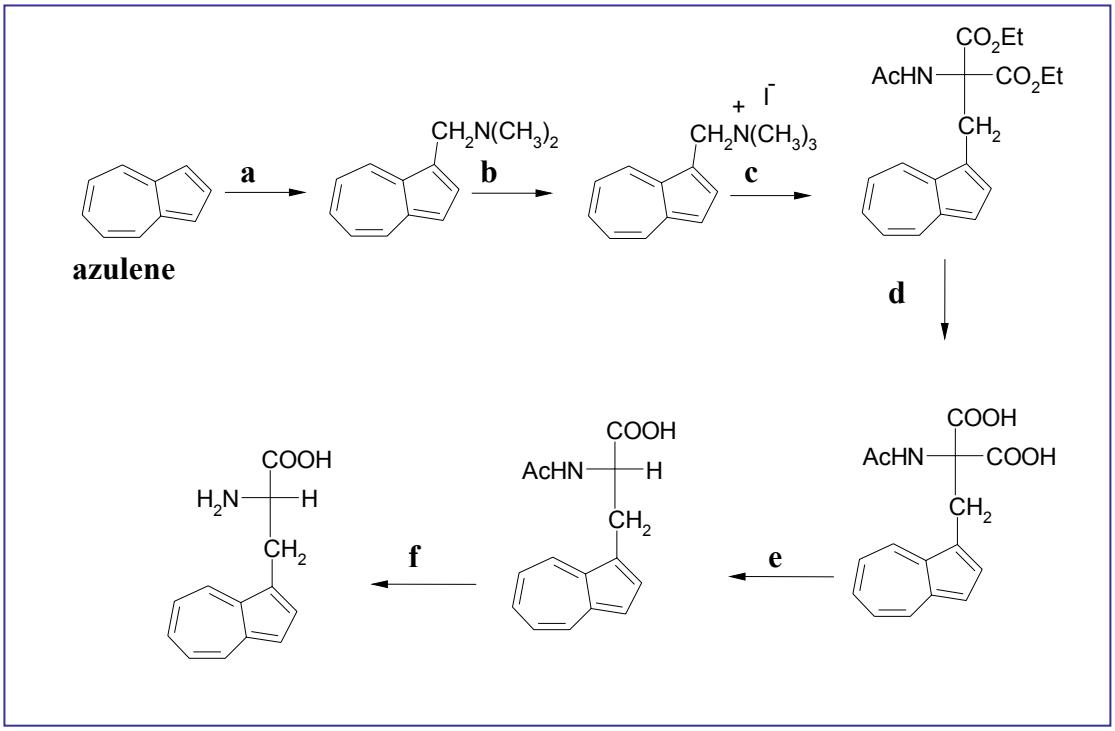


Figure 2.2. The synthesys of β-(1-Azulenyl)-L-Alanine, as it is described in the literature. **a**) $(CH_3)_2NCH_2N(CH_3)_2$, HCHO, AcOH, $CH2Cl_2$; **b**) MeI, EtOH; **c**) $AcNH(CO_2Et)_2$, NaOEt, EtOH, 2 h reflux under an argon atmosphere; **d**) 10% KOH in 50% EtOH, 3 h reflux; **e**) 0.2 N HCl in 15% THF, 8 h reflux under an argon atmosphere; **f**) enantioselective deacetylation with acylase I from *Aspergillus melleus.*

The first step was to protect the amino group at position 4 on the indole ring. For this purpose 590 mg (4-NH_2) indole (4.64 mmol) as starting product were dissolved in 5 ml pyridine and one equivalent Z-Cl was added to the solution. From this reaction, 1 g. (3.31mmol) Z-Cl-(4-NH_2)indole was obtained. Z-Cl-(4-NH_2)indole was then dissolved in 15 ml CH_2Cl_2 and put on ice. At the same time, 0.5 equivalents (66 mg) p-Formaldehyde and 0.5 equivalents Tetramethyldiaminomethane (300µl) were added to 15 ml glacial acetic acid and the solution was heated until it became clear. The acid acetic solution was cooled, added dropwise to the Z-Cl-(4-NH_2)indole containing solution and the mixture was stirred for two hours. After this time, 60 ml H_2O and 30 ml aqueous HCl 5% were added to the solution and the aqueous phase has been washed four times with dichloromethane. The combined organic layers were reextracted with water and the aqueous phases were adjusted to pH 12 with NaOH. Water was then extracted five times with diethylether. The combined ether layers were washed with water and dried over sodium sulphate and evaporated. The final yield is of 560 mg (1, 73 mmol).

The oily product was solved in 50 ml dry ethanole and 4 ml (excess) Methyl-Iodide were added. The flask was allowed to stay at 4 °C for three days, but unfortunately the crystallization process

was very slow and the amount of precipitate was rather minimal. The solution was evaporated and redissolved in ethyl acetate. Petrol ether was used as precipitant, but no significant crystallization took place.

2.4 Microbiological methods

2.4.1 Gene sequences, expression vectors and auxotrophic bacterial strains

The expression experiments with model proteins were performed in suitable auxotrophic strains transformed with expression plasmids based on T5 promotor controlled system inducible with IPTG.

2.4.2 Annexin V

The expression experiments of Annexin V were performed with two different bacterial strains. For the incorporation of Leucine analogues, the host strain DSM 1563 (*leu*$^-$) was transformed with two plasmids: pQE60-PP4 harboring the Annexin V gene sequence (*NcoI-HindII* fragment) and pREP4 which confers kanamycin resistance. For the incorporation of tryptophan analogues the bacterial strain *ATCC*49980 (*trp*$^-$) was transformed with the plasmid pQE80 containing an N-terminal His tag. For the expression of wt Annexin V were used both systems described above.

2.4.3 EGFP

For the incorporation experiments of Leucine analogues in EGFP, the host strain DSM 1563 (*leu*$^-$) was transformed with the plasmids pQE60 carrying the gene encoding for the protein sequence, and pREP4 which confers kanamycin resistance. The same system was used also for the expression of the native protein.

2.4.4 Barstar

For the expression of wt Barstar, of its variants and mutants, the EcoRI-HindIII fragment from the pKK223-3 plasmid containing the ribosome-binding site and wt-, and mutant- barstar DNA sequences was inserted into the pQE-30 vector and transformed into the expression host *E. coli ATCC49980* together with pREP4.

2.4.5 Cell transformation via electroporation

The vectors were introduced into the hosts by electroporation. Before the electrotransformation of competent *E. coli* cells, plasmids were dialyzed against distilled and deionized water by placing 10 µl of circular DNA on a dialysis membrane. After 30 min the DNA solution was pipetted from the membrane and mixed with 50 µl cell culture in an electrotransformation chamber. The mixture was electroporated at 1650 mV with time constant between 3.5 and 5.0 ms. Then the transformed cells were washed with 1 ml LB and transferred into a tube to be incubated at 37°C for at least 1 h. After incubation 20 µl and 200 µl of the transformed cell culture were plated on LB-agar plates (with suitable antibiotics) and incubated at 37°C overnight.

2.4.6 Expression test

Before starting with fermentation experiments one has to test whether the transformed cells are capable to express the target protein. For this reason, it is recommended to perform an expression test in order to choose the best colonies for further protein expression.
Up to ten colonies were picked random from the agar plates and every single colony was transferred into 5 ml LB medium added with the proper antibiotics and cells were let grow at the suitable temperature (30°C or 37°C) until OD_{600}: 0.5-0.8. At this point 1 ml cells suspension (10^9 cells) was transferred into a 1.5 ml tube, induced with 1µl IPTG and allowed to shake for at least 45 min at the proper temperature. After that, the suspension of bacteria was centrifuged for 5 min and the supernatant was discarded. The cell pellet was resuspended in 50 µl loading buffer and 50 µl H_2O to be incubated at 95°C for 1 min. After a brief sonification in order to destroy DNA, the sample was centrifuged for 5 min and 12 µl of the sample was loaded on the gel to be run at approximately 130 V and 30 mA. Staining of the gel was performed by incubation in Coomassie-Brilliant-Blue for 30 min at room temperature. By the application of the distaining solution the protein bands were made visible and the most suitable cells were used for the fermentation experiments.

2.4.7 Preparation of bacterial stock cultures

Single colonies grown on plates were pipetted to 5 ml LB media to be incubated until OD_{600} = 0.5-1.0. For the establishment of cell cultures 200 µl of the aliquots were mixed with glycerin to a final concentration of 10 %. The mixture was frozen in nitrogen and stored at -80°C. Transformed bacteria were handled in the same way by adding corresponding antibiotics to the media.

2.5 Fermentation, expression and incorporation experiments

2.5.1 Wt- proteins

Beside the incorporation experiments performed in NMM, wt-Barstar, wt-EGFP, and wt-Annexin V have also been expressed in order to have a suitable comparison system. For the expression of native proteins, the cells were allowed to grow in Luria-Bertani Medium (LB) until they reached OD_{600} 0.7-1.0. After the induction with isopropyl-β-D-thiogalactopyranoside (IPTG; 1M 1μl/ml), the cells were kept shaking overnight. The optimal fermentation temperature for Barstar is 27 °C, in order to avoid the activation of an N-terminal protease. For Annexin V and EGFP the cells were kept shaking at 30 °C.

2.5.2 Annexin V

For the incorporation experiments of 5′,5′,5′,trifluoroleucine (TFL) into Annexin V, *DSM1563* cells were allowed to grow in NMM at 30 °C with a 0.3 mM limiting concentration of Leucine as natural substrate.

After the mid-logarithmic phase of the growth (OD_{600}: 0.6-0.9), corresponding to the depletion of Leu was reached, the cells were induced with isopropyl-β-D-thiogalactopyranoside (IPTG; 1M 1μl/ml) and simultaneously the non canonical analogue TFL was added. The cells were kept shaking overnight at room temperature.

For the incorporation attempts of Tryptophan analogues, ATCC49980 *E.coli* cells were allowed to grow in NMM at 37 °C with a 0.015 mM limiting concentration of Trp. After the consumption of the natural substrate, the cells were also induced with IPTG and "fed" with the Tryptophan analogues (5 ml/L, stock solution 50mM). The cells were allowed to express at 30 °C overnight.

2.5.3 EGFP

For incorporation experiments of TFL into EGFP was followed the same protocol as for incorporation attempts of TFL in Annexin V.

2.5.4 Barstar

For the incorporation attempts of Tryptophan analogues into Barstar (both wt and mutants), *ATCC49980 E. coli* cells were allowed to grow in NMM at 37 °C with a 0.015 mM limiting concentration of Trp. After reaching the mid-logarithmic phase of the growth, protein expression

was induced with 1M IPTG (1M, 1µl/ml), the Trp analogues were added to the medium (5 ml/L, stock solution 50mM) and the cells were kept shaking at 27 °C overnight.

2.6 Protein purification

2.6.1 Purification of Annexin V

For the purification of Ax V substituted with TFL, cells were centrifuged at 4 °C and 5000 rpm for 20 min.The pellet was then resuspended in Spheroplast buffer (100ml for 1l culture) with Lysozyme (20mg per 100 ml buffer) and put on ice for 30 min. After another centrifugation (5000 rpm for 20 min) the pellet was resuspended in 60 ml UZ-buffer and centrifuged at 32000 rpm between 4h and 10 h (overnight). Liposomes (1-5 ml) and $CaCl_2$ (10 mM final concentration) were added to the surnatant and the mixture was put on ice for at least 30 min and then centrifuged (30000 rpm for 30 min). The pellet was resuspended with wash-buffer and centrifuged for 30 min (30000 rpm). The pellet was resuspended with elution-buffer (40-45 ml) and finally centrifuged for 1 hour at 40000 rpm. The protein was put in dialyze membrane (12000-16000 molecular weight cut off) and dialyzed first against water for at least 5 hours and finally overnight against 10 mM Tris/HCl pH 8.0. The purified protein was concentrated with CentriPrep and Centricon 10 (molecular weight cut off 10KDa).
For the purification of AxV with fluorotryptophan and aminotryptophan were followed the initial protocol steps described above, but after the resuspension in UZ-buffer and the centrifugation, the surnatant was loaded on a Ni-sepharose column and eluted with 100mM imidazole buffer.

2.6.2 Purification of Barstar

All Barstar mutants and variants were purified using the same protocol, as follows.
After cells were centrifuged (4200 rpm for 40 min), pellet was resuspended in 50 mM Tris/Cl with the addition of PMSF (1 mM) and lysozime (final concentration 50µg/ml) and put on ice for about 30 min. After sonication (3x60s) and centrifugation (40 min at 15000 rpm), the supernatant was discarded and the inclusion bodies were dissolved in 25-30 ml urea-buffer (7.5 M urea in 50 mM Tris pH 8.0). After centrifuging (40 min at 9000 rpm), the surnatant was dialyzed overnight against buffer (50 mM Tris/Cl+0.1 M NaCl), loaded on a Q-Sepharose column and eluted with a NaCl gradient.

2.6.3 Purification of EGFP

The cell suspension was centrifuged for 20 min at 6000 rpm and pellet was resuspended in 100 mM Sodium sulphat buffer pH 8.0. After sonication (6 x30s) and centrifugation (9000 rpm for 30 min.), the supernatant containing the soluble protein was collected. To this solution NaCl was added to a final concentration 0.5 M. The protein was loaded on a Ni-Chelat column and washed with 800 ml washing buffer. After applying 100 ml of imidazol buffer (20 mM) protein was eluted with 100 mM of the same buffer and protein fractions were collected and pooled. To obtain the protein in a purer form the protein was further purified with a phenylsepharose column, according to the following protocol. To the previously collected protein was added Ammonium sulphate to 20 % (v/v). The column was washed with Ammonium sulphate buffer and the protein loaded. The washing step was performed with 20 % Ammonium sulphate / Tris/ EDTA (30 ml) and protein was eluted with a gradient 20 % - 0 % Ammonium sulphate (60 ml).

2.7 Analytical methods

2.7.1 Analytical HPLC

In this study, HPLC has been used to detect the course of chemical or enzymatic reactions performed in order to prepare Tryptophan analogues.

Conditions:

Eluent A: 5% Acetonitrile/95% H_3PO_4 (2%)

Eluent B: 90% Acetonitrile

Flow: 1.5 ml/min

Wavelength: 210 nm

Gradient A. 0% Eluent B → 100% Eluent B in 15 min

Gradient B. 0% Eluent B → 100% Eluent B in 13 min

2.7.2 Mass Spectrometry

The mass spectra were measured with an ESI PE SCIEX API 165 (Perkin Elmer) single quadrupole MS system. The spectra for detecting the products were collected with an autosampler Series 200 (Perkin Elmer) rate of 10µl/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 was 10 µl dissolved in Acetonitrile. The eluent was 100% Acetonitrile (0.05% TFA).

2.7.3 UV Spectroscopy and molar extinction coefficients of Trp analogues

The UV spectra were measured on a UV/VIS spectrometer Lambda 19 connected with the software UV WinLab Version 2.0. The spectra were detected between 240 and 350 nm, with slit of 1.00 nm, a data interval of 0.1 nm and a scan speed of 240 nm/min.

Protein samples were always measured so to give an absorption signal in the range of 0.2~0.9 at 280 nm (linear region in Lambert-Beer law).

Since it was difficult to purify enzimatically synthesized Trp analogues to high grade, commercially available related indole analogues were used to determine ε_M (Fig. 2.1 and 2.2 and Table 2.1).This was possible since there is no significant difference in absorbance between the two compounds. 1 mg indole analogue was weighted on a precision balance and then dissolved in a solution containing 500 µl water and 500 µl HCl 1N. After 15 min in an ultrasound bath at 70°C, the solution was ready for the measurements. The molar extinction coefficient (ε) can then be easily calculated from the Lambert-Beer law:

$$A = \varepsilon lc$$

A = absorbance l = path length (1 cm) c = molar concentration ε = molar extinction coefficient in M^{-1} cm^{-1}.

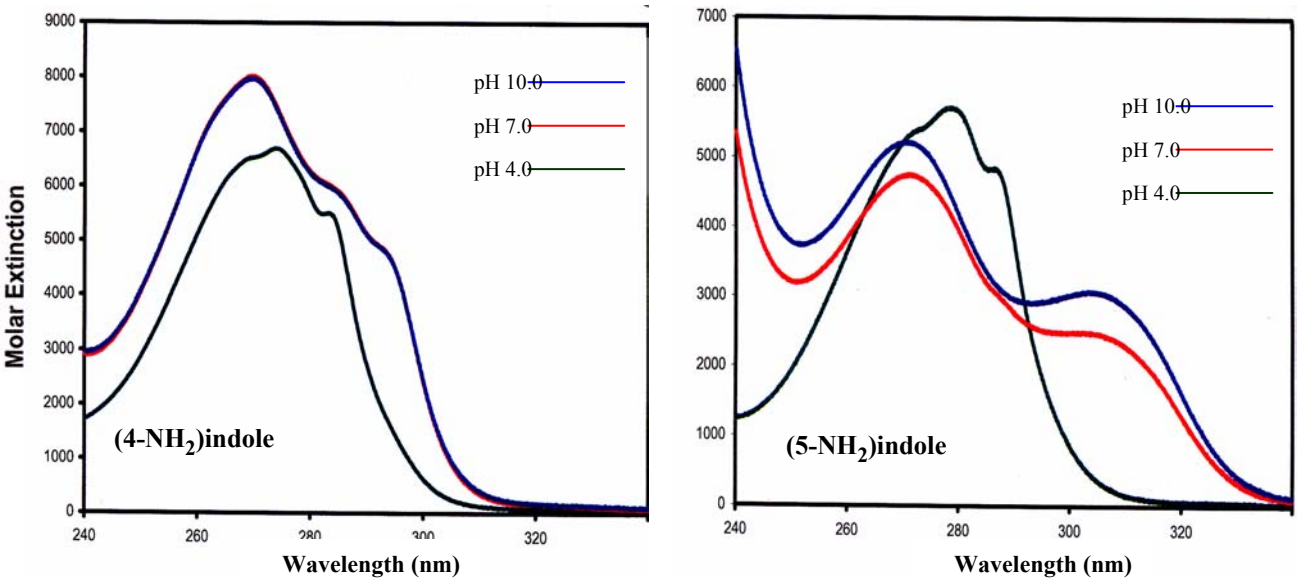


Figure 2.3. UV-absorbance spectra of 4- (left) and 5-aminoindole (right). The two indole analogues present strong pH sensitivity. As there is no significant difference in absorbance between aminoindole and aminotryptophan, the indole analogues were used as reference substance to determine ε_M (see Table 2.1).

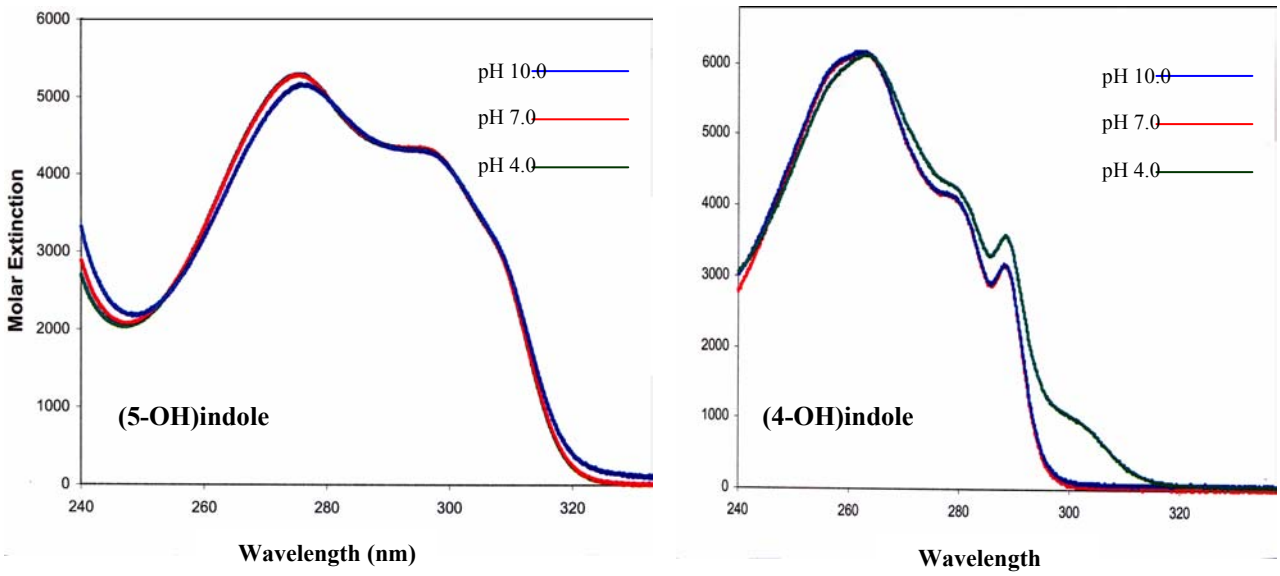


Figure 2.4. UV-absorbance spectra of 5- (left) and 4-hydroxyindole (right). Hydroxyindoles do not display pH sensitivity. As there is no significant difference in absorbance between hydroxyindoles and hydroxytryptophans, the indole analogues were used as reference substance to determine ε_M (see Table 2.1).

	(4-OH)indole		(5-OH)indole	
pH	A_{max}(nm)	ε_M(M^{-1} cm^{-1})	A_{max}(nm)	ε_M(M^{-1} cm^{-1})
4.0	262	6130	275	5300
7.0	262	6130	275	5300
10.0	262	6130	275	5150
	(4-NH_2)indole		(5- NH_2)indole	
pH	A_{max} (nm)	ε_M(M^{-1} cm^{-1})	A_{max}(nm)	ε_M(M^{-1} cm^{-1})
4.0	274	6700	278	5700
7.0	270	7900	271	4750
10.0	270	7900	270	5200

Table 2.1. The ε_M values calculated for (4-NH_2)indole, (5-NH_2)indole, (4-OH)indole, and (5-OH) indole at different pH values (see Fig.2.1 and 2,2)

2.7.4 Fluorescent Spectroscopy

The fluorescent emission spectra were recorded on Perkin Elmer Lambda 19 fluorescence spectrometer at protein concentration of 0.5-1.0 μM in the aqueous buffered solutions. The samples were excited in a range between 280-320 nm and the emission was detected in the range between 300-450 nm. All spectroscopic investigations were performed with filtered solutions (pore size 0.22 μm) at 20 °C.

2.7.5 Secundary structure determined by Circular Dichroism

The circular dichroism spectra were recorded on a Jasco J-715 spectropolarimeter with temperature controlled PFD-350S connected to the software program J-700 for Windows. The spectra were taken at 20 °C in quartz glass cuvettes with a path length of 0.1 cm. The instrument was calibrated with epiandrosteron in dioxane in the range of 360-240 nm and the linearity of the instrument was tested with camphorsulfonic acid in the range of 350-180 nm. The spectra were collected between 185 and 250 nm, with a scanning speed of 50nm/min, the response of 1 s, the

band width of 1.0 nm and the average of 10 scans were reported. The spectra were all normalised and the ellipticity was expressed as mean residue molar ellipticity $(\Theta)_R$ (deg cm^2 $dmol^{-1}$) :

$$(\Theta)_R = \Theta/lcn_R$$

where Θ is the observed ellipticity, l is the path length in cm, c is the molar concentration of the protein in solution and n_R is the number of optical active chromophores, that for a protein coincides with the number of amino acid residues.

The melting curves were measured by following the molar ellipticity at 222 nm versus the temperature, with the temperature slope of 30 °C/h, the response of 16 s and the band width of 1 nm. The wavelength at the zero point of the second derivative of the relative melting curve was taken as melting point, T_m. In the course of this experiment a CD spectrum was measured every 10 degrees in order to monitor the conformational changes.

2.7.6 Analyses of denaturation process

The conformational stability of a protein is defined as the difference in free energy (ΔG):

$$\Delta G = -RTlnK,$$

K = equilibrium constant

R = gas constant (8.31 J/mol deg)

T = temperature (K)

Melting curves were fitted using the program KaleidaGraph (Version 3.0 Sinergy Software; PCS *inc*.) on the basis of the Gibbs-Helmholtz equation for not linear regression. The thermal capacity ΔC_p (2500 cal/mol) has been empirically calculated from the amino acid sequence (Makadzde and Privalov, 1990). The van´t Hoff Plots can be used for the determination of ΔH_m values as they are linear in the melting point region (Santoro *et al*., 1988).

Entropy, ΔS_m, is defined as

$$\Delta S_m = \Delta H_m/T_m$$

as ΔG = 0 at the melting point.

When T_m, ΔH_m, and ΔC_p are known, ΔG can be easily calculated for any temperature value:

$$\Delta G = \Delta H_m (1-T/T_m) - \Delta C_p [Tm – T \ln(T/T_m)]$$

2.7.7 Nuclear Magnetic resonance Spectroscopy

The identity of the Trp analogues, obtained in solid form via cold precipitation was proofed by NMR analyses. All compounds were dissolved in D_2O:MeOD (1:1). The NMR spectra were recorded in a range between 1-10 ppm.

2.8 Electrophoretic methods

2.8.1 SDS-polyacrylamide gel electrophoresis

For a qualitative confirmation of protein expression and protein purification SDS-PAGE method of Laemmli (1970) was used. The gels of dimensions of 10.4 cm x 10 cm x 0.7 mm were of the following composition (for 8 gel plates):

Separating gel (5%): 7.5 ml AABis

5.65 ml 1 M Tris/HCl pH 6.8

31.2 ml distilled H_2O (dH_2O)

0.45 ml 10 % SDS

0.21 ml 10 % ammoniumperoxisulphate (APS)

0.08 ml N, N, N´, N´ - Tetramethylethyldiamine (TEMED)

Running gel (12%): 36 ml AABis (60ml for 20% gel)

17 ml 2 M Tris/HCl pH 8.8

36 ml dH_2O (12 ml for 20% gel)

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 % bromophenolblue

20 % glycerine

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 R250

90 ml methanol/H2O (1:1)

10 ml concentrated acetic acid

Destaining solution (5 l): 1.25 l absolute ethanol

0.4 l concentrated acetic acid

3.35 l dH_2O

2.9 X-ray Crystallography

2.9.1 Crystallization of Barstar

For the crystallization screening was applied the sparse matrix sampling method and the best results were obtained under the following conditions. Crystallization of (4-NH_2)Trp-Barstar was achieved in sitting drops at +4°C by vapor diffusion. The protein was concentrated to3.5 mg/ml in 5mM sodium phosphate buffer and 2µl of the protein solution were mixed with 1.2 µl of crystallization buffer (2.4 M Ammonium sulphate pH 9.0) and equilibrated against 200 µl reservoir buffer. The protein crystals appeared from the precipitated protein solution in 5-7 days. Crystals of (5-NH_2)Trp-Barstar were obtained in sitting drops at +4°C by vapor diffusion. The protein was concentrated to3.2 mg/ml in 5mM Sodium phosphate buffer and 2µl of the protein solution were mixed with 1.2 µl of crystallization buffer (2.5 M Ammonium sulphate pH 9.0) and equilibrated against 200 µl reservoir buffer. The protein crystals appeared from the precipitated protein solution in 7-10 days. Unfortunately, these crystals could not be used for X-ray data collection as no diffraction pattern could be obtained.

2.9.2 X-ray data collection and Structure Refinement of (4-NH_2)Trp-Barstar

Diffraction data for (4-NH_2)Trp-Barstar up to 2.0 Å resolution were collected from a single crystal on a X-ray image plate system at DESY, Hamburg. To prevent damages from X-ray exposure, a solution of 1.5 M Li_2SO_4 mixed with 21% glycerol was used as cryoprotectant and the crystal was shock-frozen in a nitrogen stream of 100K.
Reflections were integrated with the program MOSFLM (Leslie *et al*., 1998), scaled and reduced using SCALA of the CCP4 package (Collaborative computational project, 1994). For the model building was used the program MAIN (Turk, 1996) and refinement has been done with CNS (Brünger *et al*., 1998) after initial rigid body minimization. The atom replacements were determined by difference Fourier methods and for the Patterson search was used the program AmoRe (Navaza *et al*., 1994), setting the native wt-Barstar as model. The incorporation of (4-NH_2)Trp was confirmed by the appearing of additional electron density in the Fo-Fc Fourier map at the position where the non canonical amino acid has been incorporated.

2.9.3 Molecular modeling for (5-NH_2) Trp-Barstar

Since it could not be possible to obtain a clear diffraction pattern from (5-NH_2)Trp-Barstar crystals, a replacement with 5-amino-Trp has been simulated using the crystal structure coordinates of (4-NH_2)Trp-Barstar. A short run of energy minimization was performed using the SYBYL package (Tripos, Inc., St. Louis, Missouri). (5-NH_2)Trp was then inserted in the structure in place of (4-NH_2)Trp and a new refinement has been performed in order to check the possible interactions of the amino group at the new position.

2.10 Delivery of proteins containing non canonical amino acids into cells

MCF10 cells (breast epithelial human cells) were incubated in well plates at 37 °C until they reached the concentration of about 250000 units per ml. The excess of growing medium was eliminated and a mixture of lipofectamine (Invitrogen) and protein (previously incubated at room temperature for about 30 m.) was applied on the cover slips. At this point some fresh medium was added and cells were allowed to stay at 37 °C for a time between 6 and 48 hours. After this time, the medium was taken away and the cover slips were washed with PBS before the fixation solution (PBS + 2% Formaldehyde + 100mM sucrose) was applied for 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).

The fixed cells were then observed under a microscope (Axiorvert-Zeiss) with two different wavelengths. At 495 nm GFP and at 385 nm DAPI were detected.

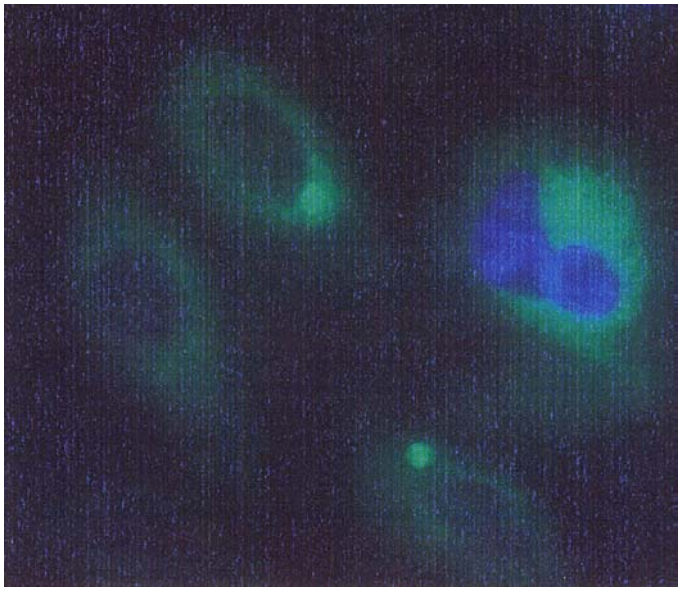


Figure 2.5. Cultured breast epithelial human cells (MCF10) after incubation with *m*-F-Tyr-EGFP (24 h). Note that EGFP is uniformly distributed in cell cytoplasm. An intact nucleus indicates that cells still did not undergo apoptosis. The protein is delivered by using lipofectamine.

2.11 Media and Buffers

2.11.1 Nutrition media

***Luria Bertani Medium (LB)*:** 1 % trypton
1 % NaCl
0.5 % yeast extract
pH 7.5
agar plates: 1.5 % bacto-agar

***New Minimal Medium (NMM)*:** 50 mM $K_2HPO_4.3H_2O$
22 mM KH_2PO_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
100 µg/ml corresponding antibiotics
50 mg/l all amino acids

Vogel-Bonner Medium (VB): 0.8% trypton
1.6% yeast extract
0.4% glycerin
0.2% NaCl

Full media were sterilized in autoclave at 120°C for 35 m.

Addition of antibiotics: 100 mg/ml ampicillin
70 mg/ml kanamycin.

Growth Medium for MCF10 cells

The medium consists in one part of Dulbecco medium and one part of F12 medium. To this basic solution were added:

Horse serum (final concentration 5%)
Epidermal Growth Factor
Cholera toxin
Insulin
Dexometazone

2.11.2 Buffers

Purification of Annexin V (without His-tag)

Spheroplast-buffer:
200mM Tris/HCl pH 8.0
0.5 mM EDTA pH8.0
750 mM Saccharose

UZ-buffer:
2 mM Tris/HCl pH 8.0
0.2 mM EDTA pH8.0
0.5 mM $MgCl_2$
10 mM NaCl

To 100 ml buffer were also added:
0.1 mg/ml RNAse
0.1 mg/ml DNAseI
2.0 mM PMSF
0.1% Triton x100
2.0 mM Benzamidiniumchloryde

Lipo-buffer:
2 mM Tris/HCl pH 8.0
0.3 mM $MgCl_2$
10 mM NaCl

Washing-buffer: Lipo-buffer + 5 mM $CaCl_2$

Elution-buffer: Lipobuffer + 20 mM EDTA

Lipos: 1g. Extract were dissolved in 4ml Chloroform in a 2 liter flask, dried under N_2 atmosphere (15 min) and resuspended in 10 ml Lipo-buffer in sonication bath

Buffers for purification of EGFP

Washing buffer:
100 mM NaP buffer pH 8.0
500 mM NaCl

Elution buffers:
20 mM Imidazole
100 mM NaP buffer pH 8.0
500 mM NaCl

100 mM Imidazole
100 mM NaP buffer pH 8.0
500 mM NaCl

Washing buffer for Phenylsepharose column:
20 % Ammonium sulfate
20 mM Tris pH 8.0
1 mM EDTA

3 RESULTS

3.1 Expression and analyses of wt proteins

The present study is mainly based on incorporation studies of non natural amino acids into model proteins, EGFP, human recombinant Annexin V and Barstar. Resulting variants of these proteins have been subsequently characterized by using following methods: circular dichroism, thermal denaturation, fluorescence emission, pH titration and X-ray crystallography. Wild-type forms of the above mentioned proteins served as control system to check the expression capacity and as a reference in investigating novel properties of the related variants.

3.1.1 Expression and analytical characterization of wt-EGFP

The expression level of EGFP protein in LB medium was always high (about 30 mg/l). Such high levels of expression were possible only with freshly transformed cells, as shown in Fig. 3.1. By using older competent cells (2-3 weeks) the expression level dropped down rapidly (about 3-4 mg/l).

Wt-EGFP protein was successfully purified and isolated as unique specie, as shown by the results of the mass analyses. The found mass value (27743.0 Da) corresponds to the theoretical mass of the wt-protein (27740.0 Da).

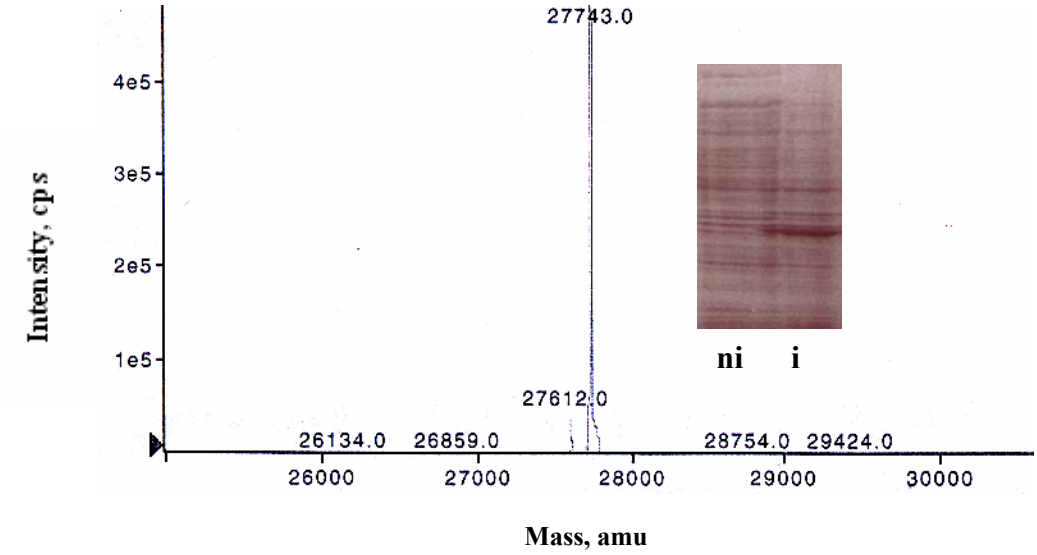


Figure 3.1. Deconvoluted ESI-MS profile of wt-EGFP. Insert: expression profiles in cell lysates of E. *coli DSM* 1563 (**i** = induced cells; **ni** = non induced cells).

3.1.2 Spectral properties of EGFP

EGFP (Ser65Thr, Phe58Leu) belongs to the class II of GFP mutants (where Ser 65 is replaced with Thr, Ala or Gly) and displays in the photophysics of its chromophore the properties of phenolate ion.

In the UV spectrum of EGFP it can be identified the contribution of two different chromophores. The absorption band at 277 nm can be attributed to Trp and Tyr, while the absorption band at λ_{max} = 488 nm is due to the chromophore.

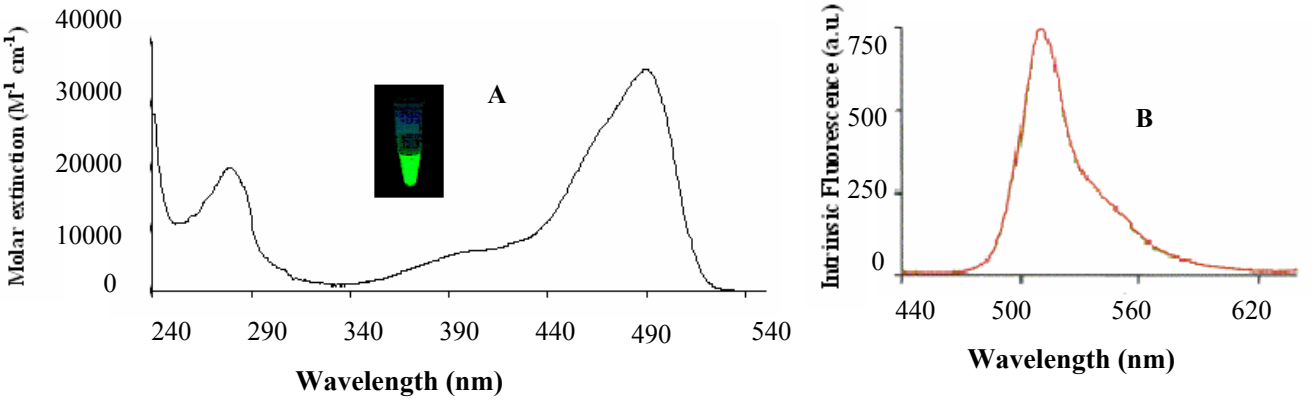


Figure 3.2. Spectroscopic features of EGFP. **A**) UV absorbance spectrum. When compared to wt-GFP from *Aequorea Victoria* (λ_{max} = 488 nm), the absorption peak is red-shifted of about 13 nm with enhanced amplitude due to the anionic form of GFP chromophore. **B**) Fluorescence emission spectrum upon excitation at λ = 488 nm.

The fluorescence emission spectrum mirrors the UV absorbance spectrum and is characterized by an emission maximum at λ = 509 nm with a Stokes shift of 21 nm.

3.1.3 Expression and mass analyses of wt-Annexin V

The expression of wt-Annexin V was performed in LB medium with a protein yield of about 100 mg/l, by using both vectors, pQE60-PP4 as well as pQE80 containing an N-terminal His-tag (NtHT).

The results of the mass analyses showed one single peak corresponding to Annexin V (calculated molecular weight: 35805 Da; found 35806 Da) For NtHT-Ax V the found molecular weight is 37680.0 Da, a value that is consistent with the calculated one (37676.0 Da).

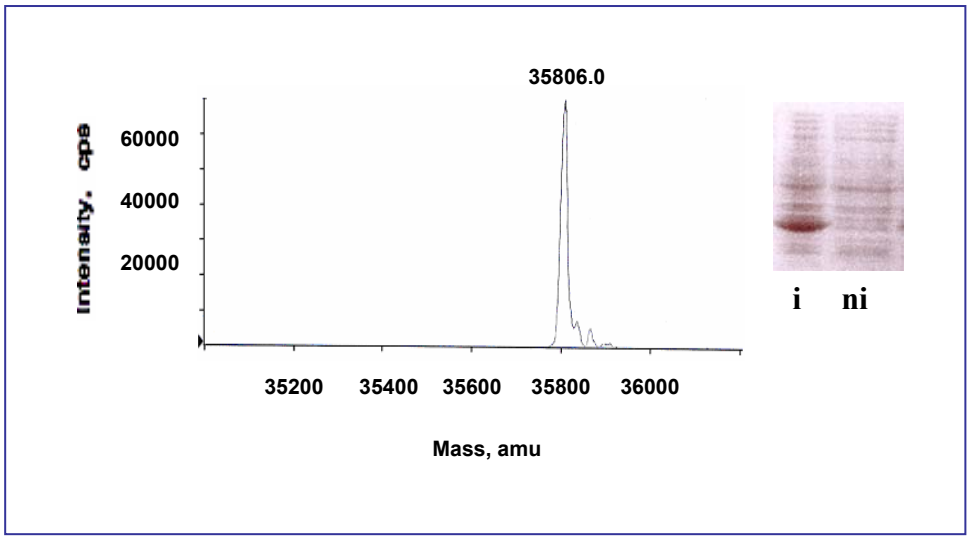


Figure 3.3. Deconvoluted mass profile of human recombinant Annexin V. Note that single peak indicates a highly homogeneous sample. Insert: expression profile of Annexin V in *E.coli ATCC 49980*. (**i** = induced cells; **ni** = non induced cells)

3.1.4 UV absorbance and Fluorescence emission spectra

Human recombinant Annexin V contains in its sequence one single Trp and 12 Tyr residues that contribute together to the spectroscopic properties of this protein. The typical UV absorbance spectrum of human Annexin V displays a maximum centred at λ = 277 nm.

The fluorescence emission spectrum presents two different maxima, depending on the excitation wavelength. When the protein is excited at λ = 280 nm the spectrum presents a broad peak in the range between 318-325 nm with a maximum positioned at 320 nm, while excitation at λ = 295 nm the position of the maximum is at 333 nm

An interesting property regarding Annexin V fluorescence is its pH dependence. At low pH (below 3.5) its fluorescence intensity is considerably enhanced (~40%) and red-shifted (~12 nm) when excited at 280 nm. This is most probably due to the molten globule like state of Annexin V at low pH values. It is well-known that in molten globule state certain residues such as Tyr and Trp become solvent exposed.

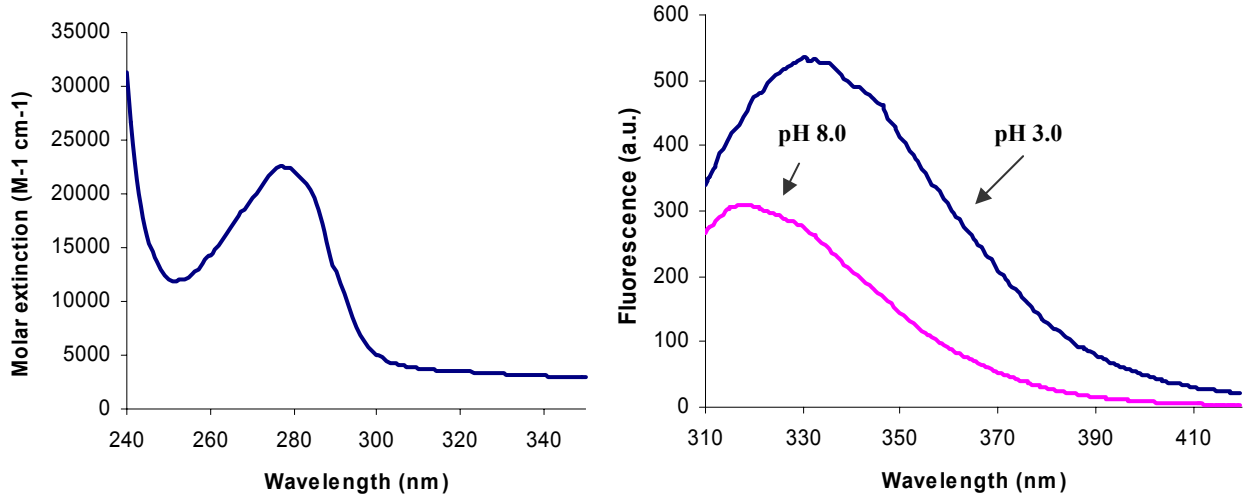


Figure 3.4. Absorption profile (left) of human Annexin V. Note that scattering between 310-350 nm is due to the high concentration of the sample. Fluorescence emission spectra (right) of human Annexin V excited at 280 nm. The emission maximum at pH 3.0 is centred at 333 nm, while at pH 8.0 is centred at 318 nm. Annexin shows the maximum in intensity at acidic pH values.

3.1.5 Thermal denaturation and CD spectra of native Annexin V

The melting curve of human Annexin V exhibits a typical cooperative sigmoidal shape with a midpoint transition at about 56 °C (324 ± 2.5K).

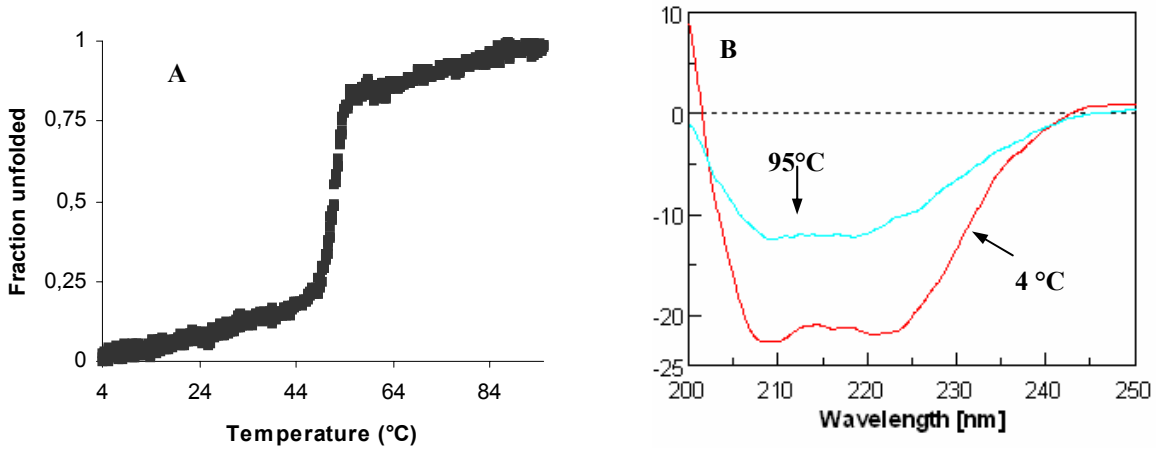


Figure 3.5. The melting curve of Annexin V (**A**) was obtained by following the dichroic intensity at 222 nm (helical minimum), upon increasing temperature from 4 °C to 95 °C (helix-to-coil transition). (**B**) Far-UV CD spectra at 4 °C (native state) and 95 °C (denatured) state. Note that denatured state is characterized by residual secondary structure (compact denatured state)

3.1.6 Expression and mass analyses of native Barstar and its mutants

The expression of Barstar, both its native and mutant forms, was performed in LB medium with bacterial cells *E. coli ATCC 49980* showing an exponential growth. Usually, after five hours the cells were ready for the induction (OD_{600} 0.6-0.9) and allowed to shake overnight. After fermentation procedure was finished, cell number has usually duplicated.

The mutants W38F and W44F showed a satisfactory level of protein expression, fully comparable to the parent protein. The double mutant W3844F, with only a single Tryptophan buried in the protein interior showed always about one third of expression level when compared to the native protein.

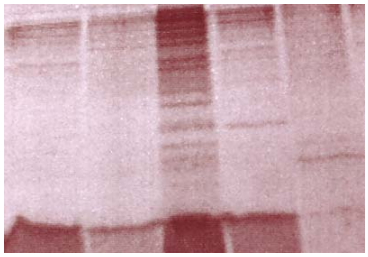


Figure 3.6. Expression of wt-Barstar and its mutants. The overexpressed native protein (P27A) and its mutants (about 10 KDa) migrate in the lower part of the gel (SDS-PAGE 20%) (**ni** = non induced cells).

As shown in Fig. 3.7, mass profiles of these proteins revealed homogeneous samples with single mass peaks.

Protein	M.W. (calculated, Da)	M.W. (found, Da)
P27A	**10253**	**10254**
W38F	**10213**	**10215**
W44F	**10213**	**10215**
W3844F	**10170**	**10173**

Table 3.1. Theoretical and found molecular weights for wt-Barstar and its mutants. Mass profiles are provided in Figure 3.7

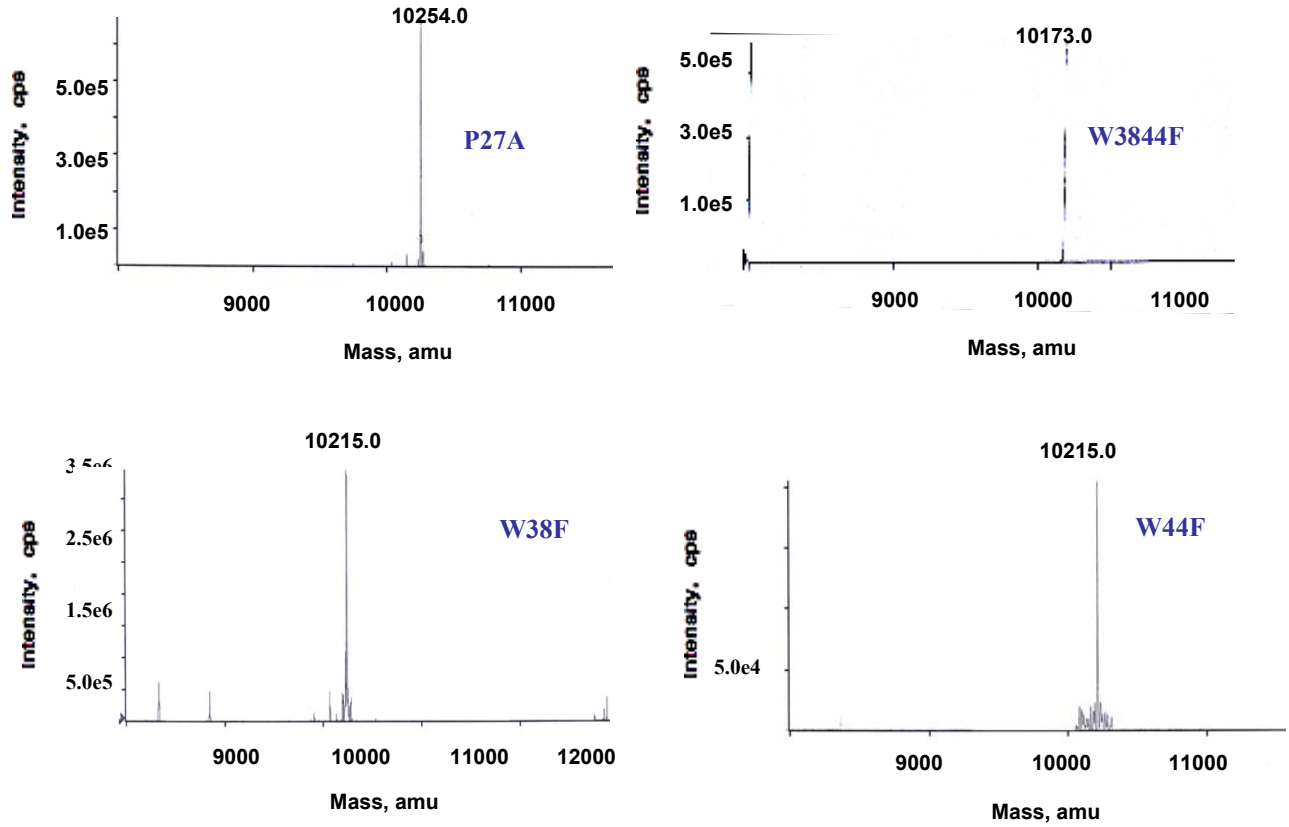


Figure 3.7. Mass spectrometric (ESI-MS) profiles of wt-Barstar (P27A) and its mutants. Found masses are in excellent agreement with the calculated mass values (see Table 3.1)

3.1.7 Fluorescence and absorbance properties of native Barstar and its mutants

Native and mutants Barstar have been spectroscopically investigated. While the absorption maximum in native and mutant proteins is identical, their intensities vary due to the different number of Trp residues. Native Barstar with three Trp residues shows the highest molar extinction (Fig. 3.8), while W3844F with only one single Trp residue has a considerable decreased intensity and maximum shifted to 277 nm, due to the increased contribution of Tyr residues to the absorption spectrum.

As already known from the literature (Nath et al., 1997), the major contribution to fluorescence emission of Barstar, comes from Trp 53, while the fluorescence of Trp 38 is completely quenched in the native protein. Therefore, as it can be seen in Fig.3.9, the maximum fluorescence intensity is displayed by native Barstar and W38F shows only a minimal decrease when compared to the native protein. As expected, the weakest intensity is showed by the double mutant W3844F. It is noteworthy that the wavelength of maximum emission is the same for wt, W44F and W3844F (λ_{max} = 330 nm), while W38F displays a 5 nm blue-shift (λ_{max} = 325 nm).

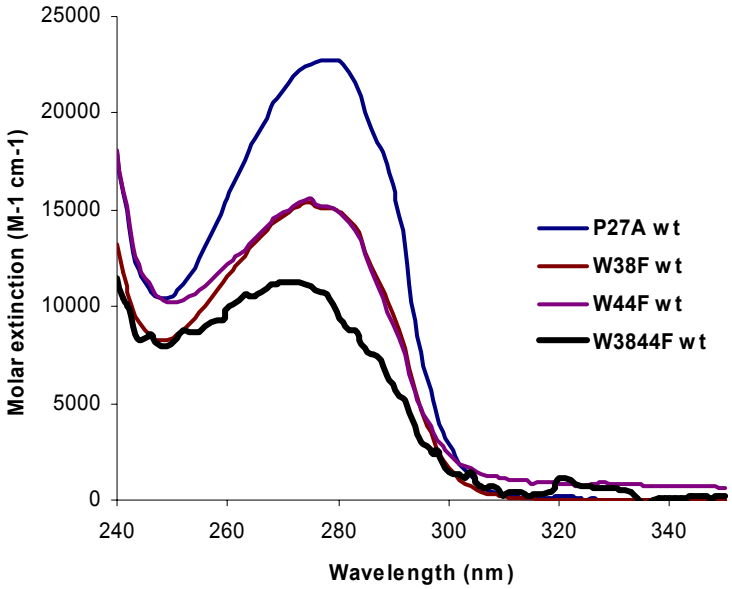


Figure 3.8. UV-spectra of native Barstar and its three mutants. There is practically no difference in the curves shape and in the absorption maximum (wavelength). Only W3844F exhibits a weak blue-shift (3 nm) and a considerable decreased intensity due to the single Trp residue that is present in the protein sequence.

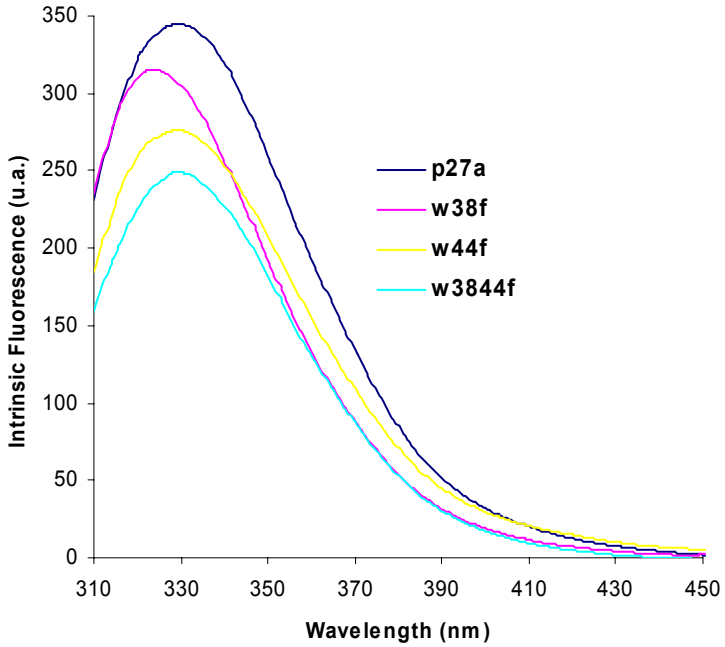


Figure 3.9. Fluorescence emission profiles of native Barstar and its mutants. All spectra were recorded at pH 7.0. The samples were excited at $\lambda = 280$ nm with slit width 5.0.

The pH sensitivity in Barstar is achieved by global substitution of all three Trp residues by aminotryptophan analogues. However, the arising question is to probe to which extent each of these three residues contribute to the pH sensitivity upon aminotryptophan incorporation. Therefore, site–directed mutagenesis experiments were performed, in order to create mutants with only two Trp residues (Barstar W38F and Barstar W44F), as well as a Barstar mutant with one single Trp residue (W3844F). These mutants also served as model proteins to study the role and influence of the different Trp residues in protein stability.

As expected, wt-Barstar and its mutants that contain not modified Trp side chains are pH insensitive, as determined by fluorescence measurements. That means that upon pH variations there are no shifts in the position of the protein emission maxima.

3.1.8 Secondary structure analyses and unfolding profiles of wt-Barstar and its mutants

As shown in Figure 3.10, mutagenesis of Trp to Phe does not influence the secondary structure of mutated proteins when compared to wt form. It should be also noted that there are no differences in the main residue ellipticities at 222 nm and this clearly indicates that the helical content does not suffer from such Trp→ Phe substitutions.

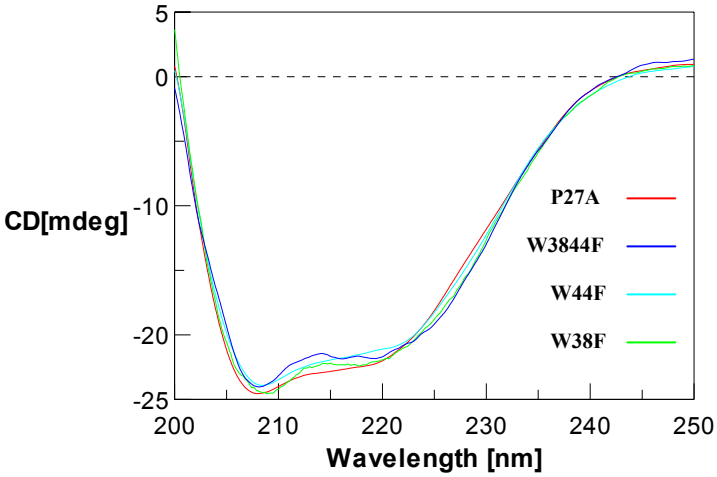


Figure 3.10. Far-UV CD spectra for pseudo-wt-Barstar and its mutants. It is evident that the replacement of Trp with Phe does not affect the helical content of the proteins. All spectra were recorded at 20 °C in 50 mM sodium phosphate buffer (pH 8.0).

In order to compare possible differences in stability among the four Barstar species, melting curves were measured from 20 °C to 95 °C at 222 nm. In addition, the CD spectra were measured every 10 degrees increase during the denaturation process, so that it could be followed the loss in secondary structure versus the increasing temperature. Wt-Barstar was expected to be the most stable form, but surprisingly it has turned out that W38F has the highest T_m value (see Fig. 3.11 and Table 3.2). However, by comparing the van´t Hoff enthalpy values for the two proteins it is clear that wt displays the highest cooperativity. The double mutant W3844F (Fig.3.12) is particularly unstable because of the lack of Trp 38 and Trp 44, as Phe residues are too small to maintain tight contacts with hydrophobic residues such as Thr 42, Val 73, Ala 77, and Leu 27.

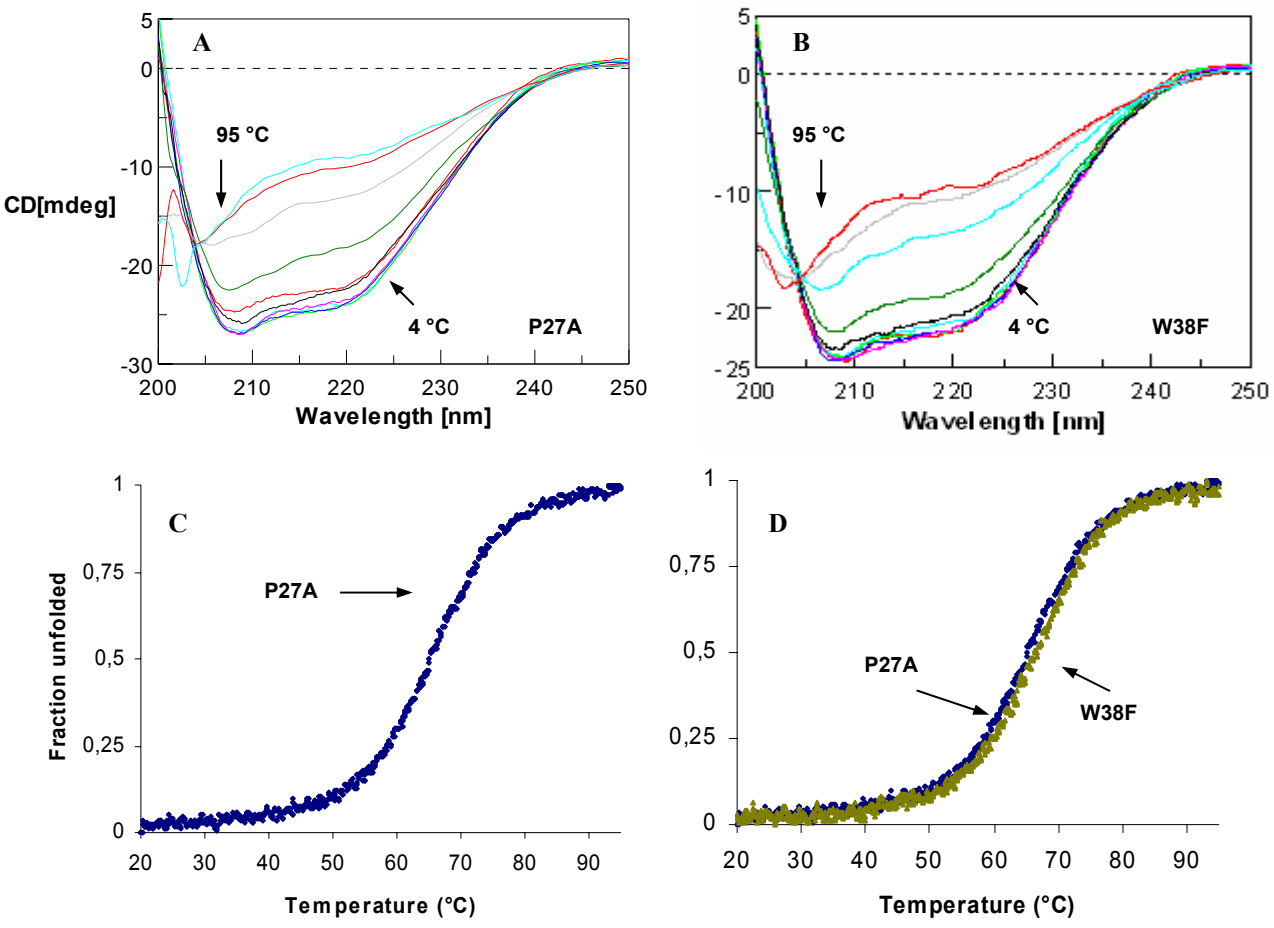


Figure 3.11. Changes in secondary structure upon heating from 20 to 95 °C (**A, B**) Secondary structure profiles of wt-Barstar and of W38F. Both Barstar species exhibit rather sharp native-to-unfolded-state transitions. (**C, D**) Melting curves of P27A (Tm = 65.62 °C) and of W38F (Tm = 67.33 °C) compared to the wt-protein.

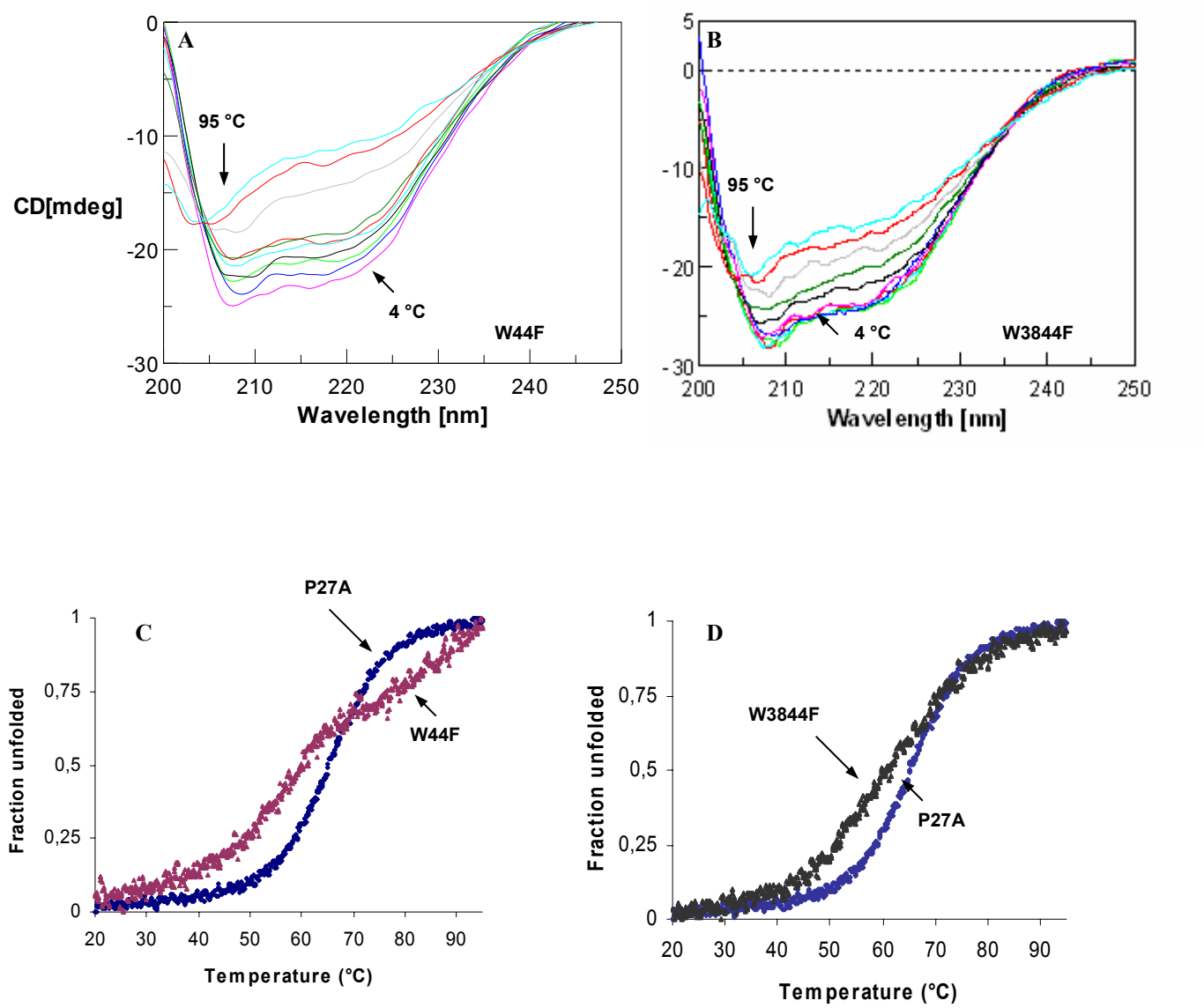


Figure 3.12. Secondary structure profiles of W44F **(A)** and W3844F (**B**) during thermal denaturation process. Comparison between W44F (Tm = 61.48 °C) and P27A melting curves (**C**) and comparison between W3844F (Tm = 60.28 °C) and P27A melting curves (**D**).

While observing the secondary structure profiles of all proteins, it will be noticed that both wt Barstar and W38F display the same sharp native-to-unfolded-state transition pattern (Fig. 3.11), while the double mutant displays a continuous transition from native to denatured state, thus revealing its less stability and cooperativity.

Protein	Tm (K)	ΔT_m(K)	ΔH_m (kJ/mol)	$\Delta\Delta H_m$(kJ/mol)
P27A	338.77 ± 0.03	-	-183.1 ± 1.9	-
W38F	340.48 ± 0.04	1.71	-176.2 ± 2.0	-6.9
W44F	334.63 ± 0.13	-4.14	-103.9 ± 2.1	-79.2
W3844F	333.43 ± 0.25	-5.34	-93.1 ± 6.3	-90.0

Table 3.2 Thermodynamic parameters derived by van´t Hoff analyses for unfolded wt-Barstar and its mutants. The ΔTm values represent the difference between native Barstar (P27A) and the related mutant melting Temperatures.

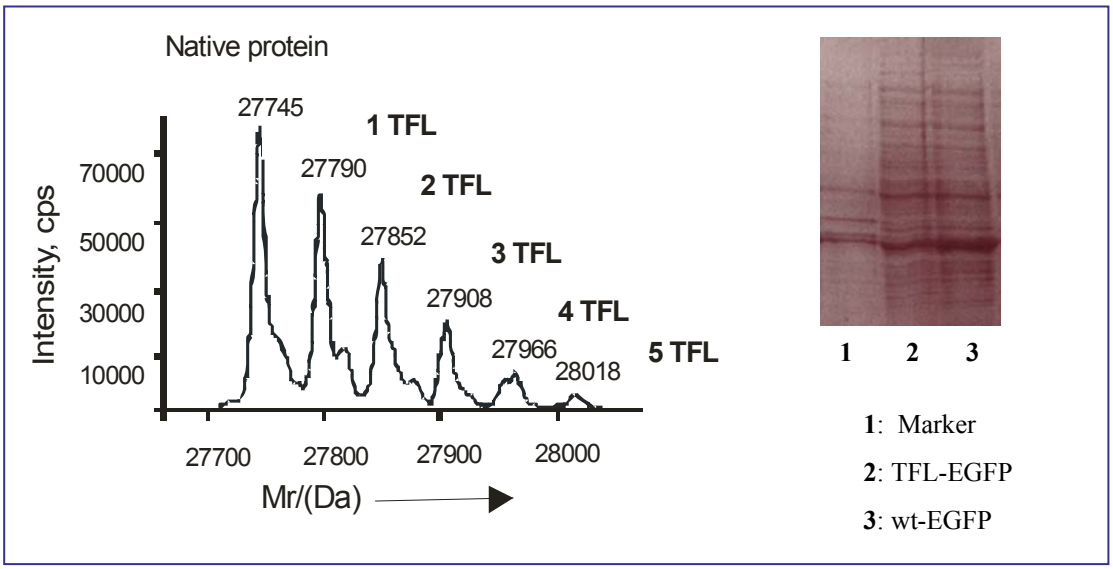


Figure 3.13. The level of substitution of Leu by TFL in EGFP checked by electrospray mass spectrometric analyses. The main peak corresponds to wt protein, while the number of substituted Leucine never exceeded five residues. Insert: Expression profiles in cell lysates of *E. coli DSM 1563*

3.2 Incorporation of non canonical amino acids into model proteins

In this study several non canonical amino acids were incorporated into EGFP, Annexin V, and Barstar in order to probe:

- their suitability as substrate for the protein synthesis
- their role in the protein folding and stability
- delivery of new spectroscopic characteristics to the entire protein
- the extent to which they can be incorporated into protein structures

3.2.1 Incorporation of 5´,5´,5´-trifluoroleucine (TFL) in EGFP

EGFP (Phe64Leu/Ser65Thr, 15 Leu, 27740 Da), was used for incorporation studies of the non canonical amino acid 5´,5´,5´-trifluoroleucine (TFL). The attempt to incorporate TFL into EGFP, using the SPI methodology was not satisfactory, as the substitution level according to the mass analyses can be regarded as marginal, even if the yield of protein expression is comparable to wild type (50 mg /l). The results of the mass analyses showed that only up to five TFL residues could be incorporated into EGFP (see Fig. 3.13).

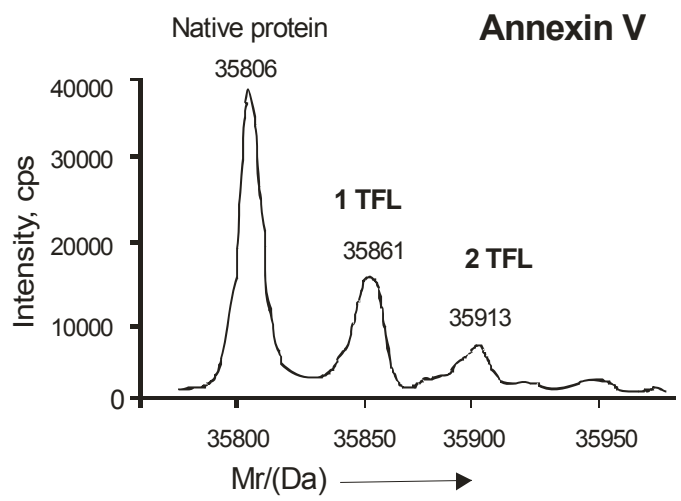
3.2.2 Incorporation of TFL into Annexin V

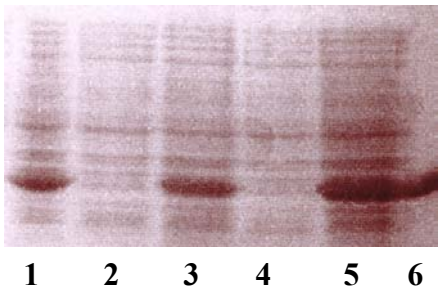
Human recombinant Annexin V contains in its sequence a single Trp residue at position 187, which plays a crucial role for the Ca^{2+}-dependent membrane binding and related channel activity (Huber et al., 1990). Therefore, Trp represents an attractive target for substitution studies with non canonical amino acids.

Human recombinant Annexin V contains 35 Leu residues in its sequence and it was chosen as target protein for the incorporation of TFL. It is already known from pioneering experiments done by Anker (Fenster and Anker, 1969) and recently by Tirrell and co-workers (Tang and Tirrell, 2001) that TFL can be used as a suitable substrate for Leu replacement experiments. Indeed, TFL is recognized and activated by Leucyl-tRNA-synthetase, even if not with the same efficiency if compared to the wild type.

This attempt was performed following our routine incorporation protocols, by allowing the bacterial cell growth in minimal medium with a limiting concentration of natural substrate, which corresponds to 0.3 mM for *DSM 1563* host strain. When the mid-logarithmic phase of the growth was reached, the cells were induced with IPTG and TFL was simultaneously added as non canonical substrate.

Even if the electrophoretic profiles of cells showed a satisfactory expression protein level, the mass spectrometry analyses clearly demonstrated that the incorporation of TFL into the target protein is only qualitative and not quantitative as expected. Indeed, the substitution level was of two out of 35 Leu residues, as showed in Figure 3.14.





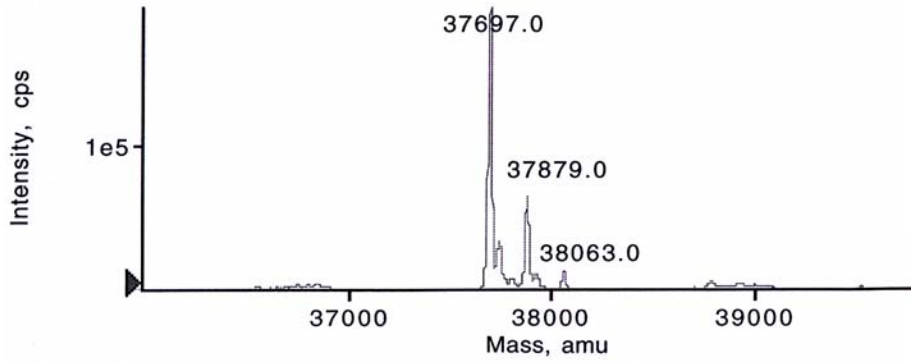
1, 3: cells after expression of TFL- AxV

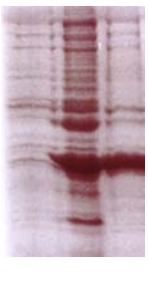
2, **4**: non induced cells

5: cells after expression of wt-AxV

6: Annexin V as marker

Figure 3.14. The level of substitution of Leu by TFL in Annexin V checked by electrospray mass spectrometric analyses. Insert: Expression profiles in cell lysates of *E. coli DSM 1563.*





1: non induced cells

2: cells after expression of NtHT-AxV with (4-F)Trp

3: cells after expression of NtHT AxV with (7-F)Trp

Figure 3.15. Deconvoluted ESI-MS profile of (4-F)Trp containing Nt-His-tag-Annexin V. Insert: Expression profiles of wt-, and variants of Annexin V in cell lysates of *E .coli ATCC49980.*

3.2.3 Incorporation of fluorotryptophans in Annexin V

The incorporation of (4-F)Trp and (7-F)Trp in Annexin V has been successfully performed with the SPI methodology. The yield of both substituted proteins is fully comparable to the wild-type and the results of mass analyses confirm that the incorporation has taken place (calculated mass: 37694.0 Da; found 37697.0 Da) (see Fig. 3.15). The presence of other peaks with a larger molecular weight is not easy to explain; most probably they are due to the formation of different non specific educts.

3.2.4 Spectral properties of Annexin V containing (4-F)Trp and (7-F)Trp

The UV-spectra of both wt-Annexin V and variants show the same emission maxima centred at 277 nm and identical intensity. Therefore, it can be concluded that the substitution of Tryptophan with (4-F)Trp and (7-F)Trp has no effects on the absorbance of Annexin V.

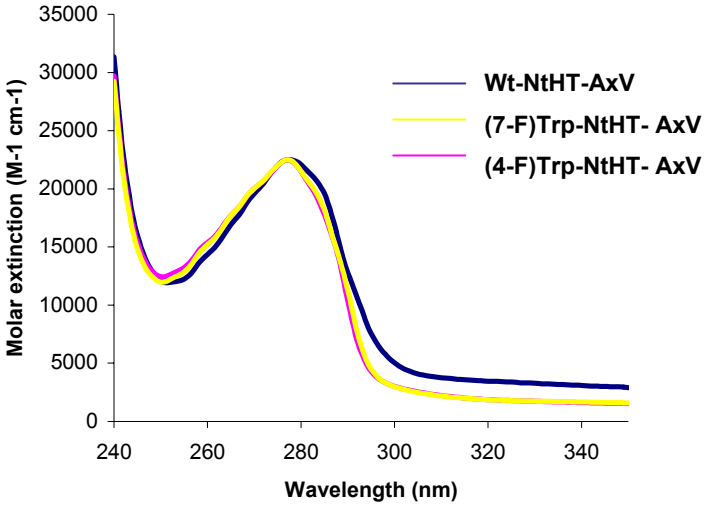


Figure 3.16. Normalized UV spectra of wt-Annexin and its fluorinated variants.

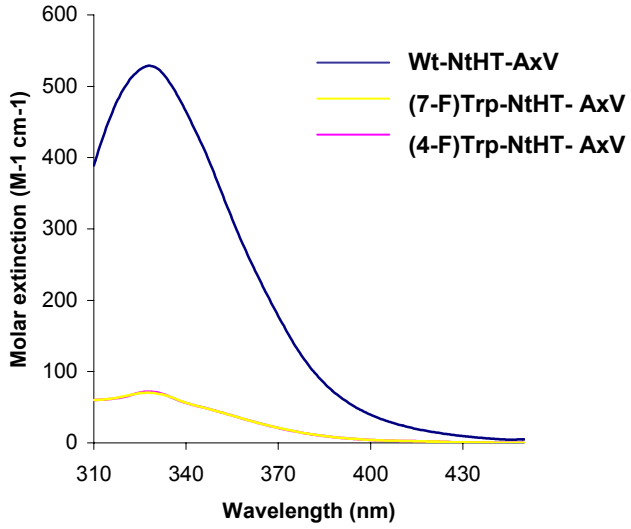


Figure 3.17. Fluorescence emission profiles of wt and variant proteins upon excitation at $\lambda = 295$ nm

In contrast to UV absorbance, the Trp fluorescence of Annexin V is abolished upon incorporation of (4-F)Trp and (7-F)Trp (Fig 3.17). Native Annexin V gives upon excitation at 295 nm an emission spectrum with maximum centred at 333 nm, while in (4-F)Trp and (7-F)Trp variants there is a total loss of Trp fluorescence, thus revealing the contribution of Tyr to the protein fluorescence emission profiles.

3.2.5 Secondary structure analyses and unfolding profiles of Annexin V and its variants

Thermal denaturation of native Annexin V and its fluoro variants was investigated by following changes in the dichroic intensities at 222 nm upon heating. To prevent protein aggregation at high temperatures, each probe was supplied with 10% glycerol to prevent protein aggregations at high temperatures.

The melting point of both native Annexin V and (4-F)Trp and (7-F)Trp variants is about 48°C (see Figure 3.18), but the large differences in van´t Hoff enthalpy values clearly indicate a higher cooperativity in unfolding of wt-Annexin V (see Table 3.3). Indeed, in other previous experiments performed in our laboratory it has been already observed that such a single atomic mutation can lead to quite large differences in ΔH_m values accompanied by small ΔT_m shifts (Minks et al., 1999; Budisa et al., 1998).

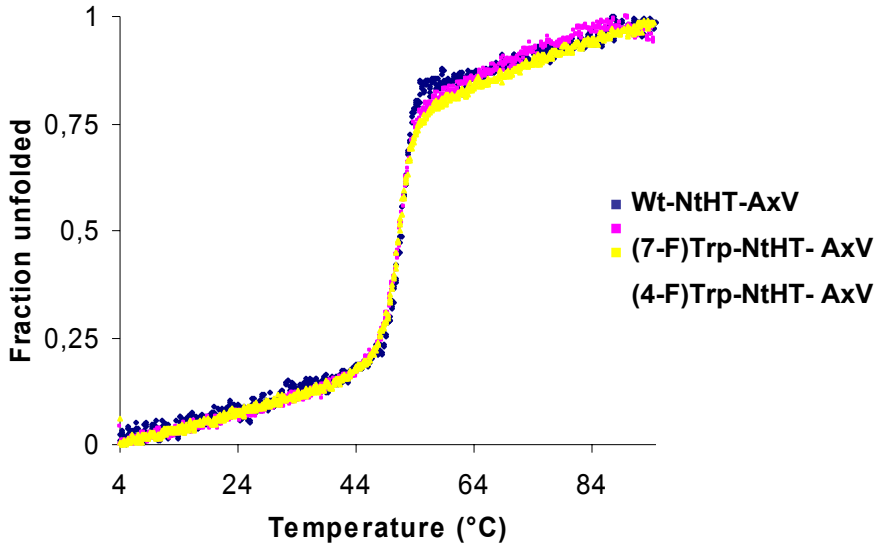


Figure 3.18. Normalized thermally-induced unfolding profiles of wt-Annexin V and its fluorinated variants. Although the T_m values are almost unchanged, the steepness of unfolding curves is different in variants when compared to wt-protein (see also Table 3.3).

Protein	Tm (K)	ΔT_m(K)	ΔH_m (kJ/mol)	$\Delta\Delta H_m$(kJ/mol)
Wt-Annexin V	324.80 ± 0.02	-	-803.5 ± 13.3	-
(4-F)Trp-Annexin V	324.12 ± 0.05	-0.68	-476.2 ± 15.9	-327.3
(7-F)Trp-Annexin V	324.59 ± 0.03	-0.31	-550.1 ± 13.5	-253.4

Table 3.3. Thermodynamic parameters derived by van´ t Hoff analyses for unfolded Annexin V and its variants

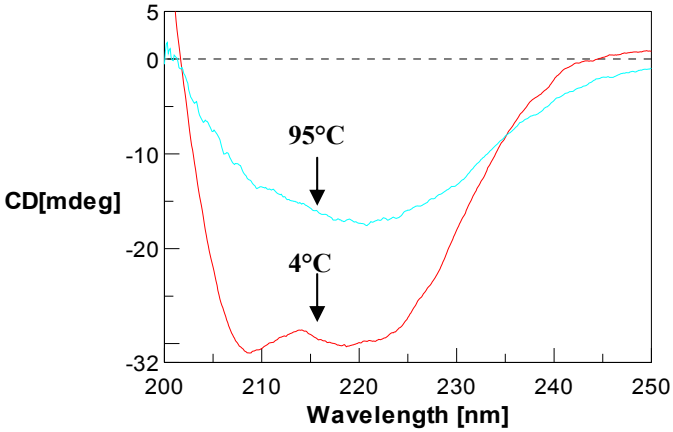
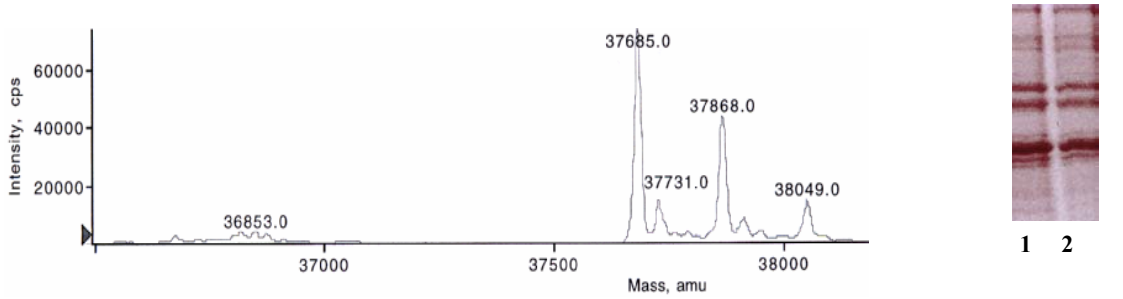


Figure 3.19. Secondary structure profiles at 4°C (native state) and 95°C (denaturated state) of (7-F)Trp-Annexin V, measured by far-UV circular dichroism.

3.2.6 Incorporation of (4-NH_2)Trp and (5-NH_2) Trp in Annexin V

The incorporation of (4-NH_2)Trp and (5-NH_2)Trp in NtHT-Annexin V was performed using the SPI method. Twelve hours after induction and addition of the non canonical substrates, the measured OD_{600} was only from 10 to 20 % higher in comparison to OD_{600} before the induction. The protein yields were also poor if compared to wt-protein.

The identity of the substituted amino-proteins was checked by mass analyses, but the results turned out to be ambiguous (Calculated mass: 37690 Da; Found: 37685). Moreover, beside the main peak, it was also detected the presence of other species with different molecular mass, most probably due to non specific educts (Fig. 3.20).



1: cells after expression of (4-NH_2)Trp-NtHT- AxV

2: cells after expression of (5-NH_2)Trp-NtHT-AxV

Figure 3.20. Deconvoluted ESI-MS profile of (5-NH_2)Trp-NtHT-Annexin V. NtHT-Annexin V substituted with (4-NH_2)Trp (not shown) displays an identical profile. Insert: expression profile in cell lysates of *E.coli ATCC49980.*

3.2.7 Spectral properties of amino-substituted Annexin V

Amino-substituted Tryptophans have peculiar spectral properties due to the presence of the highly polar amino group conjugated with the π- electrons system of the indole moiety. These unique characteristics are delivered to the entire protein upon incorporation of the single building block. Therefore, absorbance and fluorescence profiles can be used as an unambiguous analytical criterion to confirm the incorporation of aminotryptophans. As the spectroscopic mass results were ambiguous, it was very important in this case to use another method to monitor the success of the incorporation.

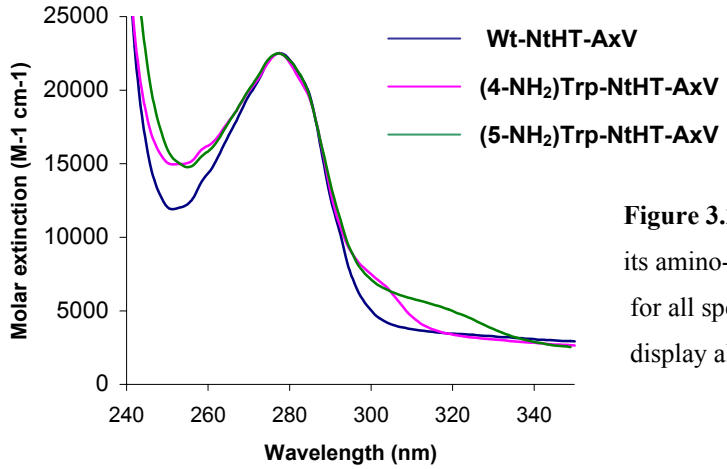


Figure 3.21. UV-spectra of native Annexin V and its amino-variants.The maximum at 277 nm coincides for all species. (4-NH_2)Trp-, and (5-NH_2)Trp-Annexin V display also a prominent red-shifted spectral shoulder.

Indeed, at pH 7 and 9.0, 5-aminoindole and to a less extent also 4-aminoindole, exhibit a large red-shifted spectral shoulder due to the delocalization of the nitrogen charge within the π-cloud of

the aromatic moiety. The typical wavelength maxima for these additional shoulders are centred around 315 nm for (5-NH_2)Trp and 305 nm for (4-NH_2)Trp (Fig. 3.22 and 3.23).

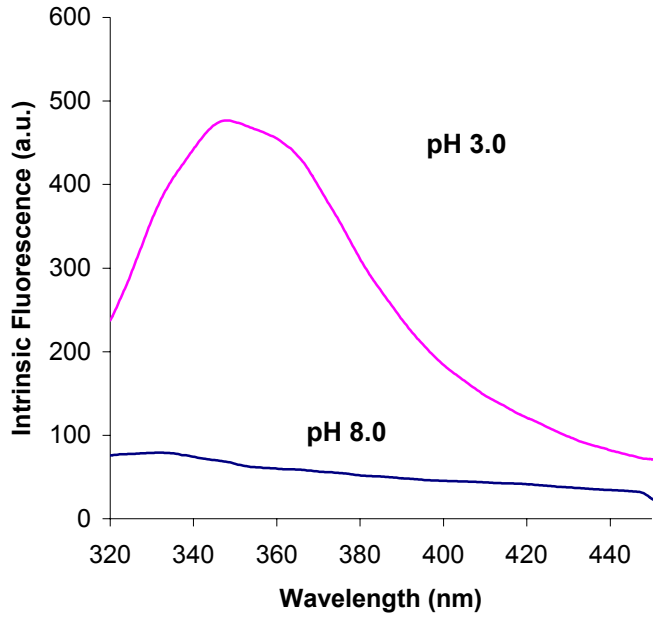


Figure 3.22. Fluorescence emission profiles of (4-NH_2)Trp-NtHT-Annexin V at pH 8.0 and pH 3.0. While the fluorescence intensity is high at low pH values, it becomes very low and blue-shifted by increasing pH of the protein sample. The samples were excited at λ = 305 nm with slit width 5.0

At pH 8.0, where the anionic form is predominant, the fluorescence of (4-NH_2)Trp-NtHT-Annexin V is almost not detectable any more. The fluorescence emission intensity of (5-NH_2)Trp-NtHT-Annexin at pH 8.0 shows the same trend, even if the decrease of the intensity is not so dramatic.

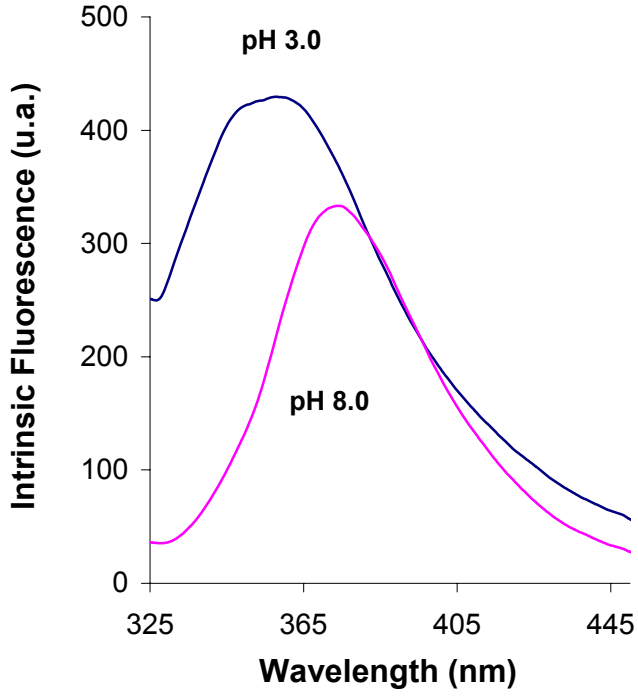


Figure 3.23. Fluorescence emission profiles of (5-NH_2)Trp-NtHT-Annexin V at pH 8.0 and pH 3.0. Note that at low pH values the fluorescence intensity is increased, while at higher pH values the emission maximum is about 20 nm red-shifted. The samples were excited at λ = 315 nm with slit width 5.0.

For (5-NH_2)Trp-NtHT-Annexin V, the emission maximum is centred at 355 nm at pH 3.0, while at pH 8.0 it is positioned at 375 nm with a 20 nm red-shift (Figure 3.23). Conversely, at pH 8.0 the incorporation of (4-NH_2)Trp causes a blue-shift. In this case, it is difficult to precisely evaluate the extent of the shift because of the lower fluorescence intensity of this protein at neutral and basic pH values.

3.2.8 Secondary structure analyses and unfolding profiles of substituted amino-Annexin

The incorporation of (4-NH_2)Trp and (5-NH_2)Trp into Annexin V led to a general destabilization of the

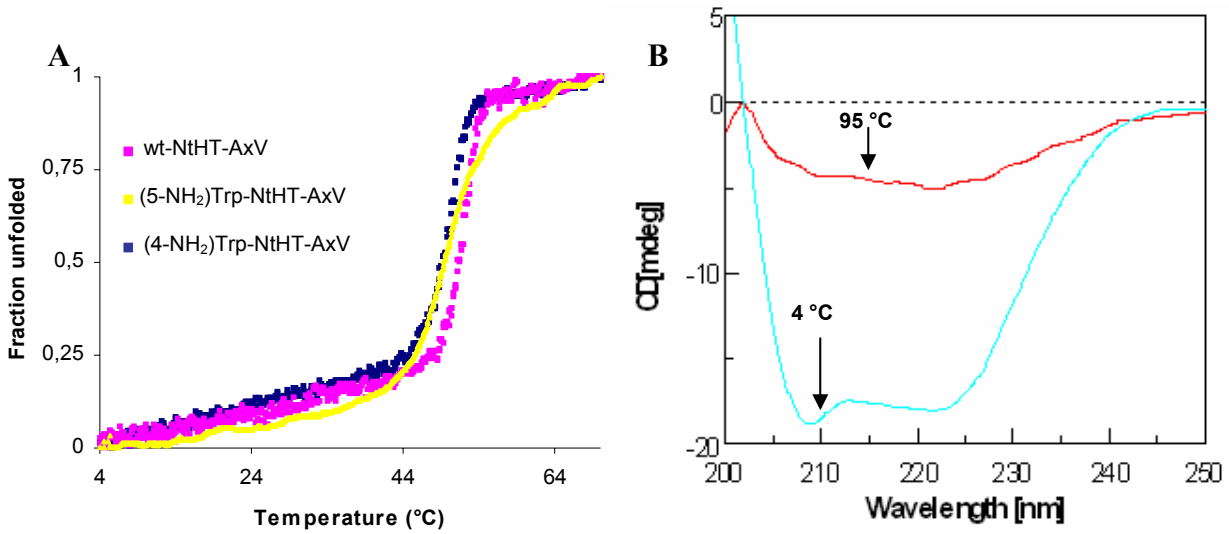


Figure 3.24. A) Melting curves of native Annexin and its amino-variants. The curves were recorded by monitoring the conformational changes at 222 nm from 4 °C to 70 °C. **B**) Secondary structure profiles of native (4°C) and denatured (95°C) state of (5-NH2)Trp-NtHT-AxV, measured by far-UV circular dichroism.

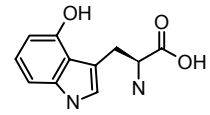
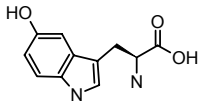
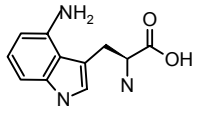
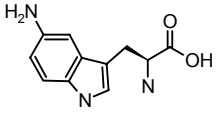
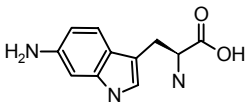
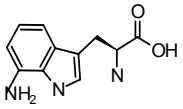
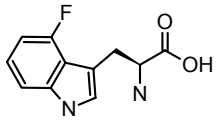
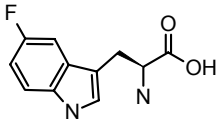
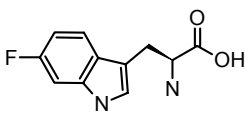
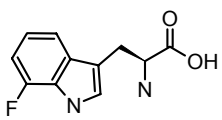
Protein	Tm (K)	ΔT_m(K)	ΔH_m (kJ/mol)	$\Delta\Delta H_m$(kJ/mol)
Wt-Annexin V	324.80 ± 0.02	-	-803.5 ± 13.3	-
(4-NH_2)Trp-Annexin V	323.04 ± 0.02	-1.76	-625.5 ± 18.2	-178
(5-NH_2)Trp-Annexin V	322.40 ± 0.03	-2.40	-353.5 ± 16.1	-450

Table 3.4. Thermodynamic parameters derived by van´t Hoff analyses of unfolded profiles of native Annexin V and its variants, presented in Figure 3.24.

substituted proteins. Once again, the melting points of Annexin variants do not differ much from the one of the native protein, but the van´ t Hoff enthalpy values are significantly dropped down (see Table 3.4). As it can be observed in Figure 3.24, the native protein displays the steepest denaturation curve, thus confirming that the wt-protein unfolding is more cooperative.

3.2.9 Incorporation of aromatic non canonical amino acids into Barstar

Barstar contains in its sequence, three Trp residues at different positions and they all contribute to different extents to protein folding and to the global protein stability. Therefore, native Barstar and its mutants were chosen as target to study the effects of the incorporation of Trp analogues in protein structure, stability and spectroscopic properties (see Table 3.5).

Tryptophan analogues	**Structure**	**Incorporation capacity**
4-hydroxytryptophan		+
5- hydroxytryptophan		+
4-aminotryptophan		+
5-aminotryptophan		+
6-aminotryptophan		-
7-aminotryptophan		-
4-fluorotryptophan		+
5-fluorotryptophan		+
6-fluorotryptophan		+
		+

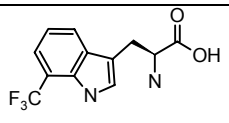
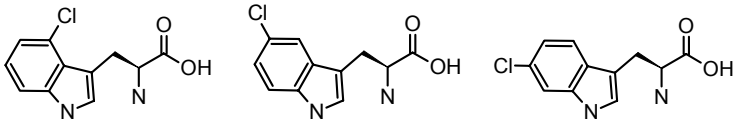
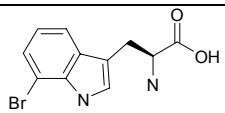
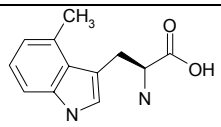
7-fluorotryptophan		+
7′-7′-7′- trifluoromethyltryptophan		-
4-chlorotryptophan		-
5- chlorotryptophan		-
6- chlorotryptophan		-
7-bromotryptophan		-
4-methyltryptophan		+

Table 3.5. Chemical formula, name and incorporation capacity of various Trp analogues used for Barstar labelling.

3.2.10 Incorporation of 4- and 5-hydroxytryptophan into Barstar

The incorporation of (4-OH)Trp and (5-OH)Trp into wt-Barstar following the routine protocols of the SPI method, led to a partial replacement of all three Trp residues and to proteins yields fully comparable to those of wild-type. As detected from the mass analyses, three profiles are to be recognized with a predominance of species containing two (10285±2.0) or three (10300±2.4) modified residues. Traces of proteins containing only one substituted Trp residue (10269±2.4) have also been found, while native protein was never detected (Figure 3.25).

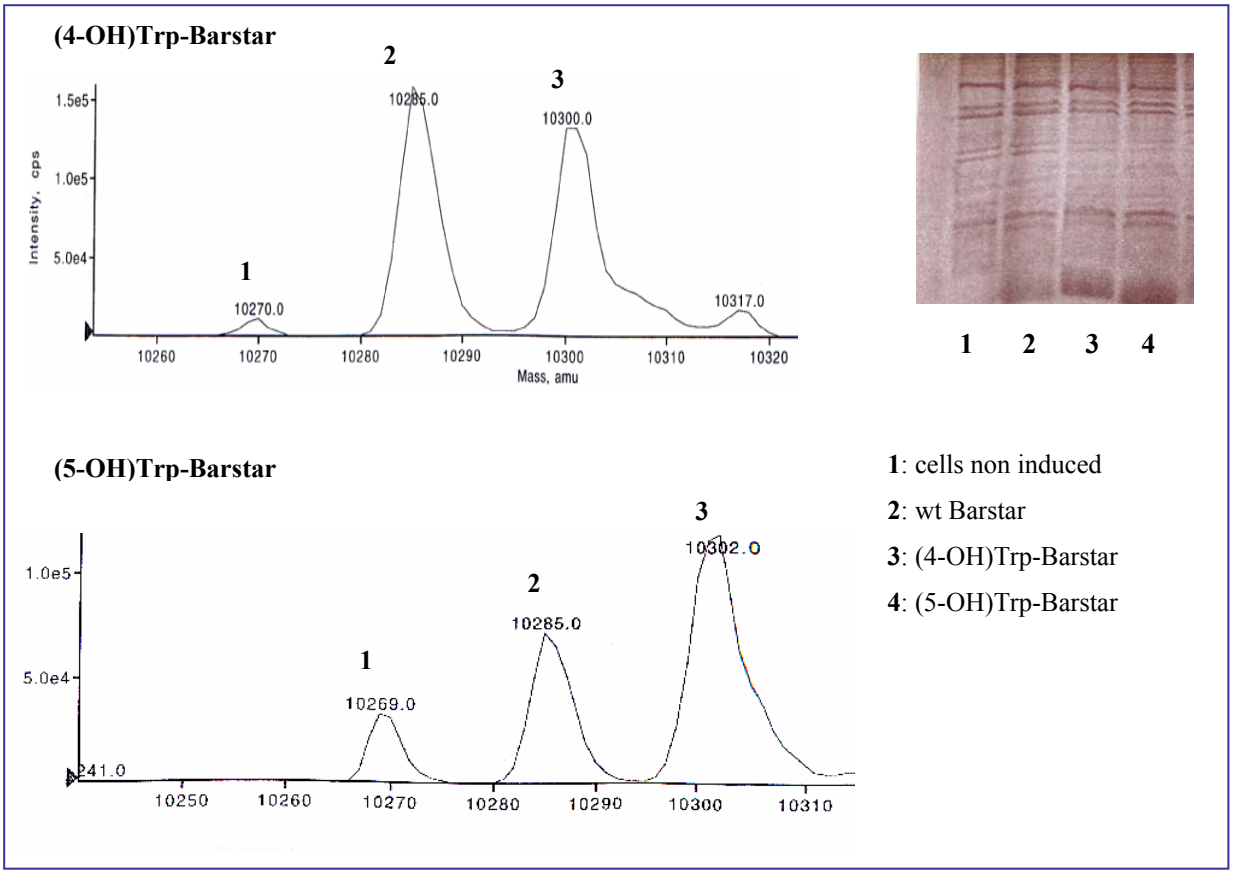


Figure 3.25. Mass spectrometric profiles of protein samples containing (4-OH)Trp and (5-OH)Trp. Insert: expression profile in cell lysates of *E.coli ATCC49980.*

3.2.11 Spectroscopic features of (4-OH)Trp and (5-OH)Trp-Barstar

The UV and fluorescence spectra of the substituted proteins have been taken in order to analyze the qualitative extent of the analogues incorporation and the delivery of new spectroscopic properties to the whole protein due to the polar hydroxyl group. In particular, the UV spectrum of (4-OH)Trp-Barstar shows a prominent shoulder at 298 nm, while (5-OH)Trp- Barstar at 310 nm, due to a pronounced enhancement of the $^{1}L_b$ transition band. This is a useful and simple preliminary criterion to distinguish both variants from the wt protein and from each other (Fig. 3.26).

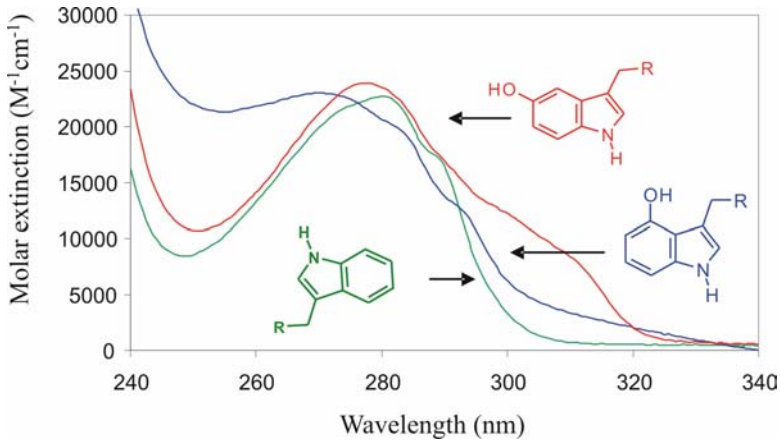


Figure 3.26. UV absorbance profiles of native and substituted Barstar. The presence of the hydroxy-group endows the protein with novel properties (Budisa et al., 2002).

In order to study the behaviour of (4-OH)Trp- and (5-OH)Trp-Barstar under different pH conditions, a pH titration was performed in the range between pH 1.0 and pH 11.0 and related fluorescent spectra have been measured.

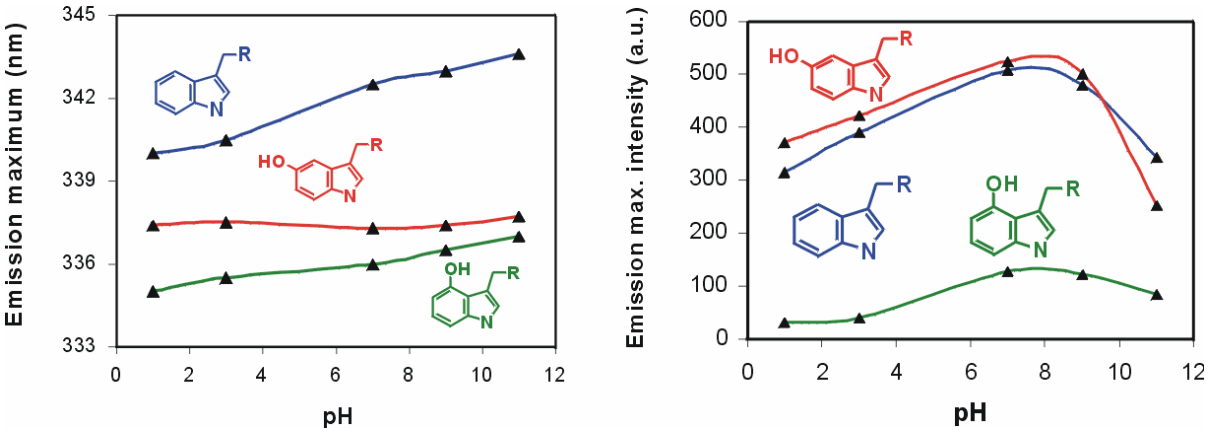


Figure 3.27. Changes in fluorescence emission of wt and substituted Barstar upon different pH values (left). Changes in relative fluorescence intensity (right). All protein samples were excited at 280 nm (Budisa et al., 2002).

By careful inspection of the plot in Fig. 3.27, it becomes evident that substituted proteins are pH insensitive, since there are no relevant shifts in the fluorescence emission maximum wavelength. The patterns showed by native and (5-OH)Trp-Barstar are almost identical concerning the intensity maxima and can be regarded as gradual and monotonic, while the poor intensity of (4-OH)Trp-Barstar can be explained through its strong tendency to aggregate (Budisa et al., 2002). The modest changes in the absorbance and emission spectra displayed by these Barstar variants in relation to pH variations, give evidence for a rather poor delocalisation of the free electron pair into the π-electron system of the indole moiety.

3.2.12 Incorporation of fluorinated Tryptophan analogues into Barstar

The Trp analogues (4-F)Trp, (5-F)Trp, (6-F)Trp, (7-F)Trp were used for incorporation studies in wt-Barstar and in the W38F, W44F, W3844F variants, while the incorporation of 7′-7′-7′-trifluoromethyltryptophan was tested only in wt-Barstar.

For 7′-7′-7′- trifluoromethyltryptophan, no expression could be detected, while all other substrates have been successfully incorporated into wt-Barstar and in W38F Barstar mutant. In Barstar mutant W44F only (5-F)Trp could not be incorporated, while in the double mutant W3844F all incorporation attempts were not successful.

3.2.13 Expression and mass analyses of fluorinated Barstars

The expression of wt and mutant Barstar was performed in NMM with 0.015mM Tryptophan as limiting concentration. After 8-10 hours, after total depletion of the natural amino acid, cells were ready for the induction (OD_{600} 0.7-0.9) and the fluorinated analogues were added. Induced cells were allowed to shake overnight and the next OD_{600} measurement always showed that cells had still duplicated twice or three times after the induction. The yield of fluorinated proteins is fully comparable to the native one, both for wt-Barstar and mutants.

The mass analyses always showed one single peak corresponding to the expected protein (fluorinated P27A: calculated mass 10307 Da; found 10307 Da; fluorinated mutants, both W38F and W44F: calculated mass: 10249 Da; found 10250 Da) (Figure3.28 and 3.29).

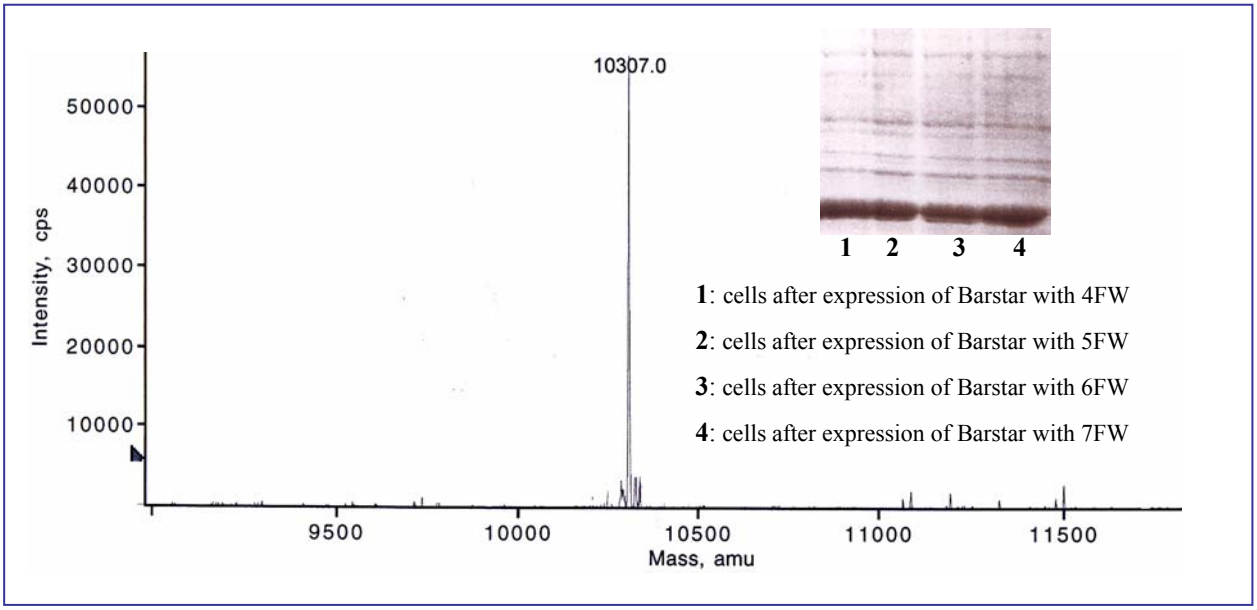


Figure 3.28. Deconvoluted mass profile of (4-F)Trp-Barstar. Insert: expression profile of (4-F)Trp-, (5-F)Trp-, (6-F)Trp-, and (7-F)Trp-Barstar.

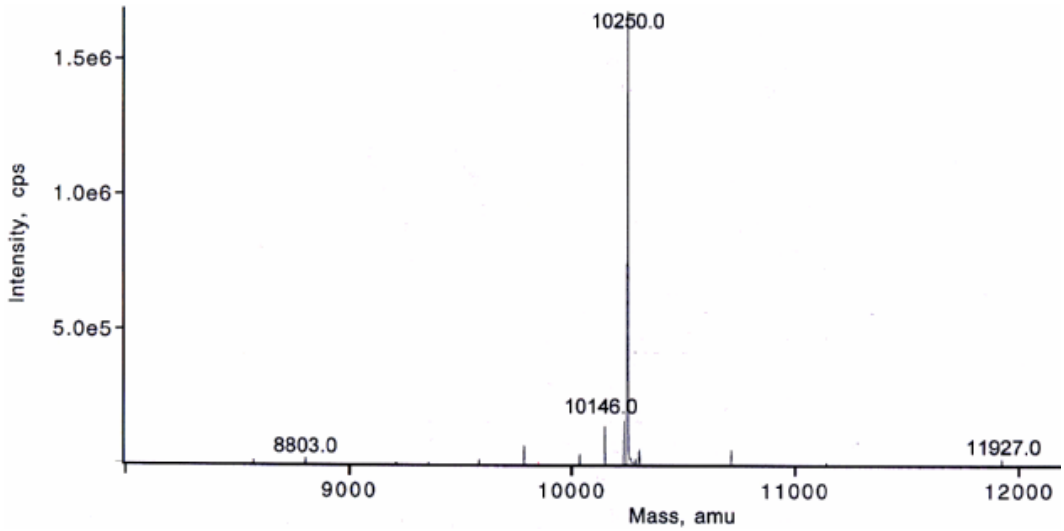


Figure 3.29. Deconvoluted mass profile of (7-F)Trp-W44F-Barstar. Protein samples containing fluorinated Barstar mutants exhibit the same profile.

3.2.14 Spectral properties of fluorinated Barstars

The incorporation of fluorotrytophans into Barstar brought evident changes in the spectroscopic features of the substituted proteins.

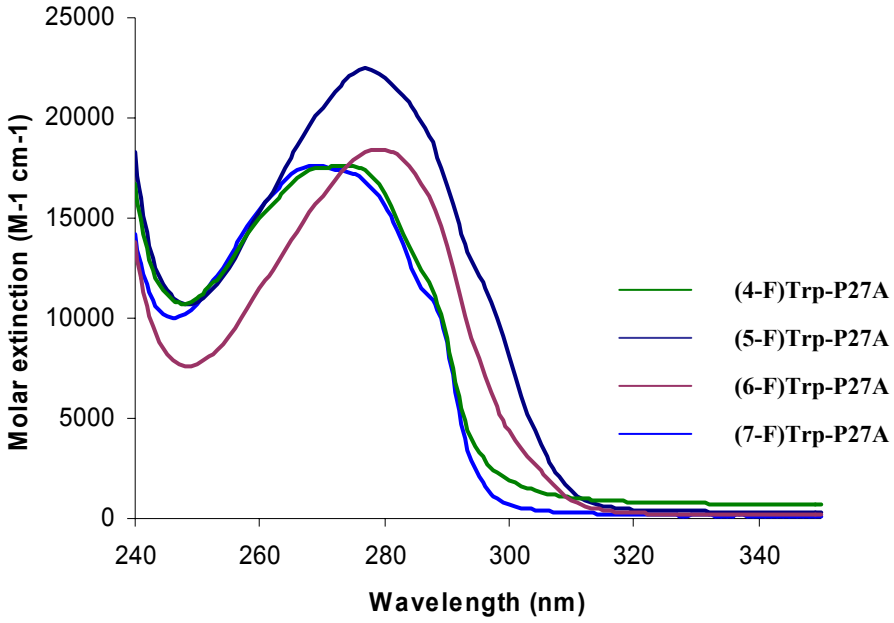


Figure 3.30. UV absorbance profiles of fluoro-substituted Barstar.

The introduction of (5-F)Trp in both wt-Barstar and its mutants caused a slight red-shift with λ_{max} positioned between 280 and 283 nm, while incorporations of (4-F)Trp and (7-F)Trp resulted in a pronounced blue-shift with λ_{max} positioned around 270 nm (Figure 3.30 and 3.31).

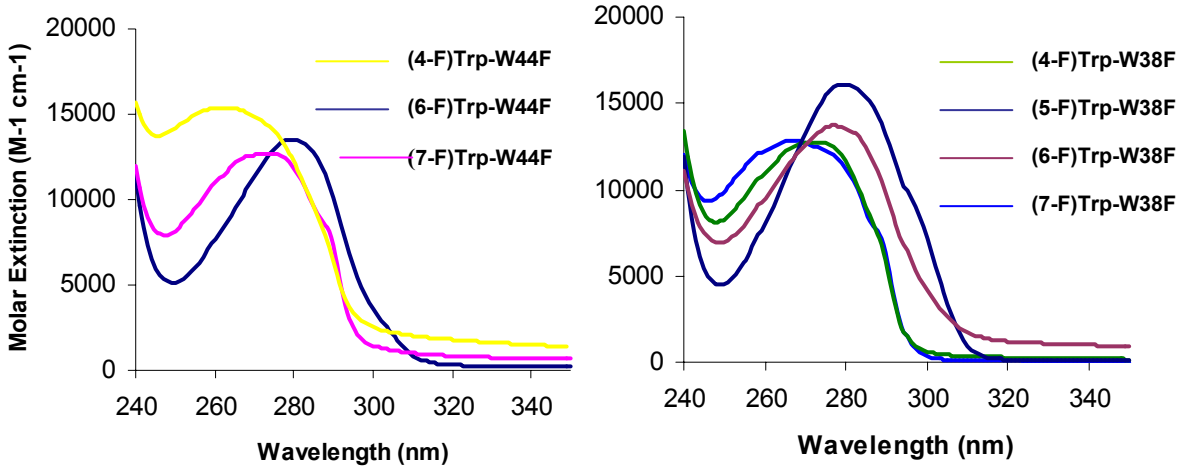


Figure 3.31. UV absorbance profiles of fluoro-substituted Barstar mutants.

The fluorescence of fluoro-substituted Barstars presents also other features in comparison to the parent proteins. The incorporation of (5-F)Trp and (6-F)Trp cause a red-shift of about 10 nm in the emission maximum (λ_{max}), while (4-F)Trp and (7-F)Trp behave as "silent fluorophors", thus revealing the contribution to fluorescence emission of other residues like Tyr or Phe.

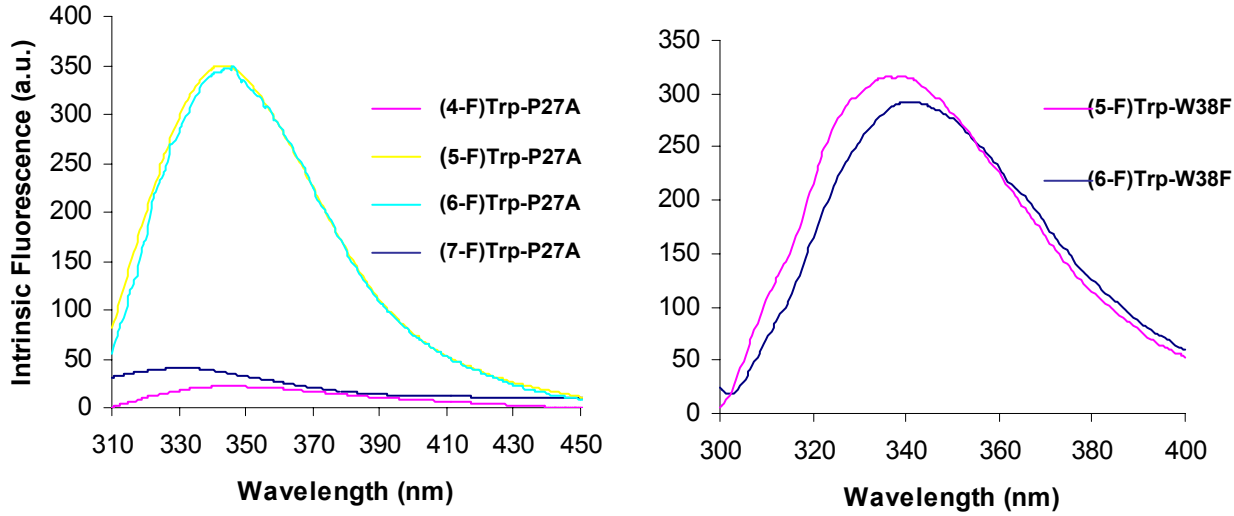


Figure 3.32. Fluorescence emission profiles of fluoro-substituted Barstar and its mutant W38F. The profile of fluorinated Barstar mutant W44F shows the same trend (red-shift about 10 nm)

3.2.15 Secondary structure analyses and unfolding profiles of fluorinated Barstars

The incorporation of fluorotryptophans into wt-Barstar and its mutants, led to different structural and thermodynamic effects. As it can be seen in Figures 3.33 and 3.34, the introduction of (4-F)Trp, (5-F)Trp, and (6-F)Trp into wt-Barstar resulted in a destabilization of the protein, while (7-F)Trp caused an increase of the melting point of about 3.5 °C.

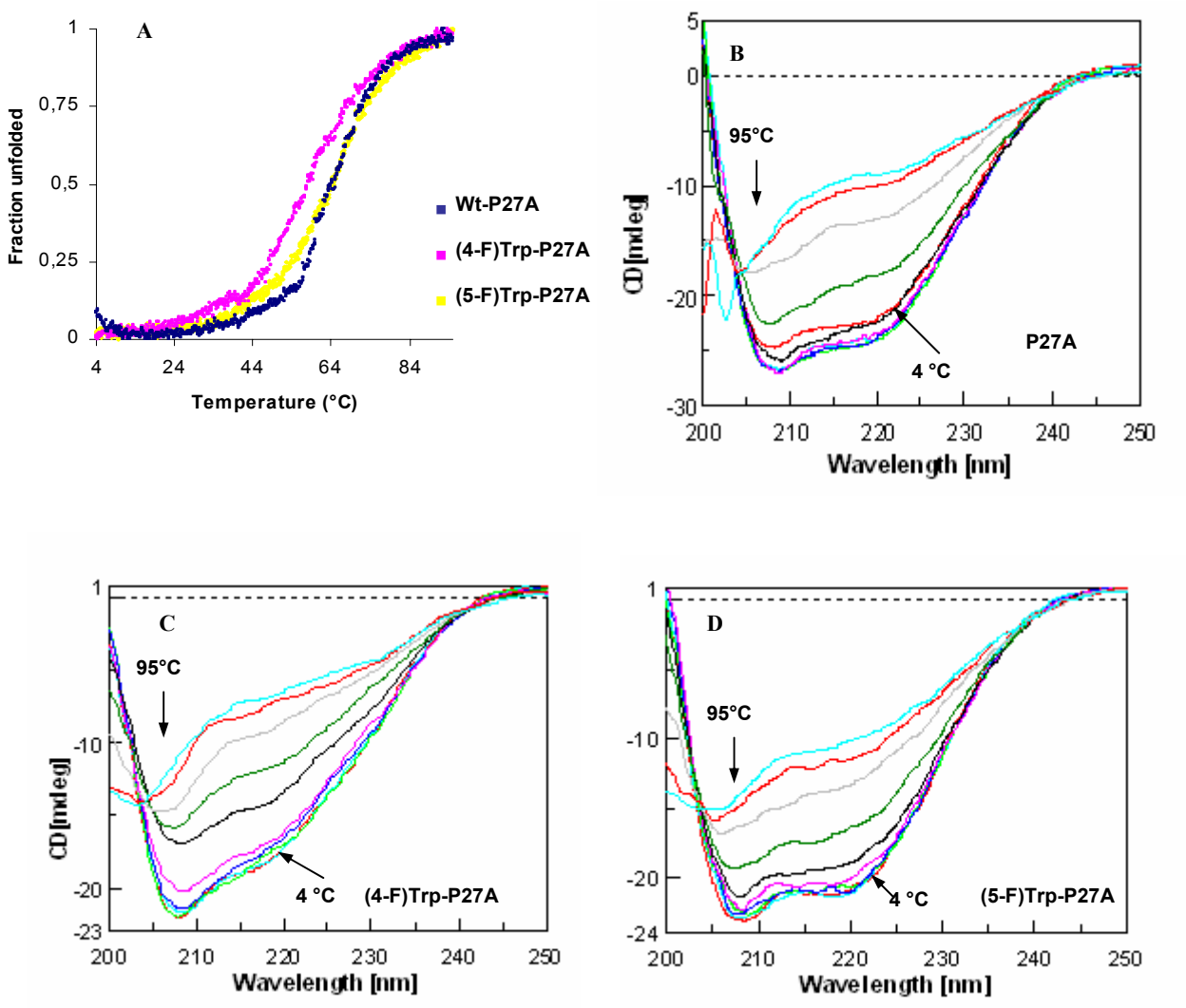


Figure 3.33. A) Unfolding profiles recorded by measurement of dichroic changes at 222 nm upon heating of the protein sample from 4 to 95 °C. Changes in secondary structure of wt-Barstar **B**) (4-F)Trp-Barstar **C**) and (5-F)Trp-Barstar **D**) upon heating in temperature range between 4-95°C.

Conversely, the enthalpy values are lowered, thus showing that the protein unfolding is less cooperative in comparison to wild type protein.

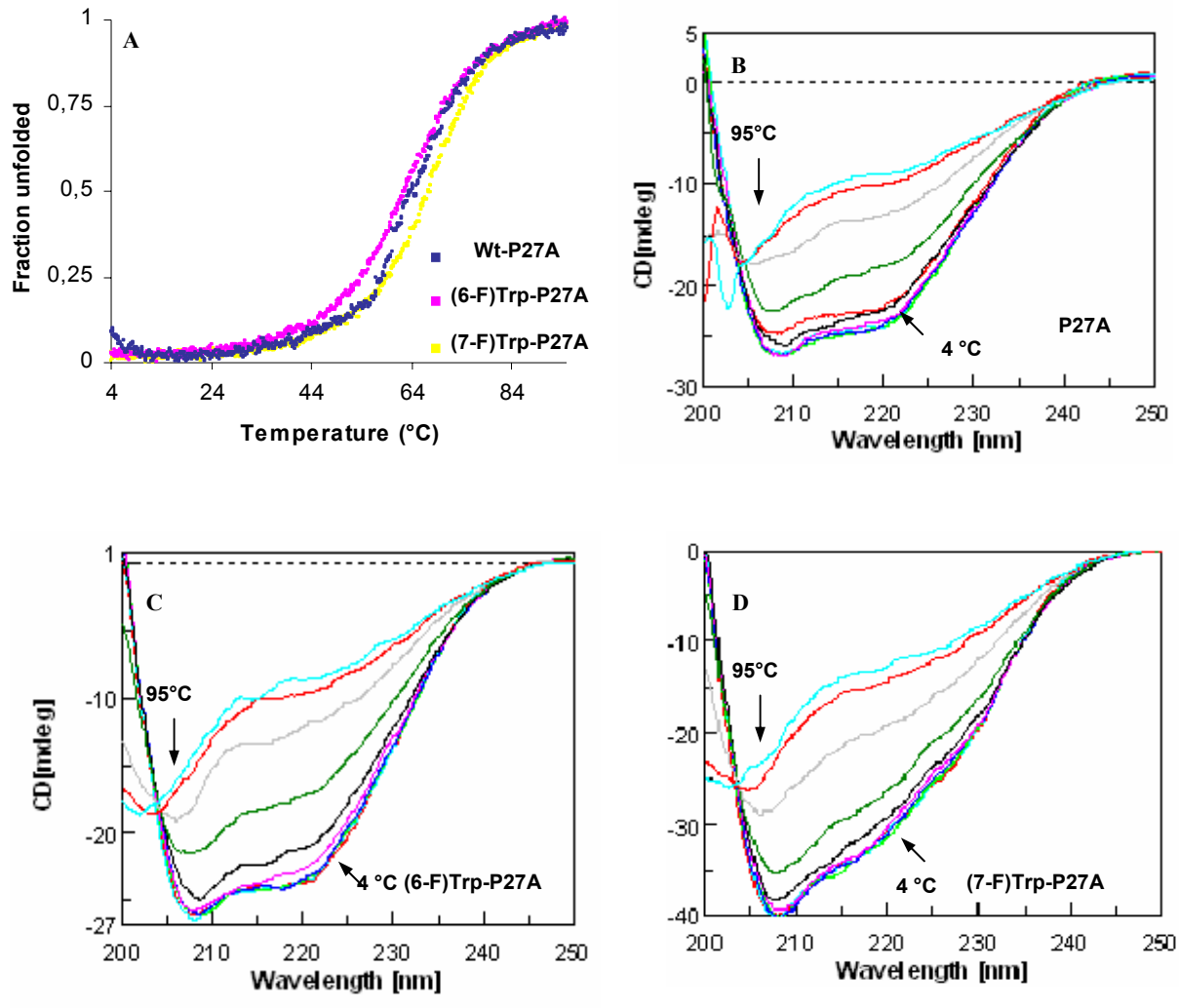


Figure 3.34. A) Unfolding profiles recorded by measurement of dichroic changes at 222 nm upon heating of the protein sample from 4 to 95 °C. Changes in secondary structure of wt-Barstar **B**) (6-F)Trp-Barstar **C**) and (7-F)Trp-Barstar **D**) upon heating in temperature range between 4-95°C.

Protein	Tm (K)	ΔT_m(K)	ΔH_m (kJ/mol)	$\Delta\Delta H_m$(kJ/mol)
P27A wt	338.77 ± 0.03	-	-183.1 ± 1.9	-
(4-F)Trp-P27A	329.02 ± 0.16	-9.75	-99.6 ± 3.7	-83.5
(5-F)Trp-P27A	337.73 ± 0.08	-1.04	-118.4 ± 1.5	-64.7
(6-F)Trp-P27A	336.33 ± 0.1	-2.44	-131.4 ± 2.7	-51.7
(7-F)Trp-P27A	342.19 ± 0.09	3.42	-147.4 ± 2.2	-35.7

Table 3.6. Thermodynamic parameters derived by van ´t Hoff analyses for wt-Barstar and its fluoro-substituted variants. Unfolding profiles are shown in Figure 3.33 and 3.34

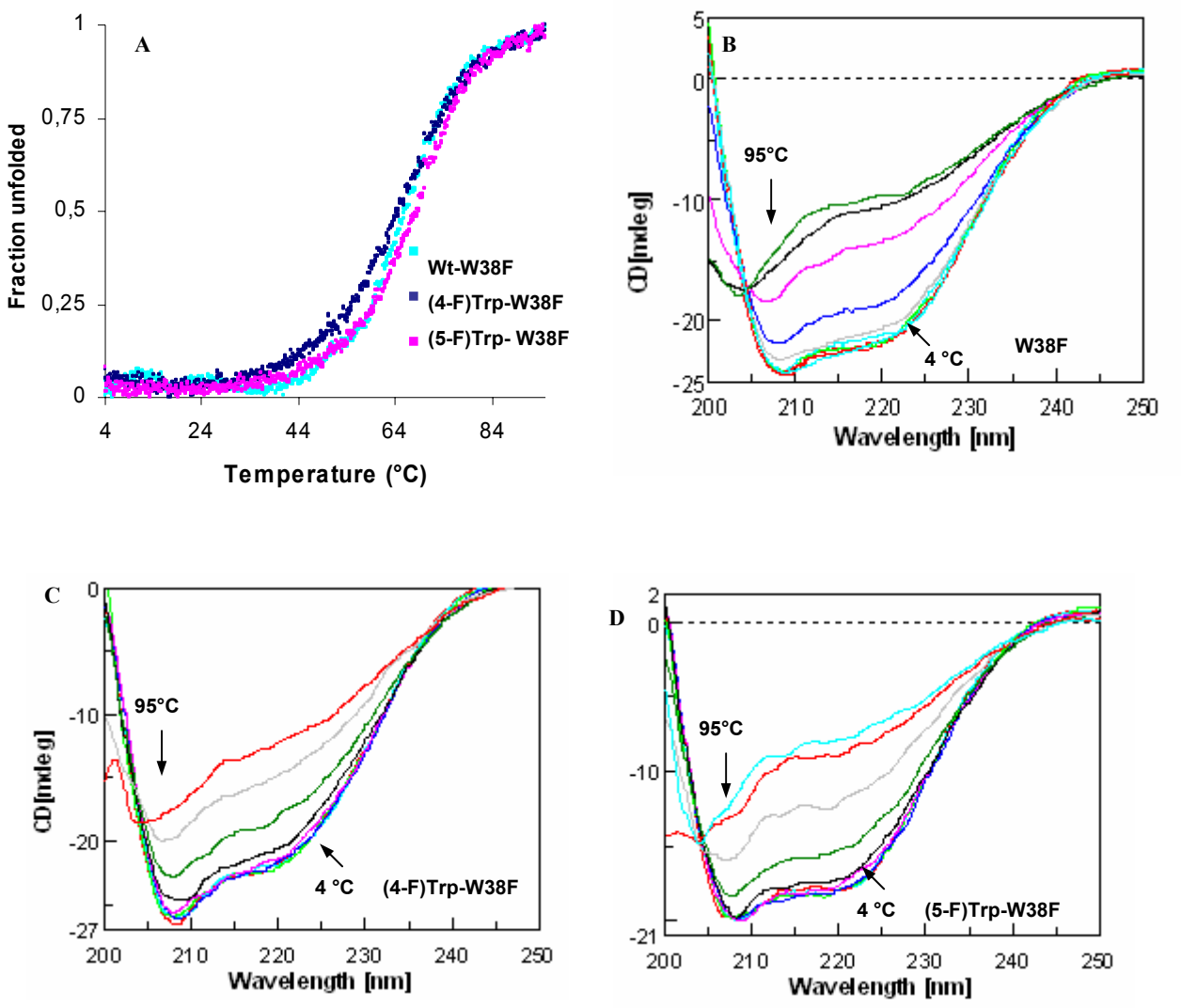


Figure 3.35. A) Unfolding profiles recorded by measurement of dichroic changes at 222 nm upon heating of the protein sample from 4 to 95 °C. Changes in secondary structure of W38F **B**) (4-F)Trp-W38F **C**) and (5-F)Trp-W38F **D**) upon heating in temperature range between 4-95°C.

The incorporation of (4-F)Trp into the mutant W38F did not lead to substantial changes in the melting point, while (6-F)Trp caused a remarkable decrease in T_m of about 5 °C. On the other hand, (5-F)Trp and (7-F)Trp increased the melting temperature of average 3.5 °C.

The incorporation of (7-F)Trp into W38F led even to a slight increase in van´t Hoff enthalpy of about 3% when compared to wild type protein (Table 3.7).

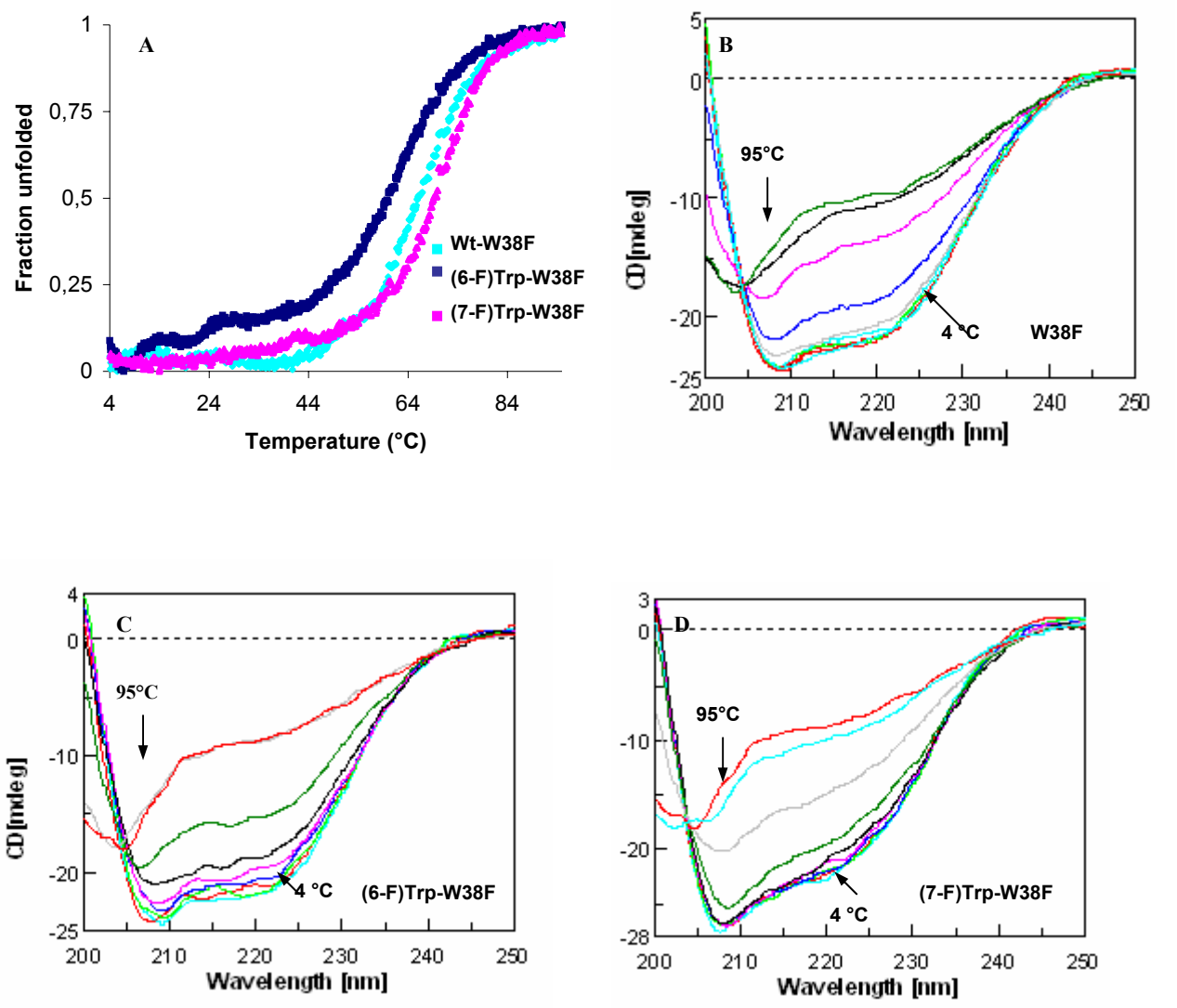


Figure 3.36. A) Unfolding profiles recorded by measurement of dichroic changes at 222 nm upon heating of the protein sample from 4 to 95 °C. Changes in secondary structure of wt-W38F **B**) (6-F)Trp-W38F **C**) and (7-F)Trp-W38F **D**) upon heating in temperature range between 4-95°C.

Protein	Tm (K)	ΔT_m(K)	ΔH_m (kJ/mol)	$\Delta\Delta H_m$(kJ/mol)
W38F	340.40 ± 0.04	-	-176.2 ± 2.0	-
(4-F)Trp-W38F	340.34 ± 0.09	-0.06	-110.8 ± 1.6	-65.4
(5-F)Trp-W38F	343.84 ± 0.21	3.44	-131.7 ± 3.7	-44.5
(6-F)Trp-W38F	335.76 ± 0.11	-4.64	-164.1 ± 4.8	-12.1
(7-F)Trp-W38F	344.82 ± 0.06	4.42	-180.9 ± 1.9	4.7

Table 3.7. Thermodynamic parameters derived by van ´t Hoff analyses for W38F and its fluoro-substituted mutants

In the case of Barstar mutant W44F, all tested fluorotryptophans led to an increase of the Tm value. In addition, (6-F)Trp and (7-F)Trp even improved dramatically the cooperativity of the substituted proteins when compared to wt-W44F.

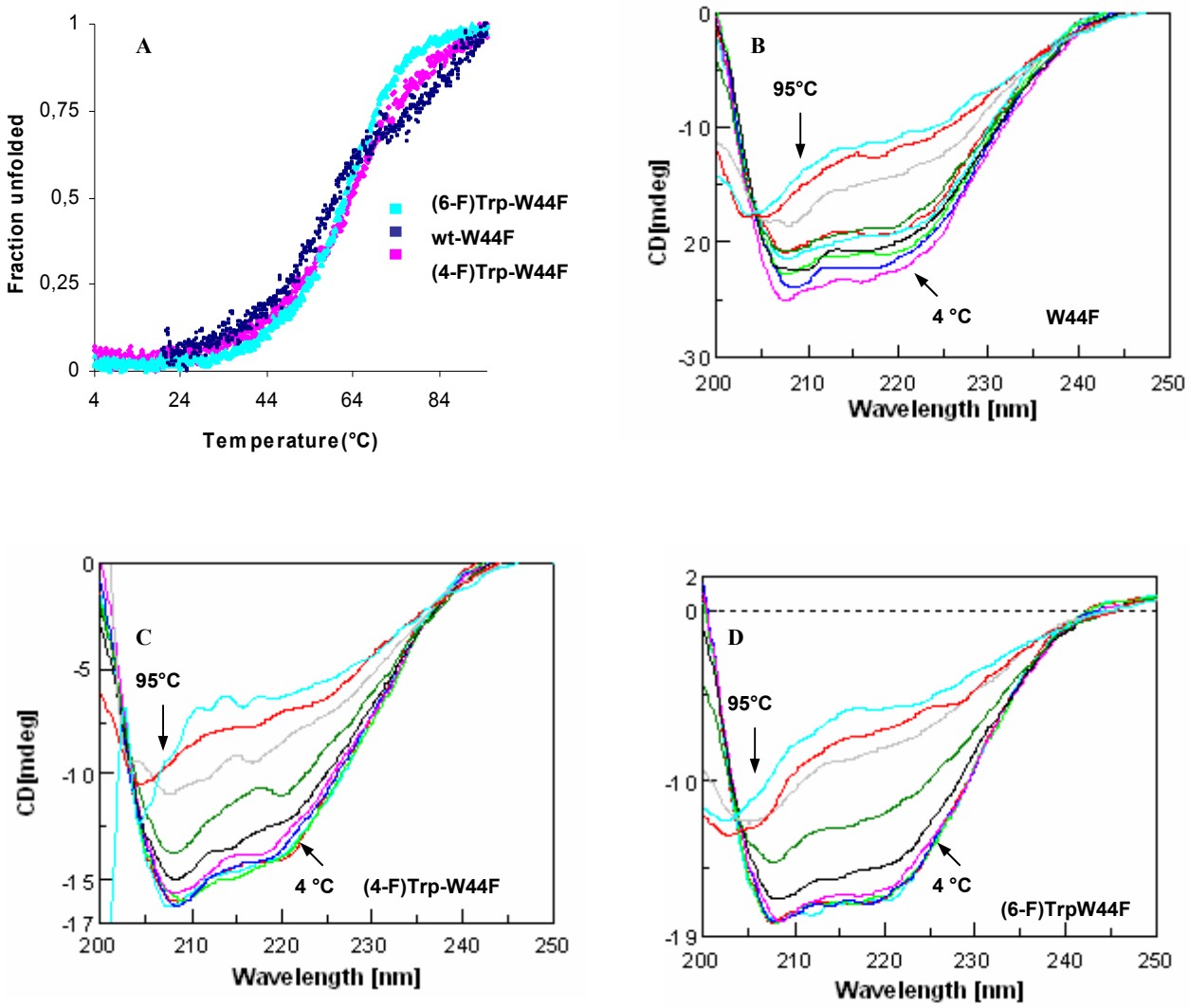


Figure 3.37. A) Unfolding profiles recorded by measurement of dichroic changes at 222 nm upon heating of the protein sample from 4 to 95 °C. Changes in secondary structure of wt-W44F **B**) (4-F)Trp-W44F **C**) and (6-F)Trp-W44F **D**) upon heating in temperature range between 4-95°C.

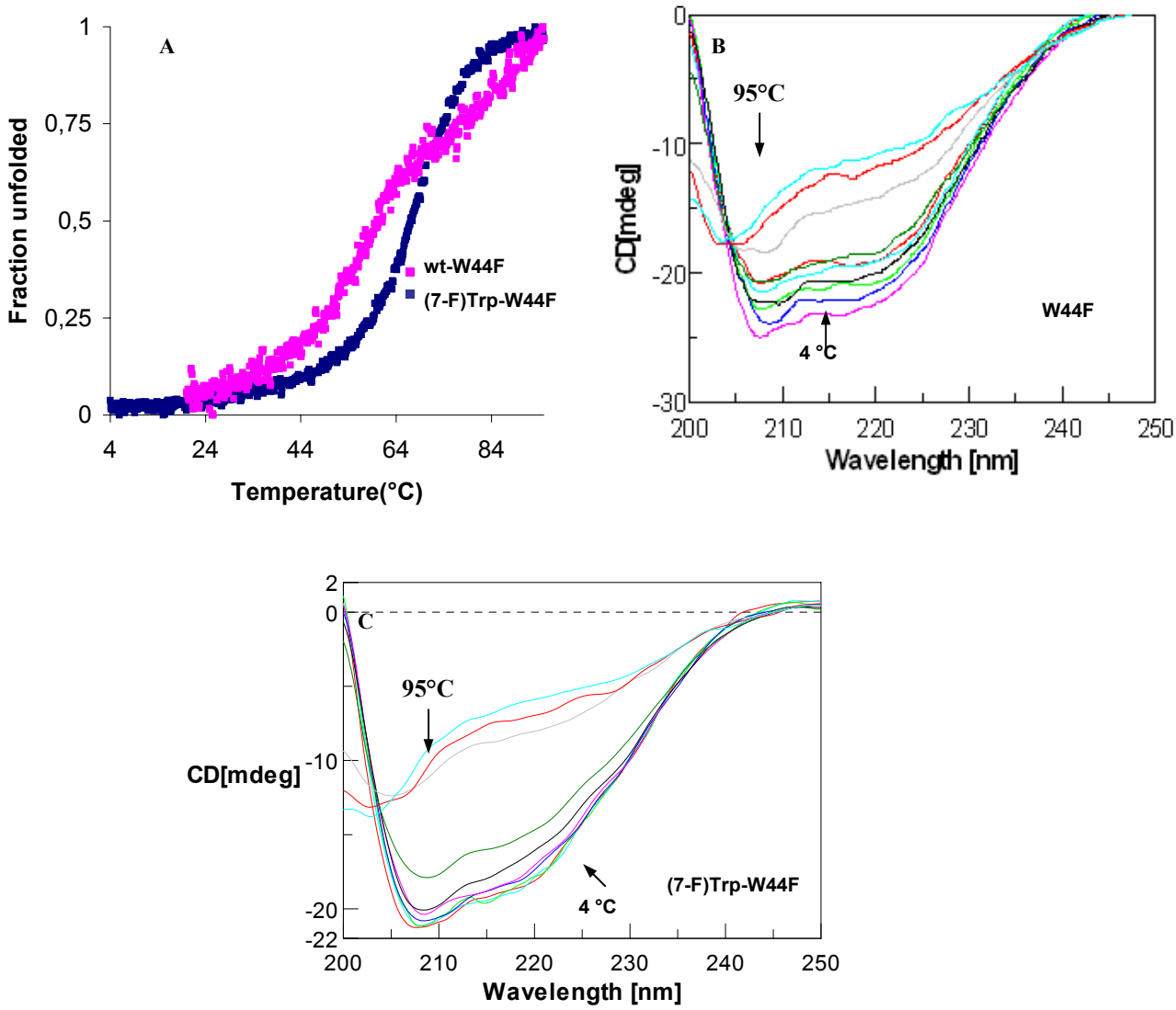


Figure 3.38. A) Unfolding profiles recorded by measurement of dichroic changes at 222 nm upon heating of the protein sample from 4 to 95 °C. Changes in secondary structure of wt-W44F **B**) (7-F)Trp-W44F **C**) upon heating in temperature range between 4-95°C.

Protein	Tm (K)	ΔT_m(K)	ΔH_m (kJ/mol)	$\Delta\Delta H_m$(kJ/mol)
W44F	334.63 ± 0.13	-	-103.9 ± 2.1	-
(4-F)Trp-W44F	337.62 ± 0.14	2.99	-103.6 ± 2.1	-0.3
(5-F)Trp-W44F	-	-	-	-
(6-F)Trp-W44F	336.79 ± 0.13	2.16	-127.5 ± 3.3	23.6
(7-F)Trp-W44F	342.33 ± 0.05	7.7	-164.5 ± 1.6	60.6

Table 3.8. Thermodynamic parameters derived by van ´t Hoff analyses for W44F and its fluoro-substituted mutants. Note the dramatic increase in van´t Hoff enthalpy of (7-F)Trp-W44F.

3.2.16 Incorporation of 4-methyltryptophan into Barstar

The incorporation of 4-methyltryptophan [(4-CH_3)Trp] into Barstar was successfully achieved exploiting SPI method with a protein yield fully comparable to wt. The results of mass analyses showed that (4-CH_3)Trp could be fully incorporated into Barstar (calculated mass: 10295 Da; found mass: 10297 Da) without any by-products.

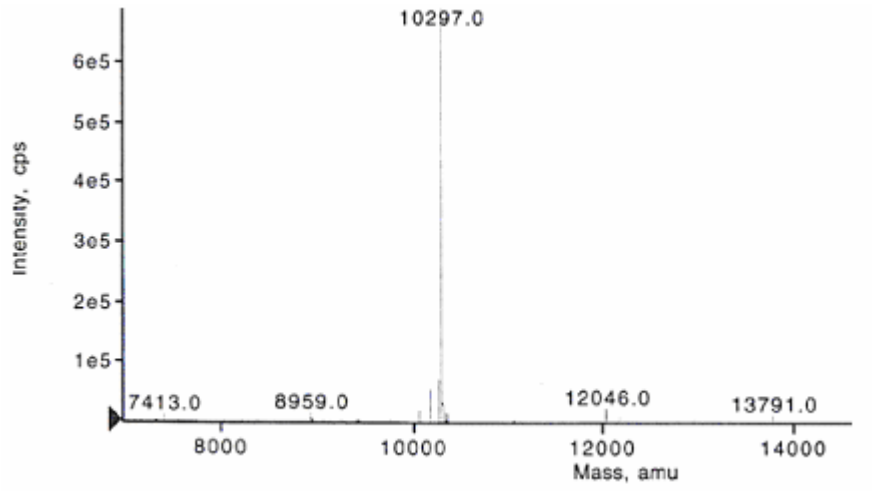


Figure 3.39. ESI-MS profile of fully substituted (4-CH_3)Trp- Barstar.

3.2.17 Spectral properties of (4-CH_3)Trp-Barstar

The spectral features of (4-CH_3)Trp-Barstar are mostly identical to those of the related wt protein. The UV absorbance maximum is slight blue-shifted of about 3 nm (λ_{max} = 277 nm), while the molar extinction remains unaltered (Fig. 3.40).

The fluorescence emission profiles are almost undistinguishable from wt protein concerning the position of the emission maximum (wt: λ_{max} = 338 nm; (4-CH_3)Trp-Barstar: λ_{max} = 342 nm), while the intensity is slightly decreased upon excitation at 295 nm.

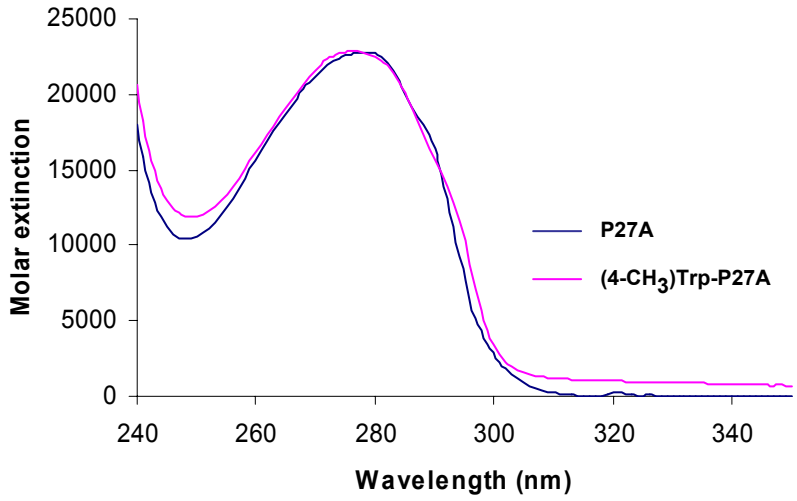


Figure 3.40. The UV absorbance profiles of wt – and (4-CH_3)Trp-Barstar are almost identical.

3.2.18 Unfolding profile of (4-CH_3)Trp-Barstar

The melting curve of (4-CH_3)Trp-Barstar was recorded by following the dichroic intensity at 222 nm in the temperature range between 4-95 °C. The presence of the methyl group seems to cause a slight increase of the melting point (T_m = 341.18 K) when compared to the related parent protein (T_m = 338.77K).

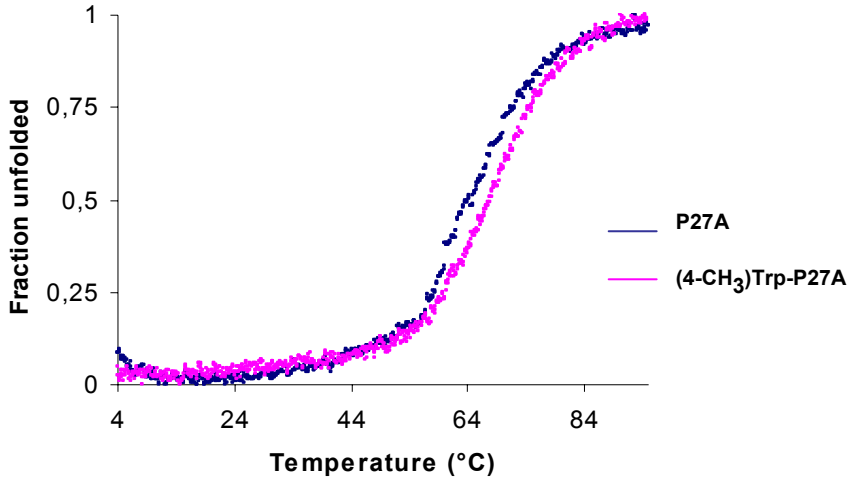


Figure 3.41. Melting curves of (4-CH_3)Trp-Barstar and wt protein. The methylated variant is slightly stabler than the parent protein, most probably due to the increased hydrophobicity of (4-CH_3)Trp.

3.2.19 Incorporation of aminotryptophans into Barstar

The analogues (4-NH_2)Trp, (5-NH_2)Trp , (6-NH_2)Trp , and (7-NH_2)Trp were used for incorporation studies in native Barstar and in the W38F, W44F, W3844F variants. Due to the strong tendency of amino-analogues to oxidize, some precautions had to be taken, such as working as much as possible under argon atmosphere, adding DTT to all media and solutions and avoiding light exposure.

For all incorporation experiments bacterial cells *E. coli ATCC49980* were allowed to grow in NMM with 0.015 mM Trp as limiting concentration. After the induction and protein expression, the OD_{600} showed that the cells had doubled when "fed" with 4-aminotryptophan, while optical density remained unchanged or the half more, when cells where forced to incorporate the other Tryptophan analogues.

The substrates (6-NH_2)Trp and (7-NH_2)Trp could not be incorporated either into wt-protein or into its variants, while (4-NH_2)Trp and (5-NH_2)Trp could be successfully incorporated only into the wt-Barstar. Indeed, a good level of protein expression was achieved also for some of the other variants, but after their purification yielded heterogeneous samples as revealed by the mass analyses. Moreover, it was not possible to separate properly the peaks during the FPLC runs, even with an isocratic gradient, so that no pure fraction could have been collected (see Figure3.42).

3.2.20 Incorporation of amino Tryptophans into native Barstar

The incorporation of (4-NH_2)Trp and (5-NH_2)Trp into Barstar has been successfully achieved following the routine incorporation protocol. The mass spectroscopic analyses confirmed that all three Trp residues were fully substituted with the amino analogues (calculated mass: 10295 Da; found mass: 10296 Da). Only one single peak was detected, thus showing that the protein was present as single species (Fig. 3.43)

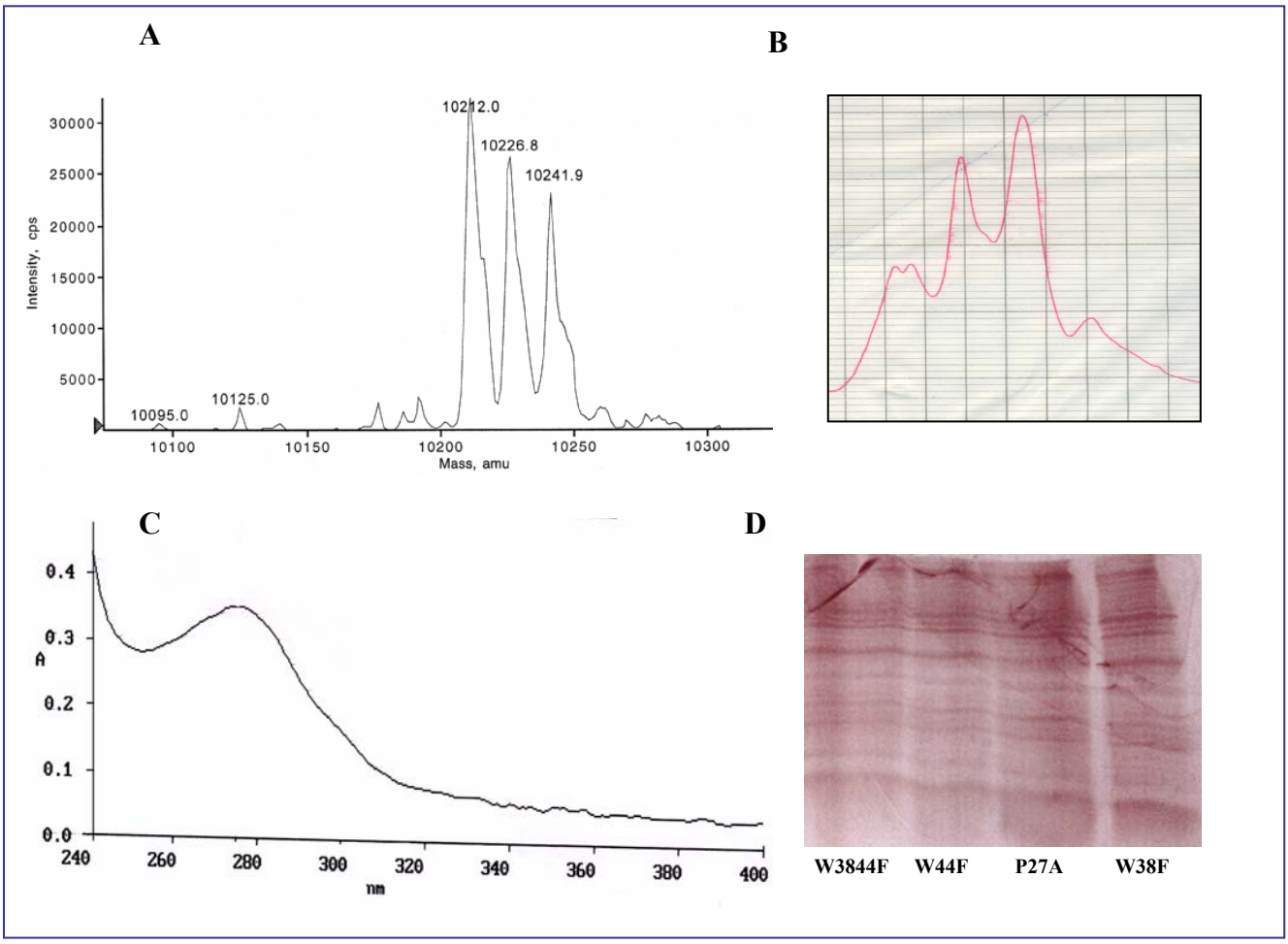


Figure 3.42. A) Deconvoluted mass profile of (4-NH_2)Trp-W44F. No homogeneous protein sample could be isolated for any amino-Barstar mutants. **B**) FPLC elution profile of (4-NH_2)Trp-W38F. An isocratic gradient (300 mM NaCl/Tris) was applied in the attempt to isolate pure protein fractions. As it can be seen in the figure, the different peaks could not be separated. **C**) UV absorbance profile of (4-NH_2)Trp-W44F. The amino-substituted protein had a strong tendency to aggregate and precipitate even at very low concentration (scattering between 310 and 350 nm). The red-shifted spectral shoulder at 305 nm indicates that at least a partial incorporation has taken place. **D**) Expression profile of (5-NH_2)Trp in both wt-Barstar and its mutants. The level of expression yield was generally very low when compared to native protein.

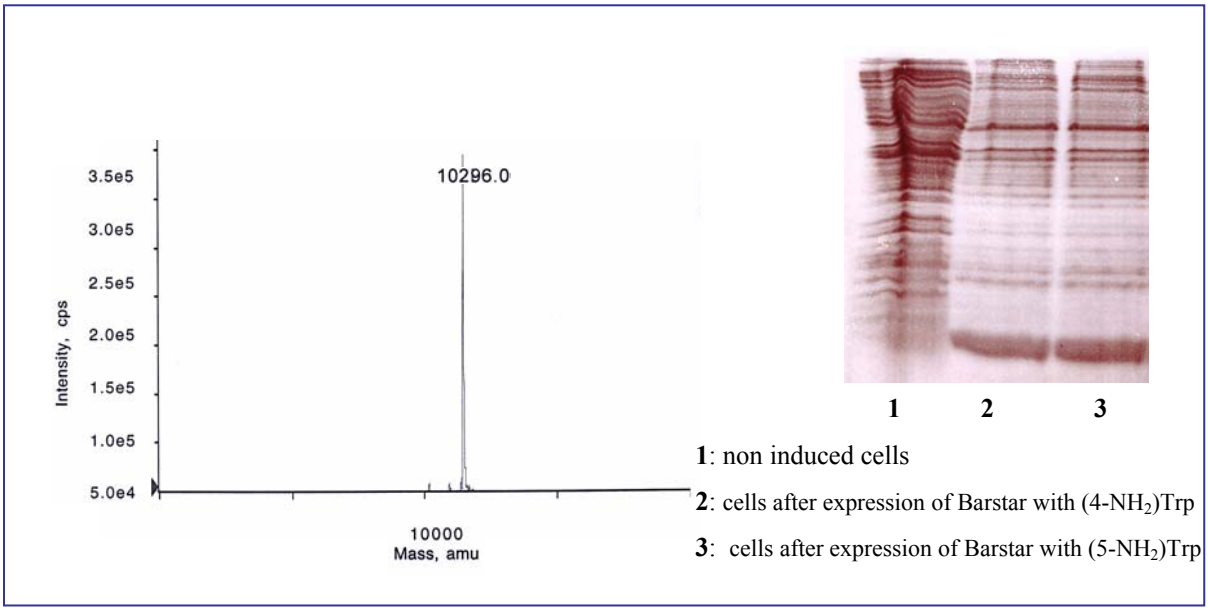


Figure 3.43. Deconvoluted mass profile of Barstar P27A substituted with (4-NH_2)Trp. Note that single peak indicates a highly homogeneous sample. Insert: expression profile of amino-Barstar in *E.coli ATCC 49980.*

The yield of amino substituted proteins was quite satisfactory and comparable to wild type for the incorporation of (4-NH_2)Trp, while the incorporation of (4-NH_2)Trp was more problematic and the expression level was not so high as in the case of wt protein.

3.2.21 Spectrocopic characterization of aminotryptophan-Barstars

The characteristic UV profiles of free amino acids (4-NH_2)Trp and (5-NH_2)Trp in solution are directly transferred in the related amino-Barstar variants upon incorporation experiments. This means that the spectral properties of these unique building blocks become intrinsic properties of the engineered proteins.

Moreover, in contrast with native Barstar, (4-NH_2)Trp- and (5-NH_2)Trp-Barstar variants displayed significantly different pH sensitivity.

Wt-, (4-NH_2)Trp-, and (5-NH_2)Trp-Barstar at pH 3.0 show an almost identical profile, as in the protonated state, the monocationic form NH^{3+} of the amino group is predominant. Conversely at neutral and basic pH values, both (4-NH_2)Trp- and (5-NH_2)Trp-Barstar show a 6 nm blue-shift (λ_{max} = 275 nm) in their absorbance maximum peak, while a second

prominent spectral shoulder appears at 305 nm for (5-NH_2)Trp-Barstar and at 300nm for (4-NH_2)Trp-Barstar.

Observing the relative fluorescence intensity behaviour pattern of native Barstar it can be noticed that the protein responds in a gradual and monotonic way to the pH changes, while 4- and (5-NH_2)Trp -Barstar show a rush and dramatic intensity decrease at increasing pH values.

While analysing the titration curves, it should be observed that at pH 3.0 the position of the emission maximum (λ_{max}) is red–shifted for both (4-NH_2)Trp- and (5-NH_2)Trp-Barstar, while at pH 7.0 and 9.0 the λ_{max} for (4-NH_2)Trp-Barstar is red-shifted and the λ_{max} for (4-NH_2)Trp-Barstar is blue-shifted (Fig. 3.44 and 3.45).

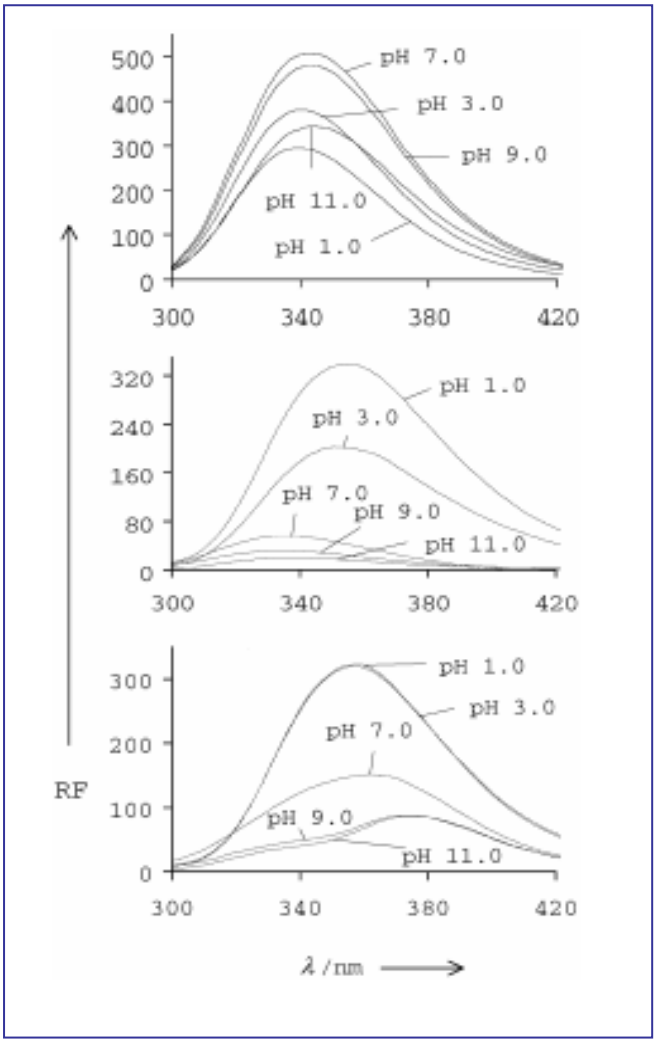


Figure 3.44. Fluorescence emission profiles of native Barstar and (4-NH_2)Trp- and (5-NH_2)Trp-Barstar upon pH titration.

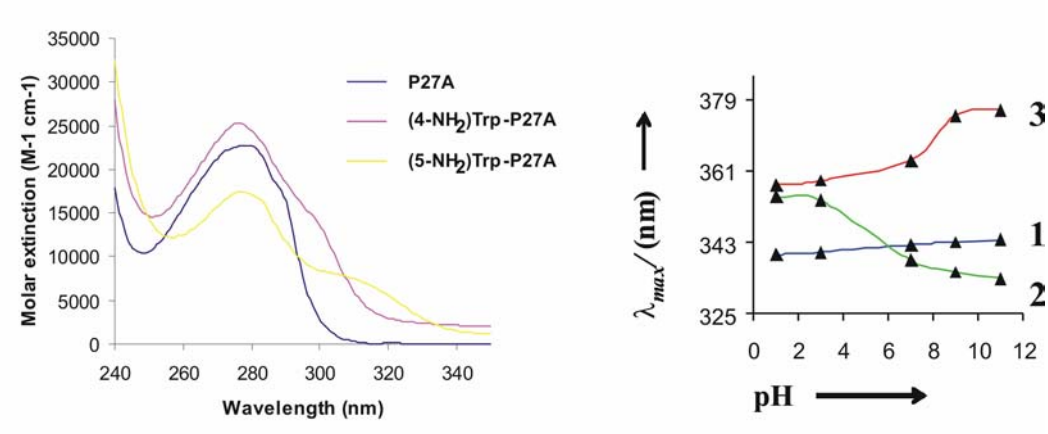


Figure 3.45 (Left) UV absorbance spectra of wt- Barstar and (4-NH_2)Trp- and (5-NH_2)Trp-Barstar. The spectra of the substituted protein resemble those of free amino acids in solution. (Right) Plot of the variations versus fluorescence emission maxima (λ_{max})

The blue-shifted cooperative titration curve for (4-NH_2)Trp-Barstar displays a transition midpoint between pH 5 and 6, while for (5-NH_2)Trp-Barstar the curve is red-shifted and the transition midpoint is positioned at pH 8.0. These different behaviours patterns are due to the changes in the charge-density distributions along the indole aromatic ring, as an important role is played by the position and the orientation of the substituent on the indole moiety. This means that (4-NH_2)-indole acts as a better proton donor than water in the excited state S_1, while for (4-NH_2)-indole is valid the opposite.

The pH sensitivity of Barstar variants obtained by incorporation of aminotryptophans derives therefore from the intramolecular charge migration that arises from cation-to-anion transitions of the aminoindoles around pH 6.0.

3.2.22 Secondary structure analyses and unfolding profiles of amino-Barstar

Melting curves of (4NH_2)Trp- and (5NH_2)Trp-Barstar were recorded following the conformational changes at 222 nm in the range between 4 - 95 °C. The introduction of the amino group into Barstar, caused a dramatic drop of the melting point for the modified proteins (see Table 3.9), thus pointing out that such a modification has brought the substituted proteins to the limit of their thermodynamic stability. This was also confirmed by the secondary structure profiles. The similar shapes and ellipticity values of wt-Barstar spectra between 4-50 °C show that the folded states are equally populated in this range of temperature.

A sharp transition occurs between 50-70 °C, denoting the transition to the denatured state. Conversely, the folded-to-unfolded transition displayed by amino-substituted Barstars occurs in an almost continuous manner. Moreover, the ratio between the two characteristic minima at 222 and 208 nm sinks from 0.92 (wt-protein) to 0.74. Amino-substituted Barstar showed also cold denaturation tendency in a temperature range between 4- 12 °C, reaching its maximum stability at about 20 °C.

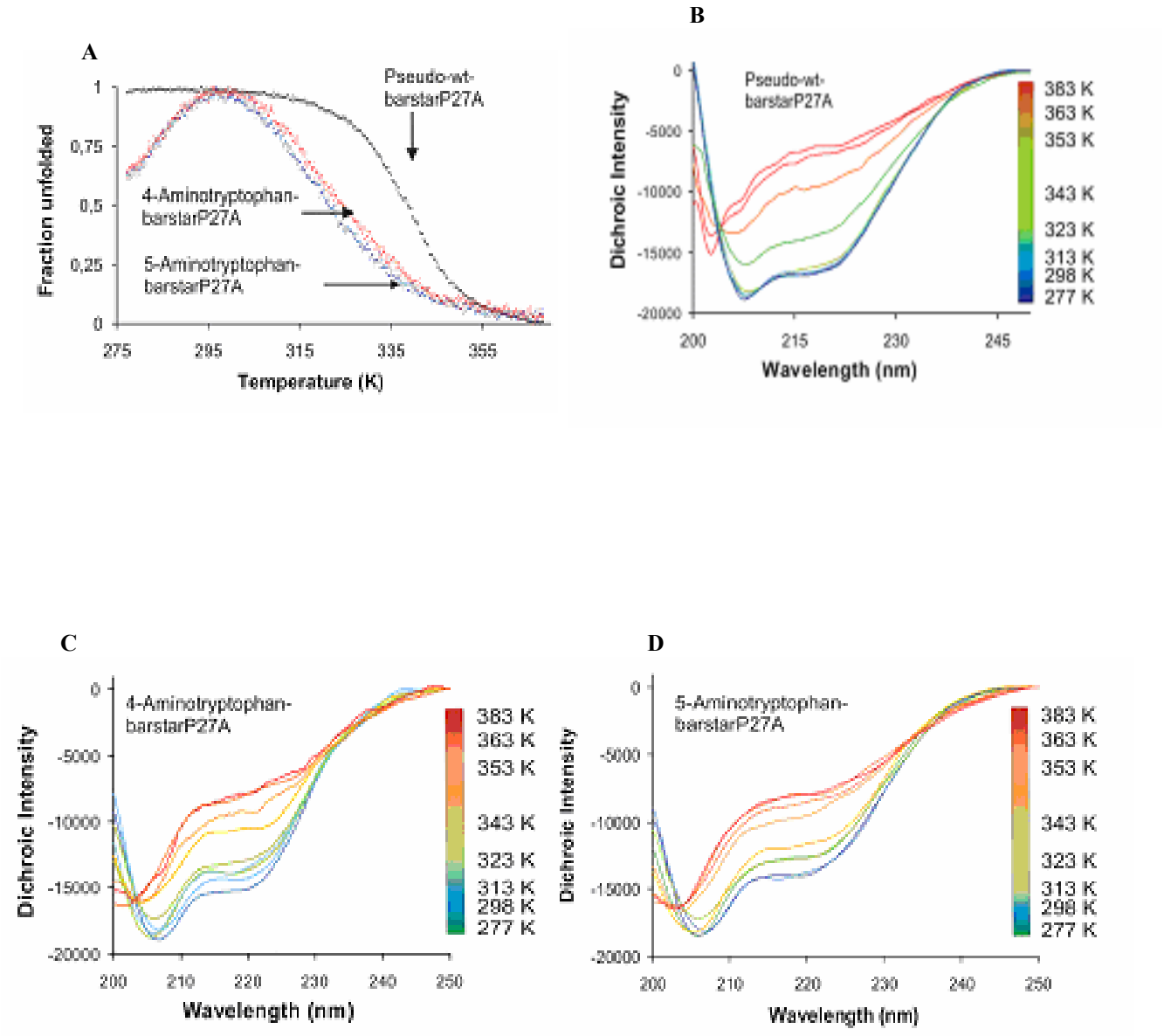


Figure 3.46. Melting curves of wt-Barstar and amino substituted Barstar variants. The incorporation of amino-Tryptophan is accompanied by a dramatic drop of the melting point (**A**). Secondary structure profiles of wt Barstar (**B**), (4-NH_2)Trp-Barstar (**C**) and (5-NH_2)Trp-Barstar (**D**). Derived thermodynamic parameters are shown in Table 3.9

Protein	Tm (K)	ΔT_m(K)	ΔH_m (kJ/mol)	$\Delta\Delta H_m$(kJ/mol)
P27A	338.77 ± 0.035	-	-183.1 ± 1.9	-
(4-NH_2)Trp-P27A	319.77 ± 0.08	-19	-101.7 ± 1.7	-81.4
(5-NH_2)Trp-P27A	315.11 ± 0.2	-23.66	-66.2 ± 1.7	-116.9

Table 3.9 . Thermodynamic parameters for unfolded wt-Barstar and variants.

Melting curves of wt-, (4-NH_2)Trp- and (5-NH_2)Trp-Barstar in the presence of 0.5 M NaCl were also recorded. Indeed, it is already known in the literature (Pradeep *et al*., 2002) that the presence of salt induces a stabilization effect due to the ionic strength. This was found to be true for both wt- and (4-NH_2)Trp-Barstar, as it resulted in an increase of the Tm value of about 10 °C and suppression of the cold denaturation effect as shown in Figure 3.47. On the contrary, for (5-NH_2)Trp-Barstar no stabilization effect could be appreciated, as this variant in presence of 0.5 M NaCl is strongly aggregated and its melting curve could not be recorded.

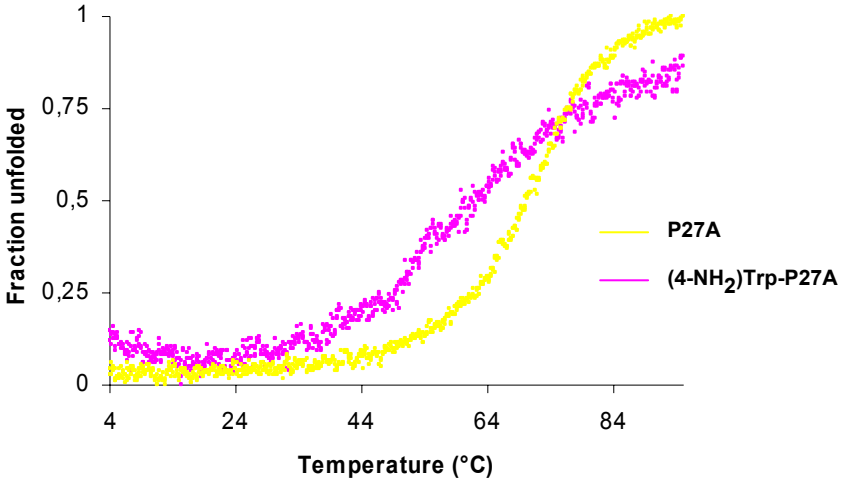


Figure 3.47. Normalized unfolding profiles of wt- and (4-NH_2)Trp-Barstar determined by measuring changes in the dichroic intensity at 222 nm as function of temperature in the presence of 0.5 M NaCl.

3.2.23 X-Ray analyses of (4-NH_2)Trp-Barstar

The three-dimensional structure of (4-NH_2)Trp-Barstar was solved in order to find a structural basis for the dramatic thermodynamic and spectroscopic changes generated upon the introduction of the highly polar amino groups into Barstar structure. Surprisingly, even if the space group and the cell constants are different (see Table 3.10), the new structure turned out to be almost identical to the wild-type.

(4-NH_2)Trp-barstar(b*C40A/C82A/P27A)		**SeTrp-barstar (b*C40A/C82A/P27A)**	
Space group	C2	Space group	R3
Cell constants	a = 85.16; b= 46.85; c = 51.06	Cell constants	a = b = 116.50; c = 36.51
Limiting resolution (Å)	2.0	Limiting resolution (Å)	2.4
Reflexions unique	10725	Reflexions unique	7134
Completeness	96.2 %	Completeness	98.2 %
R_{merge}*	11.7 %	R_{merge}*	7.9 %
Molecules per assym unit	2	Molecules per assym unit	2
Refinement:		Refinement:	
Resolution range (Å)	20 - 2.0	Resolution range (Å)	29.7 - 2.4
Crystallographic R factor†	25.21 %	Crystallographic R factor†	22.8 %
Free R factor††	30.34 %	Free R factor††	29.5 %
Rms deviations: bond lengths	0.015 Å	Rms deviations: bond lengths	0.008 Å
bond angles	1.59 °	bond angles	1.25 °
No. of non-hydrogen atoms	1438	No. of non-hydrogen atoms	1438
Solvent molecules	92	Solvent molecules	73

*R_{merge} = $\sigma[I(h)_i-\langle I(h)\rangle]/\sigma\langle I(h)\rangle$; $I(h)_i$ is the observed intensity of the ith measurement of reflection h, and I(h) the mean intensity of reflection h calculated after loading and scaling.
† R_{factor} = $\sigma[\,|F_{obs}| - |F_{calc}|\,]\,/\,|F_{obs}| \times 100$
†† R_{free} was calculated randomly omitting 5% of the observed reflections from refinement

Table 3.10. Data collection and refinement statistic for (4-NH_2)Trp-Barstar. In order to compare crystal parameters of this variant with the wt-protein, a data set from Se-Trp-Barstar was taken, as the data set for this Barstar variant are identical to those of wt-protein reported in literature (Bae et al., 2001).

The two molecules in the asymmetric unit, turned out to be a crystallographic dimer, as after shifting the molecules to the maximum contact area, the contact surface exceeds 1000 Å.
In nature, Barstar is not prone to dimerization; however it is known (Guillet et al, 1993) that under denaturant conditions, it displays a slight tendency to form dimers that tend to revert to monomers.

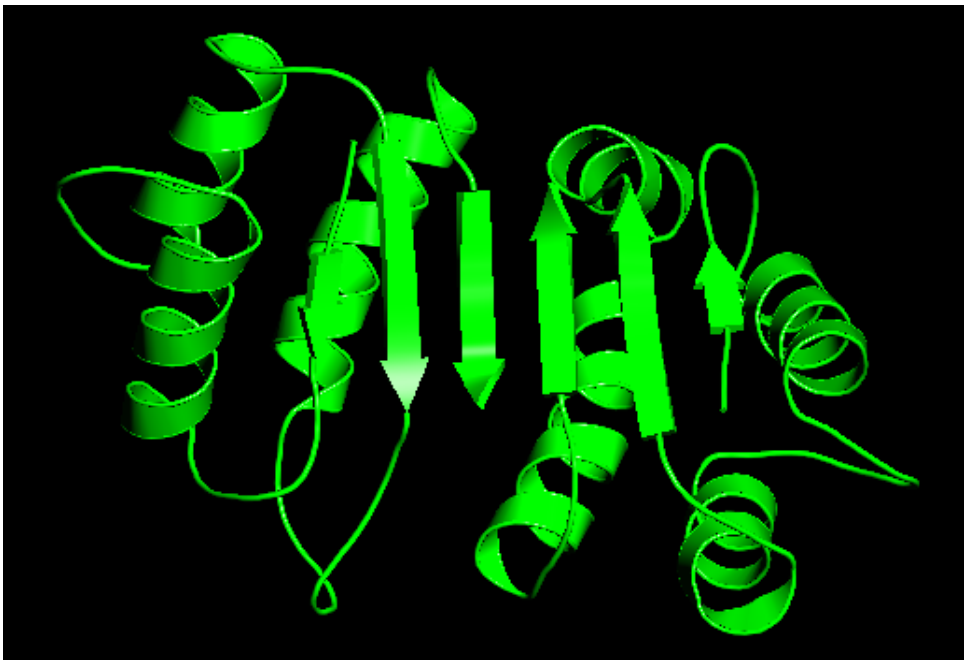


Figure 3.48. Three dimensional structure of (4-NH_2)Trp-Barstar. The two molecules in the asymmetric unit behaved as a crystallographic dimer.

The electron density was very well defined for molecule A, while for molecule B no clear electron density was to be seen from residue 58 to residue 65, most likely due to the effects induced by the crystal packing. An omit-map was calculated in order to verify the presence of the helix 3, but even if some additional electron density appeared, it was not possible to improve the current model.

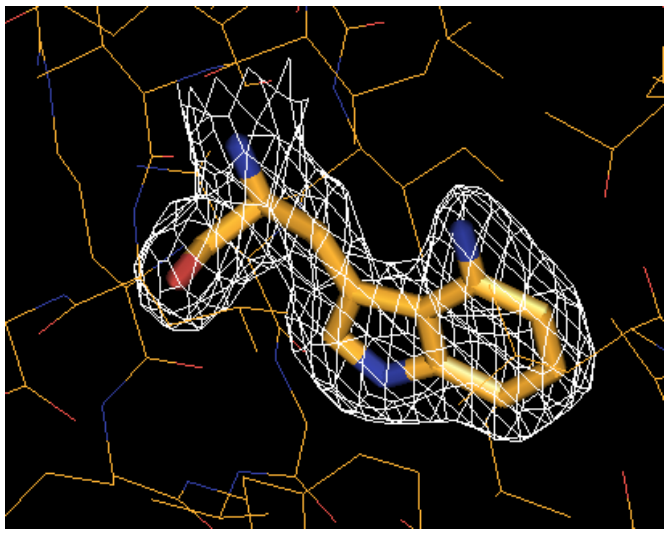


Figure 3.49. Continuous Electron density (contouring level 1σ) for (4-NH_2)Trp53 of related Barstar variant

The overall fold of (4-NH_2) Trp-Barstar is identical to that of the parent protein. The overlay of the amino-substituted protein with the parent one shows no deviations in the backbone; only the side chains of some residues on the protein surface are set apart.

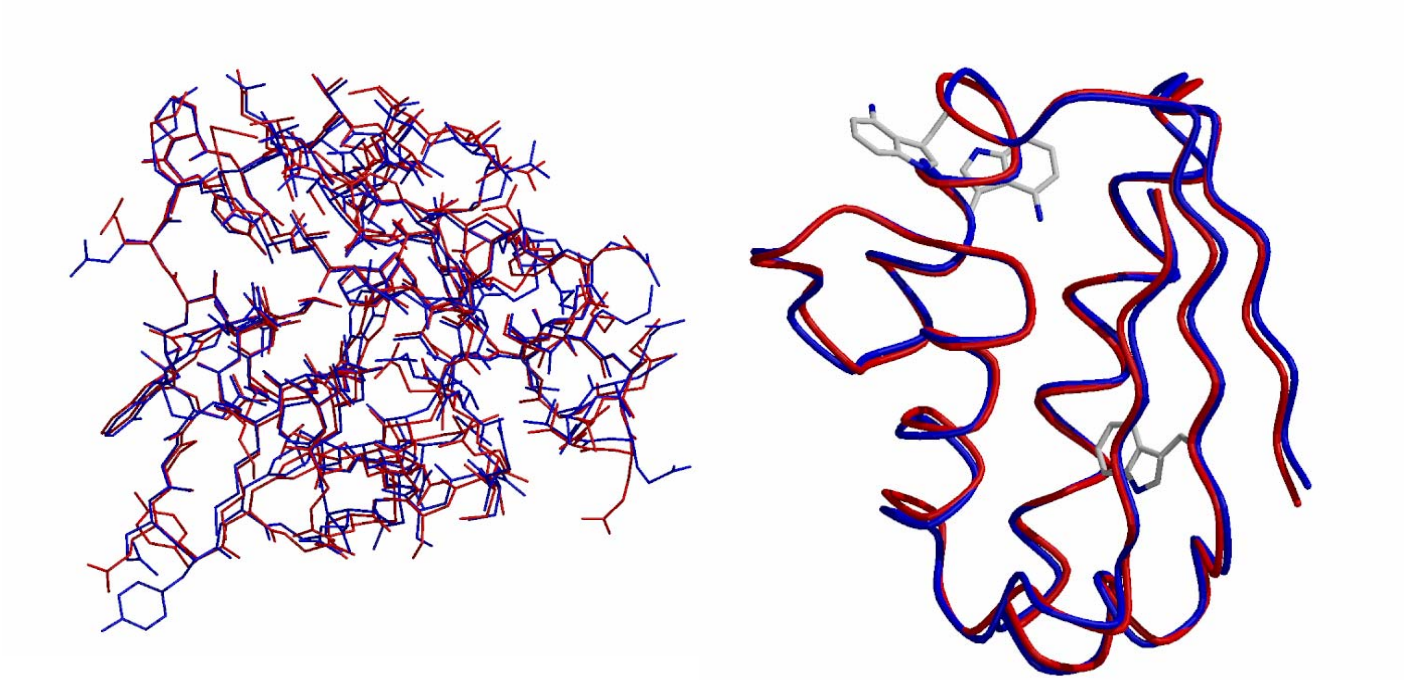


Figure 3.50. The crystal structure of (4-NH_2) Trp-Barstar (blue) superimposed to the structure of wt protein (red). (Left) Complete structure with all side chains. (Right) Backbone representation of both proteins. (4-NH_2)Trp is represented with sticks.

(4-NH_2)Trp53 is buried into protein interior in a hydrophobic cavity, while (4-NH_2)Trp38 and (4-NH_2)Trp44 are partially or completely exposed to the solvent and therefore are expected to be more flexible. Conversely, in our crystal structure, they are well defined and the difference electron density map revealed precisely the site of H→NH_2 replacement. In this way, crystallographic analysis fit perfectly with the mass-spectrometric analytical data, confirming the absence of chemical reactivity of (4-NH_2) Trp residues during oxidative refolding process (Rubini *et al*., 2004).

The sterically allowed values of ϕ and ψ angles for the peptide chain were calculated using the van der Waals distances. 133 residues (86.4%) are positioned in most favourable regions, 19 residues (12.3%) in additional allowed regions and 2 residues (1.2%) in generously allowed regions. As it can be seen in Figure 3.51, (4-NH_2) Trp44 in molecule B is placed in a generously allowed region at the border with an additional allowed region. This is not

surprising, as this residue is to be found on the surface of the protein completely exposed to the solvent.

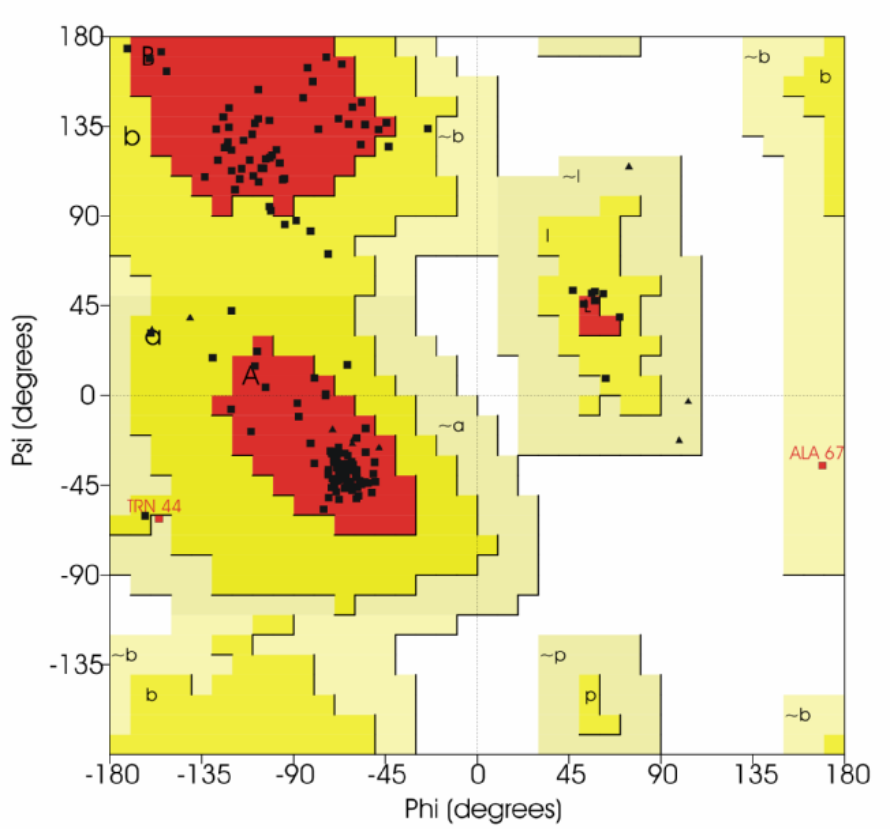


Figure 3.51. Ramachandran Plot for the model of (4-NH_2)Trp-Barstar calculated by PROCHECK. Triangles represent Gly residues, rectangles all other amino acids. In most favoured regions (A, B, L) there are 133 residues (86.4%), in additional allowed regions (a, b, l, p) 19 residues (12.3%) and in generously allowed regions (~a, ~b, ~l, ~p) 2 residues (1.3%). The dimer of (4-NH_2)Trp-Barstar contains 10 Gly and 2 Pro.

3.2.24 Modelling of (5-NH_2)Trp-Barstar

Attempts to crystallize (5-NH_2)Trp-Barstar were not successful. Therefore, its structure was modelled on the basis of the structure parameters of the closely related (4-NH_2)Trp-Barstar. Since it is well known that different substitutions on indole ring might induce substantial changes in protein structure, modelling procedures might only indicate whether there are new interactions in the local environment of (5-NH_2)Trp. The modelling performed in this study (see Materials and Methods) indicate that there are in principle no changes in the substituted Tryptophan and its surrounding between the two structures (see Figure 3.52)

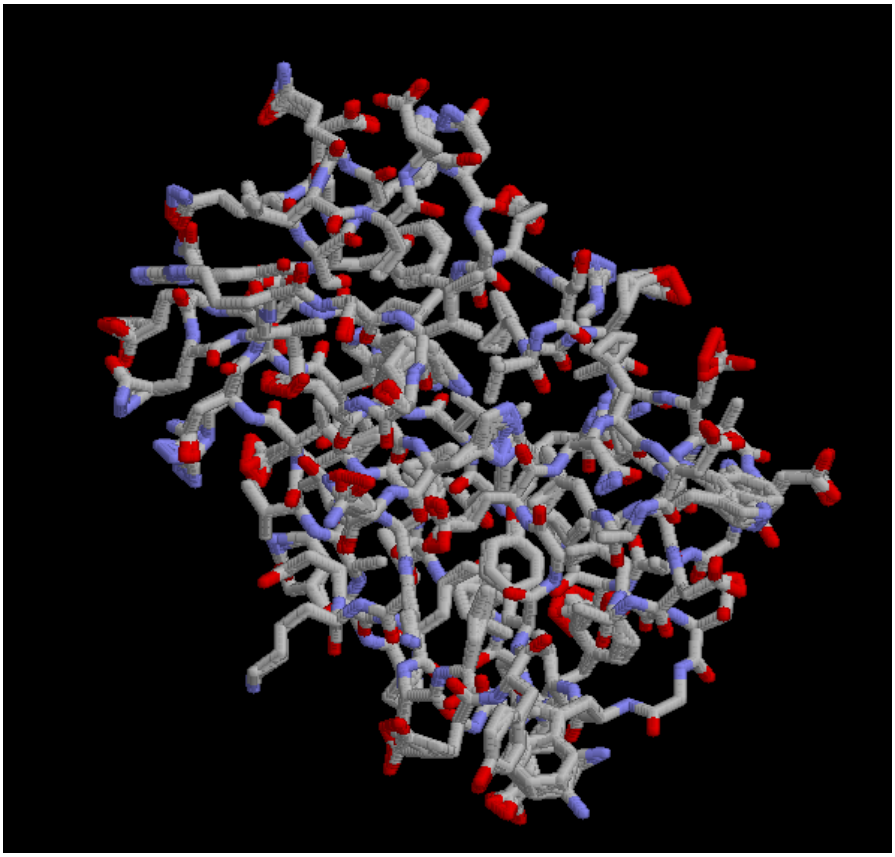


Figure 3.52. The structure of (5-NH_2)Trp-Barstar is modelled on the basis of the known structure of (4-NH_2)Trp-Barstar. The resulting model is almost identical to the parent molecule.

3.2.24 Incorporation attempts of chloride and bromide into Barstar

All attempts to incorporate chlorotryptophans and 7-bromotryptophan into Barstar were not successful. Actually, in the case of 6-chlorotryptophan a certain expression could be identified in the SDS-PAGE gel, but after the loading on the column, after the preliminary purification, no protein could be isolated. This means that even if the uptake of the analogue has taken place and the chlorotryptophan was incorporated, the folding of the protein variant did not occur (Budisa et al., 1998).

3.2.24 Lipid-mediated delivery of wt-, and (3-F)Tyr-EGFP into MCF10 cell line

The most elegant approach for intracellular drug delivery of proteins is to use lipid mediated delivery system. In this study, Lipofectamin, a commercially available transfection agent was used to deliver both (3-F)Tyr-EGFP and EGFP into MCF10 cell line. The delivery efficiency

was about 10% and intracellular localisation of both proteins was unambiguously demonstrated by fluorescence microscopy (see Fig. 2.5., Materials and methods part). The delivery experiments were done with non-lethal amounts of lipids. Our initial results have shown that cells treated with (3-F)Tyr-EGFP are more prone to undergo apoptosis and even necrosis than the cells treated with wt-protein. Unfortunately these features were not always reproducible and in some cases degradation of labelled protein does not take place at all. Future experiments with much more carefully chosen delivery system, cells and overall experimental set-up should answer the question about the utility of this approach as specific intracellulary delivery route.

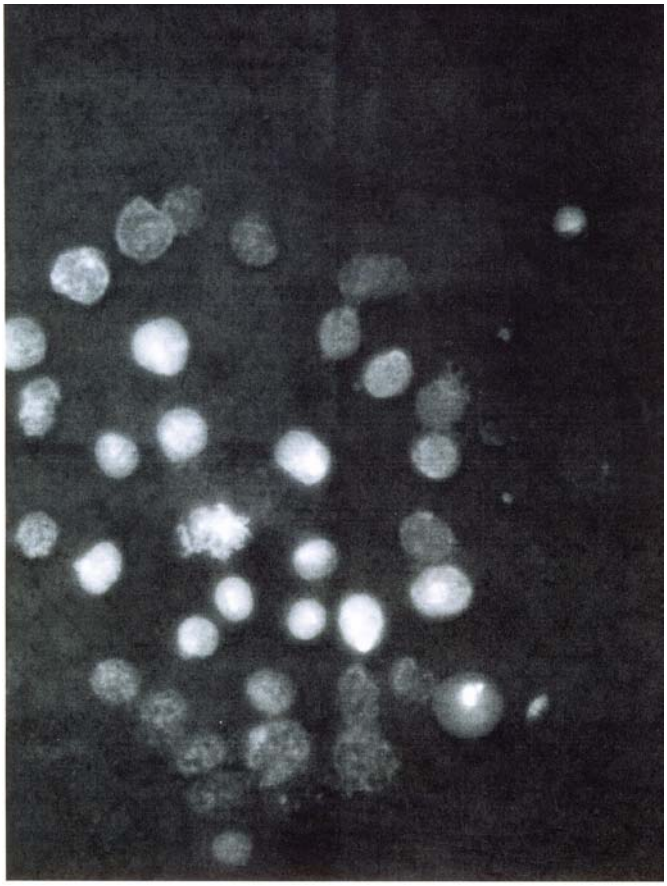


Figure 3.53. The apoptotic nuclei of the MCF10A breast epithelial cell line, 48h after liposome-mediated delivery of (3-F)Tyr-EGFP.

4. DISCUSSION

- 4.1 Limits for analogues incorporation into proteins

Although a lot of efforts have been done in order to expand the number of amino acids suitable for protein synthesis, the choice of suitable analogues and surrogates is still limited. In the auxotroph-based approaches such as the SPI-method, the non canonical amino acids must be similar in shape and size to the natural occurring molecule. Moreover, if the substituent group can ionize, it must produce the same type of ion or be un-ionized at physiological pH values According to Richmond (Richmond, 1962), four types of substitutions are expected to be well tolerated in proteins:

- hydrogen atoms in a ring system by fluorine
- phenyl ring by some other resonant ring structure
- the replacement of one type of heterocyclic ring by another
- backbone replacement of an amino acid by another with similar size and shape

Moreover, in addition to these general rules, the introduction of non canonical amino acids is restrained by three main barriers (i) metabolic toxicity, (ii) cellular proofreading and (iii) folding of the substituted proteins.

4.1.1 Metabolic toxicity

The twenty amino acids that form the natural pool do not fulfill only physical and chemical criteria that make them suitable for the protein synthesis, but also metabolic and energetic functions in order to provide an optimal balance of metabolites in the cell. This means that only a small amount of toxic substances can be tolerated under normal growth conditions. Such an example is provided by the capability of *E. coli* to biosynthesize selenocysteine and to insert it into formate dehydrogenase (Zinoni et al., 1986, 1987; Leinfelder et al., 1988). Also selenomethionine can be synthesized by *E. coli* grown on selenate and incorporated into protein (Huber et al., 1967, Huber and Criddle, 1967).

Norleucine (methionine antimetabolite) is also most likely to be present in small amounts in *E. coli* under physiological conditions, and only an accurate control can prevent the complete substitution of methionine with norleucine. Otherwise, this would result in cellular growth inhibition, host restriction, degradation of chromosomal DNA and cell death (Bogosian et al.,

1989). Such metabolic choices have to be always kept in mind by planning protein engineering with an expanded amino acid repertoire.

4.1.2 Cellular proofreading

The SPI method relies on traditional use of auxotrophic strains and exploits the absence of absolute substrate specificity of aminoacyl-tRNA synthetases. Therefore, only a limited range of analogues can be effectively recognized, activated and transferred to the cognate tRNAs to take part in the protein syntheses on the ribosome. This suggests that relaxation of substrate specificity of aminoacyl-tRNA synthetases could help to circumvent strict discrimination during the aminoacylation step. Actually, such an example is offered by nature itself, as the indirect pathways for the synthesis of Asn-tRNA^{Asn} and Gln-tRNA^{Gln} proceed via mis-aminoacylated intermediates: Asp-tRNA^{Asn} and Glu-tRNA^{Gln} respectively. The basis of this relaxed specificity in certain aaRSs has been found to derive from the absence of a loop that specifically recognizes the third position of the anticodon (Schmitt et al., 1998). For example, in *Thermus thermophilus* the lack of recognition of the last anticodon nucleotide is responsible for the non discrimination of AspRS between tRNA^{Asn} (GUC anticodon) and tRNA^{Asn} (GUU anticodon) (Becker et al., 1998). Some efforts have been done in order to enlarge the number of substrates that can be recognized and activated by aaRSs. This led for example to the incorporation of *para*-chloro-Phenylalanine and *para*-bromo-phenylalanine into a recombinant luciferase (Ibba et al., 1995). This was possible since the substrate specificity of PheRS was relaxed by Ala494Gly mutation in the amino acid binding pocket of this enzyme. However, it should be also stressed that amino acid activation and tRNA charging are not the only "bottlenecks" in protein translation. As it was demonstrated in our laboratory, certain amino acid analogues like 4- and 5-aminotryptophan, and 4- and 5-hydroxytryptophan can be successfully incorporated into Barstar, but only 4-aminotryptophan could be incorporated into *Aequorea* GFP proteins (Budisa et al., 2002, Bae et al., 2003). Similarly, both thiaproline and selenoproline can be selected and loaded onto cognate tRNAs, but only thiaproline can be incorporated into target proteins (Budisa et al., 1998). This strict selectivity clearly indicates that exist editing mechanisms that go beyond the aminoacylation steps.

4.1.3 Protein folding

The last barrier for the incorporation of non canonical amino acids into proteins is their capacity to fold in the correct manner, so that the protein functionality is retained. It should be kept in

mind that even if the translation machinery of the cell can be forced to uptake some non natural amino acids, the correct protein folding cannot be driven in the desired direction. As it was shown in this work, *E. coli* is able to incorporate 6-chlorotryptophan into Barstar, but the too bulky analogue prevents the protein to fold correctly.

In the engineering of protein interiors, the hydrophobic effect and the density packing have to be taken into account (Sandberg et al., 1991). The hydrophobic effect accounts on the fact that clustering of apolar atoms in hydrophobic cores enhances the stability of the native proteins (Kauzmann, 1959, Tanford, 1979). Models of the hydrophobic effect based on the properties of apolar liquids (Radzicka et al., 1988) indicate that the influence of core amino acids on protein stability might be predictable using a liquid-like model of the protein interior.

The density packing model suggests that protein interiors are very densely packed and that proteins are built around rigidly packed, non deformable cores of apolar side chains (Richards, 1977) Following this reasoning, it might be possible to predict which amino acid side chains would fit at a particular site in the interior of a protein, on the basis of the shape of the site and the shapes of the amino acids (Ponder et al., 1987). Indeed, the assumption that buried apolar residues contribute to protein stability might create the wrong impression that stability is a simple function of the hydrophobicity of these apolar side chains (Yutani et al., 1987, Matsumura et al., 1988, Kellis et al., 1988). In fact, as the stabilities of more proteins with apolar substitutions were determined, it turned out that the dependence of stability on hydrophobicity varies from case to case (Sandberg et al., 1991). More recently, it has also been confirmed that hydrophobic core packing does not determine the global fold of a protein, but plays a role to a certain extent in determining the stability of a protein and its conformational uniqueness (Lazar et al., 1998).

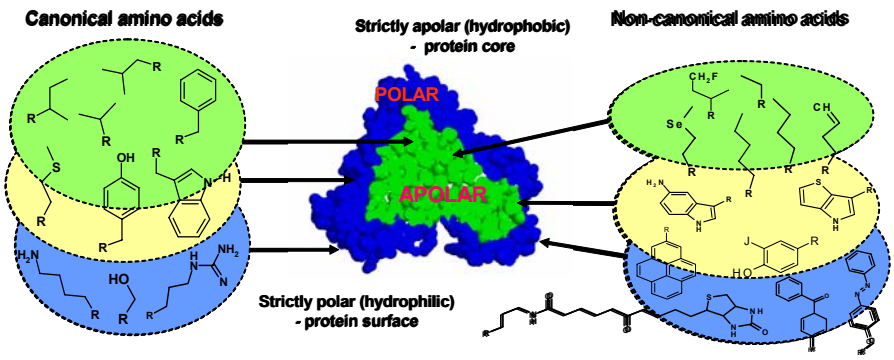


Figure 4.1. The basic rules for protein building and folding with canonical and noncanonical amino acids. The binary partitioning of polar and non-polar residues in the protein basic architectural pattern (polar-in, apolar-out) is fully reflected in the structure of the genetic code (from Budisa, 2004).

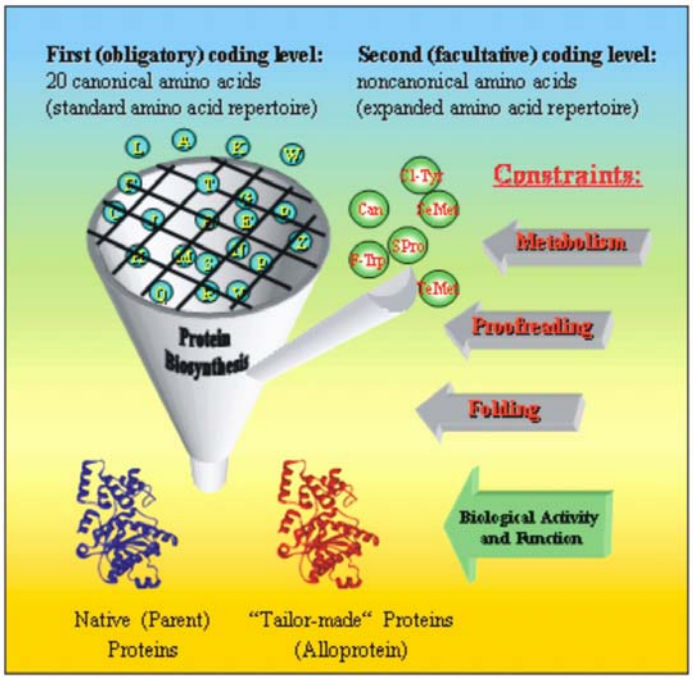


Figure 4.2. Schematic representation of the limits of the SPI method. Non canonical amino acids enter the genetic code exploiting the same roots of the canonical ones. Three main barriers have to be overcome to allow the amino acid analogues to act as substrates for the protein synthesis: metabolic toxicity, proofreading and editing mechanisms, and capability of the substituted protein to fold correctly. Proteins produced in this way are named alloproteins or "tailor-made" proteins, as they display specific roles in user defined environments. (taken from Budisa, 2004)

- 4.2 Thermodynamic penalty for the incorporation of aminotryptophans into Barstar

The substitution of Trp with (4-NH_2) Trp and (5-NH_2) Trp in Barstar brings the replacement **–H → -NH_2** at all Trp residues in the structure. Although the crystal structure of (4-NH_2) Trp Barstar is fully isomorphic to wild type, dramatic changes were observed in solution, concerning the stability and thermodynamic of the protein variants.

It is known that native proteins are stabilized by several factors, such as van der Waals interactions, hydrogen bonds, hydrophobic interactions, effect of hydrations and salt bridges (Rose et al., 1993). This delicate equilibrium can seriously be affected and by amino acid mutations. In the late 50´s, Kauzmann suggested that proteins fold by clustering non-polar groups in aqueous solution (Kauzmann, 1959). This concept is now widespread among researches in protein folding field and enriched with more detailed models, like the "framework model" or the "hydrophobic zipper", in which clusters of non polar side chains stabilize secondary structures such as α-helices and β-hairpins (Zdanowsky et al., 1999). However, the

main idea is that burial of non polar side chains out of contact with water provide the major source of the free energy change that drives folding (Baldwin, 2002). There are evidences that the initial stages of Barstar folding begin with such a hydrophobic collapse due to long range interactions formed by Trp53 buried between Phe56 and Phe74 in the protein interior (see Fig. 4.3). The crucial role played by Trp53 is also confirmed experimentally by the inability to replace this residue by site-directed mutagenesis with any other canonical amino acids without a loss of protein structural integrity (Nath et al., 1997).

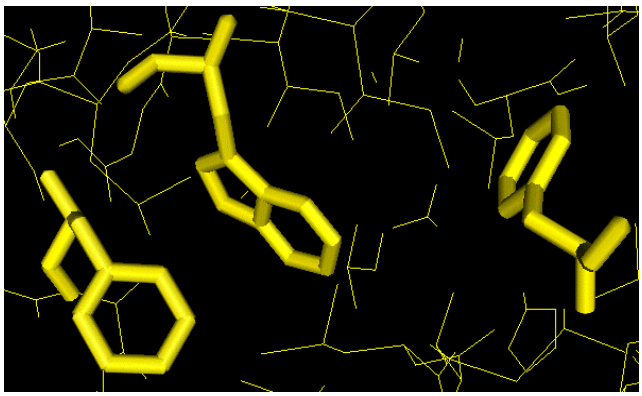


Figure 4.3. Barstar hydrophobic core with Trp53 "sandwiched" between Phe56 and Phe74. The cluster of these three hydrophobic residues is supposed to play a crucial role in the initial stages of barstar folding.

The introduction of (4-NH_2)Trp and (5-NH_2)Trp into the protein interior, do not involve sterical perturbations but results in an increased hydrophilicity in an apolar milieu. These considerations are supported by the careful inspection of the crystallographic interaction distances in the molecular environment of Trp53 versus (4-NH_2)Trp53. As it can be seen in Figure 4.4, the number of hydrophobic interactions is reduced in comparison to wild type. Therefore, these protein variants are already partially unfolded under physiological conditions and are prone to undergo phenomena like cold denaturation.

The creation of proteins endowed with novel properties using an expanded amino acid repertoire might face a problem of thermodynamic and energetic consequences of such a operation, as these new building blocks can seriously impair protein stability. The free energy of folding as the difference in the free energies of folded and unfolded states for small proteins is generally only 5-20 kcal/mol (Radzicka et al., 1988).

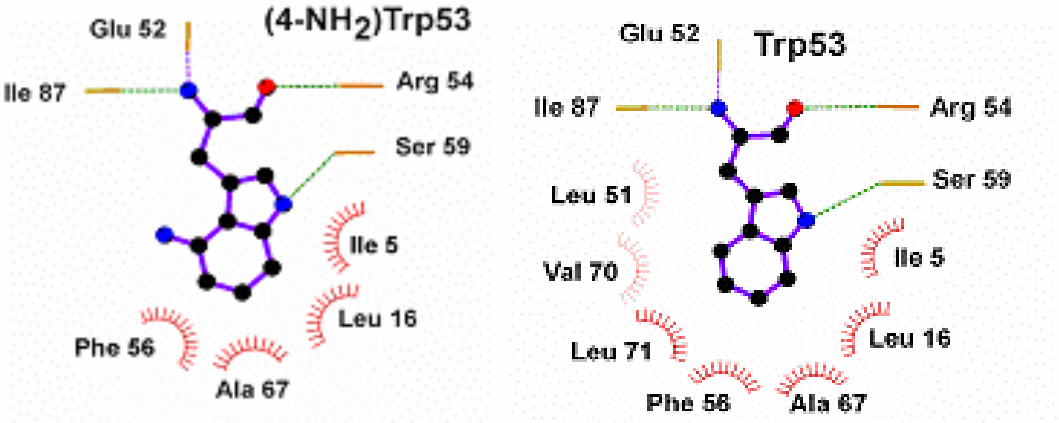


Figure 4.4. Schematic representation of Trp53 (left) and (4-NH_2)Trp53 (right) molecular environments with their interaction distances to the residues in the surrounding protein matrix drawn by LIGPLOT. The number of the hydrophobic contacts in the molecular environment of (4-NH_2)Trp53 is dramatically reduced in comparison with the native residue. This is due to the more hydrophilic nature of the aminotryptophan side chain.

This means that nature excluded a certain class of amino acids from the code repertoire through selection, since they might exhibit adverse effects on the folding and structural integrity of target proteins. In this context, experiments presented in this study clearly indicate that the natural scaffolds of hydrocarbon cores are shaped and optimized for billions of years of evolution for the canonical amino acid repertoire. Thus, the existing protein structural frameworks might not always be suitable for the design and engineering experiments by accommodation of novel non canonical amino acids, especially those of synthetic origin (Rubini et al., 2004).

Nonetheless, it has to be noted that the incorporation of 4-methyltryptophan into Barstar is followed by a stabilization of the protein. On the other hand, the dramatic decrease in the stability due to the -CH_3 → –NH_2 isosteric replacement clearly indicates that the presence of aminotryptophan is the main reason for the observed phenomena. The filling of preexistent cavities is well tolerated in many proteins, as minimizations of empty spaces within an interior without causing a steric overlap, seems to be a strong guiding principle for structure stabilization (Richards, 1997). Otherwise, the evolution of the genetic code seems to have found a compromise between mutability and variance. This means that one of the aims of nature is to preserve the proteins interior from the entrance of hydrophilic amino acids. The strict partition between polar and apolar amino acids was achieved during the evolution by assigning codons with a central U to amino acids with chemically relatively uniform, hydrophobic side chains (Budisa, 2004). Therefore, the simplest rule for protein building to be obeyed is the general "apolar in-polar out" principle.

- 4.3 Attempts to design a “Teflon” proteins

Fluorinated amino acids occupy a special place since they might serve as unique building blocks, for both classical peptide as well as ribosomal protein synthesis. The fluorine is much less bulkier than chlorine and is expected to cause less steric perturbance in the protein structure. Since fluorine is the most electronegative element, C-F bonds dramatically change electronic properties when compared with those of the C-H bond. In fact, its dipole moment is opposite to that of a C–H bond. Fluorine exerts a strong inductive effect that is sufficient to affect properties such as polarity or binding capacity of functional groups that are even distantly positioned in the structure. The 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. The global replacement of methyl by trifluoromethyl groups in proteins would provide them with a fluorous core. Such proteins should have dramatically changed different cooperative properties when compared with natural ones that possess a hydrocarbon core. Perfluorinated amino acid side chains are endowed with the capacity to induce a hydrophobic collapse through their exclusion from either aqueous or organic phase. In other words, fluorous proteins should be capable to fold in aqueous media and in addition to resist denaturation by organic solvents. This means that fluorinated proteins should fold into a tight particle with an interior hydrophobic core that is shielded from the surrounding solvent following the general “apolar in – polar out” folding principle.

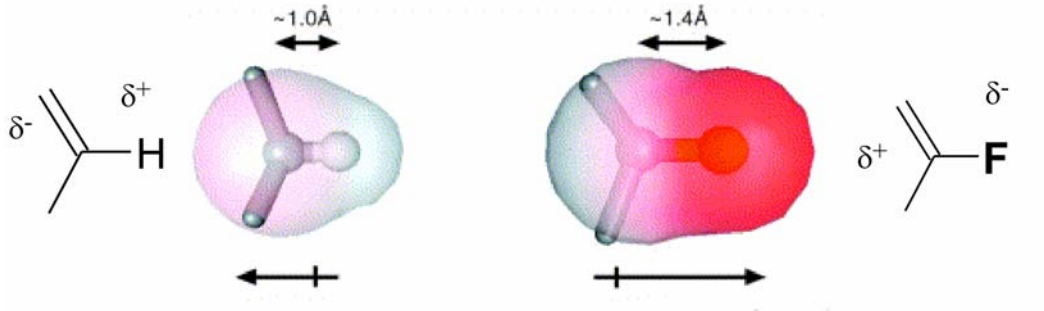


Figure 4.5. H-F versus C-F bond. Contrary to the widespread beliefs, H-F replacements are not fully isosteric: the C–F bond of 1.39 Å is significantly larger than that of the C–H bond (1.09 Å), although the van der Waals radius of fluorine (1.35 Å) is only 0.15 Å larger than that of hydrogen. Note also opposite polarities. (Modified from Neil and March, 2000).

It is conceivable that the replacement of the 5,5,5-hydrogen atoms of Leu with fluorine might be tolerated at single sites, while global substitution of all Leu side chains in Annexin V, or EGFP would significantly enlarge their fluorous hydrophobic core. Namely, the volume of the

trifluoromethyl group was estimated to be closer to that of the isopropyl group. Clearly, such an increased sterical bulk of trifluoromethyl group relative to the non fluorinated methyl group would lead to unfavorable interactions with the surrounding residues in a particular protein interior.

Taking into account all above considerations it is obvious that natural structures are designed for the accommodation of these amino acids as protein building blocks. Namely, the natural scaffolds of hydrocarbon cores have been shaped and optimized for billions of years of evolution and are, therefore, unsuitable to accommodate a large number of fluorine atoms. Therefore, the generation of "Teflon" or "non-sticking" proteins, capable to fold and function in both organic and aqueous solvents would require novel repacking of the protein core i.e. a *de novo* design.

- 4.4 Possible medicinal applications of amino acid analogues

Most of the non canonical amino acids are toxic; often this toxicity is the result of their conversion into toxic substances by a relatively complex metabolic route. Thus, they should have interesting pharmacological properties that certainly have a great potential for biomedicine. Many microorganisms, plants and some fungi strains produce unusual amino acid containing toxic substances that are physiologically highly active in mammals or other microorganisms, usually in a highly deleterious manner like cephalosporines or penicillin. Some amino acids produced by plants as secondary metabolites are structurally very similar to the canonical ones and have impressive potentials to affect the metabolism in other organisms. The early studies on noncanonical amino acids as antimetabolites extensively reviewed by Richmond, 1962 Hortin & Boime, 1983 Wilson & Hartfield, 1984 clearly demonstrated the strong *anti*-microbial, *anti*-mycotic, and *anti*-tumour activities of such substances. The roles of non canonical amino acids beyond participation in protein synthesis are (i) specific interactions with metabolic and catabolic enzymes (inhibition, activation, modulation) and (ii) interference in the process of amino acid biosynthesis, turnover, transport and storage.

The toxicity of *m*-fluorotyrosine and *p*-fluorophenylalanine is due to the formation of fluoroacetate formed via the tyrosine degradation pathway is well documented. The toxin fluoroacetate is the most ubiquitous of small class of organofluorine compounds and has been identified in more than 40 tropical and subtropical plant species, and is also produces by some microorganisms when incubated in the media containing fluoride. Whether such principle from nature can be "borrowed "in the way that it enables to kill tumor cells and spare normal ones, by use of cytotoxic amino acids? Indeed in drug design for chemotherapy the best way to achieve

the specificity of action and selectivity of delivery is to use proteins as delivery vehicles (protein shuttles). For example, it is conceivable that e.g. lethal metabolic derivation of fluorocitrate and fluoroacetate from m-fluorotyrosine can be achieved intracellulary in target tumour cells, if this cytotoxic amino acid would be specifically delivered (Figure 4.6). MCF10A the growth of the epithelial breast cells transfected (3-F)Tyr-EGFP thought lipid delivery, was blocked when compared with the same cells transfected with wt-EGFP. In this particular case, the (3-F)Tyr was most probably set free during normal cellular protein turnover. Free cytotoxic amino acids have generally two choices. The first one concerns the re-incorporation into other proteins (i.e. re-entering protein translation), and the second one the entry in the intracellular metabolism, biosynthesis and physiology.

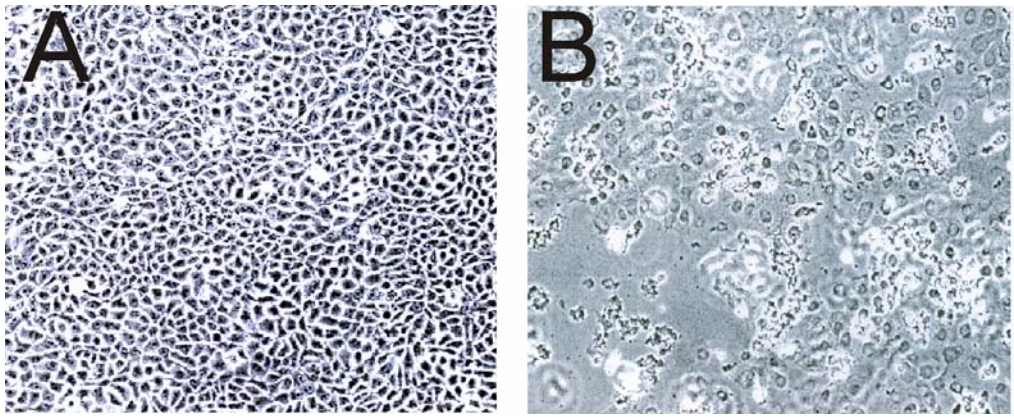


Figure 4.6. Cellular morphology of breast epithelial cell line MCF10A after liposome-mediated delivery of wt-EGFP (A) and (3-F)Tyr-EGFP (B). Cells were allowed to grow for 48 h after treatment.

It is well known from peptide chemistry that a variety of synthetically substituted small derivatives peptides could be prepared and used as carriers for some antibacterial amino acids. Taking into account nearly uniform metabolic roles of amino acid in different tissues or cell types the delivery of cytotoxic amino acid analogues into cytoplasm might lead to the inhibition of the cellular growth, most probably to the by blocking of cellular signaling leading eventually to apoptosis or necrosis. The similar principle could be applied on biosynthetic substitution proteins with drug-like amino acids. The basic advantage of this system should be non-invasiveness, i.e. the cytotoxic amino acid is covalently integrated into polypeptide in inactive prodrug form and exerts no toxicity during delivery. This is indeed, plausible for isosteric analogues or surrogates since they resemble canonical counterparts by shape, size and chemical

properties. Combined with the choice of suitable `shuttle' candidates like antibodies, cytokines, growth factors, or other tissue-specific proteins, their substitution offers the opportunity to intervene in a pathological process with a high degree of specificity and minimal perturbation of normal physiological processes. Future research should show whether such a concept might be converted into practical approach to less toxic cancer therapy.

- 4.5 Engineering of protein spectroscopic properties by Trp-analogues

Tryptophan is found in almost all proteins and its fluorescence dominates the emission of most native proteins. As it is the rarest amino acid residue in globular proteins, (having a total content of only 1%), Trp represents a site-specific intrinsic probe for studying protein structure, dynamics and function. Its conservation throughout many protein families indicates that Trp has essential functions both in the structure of many proteins as well as their means of interacting with other molecules. Since Trp is the main source of fluorescence and absorbance in proteins it was recognized very early to be a useful intrinsic probe in the structure and function of proteins and enzymes. For example, fluorescence maximum of most globular proteins is changed from 330 nm to 350 nm when the protein is denatured by urea. This is attributed to the solvent-accessibility and environment of the Trp residues in the proteins with the buried residues having their maxima close to 330 nm and exposed residues maxima near 350 nm.

This is especially pronounced in the UV spectra of proteins like Barstar (see Result Section) where 3 Trp residues dominate the spectral contribution of other 3 Tyr residues. Moreover, the fluorescence properties of Barstar are even more dramatically changed. The most prominent change was in the fluorescence quantum yield (almost four times higher) for (6-F)Trp-barstar accompanied with a red-shift of about 14 nm. Similarly, (5-F)Trp-barstar shows also a higher quantum yield in comparison with native-protein (about three times higher) with a smaller red-shift of 9 nm. On the other hand, the incorporation of (4-F)Trp or (7-F)Trp leads to total loss of the tryptophan fluorescence and reveal the contribution of the tyrosine residues with an emission maximum at 306 nm. Indeed, 4-fluorotryptophan as a non-fluorescent analogue (“silent fluorophore”) allows for identification of spectroscopic contributions of "hidden" chromophores like tyrosines or phenylalanines in proteins as was shown in (see Result section). This property of the (4-F)Trp is well known and documented in literature. Now by incorporation of (7-F)Trp in Barstar and Annexin V the repertoire of "silent" fluorophores as protein building blocks for *in vivo* translation is expanded.

The substitution of indole with electron-donating amino groups leads to an intramolecular charge transfer that is extremely sensitive to pH changes and might be a better alternative to commonly used analogue (7-Aza)Trp. The amino substituents of aromatic ring systems are less basic in the

excited than in the ground state and their protonation level in the excited state affects the resonant integration of their free electrons in conjugated ring systems causing charge migration seen as blue- or red-shifted spectroscopic bands. That means they have pH-sensitive fluorescence. This property of the aminoindole side chains of Trp analogues such as (5-NH_2)Trp and (4-NH_2)Trp is fully transferred to the target protein, upon incorporation allowing for a pH shuttle of its optical properties. Successful incorporation of (4-NH_2)Trp and (5-NH_2)Trp into Barstar as a model protein, fully confirmed these possibilities (Fig. 4.7). Their absorption profiles resemble those of free amino acids in solution and the spectra of the (4-NH_2)Trp-barstar and (5-NH_2)Trp-Barstar variants are indeed pH sensitive.

For the (4-NH_2)Trp-Barstar a blue-shifted co-operative fluorescence titration curve was obtained with a transition midpoint between pH 5 and 6 (Fig. 4.7.). Conversely, in 5-aminotryptophan-barstar these cooperative transitions of the emission maxima are red-shifted with a transition point at pH 8. These large differences have to be assigned to the changes in the charge density distributions along the indole aromatic ring, which depend upon the nature, orientation and position of the substituent in the parent indole molecule in a particular solvent. In addition, these spectral properties arise also from equilibrium between ground and excited states as well as the corresponding ionized forms as was already discussed elsewhere.

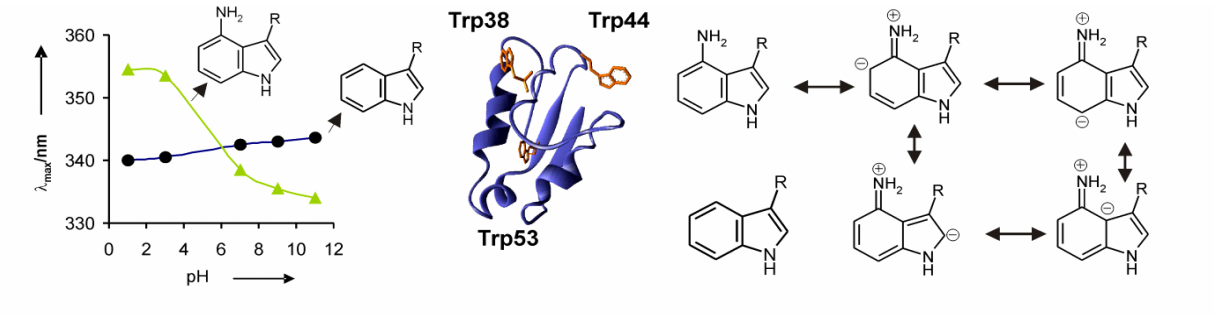


Figure 4.7. The pH-dependence of 4-aminotryptophan ((4-NH_2)Trp) fluorescence and its mesomeric structures depicting possible structures that might arise due to the charge transfer. These fluorescence properties as well as charge transfer of the (4-NH_2)Trp are fully reflected in the solution properties of related substituted protein.

4.6 Manipulating protein stability with monofluorinated Trp-analogues

In the frame of common-sense reasoning, it is difficult to imagine that the exchange of a single atom, *i.e.* atomic mutation such as H→ F, at particular residues in the folded protein structure might produce any significant effects on protein structure, stability and activity. Therefore, it is not surprising that until very recently the thermodynamic effects of atomic mutations at tryptophan residues on protein stability and folding cooperativity had not been investigated at all. On the other hand, since proteins are only marginally stable, all types of molecular interactions are important and even small interactions can contribute significantly to the stability. For example, the replacement of the Barstar mutants with different number of Trp side chains with related fluorinated analogues was demonstrated to change not only the intrinsic spectroscopic properties of this protein, but also its folding/unfolding properties. It is difficult to speculate about these various aspects without further structural studies of these mutants and variants. However, such single atom exchanges represent atomic mutations that result in slightly increased covalent bond lengths and inverted polarities in the residue side-chain structure. These almost negligible changes in the local geometry do not cause any significant alternation either in the secondary or in the tertiary structures of the mutants, which were identical to those of the wild-type protein in crystal form. On the other hand, these variants exhibit significant differences in stability and folding cooperativity, when compared to the wild-type protein. These rather large global effects, resulting from the minimal local changes, have to be attributed either to the relatively strong changes in polar interactions of the indole ring or to differences in the van-der-Waals radii or to a combination of both factors. The changes in local geometry that are below the resolution of protein X-ray crystallographic studies are probably of secondary importance in comparison to the strong electronegativity introduced by the fluorine atom. Therefore, this type of mutations provides an interesting reporter system to study cooperative functions of integrated residues by using non canonical amino acids.

4.7 Other applications of non canonical amino acids

The possibility to expand the amino acid repertoire will lead to supplementation or even full replacement of traditional methods for protein engineering and design. An expanded amino acid repertoire of the genetic code has already proven to be useful in the design of novel proteins, the rational alternation of existing proteins, and will certainly be a precious tool for understanding the precise molecular basis of protein structure-function relationships, ligand binding, conformational change and protein-protein interactions. It is now also widely accepted that incorporation of synthetic (i.e. non canonical) amino acids have a great potential for various industrial applications such as:

- design of proteins and peptides, "armed" pharmacologically active non-canonical amino acids, as vehicles for drug delivery and targeting
- enhancement or modulation of protein folding and stability
- incorporation of suitable markers into proteins for structural genomics (X-ray crystallography, NMR spectroscopy) and proteomics studies
- design of enzymes/biocatalysts with changes in substrate specificity and range of pH activity
- design of non-invasive protein-based sensors for *in vitro* and *in vivo* applications as well as non-invasive protein-based diagnostic tools

Isosteric replacements in the frame of some canonical amino acids such as Ser/Cys, Thr/Val represent an excellent tool to study the nature of the catalytic mechanism of residues directly involved in enzymatic action. However these replacements are not possible for canonical amino acids such as His or Pro. This problem can be circumvented by exploiting the opportunity to seize the wider choice of analogues offered by non canonical amino acids. Other properties such as hydrophobicity, can also be effectively probed using non canonical amino acids, as it has been done by different groups (Corringer et al., 2000, Arias, 2000), who succeeded in clarifying the role of certain amino acids at a conserved site in the ion channel region of the nicotinic acetylcholine receptor. Another example include cation-π interactions which were investigated (Ting et al., 1998) through the use of non canonical amino acids such as fluorinated derivatives of Trp, as it is known that fluorine has a significant and additive effect on the cation-π interactions. The site specific incorporation of biophysical probes such as fluorescent groups or spin labels might profit from an expanded amino acid repertoire as well. Fluorescent groups can allow static and dynamic investigations of protein function in several ways, as it has been demonstrated by several different works (Cornish et al., 1994, Stewart et al., 1997, Karginov et al., 1999, Turcatti et al., 1996).

The use of amino acids with photoreactive side chains represents another useful biophysical tool. Especially useful for this purpose are "caged" amino acid side chains (Adams et al., 1993), in which a heteroatom is protected as an *o*-nitrobenzyl group. After incorporation of the non canonical amino acid into the protein, protecting group removed by photolysis reveals the previously caged functionality. Schultz and coworkers (Cook et al., 1995) succeeded in incorporating caged serine at a crucial site in a self-splicing intein system.

One interesting observation is that α-hydroxy acids can be used to replace α-amino acids, producing a backbone ester in place of the usual amide bond. By incorporating a α-hydroxy acid

into a well-defined helix of T4 lysozyme Schultz and coworkers were able to establish the relative importance for the helical stability of two effects due to the α-hydroxy acid incorporation: the removal of the peptide bond and the decreased capability of the carbonyl to form hydrogen bonds (Koh et al., 1997).

5 REFERENCES

A

Adams, S.R., and Tsien, R.Y (1993) Controlling cell chemistry with caged compounds. *Annu. Rev. Physiol.*, **55**, 755-784

Alberti,S. (1997) The origin of the genetic code and protein synthesys. *J. Mol. Evol.,***45**, 352-358

Anderson, S., Bankier, A.T., Barrell, B.G., de Brujin, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R., and Young, I.G. (1981) Sequence and organization of the human mitochondrial genome. *Nature*, **290**, 457-465

Arias,HR (2000) Localization of agonist and competitive antagonist binding sites on nicotinic acetylcholine receptors. *Neurochem. Int*., **36**, 595-645

B

Bae, J.H., Alefelder, S., Kaiser, J.T., Friedrich, R., Moroder, L., Huber, R. and Budisa, N.(2001) Incorporation of beta-selenolo 3,2-b pyrrolyl-alanine into proteins for phase determination in protein X-ray crystallography. *Journal of Molecular Biology*, **309**, 925-936.

Bae, J.H., Rubini, M., Jung, G., Wiegand, G., Seifert, M.H.J., Azim, M.K., Kim, J.S., Zumbusch, A., Holak, T.A., Moroder, L., Huber, R. and Budisa, N. (2003) Expansion of the genetic code enables design of a novel "gold" class of green fluorescent proteins. *Journal of Molecular Biology*, **328**, 1071-1081.

Baca, M., Alewood, P.F., and Kent, S.B. (1993) Structural engineering of the HIV-1 protease molecule with a β-turn mimic of fixed geometry. *Protein Science*.,**2**, 1085-1091

Baca, M., Kent, S.B. (1993) Catalytic contribution of flap-substrate hydrogen bonds in HIV-1 protease explored by chemical synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, **90**, 11638-11642

Baldwin, R.L. (2002). Making a Network of Hydrophobic Clusters. *Science*, 295, 1657-1658

Becker, H.D. and Kern, D. (1998). *Thermus thermophilus*: A link in evolution of the tRNA-dependent amino acid amidation pathways., *Proceedings of the National Academy of Sciences of the United States of America*, 95, 12832-12837

Bock, A., Forchhammer, K., Heider, J., Leinfelder, W., Sawers, G., Veprek, B. and Zinoni, F. (1991) Selenocysteine - the 21st Amino-Acid. *Molecular Microbiology*, **5**, 515-520.

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

Borst, P. and Grivell, L.A. (1981) Small beautiful portrait of a mitochondrial genome. *Nature*, **90**, 443-444

Brawerman, G. and Ycas M. (1957) Incorporation of the Amino Acid Analog Tryptazan into the Protein of Escherichia-Coli. *Archives of Biochemistry and Biophysics*, **68**, 112-117

Brünger, A.T., Adams,P.D., Clore, G.M., DeLano, W.L., Gros, P., and Grosse-Kunstleve, R.W. (1998). Cristallography and NMR system. (CNS): a mew software system for macromolecular structure determination. *Acta Chryst*., **D54**, 905-921

Budisa, N., Minks, C., Medrano, F.J., Lutz, J., Huber, R. and Moroder, L. (1998) 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.

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

Budisa, N., Alefelder, S., Bae, J.H., Golbik, R., Minks, C., Huber, R. and Moroder, L. (2001) Proteins with beta-(thienopyrrolyl)alanines as alternative chromophores and pharmaceutically active amino acids. *Protein Science*, **10**, 1281-1292.

Budisa, N., Rubini, M., Bae, J.H., Weyher, E., Wenger, W., Golbik, R., Huber, R. and Moroder, L. (2002) Global replacement of tryptophan with aminotryptophans generates non-invasive protein-based optical pH sensors. *Angewandte Chemie-International Edition*, **41**, 4066-4069.

Budisa, N., Karnbrock, W., Steinbacher, S., Humm, A., Prade, L., Neuefeind, T., Moroder, L. and Huber, R. (1997) Bioincorporation of telluromethionine into proteins: A promising new approach for X-ray structure analysis of proteins. *Journal of Molecular Biology*, **270**, 616-623.

Budisa, N**.,** Alefelder, S., Bae, J.H., Golbik, R., Minks, C., Huber, R. and Moroder, L. (2001) Proteins with beta-(thienopyrrolyl)alanines as alternative chromophores and pharmaceutically active amino acids. *Protein Science*, **10**, 1281-1292.

Budisa, N. (2004). Prolegomena to future efforts on genetic code engineering by expanding its amino acid repertoire. *Angewandte Chemie* **45**, *In press*.

Budisa, N. (2004). Adding new tools to expressed protein ligation. *ChemBioChem. In press*.

Budisa, N. (2004). Protein Engineering and Design with an Expanded Amino Acid Repertoire. Habilitation, Technische Universität, München, Germany

C

Chapeville, F., Ehrenstein, G.V., Benzer, S., Weisblum, B., Ray, W.J. and Lipmann, F. (1962) 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**, 1086-1098.

Cohen, S.N., Chang, A.C.Y., Boyer, H.W. and Helling, R.B. (1973) Construction of biologically functional bacterial plasmids in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, **70**, 3240-3244.

Cook, S.N., Jack, W.E., Xiong, X., Danley, L.E., Ellman, J.A., Schultz, P.G., and Noren, C.J: (1995) Photochemically initiated protein splicing. *Angew. Chem. Int. Ed. Engl.,* **34**, 1629-1630

Corey, D.R. and Schultz, P.G. (1987) Generation of a hybrid sequence-specific single-stranded deoxy ribonuclease. Science, **238**, 1401-1403

Cornish, V.W., Benson, D.R., Altenbach, C.A, Hideg, K., Hubbell, W.L., and Schultz, P.G. (1994). Site-specific incorporation of biophysical probes into proteins. *Procl. Nat. Acad. Sci. USA*, **91**, 2910-2914

Corringer, P.J., Le Novere, N., and Changeux, J.P. (2000). Nicotinic receptors at the amino acid level. *Annu. Rev. Pharmacol. Toxicol*., **40**,431-458

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

Crameri, A, Whitehorn, E.A., Tate, and E., Stemmer, W.P.C.(1996). Improved Green Fluorescent Protein by molecular evolution using DNA shuffling. *Nature Biotechnology*, **14**, 315-319

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

D

Dawson, P.E., Muir,T.W., Clark-Lewis, I., and Kent, S.B. (1992) Synthesis of proteins by native chemical ligation. *Science*, **266**, 776-779

Di Giulio, M. (1997) On the origin of the genetic code. J. Theor. Biol., **187**, 573-581

Dougherty D.A. (2000). Unnatural amino acids as probes of protein structure and function. *Current Opinion in Chemical Biology,* **4**, 645-652

E

Eriani G., Delarue M., Poch O., Gangloff J. & Moras D. (1990). *Nature*,**347**, 203-206

F

Fenster, E.D. and Anker, H.S. (1969). Incorporation into plypeptide and charging on transfer ribonucleic acid of amino acid analog 5′, 5′, 5′,-triföuoroleucine by leucine auxotrophs of Escherichia coli. Biochemistry, **8**, 269-277

Frankel, A. & Roberts R.W. (2003). *RNA,* **9**, 780-786

G

Giege′, R. (2003) Genetic code expansion. *Nature Structural Biology*, **10**, 414-416

Gogarten, J.P., Senejani, A.G., Zhaxybayeva, O., Olendzenski, L., Hilario, E. (2002).Annual Review of Microbiology

Golbik, R., Fischer, G. and Fersht, A.R. (1999) Folding of barstar C40A/C82A/P27A and catalysis of the peptidyl-prolyl cis/trans isomerization by human cytosolic cyclophilin (Cyp18). *Protein Science*, **8**, 1505-1514.

Guillet, V., Lapthorn, A., Fourniat, J., Benoit, J.P., Hartley, R.W., and Mauguen, Y. (1993) Crystallization and preliminary-X-Ray Investigation of Barstar, the Intracellular Inhibitor of Barnase. *Proteins-Structure Function and Genetics*, 17, 325-328

H

Hirao, I., Ohtsuki, T., Fujiwara, T., Mitsui, T., Yokogawa, T., Okuni, T., Nakayama, H., Takio, K., Yabuki, T., Kigawa, T., Kodama, K., Yokogawa, T., Nishikawa, K. & Yokoyama S.(2002). An unnatural base pair for incorporating amino acid analogs into proteins. *Nature Biotechnol*. **20**, 177-182.

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

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

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

Huber, R.E., Segel, I.H., and Criddle, R.S. (1967). Growth of Escherichia coli on selenate. *Biochim. Biophys. Acta*, **141**, 587-599.

Huber, R.E., and Criddle, R.S. (1967). Isolation and properties of beta-Galactosidase from Escherichia coli grown on sodium selenate. *Biochim. Biophys. Acta*, **141**, 573-586.

Hutchinson, C.A. 3rd, Phillips,S., Edgell, M.H., Gillam,S., Jahnke,P & Smith M. (1978). Mutagenesis at a specific position in a DNA Sequence. *J. Biol. Chem*., **253**, 6551-6560

I

Ibba, M., and Hennecke, K. (1995). Relaxing the substrate specificity of an aminoacyl-transfer-RNA synthetase allows in-vitro and in-vivo synthesis of proteins containing unnatural amino-acids. *FEBS Letters*, **364**, 272-275

Ibba, M. and Söll, D. (2000). Aminoacyl-tRNA Synthesis. *Annual Review of Biochemistry*, **69**, 617-650

J

Jakubowski, H. & Goldman, E. (1992). Editing of errors in selection of amino acids for protein synthesis. *Microbiol. Rev*. **56**, 412-429

K

Kaiser, E.T., and Lawrence, D.S. (1984) Chemical mutation of enzyme active sites. *Science*, **226**, 505-511

Karginov, A.V., Lodder, and M., Hecht, S.M. (1999) Facile characterization of translation initiation via nonsense codon suppression. Nucleic Acid Research, **27**, 3283-3290

Kauzmann, W. (1954) Some Factors in the Interpretation of Protein Denaturation. *Advances in Protein Chemistry*, **14**, 1-63

Kellis, J.T., Jr, Nyberg, K., and Fersht A.R. (1989) Contribution to Hydrophobic interactions to protein stability, *Biochemistry*, **28**, 4914-4922

Kiick, K.L., van Hest, J.C.M. and Tirrell, D.A. (2000) Expanding the scope of protein biosynthesis by altering the methionyl-tRNA synthetase activity of a bacterial expression host. *Angewandte Chemie-International Edition*, **39**, 2148-2151.

Koh, J.T, Cornish, V.W., and Schultz, P.G. (1997) An experimental approach to evaluating the role of backbone interactions in proteins using unnatural amino acid mutagenesis. *Biochemistry*, **36**, 11314-11322

Kowal, A.K., Kohrer,C.& RajBhandari U.L. (2001). Twenty-first aminoacyl-tRNA synthetase-suppressor tRNA pairs for possible use in site-specific incorporation of amino acid analogues into proteins in eukaryotes and in eubacteria. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 2268-2273

Kukhar, V. P., and Soloshonok, V.A. (1991) *Russian Chemical Reviews*, ***60***, 850.

Kwon I., Kirshenbaum K. & Tirrell D.A. (2003). Breaking the degeneracy of the genetic code. *J. Am. Chem. Soc*.**125**, 7512-7513.

L

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

Lazar, G.A., and Handel, T.M. (1998). *Current Opinion in Chemical Biology*, **2**, 675-679

Leinfelder, W., Zehelein, E., Mandrand-Berthelot, M. A., and Bock, A. (1988). Gene for a novel transfer-tRNA species that accepts L-Serine and cotraslationally inserts selenocysteine. *Nature*, **331**, 723-725.

Leslie, A.G.W. (1998). *MOSFLM*. Cambridge, U.K.

Liu, D.R., Magliery, T.J. & Schultz P.G. (1997). Characterization of an "orthogonal" suppressor tRNA derived from E. coli tRNA(2)(Gln). *Chemistry & Biology*, **4**, 685-691.

Loidl, G., Musiol, H.J., Budisa, N., Huber, R., Poirot, s., Fourmy, D., and Moroder, L. (2000) Synthesis of beta-(1-azulenyl)-L-Alanine as a potential blue-coloured fluorescent Tryptophan analogue and its use in peptide synthesys. *Journal of Peptide Science*, **6**, 139-144

M

Makadzde, G., and Privalov, P. (1990) Heat capacity of proteins. I. Partial molar heat capacity of individual amino acid residues in aqueous solution: hydration effect. *J. Mol. Biol*., **213**, 375-384

Matsumura, M., Becktel, W.J., and Matthews, B.W. (1988). Hydrophobic stabilisation in T4 Lysozyme determined directly by multiple substitutions of Ile 3. *Nature*, **334**, 406-410

Merrifield, R.B. (1963) *J. Am. Chem. Soc.*, **85**, 2149-

Miles, E. (1977) Tryptophan synthase of *Escherichia coli*. *J. Biol. Chem*., **252**, 6594-6599

Minks, C., Alefelder, S., Moroder, L., Huber, R. and Budisa, N. (2000a) 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.

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

Muir, T.W., Sondhi, and D., Cole, P.A (1998) *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 6705-

N

Nath, U., and Udgaonkar, J.B. (1997) Folding of tryptophan mutants of Barstar: Evidence for an initial hydrophobic collapse on the folding pathway. *Biochemistry*, **28**, 8602-8610.

Navaza, J. (1994) AMoRe: an automated package for molecular replacement. *Acta Cryst*., **A50**, 157-163

Neil, E. and Marsh, G. (2000) Towards the non-stick egg: designing fluorous proteins. *Chemistry and Biology*, **7**, 153-157

Nirenberg, M.W, and Matthaei J.H (1961) The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci*., **47**, 1588-1602

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

Nowak, M.W., Kearney, P.C., Sampson, J.R., Saks, M.E., Labarca, C.G., Silverman, S.K., Zhong, W., Thorson, J, Abelson, J.N., Davidson, N., Peter G. Schultz, Dougherty D. A.& Lester H.A. (1995). Nicotinic receptor-binding site probed with unnatural amino acid incorporation in intact cells. *Science,* **268**, 439-442.

O

Osawa, S., Jukes, T.H., Watanabe, K., and Muto, A. (1992) Recent evidence for evolution of the genetic code. *Microbiol. Rev*., **56**, 229-264

P

Pardee, A.B., Shore, V.G., and Prestidge L.S. (1956) Incorporation of Azatryptophan into Proteins of Bacteria and Bacteriophage. *Biochimica and Biophysica Acta*, **21**, 406-407

Pastrnak M, Magliery TJ & Schultz P.G. (2000). A new orthogonal suppressor tRNA/aminoacyl-tRNA synthetase pair for evolving an organism with an expanded genetic code. *Helvetica Chimica Acta*, **38**, 2277-2286.

Pei, D. and Schultz P.G. (1990). Site-specific cleavage of duplex DNA with a λ repressor-staphylococcal nuclease hybrid. *J.Am.Chem.Soc*., 112, 4579-4580

Pendergrast, P.S., Ebright, Y., and Ebright R. (1994). High-specifity DNA cleavage agent. Design and application to kilobase and megabase DNA substrates. *Science*, **265**, 959-961

Ponder, J.W., and Richards, F.M. (1987). Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol*., **193**, 775-791.

R

Radzicka, A. and Wolfenden, R. (1988) Comparing the polarities of the amino acids - Side-chain distribution coefficients between the vapor-phase, cyclohexane, 1-octanol, and neutral aqueous solution. *Biochemistry*, **27**, 1664-1670.

Richards, F.M. (1997). Protein stability:still an unresolved problem. *Cell mol. life sci*., **53**, 790-802.

Richards, F.M. (1977). Area, volumes, packing, and protein structure. *Ann. Rev. Biophys. Bioeng*., **6**, 151-176

Richmond, M.H. (1962). The effect of amino acid analogues on growth and protein synthesis in microorganisms. *Bacteriol. Rev*., **26**, 398-420.

Rose, G.D., and Wolfenden., R. (1993). Hydrogen bonding, Hydrophobicity, Packing and Protein Folding. *Annu. Rev. Biophys. Biomol. Struct*., **22**, 381-415.

Ross, J.B.A., Szabo, A.G. and Hogue, C.W.V. (1997) Enhancement of protein spectra with tryptophan analogs: Fluorescence spectroscopy of protein-protein and protein-nucleic acid interactions. *Fluorescence Spectroscopy*, Vol. 278, pp. 151-190.

Rubini, M., Lepthien, S., Pal, P.P., Huber, R., Moroder, L. and Budisa, N. (2004) Thermodynamics of the expanded genetic code: Structure and stability of Barstar as pH sensor. Dynamics of Proteins, Symposium of the SFB 533, Freising, 9-11 July

S

Sakamoto, K., Hayashi,A., Sakamoto, A., Kiga, D., Nakayama, H., Soma, A., Kobayashi, T., Kitabatake, M., Takio, K., Saito, K., Shirouzu, M., Hirao, I. & Yokoyama S. (2002). Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells. *Nucleic Acids Research,* **30**, 4692-4699.

Sandberg, W.S., and Terwilliger, T.C. (1991) Repacking protein interiors. *TIBTECH*, **9**, 59-63

Santoro, M M., and Bolen, D.V. (1988) Unfolding free energy changes determined by the linear extrapolation method. *Biochemistry*, **27**, 8063

Schmitt, E., Moulinier, L., Fujiwara, S., Imanaka, T., Thierry, J.C., and Moras, D. (1998) Crystal structure of aspartyl-tRNA synthetase from *Pyrococcus kodakaraensis* KOD: archaeon specificity and catalytic mechanism of adenylate formation. *The EMBO Journal*, **17**, 5227-5237

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

Sisido, M. and Hohsaka, T. (2001) Introduction of special functions by the position-specific incorporation of nonnatural amino acids into proteins through four base codon/anticodon pairs. *Appl. Microbiol. Biotechnol*., **57**, 274-281

Spirin A.S. (2002) Ribosome as a molecular machine. *FEBS Letters*, **514**, 2-10.

Stemmer, W.P.C. (1994) DNA shuffling by Random Fragmentation and reassembly: *in vitro* recombination for molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 10747-10751.

Stewart, L.E., Collins, C.S., Gilmore M.A., Carlson, J.E., Ross, J.B.A., Chamberlin, A.R. (1997) In vitro site-specific incorporation of fluorescent probes into β-galactosidase. *J. Am. Chem. Soc.*, **119**, 6-11

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

Sykes, B.D., H., W. and Schlesininger, M.J. (1974) Fluorotyrosine Alkalyne-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.

T

Tanford,C. (1979) Interfacial free-Energy and the Hydrophobic effect. *Proceedings of the National Academy of Sciences of the United States of America*, **76**, 4175-4176

Tang, Y. and Tirrell D.A. (2001) Biosynthesis of a highly stable coiled-coil protein containing hexafluoroleucine in a engineered bacterial host. *J. Am. Chem. Soc.*, **123**, 11089-11090

Ting, A.Y., Shin, I., Lucero, C., Schultz, P.G. (1998) Energetic analysis of an engineered cation-π interaction in staphylococcal nuclease. *J. Am. Chem. Soc.*, 120, 7135-7136

Turcatti, G., Nemeth, K., Edgerton, M.D., Meseth, U., Talabot, F., Peitsch, M., Knowles, J., Vogel, H., Chollet, A. (1996) Probing the structure and function of the tachykinin neurokinin-2 receptor through biosynthetic incorporation of fluorescent amino acids at specific sites. *J. Biol. Chem.*, **271**, 19991-19998

Turk, D. (1996) MAIN 96: An interactive software for density modifications, model building structure refinement and analysis. In *Proceedings from the 1996 meeting of the International Union of Christallography Macromolecular Computing School*. (ed. Bourne, P.E. & Watenpaugh K.)

W

Weber-Ban, E., Hur O., Bagwell C., Banik, U., Yang, L.H., Miles, E.W., and Dunn M.F. (2001) Investigation of Allosteric Linkages in the Regulation of Tryptophan Synthase: The Roles of Salt Bridges and Monovalent Cations Probed by Site-Directed Mutation, Optical Spectroscopy, and Kinetics. Biochemistry, **40**, 3497-3511.

Weberndorfer, G., Hofacker, I.L., Stadler, P.F. (2003) On the evolution of primitive genetic codes. Origins of Life and Evolution of the Biosphere, **33**, 491-514.

Wilson, M.J. and Hatfield, D.L. (1984) Incorporation of modified amino acids into proteins *in vivo*. *Biochimica et Biophysica Acta*, **781**, 205-215.

Y

Yutani, K., Ogasahara, K., Tsujita, T., and Sugino, Y. (1987). Dependance of conformational stability on Hydrophobicity of the Amino Acid Residue in a Series of Variant Proteins Substituted at a Unique Position of Tryptophan Synthase α Subunit. *Proc. Natl. Acad. Sci. U.S.A.*, **84**, 4441-4444

Z

Zdanowsky, K. and Dadlez M. (1999) Stability of the residual structure in unfolded BPTI in different conditions of temperature and solvent composition measured by disulphide kinetics and double mutant cycle analysis. *J. Mol. Biol*., **287**, 433-445.

Zinoni, F., Birkmann, A., Stadtman, T.C., and Bock, A. (1987). Cotraslational insertion of selenocysteine into formate dehydrogenase from *Escherichia coli* directed by a UGA codon. *Proc. Natl. Acad. Sci. U.S.A*., **84**, 3156-3160

Zinoni, F., Birkmann, A., Leinfelder, W., and Bock, A. (1986). Nucleotide-sequence and expression of the selenocysteine-containing polypeptide of formate dehydrogenase (Formate-hydrogen-Lyase-linked) from *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A*., **83**, 4650-4654