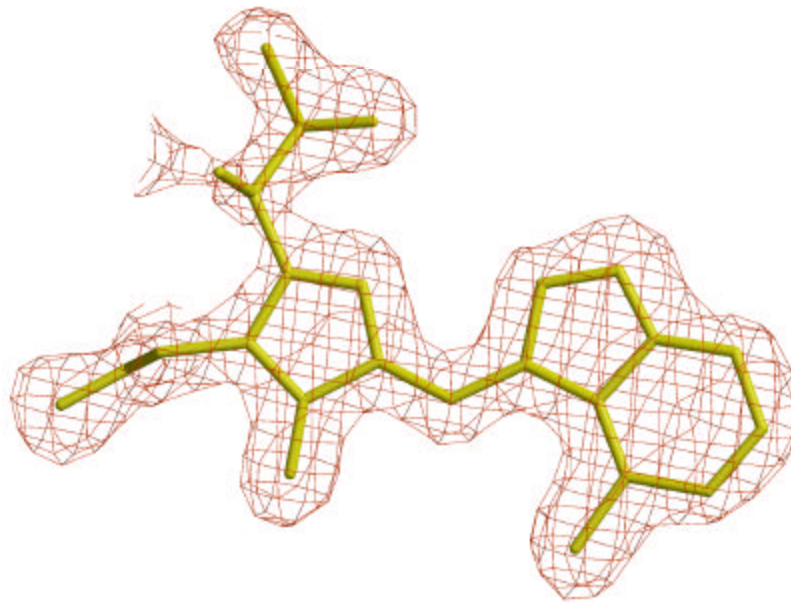Max-Planck-Institut für Biochemie
Abteilung für Strukturforschung

Studies on Tryptophan Analogues in Proteins

Jae Hyun Bae

München 2002

Max-Planck-Institut für Biochemie
Abteilung für Strukturforschung

Studies on Tryptophan Analogues in Proteins

Jae Hyun Bae

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. Johannes Buchner

Prüfer der Dissertation:

1. apl. Prof. Dr. Dr. h.c. Robert Huber

2. Univ.-Prof. Dr. Dr. Adelbert Bacher

Die Dissertation wurde am 28.10.2002 bei der Technischen Universität München eingereicht und
durch die Fakultät für Chemie am 04.12.2002 angenommen.

To God, my family and grandma

ACKNOWLEDGMENT

PUBLICATION LIST

Parts of this work are published in following publications:

Budisa, N., Alefelder, S., **Bae, J. H**., Golbik, R., Minks, C., Huber, R. & Moroder, L. (2001). Proteins with beta-(thienopyrrolyl)alanines as alternative chromophores and pharmaceutically active amino acids. *Protein Sci.* **10(7)**, 1281-1292

Bae, J. H., Alefelder, S., Kaiser, J. T., Friedrich, R., Moroder, L., Huber, R. & Budisa, N. (2001). Incorporation of beta-selenolo[3,2-b]pyrrolyl-alanine into proteins for phase determination in protein X-ray crystallography. *J. Mol. Biol.* **309(4)**. 925-936

Budisa, N., Rubini, M., **Bae, J. H.**, Weyher, E., Wenger, W., Golbik, R., Huber, R. & Moroder, L. (2002). Global replacement of tryptophan with aminotryptophans generates non-invasive protein-based optical pH sensors. *Angew. Chem. Int. Ed.* **41(21)**. 4066-4069

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. & Budisa, N. (2002). Expansion of the Genetic Code Enables Design of a Novel "Gold" Class of Green Fluorescent Proteins. *J. Mol. Biol.* Submitted.

ABSTRACT

Amino acid tryptophan (Trp) which serves as a substrate in ribosome-mediated protein biosynthesis is usually a main chromophore in proteins. Its natural abundance in amino acid sequences of proteins is very low, and it has only one coding triplet in the genetic code (UGG). On the other hand, Trp has special biophysical properties which allow for its participation in a numerous interactions in proteins ($\pi-\pi$ stacking, hydrogen bonding, cation-π interactions). These interactions play a major role in protein stability and folding, receptor-ligand interactions, enzyme-substrate binding, antigen-antibody recognition, etc. Due to the multiple type of reactions it can participate in, Trp occupies a unique position among gene-encoded amino aicds. For those reasons Trp is an attractive target amino aicd for protein engineering and design, as well as for study of protein structure and function.

In the present study additional Trp-analogues and Trp-like amino acids (i.e. surrogates) containing sulphur and selenium are chemically synthesised and *in vivo* translated into model proteins using our-laboratory-developed selective pressure incorporation (SPI) method. Sulphur-containing isosteric surrogates of Trp, L-ß-(thieno[3,2-b]pyrrolyl)-alanine ([3,2]Tpa) and L-ß-(thieno[2,3-b]pyrrolyl)-alanine ([2,3]Tpa) are pharmacologically active substances whose thieno-pyrrolyl moieties are bio-isosteric analogues of the hallucinogen and serotonin agonist N,N'-dimethyl-tryptamine. Proteins substituted in this way, due to their potential ability of selective delivery and targeting, might serve as a specific "shuttle" of even "magic bullets" for drug delivery in biomedicine.

Relatively low abundance of Trp in proteins and growing interest for incorporation of heavy-atom-containing (i.e. electron rich) amino acids into proteins in order to solve their three-dimensional structures using X-ray crystallography, motivated us to chemically synthesise and

incorporate two selenium-containing Trp-surrogates: L-ß-(selenolo[3,2-b]pyrrolyl)alanine ([3,2]Sep) and L-ß-(selenolo[2,3-b]pyrrolyl)alanine ([2,3]Sep) into model proteins, barstar and annexin V. This approach represents a useful supplementation and even a promising novel alternative to selenomethionine (SeMet) for solving the phase problem in protein X-ray crystallography.

Finally, Trp as a chromophore amino acid is the main source of UV absorbance and fluorescence of proteins. Thus, its substitution with suitable analogues would provide a new spectral windows and subsequently enable for many useful spectroscopic applications. Most recent breakthrough in this field is achieved in our laboratory, where Trp residues in barstar and GFP were successfully replaced with $(4-NH_2)$-Trp and $(5-NH_2)$-Trp respectively, using SPI method. Conversion of pH-insensible barstar into pH sensible protein species by Amino-Trp incorporation provided an example for a general approach for the design of protein-based molecular sensors. Moreover, by incorporation of $(4-NH_2)$-Trp into "cyan" class of green fluorescent proteins (ECFP) yielded a novel "golden" class of green fluorescent proteins (Gold Fluorescent Protein, GdFP). Since both proteins (ECFP, GdFP) were found to have an exceptional biophysical and spectroscopic properties, it was necessary to determine their three-dimensional structures. In this study, crystal structures of ECFP and GdFP were elucidated using molecular replacement approach, providing in this way a structural framework for understanding their unusual spectroscopic properties and open a perspective for their further future studies.

CONTENTS

FUTURE PERSPECTIVES

APPENDIX

INTRODUCTION

1   Genetic Codes and Flow of the Genetic Information

For living organisms, entire genetic information is saved in string-like nucleic acids forms (ribonucleic acid, RNA; deoxyribonuclieic acid, DNA). DNA is two-stranded molecule which folds to form a helical structure which resembles a spiral staircase (i.e. double helix) where each strand is a polynucleotide composed of **A** (adenosine), **T** (thymidine), **C** (cytidine), and **G** (guanosine) residues. In RNA molecules the **T** (thymidine), is replaced with **U** (uridine). The synthesis of these linear chains with specific sequences is polymerisation reaction where each strand has its own polarity. This polarity means that the 5'-hydroxyl (or 5'-phospho) group of the first nucleotide begins the strand and the 3'-hydroxyl group of the final nucleotide ends the strand. The two strands of DNA run anti-parallel to each other i.e. when one strand runs 5' -> 3' the other runs 3' -> 5' direction. Another prominent feature of DNA as main carrier of biological information is the complementarity between opposing strands of the double-stranded DNA molecule. That means, **A** forms two hydrogen-bonds with **T**; **C** forms three hydrogen bonds with **G**. This enables for another prominent feature of DNA – the capacity for self-replication as well as transcription into corresponding RNA forms. Two strands of DNA are not equal: coding or sense strand is transcribed into messenger RNA (mRNA) which is further decoded (via codon-anticodon interactions) at ribosome, and non-coding or antisense strand.

A flow chart is given in Scheme 1 representing a relation between two complementary DNA strands (template) and a mRNA which serves as a template replica. mRNA is translated at ribosome into corresponding protein sequence according to the rules prescribed by genetic code.

The genetic code is a set of rules which correlates sequence molecules in nucleic acids and sequence of amino acids in proteins (Table 1).

(5' -> 3') ᴧᴧᴧ**ATGGAATTCTCGCTC**ᴧᴧᴧ (Coding, sense strand)

(3' <- 5") ᴧᴧᴧ**TACCTTAAGAGCGAG**ᴧᴧᴧ (Non-coding, antisense strand)

} DNA template

↓ Transcription

(5' -> 3') ᴧᴧᴧ**AUGGAAUUCUCGCUC**ᴧᴧᴧ (mRNA, template replica)
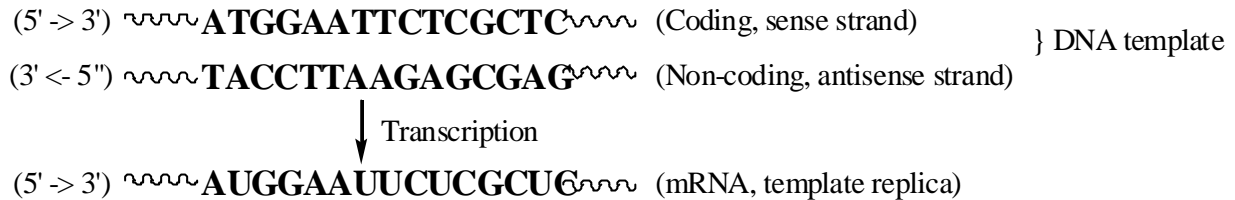
Scheme 1. The sense strand of self-replicable DNA template is transcribed into mRNA (template replica) which is translated into proteins at ribosome.

		Second Position of Codon								
		T		C		A		G		
First Position	T	TTT	Phe	TCT	Ser	TAT	Tyr	TGT	Cys	T	Third Position
		TTC	Phe	TCC	Ser	TAC	Tyr	TGC	Cys	C	
		TTA	Leu	TCA	Ser	TAA	*Ter*	TGA	*Ter*	A	
		TTG	Leu	TCG	Ser	TAG	*Ter*	TGG	Trp	G	
	C	CTT	Leu	CCT	Pro	CAT	His	CGT	Arg	T	
		CTC	Leu	CCC	Pro	CAC	His	CGC	Arg	C	
		CTA	Leu	CCA	Pro	CAA	Gln	CGA	Arg	A	
		CTG	Leu	CCG	Pro	CAG	Gln	CGG	Arg	G	
	A	ATT	Ile	ACT	Thr	AAT	Asn	AGT	Ser	T	
		ATC	Ile	ACC	Thr	AAC	Asn	AGC	Ser	C	
		ATA	Ile	ACA	Thr	AAA	Lys	AGA	Arg	A	
		ATG	Met	ACG	Thr	AAG	Lys	AGG	Arg	G	
	G	GTT	Val	GCT	Ala	GAT	Asp	GGT	Gly	T	
		GTC	Val	GCC	Ala	GAC	Asp	GGC	Gly	C	
		GTA	Val	GCA	Ala	GAA	Glu	GGA	Gly	A	
		GTG	Val	GCG	Ala	GAG	Glu	GGG	Gly	G	

Table 1. The Universal Genetic Code in represented in DNA-format. Each amino acid is coded by combination of three nucleotides (coding triplets). There are 64 coding triplets assigned to only 20 amino acids which is the cause for the degeneracy of the genetic code (i.e. most of the amino acids are assigned with more than one coding triplet).

Coding triplets presented at mRNA are *decoded* by transfer RNAs (tRNA) which interact with a ribosome-bound mRNA containing the coding sequence (*translation*). Theoretically there should be 64 different tRNAs. The fact, that there are 61 coding triplets for only 20 amino acids,

indicates that the genetic code is degenerated, i.e. usually one amino acid is coded with more than one coding triplet (with the exception of Met and Trp). For that reason, amino acids having also more than one tRNA adapter (cognate tRNA), and such tRNAs are known as *isoacceptor* tRNAs. Three triplet combinations, Umber (UGA), Amber (UAG), and Orche (UAA) serve as termination signals in protein syntheses. Remaining 61 codons are decoded on ribosome by tRNA molecules which have anticodon loops with complementary nucleotides. Each of 20 standard amino acids is enzymatically transferred onto cognate tRNA in the aminoacylation reaction catalysed by aminoacyl-tRNA synthetases (aaRS). The tRNA attached with amino acid (i.e. aminoacyl-tRNA) enters ribosome where it participates into decoding of mRNA (via codon-anticodon interactins) while attached amino acid participates in peptide bond formation (Figure 1).
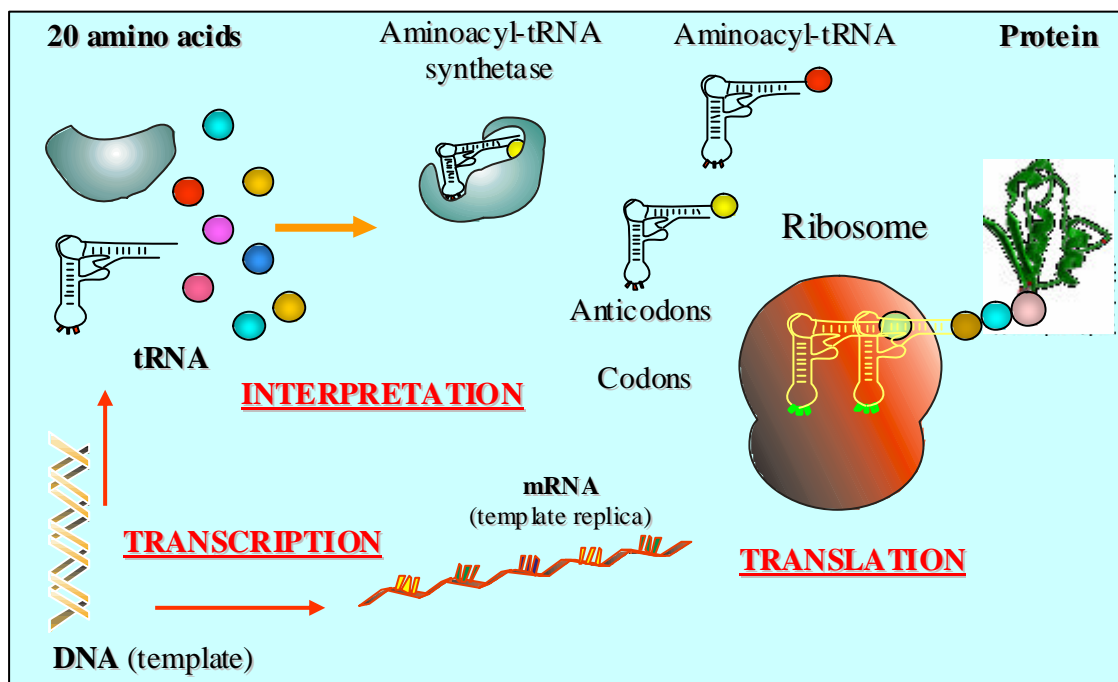
Figure 1. A general plan for template-directed, ribosome-mediated protein synthesis. The genetic information is stored in self-replicating DNA (template) from, which is subsequently transcribed (transcription) into related RNA molecules (mRNA, tRNA etc.). For the correct protein translation, two molecular recognition events are crucial: codon-anticodon interactions as ribosome (translation in *sense strictu*) and amino acid recognition and transfer onto cognate tRNA by aaRS in the aminoacylation reaction. The aminoacylation reaction can be therefore regarded as interpretation of the genetic code.

Transfer RNA molecules (tRNA) serve as adapters in protein synthesis, and are involved in two reactions: they are charged with proper amino acids in the aminoacylation reaction, and charged tRNAs participate in mRNA decoding and in peptide bond formation in ribosome. This is possible due to the special shape of tRNAs with two prominent features: anticodon loop and acceptor part for amino acid attachment. Since the formation of the peptide bond is themodynamically unfavourable, the carboxyl group of the amino acid is activated by attachment to the tRNA, forming an aminoacyl-tRNA (aa-tRNA). This reaction is catalysed by specific enzymes, *aminoacyl-tRNA synthetases* (aaRSs), which are highly selective in their recognition of both the amino acid to be activated and the related tRNA. By catalysing a specific aminoacylation reaction the aaRS provies a critical link between amino acids and codons (Figure 1), i.e. the interpretation of the genetic code.

Usually, the Genetic Code in the Table 1 has also been called "The Universal Genetic Code" since the relations between codons and amino acids are similar in almost all living beings (animals, plants, fungi, archaea, bacteria, and viruses). Although small variations in the coding rules exist in mitochondria and certain microbes, the genetic code rules are widely accepted as conservative and universal since observed variances are found in only a tiny fraction of known living beings, cells, and organelles. Finally, the amino acid repertoire is conserved in all living beings without exception.

2   Methods for the Amino Acid Repertoire Extension

2.1   Classical Site-Directed Mutagenesis

Classical site-directed mutagenesis include codon manipulations but does not affect amino acid repertoire. For example, it is possible to mutate AUG (Met) codon to UGG (Trp) codon using

properly designed oligonucleotides in PCR reaction. Therefore, site-directed mutagenesis is one of the most important tools for understanding the relationship between the structure, sequence and function of studied proteins. For example, this method has made possible the engineering of proteins with enhanced properties including stability, catalytic function and binding specificity. It was discovered by Hutchison and co-workers a few decades ago (Hutchison, *et al.*, 1978). However, it should be stressed that the changes in protein structure made by this approach are limited to the repertoire twenty encoded amino acids. Thus, the ability to incorporate non-canonical amino acids into proteins would greatly expand the ability to manipulate the structures of proteins, both to probe protein function and create proteins with new properties.

2.2 Expansion of the Amino Acid Repertoire using Suppressor-based Methodologies

The idea to use chemically misacylated tRNAs for the incorporation of non-canonical amino acids into proteins is almost four decades old (Chapeville, *et al.*, 1962; Hecht, *et al.*, 1978). The *in vitro suppression* approaches for an amino acid repertoire expansions are based on the early observations that chain terminating mutations are correctable in some bacterial strains by specific suppressor tRNAs which have altered reading capacities. The alteration was caused by mutations in their anticodons, i.e. they have the termination anticodons as their own anticodons. These tRNAs can insert different canonical amino acids in response to the *nonsense* (one of the three terminator codons appears in mRNA) or *missense* (genetic mutation involving alteration of one sense to another sense codon so that different amino acids are determined) codons or *frameshift* mutation in the parent gene (Garen, 1968; Engelhardt, *et al.*, 1965). Discovery that *misacylated* tRNAs (tRNAs which contain a different amino acid from their genetic information) participate in peptide bond formation at ribosomes (Chapeville, *et al.*, 1962) in a fashion consistent with the adapter hypothesis (Crick, 1966) was the second main prerequisite for the in vitro *suppression*

methodology. Finally, this approach requires suitable *in vitro* transcription and translation system (Spirin *et al.*, 1988).

One of the approaches developed in Sisido's laboratory is based on use of frameshift suppressor tRNAs capable of reading (readthrough) quadruplet base pairs (codons) in gene coding for target protein as it is shown in Figure 2A. Desired amino acid is incorporated in response to quadruplet codon inserted in a parent gene and decoded by chemically misacylated tRNA (Sisido, *et al.*, 1993, 1999; Hohsaka, *et al.*, 1993; Hohsaka, *et al.*, 1996; Hohsaka, *et al.*, 1999). Usually target gene sequence is optimised in terms of codon-composition, i.e. in the case that quadruplet is not decoded, one of the following coding triplets is terminal one, i.e. protein synthesis is automatically blocked (Figure 2B).

Another approach is based on readthrough (or suppression) of nonsense coding triplets as follows. A suppressor tRNA is prepared so that recognises the stop (nonsense) codon UAG (but is not recognised by the naturally occurring aminoacyl-tRNA synthetases) and chemically aminoacylated with the desired non-canonical amino acid. Conventional site-directed mutagenesis is used to introduce the stop codon UAG at the site of interest in the protein gene (Figure 2C). When the acylated suppressor tRNA and the mutant gene are combined in an *in vitro* transcription/translation system, the non-canonical amino acid is incorporated in response to the UAG codon to give a protein containing that amino acid at the specified position (Noren, *et al.*, 1989). Its general features are given in Scheme 2.

The basic limitation of all suppressor-based (or readthrough-based) approaches is the limited capacity of chemically misacylated synthetic suppressor tRNA charged with non-canonical amino acid to fully decode (suppress, readthrough) nonsense (triplet or quadruplet) codon. This is not surprising since the suppressions are exceptional phenomena which are *context-dependent*, i.e. each desired position into protein sequence cannot be always suppressed. It has been

observed and documented that *release* (i.e. intervention of so-call *release factors* (RF1 and RF2), proteins known to participate in translation termination) is so effective at some sites that the natural function of the amber stop codon is not suppressed even when the suppressor tRNA is charged with the amino acid that normally appears at that point in the wild-type protein (Kowal & Oliver, 1997). Second, chemical aminoacylations are expensive and complicated experiments. Finally, this method can be used only *in vitro*, which is associated with very low yields of mutated proteins (i.e. sub-milligram range), as well as with sub-optimal conditions for protein synthesis.
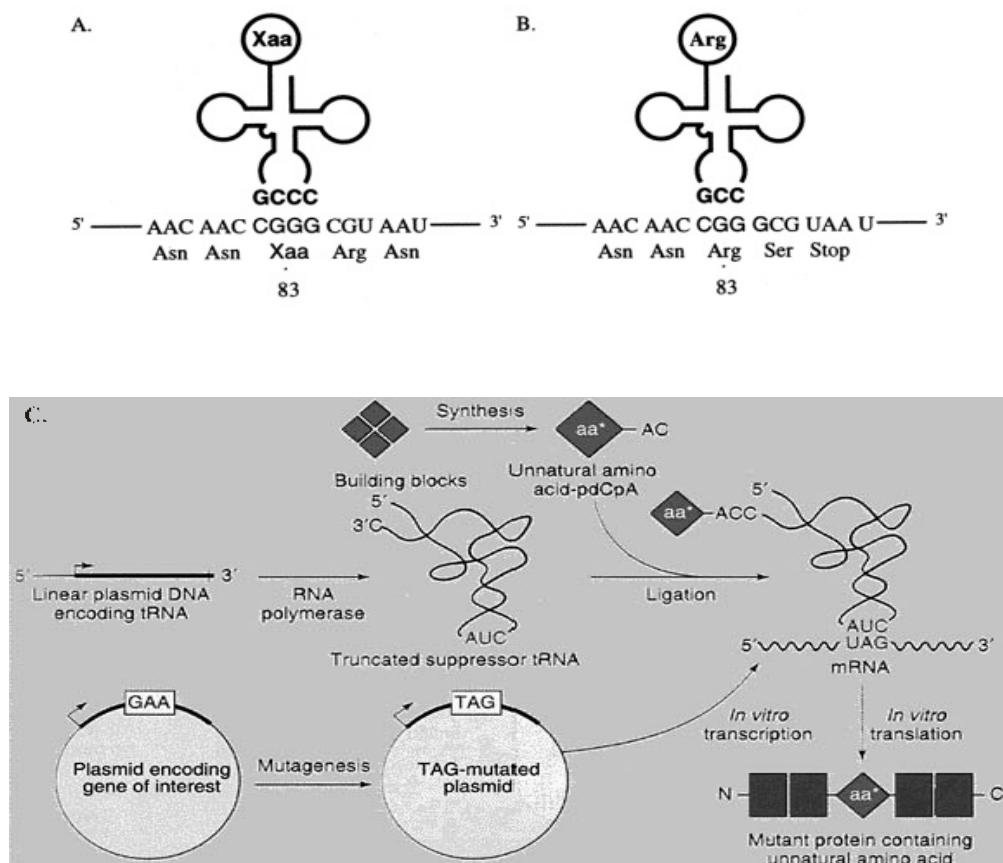
Figure 2. A) Frameshift suppression, where one quadruplet is decoded as non-cannonical amino acids (position 83). **B)** In the case that frameshift suppression is not possible, the termination of protein synthesis follows (Figure A and B are taken from Hohsaka, *et al*, 2001). **C)** General scheme for non-cannonical amino acid incorporation using UGA (Amber) suppression (Figure taken from Wang & Schultz, 2001)

2.3 Novel Development of Suppressor-based Mutagenesis: Orthogonal tRNA/aaRS pairs

Further step in development of readthrough (or suppression) based methodologies is their transfer to *in vivo* conditions. There are few reasons for that, but most importantly is that currently available in vitro systems suffering from low protein yields. In addition such general approach for the site-specific incorporation *in vivo*, where cells are taking desired non-canonical amino acids directly from the growth media, might be advantageous over existing systems and approaches. A few years ago, Schultz and co-workers (Wang & Schultz, 2001) have reported an experimental effort in this direction which could be summarised as three-step procedure as follows.

1. The construction of a suppressor tRNA that must be "orthogonal" i.e. does not interact with any of the naturally occurring aminoacyl-tRNA synthetases and which should be able to deliver non-canonical amino acid in response to a stop codon (usually UAG) in the mRNA encoding the protein of interest;

2. The design and engineering of "orthogonal" aminoacyl-tRNA synthetase (using mutagensis and screening approaches) which should recognise related orthogonal-tRNA but does not recognize any endogenous tRNAs; and

3. The screening of a library of mutants of both orthogonal tRNAs and aaRS in order to find "orthogonal" aaRS-tRNA pair capable of activating and transferring of the desired non-canonical amino acid but not any canonical one.

Very recently, experiments where sufficient orthogonal suppressor tRNA/synthetase pairs were capable to incorporate non-canonical amino acids (for example O-Methyl-Tyr) *in vivo* instead of related natural amino acids are reported (Wang, *et al*., 2001; Wang & Schultz, 2001).

Although *in vivo* transfer of suppression was done to improve mainly protein yields, in reality almost the same difficulties remained, with additional difficulties associated with the metabolic toxicity and bioavailability of the desired amino acids (i.e. their transfer into the cytoplasm). First, the orthogonal suppressor tRNA/synthetase pair for each specific non-canonical amino acid

has to be generated through the series of complicated mutagenesis and selection cycles (Lee, *et al.*, 2001; Wang & Schultz, 2001). Second, like in *in vitro* approach all positions of proteins are not equally suppressible (and many of them probably not at all). Third, the yields of mutant proteins will always depend on how much is possible to suppress desired position and this can be in the range of 0 – 100 %. Finally, this *in vivo* approach is restricted to those amino acid analogues that are closely related in size and structure to the canonical amino acids (Köhler, *et al.*, 2001). But the attempt to make proteins with "radically different" amino acids than canonical ones was one of the basic motivations for the development of this approach (Noren et al. 2989). A comparison of these approaches with the method used in our laboratory is given in Scheme 2. Since almost two years in our laboratory, there have been efforts to develop an alternative to suppressor (readthrough) based methodologies because they are statistical in nature (i.e. efficient incorporation is allowed only for some positions in protein sequences) and inefficient in terms of yields of the mutant protein. The goal of these efforts is to develop a hybrid translation system based on heterologous protein expression. This system should use sense coding triplets instead of terminal coding triplets or frameshifted DNA templates, to avoid complications arising from suppressions and the action of release factors. In addition, such hybrid translation system will be compatible with host cell metabolism and genetics. Finally it should provide position-specific translation on non-cannonical amino acids by use of orthogonal aaRS:tRNA pairs and codon-optimised DNA/RNA template sequences.

2.4   Residue -Specific Incorporation: Selective Pressure Incorporation (SPI)

From historical and practical perspectives, the most favourable approach for the incorporation of non-canonical amino acids is the expression of proteins in auxotrophic host cells, such as *ML304D* (Cowie, *et al.*, 1957), *DL42* (Hendrickson, *et al.*, 1990), *B834(DE3)* (Golden, *et al.*,

1993), or *PO1562* (Hädener, *et al.*, 1993). Based on the strict absence of sulphur in the growth medium, Cowie and Cohen (1957) reported the complete incorporation of the slightly toxic methoinine analogue, selenomethionine, in bacterial proteins. Before the work of Budisa (Budisa, *et al*., 1995), there were no reports for the high level or even quantitative replacement of more toxic or unstable non-canonical amino acids (Anfinsen, *et al.*, 1969; Gilles, *et al.*, 1989). This became possible upon development of technologies and approaches for efficient and controllable expression of recombinant proteins.

In comparison to *in vitro* suppression (see Scheme 2), SPI-procedure for analogue incorporation is very simple. It is necessary to grow the cells in a defined minimal media with a limited concentration of the canonical amino acid and to keep the target gene transcriptionally silent. After depletion of the canonical amino acid, the stationary phase of the growth is reached. At this point, the exchange of the canonical amino acid with the appropriate analogue and the simultaneous induction of the target protein will result in the accumulation of the protein containing exclusively the non-canonical amino acid analogue instead of the canonical amino acid. The SPI-method works since full selective pressure is obtained in the context where:

(1) the host strain exhibits a strong auxotrophism during the fermentation procedure;

(2) the cloned gene is under the efficient control of the promoter;

(3) the competitive expression system is able to use cytosolic enzymes to express mainly target DNA after the induction of the protein synthesis.

In the presence of background expression (*gene leakage*) before the induction of protein synthesis, the resulting protein will be incompletely labelled.

Although this simple method results in usually high amounts of recombinant protein, it is limited only to those non-canonical amino acid analogues which are in the editing tolerance range of aminoacyl-tRNA synthetases.

IN VITRO SUPPRESSION	SELECTIVE PRESSURE INCORPORATION (SPI)
BASIC PROTOCOLS:	
1) tRNA preparation (preparation of suppressor tRNA) - Chemical synthesis of pdCpA dinucleotide - Aminoacylation of pdCpA dinucleotide -Enzymatic ligation of aminoacyl-dinuclotide to suppressor tRNA	**BASIC PROTOCOLS:** - Routine transformation of intact auxotrophic expression host cells -Requirement for controllable and robust expression system - Fermentation and expression in defined minimal media - Protein mutant isolation
2) In vitro transcription-translation (protein production) - Preparation of E. coli S-30 extracts - Preparation of plasmid for use as template DNA in S-30 reactions - *In vitro* control reactions (optimisation) - *In vitro* suppression and expression - Protein mutant isolation	
ADVANTAGES - site specific incorporation -large number of noncanonical amino acids can be introduced into proteins	*ADVANTAGES* - residue-specific incorporation - wild-type protein yield levels - simple, easy reproducible technology
DRAWBACKS - extremely low protein yield levels (submiligramm range) - expensive and complicated technology, not easy to reproduce -suppressions or readthrough are exceptional, codon-context dependent phenomena	*DRAWBACKS* - limited number of non-canonical amino acids can be introduced into proteins
PERSPECTIVES - possible practical applications after overcoming these drawbacks are practically limitless	*PERSPECTIVES* - the potential of the wt-level (industrial scale) of production of substituted proteins

Scheme 2. Comparison between readthrough (suppression) based methodologies with the method used in our laboratory, SPI (selective pressure incorporation).

3   Tryptophan as Model Amino Acid for Substitution Studies using SPI Method

With the development of different methods for non-cannonical amino acids incorporation many limits of classical protein engineering based on 20 standard amino acids might be overcome. In this context, the repertoire for the tryptophan coding triplet, UGG, has been expanded with novel Trp analogues such as (4-F)-Trp, (5-F)-Trp, (6-F)-Trp, and (7-Aza)-Trp (Bronskill & Wong, 1984; Ross, *et al.*, 1997; Minks, *et al.*, 1999, 2000a; Soumillion, *et al.*, 1995). Trp indeed represents an attractive target for such replacements since it is the main source of the absorption and fluorescence of proteins. Moreover, occurrence of Trp is rare as it represents only about 1% of all residues of globular proteins (McCaul & Ludescher, 1999) and it may well provide a quasi site-specific intrinsic probe for studying protein structure, dynamics and function. Apart from its role as basic building block in ribosome-mediated protein synthesis, Trp plays an essential role in the metabolism of living cells as it is involved in the biosynthesis of hormones such as serotonin or melatonin in animals, and of indole alkaloids in plants (Phillips, *et al*, 1995). Therefore, non-canonical amino acids that mimic Trp could be of great interest as potential antagonists, drugs or antibiotics. And their incorporation into proteins could be a promising approach for production of therapeutic agents (Budisa, *et al.*, 1998b; Minks, *et al.*, 2000b).

The number of pharmacologically active or spectroscopically and structurally interesting Trp analogues or Trp-like amino acids (surrogates) is further extended in a last few years in our laboratory (Bae, *et al.*, 2001; Bae, *et al.*, 2002; Budisa, *et al.*, 2001; Budisa, *et al.*, 2002). They are [3,2]Tpa, [2,3]Tpa (pharmacologically active), [3,2]Sep, [2,3]Sep (crystallographic markers), (4-OH)-Trp, (5-OH)-Trp (metabolic precursors), (4-NH$_2$)-Trp, and (5-NH$_2$)-Trp (pH sensitive amino acids).

3   **Sulfur- and Selenium-containing Tryptophan like Amino Acids**

For the synthesis of possibly useful isosteric heavy-atom analogues of tryptophan, the nitrogen of the indole moiety appears to be a very attractive target. However, upon replacing this nitrogen with sulphur, the resulting planar molecule was not enzymatically activated for condensation with serine by the tryptophan-synthase. This suggested that the related synthetic Trp analogue would probably not be recognised and activated by the tryptophanyl-tRNA synthetase in the protein translation step. Even changes in positions adjacent to the imino group were expected to be disturbing, since the (2-Aza)-Trp was also not a substrate for protein synthesis. Conversely, (7-Aza)-Trp is recognised by cellular tryptophanyl-tRNA synthetase and incorporated into proteins (Soumillion, *et al*., 1995; Ross, *et al*., 1997). Therefore, the benzene ring of the indole moiety is apparently a much more permissive target for chemical transformations that might give compounds containing electron-rich atoms suitable for X-ray crystallography.

All benzene-based amino acids (Phe, Tyr, Trp) exhibit strong quadrupole moments that arise from the non-spherical charge distribution; introduction of heteroatoms into these systems results in novel properties not found in the parent molecule. For example, the sulphur atom in thiophene, selenium in selenophene, has an unshared pair of electrons in a *p*-orbital conjugated with the carbon-carbon double bonds, and unlike the carbon, nitrogen, and oxygen, it has vacant *d*-orbitals in the outer shell and can therefore act as an electron acceptor. The fused pyrroles are more permissive to interact with nearby charges with induction of dipoles in the system as well as additional dispersion forces, polarisabilites, exciplex formation, resonance energy transfer or with formation of charge transfer complexes. Other properties that differentiate them from benzene include enhanced hyperpolarisability and differences in aromatic delocalisation energies (benzene: 36 kcal/mol, thiophene: 29 kcal/mol, thiazole: 25 kcal/mol) (Bird, 1985).

4   X-ray Crystallography

The availability of three-dimensional structures for biologists, chemists, and biophysicists provided them with an excellent tool to assess, study, and understand proteins' functional roles in living beings. For that reason, much efforts have been dedicated to facilitate the elucidation of protein structures used different methods such as X-ray crystallography, nuclear magnetic resonance spectroscopy (NMR), electron microscopy (EM), as well as other approaches.

The pioneering works on protein structure elucidation using X-ray diffraction patterns of single haemoglobin and myoglobin crystals were made by Perutz and Kendrew (Perutz, *et al*, 1960; Kendrew, *et al*, 1960). Original proposal for the possibility for structure determination of proteins using diffraction properties of crystals was made by J. D. Bernal in 1934 (Bernal & Crowfoot, 1934). Diffraction studies of DNA fibrins in crystalline form made by R. Franklin were bases for a model DNA helical structure originally proposed by Watson and Crick in 1953 (Watson & Crick, 1953).

Since then, protein crystallography has become a well established and most powerful technique for protein structure determination. Coupled with rapid developments in molecular biology and biochemistry as well as by advances in computing technologies in the last 20 years, this has led to the solution of more than 11200 protein structures which are deposited in the Protein Data Bank in the end of 1999. More than 9100 of those structures were solved by X-ray diffraction, around 1800 by multidimensional nuclear magnetic resonance spectroscopy (NMR) and the remaining 200 by computer modelling. This data emphasises the central role of protein crystallography in the field of structural biology.

Structure solved by NMR is at present still limited to small proteins of up to 30 kDa, but it complements protein crystallography in several ways. It provides information on protein

dynamics that cannot be obtained from the rather rigid environment of a crystal lattice and it is also not dependent on the availability of protein crystals. Usually the structures determined by these two different techniques are in many cases almost identical. Differences arise especially when exposed protein regions are hindered by contacts in the crystalline lattice.

In recent years, the advances in radiation detection and computing power by the introduction of synchrotron radiation have made possible to study enzyme catalysis and associated conformational changes in the crystalline state as well as in large biological complexes such as ribosome, troteosome, nucleosome, etc. Time-resolved or Laue crystallography enables for crystallographic studies of different dynamic processes in proteins like substitute of domain motives usually not accessible by classical X-ray crystallography.

Cryo-crystallography, named also as "cold" crystallography, has been in existence almost as long as "classical" protein X-ray crystallography. Early experiments used sealed cryostats, which were essentially dewars with beryllium walls transparent to the scattered X-rays. This method is still used today for experiments that must be performed below liquid nitrogen temperatures, 77 K. Later experiments used chillers to cool air to as low as 200 K. This is a lower practical limit due to the problems associated with dehumidifying the air and removing carbon dioxide. Modern cryo-crystallography is the general methodology by which the crystal of interest is placed into a protective solution (a cryoprotectant solution) and frozen with minimal damage, resulting in data collection at cryogenic temperatures (liquid nitrogen temperatures).

Cold, but not solid state, crystallography has a number of special advantages: the data collection process can be conducted with fewer worries about crystal decay, data with high signal-to-noise ratios can be collected, and optimal crystals can be more readily selected.

A detailed discussion of other aspect of protein crystallography is far beyond the scope of this study and can be found in relevant textbooks and articles (Blundell & Johnson, 1994; Drenth, 1994; McRee, 1993).

5.1 Phase Problem in solving the Protein Structures with X-ray Crystallography

Since it is possible to detect X-rays which are scattered cooperatively by their interactions with electrons of a large number of the identical protein molecules arranged in the unit cell of a crystal lattice, the positions and intensities of such reflections contain enough information for accurate structure determination. As the result of the data collection is a list of intensities for every reflection in the three-dimensional reciprocal lattice (I_{hkl}). The amplitude of the scattered wave can be considered as Fourier transform of the density of elementary scatterers which may reveal the electron distribution *i.e.* electron densities. Therefore, a unique solution is possible, but not only the X-ray intensities but also their phases must be known. However, the phase information is lost during the measurements – it is not possible to measure the phases of the radiation whose wavelength is in the order of a magnitude of 10^{-10} meter. For that reason, for each reflection a phase angle (a_{hkl}) must be determined (*phase problem*). The phase problem still remains one of the main obstacles in solving novel protein structures, in spite of the tremendous progress in the theoretical background, experimental techniques, and the accumulation of structural knowledge. Therefore, a variety of the approaches were developed to overcome the phase problem in X-ray crystallography:

a. Multiple Isomorphous Replacement (MIR)

b. Multiwavelength Anomalous Dispersion (MAD)

c. Patterson Search or Molecular Replacement (MR)

d. Direct Methods

In the absence of any previous phase information, the MIR method is certainly the most important method for the solution of the phase problem in protein crystallography. With the availability of tuneable synchrotron radiation, the MAD method (Hendrickson, *et al*, 1988) is becoming more and more relevant for the solution of protein structures. If a related protein structure is available the method of Patterson search (Molecular Replacement, MR) can be used (Hoppe, 1957a; Hoppe, 1957b; Rossman & Blow, 1962). The prerequisite for this method is the availability of a structural homologue for the protein under investigation. Thus, the related structure (phasing model) is used to obtain phase information after orienting the model in the unit cell of the new protein.

5.2   Multiple Isomorphous Replacement: MIR

The method of multiple isomorphous replacement (MIR) involves the introduction of a new X-ray scatterer of high atomic number (*heavy atom*) into the macromolecule under study without disrupting its structure or packing in the crystal. This can be achieved either by soaking the heavy atom compound into the crystal or by co-crystallising it together with the protein. It is particularly important not to disturb the crystal structure in this process. Therefore, soaking with the help of the large solvent channels present in protein crystals is usually the method of choice. Addition of one heavy atom to a macromolecule indtroduces differences in the diffraction pattern of the derivative crystal in relation to that of the native (Figure 3).

$$F_{PH} = F_P + F_H$$

F_P corresponds to the structure factor of the native protein. It represents a vector with a structure factor amplitude F_P and a_P. F_{PH} of the heavy atom derivative can be derived from F_P by the

vector addition of F_H, which is the contribution of the heavy atom(s) to the structure factor of the

derivative. The position of the heavy atom can be determined from a difference Patterson map

between the derivative and the native protein crystals. After adequate refinement of this heavy

atom position(s) the protein phases can be calculated:

$$a_P = a_H + cos^{-1}((F^2_{PH} - F^2_P - F^2_H)/(2F_P F_H)) = a_P \pm a'$$
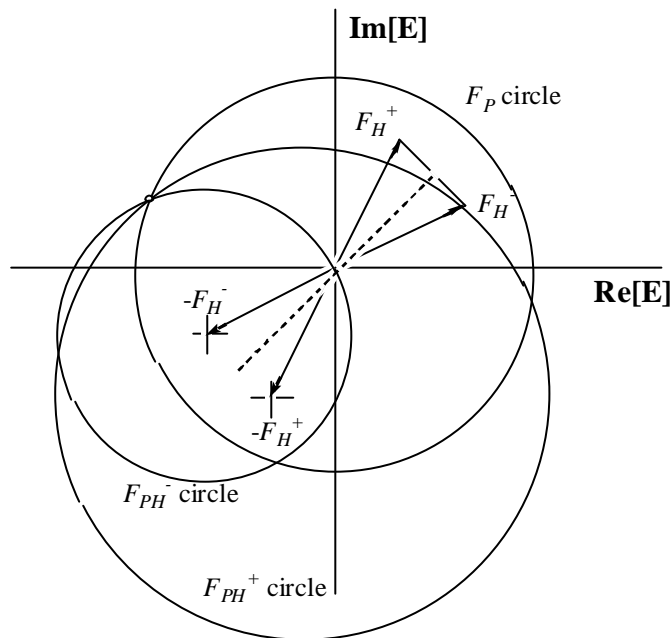
Figure 3. A Harker diagram for the deduction of protein phase angles by anoamlous scattering (after Drenth, 1994). $|F_P|$ is the structure factor amplitude for the reference (remote) dataset and $|F_{PH}^+|$ and $|F_{PH}^-|$ for the Friedel motes of the dataset with maximised anomalous contribution (f''). The contribution of the metal to the structure factor is F_H^+ for one member of Friedel pair and F_H^- for the other member. Due to the anomalous scattering component, these two structure factors are not symmetric to the horizontal axis. For clarification, the F_H vectors have been drawn with opposite phase angles such that the three circles intersect in a single point. The dashed line then indicates the direction of the nonanomalous scattering part of the heavy atoms.

As shown in the previous equation, there are always two possible values for a_P, which cannot be

distinguished with a single isomorphous derivative. The phase ambiguity for acentric reflections

can be illustrated geometrically in a Harker diagram (Harker, 1956; Figure 3). The ambiguity in

the phase is theoretically solved when a second, unrelated heavy atom derivative is obtained. In

practice, experimental data are affected by a number of errors arising from the crystal growth procedure, the soaking (or co-crystallisation) process, non-isomorphism, and errors derived from the collection of X-ray diffraction data. To obtain an interpretable electron density map, usually more than two isomorphous heavy atom derivatives are necessary. Additionally, the phase ambiguity can be solved by incorporating anomalous scattering data, density modification methods (e.g., solvent flattening) or averaging the calculated electron density when non-crystallographic symmetry is present.

5.3 Multiwavelength Anomalous Dispersion: MAD

The classical derivation of the physical basis for anomalous scattering is developed by treating the scattering interaction as a pair of coupled oscillators. One oscillator is the *e* component incident X-ray wave; the other is the scattering atom treated as a dipole oscillator with frequency equal to the observed value of an absorption edge. This treatment is expounded at length by James (1948), but does not lend itself to an intuitive summary. Here is a basic description in terms of the scattering cross section of an atom as seen by an incident X-ray photon.

When the incident photon has relatively low energy:
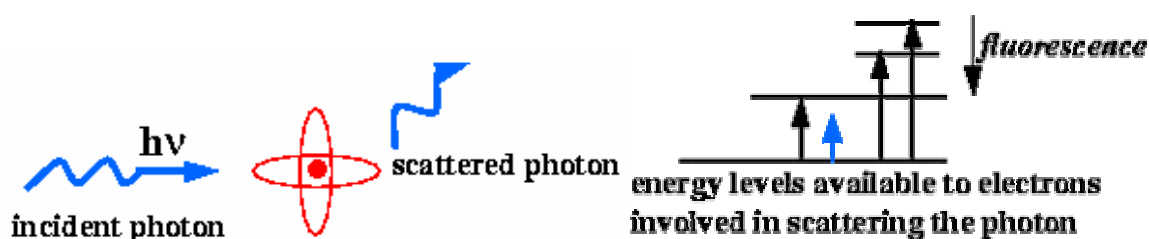
Figure 4. Diagram for the normal scattering pattern (Pciture taken from Merritt, www.bmsc.washington.edu/scatter)

a.   The photon is either scattered or not, but is not absorbed as it has insufficient energy to excite any of the available electronic transitions.

b.   The scattering cross-section of the atom (or the probability that the photon is scattered) may be adequately described in using the normal atomic scattering coefficient f^0 only.

c.   The photon scatters with no phase delay (imaginary, or f'', component is 0).

When the incident photon has high enough energy:
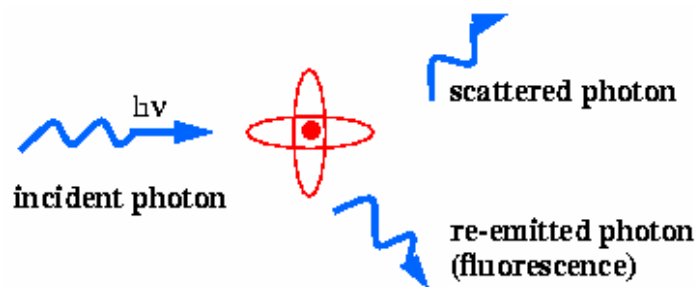
Figure 5. Diagram for the anomalous scattering (Pciture taken from Merritt, www.bmsc.washington.edu/scatter)

a.   Some photons are scattered normally.

b.   Some photons are absorbed and re-emitted at lower energy (fluorescence).

c.   Some photons are absorbed and immediately re-emitted at the same energy (strong coupling to absorption edge energy).

d.   The scattered photon gains an imaginary component to its phase (f'' scattering coefficient becomes non-zero); i.e. it is retarded compared to a normally scattered photon.

This effect is most easily measured as a function of X-ray energy by noting either the sharp increase in absorption or in fluorescence. The imaginary scattering component f'' is proportional to these directly measurable quantities. The real scattering component f' is related to f'' via the Kramers-Kronig relationship.

The breakdown of Friedel's law caused by the imaginary component of anomalous scattering (*f''*) has been used for many years as a source of phase information in macromolecular crystallography. Wavelength-tunable synchrotron radiation allows the real component (*f'*) to be used as well, proving the opportunity for direct phasing through combination of the orthogonal effects of *f'* and *f''*. To exploit the *f'* dispersive component, measurements must be made at multiple wavelengths. Both the real and the imaginary parts of the anomalous scattering vary with wavelength, especially in the region of the absorption edge of the anomalous scatterer. The differences in X-ray diffraction pattern at different wavelengths can be used to calculate estimates of the phases of the protein by a method that is equivalent to MIR. The optimal wavelengths for MAD experiments using anomalous scatterers correspond to a minimum *f'* (inflection point of the absorption edge) and to a maximum *f''*, and a point above the edge at which *f'* is different from its minimum value and *f''* is still large. Although the differences in intensities will be less than for the isomorphous replacement method, MAD is insensitive to systematic errors due to non-isomorphism. Data collection on frozen crystals reinforces the advantages of multiwavelength data collection because there is virtually no radiation damage. Particular care must be taken also in the scaling of the data sets and in the calculation of the absorption effects as different wavelength.

6   Phase Evaluation for Protein Structures using Selenium-containing Amino Acids

The ever increasing number of protein structures resolved by the MAD method is made possible primarily by the simplicity of the *in vivo* selenomethionine (SeMet) incorporation into recombinant proteins directed by the Met AUG-codon. Due to its moderate toxicity, SeMet

(Cowie & Cohen, 1957) together with trifluoroleucine (Richmond, 1962) is the only known non-canonical amino acid that supports the growth of auxotrophic *E. coli* strains *in vivo* in a defined minimal medium. Other non-canonical amino acids that contain even heavier atoms than selenium, e.g. telluromethionine (TeMet), are highly toxic and do not support cellular growth at all (Budisa, *et al.*, 1995). In fact, nearly all amino acids outside the 20 canonical gene-encoded ones are efficient inhibitors of cellular growth (Budisa, *et al.*, 1999). The traditional *in vivo* incorporation approach (Richmond, 1962) has shown that auxotrophic *E. coli* strains can be used to circumvent problems associated with the toxicity of amino acid analogues as well as with the preferred usage of native amino acids. However, low incorporation levels and inadequate protein yields remained the major stumbling blocks for the wider use of this method for protein-mutant expression. A significant improvement was achieved by developing the selective pressure incorporation (SPI) method with which a full preference for the incorporation of non-canonical amino acids over canonical ones can be obtained (Budisa, *et al.*, 1998; Minks, *et al.*, 2000).

7 Model Proteins

For this study, three different model proteins, human annexin V, barstar, and green fluorescent protein variants, were used.

7.1 Human Annexin V

A suitable model protein for replacement experiments should have a known three-dimensional structure and well established biochemical, biophysical, genetic and kinetic properties. In addition, a controllable, robust and efficient plasmid-directed expression should be available in

suitable auxotrophic host strain. A model protein that meets above mentioned conditions is recombinant human annexin V (AxV) which is expressed in soluble form under the control of the T5 promoter/polymerase system. AxV binds in a calcium-dependent manner to acidic phospholipid membrane head groups (Huber, *et al.*, 1990; Berendes, *et al.*, 1993; Liemann & Huber, 1997). AxV contains only one, but essential Trp residue (W187) which in the crystalline state can have two molecular conformations: one where Trp187 is buried in the hydrophobic niche of the domain III (Figure 38) and another where Trp187 is completely exposed to the bulk solvent (Figure 6a; Berendes, *et al.*, 1993; Concha, *et al.*, 1995). Minks and co-workers demonstrated that this residue can be replaced with fluoro-substituted Trp-analogues without affecting the protein structure in solution and in crystals, although the thermodynamic properties were drastically altered (Minks, *et al.*, 1999).

A)Human annexin V   B)Barstar (b*C40A/C82A/P27A)   C)Green Fluorescent Protein
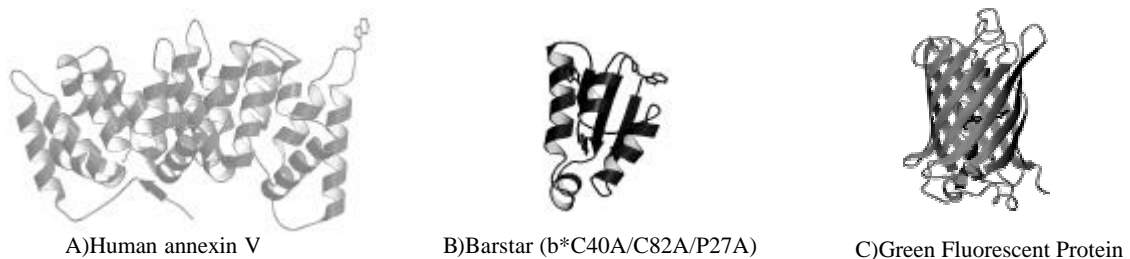
Figure 6. a) Human recombinant Annexin V (AxV) **b)** Pseudo wild-type Barstar (b*) **c)** Green Fluorescent Protein (GFP)

7.2 Barstar

Barstar (Figure 6b), a small bacterial ribonuclease inhibitor is an excellent model protein for incorporation studies since it meets all the criteria mentioned above : both high-resolution crystal and solution structures are known (Martin, *et al.*, 1999); it undergoes completely reversible unfolding either upon thermal or chemical denaturation. Therefore, it is frequently used as model protein for protein folding, stability and dynamic studies (Nath & Udgaonkar, 1997). Barstar

contains three Trp residues at positions 38, 44, and 53. In present study, the replacement experiments were performed using the pseudo-wt mutant C40A/C82A/P27A/W38F (barstar, b*) which is expressed in form of inclusion bodies and is refolded into the native form during purification procedure (Golbik, *et al.*, 1999). This mutant (barstar, b*) contains only two Trp residues: W44 which is partially buried and W53 completely buried in the protein hydrophobic core (Figure 6b). The latter Trp residue is essential for the protein conformational integrity as it cannot be replaced by any of 20 canonical amino acids by routine DNA mutagenesis (Nath & Ugaonkar, 1997).

7.3 Green Fluorescent Proteins

The basic property of Green Fluorescent Protein (GFP; Figure 6c) from jellyfish *Aequorea victora* and its variants is that the chromophore (4-(p-hydroxybenzilidene)imidazolid-5-one) is completely coded in its amino acid sequence. It is autocatalytically formed in the post-translational reaction between side chains of residues 65-67 and only external requirement is the presence of oxygen. Such resonantly stabilised π-conjugated system through its dynamic interactions with the molecular environment allows for visible absorption and characteristic green fluorescence of this protein. This unique fluorescence is basically provided by specific protein molecular architecture consisting of 11 ß-sheet barrel like structure forming the walls of can and an a-helix which runs diagonally through the can. The chromophore is in the centre of this structure, enclosed in the can and completely protected from environment. The loss of fluorescence in the denatured GFP indicates that only properly folded protein provides a milieu for its unique optical properties.

Although the detailed mechanism for the formation of the chromophore in GFP is still not experimentally demonstrated, the mature form is well known and structurally characterised. The

production of the novel variants of GFP chromophore is straightforward since residues in the positions 65-67 can be replaced using site directed mutagenesis by any other from remaining 19 canonical amino acids. Currently available GFP variants are exclusively produced by classical protein engineering. Taking their chromophore photophysics as criteria, Tsien divided all known native and engineered GFPs into seven classes (Tsien, 1998). Class I include wt-GFP (Ser65-Tyr66-Gly67) consisting of equilibrated mixture between the phenol and phenolate ion. Class II consists of mutants (Ser65 replaced with Thr, Ala or Gly) with structure containing exclusively phenolate ion (e.g. commercially available "enhanced green fluorescent protein" EGFP; Phe64Leu/Ser65Thr) while class III has neutral phenol exclusively (e.g. Thr203Ile). The mutants belonging to class IV was rationally designed from the available crystal structures with expectation that introduction of polar residues and induction of π-π interaction would reduce excited state energy and subsequently increase excitation and emission wavelengths. Indeed Thr203Tyr mutant meets all these requirements and is found to be the best mutant for various dynamic and tracing studies (e.g. commercially available "enhanced yellow fluorescent protein" EYFP; Ser65Gly/Val68Leu/Ser72Ala/ Thr203Tyr). The mutation Tyr66Trp represent class V with a characteristic blue-green emission known as "enhanced cyan fluorescent protein" (ECFP; Ser65Thr/Tyr66Trp) where an indole ring in the chromophore extents toward Val150 and Phe156 and its accommodation is probably favoured with other compensatory mutations (Phe64Leu/Asn146Ile/Met153Thr/Val163Ala). Classes VI and VII include chromophore phenol moiety replacement with other two available aromatic canonical amino acids: His (Tyr66His; "Blue" fluorescent protein) and Phe (Tyr66Phe) with lower fluorescence excitation and emission values. Most of the mutants belonging to these GFP classes (schematically represented in Figure 7) contain only one Trp residue in their amino acid sequences (Trp57) with the exception of class V ("cyan"; fluorescent proteins) which have Trp66 as integral part of the chromophore in the

protein mature form. As suitable models for Trp-analogues incorporation the EGFP (Class II)
with Trp57 (outside from chromophore) and ECFP (ClassV) with indole moiety in the
chromopore were used.
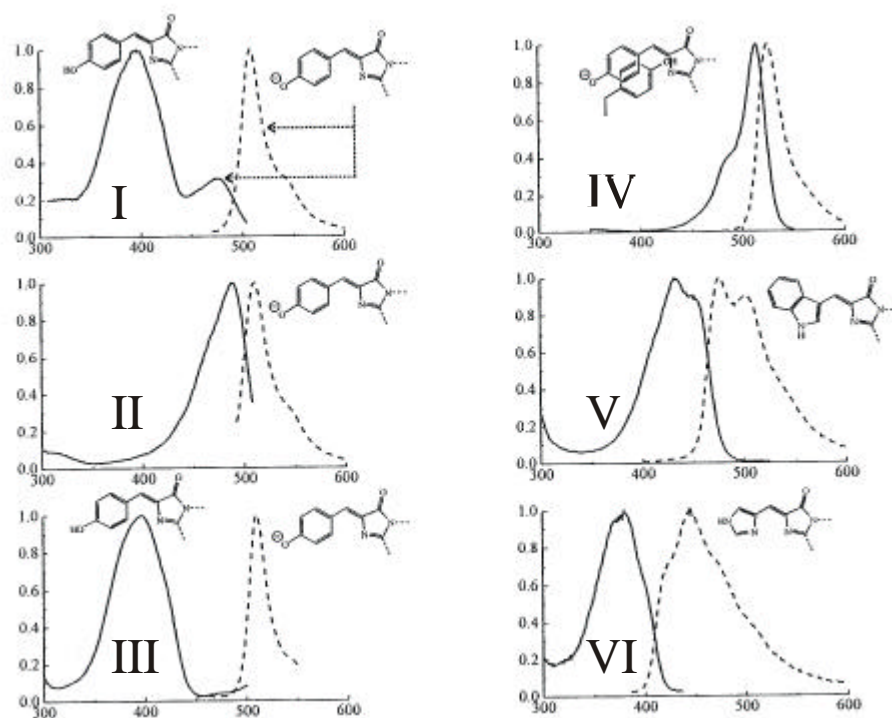
Figure 7. Main spectral classes of GFP Proteins (Tsien, 1998). Class **I** (e.g. wt-GFP) has excitation values at 395 and
475 nm (Phenol- and Phenolate-form). In Class **II** (e.g. EGFP) Phenol-form is missing; in Class **III** (e.g "Saphir")
the Phenolate-form dominates. In class **IV** (Thr203Tyr mutants, e.g. EYFP) is spectrum through π-Stacking of
Tyr203 with Chromophore red-shifted. Class **V** has an Indole moiety in Chromophore (e.g. ECFP), while Class **VI**
has Imidazole in Chromophore (For example "Blue Fluorescent Protein" EBFP). Finally Class **VII** (not shown here)
that has phenyl-moiety in chromophore (Tyr66Phe) is normally not interesting for practical applications due to large
blue-shift in fluorescence (Figure taken from Tsien, 1998).

8   Motivation for Structural Studies of Proteins substituted with Tryptophan Analogues

Green fluorescent protein from jellyfish *Aequoria victora* has recently become a standard reporter
in cellular and molecular biology as well as a model for studies of chromophore photophysics.
The placement of any aromatic amino acid in the position 66 is crucial for the emergence of a
spectrally active GFP chromophore in the context of the intact protein matrix. However, in order

to gain GFPs with an altered spectral window by chromophore redesign with aromatic amino acids, the standard genetic code limits their number to Trp, Tyr, His and Phe canonical residues. It should be therefore not surprising that despite years of effort using classical protein engineering approaches, no significantly red-shifted variants of *av*GFP were found (i.e. not beyond 530 nm). In other words, currently available GFP variants are exclusively produced by classical protein engineering procedures. Among them, "yellow" class (Class IV) of GFPs (for example enhanced yellow fluorescent protein - EYFP) is found to be most red shifted mutant designed by conventional mutagenesis methods to date (λ_{em} = 527 nm).
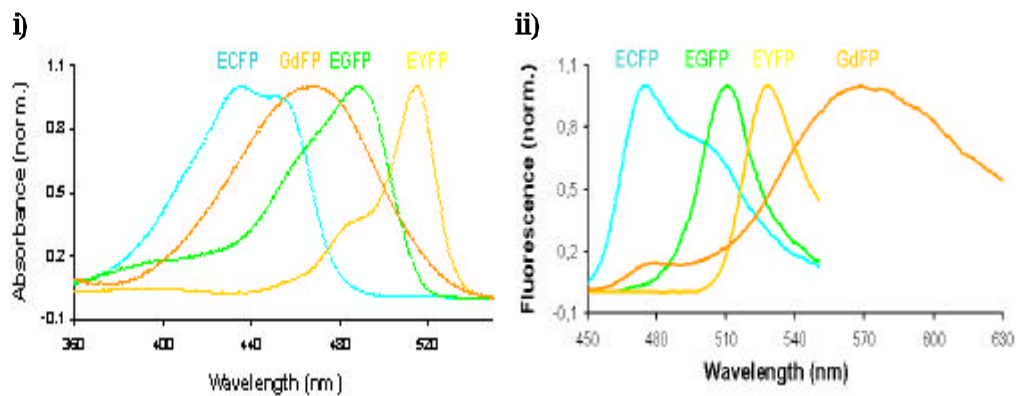
Figure 8. Breaking through the limits of classical GFP engineering: Normalised absorption **(i)** and fluorescence **(ii)** emission spectra for three typical members of major classes of GFP mutants compared with GdFP. Note a large Stokes shift of GdFP (about 100 nm).

Recently in our laboratory a breakthrough of these limits is demonstrated by introduction of electron-donating amino-Tryprophan into "enhanced cyan fluorescent protein", ECFP. The resulting "gold" fluorescent protein, GdFP, is the most red shifted *av*GFP-variant known to date (69 nm when compared to parent ECFP and 54 nm when compared with "enhanced yellow fluorescent protein", EYFP; Figure 8). GdFP has not only different colour but also dramatically increased thermostability, and stability of monomer state over a long period of time.

Normally, Tyr66 mutants derived from *Aquorea* wt-GFP have only one Trp residue which exhibits no trytophan fluorescence. Available tertiary structures reveal relative rigidity (i.e.

immobility) of both Trp57 and chromophore which are separated by about 15 Å. Total quenching of intrinsic Trp-fluorescence in the native GFP is probably due to the fluorescence resonance energy transfer between these two chromophores. In this context, the question arises as whether the transformation from "cyan" to "gold" can be attributed solely to replacement of the Trp at position 66 with electron-donating amino-Trp or to the both Trp residues (Trp57 and Trp66) of ECFP. The proof can be derived using EGFP as test system with only one Trp-residue (Trp57). $(4-NH_2)$-Trp incorporation into EGFP results in substituted protein $((4-NH_2)$-Trp-EGFP) with spectral properties indistinguishable from parent protein EGFP (Figure 9), clearly indicating that amino-group at Trp57 does not affect spectral properties of the protein chromophore.

Finally, the practical applications of these experimental results are easily conceivable. With the GFP mutants available to date, only a pair of two variants can be excited with one colour and detected at two different wavelengths. Figure 10, where three populations of Trp-auxotrophic *E. coli ATCC49980* strain expressing ECFP, EGFP, and EGdFP respectively, clearly shows that tailor-made variants like EGdFP employing non-canonical amino acids open new ways for multiple labelling applications and design of genetically encodable, physiological indicators.

These experiments and observations made in our laboratory represented a basic motivation to elucidate three-dimensional structures of GdFP and its parent protein ECFP. Since they are expressed and purified using Trp-auxotrophic *E. coli* strain *ATCC 49980* in relatively large quantities ($\approx$30 mg per liter of the culture), this enabled successful crystallisation, structure elucidation, thus providing insights into the nature of the chromophore dynamics. The most intriguing question was if there is any relation between their unique spectral properties of GdFP and ECFP and their structures, especially structural configurations around the chromophores. In order to investigate possible structural bases for these novel and spectacular optical properties of

"gold" fluorescent protein (GdFP), its crystallisation and structure determination (together with its parent protein, ECFP) was performed.
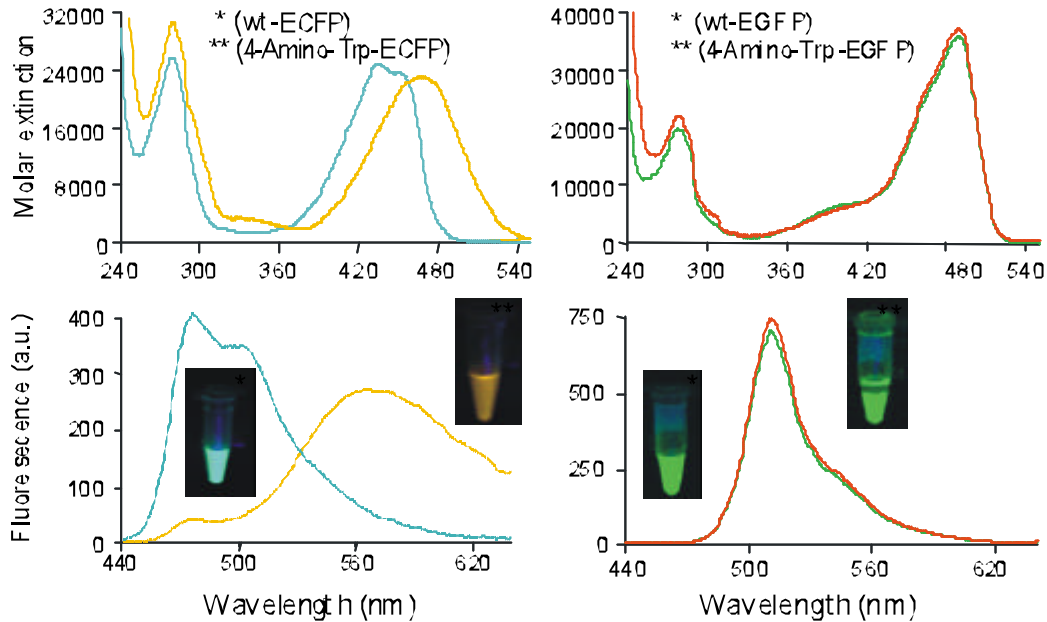
Figure 9. Role of Trp57 for the unique spectral properties of GdFP. Spectral properties of EGFP when single Trp57 is substituted with (4-NH$_2$)-Trp are unchanged. Conversely, in ECFP Trp57 and chromophore Trp66 replacement with (4-NH$_2$)-Trp yielded the novel protein variant GdFP with golden colour, clearly indicating that the colour conversion is induced by replacement at position 66. All proteins (2 μM) were excited at 305 nm and fluorescence was measured by using excitation/emission slits of 5.0 nm.
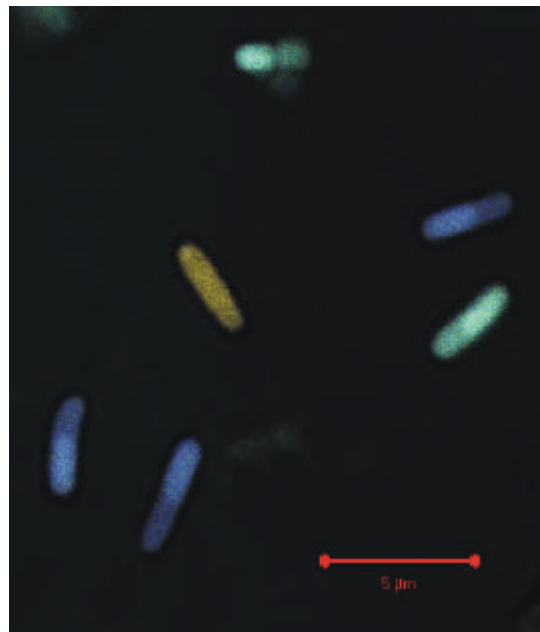
Figure 10. Practical potentials of novel "gold" class of GFPs. Three populations of Trp-auxotrophic *E. coli ATCC49980* strain expressing ECFP, EGFP, and GdFP are simultaneously excited at 457 nm and imaged on a Zeiss LSM 510 confocal microscope using standard emission filters (Courtesy of Dr. A. Zumbusch).

MATERIALS AND METHODS

1 Chemical Methods and Materials

1.1 Solvents and reagents

The solvents used were from Aldrich, Fluka and Merck. When necessary, these solvents were purified through standards procedures to obtain the desired quality. The acetonitrile for HPLC was LiChrosolv Acetonitrile from Merck. The reagents used were from Sigma Aldrich, Fluka, Langcaster or Merck. The selenophene-3-carboxylic acid (See *Appendix*) was purchased from Specs and Bio Specs (The Netherlands).

1.2 Analytical HPLC

The analytical HPLC was a Waters system assembly composed of two 515 HPLC pumps, a Waters pump control Module, a 717 plus Autosampler, a 996 Photodiode array detector. The software connected to the system was the Program Millennium version 3.00 supplied by Waters. The column used were the ET 125/4 Nucleosil 100-5 C_8 from Macherey-Nagel GmbH & Co KG(Germany), X-Terra-TM-C_8 5µm 3.9*150mm from Waters Xterra Columns (USA), and Chromolith C_{18} 100*4.6mm from Merck KGaA (Darmstadt, Germany).

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 13 min with ET 125/4 Nucleosil 100-5 C_8

Gradient B: 0 % Eluent B → 100 % Eluent B in 15 min with X-Terra-TM-C_8 5µm 3.9*150mm

Gradient C: 0 % Eluent B ➔ 100 % Eluent B in 13 min with ET 125/4 Nucleosil 100-5 C_8

Gradient D: 0 % Eluent B ➔ 100 % Eluent B in 6 min with Chromolith C_{18} 100*4.6mm

Gradient E: 0 % Eluent B ➔ 100 % Eluent B in 15 min with X-Terra-TM-C_8 5µm 3.9*150mm

1.3 Mass Spectro metry

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 auto-sampler 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 10µl dissolved in Acetonitrile. The Eluent was 100% Acetonitrile (0.05% TFA).

1.4 UV Spectroscopy

The UV spectra were taken on an UV/VIS spectrometer Lambda 19 (Perkin Elmer) connected with the software UV WinLab Version 2.0. The spectra were detected between 240 and 600 nm, with slit of 1.00 nm, a data interval of 0.1 nm and a scan speed of 240 or 480 nm/min.

All the cuvettes used for measurements were of quartz glass of Hellma GmbH (Baden, Germany). Protein samples were always measured so to give a absorption signal in the range of 0.2 ~ 0.8 at 280 nm (linear region in Beer-Lambert law).

1.5 Fluorescence Spectroscopy

The fluorescence emission spectra were recorded on Perkin Elmer Lambda 19 fluorescence spectrometer at protein concentrations of 2 µM in the aqueous buffered solutions. Several excitation wavelengths were used in all experiments (Table 2), and the emission was detected in

the range of 300 – 600 nm or 300 – 450 nm (GFP variants or AxV (and barstar), respectively).

All spectroscopic investigations were performed with filtered solutions (pore size 0.2 μm) at 25 °C.

Protein	Excitation Wavelength (l_{exc}, nm)
EGFP	305, 488
ECFP	305, 452
GdFP	305, 472
AxV	280, 295
Barstar (b*)	280, 295

Table 2. Fluorescence excitation (l_{exc}) values at 25 °C for native and mutant proteins.

1.6 Circular Dichroism

The circular dichroism (CD) spectra were recorded on a Jasco J-715 spectropolarimeter with temperature controller 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 50 nm/min, the response of 1 sec, the band width of 1.0 nm, and the average of 10 scans was reported. The spectra were all normalised and the ellipticity was expressed as mean residue molar ellipticity $[\Theta]_R$ (deg cm^2 dmol^{-1}), $[\Theta]_R = \Theta / l \cdot c \cdot n_R$, 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/hr, the response of 16 sec, 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

15 degrees in order to monitor the conformational changes. At the end of the melting curve the temperature was lowered at the initial value, 5 °C and a CD spectrum was taken then to verify the reversibility of the process.

1.7 Nuclear Magnetic Resonance spectroscopy

The ^{1}H-NMR measurement was performed on a Bruker DRX400 spectrometer. All compounds were dissolved in CD_3OD or CD_3Cl. The NMR spectra were recorded in the range of 1 to 10 ppm.

1.8 Chemical Syntheses of Tryptophan Analogues

The trytophan analogues used in the present study were essentially made previously published procedures (Welch & Phillips, 1999; Figure 11). The detailed description of the syntheses are reported in *Appendix*.

2 Biological Methods

2.1 Gene Cloning and Transformation

2.1.1 AxV and Barstar

T5 promotor-based expression system (Qiagen) inducible with IPTG was used for protein expressions. The expression host *E. coli ATCC49980* was routinely co-transformed with two plasmids: ampicillin resistant pQE-60-PP4 harbouring the AxV gene sequence (*NcoI-HindII* fragment), and kanamycin resistant pREP4 containing a repressor gene *lacI^q*. The plasmid with the gene for Barstar was constructed as follows: *EcoRI-HindIII* fragment from pKK223-3

plasmid (Pharmacia) containing ribosome binding site and barstar (C40A/C82A/P27A/W38F) DNA sequence was inserted into pQE-30 vector (Qiagen) resulting in pQIA-30b* which is together with pREP4 transformed into the expression host.

2.1.2 EGFP and ECFP

The genes encoding for EGFP and ECFP (Invitrogen), were inserted in a pQE80 vector (Qiagen) harbouring Nt-His-Tagged gene sequence, and expressed in tryptophan auxotrophic *E. coli ATTC49980* strain which served as expression host.

2.1.3 Trp-Synthase

The *E coli* strain, *CB149*, containing the vectors (pSTB7, pEBA10) for Trp- synthase was provided by generous help of Prof. E. Wilson-Milles.

2.2 Fermentation, Expression, and Incorporation Experiments

In all fermentation experiments, the expression host, *E. coli ATCC49980*, was grown to reach the values of cellular density corresponding to the mid-logarithmic phase of the normal growth. This is the optimal point for efficient expression and concomitant high-level amino acid replacement in the target protein. The experimentally determined limiting concentration essential for successful incorporation of a non-canonical amino acid was 0.015 mM L-Trp which led to Trp depletion in the culture medium in the mid-logarithmic phase of the growth (OD_{600}: 0.7-1.0). The cells were grown in minimal medium (NMM) with 0.015 mM L-Trp as natural substrate until its exhaustion, followed by simultaneous addition of a non-canonical amino acid substrate and by target-gene induction with isopropyl-β-D-thiogalactopyranoside (IPTG). Under these conditions optimal expression of the labelled proteins was achieved.

Light sensitive and chemically unstable Trp analogues, [3,2]Sep, and [2,3]Sep, when protected from light proved to be stable, and despite exposure to aerobic conditions a stronger degradation of these amino acids were not observed.

2.3 Protein Purifications

2.3.1 Purification of Trp-Synthase

Trp-synthase for the enzymatic reaction to make Trp-like analogues was purified according to the protocol of Miles and co-workers (Miles, *et al.*, 1989). Its activity was checked with 100 μl of 0.1 M L-serine, 100 μl of 0.1 M indole, and 10 mM PLP in 2.62 ml 0.1 M Tris pH 7.8. The protein was stored in 50 mM Tris buffer (pH 7.8) at +4°C.

2.3.2 Purification of AxV with [3,2]Tpa, [2,3]Tpa, and [3,2]Sep

The mutant protein of AxV was purified as the wild type forms (Budisa, *et al.*, 1995). Bacterial cells were opened by an osmotic shock and the protein was bound reversibly to liposomes in the presence of calcium. After release from liposomes by EDTA, the protein was further purified by ion-exchange chromatography on a DEAE-Sepharose column buffered with 20 mM BisTris, pH 6.0, and eluted with a linear NaCl gradient (0-200 mM). The yield was 2-10 mg protein/L medium. The purity of the recombinant proteins was checked by SDS-PAGE (Commassie and silver staining) and HPLC-profile analyses.

2.3.3 Purification of Barstar with [3,2]Tpa, [2,3]Tpa, and [3,2]Sep

The mutant protein of barstar was purified as the wild type forms (Golbik, *et al.*, 1999). The protein was overexpressed in *E. coli ATCC49980* as inclusion bodies. Cells were opened by

lysozyme digestion and sonication. After centrifugation at 15,000 rpm for 40 min the pellet was dissolved in 50 mM Tris-HCl, pH 8.0 containing 7.5 M urea. The suspension was centrifuged at 9,000 rpm for 40 min and the supernatant dialysed three times against 4 L, 50 mM Tris-HCl, 0.1 M NaCl, pH 8.0. The cloudy solution was centrifuged at 15,000 rpm for 40 min, and the supernatant containing refolded protein was applied to a HiLoad^{TM} 26/10 Q Sepharose HP (Pharmacia, Piscataway, New Jersey) anion exchange column on a FPLC. Barstar was eluted with a linear NaCl gradient at about the concentration of 300 mM. Further purification was performed by gel filtration on HiLoad^{TM} 26/60 Superdex 75 (Parmacia) with 50 mM Tris-HCl buffer, pH 8.0. The purity of the recombinant proteins was checked by SDS-PAGE (Commassie and silver staining) and HPLC-profile analyses.

2.3.4 Purification of EGFP, ECFP, and GdFP

Native, mutant and variant and GFPs were purified through two columns (i) Ni-NTA Agarose (Qiagen) via imidazole gradient (0 – 100 mM) in 100 mM Na-phosphate buffer pH 8.0 and 0.5 M NaCl, and (ii) phenyl-sepharose (Pharmacia) with ammonium sulphate gradient 20 – 0% in 20 mM TrisCl pH 8.0 and 1 mM EDTA.

3 X-ray Crystallography

3.1 Crystallisation conditions of Model Proteins

3.1.1 Crystallisations of AxV and Barstar with [3,2]Sep

Crystallisation of [3,2]Sep-AxV was achieved in hanging drops at room temperature by vapour diffusion against 2.0 M ammonium sulphate buffered with 0.1 M Tris/HCl (pH 8.5). The protein

was dialysed against 10 mM Tris/HCl buffer (pH 8.5) and concentrated to 10 - 20 mg/ml. 10 µl protein solution was mixed with: 1 µl 20 mM CaCl₂ solution and 1.5 µl of 2.0 M buffered ammonium sulphate precipitant. This solution was supplemented with agarose to a final concentration of 0.1 % or 0.05 % according to the procedure described elsewhere (Robert, *et al.*, 1991). Preparation of the drops in agarose was achieved in three steps as follows. First, 2 % agarose was prepared in distilled water; second, the molten agarose was mixed with reservoir puffer; e.g. for preparation of 10 ml of precipitant in 0.2 % agarose it was necessary to mix 3 ml reservoir puffer, 2 ml 20 mM Ca^{2+}, 4 ml water and 1 ml 2 % agarose. This mixture was liquated and stored at 42 °C. Finally, 10 µl of [3,2]Sep-AxV (10-20 mg/ml) was mixed with 10 µl of precipitant in 0.2 % agarose at 42 °C and the resulting hanging drops (with final concentration of agarose of 0.1 %) were left to cool down. Crystals grew after 5-9 days and were harvested into 3 M ammonium sulphate, 0.1 M Tris/HCl (pH 8.5) and 1 mM CaCl₂.

For barstar, crystallisation was achieved using the sitting drop method at room temperature (18 – 20 °C) in "Cryschem" plates from Hampton Research Ltd. The protein crystals appeared from the precipitated protein solution (2.5 mg/ml) in 3-5 days in 0.9-1.1 M sodium citrate buffered with 50 mM Tris/HCl buffer, pH 8.0. These crystallisation conditions are similar to those of cysteine -free barstar and its different mutants (Martin, *et al.*, 1999).

3.1.2   Crystallisations of ECFP and GdFP

The green fluorescent protein variants, ECFP and GdFP, have been crystallised in the conditions of 14% (w/v) polyethylene glycol (PEG) 1000 within 1 or 2 days, and 0.1 M HEPES pH 7.5, 5% PEG 8000, 10% ethylene glycol. For the crystallisation, the hanging drops were made of 1 µl, or 1.5 µl of protein solutions with the concentration of 20 mg/ml and 1 µl of precipitant solution at 20°C. Drops were equilibrated against 0.2 ml of precipitant solution.

3.1.3   Crystallisations of Thieno-, and Selenolo-Trp analogues EGFP variants

EGFP variants containing [3,2]Tpa, [3,2]Sep, and [2,3]Sep were crystallised in several conditions, such as 12 % (w/v) PEG 20000 in 0.1 M MES pH 6.5. The hanging drop experiment was performed to get crystals. 1 µl of protein solution (with the concentration of 10 mg/ml) was mixed with the same volume of precipitant solution at $20^{°}$C. Drops were equilibrated against 0.2 ml of precipitant solution for 24 hours.

3.2   X-ray data collection and Structure refinement of Model Proteins

3.2.1 AxV and Barstar with [3,2]Sep

Diffraction data for [3,2]Sep-annexinV up to 2.3 Å resolution were collected from a single crystal on an X-ray image plate system (Mar Research, Hamburg, Germany) using CuK_{α}-radiation generated by a Rigaku rotating anode operated at 5.4 kW in a thermo stated room at 18 °C. Since crystals of barstar, suffered from X-ray exposure, they were transferred to their mother solution containing 20% glycerol as a cryoprotectant and shock-frozen in a nitrogen stream of 100 K. A complete data set for the for [3,2]Sep-barstar, was recorded from a single crystal under a nitrogen stream at 100 K on an image plate system (Mar Research, Hamburg) to a resolution of 2.4 Å. Reflections were integrated with the program MOSFLM (Lesile, *et al.*, 1998), scaled and reduced using SCALA of the CCP4 package (Collaborative computational project, 1994). Model building was performed with MAIN (Turk, 1992) and refinement with CNS (Brünger, *et al.*, 1998) after initial rigid body minimisation, positional refinement and simulated annealing was alternated with electron density map inspections for several cycles. The atom replacements were determined by difference Fourier methods.

3.2.2 GFP Variants

The structures of GFP variants were solved by the molecular replacement technique. As the platform for the molecular replacement, one of the GFP structures (EGFP, 1EMG) was used. Data sets of GFPs were collected with the resolution range between 1.15 Å and 2.9 Å. X-ray diffraction data of one of ECFP crystals and the GdFP crystal were collected at DESY (Germany), and the other ECFP, [3,2]TpaGFP, [3,2]SepGFP, and [2,3]SepGFP crystals were collected on an X-ray image plate system (Mar Research, Hamburg, Germany) using CuKα-radiation generated by a Rigaku rotating anode at 5.4kW in a cryo protected condition at 100°K. All crystals were transferred to their mother solution containing 20% (w/v) meso-erythritol as a cryoprotectant and shock-frozen in a nitrogen stream.

Reflections were integrated with the program DENZO, scaled and reduced using SCALEPACK (Otwinoski & Minor, 1997). Model building and refinement were performed with CNS (Brünger, *et al.*, 1998). The initial model was refined alternating automatic minimisation protocols performed with CNS inspecting visual electron density map and manually adjusted using the program O (Jones & Kjelgaard, 1991).

RESULTS

1 Chemical Syntheses of Indole Analogues

1.1 Standard Methods

The overall strategy applied for the synthesis of indole analogues follows essentially previously published (Phillips, *et al*, 1999; Figure 11, 37). Modifications of these procedures resulted in increased product yields of the whole steps in the range from 60 % to almost 100 % (see *Appendix*).
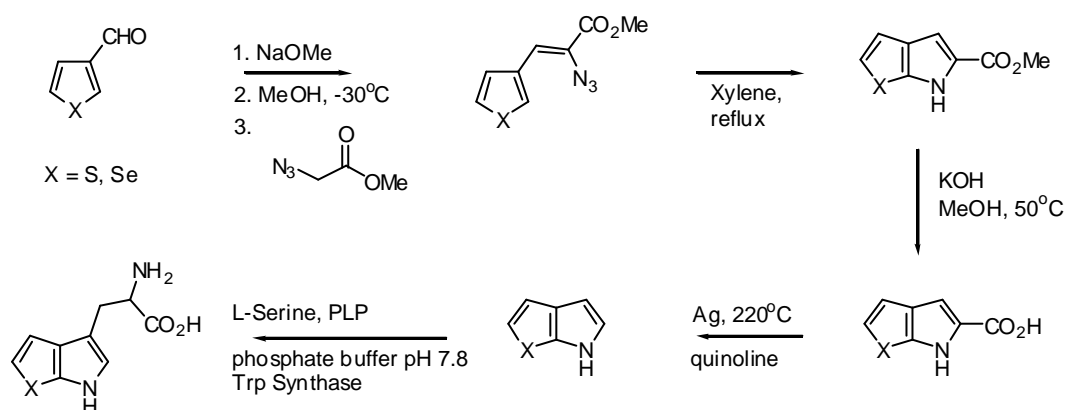
Figure 11. General synthetic route to ß-[2,3]pyrrolealanines.

Starting compounds for the synthesis of [3,2]Tpa, and [2,3]Tpa, i.e. thiophene-2-carboxaldehyde and thiophene-3-carboxaldehyde, respectively, were commercially available, while the starting materials for the synthesis of [3,2]Sep and [2,3]Sep, i.e. selenophene-2-carboxaldehyde and selenophene-3-carboxaldehyde, were synthesised from selenophene. Transformation of selenophene into selenophene-2-carboxaldehyde proceeded straight forwardly. Conversely, there is no direct route for the synthesis of selenophene-3-carboxaldehyde due to the reactivity of the

position 3 of the ring. To solve this problem, novel synthetic approaches were explored (detailed descriptions are presented in the next section).

All the final compounds, β-pyrrole-alanines, were not purified and characterised due to their instability. They were used as reaction mixture directly in the fermentation for the incorporation experiments to obtain the desired labelled proteins. These mixtures did not affect significantly the viability of the expressing cells in culture, nor did they have influence on the analogue-incorporation and protein expression (Figure 14, 15).

1.2 New Synthetic Route to Selenophene-3-carboxaldehyde
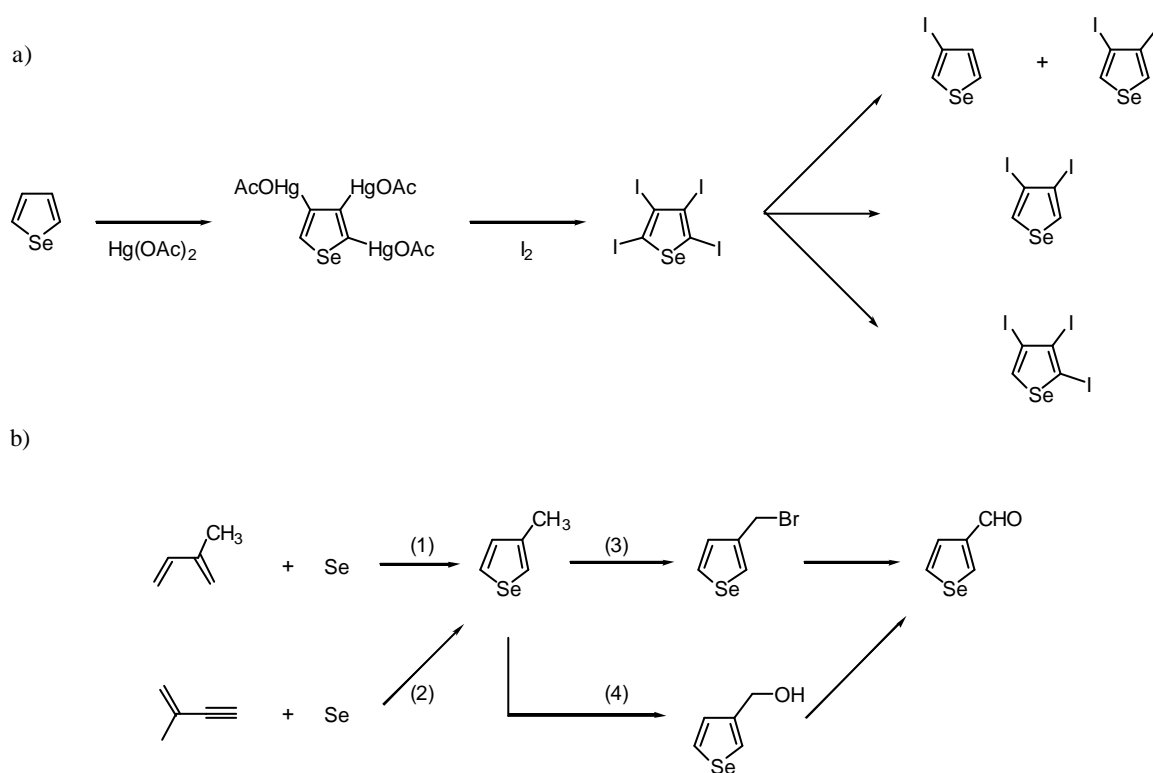
Figure 12. Synthetic approaches for selenophene-3-carboxaldehyde. **a)** Classical route to the desired compounds by Paulmier. **b)** Synthesis via 3-methylselenophene. [(1) Barbey, *et al*., 1989, (2) Kulik, *et al*., 1983, (3) Gronowitz, *et al*., 1976, (4) Grissom, *et al*., 1994]

For the synthesis of selenophene-3-carboxaldehyde as educt for the preparation of ß-3-selenolo[2,3-b]pyrrolyl-L-alanine several routes were investigated. At first, the protocol reported by Phillips (Phillips, *et al*, 1999) on the basis of the Paulmier approach (Paulmier, *et al*, 1967) was applied. Following this procedure, a mixture of 3-Iodoselenophene and 3,4-diiodoselenophene was obtained unacceptable low yield (ca. 10%), which in addition required separation (Figure 12a). For this reason alternative synthetic routes were explored.

In this context procedures for the synthesis of selenophene-3-carboxaldehyde via 3-methylselenophene (Figure 12b) reported Barbey and Kulik (Barbey, *et al*, 1989; Kulik, *et al*, 1983) were examined. Indeed, 3-methylselenophene could be prepared good yields (ca. 80%), but the subsequent side-chain bromination to 3-(bromomethyl)-selenophene failed, as already reported by other laboratories (Gronowitz, *et al*, 1976).

Primary alcohols are readily converted into aldehyde (Marx, *et al*, 1984). Therefore transformation of 3-methylselenophene into 3-(hydroxymethyl)-selenophene was attempted with several different methods (Grissom, *et al*, 1994), but all failed.
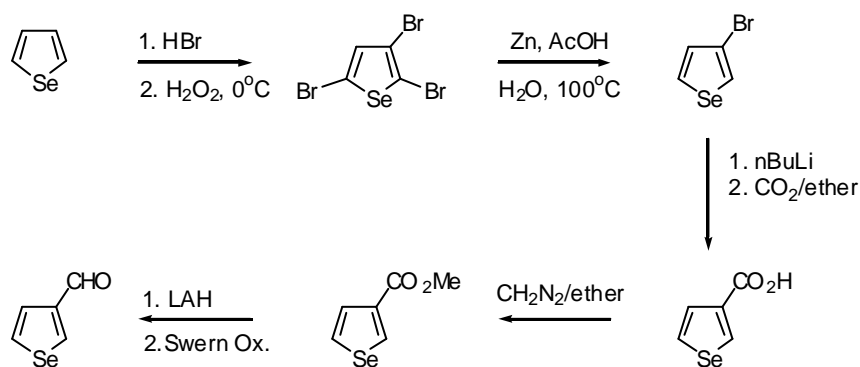
Figure 13. The novel rote for the synthesis of selenophene-3-carboxyaldehyde from selenophene.

Finally a new synthetic route to 3-(hydroxymethyl)-selenophene was elaborated (Figure 13).

Multibromination of selenophene with HBr led to 2,3,5-tribromoselenophene (Keegstra, *et al*,

1988), without isolation, the crude product of the multibromination reaction was refluxed with

acetic acid and zinc powder to yield mixture containing predominatly 3-bromoselenophene

(Hallberg, *et al*, 1981). This mixture was carefully vacuum-distilled to isolate the desired

product, 3-bromoselenophene in the yield of 56% over the two steps.

The carboxylation reaction was performed to obtain selenophene-3-carboxylic acid (Gronowitz,

et al, 1972), which was esterified (Hedlicky, 1971), reduced with lithium aluminium hydride

(LAH), fianlly Swern oxidation (Marx, *et al*, 1984) proceeded the selenophene-3-

carboxaldehyde.

The subsequent steps for the synthesis of ß-3-selenolo[2,3-b]pyrrolyl-L-alanine were identical to

the protocol of Phillips (Phillips, *et al*, 1999).

2   [3,2]Tpa and [2,3]Tpa in Proteins

2.1 Fermentation and Expression of Proteins with [3,2]Tpa and [2,3]Tpa

The *E. coli ATCC49980* strain exhibits a typical growth curve in the minimal medium with 0.03

mM tryptophan. The optimal limiting concentration essential for successful incorporation of the

both β-(thienopyrrolyl)alanines is 0.015 mM Trp. This concentration leads to Trp depletion in

the culture in the mid-logarithmic phase of the growth (OD_{600}: 0.7-1.0), a fact that is almost ideal

for induction of protein expression. Thus, the cells were grown in the minimal medium (NMM)

with 0.015 mM Trp as natural substrate (Minks, et al., 1999) until its exhaustion, followed by

simultaneous addition of the β-(thienopyrrolyl)-alanines and target gene induction with IPTG.

Under these conditions, optimal production of the labelled proteins was achieved in yields comparable with those of the wt-proteins (10-30 mg/L in the case of AxV and 50-100 mg/L in the case of b*).
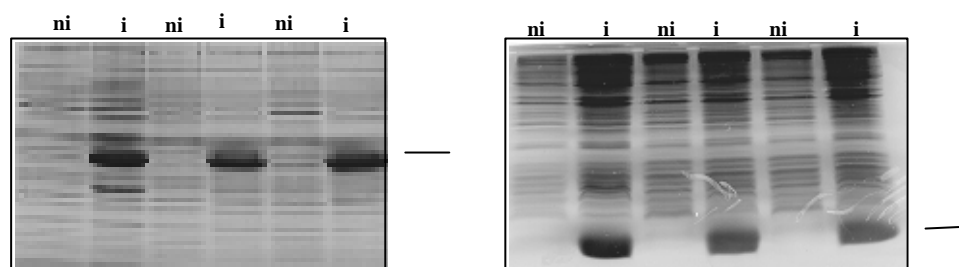
Figure 14. Expression of model proteins by SPI-method. Analysis of the expression profiles in cell lysates of *E. coli ATCC49980* with over-expressed AxV (left) and barstar (right) in the defined minimal medium with the Trp, [2,3]Tpa and [3,2]Tpa, respectively. *ni*- non induced cells, *i* – cells induced for protein expression in the presence of Trp, [2,3]Tpa and [3,2]Tpa. Arrows indicate position of over-expressed substituted proteins (left: AxV, right: barstar).

2.2 Analytical Characterisation and Spectroscopic Properties of the Protein Mutants

In routine bio-expression protocols based on the SPI method, the almost quantitative incorporation of both β-(thienopyrrolyl)-alanines in AxV and barstar was readily achieved (Figure 14). To check incorporation of these amino acids, the fluorescence profiles of the protein mutants were recorded, since replacement of the benzene ring of the indole moiety with thiophene provides an efficient fluorescence quenching. Thus, the absence of the characteristic Trp emission of fluorescence profile was used as qualitative analytical criteria to monitor successful labelling.

Conversely, for quantitative analysis mass spectrometry was applied, since the molecular mass difference between Trp and both thia-variants is sufficiently large (6 Da) to be determined experimentally. AxV (expected mass: 35809.2 Da; experimental value: 35808 ± 2.0 Da) contains only one Trp residue which upon replacement with the β-(thienopyrrolyl)alanines leads to a slight increase in the protein molecular mass (expected mass: 35815.2 Da; experimental

value: 35814 ± 1.0 Da for [3,2]Tpa-AxV and 35815 ± 2.0 Da for [2,3]Tpa-AxV). The barstar

mutant (expected mass: 10214 Da; experimental value: 10214 ± 3.2 Da) contains two Trp

residues, thus the mass differences upon substitution are much more suitable (expected mass:

10226.2 Da; experimental value: 10226 ± 2.0 Da for both protein variants; Figure 15). These

data confirm also that thia-analogues once incorporated into proteins are stable. This is

especially pronounced in the case of barstar which is refolded under air-oxygen prior to its

purification.
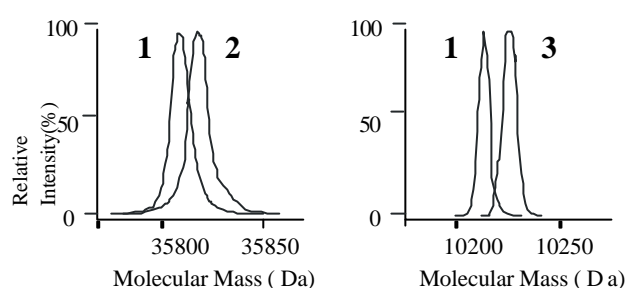
Figure 15. Mass spectrometric profiles of substituted proteins. Two different electrospray mass-spectrometric measurements were superimposed at the same mass-scale (left: native and [2,3]Tpa-AxV; right: wt-, and [3,2]Tpa-b*).

As expected from the UV spectra of the two β-(thienopyrrolyl)-alanines (Table 3, Figure 16A),

the optical properties of the protein mutants are affected by their incorporation at the Trp-

positions. The UV-spectra of the AxV mutants are not dramatically changed due to dominant

contribution of the Tyr residues to the overall absorbance (Figure 16C). Conversely in the case

of the barstar mutants, the UV spectral properties of the β-(thienopyrrolyl)-alanines are fully

reflected in the spectra of the mutant proteins (Figure 16B) with a blue shift of the absorption

maximum (Table 3). Such blue shift originates from the larger number of tryptophans that are

presented in the structures (dominates over tyrosines).
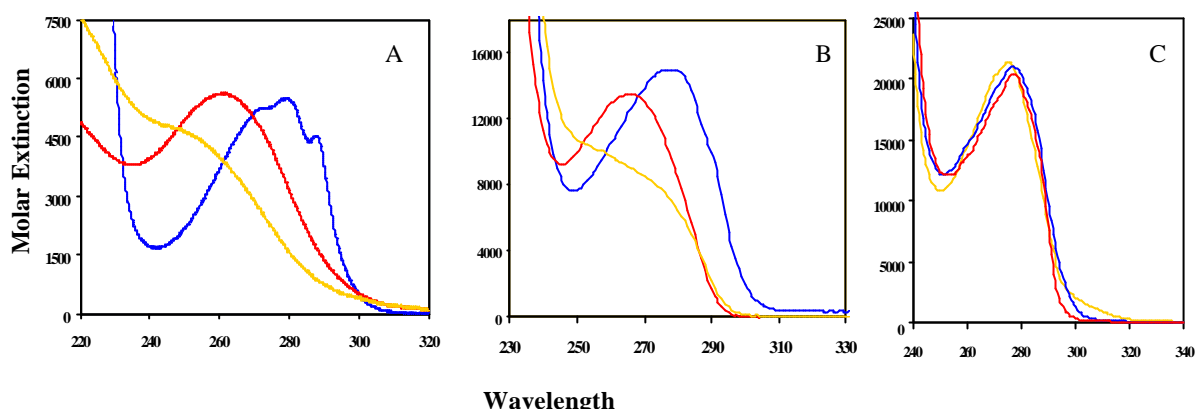
Figure 16. Ultraviolet-absorbtion (UV) spectra of (**A**) free amino acids Trp and its thia-surrogates and of native proteins and related mutants. (**B**: b*, **C**: AxV) at the 20 °C (blue: Trp and parent proteins; yellow: [2,3]Tpa and [2,3]Tpa-proteins; red: [3,2]Tpa and [3,2]Tpa-proteins).

2.3 Conformational Analysis of the Protein Mutants in Solution

The far-UV CD spectra of AxV and its mutants exhibit the typical pattern of a-helical proteins with the two characteristic minima at 222 nm and at 208 nm of similar intensity (Figure 17A insert). The signal ratio between these two minima $([\theta]_R)_{222}/([\theta]_R)_{208}$ is about 0.96 for the wt-protein and the [2,3]Tpa-mutant and only slightly reduced for [3,2]Tpa-AxV (0.95). Similarly, the parent barstar mutant and its [2,3]Tpa-barstar analogue (Figure 17A) exhibit almost identical spectra and signal ratios $(([\theta]_R)_{222}/([\theta]_R)_{208})$ of about 0.90 while for the [3,2]Tpa-barstar protein this ratio is significantly increased (0.97).

The near-UV CD spectrum of wt-AxV shows the contribution of the Phe residues with the sharp fine structure (255 - 270 nm), of the Tyr residues with the maximum centred between 275 and 282 nm, and of the Trp residue with its maximum around 290 and 305 nm (Minks, *et al.*, 1999). Replacement of the Trp residues in both proteins by the two β-(thienopyrrolyl)-alanines leads expectedly to spectral alterations in the aromatic region which mainly arise from the intrinsic dichroic properties of the Trp variants. A comparative analysis of the near-UV CD spectra of Trp and the β-(thienopyrrolyl)-alanines as free amino acids revealed that the spectra of Trp and

[2,3]Tpa are similar while the spectrum of [3,2]Tpa exhibits a strong positive band between 280 - 250 nm (data not shown). This is fully reflected in the spectra of the both mutant proteins and especially for barstar where the contribution of Trp residues to the overall protein spectrum is dominant (Figure 17).
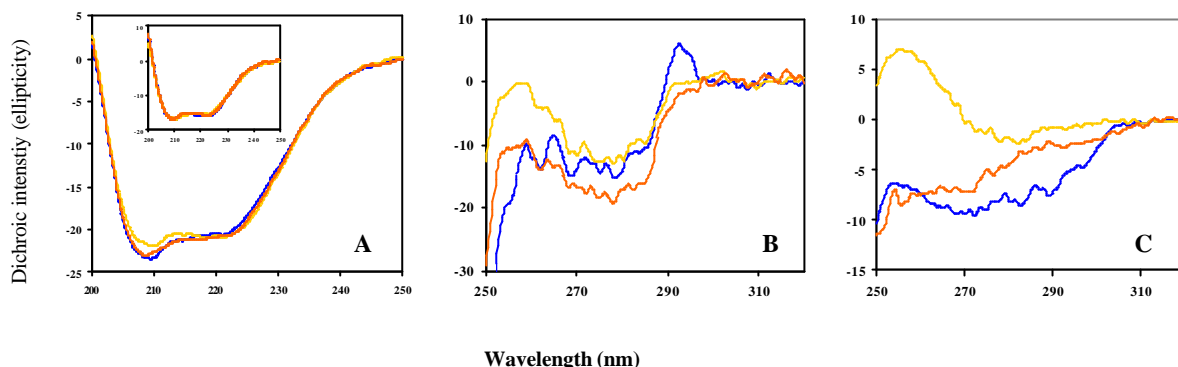
Figure 17. Far-UV CD profiles of the parent and substituted barstar (**A**) and AxV (insert) protein forms at the 20 °C. The mean residual ellipticity ($[\theta]_R$) is expressed in *deg cm² dmol⁻¹*. Near-UV CD spectra for native and mutant proteins at 20 °C: (**B**) AxV, (**C**) barstar. The mean molar ellipticity ($[\theta]_M$) is expressed in *deg cm² dmol⁻¹*. Blue lines represent spectra of wt-protein forms; yellow [3,2]Tpa-mutants and red lines are [2,3]Tpa-protein mutants.

Residue	Free amino acid [c]		AxV [d]		b* (C40A/C82A/P27A/W38F) [c]	
	ε_M	l_{max}	ε_M	l_{max}	ε_M	l_{max}
L-Trp	5579 ± 257	280	21500 ± 931	277	14920 ± 531	280
[3,2]Tpa	5632 ± 373	260	21500 ± 791	277	13060 ± 197	261
[2,3]Tpa	4897 ± 169	241	22800 ± 287	277	10712 ± 331	250

Table 3. Absorption maxima (l_{max})[a] and molar extinction coefficients (ε_M)[b] of L-Trp, [3,2]Tpa and [2,3]Tpa as free amino acids in buffered solution and of AxV and barstar related mutants produced by Trp-replacements. [a]expressed in *nm*; [b]expressed in $M^{-1} cm^{-1}$, [c]in 50 mM Na-dihydrogenphosphate pH 8.0; [d]in 10 mM TrisCl pH 8.0

The near-UV CD spectra of [2,3]Tpa-containing proteins are similar to those of the native proteins (Figure 17B, C) with only small differences in the dichroic intensity of the positive bands. This is stronger for wt-AxV than for [2,3]Tpa-AxV (Figure 17B), but weaker for the parent barstar protein than for the [2,3]Tpa-mutant (Figure 17C).

On the other hand, replacing the Trp residues with [3,2]Tpa in both proteins results in drastic changes of the near-UV CD spectra (Figure 17B, C). In the near-UV CD spectrum of [3,2]Tpa-AxV, this change dominates the spectral region between 250 and 268 nm while between 270 and 290 nm the intensities are only slightly decreased compared to the spectrum of the wt-protein (Figure 17B). A similar, but more pronounced effect is observed for [3,2]Tpa-barstar (Figure 17C) where the changes could originate from a strong local alteration of the spatial array of neighbouring aromatic residues, the intrinsic spectral nature of the [3,2]Tpa as chromophore or a combination of both effects. However, in absence of spectral information from model peptides, any attempt to explain the observed effects in a more reliable and detailed manner would be too speculative.

A comparison of the far-UV CD spectra of all AxV-proteins shows that the overall dichroic patterns are nearly identical (Figure 17A), whereas for [3,2]Tpa-barstar the spectrum is slightly different around 208 nm when compared with native protein variant. This could be indicative of small changes in the secondary structure of this mutant. However, the alterations detected in the aromatic region can affect the far-UV CD spectrum without changes in secondary structure, as it was already observed in other systems (Woody & Dunker, 1999).

2.4 Thermal Denaturation

Thermal unfolding of both model proteins is known to occur as a two-state transition from native to denatured state (Golbik, *et al.*, 1999; Minks, *et al.*, 1999). Compared to the parent proteins, all mutants containing [3,2]Tpa or [2,3]Tpa, respectively, are characterised by lower thermal stabilities (Figure 18). Thereby the proteins substituted with [3,2]Tpa show not only the lowest T_m-values, but also affected the cooperativity of their unfolding process as indicated by changes

in the steepness of the transition profiles (Figure 18A, C) and consequently by the van't Hoff

enthalpies (Table 4). This effect is more pronounced in the case of the barstar mutant where

replacement of the structurally critical and buried Trp53 with [3,2]Tpa possibly provokes local

perturbations of the hydrophobic environment. Incorporation of [2,3]Tpa -residues into the model

proteins leads in both cases to folding cooperativities which are even enhanced by almost 15 % if

compared with the parent proteins (Table 4). It has previously been observed that isosteric

replacements of Met and Trp are accompanied by rather large differences in van't Hoff enthalpies

despite the moderate shifts of the T_m values (Budisa, *et al.*, 1998a; Minks, *et al.*, 1999).
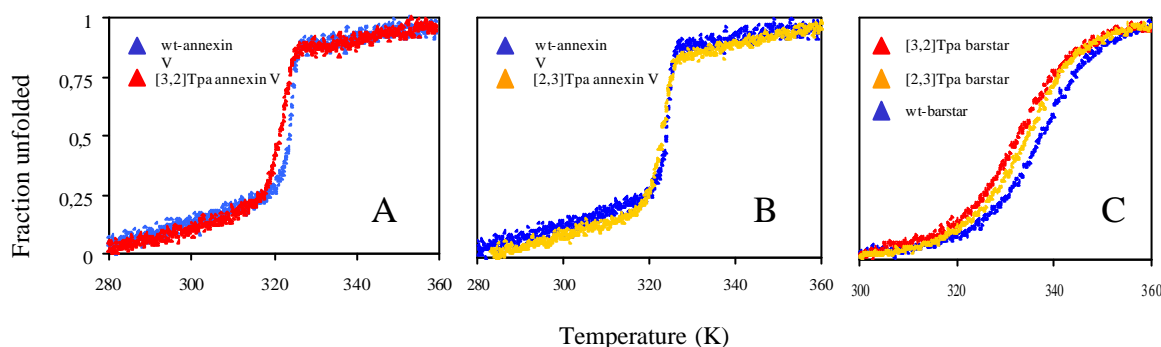
Figure 18. Thermal denaturing of wt and mutant protein forms of AxV (**A** and **B**) and barstar (**C**). Fractions of denatured proteins are calculated from CD data monitored at 222 nm as derived thermodynamic parameters are presented in Table 4.

Protein variant	T_m value (K)	DT_m	DH_m (kJ/mol)
wt-AxV	323.76 ± 0.11	-	703.92 ± 18.8
[2,3]Tpa-AxV	323.70 ± 0.03	- 0.06	817.32 ± 17.7
[3,2]Tpa-AxV	322.69 ± 0.03	- 1.07	677.06 ± 14.1
wt-barstar	337.28 ± 0.20	-	154.14 ± 9.51
[2,3]Tpa- barstar	335.57 ± 0.23	- 1.71	181.90 ± 14.7
[3,2]Tpa- barstar	332.45 ± 0.18	- 4.83	153.63 ± 9.45

Table 4. Thermodynamic parameters of the denaturation of the wild-type and substituted protein variants of AxV and barstar. Related curves are presented in Fig. 18. DT_m is the difference between T_m of the native and substituted proteins.

The findings on thermal stabilities agree with the observed changes in the near-UV CD spectra of both proteins (Figue 17). The [3,2]Tpa-mutants which exhibit significantly enhanced dichroism in the near-UV are characterized by markedly lower T_m values and cooperativities than the [2,3]Tpa-mutants (Table 4). Conversely, the higher cooperativity in unfolding of the [2,3]Tpa-mutants would suggest an enhanced order of the internal architecture.

2.5   Biological Activities

Mutants of both proteins are biologically active in the qualitative activity tests. The barstar mutants inhibit the RNase activity of barnase in standard inhibition assays on RNA agar containing plates (Golbik, *et al.*, 1999) as the parent variant. Similarly, the AxV mutants are able to bind to liposomes in presence of higher Ca^{2+} concentrations (Berendes, *et al.*, 1993).

3   [3,2]Sep in AxV and Barstar

3.1   Fermentation and Expression of proteins with [3,2]Sep

E. coli ATCC49980 exhibits a typical growth curve in minimal medium with 0.03 mM L-Trp; conversely, [3,2]Sep does not support cellular growth of this Trp-auxotrophic *E. coli* strain in minimal medium (Figure 19). As shown in Figure 19, the growth of this strain is strongly inhibited from the beginning of the fermentation in defined minimal medium that contains 0.015 mM (2.7 µg/ml) of the seleno-analogue as sole source of Trp. Interestingly, the Trp-auxotroph *E. coli* strain *SVS370* was reported to grow on glucose minimal medium at amounts of 1 - 50 µg/ml of the "seleno-Trp" with a growth rate that is several-fold slower than that with L-Trp under identical conditions (Welch & Phillips, 1999b).
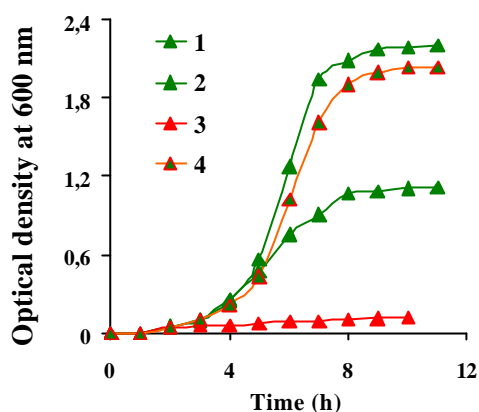
Figure 19. Growth curves at 30 °C of the transformed Trp-auxotrophic *E. coli* strain *ATCC49980* in NMM with different L-Trp and [3,2]Sep concentrations. Four different cultures grown in NMM-medium were supplied with following amounts of Trp as native substrate and [3,2]Sep: (**1**) 0.03 mM L-Trp; (**2**) 0.015 mM L-Trp; (**3**) 0.015 mM [3,2]Sep; (**4**) 0.015 mM [3,2]Sep + 0.015 mM L-Trp. The growth was followed by measuring the changes in the optical density at 600 nm (OD_{600}) by UV spectroscopy.

3.2 Analytics, Stability and Spectral Properties of the [3,2]Sep-containing Proteins

High-level expression and almost quantitative labelling of both barstar and AxV with [3,2]Sep were achieved by the SPI-method (Figure 20). The ability of [3,2]Sep to suppress the Trp-characteristic fluorescence of labelled proteins can be used as a qualitative criterion for successful labelling like in the cases of [3,2]Tpa and [2,3]Tpa. Quantitative estimation was performed by mass spectrometry, since the molecular mass difference between Trp and [3,2]Sep is large enough (50 Da) to be determined reliably (Figure 21; Table 5). These findings fully confirmed that [3,2]Sep once incorporated into proteins is sufficiently stable and resistant to oxidation.
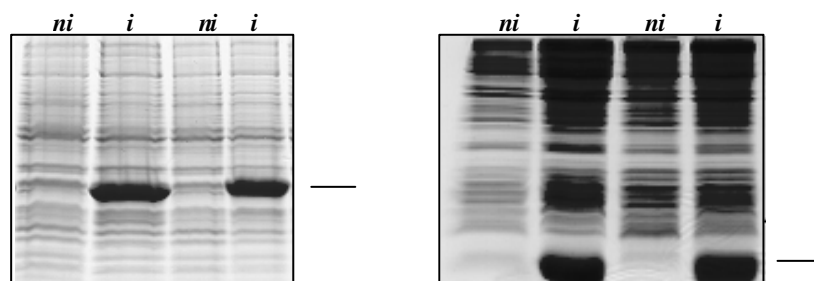
Figure 20. Expression profiles of proteins labelled by SPI-method. Analysis of the expression profiles in cell lysates of *E. coli ATCC49980* with AxV (left) and barstar (right) in defined minimal medium, NMM with Trp (second lane) and [3,2]Sep (forth lane); *ni*: non induced cells, *i*: induced cells in the presence of canonical and non-canonical substrates.
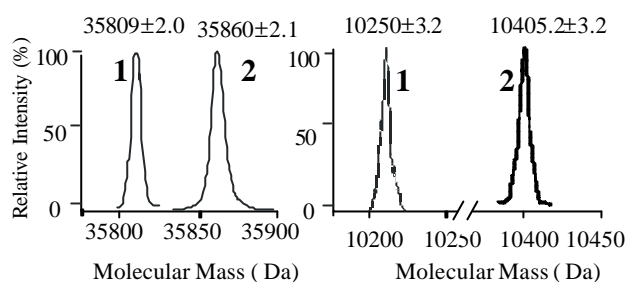
Figure 21. Mass spectrometric profiles of substituted proteins. In all mass spectra of related mutant proteins signals from native species were not detectable. Data are summarised in the Table 5.

The thermodynamic stability measured through temperature-induced melting profiles of substituted proteins (T_m) is not changed dramatically upon replacement: for AxV it is lowered by about 2 K and for barstar by about 3 K as determined by circular dichroism (data not shown). Similar marginal effects were found on the stability and folding cooperativity in SeMet- and even TeMet-substituted proteins (Budisa, *et al.*, 1997), reflecting in this way a remarkable plasticity of the protein, i.e. the ability to retain its native state by tolerating small sequence modifications.

Proteins	Calculated mass (Da)	Measured mass (Da)
AxV	35809.2	35809 ± 2.0
[3,2]Sep-AxV	35859.8	35860 ± 2.1
Barstar	10253	10250 ± 3.2
[3,2]Sep-Barstar	10403	10405.2 ±3.2

Table 5. Calculated and measured molecular weights of wt- and [3,2]Sep-containing model proteins.

Studies of Budisa and co-workers with TeMet proved this amino acid analogue in AxV to be oxygen-sensitive and thus stable only in the interior compartments of proteins expressed in the soluble form, while surface-exposed residues experienced degradation (Budisa, *et al.*, 1997). Similarly, the higher chemical reactivity of SeMet compared to Met and thus, an enhanced susceptibility to oxidation was reported (Smith, *et al.*, 1998). However, SeMet can easily be

incorporated into proteins even when expressed in inclusion-body form since oxidation can be avoided by carefully performed refolding (Piper, *et al.*, 1997). The same holds true for [3,2]Sep as well, as demonstrated with the [3,2]Sep-barstar which is expressed in denatured form and refolded into the native state during purification.

Although barstar refolding was performed without special precautions, i.e. in air-oxygen presence, oxidation of [3,2]Sep residues in the mutant protein could be excluded by mass-spectrometric and X-ray analyses. It should be, however, noted that only Trp53 of barstar is placed in a solvent-free hydrophobic environment, while the two other residues (Trp44, Trp38) are partially or fully exposed to solvent like [3,2]Sep-187 in AxV. This exceptional stability of [3,2]Sep in the mutant proteins might be a considerable advantage of this selenium-amino acid over SeMet in substitution experiments for crystallographic phasing.

Since the UV absorbance of proteins is mainly dominated by the spectral contributions of Trp residues and to lesser extent by Tyr residues, changes in the absorption spectrum can offer an alternative indication for successful [3,2]Sep incorporation into proteins.

As shown in Figure 22, the overall shape and absorption maximum of the spectrum ($\lambda_{max} = 277$ nm) of [3,2]Sep-AxV is not significantly affected by the substitution due to the Tyr dominance in AxV (13 Tyr and only one Trp). However, the intensity of the absorption profile of [3,2]Sep-AxV is significantly increased ($\varepsilon_M = 28400$ $M^{-1}cm^{-1}$) when compared with that of the native protein ($\varepsilon_M = 22500$ $M^{-1}cm^{-1}$). On the other hand, in barstar (b*C40A/C82A/P27A) with only three Tyr and 3 Trp residues, substitutions caused indeed significant changes in the position (blue shift by 9 nm, $\lambda_{max} = 271$ nm) and especially in the intensity at the absorption maximum. Native barstar has an extinction coefficient $\varepsilon_M = 20460$ $M^{-1}cm^{-1}$ at 280 nm while for [3,2]Sep-barstar ε_M is 41580 $M^{-1}cm^{-1}$ at $\lambda_{max} = 271$ nm. Molar extinction coefficients are calculated for the whole

absorption range and are expressed as $M^{-1}cm^{-1}$. In general, the spectral properties of [3,2]Sep-containing proteins (Figure 22) are the best indications for its incorporation into the structure, representing an easy, simple and fast test for successful protein derivatisation.
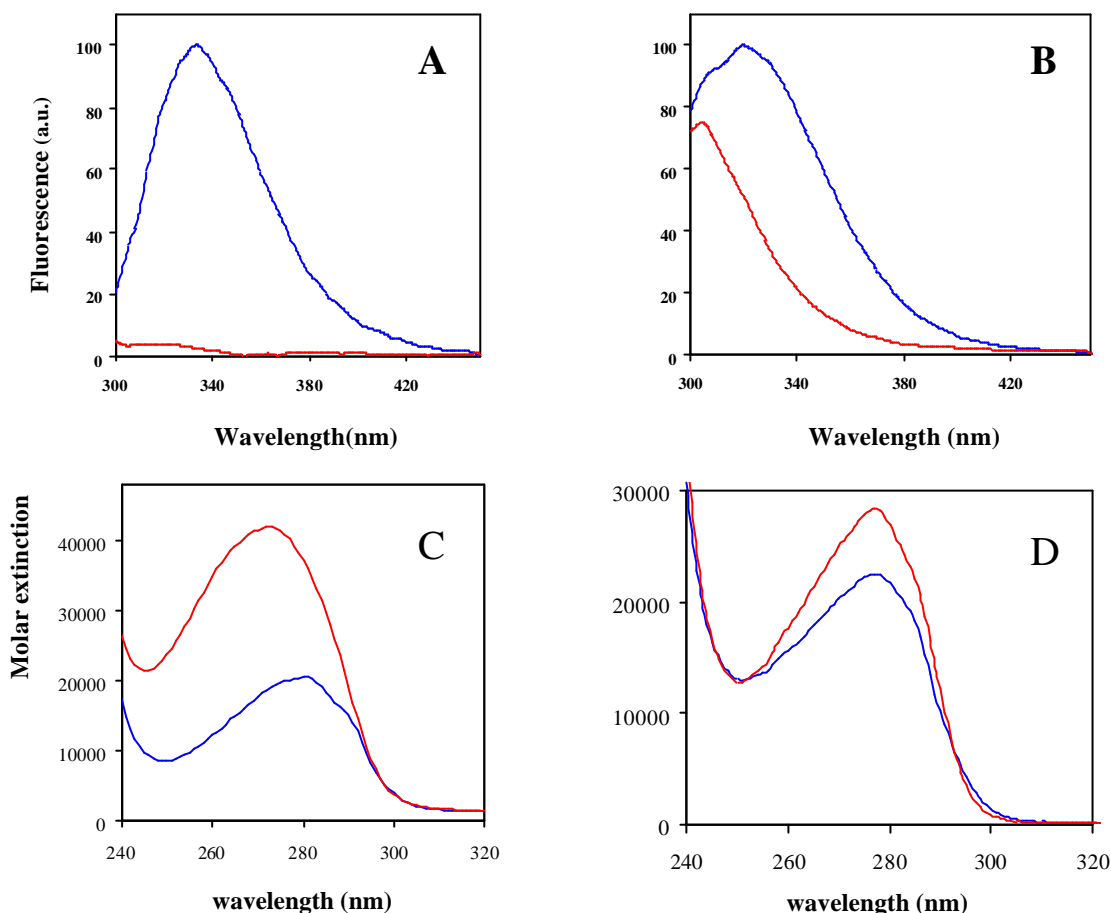
Figure 22. A), B). Fluorescence emission spectra of native and mutant proteins of AxV (**A**) and barstar (**B**) excited at 280 nm (blue: native proteins; red: [3,2]Sep-proteins). **C), D).** UV-absorption spectra of native proteins and their mutants. (**C**: barstar, **D**: AxV) at 20 °C (blue: wt-proteins; red: [3,2]Sep-proteins).

3.3 Crystal Structures of [3,2]Sep-Containing AxV and Barstar

The structure of [3,2]Sep-AxV, as determined at 2.3 Å resolution, revealed full replacement of Trp187 by [3,2]Sep (Figure 23). In wt-AxV crystals (R3, a = b = 99.77 Å, c = 96.6 Å with one molecule per asymmetric unit) the Trp187 is buried inside the protein domain III (Huber, *et al.*, 1990). [3,2]Sep-AxV was found to crystallise preferentially in the same space group (R3) but

with different lattice constants (a = b = 157.31 Å, c = 35.91 Å) and with a surface-exposed Trp187. This molecular form was reported for the Glu95Ser AxV mutant (Berendes, *et al.*, 1993) as well as for rat AxV (Concha, *et al.*, 1993). A Ca^{2+}-binding site was found to be located in the loop with an exposed [3,2]Sep187. This residue forms crystal contacts to a symmetry related molecule in the same manner as Trp187 in the Glu95Ser mutant (Berendes, *et al.*, 1993). Other significant structural alterations compared to the Glu95Ser mutant or rat AxV were not detected for [3,2]Sep-AxV, thus confirming that the replacement of Trp with [3,2]Sep is fully isomorphous.

The crystal structure of the barstar mutant, determined at 2.4 Å resolution, comprises two molecules per asymmetric unit. The overall fold of the [3,2]Sep-barstar is practically identical to that of the parent protein despite the replacement of all three Trp-residues. Since the solvent accessible Trp38 and Trp44 are in contact with the ribonuclease barnase in the crystallographic complex (Giulliet, *et al.*, 1993), they are expected to be more flexible. However, in the crystal structure they are well defined and the difference electron density map revealed precisely the site of replacement (Figure 24). Correspondingly, crystallographic analysis fully confirmed the mass spectrometric data of the mutant protein in terms of stability of the [3,2]Sep residues.
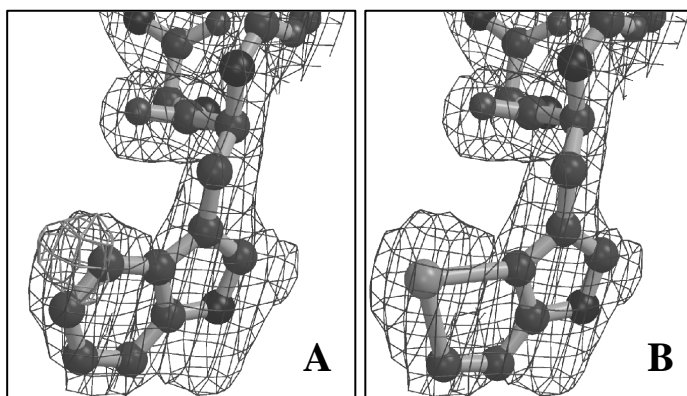
Figure 23. **A**) Difference electron density map (Fo-Fc) of residue Trp187 of AxV contoured at 3.0 σ is superimposed on the 2Fo-Fc electron density map (1.3 σ). **B**) Electron density (2Fo-Fc) map at 1.3 σ with the [3,2]Sep-side chain model structure at position 187 in AxV.
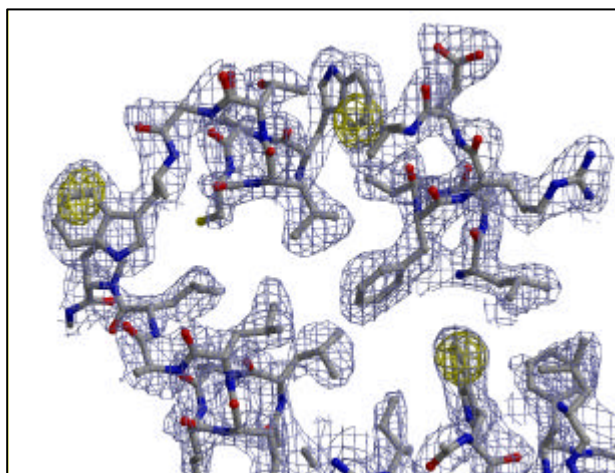
Figure 20. Portion of the electron density map of barstar which includes residues Trp53 (right), Trp44 (above) and Trp38 (left). Difference electron density map (F_o-F_c) between Trp and [3,2]Sep contoured at 3.5 σ (red) is superimposed on to the $2F_o$-F_c electron density contoured at 1 σ (blue).

4   X-ray Crystal Structure of EGFP with [3,2]Tpa, [3,2]Sep, and [2,3]Sep

4.1 Expression of EGFP with [3,2]Tpa, [3,2]Sep and [2,3]Sep

The expression experiments of recombinant wt-, [3,2]Tpa-, [3,2]Sep-, and [2,3]Sep-containing EGFP were performed in Trp-auxotrophic *E. coli ATTC49980* host cells. The cells were first grown in the minimal medium in the presence of 0.015 mM L-Trp as natural substrate until its depletion from the culture medium in the mid-logarithmic phase of the growth (OD_{600}: $0.5 - 0.8$). At that point, the non-canonical analogue was added and the translation of the target gene product induced. Under these conditions optimal expression and high-level labelling of target proteins were achieved. This was confirmed by mass spectrometry which showed that the calculated and experimentally found values were in good agreement (Table 6). On the other hand, UV and fluorescence properties of chromophore of these mutants were indistinguishable from those of parent EGFP protein since only Trp57 was labelled (and it is positioned outside of the chromophore).

	Mass	
	Expected (Da)	Found (Da)
[3,2]Tpa-EGFP	27749.6	27751.4 (±3.0)
[3,2]Sep-EGFP	27793.6	27797.8 (±3.0)
[2,3]Sep-EGFP	27793.6	27798.8 (±3.0)

Table 6. The analytical parameter (mass spectrometry) of EGFP with [3,2]Tpa, [3,2]Sep, and [2,3]Sep.

4.2 Crystallisation conditions of EGFPs with [3,2]Tpa, [3,2]Sep, and [2,3]Sep

EGFPs substituted with [3,2]Tpa, [3,2]Sep, and [2,3]Sep were crystallised in several crystallisation conditions. [3,2]Tpa-EGFP crystallised better in 0.2 M ammonium acetate, 0.1 M tris-sodium citrate dihydrate pH 5.6, and 30 % (w/v) PEG 4000, the EGFP with [3,2]Sep, in 0.2 M lithium sulphate monohydrate, 0.1 M Tris hydrochloride pH 8.5, and 30 % (w/v) PEG 4000, and the EGFP with [2,3]Sep, in 0.2 M calcium acetate hydrate, 0.1 M sodium cacodylate pH 6.5, and 18 % (w/v) PEG 8000 in the time of 1 or 2 days. The crystal pictures of these GFP variants are shown in the figure 25.

Figure 25. Crystals of ECFP, EGFP, and GdFP, respectively.

4.3 Overall structures of EGFPs with [3,2]Tpa, [3,2]Sep, and [2,3]Sep

The overall structures of EGFPs containing [3,2]Tpa, [3,2]Sep, and [2,3]Sep were almost identical to the known structure of parent EGFP (Örmoe, *et al.*, 1996; Figure 27). The overall structure (β-barrel can) and the topology of the chromophore of the parent EGFP were retained in those of [3,2]Tpa-, [3,2]Sep-, and [2,3]Sep-EGFP where Trp57 was only target of substitution.

These results confirm the observation on other proteins (for example barstar and AxV in this study) that isosteric replacement of Trp residue with its analogues usually does not affect their overall structures significantly.
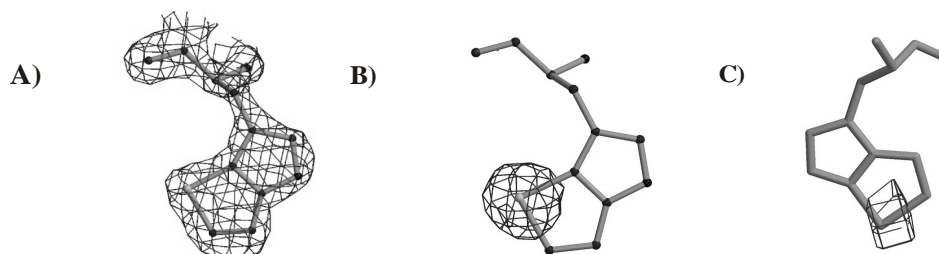
Figure 26. Different Fourier-maps for residue Trp57 replaced with [3,2]Tpa (**A**, 2Fo-Fc, 1.5 s), [3,2]Sep (**B**, Fo-Fc, 4.0 s), and [2,3]Sep (**C**, Fo-Fc, 4.5 s)

Full replacement of EGFP single tryptophan residue at position 57 with [3,2]Tpa, [3,2]Sep, or [2,3]Sep found by analytical methods (Table 6) was fully confirmed in its structure as well (Figure 26). Namely, calculated electron densities clearly reveal replacement sites (i.e. benzyl → thiophene or selenophene moiety) at the indole side chain of Trp57.

Figure 27. Superimposed overall structures of GFPs, wt-(green), [3,2]Tpa-(red), [3,2]Sep-(blue), and [2,3]Sep-(yellow) EGFPs. The ß-barrel structure of the GFP is shown with its chromorphore located in the center of the barrel and Trp57. The chromorphore and the barrel are perpendicular.

Sulphur and selenium are known as static quenchers of fluorescence (Budisa, *et al*, 2001). These elements are responsible for the loss of fluorescence in cases of the [3,2]Tpa-, [3,2]Sep-, and [2,3]Sep proteins (Figure 16, 22) where Trp residues are main source of protein fluorescence. On the other hand, [3,2]Tpa-, [3,2]Sep-, and [2,3]Sep-EGFP variants did not loose their fluorescence, which should not be surprising since the quenching of intrinsic Trp-fluorescence in GFPs is well documented and attributed to the fluorescence resonance energy transfer between these two chromophores (Tsien, 1998).

For MIR, it is necessary that the native protein and its heavy atom derivative have to be isomorphous, which means that all the atoms in both proteins are located to the same positions. The structure of EGFP containing [3,2]Sep shows that there is no major change around the trytophan residue Trp57 which are replaced with its surrogates (Figure 28). But the incorporation of isosteric isomer [2,3]Sep turned out to induce local changes in side chain geometry of substituted Trp57 in EGFP. The geometry of pyrrolyl side chain is somewhat distorted from that of the wild type indole ring (Figure 28) while the backbone of the protein remains unchanged (Figure 27). Obviously more structures containing [2,3]Sep should be solved to see if this observation is general or exceptional.
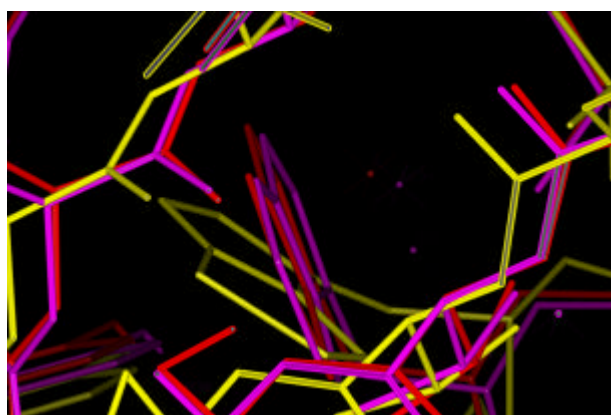
Fugure 28. Trp57 and its surroundings of EGFP. super imposed with parent- (magenta), [3,2]Sep- (red), and [2,3]Sep- (yellow) EGFP showing that [2,3]Sep is not exactly isomorphous to the parent structure while [3,2]Sep is isomorphous at Trp57 in EGFP.

5   ECFP and GdFP

5.1   Expression of ECFP and GdFP

The gene encoding for ECFP (Invitrogen) was inserted in a pQE80 vector (Qiagen) harbouring Nt-His-Tagged gene sequence, and expressed in tryptophan auxotrophic *E. coli ATTC49980* strain which served as expression host. The high substitution levels with the Trp analogues have been confirmed by electrospray mass spectrometric analyses. Both Trp residues in ECFP were replaced to high level (parent-ECFP : 27707 ($\pm$2.9) Da; GdFP : 27739 ($\pm$3.1) Da). They corresponded to the expected masses. Small amounts (2-5 %) of wt-protein are present in all labelled samples; it is well known that even good controllable expression systems are not perfectly "silent" before induction of the target gene expression.

5.2   Spectral Properties of ECFP and GdFP

UV and fluorescent spectra of ECFP and GdFP were found to have unique features. The UV-profile of ECFP as known (Tsien, 1998) has two characteristic absorption maxima, while GdFP has only one band, but a much broader and red-shifted absorption maximum. The UV-profile of GdFP shows absorbtion maximum at 466 nm which brings a red shift for 14 nm when compared with parent ECFP whose second absorbtion maximum is positioned at 452 nm (Table 7).

Protein	*UV (λ_{max})*	*Fluorescence (λ_{max})*	*Mass (Da)* expected	*found*
EGFP	488 nm	509 nm	27743.6	27743 ($\pm$3.0)
ECFP	434 nm, 452 nm	476 nm, 505 nm	27707.0	27707 ($\pm$2.9)
GdFP	466 nm	574 nm	27741.0	27739 ($\pm$3.1)
(4-NH$_2$)-Trp-EGFP	488 nm	509 nm	27760.6	27758 ($\pm$2.5)

Table 7. Spectral (UV-absorbtion and fluorescence emission maxima), and analytical (mass spectrometry) parameters of ECFP and EGFP and their variants substituted with (4-NH$_2$)-Trp.

Fluorescence emission of ECFP is characterised by two maxima in its UV absoption spectra, while those of GdFP has one broad peak (Figure 8) at 574 nm with a weak shoulder ($\lambda_{max1} = 480$ nm) representing a contamination with the parent protein (Section 5.1) . The most dramatic change produced by incorporation experiment is a large red shift (69 nm) in GdFP fluorescence ($\lambda_{max} = 574$ nm) when compared with parent protein ECFP ($\lambda_{max2} = 509$ nm). This is main source for conversion into characteristic golden fluorescence colour of GdFP with an emission maximum similar to that of *dsRed* from *Discosoma striata* (Matz et al, 2000). However it should be kept in mind that the chromphore of *dsRed* is additionally lengthened by conjugation of a fourth amino acid which is responsible for its longer wavelength emission maximum. On the other hand, GdFP chromophore has the same length as the other *Aequoria victoria* chromophores, but an emission wavelength similar to that of the more extended *dsRed* chromophore. Therefore the large red shift can be explained only by GdFP chromophore extension with an additional electron-donating amino group. To exclude the possibility that amination of Trp57 might influence observed spectral parameters in GdFP, EGFP was used as a test system since contains only Trp57. Table 7 clearly shows that (4-NH_2)-Trp incorporation into EGFP results in a substituted protein ((4-NH_2)-Trp-EGFP) with spectral properties indistinguishable from the parent protein EGFP. This confirms additionally that the amino group at Trp57 does not affect the spectral properties of the protein chromophore.

5.3   X-ray data collection and Structure refinement of ECFP and GdFP

The X-ray diffraction data (Figure 29) of ECFP and GdFP were collected, and refined with the support of computational programs (DENZO, CNS). The resolutions of crystals of both proteins were 1.15 Å, and 1.69 Å, respectively. The statistics were summarised in Table 8.
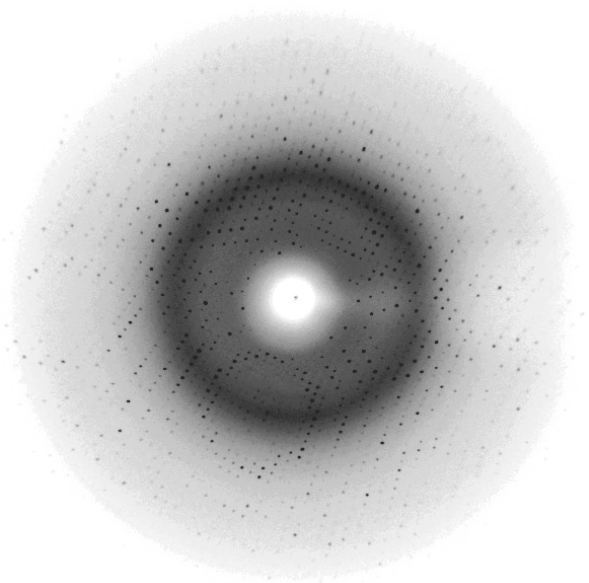
Figure 29. Diffraction pattern of ECFP.

Crystal	ECFP_115	ECFP_175	GdFP
Source	DESY	CuKa	DESY
Wavelength, Å	1.05	1.54178	1.05
Resolution Range, Å	8-1.15	10-1.75	8-1.69
Completeness, %	92.4/84.0(1.17-1.15)	89.9/88.1(1.83-1.75)	90.8/83.5(1.69-1.72)
R_{sym}[1], %	3.9/24.7	3.8/22.4	4.3/33.0
No. Of protein atoms/waters	1800/166	1792/84	1751/94
R_{factor}[2](%)/R_{free}[3](%)	21.5/22.8	22.6/26.1	20.8/24.3
rms bond lengths, Å	0.006055	0.007944	0.007074
rms angles, Å	1.54226	1.56195	1.58621
Refleection used(I/Sigma >0)	73514	22567	22318

[1]$R_{sym} = \sigma[I(h)_i - \langle I(h) \rangle]/\sigma\langle I(h) \rangle$; $I(h)_i$ 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.

[2]$R_{factor} = \sigma[|F_{obs}| - |F_{calc}|] / |F_{obs}| \times 100$

[3]R_{free} was calculated randomly omitting 10 % (barstar 5%) of the observed reflections from refinement and R factor calculation

Table 8. Statistical summary of ECFP and GdFP structure determination and refinement.

In spite of the high qualities of the diffraction data, side chains of some residues at the protein surface and flexible regions were not visualised (Glu132, Asp133, Lys156, Gln157, Lys158 for ECFP, and Ser2, Lys3, Glu5, Glu6, Lys52, Lys79, Arg80, Glu132, Lys156, Gln157, Lys158, Glu172, Asp190, Gln204, Ile229 for GdFP). They all were replaced to alanine in the coordinates. Ile229 for GdFP was deleted because of the lack of the electron density. For the 1.15 Å resolution coordinate of ECFP, Met1 was added.

5.4   Overall X-ray Crystal Structure

The calculated structures of ECFP and GdFP are almost identical to that of EGFP. On the other hand, one significant change could be detected in the electron densities in the chromophore vicinity of ECFP not found in EGFP as well as other known GFP structures. The β-strain from Asn144 to Asn149 of ECFP in the β-barrel structure is twisted from that of EGFP when compared. It is due to the flipping of two residues, Tyr145 and His148. Dominating density (major conformation), is from those where Tyr145 side chain is toward inside of the protein and His148 side chain toward outside. In contrast, both Tyr145, and His148 are toward inside in case of EGFP. In addition, the electron density map for the ECFP shows that the alternative conformation (minor conformation, due to the less prominent density) where Tyr145 is moved outside, while His148 inside. It indicates that this region of ECFP is flipping between two conformations. Conversely, the same region of GdFP is rigid like EGFP but with the same conformation like the predominant ECFP conformation, Tyr145 inside, and His148 outside. GdFP electron density map doesn't reveal any additional conformation like ECFP (Figure 30, 32a).

5.5 Hydrogen Bond Network around the Backbone from Asn144 to Asn149

There are several interactions between the hydroxyl group of tyrosine of EGFP chromophore and the β-strain of Asn144 to Asn149, especially with His148 ring nitrogen, while Asn146 has one hydrogen bond with amide oxygen of Arg168. These hydrogen bonds make this β-strain rigid. In contrast, the electron density map of ECFP revealed that there are two conformations in this region (Figure 30), and one conformation is predominant (**A**, (cyan) in Figure 30), and the other marked as B is minor conformation.
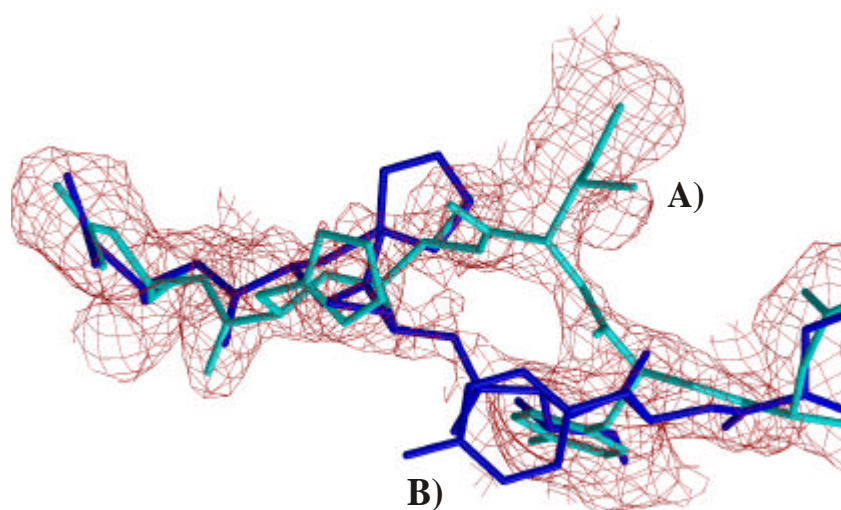
Figure 30. Two conformations of ECFP backbone. The Contour level for the electron density is 0.5. The cyan (**A**) is major conformation, and the blue (**B**) is the minor.

Since the chromophore of ECFP doesn't have the hydroxyl group to interact with the His148 ring nitrogen, these specific hydrogen bonds (which are found in ECFP) are lost. As a result, His148 is flipped out from the inside of the molecule in the predominant form (Figure 30). Furthermore, ECFP is mutated at Asn146 to isoleucine. It explains the loss of the hydrogen bond between Ile146 and amide oxygen of Arg168 which is also presented in EGFP. Therefore, ECFP backbone loses those stabilising hydrogen bonds which are most probably the reasons for losing

rigidity of the region, and the emergence of an additional conformation. For the predominant conformation (**A**), four water molecules and three adjacent residues are in contacts through direct hydrogen bonds. For the minor conformation, only two residues are in contacts (Figure 31).

In GdFP, the backbone is well defined with electron density. Like ECFP it loses the hydrogen bond contacts with the chromophore due to the lack of the hydroxyl group in the GdFP chromophore, and the mutation of Asn146Ile causes additional loss of a hydrogen bond. But it has other contacts which are not found either in EGFP of ECFP. Five water molecules and four adjacent residues are in contacts with the backbone. One of them is extended to Gln69 through two water molecules like the major conformation of ECFP (Figure 32b).
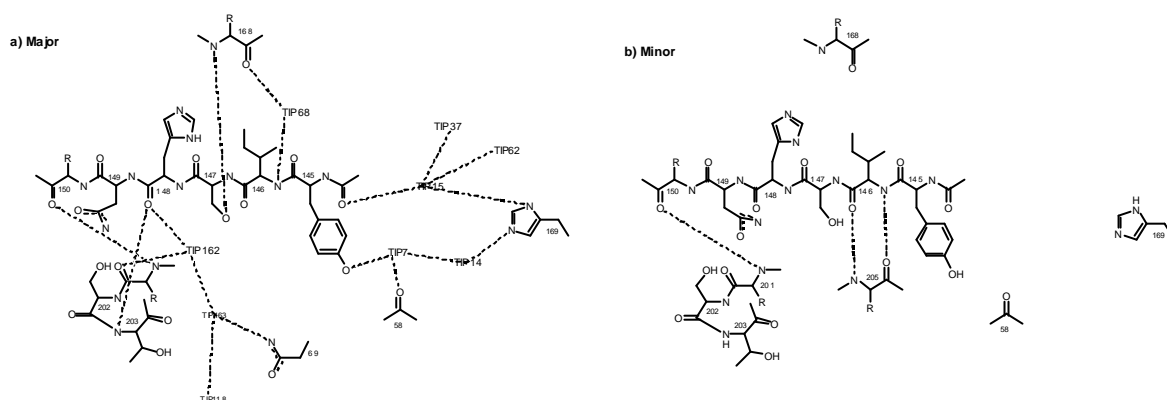
Figure 31. Hydrogen bond network of backbone 145-150 amino acid in ECFP surrounding its chromophore. **A)** the major conformation of the backbone **B)** the minor conformation of the backbone.
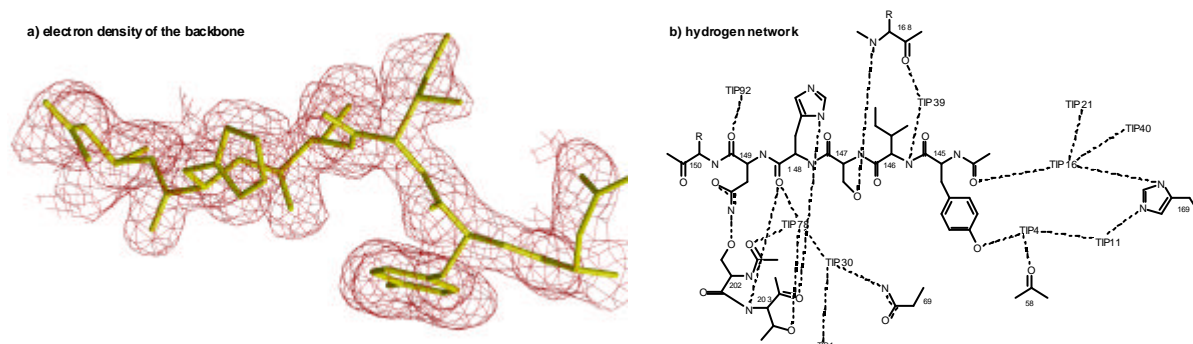
Figure 32. **A)** Difference Fourier map (2Fo-Fc, 1.0 s) of GdFP and **B)** the diagram of the hydrogen network of the residue 145-150 surrounding GdFP chromophore.

5.6   Chromophores are well defined with electron density

The planar resonant structures of chromophres of ECFP and GdFP are defined with electron density maps.   They are well-buried nearly at the center of the β-barrel (Figure 27), and perpendicular to the barrel.   Around the chromophres, there are fixed water molecules.   Another unique property of GdFP chromophore is its small shift toward Phe165.   This is a unique structural feature not observed in parent ECFP.

The hydrogen bond network around the chromophore is extended to the entire molecules, and the fluorescence of the protein is believed to be related to this whole network (Tsien, 1998).   In all three proteins, the chromophores are surrounded by hydrophilic residues and water molecules, and make hydrogen bonds among them (Figure 33, 34). Although the hydrogen bond networks around the chromophores in all three proteins are nearly the same, several side chain rearrangements are observed from the available structures.   For example, GdFP has the water molecules around the chromophore, but they are not hydrogen bonded in the same manner as those in ECFP and EGFP because of side chains shifts of residues around (Figure 34).
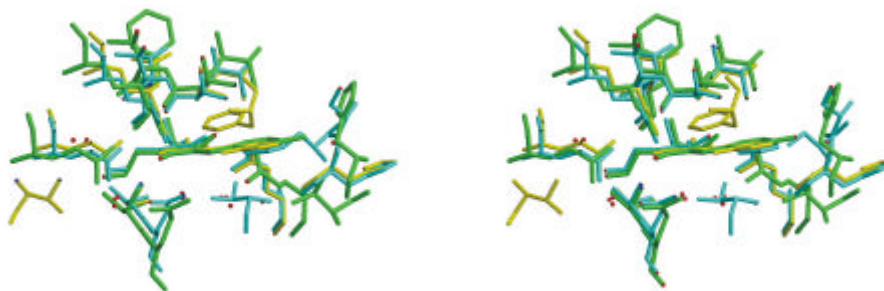
Figure 33. The stereoview of the overall feature around the chromophores. EGFP (green), ECFP (cyan), and GdFP (yellow) are presented.

In EGFP, the hydroxyl group of tyrosine of the chromophore makes three hydrogen bonds with water (TIP316), the ring nitrogen of His148, and O? of Thr203. The water, TIP316 also makes

the contact with seryl hydroxyl group of Ser205 (Figure 34C). And one of nitrogens of Arg96 interacts with carbonyl oxygen of imidazol derivative ring of the chromophore. The hydrogen bond network from one side of the chromophore is connected to the other side through the interactions from amide nitrogen of Val68 to Oγ of Thr65. This connection is constructed through two water molecules, TIP395, TIP374, and two oxygen atoms of the side chain of Glu222. One of the imidazol ring nitrogens of the chromophore makes hydrogen bond with Oγ of Thr65. The side chain of Gln69 is connected to one of the side chain oxygen of Glu222 in the aid of two water molecules, TIP370 and TIP374. The oxygen atom of the side chain of Gln94 is connected to amide nitrogen of Gln69 through water molecule, TIP346. This oxygen further makes a hydrogen bond with Nδ of Asp185, and Oδ of Asp185 with Nε of Gln183. The Nε of Gln183 makes two hydrogen bonds with two nitrogen atoms of the side chain of Arg96. In the reference (Tsien, 1998; 1EMA) there is another connection between Nε of Gln183 and Oε of Gln69.

In ECFP, the tyrosine has been replaced with an tryptophan in the chromophore. This makes the partial loss of the hydrogen bond network which is occurred in EGFP. On the other hand, ECFP chromophore has ring nitrogen in the indole moiety which directly makes one hydrogen bond with Oγ of Ser205 unlike EGFP. ECFP also has the contact between amide nitrogen of Val68 and Oγ of Thr65 via two water molecules, TIP55, TIP117, and one of the two oxygens of the Glu222 side chain. In addtion, the threonyl hydroxyl group of Thr203 contacts with the same Oγ of Thr65 through two water molecules, TIP118, and TIP117. The hydrogen bond network from one side of the chromophore is connected to the other side through the interactions from amide nitrogen of Val68 to Oγ of Thr65 same as in EGFP. Two water molecules are involved, TIP117, and TIP55. The TIP117 has other connections to Oγ of Thr203 via TIP118, and Nε of Gln69 via

TIP118, and TIP163. One of imidazol ring nitrogens of the chromophore again makes hydrogen

bond with O? of Thr65 like in EGFP (Figure 34A).

These hydrogen bond connections are the possible route of fluorescence energy relaxation

(Zimmer, 2002). If there is a longer hydrogen connection or more, the excited state energy can

be relaxed through this extended route. There is one additional hydrogen bond which is

connected to this network, between amide oxygen of Gly67 and TIP11.
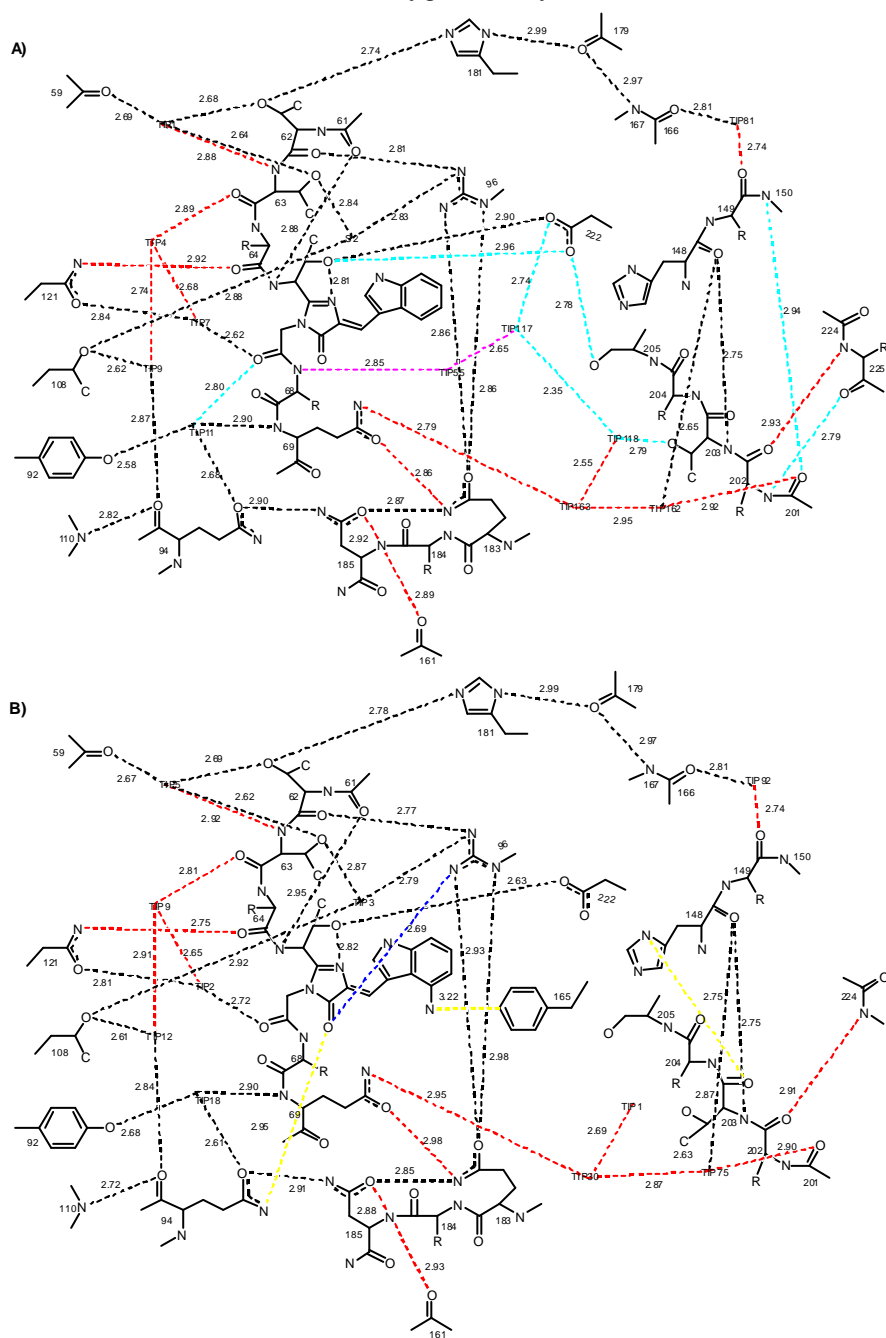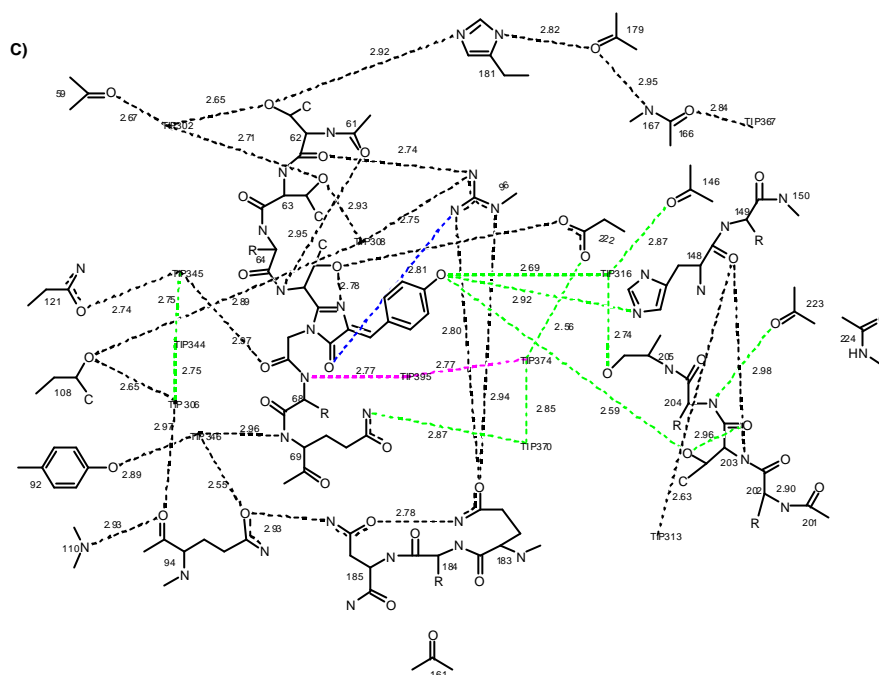
Figure 34. The hydrogen networks around the chromophores **A)** major conformation of ECFP **B)** GdFP **C)** EGFP. The black lines are the interactions which are present in all three while green lines are the interactions only in EGFP, cyan only in ECFP, yellow only in GdFP, red in EGFP and ECFP, blue in EGFP and GdFP, and magenta in ECFP and GdFP, respectively.
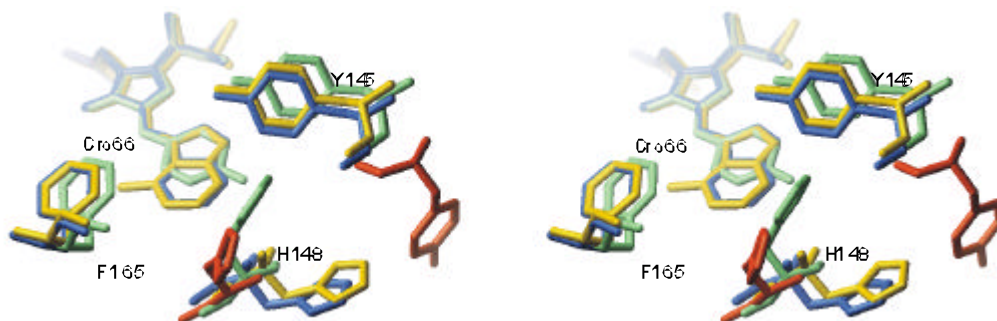
Figure 35. Stereo-view of the chromophores in EGFP (green) ECFP (blue) and GdFP (yellow) and their environments. Tyr145 and His148 of the ECFP minor-form are shown in red. Note a slight shift of the amino indole moiety of the GdFP chromophore toward the Phe165, a property unique among other GFPs.

In GdFP, there is an amino group in the chromophore indole ring (Figure 35). The amino group is at 4 position of the ring. According to the X-ray crystal structure, this chromophore is a bit shifted toward Phe165, which is the opposite direction of the twisted backbone region (Figure 35). This may indicate the interactin between the amino group and p-electron which is quite well documented in literature (Burley & Petsko, 1986). The requirement for topology of this

interaction to be axial with the distance in the range of 3 ~ 6 Å is satisfied perfectly in GdFP. The phenyl ring of Phe 165 and amino group of the chromophore lie axially, and the distance between them is in the range of 3.2 ~ 4.5Å with ring carbons. (It is extended with the p-p interaction between Phe 165 and His181.) Even hydrophobic interaction between them is found (Figure 36). This hydrophobic interaction is also found in case of EGFP, but not in ECFP.
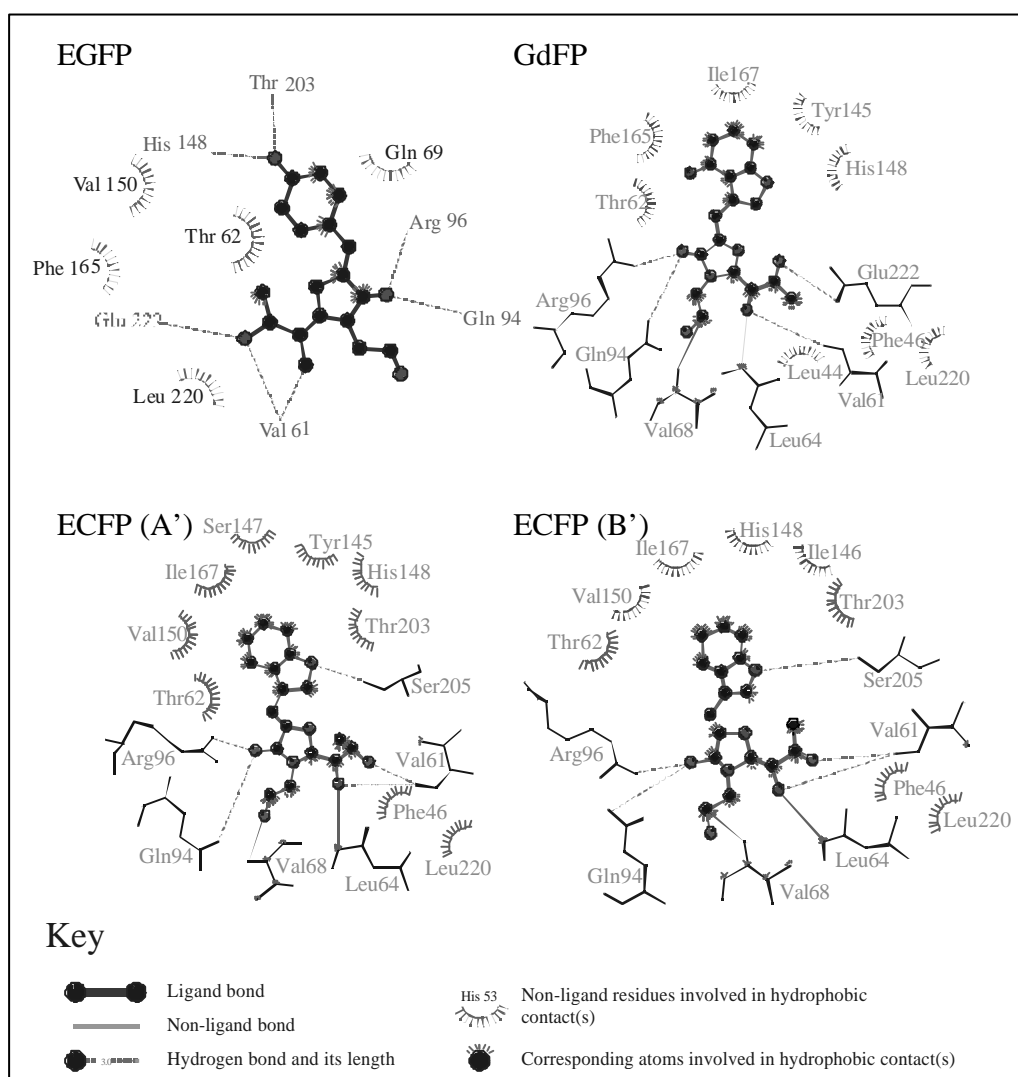
Figure 36. Representation of EGFP, ECFP and GdFP protein matrix-chromophore interactions plotted by using LIGPLOT program (Wallace, 1995). **Upper panel**: EGFP has two hydrogen bonds between hydrophilic residues and phenol ring of the chromophore. The number of the interaction with the GdFP chromophore is reduced to His148, Thr62, Ile167, Tyr145 with a novel interaction with the π-cloud of Phe165. **Lower panel**: ECFP major configuration A' where residues His148, Thr62, Ile167, Thr203, Val150, Tyr145 and Ser145, are in the hydrophobic contacts while Ser205 is in hydrogen bond contact with the chromophore. In the minor conformation B', Ile146 makes a hydrophobic contact with the chromophore while Tyr145 does not make any contact with the chromophore (Courtesy by Dr. M. K. Azim).

DISCUSSION

1 Making Seleno-Surrogates of Trp: New Pathway to Selenophene-3-carboxyaldehyde

Working protocols used in this study are described by Phillips and co-workers (Phillips *et al.*, 1995, Figure 11, 37). However, the attempts to synthesise the starting material, selenophene-3-carboxyaldehyde, using Phillips' protocols were not fully successful since yield had low yield (less than 20 %). For that reason, a new synthetic pathway to achieve selenophene-3-carboxyaldehyde as the starting material for the synthesis of ß-3-selenolo[2,3-b]pyrrolyl-L-alanine was found.

The key step for achieving the target molecule was to make 3-haloselenophene which could be easily transformed to selenophene-3-carboxyaldehyde. Several typical methods were investigated (Figure 12), but failed. An old method was investigated to make the target product (Figure 13), and succeeded to get it. As shown in Figure 13, the method was a very typical chemical method, but the products of the first step (the multibromination) were not detectable with mass spectroscopy, the method, which was used here for detecting the chemical products. The products in Figure 13 were detected by NMR (nuclear magnetic resonance spectroscopy; see *Appendix*).

Even though 3-haloselenophene was achieved, there was another concern for the goal. The reactivity of 3-haloselenophene is quite less than that of 3-halothiophene, the sulphur homologue of selenophene. This means that the reaction to make thiophene-3-carboxyaldehyde used by Phillip and co-workers (1995) was not suitable for 3-haloselenophene. Another way was searched, and finally found. This method was based on the work of Gronowitz and co-workers (1972; Figure 13). When 3-lithiumselenophene transformed from 3-haloselenophene by *n-*

butyllithium was applied to the powdered dry ice (CO_2), selenophene-3-carboxylic acid was formed. This acid could be further modified to ester (Hedlicky, 1971), and finally to the aldehyde (Marx, *et al*, 1984). Using this newly developed procedure, the total yield of 3-haloselenophene is increased to 54 % of overall yield. Even though this procedure requires additional steps, when compared with Phillips' protocols, it brings substantial increase in the product yield and thus can be considered as great improvement.
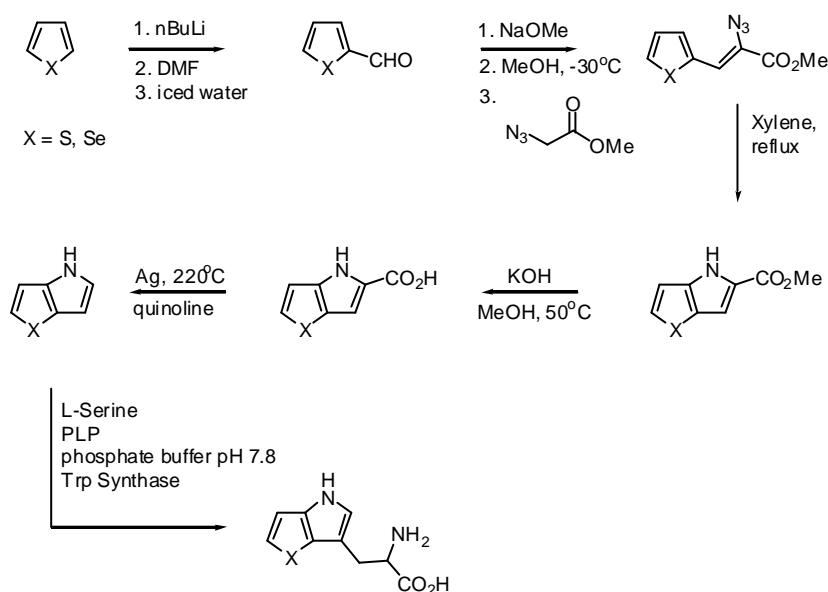
Figure 37. General scheme of the synthetic route to β-[3,2]pyrrolealanines

2 Tryptophan Analogues incorporated into Proteins

2.1 [3,2]Tpa and [2,3]Tpa

2.1.1 Structural Stability of β-(thienopyrrolyl)alanine Proteins

The aromatic amino acids Phe, Tyr and Trp harbour in their chemical structures two properties, i.e. hydrophobicity, being composed of hydrocarbon units, and polarity, the ability to bind ions

through cation-p interactions, that are often considered to be mutually exclusive (Dougherty, 1996). Due to these properties the aromatic residues are generally placed in the interior of proteins or interact with cell membranes. For example, Trp187 of AxV is closely packed in the hydrophobic niche of the domain III (Figure 38) and upon addition of Ca^{2+} in presence of lipid bilayers, it undergoes large local conformation changes with insertion into the membrane (Concha, *et al.*, 1993; Sopkova, *et al.*, 1994; Liemann & Huber, 1997).

When Trp187 of AxV is replaced with β-(thienopyrrolyl)-alanines, the secondary structure of the mutants is identical to that of the parent protein, while the near-UV CD spectra differ (Figure 17B). The differences may originate from coupled-oscillator interactions between the aromatic side chains as it is often observed with other proteins (Woody & Dunker, 1999). Indeed, the neighbouring Phe194 which is in van der Waals contact with the C4 position of Trp187 (Huber, *et al.*, 1990) could be involved in such interactions. Alternatively, the observed effects could also result just from the intrinsic properties of [3,2]Tpa.

Figure 38. Ribbon plots of AxV based on its PDB-coordinates with W187 buried in the hydrophobic pocket at the convex side of the molecule (Huber et al., 1990).

Both [3,2]Tpa-AxV and [2,3]Tpa-AxV crystallise preferentially in the molecular form where the mutated side chain at position 187 is exposed to the surface ("open form") like some other AxV mutants (Berendes, *et al.*, 1993) and rat AxV (Concha, *et al.*, 1993; data not reported).

On the other hand, wt-AxV crystallises preferentially in the "closed form" where Trp187 is buried in a hydrophobic core (Huber, *et al.*, 1990). Therefore, the relatively small differences in the unfolding profiles between native AxV and mutants could be explained as follows. Upon unfolding of the wt-AxV, the indole moiety of Trp187 undergoes a transition from a hydrophobic to a water-solvent environment, while the β-(thienopyrrolyl)-alanine residues are already in contact with the bulk water. Indeed, it is well known that only buried residues contribute significantly to experimental folding parameters (Shortle & Dill, 1991).

With this insight, it is not surprising that in barstar mutants thermodynamic parameters are changed to greater extents than in AxV (Table 4, Figure 18C). Namely, barstar residue 44 is partially and residue 53 completely buried into protein interior. Aromatic residues in proteins are often found to form a network of three or more interacting side chains and these interactions are supposed to serve as nucleation sites in protein folding pathways and as main stabilising forces of the tertiary structures (Nath & Ugaonkar, 1997). In barstar, this clustering occurs around Trp53 which is "sandwiched" between Phe56 and Phe74. Site-directed mutagenesis demonstrated that Trp53 is predominantly contributing to the absorption and fluorescence of barstar upon unfolding (Nath & Udgaonkar, 1997). The near-UV CD spectrum of the [3,2]Tpa-mutant (Figure 17C) would suggest an exceptionally strong coupled-oscillator interaction between the [3,2]Tpa residue and the side chain of neighbouring Phe. On the other hand, it is difficult to conceive large structural rearrangements upon the substitution since it is well documented that the environment of Trp53 is rigid and devoid of any flexibility (Nath & Udgaonkar, 1997). As expected, the X-ray analysis of the crystals of analogues β-(selenolo[3,2-*b*]pyrrolyl)-alanine-containing barstar-mutant fully confirmed these observations.

Like in the case of AxV, the differences in the near-UV CD spectral properties of the barstar variants correspond well with the unfolding profiles. The [3,2]Tpa-mutant is significantly less

stable in terms of T_m then the [2,3]Tpa-mutant and even less than the parent barstar (Table 4). Taking into account of the dominant role of Trp53 for barstar folding and assuming that the local geometry is not significantly changed upon substitution, the packing interactions might be responsible for the observed differences. Such reasoning is based on the assumption that two isosterically shaped moieties may occupy a cavity in a different mode due to their different van der Waals interactions that may exert a strong impact on the restricted internal architecture of the protein. Previous thermodynamic studies on crystallographically isomorphous proteins with Met and Trp isosteric analogues confirmed this as well (Budisa, *et al.*, 1998a; Minks, *et al.*, 1999). Since both [2,3]Tpa and [3,2]Tpa are mutually isosteric, the most conceivable explanation for the differences observed in related protein mutants might derive from their differentiated physico-chemical properties. These differences are caused by the stereochemical position of the sulphur atom relative to the protonated nitrogen in the thienopyrrolyl moiety. Indeed, theoretical molecular-orbital calculations indicate that the stabilities of these positional isomers of thienopyrroles should differ significantly (Milun & Trinajstic, 1977). Thus, by co-translational incorporation of these Trp-surrogates into proteins, their different properties are transmitted, integrated and modulated into the structures of the related mutants.

It is therefore to expect that the replacement of the benzene ring of indole with a thiophene results in altered interactions responsible for the structure of biological macromolecules and for mediating processes such as receptor-ligand interactions, enzyme-substrate binding and antigen-antibody recognition. This was well exemplified in the studies of structure-activity relationships with the enzyme tryptophan-indole lyase. For [2,3]Tpa a k_{cat}/K_M value of one order of magnitude higher (8.6×10^3 $M^{-1}s^{-1}$) than that of [3,2]Tpa (1.2×10^3 $M^{-1}s^{-1}$) was determined and this was attributed to the differences in electronic composition (Sloan & Phillips, 1996).

2.1.2 Biophysical Properties of Thieno-surrogates of Trp

Thiophene-based substances have attracted widespread interests in material science since they may exhibit many useful properties like improved optical transparency or good thermal stability that result from inductive effects of the electron-rich sulphur (Kothakota, *et al.*, 1995; Bretiung, *et al.*, 2000). The rather unusual optical and thermodynamic properties of the proteins containing β-(thienopyrrolyl)alanines may therefore be better explained and understood if more information from suitable model peptides would be available.

2.1.3 [3,2]Tpa and [2,3]Tpa: Chromophores with new spectral windows

Although Trp as the main chromphore in proteins has the unique advantage of an intrinsic probe, it is less suitable for investigating protein/protein or protein/nucleic acids interactions since the absorption spectra of nucleic acids overlap that of Trp and thus prevent the assignment of the spectral contribution of Trp residues to the total signal output. Similar difficulties arise when protein/protein interactions are investigated since many interacting proteins in multiprotein complexes contain Trp residues and thus, their absorption and fluorescence signals are more or less indistinguishable (Ross, *et al.*, 1997).

A very common approach to study the functional role of Trp in proteins is to use site-directed mutagenesis. Thereby, in most of the cases Trp residues are mutated to Phe in the attempt to minimise structural perturbations by replacing one aromatic planar moiety with another. However, this strategy is limited since Trp residues are often essential for the structural integrity and functionality of proteins like in the case of Trp187 in AxV and Trp53 in barstar (Figure 6A, C) and therefore cannot be replaced by any of the remaining 19 canonical amino acids. Even if such replacements are possible, local structural perturbations could alter the spectral

contributions of the remaining chromophores, since the most similar canonical counterparts Phe or Tyr always bring relative large alterations both in the size and charge.

By the use of non-canonical Trp analogues much more subtle alterations are expected that could facilitate interpretation of the experimental data by addressing issues like spectral overlap, better sensitivity to small perturbations of the environments of substituted aromatic side chains and most importantly novel spectral windows. Up to date, five non-canonical analogues of Trp were reported to be incorporated into the proteins, i.e. (4-F)-Trp, (5-F)-Trp, (6-F)-Trp, (7-Aza)-Trp and (5-OH)-Trp (Sommulion, *et al.*, 1995; Ross, *et al.*, 1997; Minks, *et al.*, 1999). All these Trp analogues bring about the lowest possible level of structural alterations, i.e. single atom exchanges like H→F or =CH– → –NH–, thus providing "atomic mutations" for studying protein folding, activities, dynamics and stability (Budisa, *et al.*, 1998a; Minks, *et al.*, 1999).

With β-(thienopyrrolyl)-alanines incorporated into proteins two notable spectroscopic properties are achieved, i.e. altered absorption profiles at least when the Trp contribution is dominant as in the case of the barstar mutant (Figure 22), and an efficient static fluorescence quenching. Besides the two β-(thienopyrrolyl)-alanines as isosteric analogues of Trp reported by Alefelder and co-workers (Alefelder, *et al.*, 2001), (4-F)-Trp was the only known non-fluorescent Trp analogue incorporated into proteins *in vivo* (Bronskill & Wong, 1988). Therefore, the repertoire of "silent" fluorophores as protein building blocks for *in vivo* translation are increased as needed for investigating protein/DNA interactions or multi-protein assemblies.

2.1.4 Chemistry and Pharmacology of Thienopyrroles

Bioisosteres are isosteric molecules that have near-equal shapes and volumes, approximately the same distribution of electrons and exhibit similar or antagonistic properties in biological systems.

Such compounds are abundantly found in nature like for example alkaloid bearing plants (Burger, 1991). It has been attempted to produce the indole isosteres where the imino group of the indole moiety is replaced with other heteroatoms (sulphur or oxygen). Unfortunately, these experiments resulted in compounds that did not predictably retain biological properties analogous to their indole counterpart. First, these planar systems were not activated in the enzymatic condensation reaction with serine by Trp-synthase (unpublished data). Second, even if the related Trp analogues would have been synthesised, probably they would not act as substrates for activation by tryptophanyl-tRNA synthetase (TrpRS) in protein translation. Finally, even changes at the neighbouring positions of the protonated nitrogen of indole were disturbing since the introduction of an additional nitrogen into the indole position 2 leads to (2-Aza)-Trp that is not recognized as substrate for protein synthesis. On the other hand, introduction of a nitrogen in position 7 of the indole results in (7-Aza)-Trp which is recognised by cellular TrpRS and incorporated into proteins. These findings illustrate the universal biological significance of the indole imino function and thus suggest that the benzene ring of indole is a much better target for chemical transformations that might lead to biologically interesting isosteric compounds.

The biomedical potentials of aromatic systems consisting of a pyrrole nucleus fused to a thiophene are fully recognised due to their similarities to indoles (Gronowitz, *et al.*, 1976). Indeed, it is well established that thieno[3,2-*b*]pyrrole and thieno[2,3-*b*]pyrrole are bioisosteric analogues of the hallucinogen and serotonin agonist, *N,N´*-dimethyltryptamine (Blair, *et al.*, 1999). For the matter of that, a further step for their wider utility in biomedicine could result from their incorporation into suitable proteins since pharmaceutically active substances could convert protein mutants into useful therapeutic or even diagnostic tools as recently proposed (Budisa, *et al.*, 1998b). Namely, recombinant proteins that contain such pharmaceutically active

amino acids could act as specific "shuttles" or even "magic bullets" due to their potential ability of selective delivery and targeting in the human body (Minks, *et al.*, 2000b).

2.1.5 Overall Structure of [3,2]Tpa-containing EGFP

Medicines are chemical compounds which selectively affect sick cells and tissues while sparing normal ones. However in many cases, these compounds in undesired way interfere with normal physiological pathways in the living cells. This has raised the issue of delivering drugs directly to the target tissue in the way that diseased cells are inhibited while normal ones are spared (Budisa, *et al.*, 1998b). To achieve this, it is necessary to have a carrier which is selective in delivery and specific in biological action (Minks, *et al.*, 2000b). Proteins substituted isomorphously with cytotoxic or pharmacologically active amino acids might represent best possible vehicles to achieve these goals. The carrier protein for the drug delivery is required to specifically interact to the target, such as receptors. And it also has to be structurally unchanged when modified with desired compounds. This also means that the compound should not cause significant structural change when it is incorporated into proteins.

[3,2]Tpa and [2,3]Tpa, which are the sulphur-containing surrogates of tryptophan, are pharmaceutically active compounds (Burger, 1991). They can be delivered into the specific targets via a carrier protein when they are attached chemically or biologically. Here, with EGFP, one of GFP variants which contains only one tryptophan residue Trp57, pharmaceutically active compounds, [3,2]Tpa and [2,3]Tpa, were incorporated successfully, and the global structure of the protein was not changed (Figure 39). This result demonstrates that these compounds indeed can be integrated in prodrug-form into the carrier protein without its large structural changes. Therefore crystallographically determined three-dimensional structures of [3,2]Tpa-, and

[2,3]Tpa–containing EGFP might be seen as a proof-of-principle for the concept of proteins as specific drug-carriers for drug delivery and targeting.

Figure 39. Close look at Trp 57 in parent (green) and [3,2]Tpa-containing (blue) EGFP. Both are superimposed at Trp 57 showing that the overall structures are isomorphous.

2.2 [3,2]Sep and [2,3]Sep as a Useful Vehicle for MAD Phasing

2.2.1   Chemistry and Metabolism of [3,2]Sep in Living Cells

For the synthesis of possibly useful isosteric heavy atom analogues of tryptophan, the nitrogen of the indole moiety appears to be a very attractive target. But as described before, the benzene ring of the indole moiety is apparently a much more permissive target for chemical transformations that might give compounds containing electron-rich atoms suitable for X-ray crystallography.

ß-Selenolo[3,2-*b*]pyrrolyl-L-alanine is efficiently prepared from serine and selenolo[3,2-*b*]pyrrole by Trp-synthase (Figure 11). This confirms the recognition of the fused pyrroles by the enzyme in a manner similar to the natural indole substrate (Welch & Phillips, 1999a). Replacement of the benzene ring of Trp with selenophene, however, was expected to provoke

significant alterations of physicochemical properties in the aromatic ring, thus possibly affecting interactions of the indole moiety in proteins (p-p stacking, hydrogen bonding, cation-p interactions). Such interactions are known to play a major role in protein stability and folding pathways as well as in mediating processes such as receptor-ligand interactions, enzyme-substrate binding and antigen-antibody recognition (Doucherty, 1996). In other words, Trp is a unique amino acid because of the multiple types of interactions it can participate in, and therefore the possibility of local or even global changes in the structure of proteins containing [3,2]Sep residues could not be ruled out. For example, the unique Trp187 of AxV is found in a buried position ("closed" form) in a crystal structure (Huber, *et al.*, 1990) while some of its mutants studied by X-ray crystallography have these residues in "open" form (i.e. on the surface of the protein) (Sopkova, *et al.*, 1994; Liemann & Huber, 1997). These studies have shown how single point mutations in its close surrounding are often sufficient to provoke its large-scale conformational transitions. Similarly, the replacement of the benzene moiety of Trp with selenophene at position 187 of AxV probably produced the same effect as well. Therefore, [3,2]Sep-AxV crystals are isomorphous relative to the Glu95Ser mutant or rat AxV but not to that of wt-protein.

2.2.2 [3,2]Sep as Supplementation for SeMet in Protein X-Ray Crystallography

About a decade ago it has been anticipated (Hendrickson, *et al.*, 1990; Hendrickson, 1991) that molecular biology would aid in evaluating phase angles of the diffracted X-ray waves by allowing incorporation of suitable diffraction labels into proteins. In this context, the SPI-method was found to represent a simple and efficient approach. It also allowed the incorporation of the [3,2]Sep amino acid into proteins, thus confirming that expression of related protein mutants

should be technically rather simple and straightforward. Correspondingly, the current wide use of SeMet for phase determinations by MAD could be supplemented by the use of [3,2]Sep since this amino acid brings additional advantages. These advantages can be discussed in terms of relative abundance and preferred topological positions of Met-, and Trp-residues in proteins.

Both Trp and Met are relatively rare amino acids coded by single codons of the genetic code and are not strictly polar or apolar but combine both properties which are reflected in their promiscuous topological distribution in proteins (Budisa, *et al.*, 1999b). Sequence comparisons usually assign similar hydrophobicities to methionine, leucine and isoleucine residues in proteins (Dayhoff, 1972; Rose, *et al.*, 1985), although the methionine side chain is significantly more flexible than those of both leucine and isoleucine with their branched and more rigid side chains. Methionine residues participate in protein structures with both hydrophobic interactions and hydrogen bonding; they are relatively rare (1.5 % of all residues in proteins of known structures), and usually located in positions inaccessible to the bulk solvent with only 15 % of all Met residues exposed to the surface (Rose, *et al.*, 1985) while the natural abundance of Trp is even lower (1.2 %) as discussed earlier (Dayhoff, 1972). But these statistical considerations should be taken cautiously; it must be always kept in mind that natural amino acids have specific abundance patterns in particular proteins that usually reflect their specific roles in folding, structural organisation and biological function.

One possible complication deriving from the use of SeMet might come out when the number of selenium sites in the related protein crystals is too large to be successfully located either by direct or Patterson search methods. In these cases, the use of Trp analogues/surrogates like [3,2]Sep could be a promising alternative since UGG codons of Trp are generally much less abundant than AUG codons for Met residues in protein genes (McCaul & Ludescher, 1999; Dayhoff, 1972; Rose, *et al.*, 1985). On the other hand, the total anomalous signal depends on the number of the

well ordered anomalous scatterers and it is empirically established as a rule of thumb that well positioned Se and Br atoms give reasonable dispersive differences by one Br or Se atom per 75-90 residues (Boggon & Shapiro, 2000). However, it cannot be excluded that in some cases even lower number of either Se or Br atoms could give larger dispersive differences sufficient for phase calculations. In the case of the model proteins here, methionine-free barstar (79 residues) with its three Trp-residues is an almost ideal candidate for [3,2]Sep-incorporation for *ab initio* phase determination by MAD while AxV (320 residues), with only one Trp, is more suitable for SeMet-incorporation since it contains 7 Met residues.

Larger proteins like Protease P1 (961 residues, 30 Met and 12 Trp) could also be easily labelled with [3,2]Sep (unpublished material). In this particular case, the replacement of Trp-residues introduces a smaller but sufficient number of Se atoms, the positions of which might certainly be easier determined than those from SeMet in the phase calculations by MAD. Indeed, there are many other examples of protein sequences with large differences in the Trp/Met ratios, where labelling with [3,2]Sep could be advantageous over SeMet and *vice versa*. For example, methionine-tRNA synthetase (750 residues) from *S. cerevisiae* contains 9 Met and 12 Trp residues while its *E. coli* counterpart (650 residues) has 20 Met and only 9 Trp residues. One has to decide for every single case of MAD experiment which amino acid is most suitable for derivatisation. Additional interesting examples are proteins that do not contain enough either Trp-, or Met-residues for phase calculations by MAD, like immunoglobulin kappa chain from human milk (218 residues) that have only one Met-, and two Trp-residues. In this particular case, only simultaneous incorporation both SeMet and [3,2]Sep by use of suitable expression auxotrophic host might provide enough Se atoms for a successful MAD experiment. However, it is possible to change natural abundance of Trp-, and Met-residues in proteins by using recombinant DNA technology to arbitrarily introduce or eliminate Met AUG or Trp UGG codons

from gene sequences. In the combination with SPI-method for production of both SeMet- and [3,2]Sep-protein variants, site-directed mutagenesis can help in preparing various mutants that contain suitable number of diffraction labels.

Additional difficulties in the use of SeMet may arise from flexibility when positioned at the protein surface leading to high crystallographic B-factors. On the other hand, aromatic side chains are often part of the rigid protein interior. They often participate in networks of three or more interacting side chains which are supposed to serve as nucleation sites in protein folding and as main stabilising forces of tertiary structures (Gopalan, *et al.*, 1997). In barstar, such a clustering occurs around Trp53 which is sandwiched between Phe56 and Phe74. Site-directed mutagenesis in combination with extensive biophysical studies clearly showed that Trp53 is in a rigid environment (Nath & Udgaonkar, 1997), a fact that is in full agreement with the crystal structure analysis of [3,2]Sep-barstar, as shown in Figure 24. Therefore, in cases where SeMet is less suitable for diffraction analysis because of the atomic mobility factors, [3,2]Sep could be the alternative of choice.

Another possible complication might arise from partial oxidation of SeMet, resulting in the Se? oxo form which is not suitable for phasing whereas the related dioxole-form can indeed be used for the MAD experiment (Smith & Thompson, 1998). In this context, an exceptional stability of [3,2]Sep at the surface positions in barstar during the refolding procedure could be advantageous for other proteins as well. The most probable reason for this stability is resonance stabilisation of Se atom in the aromatic ring, which is then less prone to oxidation than the dialkylselenides like in SeMet.

Taking into account the stability of [3,2]Sep in routine fermentation and crystallisation experiments and the relative low abundance of Trp residues in proteins, *in vivo* incorporation of

[3,2]Sep is certainly a useful addition to the arsenal of X-ray crystallographic phasing by MAD and MIR.

2.2.3   First Incorporation of [2,3]Sep into Proteins: A Novel Derivative for MAD and MIR

The applicative potentials for protein labelling with [3,2]Sep should not be rigidly limited only to phase calculations by the MAD method. In some cases, it is conceivable that both SeMet and [3,2]Sep could be used as conventional derivatives for the standard multiple isomorphous replacement (MIR) method for phase calculations. For example, since the bioincorporation of these analogues (or surrogates) by the SPI-method is residue-specific, this feature can especially be useful in a chain-tracing procedure which can greatly improve the quality of electron densities and subsequently facilitate their interpretation.   Moreover, ß-selenolo[3,2-*b*]pyrrolyl-L-alanine ([3,2]Sep) has another positional isomer, i.e. ß-selenolo[2,3-*b*]pyrrolyl-L-alanine ([2,3]Sep). These two isosteric isomers differ from each other by the stereochemical position of the Se atom relative to the protonated nitrogen in the selenolopyrrolyl moiety (Se atoms are about 4.5 Å apart from each other). This isosteric analogue could indeed be used as a second derivative in MIR phasing (Figure 26C).

2.3   (4-NH$_2$)-Trp in ECFP: A Novel "Gold" Class of GFPs

2.3.1   Spectral Properties of GdFP

In a simple model, it is expected that an increase in the size of the delocalised p-system results in an increase in the excitation wavelength. This is observed for the following mutations leading to

neutral chromophores in GFPs: Phe66 (l_{max}=355 nm) < His66 (l_{max}=386 nm) < Tyr66 (l_{max}=395 nm) < Trp66 (l_{max}=436 nm) (Palm & Wlodawer, 1999). The same effect is observed if the chromophore is additionally lengthened by conjugation of a fourth amino acid as is the case for the recently discovered *dsRed* chromophore (Matz, *et al.*, 1999). It is surprising that GdFP chromophore has an emission wavelength similar to that of the more extended *dsRed* chromophore while it has the same length as the other *av*FP chromophores. This effect must be due to the electron-donating properties of the amino-substituted Trp. For GdFP, the question arises whether the transformation from "cyan" to "gold" can be attributed solely to the replacement of the Trp at position 66 with (4-NH$_2$)-Trp or whether there is also an influence from the second Trp in the protein at position 57. Tyr66 *av*FPs have only one Trp residue which exhibits no tryptophan fluorescence. In order to investigate the influence of Trp57 on fluorescence, EGFP was used as a test system which contains only Trp57. Figure 9 shows that (4-NH$_2$)-Trp incorporation into EGFP results in a substituted protein ((4-NH$_2$)-Trp-EGFP) with spectral properties indistinguishable from the parent protein EGFP. This clearly shows that the amino group at Trp57 does not affect the spectral properties of the protein chromophore.

A second interesting point is the question to which extent the spectral properties of the GdFP chromophore are affected by the sterical and the electronic properties of the amino acid at position 66. The incorporation of (4-Me)-Trp into ECFP is essentially an **-NH$_2$ $\rightarrow$ -CH$_3$** isosteric replacement and represents the best possible way to clarify this issue. The (4-Me)-Trp-ECFP variant has a slightly red-shifted absorbance and fluorescence spectrum (4 nm), but no significant differences in the quantum yields and extinction coefficients are observed (Chalfie, *et al.*, 1994). As shown in Figure 9, the dramatic spectral changes in GdFP are solely due to the presence of the electron-donating amino group.

Large Stokes' shifts exhibited by many GFP variants are commonly explained by excited state proton transfer (ESPT) (Chattoraj, *et al.*, 1996; Lossau, *et al.*, 1996). In the case of GdFP, a similar explanation for the observed Stokes shift of more than 100 nm seems unlikely because of the lacking acidity of the chromophore amino group (Waluk, *et al.*, 1988) as compared to the phenol group in other GFP mutants. Only the protonation of the amino group in the ground state could give rise to an ESPT similar to that observed in other *av*FPs. However, time resolved experiments do not give any indication for ESPT (Dr Andreas Zumbusch, data not shown).

Time resolved experiments (Dr. A. Zumbusch, personal communication) show that the X-ray structure of the protein does not indicate the presence of a possible proton acceptor in the chromphore's vicinity. Based on these results it seems unlikely that the Stokes' shift of GdFP is due to ESPT. Other explanations for the large observed Stokes' shift should therefore be considered.

Lone pair electrons of the amino nitrogen are conjugated to the delocalised p-system of tryptophan. Solution studies of amino-tryprophans unambiguously confirm that the polarizable amino group leads to resonance stabilisation of the excited state (Sinha, *et al.,* 1987). In addition to the amino-tryptophan, the GdFP chromophore contains an imidazolidinone moiety which serves as an electron acceptor. Donor-acceptor type fluorophores are known to exhibit large Stokes' shifts (Chattoraj, *et al.*, 1996; Lossau, *et al*., 1996). The spectral red shift observed in GdFP is thus most probably a consequence of the intramolecular charge migration presented as a series of canonical structures in Figure 40.
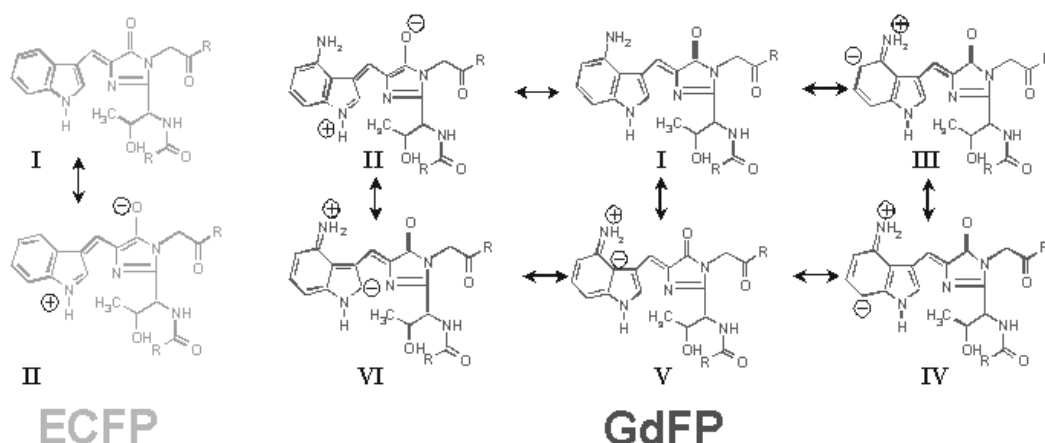
Figure 40. Possible resonance isomeric canonical structures of GdFP chromophore. Note that the number of possible isomers is far larger for GdFP as for ECFP. This indicates that charge migration is possible to higher extent in GdFP chromophore, as well as its stronger resonant stabilisation when compared with ECFP.

2.3.2   Conformational Change at the Backbone Asn144 to Asn149 of GdFP

There was so far not satisfactory explanation for the double-humped peaks in its absorption and emission spectra of ECFP (Tsien, 1998). It was suggested that the origin of double peaks might be the equilibrium state of vibrational levels or other quantum states in the excited state (Zimmer, 2002). But it can be explained by the conformational changes according to the X-ray crystal structure (Figure 30).

The availability of the high resolution crystal structures of *av*GFPs (Örmoe, *et al.*, 1996; Yang, *et al.*, 1996) offers a framework for a plausible model correlating the chromophore environment configuration and its spectroscopic function. The absorption and emission spectra of ECFP exhibit two slightly wavelength-shifted maxima, whereas GdFP shows only a single peak (Figure 7). The observation of these spectral changes upon the insertion of amino-tryptophan into the ECFP chromophore motivated us to elucidate their three-dimensional structures in order to obtain structural frameworks for a more detailed understanding of these phenomena. As expected, both

the crystal structures of ECFP (1.15 Å) and of GdFP (1.69 Å) monomers reveal mostly the same overall topology of EGFP.

In ECFP, the mutagenesis of one residue which is tyrosine 66 to tryptophan made a change of one β-strain (Asn144 to Asn149). This strain becomes less stable by losing the hydrogen bonding contact between the ring nitrogen of His148 and the chromophore (Figure 31, 34). In EGFP, this hydrogen bond holds the same strain rigid (Figure 34C), and another hydrogen bond network between hydroxyl group of Tyr145 and the ring nitrogen of His169 through water molecules stabilises the backbone conformation around Asn144 and Tyr145. The same network is also visible in the predominant conformation of ECFP and in that of GdFP. Also the Nδ and the amide nitrogen of Asn146 interact with the amide oxygen of Arg168. These hydrogen bonds stabilise the conformation of Asn146 in EGFP. This position is so important for stability of EGFP that its mutation Asn146Ile in ECFP abolishes all hydrogen bonds of this residue. In the minor conformation of ECFP, it is even worse and almost all interactions are lost (Figure 31). It has only the hydrogen bonds between the amide nitrogen and oxygen of Ile146 and the amide oxygen and nitrogen of Ser205, respectively. With these two conformational isomers of ECFP, the unusual spectral property of ECFP might be explained. The absorption spectrum and the fluorescent spectrum of ECFP show that there are two transition states in ECFP (Figure 9, 41). These two peaks are due to the possibility of two conformational isomers in the ECFP chromophore's vicinity.

Tyr145 and His148 are also involved in the direct interaction with the ECFP chromophore as shown in the plot of Figure 34 and 36. By contrast, in the minor conformation, B', Tyr145 is turned outside and His148 inside the protein. Tyr145 is not in the range of hydrophobic contacts with the chromophore as mentioned above. This is a strong indication that these side chains flip between these two conformations (Figure 41) in agreement with the recently proposed model

(Seifert, *et al.*, 2002). On the other hand, the electron density of the GdFP chromophore environment revealed only one conformation in agreement with a single fluorescence and absorption peak.

GdFP also loses the contacts due to the mutagenesis of Asn146Ile, and the interaction between the chromophore and His148. But, not like ECFP, it makes other contacts with adjacent backbone residues through hydrogen bonds (Figure 32B). Like in ECFP, the amide nitrogen of Ile146 makes the hydrogen bond with amide oxygen of Arg168 in the aid of one water molecule, and the amide nitrogen of this residue makes the contact with hydroxyl group of Ser147. These contacts hold the region rigid. Also the amide oxygen of His148 interacts with one water molecule TIP78 which makes hydrogen bond with the amide oxygen of Leu201, O? of Thr203, and another water molecule TIP30. TIP30 makes hydrogen bonds with one water molecule TIP1 and Nε of Gln69. Nδ of Asn149 also interacts with hydroxyl group of Ser202. But an additional hydrogen bond network is made between ring nitrogen of His148 and amide oxygen of Thr203. This additional interaction seems to be important to make the backbone rigid because it is the only one difference between ECFP and GdFP. The electron density shows that there is no evidence of existence of any additional conformation of GdFP in this region (Figure 32A). This observation is in good agreement with the measured spectral properties of GdFP (Figure32).

2.3.3 Structural Framework for understanding the Differences between the Chromophores

The literature indicates that the quantum yield is decreased when the chromophore has more possibility to twist out of the axis of the chromophore plane (Zimmer, 2002). If the vibration and rotation around the exo-methylene double bond of the chromophore is prevented, the fluorescence of the protein is not reduced. It is therefore clear that the increased flexibility of the

choromophore reduces the quantum yield of the protein. In addition, the charge transfer is also considered as an important fact of reducing the quantum yield in GFPs (Tsien, 1998).
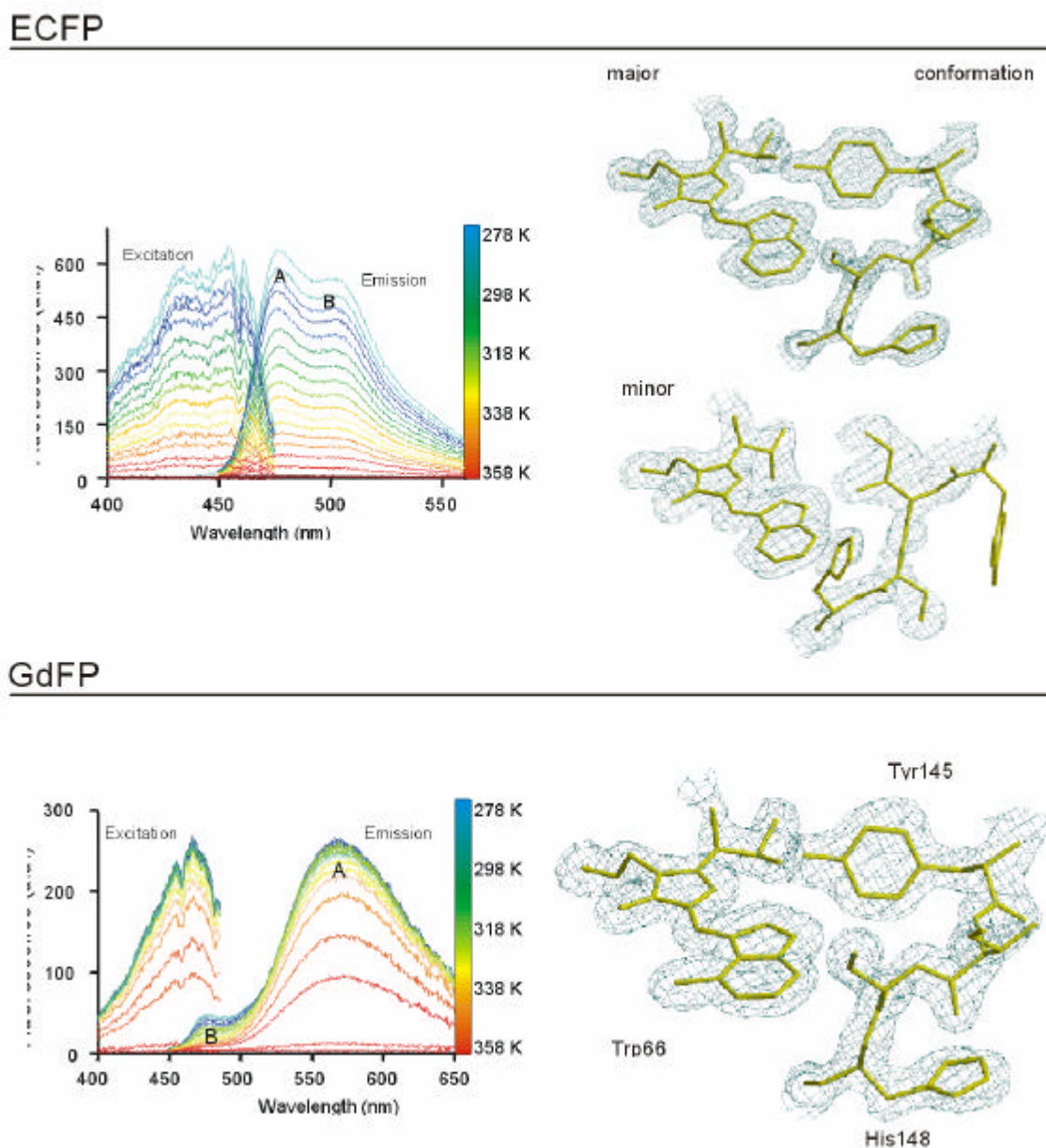
Figure 39. Correlation of structural, thermodynamic and fluorescence properties of ECFP and GdFP. In both proteins, peaks are assigned as **A** (stronger intensity) and **B** (weaker intensity; The shoulder at 480 nm (peak B) in the GdFP spectra is due to a ECFP contamination).

To explain the photophysical behavior of EGFP, the concept of photoisomerisation of the chromophore is introduced (Zimmer, 2002). There are three or four states of the chromophore in

equilibrium which have different photopysical properties. According to the three-state model, T203 has a key role for the property changes.

When the hydroxyl group of tyrosine is deprotonated at higher pH, this negatively charged oxygen can make the additional hydrogen bond with the rotated the hydroxyl group of Thr203. This additional contact of the EGFP chromophore is related to the unusual spectral property of the EGFP protein that the absorption spectrum has additional peak around 488 nm unlike non-fluorescent proteins (Figure 7). This absorption peak is emitted as the fluorescence at 507 nm. There is 19 nm wavelength shift due to the relaxation through the interactions such as hydrogen bonds (Zimmer, 2002).

In ECFP, there is no hydroxyl group in the chromophore. The chromophore can not make any hydrogen bonding network with the hydroxyl group of Thr203 in any pH range unlike that of EGFP. The X-ray crystal structure of ECFP shows that Thr203 is moved away from the chromophore.

The interactions between oxygen atom and p-electron system are well documented (Vargas, *et al.*, 2000, Burley & Petsko, 1988). This interaction occurs between the partially positive charged ($\delta+$) ring proton and the partially negative charged ($\delta-$) oxygen atom. They have to be positioned equatorial, which means in the same plane. But, in ECFP structure, indole rins anf the hydroxyl group of Thr203 are not located equatorially but axially. The orientation of two residues suggests that the interaction between the p-electron cloud ($\delta-$) and the partially positive charged ($\delta+$) hydrogen of the hydroxyl group of Thr203. The shortest distance between two residues, 4.3 Å, is quite reasonable to be considered interacting each other (Burley & Petsko, 1986). The modeling experiment shows that they are in contact (Figure 34). But the stabilisation energies of the hydrogen bond (3.6Kcal/mol) is larger than that of the between the p-electron cloud ($\delta-$) and the partially positive charged ($\delta+$) hydrogen of the hydroxyl group of Thr203 (less

than 2.5 Kcal/mol). The stabilizing effect of this interaction is relatively smaller than that of the hydrogen bond.

In GdFP (Figure 35), the chromophore is a bit shifted toward Phe165. This brings two residues in the range of 3.2 Å to 4.5 Å between amino group of the indole ring of chromophore and the carbon atoms in benzene ring of Phe165. When the amino group of indole ring is protonated, or its electron pair is participates in the resonance conformation, the sp^3-, or p-orbital of the amino group is located parallel to p-cloud of benzene ring. It can interact with the p-cloud of benzene ring of Phe165. It opens the route for the energy relaxation of the chromophore through the interaction. Energy relaxation results the red shift of the emission spectral property and reduction of quantum yield.

The indole ring of the ECFP chromophore has only one hydrogen bond between the hydroxyl group of Ser205 and the indole ring nitrogen. By this hydrogen bond and several hydrophobic interactions, the chromophore side ring is held. The EGFP has at least three hydrogen bonds between hydroxyl group of chromophore and surrounded residues to hold its chromophore side chain rigid. This brings that the ECFP chromophore side chain has more possibility thermodynamically to twist along with the exo-methylene axis of the chromophore than that of EGFP. The increased possibility of the internal movement of the chromophore, such as rotation and vibration, makes ECFP lose its quantum yield.

The charge separation is one of the facts reducing the quantum yield (Zimmer, 2002). The polarisability (1.44) of the amino group (NH_2) is greater than that (0.733) of the hydroxyl group (OH). It is interpreted that an amino group can be easily ionised than a hydroxyl group does. The GdFP chromophore has an amino group at the 4 position of the indole ring. This amino group of the indole ring in GdFP chromophore can be ionised easier than hydroxyl group in EGFP chromophore does. It can cause the charge separation in the GdFP chromophore while the

chromophore in ECFP can not make it due to the lack of the amino group. All possible resonance forms are drawn in Figure 38.

2.3.4   Aggregation Behaviour and Thermodynamic Stability of GdFP

It is observed that GdFP samples are relatively stable over long periods of time at room temperature as well as at +4 °C. For that reason aggregation tendencies of ECFP and GdFP as a function of time were examined closely, using sedimentation velocity measurements (Schuck, 2000). In fresh ECFP samples, the molecules are predominantly in the monomeric state (88%), while only a small fraction is dimerised (8%). The data from the sedimentation velocity experiments are in full agreement with the NMR-titration data reported recently (Seifert, *et al.*, 2002). The incubation of ECFP in buffered aqueous solutions at +4°C for six weeks resulted in a dramatic decrease of the monomeric species (44% monomers; 16 % dimers and the rest are higher aggregates). After 6 months under the same conditions all ECFP samples were almost completely aggregated. In contrast, there is almost no difference between freshly prepared and GdFP samples stored for 6 month at + 4°C (over 90 % monomers, data not shown, Dr. G. Wiegand).

It is widely believed that the intact hydrophobic patch, Ala206, Leu221 and Phe223, has to be responsible for dimerisation of GFPs β-barrels structures (Yang, *et al.*, 1996; Zacharias, *et al.*, 2002). Since this patch is unchanged in ECFP and GdFP, it is excluded as the origin of the vastly different aggregation behaviours of both proteins. Instead, the observed structural differences between ECFP and GdFP can offer a plausible model. In particular, the minor conformation of the ECFP chromophore environment which is not found in GdFP (with Tyr145 flipped outside and His148 moved closer to the chromophore) may be a decisive factor for the higher oligomerisation tendency.

Another striking difference between ECFP and GdFP is the strongly increased stability of fluorescence intensity against temperature in GdFP. In general, the fluorescence intensity in GFPs is a direct function of proper folding of protein matrix around the chromophore (Tsien, 1998; Chattoraj, *et al.*, 1996; Lossau, *et al.*, 1996). Obviously, GdFP is not only much more stable than ECFP, but also has much more co-operative unfolding (Figure 41). It is well known that the aromatic residues in proteins often form a network of mutually interacting aromatic side chains which globally stabilises the tertiary structure of the protein (Minks, *et al.*, 1999). Indeed, in the GFP structure such a cluster can be identified. It consists of the chromophore itself, Phe165, His148, and Tyr145 (Figure 41). These side chains are not only rigid in GdFP, but are also involved in hydrophobic contacts with the chromophore (Figure 32B). In addition, the newly incorporated amino group is involved in interaction with Phe165, which is unique among all known *av*FPs. Resonant stabilisation of the chromophore itself upon its amination stabilises this hydrophobic cluster. These interactions together are caused by the presence of a single amino group in the chromophore and are responsible for the large global effects on thermodynamics, low aggregation tendency and spectral properties of GdFP.

FUTURE PERSPECTIVES

1   Design of a Novel "Hydrid" Tanslation System that works with an Expanded Amino Acid Repertoire

It is widely believed that desired effects on GFP mutants such as a high extinction coefficients, quantum yields and efficient folding could mainly be achieved via multiple mutations in target sequences (Palm & Wlodawer, 1999). In contrast, GdFP clearly shows that a simple exchange of atomic groups i.e. **-NH₂ → -CH₃** might yield structurally identical but spectroscopically and thermodynamically different GFP mutants. This suggests the possibility for a rational strategy in designing GFPs with altered and improved fluorescence behaviour and increased stability of the monomeric state offering unprecedented opportunities to understand and manipulate relationships between protein structure and spectroscopic function. It would certainly be interesting to see whether future experiments on substitution with amino-Trp analogues on known red-emitting FP like *dsRed* will result in further red-shifted mutants. Moreover, new generations of red-shifted FPs are also conceivable with the expansion of the possibilities for introduction of the novel substituents like cyano, nitro, nitroso etc. in chromophores formed by indole, imidazole, phenol and even other aromatic rings. Using such substances as substrates for protein synthesis (i.e. an expanded amino acid repertoire) in combination with classical random and combinatorial mutational approaches might well represent a route toward novel families of tailor-made GFPs useful as multiple labels, reporters, resonance energy transfer acceptors and models for studies of light-induced chromophore dynamics

The fundamental problem of the genetic code engineering derives from the existence of the quality control mechanisms that are established at almost all levels of the genetic message transmission (Budisa, *et al.*, 1999; Wang & Schultz, 2002; Minks, *et al.*, 2000; Kiick, *et al.*,

2000). For example, by using the standard SPI-approach, it is possible to introduce only (4-NH_2)-Trp into *Aequorea* GFP proteins, but not (5-NH_2)-Trp nor hydroxy-Trp counterparts, although all these substances could be incorporated into barstar (Budisa, *et al.*, 2002). The reasons for such a behaviour are unclear, but might be due to certain still unknown editing mechanisms during the ribosomal synthesis or co-translational folding (Budisa, *et al.*, 1999). Similar findings were reported for thia-, and seleno-prolines, although they can be loaded on cognate tRNAs during translation (Budisa, *et al.*, 1998b).

While the design of GdFP will certainly serve as a model for the generation of novel GFP classes, its broader *in vivo* use as a marker for cellular biology is currently limited due to the necessity of special incorporation protocols. On the other hand, its intracellular synthesis can be assisted by the supplementation of the system with plasmids encoding modified translation components (aminoacyl-tRNA synthetases, tRNAs, etc.) which are compatible with the endogenous host translational apparatus.

The inability of a particular amino acid to be incorporated into target proteins by using simple approaches like SPI, is often an indication of the action of proofreading mechanisms during translation. In such cases, an additional component for the biosynthetic machinery is required, most probably a suitable aminoacyl-tRNA synthetase with a relaxed or switched substrate specificity or with deleted editing activity or even a novel enzyme:tRNA orthogonal pair. Although the modification of these components for *in vivo* use is expected to be difficult due to the already mentioned high intrinsic fidelity of the whole translation process, much hope in this research direction brings a recent progress made in design of "orthogonal" translation components (Wang & Schultz, 2002). These and other attempts to engineer a "hybrid" translation system with modified enzymes, tRNAs and other components compatible with endogenous host transitional machinery would certainly yield the possibilities to create, among others, novel GFP
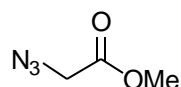
protein classes, variants and mutants with structural and functional diversification widely transcending those found in nature. It is therefore not so difficult to foresee a wider use of such method in molecular and cellular biology.

APPENDIX

1 Chemical Syntheses of Compounds


1.1 Synthesis of ß-3-Thieno[3,2-b]pyrrolyl-L-alanine

1.1.1 Methyl azidoacetate

Methyl bromoacetate (306 g, 2 mol) was dissolved in 60 ml ethanol and 180 ml water. Sodium azide (143 g, 2,2 mol) was added and the resulting solution was refluxed for 4 hours. After the reflux slightly yellowish solution was steam distiled, the organic phase in an seperation funnel seperated and stored over Na_2SO_4. Yield was 44 % (101g). the produst was confirmed by IR. IR: 2107cm^{-1}, 1748 cm^{-1}

1.1.2 Methyl 2-azido-3-(2'-thienyl)acrylate
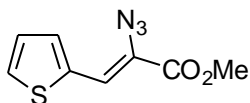
Sodium methoxide (1.01g, 18.7 mmol) was dissolved in dry methanol (1.0 M, 8.92 ml), and the solution was cooled at 0°C. A mixture of thiophen-2-carboxaldehyde (1.00 g, 8.92 mmol) and methyl azidoaceate (2.57 g, 22.3 mmol) was added dropwise. The reaction mixture was stiired at

0°C until the thin layer chromatography (TLC) indicated that all the aldehyde has been consumed (approx. 4hours). The reaction mixture was poured into iced-water and extracted with ethyl acetate (3 x 100 ml). The combined organic extract was washed with water, dried, evaporated and purified by column chromatography with silica gel (2:1 n-haxane/CH_2Cl_2) to yield 1.36 g (72.8 %) as light yellow crystals. HPLC (Gradient A) t_r 11.72 min; IR: 2117.05 cm^{-1}, 1707.90 cm^{-1}, 1611 cm^{-1}

1.1.3 5-Carbomethoxythieno[3,2-b]pyrrole (Murakami, *et al*, 1997)
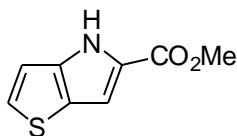
A solution of the methyl 2-azido-3-(2'-thienyl)acrylate (2.9 g, 13.9 mmol) in xylene (40 ml) was refluxed for 2 hours. The solvent was evaporated to yield 2.5 g (99%) as red/brown solid. HPLC (Gradient D) t_r 4.18 min; ^{1}H NMR (400 MHz, CDCl$_3$) d 7.50 (dd, J = 0.8, 5.1 Hz, 1H), 7.33 (d, J = 3.7 Hz, 1H), 7.17 (s, 1H), 7.07 (dd, J = 3.7, 5.1 Hz, 1H), 3.90 (s, 3H)

### 1.1.4 5-Carboxythieno[3,2-b]pyrrole (Wensbo, *et al*, 1995)
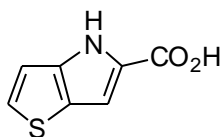
To 1.80 g (10.0 mmol) of 5-carbomethoxythieno[3,2-b]pyrrole in methanol was added potassium hydroxide (1.68 g, 30.0 mmol). The mixture was stirred at 40 ^{o}C until the TLC indicated the end of the saponification (approx. 10 hours). The solvent was evaporated, the residue dissolved in ether, and washed with 5 % KHSO$_4$. The organic phase was separated, dried over Na$_2$SO$_4$, and again evaporated to yield 1.42 g (85%) of 5-carboxythieno[3,2-b]pyrrole as brown solid. HPLC (Gradient B) t$_r$ 8.40 min; ESI-MS: m/z = 168.0 (M+H$^+$); Calcd. For C$_7$H$_5$N$_1$O$_2$S$_1$: 167.18

1.1.5 Thieno[3,2-b]pyrrole (Jones, 1993)
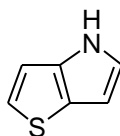
To a solution of 5-carboxythieno[3,2-b]pyrrole (366 mg, 2.19 mmol) and quinoline (2.5 ml) was added silver (24 mg, 0.22 mmol). The reaction mixture was heated at 220 ^{o}C for 1 hour under argon atmosphere, cooled, diluted in 50 ml ether and filtered off the solid. The filtrate was acidified with 3 N HCl, extracted withether, washed with NaHCO$_3$, and dried over Na$_2$SO$_4$. Evaporation of the solvent and chromatography (3:2 CH$_2$Cl$_2$/n-hexane) of the residue gave thieno[3,2-b]pyrrole (156 mg, 58%) as brown oil. HPLC (Gradient B) t$_r$ 9.04 min

1.1.6 ß-3-Thieno[3,2-b]pyrrolyl-L-alanine (Phillips, 1995)
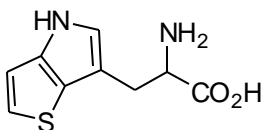
To a suspension of thieno[3,2-b]pyrrole (156 mg, 1.27 mmol), L-serine (200 mg, 1.9 mmol) and pyridoxal-5'-phosphate (0.5 mg, 1.9 µmol) in 60 ml 0.1 M potassium phosphate buffer (pH 7.8) was added 7 mg of the purified $a_2\beta_2$ tryptophan synthase complex. The reaction mixture was shaked gently under argon atmosphere at 37 ^{o}C until the HPLC indicated that all the thieno[3,2-b]pyrrole has been consumed (approx. 12 hours). After this time, the reaction mixture was filtered through a Centriprep-10 (Amicon) to remove the tryptophan synthase complex, and concentrated. The yield was calculated through the UV profile (UV (H$_2$O), λ_{max}= 261 nm (log e = 3.75) of ß-3-thieno[3,2-b]pyrrolyl-L-alanine (230 mg, 86%)). HPLC (Gradient B) t$_r$ 2.60 min; ESI-MS: *m/z* = 211.2 (M+H$^+$); Calcd. For C$_7$H$_5$N$_1$O$_2$S$_1$: 210.25

## 1.2 Synthesis of ß-3-Thieno[2,3-b]pyrrolyl-L-alanine

1.2.1 Methyl 2-azido-3-(3'-thienyl)acrylate
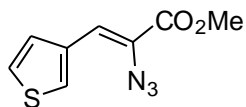
Sodium methoxide (5.06 g, 93.6 mmol) was dissolved in dry methanol (1.0 M, 44.6 ml), and the solution was cooled at 0°C. A mixture of thiophen-3-carboxaldehyde (5.00 g, 44.58 mmol) and methyl azidoaceate (7.70 g, 66.9 mmol) was added dropwise. The reaction mixture was stiired at 0°C until the thin layer chromatography(TLC) indicated that all the aldehyde has been consumed (approx. 4hours). The reaction mixture was poured into iced-water and extracted with ethyl acetate (3 x 100 ml). The combined organic extract was washed with water, dried, evaporated

and purified by column chromatography with silica gel (2:1 n-haxane/CH_2Cl_2) to yield 4.92 g (52.7 %) as light yellow crystals. HPLC (Gradient A) t_r 11.72 min

### 1.2.2 5-Carbomethoxythieno[3,2-b]pyrrole
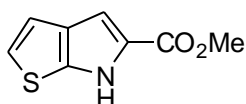
A solution of the methyl 2-azido-3-(3'-thienyl)acrylate (4.92 g, 23.5 mmol) in xylene (59 ml) was refluxed for 2 hours. The solvent was evaporated to yield 4.26 g (100 %) as red/brown solid. HPLC (Gradient A) t_r 9.85 min; ESI-MS: m/z = 182.0 (M+H$^+$); Calcd. For $C_8H_7N_1O_2S_1$: 181.21

1.2.3 5-Carboxythieno[2,3-b]pyrrole
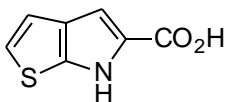
To 4.26 g (23.5 mmol) of 5-carbomethoxythieno[2,3-b]pyrrole in methanol was added potassium hydroxide (3.96 g, 70.5 mmol). The mixture was stirred at 40 ^{o}C until the TLC indicated the end of the saponification (approx. 10 hours). The solvent was evaporated, the residue dissolved in ether, and washed with 5 % $KHSO_4$. The organic phase was separated, dried over Na_2SO_4, and again evaporated to yield 2.81 g (71.5 %) of 5-carboxythieno[3,2-b]pyrrole as brown solid.

HPLC (Gradient C) t_r 6.67 min; ESI-MS: m/z = 168.2 (M+H$^+$); Calcd. For $C_7H_5N_1O_2S_1$: 167.18

1.2.4 Thieno[2,3-b]pyrrole
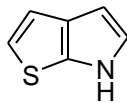
To a solution of 5-carboxythieno[2,3-b]pyrrole (480 mg, 2.87 mmol) and quinoline (7.18 ml) was added silver (22 mg, 0.20 mmol). The reaction mixture was heated at 220 °C for 1 hour under argon atmosphere, cooled, diluted in 50 ml ether and filtered off the solid. The filtrate was acidified with 3 N HCl, extracted with ether, washed with NaHCO₃, and dried over Na₂SO₄. Evaporation of the solvent and chromatography (3:2 CH₂Cl₂/n-hexane) of the residue gave thieno[2,3-b]pyrrole (290 mg, 82.2 %) as brown oil. HPLC (Gradient A) t_r 7.67 min; ESI-MS: m/z = 124.0 (M+H⁺); Calcd. For $C_6H_5N_1S_1$: 123.17

### 1.2.5 ß-3-Thieno[2,3-b]pyrrolyl-L-alanine
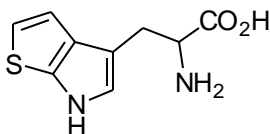
To a suspension of thieno[2,3-b]pyrrole (290 mg, 2.35 mmol), L-serine (371 mg, 3.53 mmol) and pyridoxal-5'-phosphate (187 mg, 705 µmol) in 100 ml 0.1 M potassium phosphate buffer (pH 7.8) was added 7 mg of the purified $a_2ß_2$ tryptophan synthase complex. The reaction mixture was shaked gently under argon atmosphere at 37 °C until the HPLC indicated that all the thieno[2,3-b]pyrrole has been consumed (approx. 12 hours). After this time, the reaction mixture was filtered through a Centriprep-10 (Amicon) to remove the tryptophan synthase complex, and concentrated. The yield was calculated through the UV profile (UV (H₂O), λ_{max} = 242 nm (log e =

3.69) of ß-3-thieno[2,3-b]pyrrolyl-L-alanine (428 mg, 86.6 %)). HPLC (Gradient B) t_r 2.60 min;

ESI-MS: *m/z* = 211.0 (M+H⁺); Calcd. For $C_9H_{10}N_2O_2S_1$: 210.25

## 1.3 Synthesis of ß-3-Selenolo[3,2-b]pyrrolyl-L-alanine

1.3.1 Selenophene-2-carboxaldehyde (Archer, W. J. *et al*, 1983)
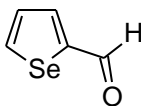
To a solution of selenophene (2.97 g, 15.77 mmol) in THF (1.0 M, 15.8 ml) nBuLi (10.8 ml, 17.35 mmol) was added dropwise at 0°C for 1 hour. The suspention was stirred for an additional hour at 0°C. DMF was added to the reaction mixture at 0°C, and stirred for one hour. The reaction mixture was poured into 25 ml of iced-water, and acidified with diluted HCl solution. Reaction mixture was extracted twice with 20 ml of ether, dried over Na_2SO_4, filtered, and evaporated to get yellow oil (2.48 g, 98.9 %). HPLC (Gradient C) t_r 5.07 min; ESI-MS: *m/z* = 161.0 (M+H⁺); Calcd. For $C_5H_4O_1Se_1$: 159.05
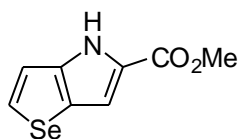
1.3.2 Methyl 2-azido-3-(2'-selenyl)acrylate
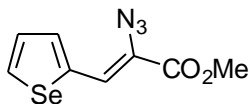
Sodium methoxide (714 mg, 13.2 mmol) was dissolved in dry methanol (1.0M, 6.29 ml), and the solution was cooled at 0°C. A mixture of selenophen-2-carboxaldehyde (1.00 g, 6.29 mmol) and

methyl azidoaceate (2.90 g, 25.2 mmol) was added dropwise. The reaction mixture was stiired at

0^oC until the thin layer chromatography(TLC) indicated that all the aldehyde has been consumed

(approx. 4hours). The reaction mixture was poured into iced-water and extracted with ethyl

acetate (3 x 100 ml). The combined organic extract was washed with water, dried, evaporated

and purified by column chromatography with silica gel (2:1 n-haxane/CH_2Cl_2) to yield 1.40 g (

87.0 %) as light yellow crystals. HPLC (Gradient C) t_r 10.17 min; ESI-MS: m/z = 230.0 (M-N$_2$);

Calcd. For $C_8H_7N_3O_2Se_1$: 256.12
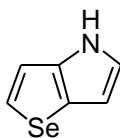
1.3.3 5-Carbomethoxyselenolo[3,2-b]pyrrole

A solution of the methyl 2-azido-3-(2'-selenyl)acrylate (400 mg, 1.56 mmol) in xylene (7.80 ml)

was refluxed for 2 hours. The solvent was evaporated to yield 355 mg (100 %) as red/brown

solid. HPLC (Gradient C) t_r 8.07 min; ESI-MS: m/z = 229.8 (M+H$^+$); Calcd. For $C_8H_7N_1O_2Se_1$:

228.11

### 1.3.4 5-Carboxyselenolo[3,2-b]pyrrole
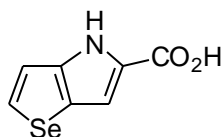
To 6.40 g (28.1 mmol) of 5-carbomethoxyselenolo[3,2-b]pyrrole in methanol was added potassium hydroxide (4.73 g, 84.3 mmol). The mixture was stirred at 40 ^{o}C until the TLC indicated the end of the saponification (approx. 10 hours). The solvent was evaporated, the residue dissolved in ether, and washed with 5 % KHSO$_4$. The organic phase was separated, dried over Na$_2$SO$_4$, and again evaporated to yield 4.69 g (77.9 %) of 5-carboxyselenolo[3,2-b]pyrrole as brown solid. HPLC (Gradient C) t$_r$ 6.50 min; ESI-MS: *m/z* = 216.0 (M+H$^+$); Calcd. For C$_7$H$_5$N$_1$O$_2$Se$_1$: 214.08

1.3.5 Selenolo[3,2-b]pyrrole

To a solution of 5-carboxyselenolo[3,2-b]pyrrole (375 mg, 1.75 mmol) and quinoline (4.38 ml) was added silver (13 mg, 0.12 mmol). The reaction mixture was heated at 220 ^{o}C for 1 hr under argon atmosphere, when cool diluted in 50 ml ether and the solid filtered off. The filtrate was extracted with 3N HCl, NaHCO$_3$, and dried over Na$_2$SO$_4$. Evaporation of the solvent and chromatography (3:2 CH$_2$Cl$_2$/n-hexane) of the residue gave Selenolo[3,2-b]pyrrole (250 mg, 83.9 %) as brown oil. HPLC (Gradient A) t$_r$ 8.09 min; ESI-MS: *m/z* = 172.0(M+H$^+$); Calcd. For C$_6$H$_5$N$_1$Se$_1$: 170.07

### 1.3.6 ß-3-Selenolo[3,2-b]pyrrolyl-L-alanine
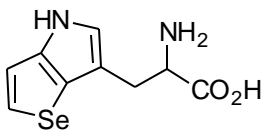
To a suspension of selenolo[3,2-b]pyrrole (120 mg, 0.706 mmol), L-serine (111 mg, 1.06 mmol) and pyridoxal-5'-phosphate (58 mg, 0.22 mmol) in 50 ml 0.1 M potassium phosphate buffer (pH 7.8) was added 7 mg of the purified $a_2\beta_2$ tryptophan synthase complex. The reaction mixture was shaked gently under argon atmosphere at 37 °C until the HPLC indicated that all the selenolo[3,2-b]pyrrole has been consumed (approx. 12 hours). After this time, the reaction mixture was filtered through a Centriprep-10 (Amicon) to remove the tryptophan synthase complex, and concentrated. The yield was calculated through the UV profile (UV (H_2O), λ_{max} = 267 nm (log e = 4.11) of ß-3-selenolo[3,2-b]pyrrolyl-L-alanine (145 mg, 79.7 %)). HPLC (Gradient B) t 2.52 min; ESI-MS: m/z = 259.0 (M+H$^+$); Calcd. For $C_9H_{10}N_2O_2Se_1$: 257.15

1.4   Synthesis of ß-3-Selenolo[2,3-b]pyrrolyl-L-alanine

1.4.1   2,3,5-Tribromoselenophene (Keegstra, *et al*, 1988)
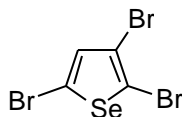
To a suspension of selenophene (5.00 g, 38.16 mmol), and hydrobromic acid (10.81 g, 133.6 mmol) was added H_2O_2 (3.89 g, 114.5 mmol) dropwise at 0°C. The mixture was additionally stirred for 30 minutes, and heated to remove excess HBr. The organic layer was separated from water, and extracted with ethyl acetate (10 ml x 2). The organic layer was combined, dried over Na_2SO_4, filtered, and evaporated to yield the brown oil. This was used further without characterisation. HPLC (Gradient E) t_r 10.85 min

1.4.2 **3-Bromoselenophene** (Hallberg, *et al*, 1981)
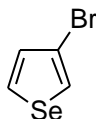
To the suspension of 2,3,5-tribromoselenophene in H_2O (1.2 M, 31.8 ml) were added zinc powder (9.975 g, 152.6 mmol) and acetic acid (9.164 g, 152.6 mmol), and stirred at 100°C overnight. The reaction mixture was filtered, and washed with diluted HCl solution. The organic layer was extracted with CH_2Cl_2 (20 ml x 2), dried over Na_2SO_4, filtered, evaporated, and distilled with reduced pressure to yield the brown oil (4.50 g, 56.2 %). This was used without characterisation. HPLC (Gradient E) t_r 8.39 min

1.4.3 **Selenophene-3-carboxylic acid** (Gronowitz, *et al*, 1972)
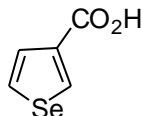
To a suspension of 3-bromoselenophene (2.432 g, 11.58 mmol) in dry ether (1.0 M, 11.58 ml) was added nBuLi (7.96 ml, 12.74 mmol) dropwise at –70°C. The mixture was stirred for 20 minutes. The excess of powdered dry ice was added into the mixture, and stirred for a while. The reaction was quenched with water. The reaction mixture was washed with ether twice to remove any unreacted 3-bromoselenophene. The water layer was acidified with HCl, extracted with ether, dried over Na_2SO_4, filtered, and evaporated to yield yellow solid (2.08 g, 99.1 %). HPLC (Gradient E) t_r 5.92 min; FAB-MS: *m/z* = 175.0 (M-H$^+$); Calcd. For $C_5H_4O_2Se_1$: 175.05; ^{1}H NMR (400 MHz, CDCl$_3$) d 9.05 (dd, *J* = 0.9, 2.5 Hz, 1H), 8.01 (dd, *J* = 2.5, 5.4 Hz, 1H), 7.85 (dd, *J* = 0.9, 5.4 Hz, 1H)

1.4.4 3-Carbomethoxyselenophene (Hedlicky, 1971)
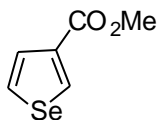
Diazomethane was prepared following the work of Hudlicky (Hudlicky, 1971). To a suspension of selenophene-3-carboxylic acid (1.966 g, 11.23 mmol) in dry ether (1.0 M, 11 23 ml) was added diazomethane in ether (10.04 mmol) dropwise at 0^oC. Solvent was evaporated to yield the yellow oil (2.106 g, 99.2 %). HPLC (Gradient E) t_r 6.55 min; ^{1}H NMR (400 MHz, CDCl$_3$) d 8.89 (dd, J = 1.1, 2.5 Hz, 1H), 7.97 (dd, J = 2.5, 5.5 Hz, 1H), 7.80 (dd, J = 1.1, 5.5 Hz, 1H), 3.86 (s, 3H)

1.4.5 3-(Hydroxylmethyl)-selenophene
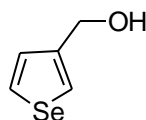
To the suspension of 3-carbomethoxyselenophene (2.106 g, 11.14 mmol) in dry ether (0.2 M, 55.7 ml) was added lithium aluminium hydride (LAH) at 0^oC. The mixture was stirred at room temperateure for an hour. The reaction was quenched by adding the water dropwise until LAH was deactivated. The mixture was filtered, and evaporated the solvent to yield the yellow oil quantitatively. HPLC (Gradient E) t_r 3.23 min; ^{1}H NMR (400 MHz, CDCl$_3$) d 8.00 (ddd, J = 0.3, 2.5, 5.4 Hz, 1H), 7.87 (m, 1H), 7.36 (ddd, J = 0.3, 1.2, 5.4 Hz, 1H), 4.64 (dd, J = 0.3, 0.8 Hz, 1H)

1.4.6 Selenophene-3-carboxaldehyde (Marx, *et al*, 1984)
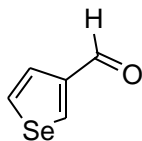
To the suspension of oxalyl chloride (2.121 g, 16.71 mmol) in dry CH_2Cl_2 (30 ml) was added dimethylsulfoxide (DMSO; 2.611 g, 33.42 mmol), and stirred for 30 minutes at -78^oC. 3-(Hydroxylmethyl)-selenophene (1.794 g, 11.14 mmol) in dry CH_2Cl_2 (26 ml) was added, and stirred for 30 minutes at -78^oC. To the reaction mixture was added triethylamine (4.509 g, 44.56 mmol), and stirred for 15 minutes at -78^oC. The mixture was warmed up to room temprature, and stirred for an hour. The reaction was quenched by adding water. The organic layer was separated, and the aqueous layer was extracted with CH_2Cl_2 (20 ml x 2). The organic layer was combined, dried over Na_2SO_4, filtered, and evaporated to yield yellow oil (1.741 g, 98.3 %). HPLC (Gradient E) t_r 4.50 min; ^1H NMR (400 MHz, $CDCl_3$) d 9.88 (m, 1H), 8.93 (dd, J = 1.2, 2.4 Hz, 1H), 8.05 (ddd, J = 0.9, 2.4, 5.5 Hz, 1H), 4.81 (ddd, J = 0.4, 1.2, 5.5 Hz, 1H)

1.4.7 Methyl 2-azido-3-(3'-selenienyl)acrylate
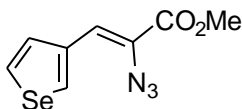
Sodium methoxide (1.183 g, 21.90 mmol) was dissolved in dry methanol (1.0 M, 10.95 ml), and the solution was cooled at 0^oC. A mixture of selenophen-3-carboxaldehyde (1.741 g, 10.95 mmol) and methyl azidoaceate (5.041 g, 43.80 mmol) was added dropwise. The reaction mixture was stiired at 0^oC until the thin layer chromatography (TLC) indicated that all the aldehyde has

been consumed (approx. 4hours). The reaction mixture was poured into iced-water and extracted with ethyl acetate (3 x 100 ml). The combined organic extract was washed with water, dried, evaporated and purified by column chromatography with silica gel (2:1 n-haxane/CH_2Cl_2) to yield 2.480 g (88.4 %) as light yellow crystals. HPLC (Gradient E) t_r 12.00 min; ^{1}H NMR (400 MHz, $CDCl_3$) d 8.61 (m, 1H), 7.97 (dd, J = 2.5, 5.5 Hz, 1H), 7.75 (dd, J = 1.1, 5.5 Hz, 1H), 6.93 (s, 1H), 3.89 (s, 3H)

1.4.8 5-Carbomethoxyselenolo[2,3-b]pyrrole
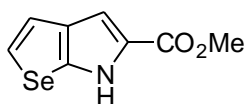
A solution of the methyl 2-azido-3-(3'-selenyl)acrylate (2.225 g, 8.687 mmol) in xylene (23 ml) was refluxed for 2 hours. The solvent was evaporated to yield 1.636 g (82.5 %) as red/brown solid. HPLC (Gradient E) t_r 9.79 min; ESI-MS: m/z = 230.00 (M+H$^+$); Calcd. For $C_8H_7N_1O_2Se_1$: 228.11

1.4.9 5-Carboxyselenolo[2,3-b]pyrrole
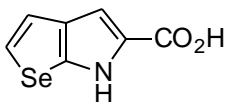
To 2.01 g (9.68 mmol) of 5-carbomethoxyselenolo[2,3-b]pyrrole in methanol was added potassium hydroxide (1.63 g, 29.1 mmol). The mixture was stirred at 40 ^{o}C until the TLC

indicated the end of the saponification (approx. 10 hours). The solvent was evaporated, the residue dissolved in ether, and washed with 5 % KHSO$_4$. The organic phase was separated, dried over Na$_2$SO$_4$, and again evaporated to yield 1.33 g (81.6 %) of 5-carboxyselenolo[3,2-b]pyrrole as brown solid. HPLC (Gradient E) t$_r$ 7.79 min; ESI-MS: m/z = 216.00 (M+H$^+$); Calcd. For C$_7$H$_5$N$_1$O$_2$Se$_1$: 214.08; ^{1}H NMR (400 MHz, CDCl$_3$) d 7.46 (d, J = 5.8 Hz, 1H), 7.14 (d, J = 5.8 Hz, 1H), 7.06 (s, 1H)

1.4.10 Selenolo[2,3-b]pyrrole
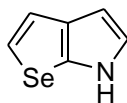
To a solution of 5-carboxyselenolo[2,3-b]pyrrole (1.03 g, 4.81 mmol) and quinoline (12 ml) was added silver (36 mg, 0.35 mmol). The reaction mixture was heated at 220 ^{o}C for 1 hour under argon atmosphere, cooled, diluted in 50 ml ether and filtered off the solid. The filtrate was acidified with 3 N HCl, extracted with ether, washed with NaHCO$_3$, and dried over Na$_2$SO$_4$. Evaporation of the solvent and chromatography (3:2 CH$_2$Cl$_2$/n-hexane) of the residue gave selenolo[2,3-b]pyrrole (480 mg, 58.7 %) as brown oil. HPLC (Gradient E) t$_r$ 12.32 min; ESI-MS: m/z = 172.00 (M+H$^+$); Calcd. For C6H$_5$N$_1$Se$_1$: 170.07

1.4.11 ß-3-Selenolo[2,3-b]pyrrolyl-L-alanine
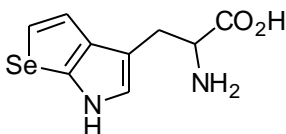
To a suspension of selenolo[2,3-b]pyrrole (480 mg, 2.82 mmol), L-serine (445 mg, 4.23 mmol) and pyridoxal-5'-phosphate (224 mg, 846 μmol) in 100 ml 0.1 M potassium phosphate buffer (pH 7.8) was added 7 mg of the purified $a_2\beta_2$ tryptophan synthase complex. The reaction mixture was shaked gently under argon atmosphere at 37 °C until the HPLC indicated that all the selenolo[2,3-b]pyrrole has been consumed (approx. 12 hours). After this time the reaction mixture was filtered through a Centriprep-10 (Amicon) to remove the tryptophan synthase complex, and concentrated. The yield was calculated through the UV profile (UV (H_2O), λ_{max} = 248 nm (log e = 3.79) of ß-3-selenolo[2,3-b]pyrrolyl-L-alanine (300 mg, 41.4 %)). HPLC (Gradient E) t_r 2.50 min; ESI-MS: *m/z* = 259.00 (M+H$^+$); Calcd. For $C_9H_{10}N_2O_2Se_1$: 257.15

References

Archer, W. J., Cook, R. & Taylor, R. (1983). Eletrophilic aromatic substitution. Part 34. Partial factors for detritiation of dithieno[1,2-b:4,3-b']benzene, dithieno[1,2-b:3,4-b']benzene, and dithieno[2,1-b:3,4-b']benzene. *J. Chem. Soc. Perkin Trans. II.* 813-819

Bae, J. H., Alefelder, S., Kaiser, J. T., Friedrich, R., Moroder, L., Huber, R. & Budisa, N. (2001). Incorporation of beta-selenolo[3,2-b]pyrrolyl-alanine into proteins for phase determination in protein X-ray crystallography. *J. Mol. Biol.* **309(4)**. 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. & Budisa, N. (2002). Expansion of the Genetic Code Enables Design of a Novel "Gold" Class of Green Fluorescent Proteins. *Nature Biotechnol.* submitted

Barbey, G., Dian, G., Merlet, N., Outurquin, F. & Paulmier, C. (1989). Direct synthesis of 3 methylselenophene and 3,4-dimethylselenophene. *Synthesis* **12**. 181-182

Berendes, R., Voges, D., Demange, P., Huber, R. & Burger, A. (1993). Structure-function analysis of the ion channel selectivity filter in human annexin V. *Science* **262**. 427-430

Bernal, J. D. & Growfoot, D. C. (1934). X-ray photograph of crystalline pepsin. *Nature* **133**. 794-795

Bird, C.W. (1985). A new aromaticity index and its application to five-member ring heterocycles. *Tetrahedron* **41.** 1409-1414

Blair, J., Marona-Lewicka D., Kanthasamy, A., Lucaites, V., Nelson, D. & Nichols, D. (1999). Thieno[3,2-*b*]- and Thieno[2,3-*b*]-pyrrole bioisosteric analogues of the hallucinogen and serotonin agonist N,N-Dimethyltrytamine. *J. Med. Chem.* **42.** 1106-1111

Blow, D. H. & Crick, F. H. C. (1959). The treatment of errors in the isomorphous replacement method. *Acta Cryst.* **12**. 794-802

Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography* Academic Press, London

Boggon, T. J. & Shapiro, L. (2000). Screening for phasing atoms in protein crystallography. *Structure* **8**. 143-149

Bragg, W. L. & Bragg, W. H. (1913). The diffraction of X-rays by crystals. *Proc. Roy. Soc.* **A88**. 428-438

Breitung, E. M., Shu, C. F. & McMahon, R. J. (2000). Thiazole and thiophene analogues of donor-acceptor stilbenes: Molecular hyperpolarizibilities and structure-property relationships. *J. Am. Chem. Soc.* **122**. 1154-1160

Bruchez Jr, M., Moronne, M., Gin, P., Weiss, S., Alivisatos, A. P. (1998). Semiconductor Nanocrystals as Fluorescent Biological Labels. *Science* **281**. 2013-2016

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

Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J. & Huber, R. (1995). High level biosynthetic substitution of methionine in proteins by its analogues 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in *Escherichia coli*. *Eur. J. Biochem.* **230**. 788-796

Budisa, N., Karnbrock, W., Steinbacher, S., Humm, A., Prade, L., Moroder, L. & Huber, R. (1997). Bioincorporation of telluromethionine into proteins: A promising new approach for X-ray structure analysis of proteins. *J. Mol. Biol.* **270**. 616-23

Budisa, N., Huber, R., Golbik, R., Minks, C., Weyher, E. & Moroder, L. (1998a). Atomic mutations in annexin V. Theromdynamic study of isomorphous protein variants. *Eur. J. Biochem.* **53.** 1-9

Budisa, N., Minks, C., Medrano, F. J., Lutz, J., Huber, R. & Moroder, L. (1998b). 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. *Proc. Natl. Acad. Sci. USA* **95**. 455-459

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

Budisa, N., Moroder, L. & Huber, R. (1999b). 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. & Moroder, L. (2001). Proteins with β−(Thienopyrrolyl)alanines as alternate chromophores and pharmaceutically active amino acids. *Prot. Sci.* **10**. 1281-1292

Budisa, N., Rubini, M., Bae, J. H., Weyher, E., Wenger, W., Golbik, R., Huber, R., & Moroder, L. (2002). Global replacement of tryptophan with aminotryptophans generates non-invasive protein-based optical pH sensors. *Angew. Chem. Int. Ed.* In press.

Burger, A. (1991). Isosterism and bioisosterism in drug design. *Prog. Drug. Res.* **37.** 287-381

Burley, S. K. & Petsko, G. A. (1986). Amino-aromatic interactions in proteins. *FEBS Lett.* **203**. 139-143

Burley, S. K. & Petsko, G. A. (1988). Weakly polar interactions in proteins. *Adv. Protein Chem* **39**. 125-189

REFERENCES

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W. & Prasher, D.C. Green Fluorescent protein as a marker for gene-expression. *Science* **263**. 802-805

Chan, W. C. W. & S. Nie. (1998). Quantum Dot Bioconjugates for Ultrasensitive Nonisotopic Detection. *Science* **281**. 2016-2018

Chapeville, F., Lipmann, F., von Ehrenstein, G., Weisblum, B., Ray, W. J. & Benzer, S. (1962). The Secondary Structure of Complementary RNA. *Proc. Natl. Acad. Sci. USA* **48**. 1086-1092

Chattoraj, M., King, B. A., Bublitz, G. U. & Boxer, S. G. (1996). Ultra-fast excited state dynamics in green fluorescent protein: Multiple states and proton transfer. *Proc. Natl. Acad. Sci. USA.* **93**. 8362-8367

COLLABORATIVE COMPUTATIONAL PROJECT (1994). The CCP4 Suite: Programs for Protein Crystallography. *Acta Cryst.* **D50**. 760-763

Concha, N. O., Head, J. F., Kaetzel, M. A., Dedman, J. R. & Seaton, B. A. (1993). Rat annexin V crystal structure: Ca^{2+}-induced conformational changes. *Science* **261**. 1321-1324

Cowie, D. B. & Cohen, G. N. (1957). Biosynthesis by *Escherichia coli* of active altered proteins containing selenium instead of sulphur. *Biochim. Biophys. Acta* **26**. 252-261

Crick, F. H. C. (1966). The genetic code- yesterday, today, and tomorrow. *Cold Spring Harbor Symp. Quant. Biol.* **31**. 3-9

Cromer, D. T. & Leberman, D. (1970). Relativistic calculation of anomalous scattering factors for X-rays. *J. Chem. Phys.* **53**. 1891-1898

Dayhoff, M. O. (1972). Atlas of protein sequence and structure. Vol. 5. National Biomedical Research Foundation, Washington D. C.

Dill, K.A.& Shortle, D. (1991). Denatured states of proteins. *Annu. Rev. Biochem.* **60.** 795-825

Dougherty, D. A. (1996). Cation-π interactions in chemistry and biology: A new view of benzene, Phe, Tyr and Trp. *Science* **271**. 163-168

Drenth, J. (1994). *Principles of Protein X-ray Crystallography*. Springer Verlag. New York

Engelhardt, D. L., Webster, R. E., Wilhelm, R. C. & Zinder, N. (1965). In vitro studies on the mechanism of suppression of a nonsense mutation. *Proc. Natl. Acad. Sci. USA* **54**. 1791-1797

Esnouf, R. M. (1997). An extensively modified version of Molscript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.* **15**. 132-134

Ewald, P. P. (1921). Das "reziproke Gitter" in der Strukturtheorie. *Z. Kristallogr.* **56**. 129-156

Garen, A. (1968). Sense and nonsense in the genetic code. Three exceptional triplets can serve as both chain-terminating signals and amino acid codons. *Science* **160**. 149-159

Giulliet, V., Lapthorn, A., Hartley, R. W. & Maugen, Y. (1993). Recognition between a bacterial ribonuclease, barnase, and its natural inhibitor, barstar. *Structure* **1**. 165-177

Golbik, R., Fischer, G. & Fersht, A. R. (1999). Folding of barstar C40/C82A/P27A and catalysis of the peptydil-prolyl *cis/trans* isomerisation by human cytosolic cyclophilin (Cyp18). *Prot. Sci.* **8**. 1505-1514

Gopalan, V., Golbik, R., Schreiber, G., Fersht, A. R. & Altman, S. (1997). Fluorescence properties of a Trp residue in an aromatic core of the protein subunit of ribonuclease P from *Escherichia coli*. *J. Mol. Biol.* **267**. 765-769

Green, D. W., Ingram, V. M. & Perutz, M. F. (1954). The structure of Hemoglobin. IV. Sign determination by the isomorphous replacement method. *Proc. Roy. Soc.* **A225**. 287-307

Grissom, J. W., Klingberg, D., Meyenburg, S., & Stallman, B. L. (1994). Enediyne- and tributyltin hydride-mediated aryl radical additions onto various radical acceptors. *J. Org. Chem.* **59**. 7876-7888

Gronowitz, S. & Frejd, T. (1972). Ring-opening reactions of heterocyclic metal-organics. III. A comparison of 3-selenienyllithium and 3-thienyllithium. *Int. J. Sulfur Chem.* **A2.** 165-167

Gronowitz, S., & Frejd, T. (1976). On the side-chain bromination of 3-methylselenophene and 3-methylthiophene. *Syn. Comm.* **6**. 475-480

Gronowitz, S., Westerlund, C., and Hörnfeldt, A.B. (1976). The synthetic utility of heteroaromatic azido compounds: Preparation of some Furo-, Thieno-, and selenolo[3,2-b]pyrroles. *Acta Chem. Scand.* **B30**. 391-395

Guex, N. & Peitsch, M. C. (1996). Swiss-PdbViewer: A fast and easy-to-use PDB viewer for Macintosh and PC. *PDB Quart. Newslett.* **77**. 7-10

Hallberg, A., Liljefors, A. & Pedaja, P. (1981). A simplified synthesis of 3-bromoselenophene and some 3-bromothiophenes. *Syn. Comm.* **11**. 25-28

Harker, D. (1956). The determination of the phases of the structure factors on non-centrosymmetric crystals by the method of double isomorphous replacement. *Acta Cryst.* **9**. 1-9

Hecht, S. M., Alford, B. L., Kuroda, Y. & Kitano, S. (1978). "Chemical aminoacylation" of tRNA's. *J. Biol. Chem.* **253**. 4517-4520

Hendrickson, W. A., Smith, J. L., Phizackerley, R. P. & Merrit, E. A. (1988). Crystallographic structure analysis of lamprey hemoglobin from anomalous dispersion of synchrotron radiation. *Proteins*. **4**. 77-88

Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **9**. 1665-1672

Hendrickson, W. A. (1991). Determination of the macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**. 51-58

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

Hohsaka, T., Ashizuka, Y., Murakami, H. & Sisido, M. (1996). Incorporation of nonnatural amino acids into streptavidin through *in vitro* frame-shift suppression. *J. Am. Chem. Soc.* **118**. 9778-9779

Hohsaka, T., Kajihara, D., Ashizuka, Y., Murakami, H. & Sisido, M. (1999). Efficient incorporation of nonnatural amino acids with large aromatic groups into streptavidin in in vitro protein synthesizing systems. *J. Am. Chem. Soc.* **121**. 34-40

Hoppe, W. (1957a). Die Faltmolekülmethode und ihre Anwendung in der Röntgenorgaphischen Konstitutionanalyse von Biflorin. *Z. Elektrochem.* **61**. 1076-1083

Hoppe, W. (1957b). Die Faltmolekülmethode: eine neue Methode zur Bestimmung der Kristallstruktur bei ganz oder teilweise bekannten Molekülstrukturen. *Acta Cryst.* **10**. 750-751

Huber, R. (1965). Die Automatisierte Faltmolekülmethode. *Acta Cryst.* **A19**. 353-356

Huber, R., Römisch, J. & Paques, E. P. (1990) The crystal and molecular structure of human annexin V, an anticoagulant protein that binds to calcium and membranes. *EMBO J.* **9**. 3867-3874

Hudlicky, M. (1971). An improved apparatus for the laboratory preparation of diazomathane. *J. Org. Chem.* **45**. 5377-5378

Hutchison, 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

Jones, G. B. & Chapman, B. J. (1993). Decarboxylation of indole-2-carboxylic acids: Improved procedures. *J. Org. Chem.* **58**, 5558-5559

Jones, T. A. & Kjelgaard (1991). O-The Manual, Uppsala, Sweden

Keegstra, M. A. & Brandsma, L. (1988). Convenient high-yield procedures for 2-bromothiophene and 2,5-dibromothiophene. *Synthesis* **11**. 890-891

Kendrew, D. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C. & Shore, V. C. (1960). Structure of myoglobin: A three-dimensional fourier synthesis at 2 Å resolution. *Nature* **185**. 422-427

Koide, H., Yokoyama, S., Kawai, G., Ha, J.M., Oka, T., Kawai, S., Miyake, T., Fuwa, T. & Miyazawa, T. (1988). Biosynthesis of a protein containing a nonprotein amino acid by *Escherichia coli. Proc. Natl. Acad. Sci. USA.* **85.** 6237-6241

Koradi, R., Billeter, M. & Wuthrich, K. (1996). MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graphics.* **14**. 51-55

Kowal, A. K. & Oliver, J. S. (1997). Exploiting unassigned codons in Micrococcus luteus for tRNA-based amino acid mutagenesis. *Nucleic Acids Res.* **25**. 4685-4689

Kothakota, S., Mason, T., Tirrell, D.A. & Fournier, M.J. (1995). Biosynthesis of a periodic protein containing 3-thienylalanine: a step toward genetically engineered conducting polymers. *J. Am. Chem. Soc.* **117.** 536-537

Köhrer, C., Xie, L., Kellerer, S., Varshney, U. & RajBhandary, U. L. (2001). Import of amber and ochre suppressor tRNAs into mammalian cells: A general approach to site-specific insertion of amino acid analogues into proteins. *Proc. Natl. Acad. Sci. USA.* **98**. 14310-14315

Kraulis, P. (1991). MOLSCRIPT: A program to produce both detailed and schematic plots of proteins. *J. Appl. Cryst.* **24**. 946-950

Kulik, W., Verkruijsse, H. D., de Jong, R. L. P., Mommes, H. & Brandsma, L. (1983). Dimetalation of isopropenylacetylene. Application in the synthesis of 3-methylselenophene, 3-methylene-2,3-dihydroselenophene and the tellurium analogues. *Tetrahedron Lett.* **24**. 2203-2204

Lee, E., Wu, Y., Xia, G., Schultz, P.G. & Romesberg, F. (2001). Efforts toward Expansion of the Genetic Alphabet: Replication of DNA with Three Base Pairs. *J. Am. Chem. Soc.* **123**. 7439-7440

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

Liemann, S. & Huber, R. (1997). Three-dimensional structure of annexins. *Cell. Mol. Life. Sci.* **53**. 516-521

Liu, D. R. & Schultz, P. G. (1999). Progress toward the evolution of an organism with an expanded genetic code. *Proc. Natl. Acad. Sci. USA* **96**. 4780-4785

Lossau, H., Kummer, A., Heinecke, R., Pollinger-Dammer, F., Kompa, C., Bieser, G., Jonsson, T., Silva, C. M., Yang, M. M., Youvan, D. C. & Michel-Beyerle, M. E. (1996). Time-resolved spectroscopy of wild-type and mutant Green Fluorescent Proteins reveals excited state deprotonation consistent with fluorophore-protein interactions. *Chem. Phys.* **58**. 1-16

Martin, C., Hartley, R. W. & Mauguen, Y. (1999). X-ray structural analysis of compensating mutations at the barnase-barstar interface. *FEBS Lett.* **452**. 128-132

Marx, M. & Tidwell, T. T. (1984). Reactivity-selectivity in the Swern oxidation of alcohols using dimethyl sulfoxide-oxalyl chloride. *J. Org. Chem.* **49**, 788-793

Matz, M. V., Fradkov, A. F., Labas, Y. A., Savistky, A. P., Zaraisky, A. G., Markelov, M. L. & Lukyanov. S. A. (1999). Fluorescent proteins from nonbioluminescent Anthozoa species. *Nature Biotechnol.* **17**. 969-973

McCaul, C. P. & Ludescher, R. D. (1999). Room temperature phosphoroscence from tryptophan halogenated analogs in amorphous sucrose. *Photochem. Photobiol.* **70**. 166-171

McPherson, A. (1990). Current approaches to macromolecular crystallization. *Eur. J. Biochem.* **189**. 1-23

McRee, D. E. (1993). *Practical Protein Crystallography.* Academic Press Inc., San Diego

Merritt, E. A. & Bacon, D. J. (1997). Raster3D: Photorealistic molecular graphics. *Meth. Enzymol.* **277**. 505-524

Miles, E. W., Kawasaki, H., Ahmed, S. A., Morita, H., Morita, H., & Nagata, S. (1989). The β subunit of trzptophan aynthase. *J. Biol. Chem.* **264**. 6280-6287

Milun, M. & Trinajstic, N. (1977). On the aromatic stability of positional isomers consisting of bicyclic systems composed entirely of five-membered heterocycles. *Croat. Chem. Acta.* **49**. 107-113

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

Minks, C., Huber, R., Moroder, L. & Budisa, N. (2000a). Non-invasive tracing of recombinant proteins with "Fluorophenylalanine-Fingers". *Anal. Biochem.* **284.** 29-34

Minks, C., Alefelder, S., Huber, R., Moroder, L. & Budisa, N. (2000b). 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

Murakami, Y., Watanabe, T., Suzuki, H., Kotake, N., Takahashi, T., Toyanari, K., Ohno, M., Takase, K., Suzuki, T. & Kondo, K. (1997). New findings on the Hemetsberger-Kittel reaction (Synthetic studies on indoles and related compounds. XLIII). *Chem. Pharm. Bull.* **45**. 1739-1744

Nath, U. & Udgaonkar, J. P. (1997). Folding of tryptophan mutants of barstar: Evidence for an initial hydrophobic collapse on the folding pathway. *Biochemistry* **36**. 8602-8610

Nicholls, A., Bharadwaj, R. & Honig, B. (1993). GRASP-Graphical representation and analysis of surface properties. *Biophys. J.* **64**. A166-A166

Noren, C. J., Anthony-Cahill, 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

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

Othwinowski, Z. & Minor, W. (1996). Processing of X-ray diffraction data collected in oscillation mode. *Meth. Enzymol.* **276**. 307-326

Palm GJ, & Wlodawer A. (1999). Spectral variants of green fluorescent protein. *Method. Enzymol.* **33**. 378-394

Patterson, A. L. (1934). A fourier series method for the determination of the components of interatomic distances in crystals. *Phys. Rev.* **46**. 372-376

Paulmier, C. & Pastour, P. (1967). Sur les dérivés iodés du sélénophéne; le formyl-3-sélénophéne. *C. R. Acad. Sc. Paris Série C.* **265**. 926-928

Perutz, M. F., Rossman, M. G., Cullis, A. F., Muirlhead, H., Will, G. & North, A. C. T. (1960). Structure of hemoglobin: A three-dimensional fourier synthesis at 5.5 Å resolution, obtained by X-ray analysis. *Nature* **185**. 416-422

Phillips, G. N. (1998). *Green Fluorescent Protein*. Wiley-Liss, New York

Phillips, R. S., Cohen, L. A., Annby, U., Wenso, D. & Gronowitz, S. (1995). Enzymatic synthsis of thia-L-tryptophans. *Bioorg. Med. Chem. Lett.* **5.** 1133-1134

Piper, M., Betz, M., Budisa, N., Gomis-Ruth, F.X., Bode, W. & Tschesche, H. (1997). Expression, purification, characterization, and X-ray analysis of selenomethionine 215 variant of leukocyte collagenase. *J. Prot. Chem.* **16.** 637-650

Prendergast, F. (1999). Biophysics of the green fluorescent protein. *Methods Cell Biol.* **58**, 1-18

Rennert, O. M. & Anker, H. S. (1963). On the incorporation of 5´5´5´-Trifluoroleucine into protein of *Escherichia coli*. *Biochemistry* **2.** 471-476

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

Robert, M. C., Provost, K. & Lefaucheux, F. (1991). Crystallisation in gels and related methods. In: *Crystallisation of Nucleic Acids and Proteins - A Practical Approach* (Ducruix, A. & Giege, R., eds) Iril Press at Oxford University Press. Oxford, New York, Tokio.

Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H. & Zehfus, M. H. (1985). Hydrophobicity of amino acid residues in globular proteins. *Science.* **229**. 834-838

Ross, J. B., Szabo, A. G. & Hogue, C. W. (1997). Enhancement of protein spectra with tryptophan analogues: fluorescence spectroscopy of protein-protein and protein-nucleic acid interactions. *Method. Enzymol.* **278**. 151-190

Rossman, M. G. & Blow, D. M. (1962). The Detection of subunits within the crystallographic asymmetric unit. *Acta Cryst.* **15**. 24-31

Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys.* **78**. 1606-1619

Seifert, M. H. J., Ksiazek, D., Azim, M. K., Smialowski, P., Budisa, N. & Holak, T. A. (2002). Slow exchange in the chromophore of a green fluorescent protein variant. *J. Am. Chem. Soc.* **124**. 7932-7942

Sinha, H. K., Dogra, S. K. & Krishnamurthy, M. (1987). Excited-state and ground-state proton-transfer reactions in 5-aminoindole. *Bull. Chem. Soc. Jpn.* **60**. 4401-4407

Sloan, M. & Phillips, R. S. (1996). Effects of α-deuteration and of aza and thia analogs of L-tryptophan an formations of intermediates on the reaction of *Escherichia coli* tryptophan idole-lyase. *Biochemistry* **35**. 16165-16173

Smith, J. L. & Thompson, A. R. (1998). Reactivity of selenomethionine – dents in the magic bullet? *Structure* **6**. 815 – 819

Sopkova, J., Gallay, J., Vincent, M., Pancoska, P. & Lewit-Bentley, A. (1994). The dynamic behaviour of annexin V as a function of calium ion binding: a circular dichroism, UV absorption, and steady-state and time-resolved fluorescence study. *Biochemistry* **33**. 4490-4499

Soumillion, P., Jespers, L., Vervoort, J. & Fastrez, J. (1995). Biosynthetic incorporation of 7-aza tryptophan into the phage lambda lysozyme. *Protein Eng.* **8**. 451-456

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

Stout, G. H. & Jensen, L. H. (1989). *X-ray Structure Determination.* John Wiley. New York

Stryer, L. (1988). *Biochemistry* 3rd Ed. W. H. Freeman and Company. New York

Thomson, W. (1897). On the thermo-elastic, thermomagnetic and pyrœlectric properties of matter. *Q. J. Math.* **1**. 57-77

Thorson, J. S., Cornish, V. W., Barret, J. E., Cload, S. T., Yano, T. & Schultz, P. G. (1998). A biosynthetic approach for the incorporation of unnatural amino acids into proteins. In: *Protein Synthesis: Methods and Protocols* (Martin, R., ed.). Humana Press.Totowa, New Jersy

Tsien, R. Y. (1998). The green fluorescent protein. *Annu. Rev. Biochem.* **67**. 509-544

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 Crystallography Macromolecular Macromolecular Computing School* (ed. Bourne, P.E. & Watenpaugh, K.)

Vargas, R., Garza, J., Dixon, D. A. & Hay, B. P. (2000). How strong is the C^a-H$\cdots$O=C hydrogen bond? *J. Am. Chem. Soc.* **122**. 4750-4755

Wallace, A.C., Laskowski, R. A. & Thornton, J.M. LIGPLOT - A program togenerate schematic diagrams of protein legand interactions. *Prot. Eng.* **8**. 127-134

Waluk, J., Rettig, W. & Spangetlarsen, J. (1988). Ground-state and excited-state protonation of aminoquinoxalines. *J. Phys. Chem.* **92**. 6930-6935

Wang, L. & Schultz, P.G. (2001). A general approach for the generation of orthogonal tRNAs. *Chemistry and Biology.* **8**. 883-890

Wang, L. & Schultz, P.G. (2002). Expanding the genetic code. *Chem. Comm.* **1**. 1-11

Watson, J. D. & Crick, F. H. (1953). Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. *Nature* **171**. 737-740

Welch, M. & Phillips, R. S. (1999a). Improved syntheses of [3,2-b]- and [2,3-b]-fused selenolo- and thienopyrroles, and of furo[3,2-b]pyrrole. *Heterocyclic Comm.* **5**, 305-310

Welch, M. & Phillips, R. S. (1999b). Enzymatic syntheses of 6-(4H-selenolo[3,2-b]pyrrolyl)-L-Alanine, 4-(6H-selenolo[2,3-b]pyrrolyl)-L-alanine, and 6-(4H-furo[3,2-b]pyrroyl)-L-alanine. *Bioorg. Med. Chem. Lett.* **9**. 637-640

Wensbo, D., Annby, U. & Gronowitz, S. (1995). Indole-3-acetic acids and hetero analogues by one pot synthsis including Heck cyclisation. *Tetrahedron* **51**. 10323-10342

Wilson Miles, E., Kawasaki, H., Ahmed, A.S., Morita, H., Morita, H. & Nagata, S. (1989). The β-subunit of tryptophan synthase. *J. Biol. Chem* **264**. 6280-6287

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

Zacharias, D.A., Violin, J.D., Newton, A.C. & Tsien, R.Y. (2002). Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science*. **296**. 913-916

Zimmer, M. (2002). Green fluorescent protein (GFP): Applications, structure, and related photophysical behaviour. *Chem. Rev.* **102**. 759-781

Curriculum vitae

Born on July 26, 1971 in Seoul, Republic of Korea.

1978-1990   Primary education and high school in Seoul (Korea)

1990-1997   Studies in Chemistry at Sogang University (Seoul, Korea)

Bachelor Degree in Chemistry

1992-1993   Military service

1997-1999   Studies in Organic Chemistry at Sogang University (Seoul, Korea)

Master Degree (Diploma) in Chemistry

1999-2002   Ph.D. student at Max-Planck Institute for Biochemistry in Martinsried (Germany)