TECHNISCHE UNIVERSITÄT MÜNCHEN

MAX-PLANCK-INSTITUT FÜR BIOCHEMIE

In vivo Tandem Labeling of Proteins: Combining Chemical Orthogonality with Intrinsic Blue Fluorescence

Sandra Lepthien

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. Michael Groll
Prüfer der Dissertation:	1. Priv.-Doz. Dr. Nediljko Budisa
	2. Univ.-Prof. Dr. Johannes Buchner
	3. Univ.-Prof. Dr. Thomas Kiefhaber

Die Dissertation wurde am 24.09.2008 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 15.12.2008 angenommen.

Parts of this work were published as listed below:

Lepthien S, Hoesl MG, Merkel L, and Budisa N (2008) Azatryptophans endow proteins with intrinsic blue fluorescence. *Proc Natl Acad Sci USA* 105:16095-16100.

Giese C, **Lepthien S**, Metzner L, Brandsch M, Budisa N, and Lilie H (2008) Intracellular uptake and inhibitory activity of aromatic fluorinated amino acids in human breast cancer cells. *ChemMedChem* 3:1449-1456.

Lepthien S and Budisa N, *In vivo* tandem incorporation of reactive-handle and fluorescence-probe into proteins, *In Preparation*

Lepthien S, *et al.* (2006) *In vivo* engineering of proteins with nitrogen-containing tryptophan analogs. *Appl Microbiol Biotechnol* 73:740-754.

Rubini M*, **Lepthien S***, *et al.* (2006) Aminotryptophan-containing barstar: Structure-function tradeoff in protein design and engineering with an expanded genetic code. *Biochim Biophys Acta - Proteins and Proteomics* 1764:1147-1158. *both authors equally contributed to this work

Lepthien S and Budisa N (2006) High performance fluorescence scanning of tryptophan with Tecan's Safire2™ microplate reader. *Tecan Journal* 2:13-15.

Abstract presentations:

XX. International Congress of Genetics, *Blue Fluorescent Proteins*, August 12-17, 2008, Berlin, Germany

Department Retreat, *Recombinant Protein Production in Pichia pastoris*, October 26-29, 2006, Linz, Austria

6. Graduate Retreat, *Structure-function tradeoff in protein design and engineering with an expanded genetic code*, June 26-28, 2006, Ringberg Castle, Ringberg, Germany

Poster presentations:

Wiltschi B, Wenger W, Merkel L, Cheburkin Y, Wolschner C, **Lepthien S**, and Budisa N, BMBF BioFuture Meeting – 5. Presentation, February 02-03, 2006, Berlin, Germany

Lepthien S, Pal PP, and Budisa N, *Expansion of the genetic code enables design of a novel "gold" class of green fluorescent proteins*, GBM Annual Fall Meeting, September 19-22, 2004, Munster, Germany

Danksagung

Zuallererst und im Besonderen möchte ich mich bei meinem Betreuer und Mentor Herrn PD Dr. Nediljko Budisa bedanken, der mich stets unterstützt und ermutigt hat. Sein Enthusiasmus, seine Kreativität und wissenschaftlichen Ideen, waren in höchstem Maße inspirierend und motivierend. Nediljko, es war mir eine Ehre, vor allem aber eine Freude, mit Dir auf dem unerschöpflichen Gebiet der Erweiterung des genetischen Codes zusammenzuarbeiten!

Herrn Prof. Dr. Johannes Buchner und Herrn Prof. Dr. Thomas Kiefhaber danke ich für ihre Übernahme des Korreferats.

Mein Dank geht auch an Herrn Prof. Dr. Hauke Lilie von der Martin-Luther-Universität Halle-Wittenberg für die gute Zusammenarbeit an unserem gemeinsamen Projekt.

Bei Herrn Prof. Dr. Robert Huber und Herrn Prof. Dr. Dieter Oesterhelt bedanke ich mich für die Möglichkeit, in ihren Abteilungen zu arbeiten, und die hervorragende Infrastruktur zu nutzen. Den Kollegen aus beiden Abteilungen danke ich für eine schöne Zeit und ihre Unterstützung.

Herrn Prof. Luis Moroder danke ich für seinen umfassenden Rat bei chemischen und biophysikalischen Fragestellungen. In diesem Zusammenhang gilt mein besonderer Dank auch Herrn Hans-Jürgen Musiol, Herrn Dr. Stephan Uebel und den anderen Mitgliedern der Core-Facility für ihren unermüdlichen Einsatz in der Protein- und Aminosäureanalytik. Vor allem und ganz herzlichst danke ich aber Frau Elisabeth „Lissy" Weyher-Stingl für ihren unübertroffenen Erfahrungsschatz auf dem Gebiet der Spektrometrie und Spektroskopie, ohne den diese Arbeit sicher nicht zustande gekommen wäre.

Herrn Dr. Frank Siedler und Frau Sigrid „Siggi" Bauer danke ich für den ersten Nachweis von 4-Azatryptophan im Protein, Herrn Reinhard „Reini" Mentele für die *N*-terminalen Sequenzierungen. Meinen Kollegen Herrn Lars Merkel und Herrn Dr. Shouliang Dong danke ich für die Synthese von Aminosäureanaloga, die für diese Arbeit unverzichtbar waren.

Meinen liebsten Kollegen „den Budisas“ danke ich für ihre Unterstützung im Labor und vor allem für eine wundervolle Zeit! Ich danke Frau Waltraud „Traudl“ Wenger, die mich den ganzen Weg begleitet hat und mir die beste Lehrmeisterin war, und Frau Petra Birle und Frau Tatjana Krywcun, zusammen die DNA-Experten „P+T“. Frau Dr. Birgit Wiltschi danke ich für ihren kritischen Blick beim Korrekturlesen, nicht nur von dieser Arbeit, und ich danke Dr. Yury Cheburkin, der immer einen praktischen Trick aus dem Hut zaubern konnte.

Für eine schöne Zeit, an die ich mich gerne zurückerinnern werde, und jede Menge Spaß möchte ich mich bei meinen allerliebsten Kollegen und Freunden bedanken. Frau Christina „Tina“ Wolschner danke ich für Ihre Freundschaft und besonders für die Zeit, die wir außerhalb des Instituts verbracht haben. I also would like to thank my dearest friend Ms Dr. Mekdes Debela for her help and support whenever I needed it. Thank you, Mekdes, and take care! An dieser Stelle danke ich auch Herrn Dr. Peter Göttig, der immer Zeit für allerhand „Spezialaufträge“ fand. Herrn Lars Merkel gebührt Dank für die Beantwortung all meiner chemischen Fragen und Herrn Michael „Michi“ Hösl, der einen großen Teil zu dieser Arbeit beigetragen hat, wünsche ich nach einer exzellenten Masterarbeit als meinem Nachfolger alles Gute für seine Doktorarbeit.

Mit ganzem Herzen danke ich meiner Familie, meinen Eltern Erika und Thomas, meinem Bruder Dirk und Markus für ihre Liebe, Unterstützung und ganz besonders für die Chance, meine Träume zu verwirklichen.

Table of contents

I Nomenclature and Definitions

I.I Nomenclature

Abbreviations used throughout the thesis are according to the recommendations of the IUPAC-IUB Commission of biochemical nomenclature and of the ACS Style Guide.

Furthermore, the following abbreviations were used:

A	Adenine
AARS	Aminoacyl-tRNA-synthetase
Aha	Azidohomoalanine
anxA5	Human Annexin A5
APS	Ammonium persulfate
Arg	Arginine
(Aza)Ind	Azaindole
(Aza)Trp	Azatryptophan
(2-Aza)Trp	2-azatryptophan
(4-Aza)Trp	4-azatryptophan
(5-Aza)Trp	5-azatryptophan
(6-Aza)Trp	6-azatryptophan
(7-Aza)Trp	7-azatryptophan
(4-Aza)Ind	4-azaindole
(5-Aza)Ind	5-azaindole
(7-Aza)Ind	7-azaindole
Azulene	Azulenyl-alanine
C	Cytosine
CCHC	Copper(I)-catalyzed Huisgen cycloaddition

CD	Circular dichroism
ECFP	Enhanced cyan fluorescent protein
E. coli	*Escherichia coli*
EGFP	Enhanced green fluorescent protein
ε_M	Molar extinction coefficient
ESI-TOF	Electrospray ionization-time of flight
FITC	Fluorescein-isothiocyanat
FRET	Fluorescence/Förster energy resonance transfer
G	Guanine
GFP	Green fluorescent protein
GdFP	Gold fluorescent protein
His	Histidine
Hpg	Homopropargylglycine
Ile	Isoleucine
Ind	Indole
IPTG	Isopropyl-β-D-thiogalactopyranoside
λ_{Ex}	Excitation wavelength
LC	Liquid chromatography
Met	Methionine
MetRS	Methionyl-tRNA-synthetase
mRNA	Messenger RNA
MS	Mass spectrometry
OD_{600}	Optical density at λ = 600 nm
PAGE	Polyacrylamide gel electrophoresis
PEG	Polyethylene glycol
Phe	Phenylalanine

PLP	Pyridoxal-5´-phosphate
Pro	Proline
ψ-b*	Cysteine-free pseudo wild-type barstar (mutations Pro28Ala/Cys41Ala/Cys83Ala)[1]
Q-TOF	Quadrupole-time of flight
QY	Quantum yield
R	$H_2CCH(NH_2)COOH$
RNA	Ribonucleic acid
RP-HPLC	Reverse phase-high performance liquid chromatography
rRNA	Ribosomal RNA
rpm	Revolutions per minute
SDS	Sodium dodecyl sulfate
SeMet	Selenomethionine
Ser	Serine
snRNA	Small nuclear RNA
SPI	Selective pressure incorporation
T	Thymine
TEMED	*N*,*N*,*N*′,*N*′-tetramethylethylenediamine
tRNA	Transfer RNA
Trp	Tryptophan
TrpRS	Tryptophanyl-tRNA-synthetase
TrpS	Tryptophan synthase
Tyr	Tyrosine
U	Uracil
UV_{280}	Ultraviolet light at λ = 280 nm

I.II Definitions

Canonical amino acids are named according to the standard three letter code, e.g. tryptophan (Trp), methionine (Met), tyrosine (Tyr).

Non-canonical amino acids are generally denoted by the standard three letter code and a prefix which characterizes the non-canonical functional group, e.g. 4-azatryptophan [(4-Aza)Trp].

Mutant denotes a protein, in which the wild-type sequence is changed by site-directed mutagenesis in the frame of the 20 canonical amino acids.

Variant denotes a protein, in which one or more canonical amino acids from a wild-type or mutant sequence are replaced by non-canonical ones. The respective incorporated non-canonical functional group is used as prefix of the protein name, e.g. pseudo wild-type barstar (ψ-b*) with incorporated homopropargylglycine (Hpg) and 4-azatryptophan [(4-Aza)Trp] is called Hpg-(4-Aza)Trp-ψ-b*.

1 Summary

The aim of this study was to challenge two goals: First to explore the possibility for the *in vivo* expression of intrinsically colored proteins without the need of further posttranslational modifications or chemical transformations without externally added reagents, and second to further equip these blue-colored fluorescent proteins with a reactive handle for individual chemical functionalizations.

The first goal was achieved by the incorporations of biocompatible azaindoles/azatryptophans as optical probes into proteins by the selective pressure incorporation method. These amino acid analogs represent almost ideal isosteric substitutes for natural tryptophan in cellular proteins. In order to realize this concept, the single tryptophan residue in human annexin A5 was substituted with 4- and 5-azatryptophan using tryptophan-auxotrophic *Escherichia coli* cells. Both cells and proteins with these fluorophores possess intrinsic blue fluorescence detectable upon routine UV irradiations. 4-azaindole was identified as a superior optical probe due to its pronounced Stokes-shift of ~130 nm. Its incorporation into proteins is the most straightforward method for the conversion of naturally colorless proteins and cells into their blue counterparts.

The second goal was achieved by *in vivo* double labeling of one protein with two distinct non-canonical amino acids. This represents the first simultaneous/tandem incorporation of amino acid analogs into protein using polyauxotrophic *Escherichia coli* cells. For this purpose, both the blue-fluorescent 4-azatryptophan and the methionine analog homopropargylglycine were simultaneously incorporated into pseudo wild-type barstar. Homopropargylglycine provides an alkyne group as reactive handle for specific bioorthogonal chemical transformations. The tandem labeling approach provides not only a fluorescent probe for straightforward detection but also a reactive tag for individualized modifications of the target proteins. This demonstrates new methodological possibilities of the selective pressure incorporation and moreover provides a new dimension of rational protein engineering and design with an expanded genetic code.

2 Zusammenfassung

Diese Arbeit verfolgte zwei herausfordernde Ziele. Zunächst sollte die Möglichkeit untersucht werden, ob die Expression von intrinsisch-fluoreszierenden Proteinen ohne weitere posttranslationale Modifizierungen oder chemische Transformationen *in vivo* möglich ist. In einem zweiten Schritt sollten diese blau-fluoreszierenden Proteine zusätzlich mit einer reaktiven Gruppe ausgestattet werden, die individuelle Funktionalisierungen erlaubt.

Das erste Ziel wurde mithilfe des Einbaus von biokompatiblen Azaindolen/Azatryptophanen als optische Sonden in Proteine erreicht. Diese Aminosäure-Analoga repräsentieren isostere Substituenten für Tryptophan in zellulären Proteinen. Mittels der so genannten Selektionsdruck-Einbaumethode wurde in Tryptophan-auxotrophen *Escherichia coli*-Zellen der einzige Tryptophanrest des humanen Annexin A5 gegen 4- und 5-Azatryptophan ausgetauscht. Nach Bestrahlung mit UV-Licht zeichneten sich sowohl Zellen als auch Proteine durch intrinsische blaue Fluoreszenz aus. 4-Azaindol wurde vor allem wegen seines ausgeprägten Stokes-Shifts von ~130 nm als überragende Fluoreszenzsonde identifiziert.

Das zweite Ziel wurde durch den ersten *in vivo* Doppeleinbau von zwei unterschiedlichen nicht-kanonischen Aminosäuren in ein einzelnes Protein erreicht. Der Einbau erfolgt simultan und wird als Tandemeinbau bezeichnet. Zu diesem Zweck wurde nicht nur das blau-fluoreszierende 4-Azatryptophan, sondern auch das Methioninanalogon Homopropargylglycin simultan in pseudo-Wildtyp Barstar eingebaut. Homopropargylglycin besitzt eine spezifische reaktive Gruppe für bioorthogonale chemische Transformationen. Auf diese Weise wird nicht nur ein fluoreszierender Marker für die einfache und schnelle Detektion, sondern auch eine reaktive Seitenkette für individuelle Modifizierungen des Zielproteins zur Verfügung gestellt.

Diese Arbeit präsentiert neue methodologische Möglichkeiten für die Selektionsdruck Einbaumethode und eröffnet darüber hinaus eine neue Dimension des rationellen Proteindesigns mit einem erweiterten genetischen Code.

3 Theoretical Background

3.1 The flow of the genetic information

The macromolecule desoxyribonucleic acid (DNA) was first described by Miescher, who isolated the DNA of a human cell in 1871.[2] His work established the fundamental basis for today's biochemical and molecular biological knowledge. In 1944, Avery and his coworkers succeeded in the identification of DNA as the carrier of the genetic information.[3] They have demonstrated the transfer of genetic information between two distinct *Pneumococcus* strains. Based on the work of Franklin and Chargaff, a few years later, Watson, Crick, and Wilkinson presented the groundbreaking three-dimensional model of the structure of DNA. It proposed a concept of base pairing of the four DNA bases adenine (A), thymine (T), guanine (G), and cytosine (C).[4-8] Their double-helix model was especially suitable to explain how the DNA molecule encodes, amplifies, and mutates the genetic information.

In 1957, Crick proposed his well-known concept of the "central dogma of molecular biology", which makes clear that the flow of the genetic information from DNA towards protein is unidirectional and no transfer of the genetic information is possible from protein back to DNA.[9] In this respect, DNA, the carrier of the genetic information, is first transcribed into ribonucleic acid (RNA), where T is replaced by uracil (U), by the cellular transcription machinery. Diverse RNA molecules, e.g. transfer RNA (tRNA), ribosomal RNA (rRNA), or small nuclear RNA (snRNA), participate in a number of cellular processes. Importantly, messenger RNAs (mRNAs) are translated into their related amino acid sequence at the ribosome (chapter 3.2) in a process known as translation. On the mRNA, units of three nucleobases, the so-called coding triplet, encode for one amino acid each in the related protein. If the four bases A, U, G, and C are assembled in groups of three, 64 combinations are possible. The standard genetic code relates 61 codons to a pool of 20 canonical amino acids while three codons (UAA (ochre), UGA (opal), and UAG (amber)) are stop codons that act as translation termination signals (Fig. 1).

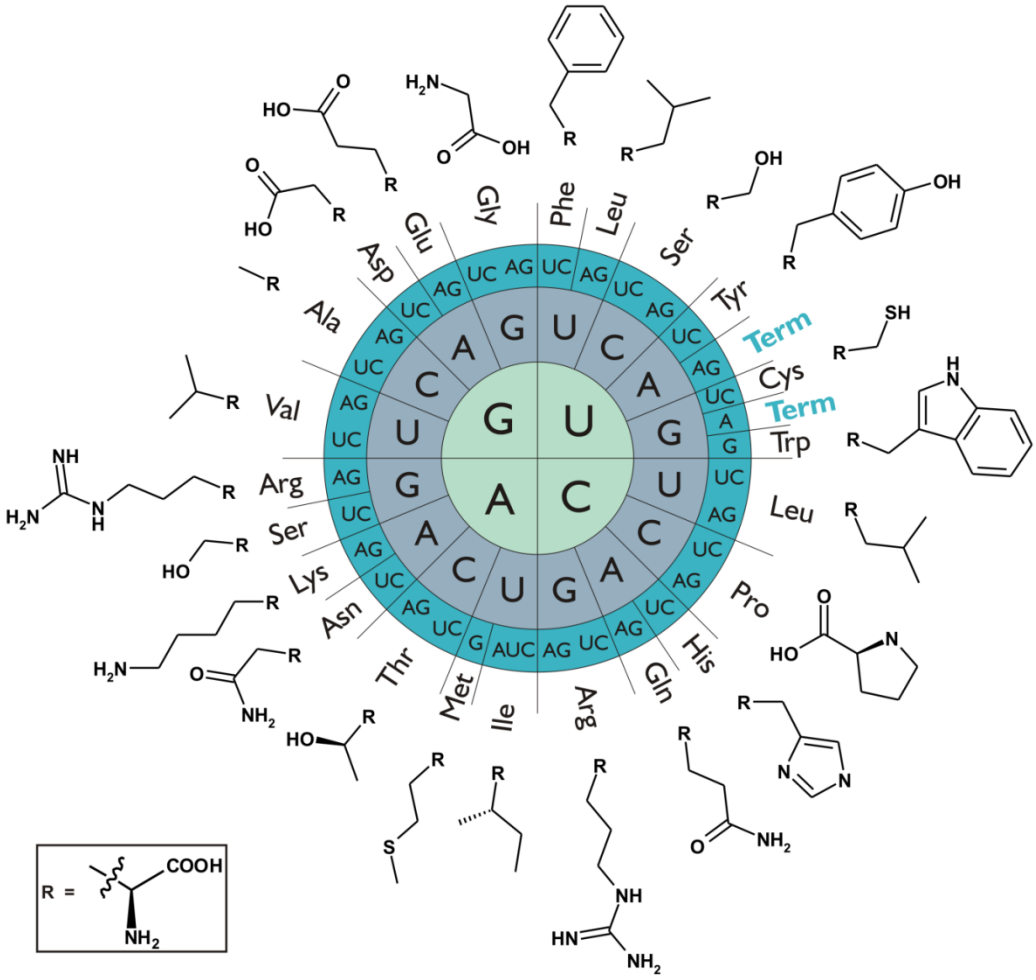


Fig. 1 **Radial presentation of the standard genetic code (RNA format).** Note that 64 triplets are encoding for 20 different amino acids and three stop codons (denoted as Term; UAA, UGA, and UAG). Ala: Alanine; Arg: Arginine; Asn: Asparagine; Asp: Aspartic acid; Cys: Cysteine; Gln: Glutamine; Glu: Glutamic acid; Gly: Glycine; His: Histidine; Ile: Isoleucine; Leu: Leucine; Lys: Lysine; Met: Methionine; Phe: Phenylalanine; Pro: Proline; Ser: Serine; Thr: Threonine; Trp: Tryptophan; Tyr: Tyrosine; Val: Valine.

The genetic code is called degenerated or redundant, since some amino acids are encoded by more than one codon. In most cases, the synonymous triplets only differ in the third base. Even after a single base mutation, still the correct amino acid is incorporated into the growing polypeptide chain during translation. In 1966, this was also formulated by Crick as "Wobble" hypothesis.[10] Generally, amino acids with a high abundance in proteins, e.g. arginine, leucine, or serine, are encoded by more than one triplet, whereas the rare amino acids tryptophan (Trp) and methionine (Met) are only related to one codon (Trp: UGG; Met: AUG).

Although all living organisms use the same 20 canonical amino acids to execute the rules of the genetic code, there are few deviations in its interpretation. For instance, the genetic code in mitochondria and chloroplasts and in some ciliates uses different amino acid assignments, e.g. UAA and UAG encode for glutamine.[11] Besides the 20 canonical amino acids, two additional amino acids, selenocysteine, in archaea, bacteria, and mammals, and pyrrolysine in archaea, can be incorporated into the nascent polypeptide chain in response to UGA or UAG (at stop codon) in a certain sequence context.[12,13] Furthermore, their incorporation requires a specific folding of the mRNA and the presence of specific translation factors.

Another distinctive feature of the genetic code is that the chemical nature of the amino acid is related to the middle base of its coding triplet (codon). Consequently, in the event of mutations of the first and/or last base of a triplet, at least, the nature of the amino acid is preserved. In this respect, the structure of the genetic code and the principles of protein folding are related so that the impact of mutations is minimized (concept of error minimization).[14]

3.2 Quality control mechanisms in aminoacylation and translation

Protein translation is mediated by a large protein-RNA complex, the ribosome, which catalyzes the synthesis of a polypeptide chain, dictated by the order of the nucleotide sequence of the mRNA. Covalently bound to tRNA molecules, the appropriate amino acids are provided for protein synthesis. These act as adaptors between the coding triplets of the nucleic acids and the amino acid sequence of the nascent polypeptide. The correct translation of the genetic information from the nucleic acid into a protein is controlled by two independent proofreading mechanisms: First the charging of the amino acids onto the corresponding tRNAs by the aminoacyl-tRNA-synthetases (AARS), and second the appropriate base pairing of the coding triplet of the mRNA with the anticodon of the tRNA. Thus, the execution of the genetic code, the linkage between a certain anticodon (hence, codon) and the corresponding amino acid is actually performed by the AARS, which are the exclusive interpreters of the genetic code. The ribosome does not check whether the correct amino acid is attached to the acceptor stem of the tRNA, as already demonstrated by Lipmann and coworkers

in 1962.[15] The importance of this finding will be discussed in detail in chapter 3.3.2. In general, for each canonical amino acid only one AARS exists, which catalyzes the aminoacylation reaction between the amino acid and the appropriate tRNA. This two-step reaction consumes ATP as energy and requires Mg^{2+} ions as cofactors. In the first step, the amino acid is activated by the covalent addition of AMP, forming an aminoacyl-adenylate. This is followed by a second step, where the amino acid is transferred to the cognate tRNA (formation of aminoacyl-tRNA), while the AMP is released. The charged aminoacyl-tRNA enters the ribosomal translation process via codon-anticodon base pairing at the A position of the ribosome and further participates in the translation process. Each ribosomal translation cycle ends with the release of the uncharged tRNA, which is now available for the next aminoacylation reaction.[11] Approximately, 10-20 polymerization reactions are performed per second by the ribosome, thus, polypeptide biosynthesis is five times slower than the RNA synthesis.[16]

3.3 Code engineering

3.3.1 Motivation

The chemical functions introduced into polypeptides by the 20 canonical amino acids are limited to simple aliphatic, aromatic, basic, acidic, hydroxyl-, and sulfur-containing groups (Fig. 1). To overcome these limitations, nature introduces complex posttranslational modifications.[17,18] However, a number of very interesting functionalities and features are still missing. For example, halogen, alkene, alkyne, keto, cyano, azido, nitroso, or silyl groups would certainly offer a vast variety of new chemical reaction partners, such as for posttranslational modifications. The expansion of the canonical pool of amino acids with these functional groups would highly increase the available tools for designing tailor-made proteins.

3.3.2 Historical background and experimental approaches

Engineering of the genetic code is based on the reassignment of the coding triplets. In this way, codons are no longer encoding for the canonical amino acid as it is depicted in Fig. 1, but for a new or synthetic non-canonical amino acid. Basically, the meaning of sense and stop codons can be reassigned. In the

1950s, Cowie and Cohen performed experiments aiming at the mis-incorporation of amino acids into proteins using amino acid auxotrophic bacterial cells.[19] Their approach is denoted as auxotrophy-based multiple-site mode of incorporation and is the basis of the selective pressure incorporation method (SPI), which will be discussed in detail in the next chapter 3.4. SPI enables the efficient translation of mis-acetylated tRNAs in response to the sense codons in the target gene.[14,20] For example, the incorporation of selenomethionine and telluromethionine instead of Met into proteins and the substitution of amino acids with ^{19}F-labeled aromatic amino acids for structure determination by X-ray crystallography and for NMR studies respectively, are classical examples for code expansion with this method.[21,22]

Another very important step towards the incorporation of non-canonical amino acids represented the experiments of Lipmann and his coworkers in 1962.[15] By using mis-aminoacylated Ala-tRNA^{Cys}, generated by the chemical reduction of Cys-tRNA^{Cys}, they succeeded in a codon-specific incorporation of alanine instead of cysteine *in vitro*. Currently, this idea is the basis for the production of numerous chemically or enzymatically aminoacylated tRNAs called as suppressor-tRNAs, since they suppress the original meaning of their cognate coding triplets.

The reassignment of translation termination signals (stop codon suppression-based methodologies) can be applied for isosteric amino acid replacements *in vivo* and *in vitro*.[23-25] Once the appropriate orthogonal AARS/tRNA pairs for amino acid analogs have been developed, they would theoretically allow the incorporation of one –in best case two– analogs in response to a single stop codon. The term orthogonal in this respect means that the AARS/tRNA pair does not interfere with the AARS and tRNAs of the expression host. However, the generation of an orthogonal AARS is tedious and time-consuming. The natural substrate specificity of an existing AARS towards the cognate canonical amino acid must be destroyed and the novel specificity for the non-canonical analog grafted onto the enzyme by directed evolution.[26] The groups of Sisido, Schultz and Tirrell have made major contributions in this field.[23-25] Nevertheless, this approach is still at the proof-of-principle level due to low product yields that originate from inefficient stop codon suppression and a technically expensive and

highly sophisticated experimental setup. Beside these, there are other approaches for the production of unnatural proteins and peptides, which are, for example, based on the use of ribozymes, modified ribosomes, novel base pairs, or reconstituted translation.[27,14]

The incorporation of non-canonical amino acids is expected to change the enzyme activity, the spectroscopic, and/or thermodynamic properties of protein. Furthermore, this may lead to novel applications in structure determination, diagnostics, and may open prospects for new drug discovery and medical treatments. Recent studies demonstrate that non-canonical amino acid incorporations turn neutral proteins into pH-shift- or redox-sensors for protein stability, folding, and dynamics. In general, the intrinsic properties of the incorporated non-canonical amino acid are directly transferred to the protein.[28,29] A number of different groups have also reported anti-microbial, anti-fungal, and anti-tumorigenic effects of non-canonical amino acids in proteins.[30-33] This medical aspect has been underestimated in the past but certainly will play an important role in prospective medical applications.

3.4 Selective pressure incorporation method (SPI)

In this study, the selective pressure incorporation (SPI) method was applied for the incorporation of non-canonical amino acids into proteins, which is based on the work of Cowie and Cohen (chapter 3.3.2).[19] For high yield incorporation of non-canonical amino acids via SPI in proteins, a number of prerequisites have to be fulfilled, which will be discussed in the following.

First the bacterial expression host cells have to be auxotrophic for the amino acid, which should be exchanged. Only with auxotrophic cells, grown in a minimal medium with defined nutrient composition, a quantitative incorporation of the novel amino acid is obtained. In this study, not only one but two non-canonical amino acids were incorporated in parallel. For this purpose, vital and fast-growing bacterial cells with at least two stable auxotrophies were necessary.

Second the gene of the model protein has to be cloned in a tightly controlled but highly inducible expression vector. Before expression induction, the promoter should not be leaky due to undesirable parent protein expression, but highly inducible for high yield expression of the fully substituted target protein.

Third a minimal medium with defined amino acid concentration is required for fermentation. This enables control of the amino acid composition, i.e. depletion of the target amino acid, and facilitates the pressure on the bacterial cells to incorporate the desired non-canonical amino acid.

Finally the non-canonical amino acid has to be available in excess for optimal incorporation conditions. Before this amino acid analog can participate in protein translation, it has to be uptaken by the cells in an active or passive way. This issue will be discussed in detail in the following chapter 3.5.

In summary, SPI is a reliable method for the high yield expression of proteins with various non-canonical amino acids, in which only sense-codons are reassigned. Nevertheless, its incorporation mode is not site-directed, i.e. all residues of the respective canonical amino acid are substituted for the non-canonical amino acid (incomplete substitution is also possible). By using proteins, which comprise only a small number of the target amino acid, a so-called pseudo site-directed incorporation mode is achieved. This mode applies to the proteins used in this study, the human annexin A5 (anxA5; one Trp) and ψ-b*, the cysteine-free mutant of barstar from *Bacillus amyloliquefaciens* (three Trp and one Met; *vide infra* chapters 6.1.1 and 6.1.2).

3.5 Amino acid uptake and metabolic stability

The efficient uptake of exogenous non-canonical amino acids into the expression host is the first crucial obstacle to overcome before the analogs can participate in protein translation. *In vivo* every cell membrane contains versatile amino acid transport systems, which differ significantly with respect to their substrate spectrum and transport mechanism.[34,35] Most transport systems are characterized by a very flexible substrate specificity and they enable the uptake of a family of similar substances. This lack of specificity is especially utilized in the field of protein and metabolic engineering.[36] The uptake of a non-canonical amino acid is absolutely crucial for its later successful incorporation. Fortunately, the existing broad substrate recognition of amino acid transporters proved to be sufficient for this purpose without need for direct engineering of these systems.

The incorporation capacity of a non-canonical amino acid is certainly highly influenced by its intracellular accumulation. Recently, our group in collaboration

with Hauke Lilie at the Martin-Luther University of Halle-Wittenberg reported that fluorotryptophans are accumulated in the breast cancer cell line MCF-7 *in vitro* up to their 70fold medium concentration (Fig. 2).[37] Although these substances cause irreversible growth inhibition in eukaryotic cell cultures with IC_{50} values in the low micromolar range, fluorotryptophans are generally well tolerated by bacterial cells and probably also intracellularly accumulated in high levels. They have been among the first non-canonical amino acids incorporated into proteins and their influence on protein folding and stability have also been exploited.[22,38]

Next to our data, also the reports of other groups indicate that the uptake systems are not capable to distinguish between chemically and sterically similar amino acids during the substrate recognition process.[37,27]

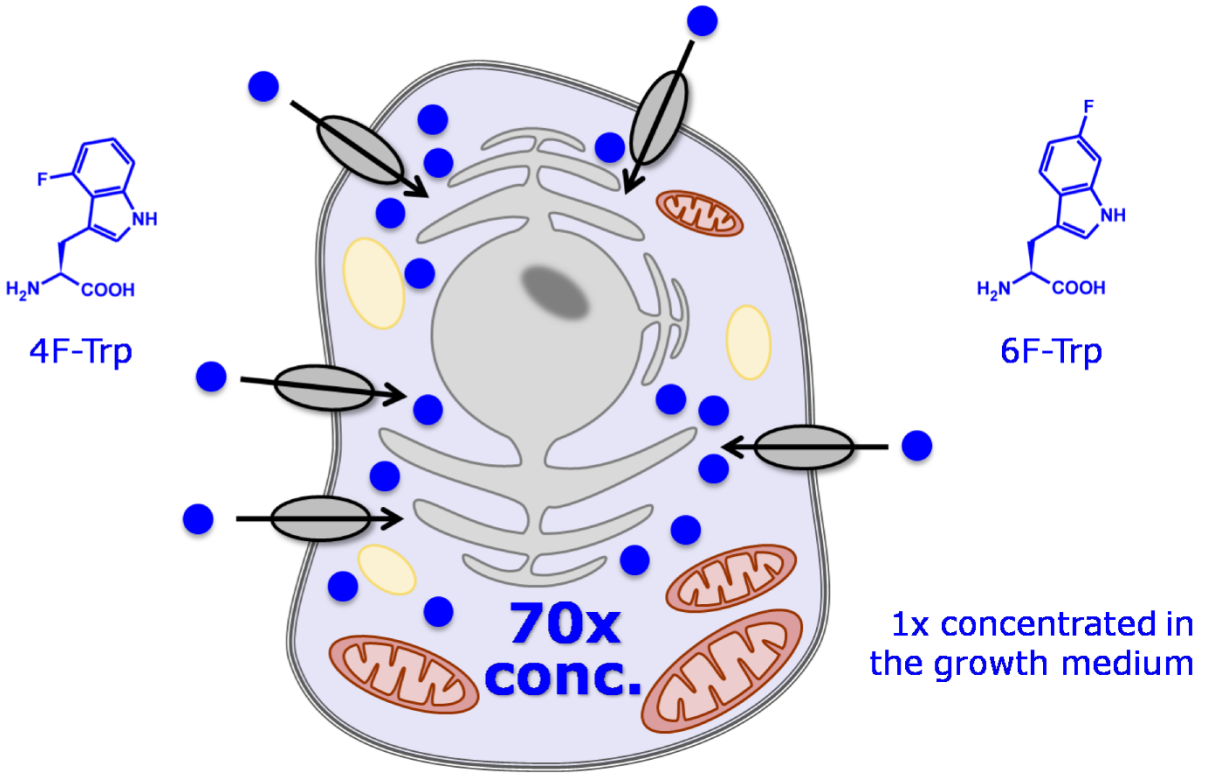


Fig. 2 **Import of amino acid analogs via the active transport system L of eukaryotic cells resulting in their 70fold intracellular accumulation.** We discovered that 4-fluorotryptophan (4F-Trp) and 6-fluorotryptophan (6F-Trp) are effectively uptaken by the transport system L of MCF-7 cells and are exhibiting a cytostatic effect *in vitro*.[37] Similar systems exist in bacterial cells and the uptake and accumulation of non-canonical amino acid analogs can be assumed as well.[36] The figure was kindly provided by Dr. B. Wiltschi.

However, once inside the cell, the participation of non-canonical amino acids in intracellular chemical reactions and transformations cannot be excluded. These reactions may change their chemical nature as amino acid or transform them into toxic compounds.

In particular, non-canonical amino acids can interact with catabolic enzymes, act as irreversible inhibitors, or interfere with the canonical amino acid biosynthesis.[39] Especially in higher organisms, problems may occur as amino acids serve as precursors for hormones and play an important role in the transmission of signals in the nerve system, e.g. serotonine.[40] A number of studies reported that non-canonical amino acids are interfering with the natural metabolism of cells.[27] Their successful incorporation into the target protein sequence requires careful considering of all these side-effects prior to the experiment.

3.6 Substrate promiscuity in amino acid activation and tRNA charging

After the efficient uptake and high level intracellular accumulation of the non-canonical amino acid, the respective AARS has to charge the analog onto the cognate tRNA. By using non-canonical amino acids, which are very similar with respect to their chemical nature and structure of a certain canonical amino acid, the chances are certainly higher that the analog is accepted by the AARS. Beside positive recognition, the enzymes have to exclude the other non-cognate amino acids of the intracellular pool in order to ensure translation fidelity. High precision of this recognition process and accurate proofreading is essential for the correct translation of the genetic information into the required polypeptide sequence. In this context, two factors are of particular importance: First the affinity of the enzyme towards its substrate and second the relative concentration of competing substrate analogs such as non-canonical amino acids. The ability to either catalyze one specific reaction with a number of similar substrates or to catalyze a range of chemical reactions depending on the reaction partner is called catalytic promiscuity.[41] For the expansion of the genetic code using the SPI method, this lack in absolute substrate specificity of the AARS for the charging of the non-canonical amino acids to the endogenous tRNAs is exploited. Furthermore, applying a selective pressure to the cells, which means

the depletion of the canonical amino acid and the parallel induction of protein expression, the cellular protein translation machineries of the cells are forced to accept the alternative substrates.

In the frame of this study above all, the catalytic promiscuity of the tryptophanyl-tRNA-synthetase (TrpRS; EC 6.1.1.2) and –in case of the tandem labeling approach– also of the methionyl-tRNA-synthetase (MetRS; EC 6.1.1.10) were of particular importance. In my previous studies (Diploma thesis), the kinetics of the TrpRS activation reaction towards Trp and its amino-analogs were studied.[42] In addition, a radioactivity assay was applied to examine the charging of amino acid analogs onto the respective tRNA. These studies showed a rather different catalytic performance of the enzyme, which is depending on the selected Trp analog. Nevertheless, the ability to accept the non-canonical amino acids as substrates was always determined. Furthermore, successful reports on incorporation experiments with non-canonical amino acids by the exploitation of the catalytic promiscuity of natural or mutant AARSs prove the importance of this aspect.[43,44] A straightforward incorporation experiment via SPI is also the simplest way to check whether a non-canonical amino acids is translationally active or not since it shows if the analog is capable to pass all cellular checkpoints of quality control. Recent approaches by using orthogonal pairs of AARS and tRNA from other organisms than the host will certainly enhance the repertoire of the host cell to accept non-canonical amino acids (chapter 3.4).[23-25]

3.7 Ribosome flexibility

As already mentioned, the translation of the RNA sequence into a polypeptide is catalyzed by the ribosome (chapter 3.2). Each amino acid enters the translation process as aminoacyl-tRNA complex and is identified by codon-anticodon base pairing. The nascent polypeptide chain leaves the ribosome through a channel, which shows a high flexibility when enabling the passage of the 20 different canonical amino acids. Recent reports on the successful incorporation of non-canonical amino acids with coumarin- or azophenyl-moieties, demonstrated the ability of the ribosome to accept even bulky and chemically extraordinary side chains as substrates for translation.[45] Nevertheless, the flexibility of the ribosome is limited. The exclusion of D-amino acids due to

kinetic proofreading and the very different performance towards bulky aromatic side chains demonstrate that ribosome flexibility can considerably limit the incorporation efficiency of non-canonical amino acids.[46,47] Recently, Hecht and coworkers succeeded in the incorporation of D-amino acids into proteins by mutating the 23 S rRNA of the *E. coli* ribosome.[48] In the future, the ribosome as target for engineering approaches may play an important role in the expansion of the amino acid repertoire.

3.8 Effect of non-canonical amino acids on protein folding and stability

Protein folding is influenced by different factors: van der Waals interactions, hydrogen bonds, hydrophobic interactions, hydration effects of non-polar groups, and salt bridges.[49,28] These interactions play a major role in protein folding and especially the aromatic amino acid Trp participates in a number of interactions, which can stabilize the three-dimensional structure of proteins (chapter 3.9). In general, aromatic amino acids are able to bind ions via cation-π interactions and are characterized by hydrophobicity. Localized inside the protein, they often contribute to the hydrophobic core stabilization of globular protein. Although mostly regarded as hydrophobic amino acid, Trp´s hydrophobic/hydrophilic nature is controversial and very differently discussed in the literature.[50,51] In general, the location of an amino acid in a protein structure correlates with its hydrophobicity.

Among all amino acids, only Trp and Met have no strict preference for a protein´s interior or its surface.[14,29] In protein interior, Trp often participates in a network of aromatic side chains via π-π-stacking interactions, located at the surface, it usually plays a crucial role in the catalytic reaction of an enzyme or participates in protein-protein interactions. In this respect, the small nuclease inhibitor barstar, which was also used as model protein in this study, is of particular interest (chapter 6.1.2).[1] It contains three Trp residues: Trp53 is located in the protein interior, Trp38 is partially solvent exposed, and Trp44 is fully surrounded by the solvent.[28] In the barstar core, Trp53 is sandwiched between two phenylalanine residues, and these residue clusters are known to serve as nucleation sites in protein folding (Fig. 3).[52,53]

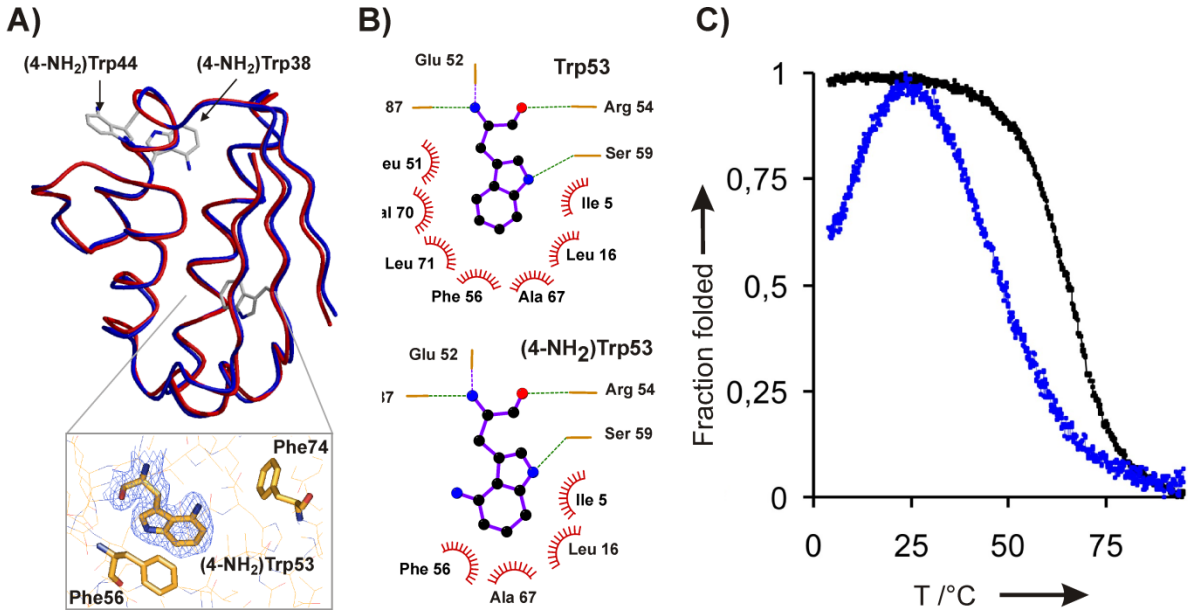


Fig. 3 **The thermodynamic price of the introduction of non-canonical amino acids into protein (Rubini *et al.*).[28] A)** Superimposed backbone structures of ψ-b* (blue) and its 4-aminotryptophan variant (red). The 4-aminotryptophan residues at the positions 38, 44, and 53 are shown as sticks. The surrounding of 4-aminotryptophan53 with the neighboring phenylalanine residues 56 and 74 is presented in the electron density. **B)** LIGPLOTv4.4.2[54] presentation of the molecular environment of Trp53 (above) and 4-aminotryptophan53 (below) of ψ-b* and its variant, respectively. The number of hydrophobic interactions of 4-aminotryptophan is significantly reduced. **C)** Thermally induced unfolding curve of ψ-b* (black) and its 4-aminotryptophan variant (blue). The melting profile of 4-aminotryptophan-ψ-b* is characterized by a decreased T_m value, a reduced folding cooperativity, and cold denaturation-like phenomenon, i.e. a maximum in stability at ~25 °C. These data are presented in detail in Rubini *et al.*[28]

This particular Trp53 residue is essential for maintaining barstar´s functional and structural integrity, thus it could not be exchanged by any other canonical amino acid.[55,56] Nevertheless, the substitution of all Trp residues by 4-aminotryptophan was successful.[28] The introduction of the amino group into the barstar structure revealed an isomorphous crystallographic structure, which showed no significant effect of 4-aminotryptophan on the protein integrity (Fig. 3A). However, this addition of a hydrophilic group in the hydrophobic core of the tightly packed protein reduced the number of hydrophobic interactions of Trp53 (Fig. 3B). This destabilizing effect was in agreement with the thermally induced unfolding profiles of 4-aminotryptophan-barstar (Fig. 3C). When

compared to the parent protein, these protein variants show a significantly decreased T_m value (ΔT_m = -20 °C), reduced unfolding cooperativity, and substantial loss of free energy of unfolding.[28] Moreover, the melting profiles are characterized by a phenomenon called cold denaturation, i.e. maximum in stability at ~25 °C.[57,58] Cold denaturation usually takes place below the freezing point of water, which is the reason why this phenomenon is difficult to study in detail. We assumed that the severe loss in hydrophobic interactions of the crucial Trp53 residue is the main reason for this effect but cannot exclude other factors. This behavior is traditionally studied by the addition of denaturants or a high hydrostatic pressure, nevertheless the incorporation of 4-aminotryptophan into barstar may represent a novel tool to examine this thermodynamic phenomenon.[57,58]

Although the presence of the amino group endowed barstar with the novel feature of pH sensitivity, the thermodynamic price is relatively high and represents a typical structure-function tradeoff. This is only one example for the thermodynamic price, which has to be paid in general for introducing new biophysical and chemical features to native proteins. Both the thermodynamic price and methodological benefit should always be balanced and have to be considered in every particular attempt of rational protein design.

3.9 Tryptophan – a unique amino acid in the genetic code

3.9.1 Tryptophan chemistry and metabolism

The canonical amino acid Trp is particularly interesting as target amino acid for protein engineering and design. Encoded by only one triplet (UGG), it has already been substituted by various different analogs.[29] These analogs are generated biosynthetically or by organic chemistry and demonstrate rich and multi-faceted indole (Ind) chemistry. In addition, Trp features a number of characteristics, which makes it an ideal target for SPI substitution approaches. In protein, Trp is a rare amino acid with an abundance of ~1% of all amino acid residues. It is the main source of the protein´s absorbance and fluorescence, comprising unique biophysical and spectral properties. Furthermore, it is especially involved in physical and chemical interactions, which are essential for protein stability and folding, e.g. hydrogen bonding, cation-π, and π-π-stacking

interactions.[27] The role of Trp in protein folding and stability is discussed in detail in chapter 3.8. Very often it is found at crucial positions in the active site of enzymes or on surface motifs interacting with membranes or other proteins. In addition, the canonical amino acid Trp plays an essential role in the metabolism of many living cells. In animals, it serves as a precursor for the biosynthesis of the hormones serotonine and melatonine, while plants require Trp for the production of alkaloids.[40,59] Many pharmacologically active bioisosteres are Ind-based substances, such as [3,2-*b*]-pyrrole and thieno[2,3-*b*]-pyrrole, analogs of the hallucinogen and serotonine agonist *N,N´*-dimethyltryptamine.[40]

3.9.2 Tryptophan as optical probe in protein biophysics

The absorbance and fluorescence profiles of most proteins are dominated by Trp and further influenced by phenylalanine and tyrosine. Trp´s intrinsic fluorescence is highly sensitive to its local environment, e.g. solvent changes, and seems to be the only amino acid residue, which donates electrons in the excited state, and is therefore susceptible to collisional quenching.[60] Thus, Trp enables the spectral measurement of structural changes and protein dynamics, but the interpretation of these data become complicated if more than one residue is involved.

While the long-wavelength absorbance (240 – 300 nm) of most other fluorophores, e.g. of tyrosine, consists of only a single electronic transition to the first singlet state S_1, the spectral features of Trp and its analogs are characterized by two nearly isoenergetic excited states, ^1L_a and ^1L_b (Fig. 4).[60] According to Platt, the relative energies and oscillator strengths of these two π-π* transitions are influenced not only by the solvent but also by the nature and position of substituents on the Ind moiety.[61]

The ^1L_b state is more highly structured than the ^1L_a state and represents the lower energy state in vapor phases and nonpolar solvents. The ^1L_a transition, which is strongly influenced by local changes, is broader, less structured, and stabilized by polar solvents. In compliance with Kasha´s rule, the fluorescence emission occurs always from the respective lowest energy state. Therefore, both electronic transitions show a different behavior in respect to absorbance, emission, anisotropy, and solvent polarity.[60]

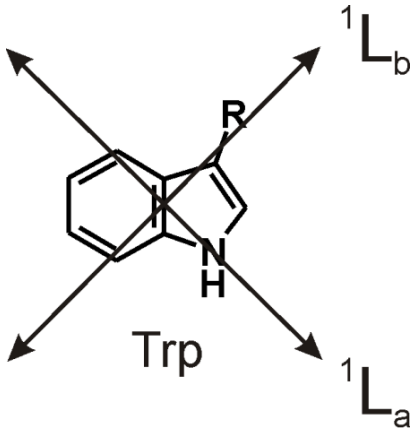


Fig. 4 Electronic absorbance transitions in Trp.[60] Transition dipoles are depicted as arrows and nearly perpendicular to each other. R denotes $H_2CCH(NH_2)COOH$.

Beside charge distribution in the excited state, i.e. change in the dipole moment, these are the reasons for Trp´s high sensitivity to variations in its molecular neighborhood (see also 3.9.4).

The addition of electron donating or electron withdrawing groups at distinct positions of the Ind moiety affects each of the two transition states in a different way. For example, a hydroxyl group at position 5 leads to a significant enhancement of the ^1L_b transition band and results in a shoulder in the absorbance profile at 300 nm.[60] In contrast, Sinha and coworkers reported the absorbance spectra of Ind with an amino group at position 5 to result from enhancements of both ^1L_a and ^1L_b transitions (main band at around 275 nm and a shoulder towards 315 nm).[62] Furthermore, hydroxyl and amino substituents increase the pH sensitivity of Ind´s fluorescence.[63,28] In the case of 4-aminotryptophan, the fluorescence maximum is shifted towards shorter wavelengths with increasing pH, while the maximum of 5-aminotryptophan is shifted to longer wavelengths with increasing pH.[28] Considering the possibility to turn every protein in a pH-sensor, it is reasonable to expect non-canonical analogs of Trp to serve as non-invasive tools for *in vivo* or *in vitro* sensors of protein-membrane, protein-protein, protein-ligand, or enzyme-substrate interactions (see also chapter 3.9.4).[29,14]

3.9.3 Tryptophan as target for protein engineering and design

As already discussed in chapter 3.9.1, Trp is one of the most suitable targets for engineering the spectral properties of proteins. However, Trp is not always qualified as optical probe due to its complicated photophysics and the unfavorable overlap between nucleic acid and protein fluorescence spectra (chapter 3.9.2).[60,64] For these reasons, alternative substances with distinct optical probes are needed. In addition, very often Trp cannot be exchanged by classical site-directed mutagenesis using canonical aromatic amino acids; this often leads to protein unfolding. Therefore, best candidates for the substitution of Trp are non-canonical Trp-analogs.

The diverse chemistry of Ind substitutions offers numerous analogs of Trp with intriguing physicochemical properties. However, the chemical substitution of Ind is limited. For example, the nitrogen atom at position 1 of the Ind moiety has been considered as attractive target for chemical functionalization. However, upon replacing this nitrogen with other heteroatoms (carbon, sulfur, or oxygen) the resulting Trp analog could not participate in protein synthesis.[65] These synthetic Ind analogs have neither served as substrates for the tryptophan synthase (TrpS, EC 4.2.1.20, see below) nor have their Trp derivatives been charged onto the respective tRNA by the enzyme TrpRS. Similarly, the Trp-like blue amino acid azulenyl-alanine (azulene) has not been a substrate for the native translation machinery even if it has been chemically charged onto tRNA.[66] Consequently, the related synthetic Trp analogs without –NH- within the ring system at position 1 have not served as substrates for Trp´s metabolism and translation.[64]

Thus, the benzene ring of the Ind moiety is apparently a much more permissive target for chemical transformations than the pyrrole ring in order to gain chemical and structural diversifications. The attempts to substitute Trp with its analogs started in the 1950s, when Prestidge and Pardee, Brawerman, and Ycas reported experiments on the incorporation of 7-azatryptophan [(7-Aza)Trp] and 2-azatryptophan [(2-Aza)Trp] into the proteome of *Escherichia coli* (*E. coli*).[67,68]

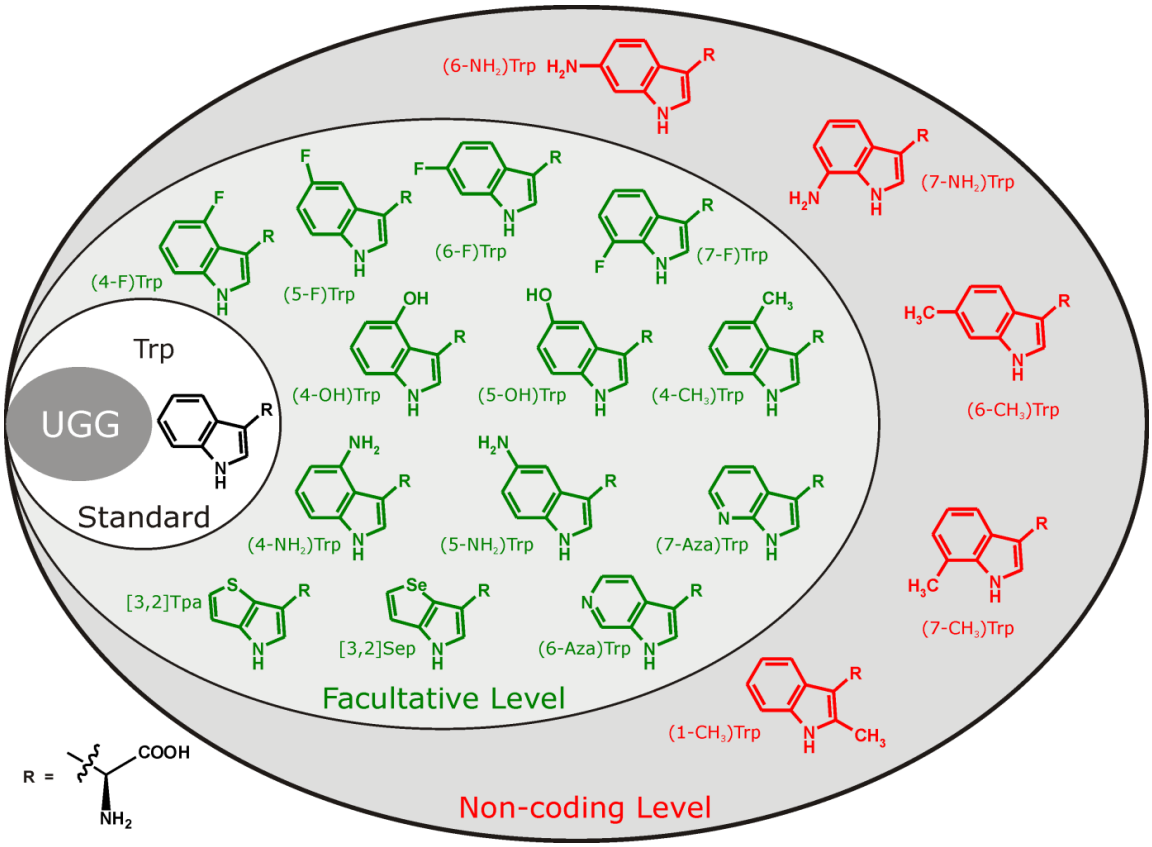


Fig. 5 **Expanded genetic code of Trp.** This scheme depicts examples of Trp analogs and surrogates, which have already been successfully incorporated into proteins instead of Trp (green). These non-canonical amino acids belong to the "Facultative coding Level". Trp analogs and surrogates, which have not been incorporated into any protein yet (red), belong to the "Non-coding Level". Trp: tryptophan; (4-F)Trp: 4-fluorotryptophan; (5-F)Trp: 5-fluorotryptophan; (6-F)Trp: 6-fluorotryptophan; (7-F)Trp: 7-fluorotryptophan; (4-OH)Trp: 4-hydroxytryptophan; (5-OH)Trp: 5-hydroxytryptophan; (4-CH_3)Trp: 4-methyltryptophan; (4-NH_2)Trp: 4-aminotryptophan; (5-NH_2)Trp: 5-aminotryptophan; (7-Aza)Trp: 7-azatryptophan; [3,2]Tpa: β-(thieno[3,2-*b*]pyrrolyl)-L-alanine; [3,2]Sep: β-(selenolo[3,2-*b*]pyrrolyl)-L-alanine; (6-Aza)Trp: 6-azatryptophan. R denotes $H_2CCH(NH_2)COOH$.

One decade later, Schlesinger and Schlesinger reported the enzyme alkaline phosphatase to be the first protein, specifically *in vivo* labeled with these amino acids and introduced fluorotryptophan-substituted proteins as tools for ^{19}F-NMR analyses.[69,70] In the last decade, Szabo, Ross, and their coworkers further extended this repertoire by introducing 5-hydroxytryptophan as a useful intrinsic fluorescent probe. Recently, our group demonstrated that selenophen- and thienyl-containing Trp-surrogates represent interesting pharmacologically active

substances and markers for protein X-ray crystallography.[64,71,72,27] Fig. 5 shows examples of Trp analogs (green: incorporated successfully; red: still under examination).

Trp analogs can also be added to the cell medium as analogs of Trp´s metabolic precursor Ind. This demonstrates simple but effective coupling of the expanded genetic code with the existing metabolic pathway. After the cellular uptake of Ind (or the Ind analog), the synthesis of the related Trp is catalyzed by the enzyme TrpS. Finally the analog can be charged onto the $tRNA^{Trp}$ by TrpRS, as already described in chapter 3.2, and further participate in the translation cycles of the ribosome.

3.9.4 Nitrogen-containing tryptophan analogs

Among the Trp analogs, nitrogen-containing Trp or Ind analogs occupy a central position. Compared to aliphatic amines and imines, they have distinct chemical properties and resemble the purine bases of DNA, sharing their capacity of pH-sensitive intramolecular charge transfer (Fig. 6).[29] The nitrogen atom can be located either exocyclic or endocyclic of the Ind moiety.

In the 1970s and 1980s, studies with aminoindoles in solution demonstrated that the addition of electron-donating atoms leads to intramolecular charge transfer, which is highly sensitive to pH changes.[63,62] Amino substituents, i.e. exocyclic nitrogen atoms, of aromatic ring systems are less basic in the excited than in the ground state since the protonation in the excited state affects the resonant integration of free electrons into conjugated ring systems. This causes a charge migration, which can be seen as a blue- or red-shifted spectroscopic band. For that reason, the presence of an amino group at the Ind ring of Trp results in a spectral shoulder in absorbance and a significant shift of the fluorescence emission maximum when compared to Trp. The emission maximum can be either red- or blue-shifted, depending on the position of the electron-donating amino group in the Ind moiety. All Trp residues of barstar (chapter 6.1.2) have successfully been replaced by both 4- and 5-aminotryptophan via SPI (chapter 3.4).[73,28]

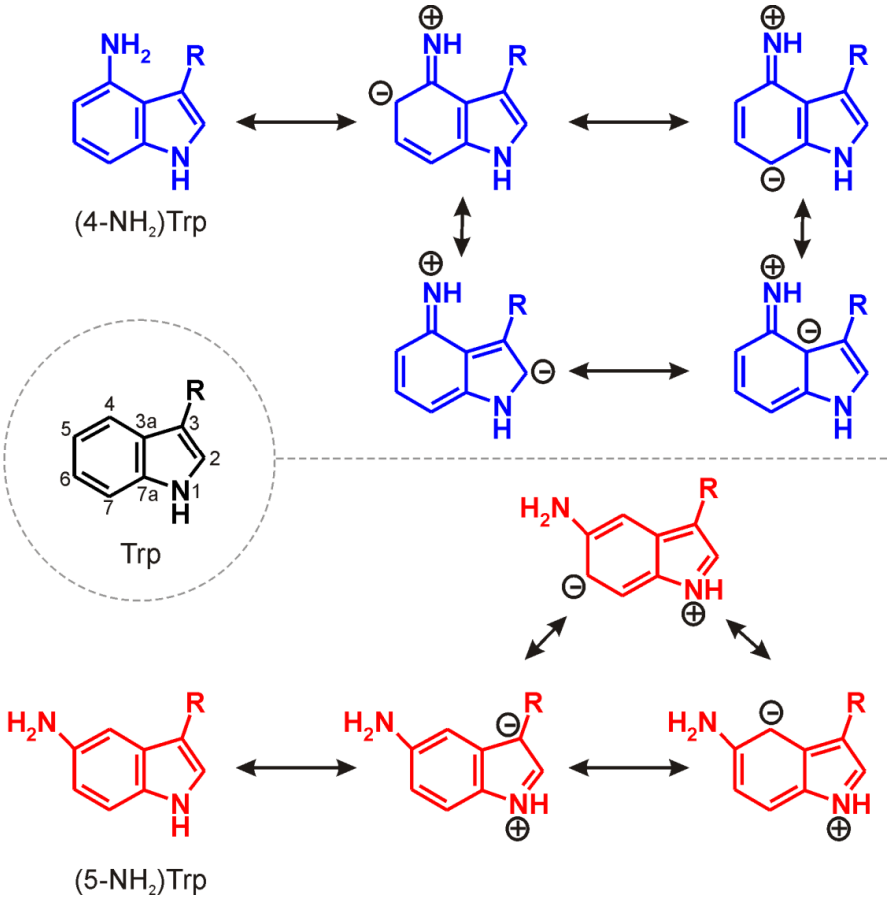


Fig. 6 **Charge transfer of aminotryptophans.** 4- and 5-aminotryptophan are exocyclic nitrogen-containing Trp analogs and are capable of charge transfer. This feature, which is present in nucleobases as well, is a consequence of the presence of an amino group as good electron donor in the indolic moieties. Such augmentation of the indole ring system facilitates charge transfer and leads to an increased number of mesomeric structures. R denotes $H_2CCH(NH_2)COOH$.

The pH sensitivity pronounced in the fluorescence emission spectra of the substituted proteins originates exclusively from cation → anion transitions of the aminoindoles. Thus, any pH insensitive protein, in respect of fluorescence, can be converted into a pH sensitive, since these properties result from the intrinsic properties of the incorporated aminotryptophans.[29,14]

Above all, the incorporation of 4-aminotryptophan into the enhanced cyan fluorescent protein (ECFP) represents a paradigm for the generation of tailor-made proteins via expanding the genetic code.[74] ECFP is a mutant of the green fluorescent protein (GFP) of *Aequorea victoria*, in which the Tyr66 was replaced by a Trp residue.[75,76] The chromophore is post-translationally formed in an

autocatalytic way of the residues 65-67. By introducing an exocyclic electron-donating amino group into the chromophore of ECFP, new perspectives for designing protein-based sensors have been opened up. Before, it was commonly perceived that properties such as high extinction coefficients, quantum yields, a monomeric state, and efficient folding, could only be achieved via multiple mutations in the target gene sequence.[77] Strikingly, the substitution of both Trp-residues (Trp57 and Trp66) in ECFP via SPI-method with 4-aminotryptophan resulted in a gold fluorescent protein (GdFP). This variant exhibits the largest red-shift in emission and the highest thermodynamic stability among all autofluorescent proteins known to date. Its emission wavelength (~570 nm) is similar to the ones of *ds*Red of *Discosoma striata* and of *as*FP595 of *Anemonia sulcata*.[78-80]

However, the broader *in vivo* use of GdFP as a fluorescent marker for cellular biology is currently limited due to its relatively large size and also the need for changes on the DNA level of the target sequences.[74,29] Furthermore, the extraordinary spectral properties of the GFP derived fluorescent proteins result from the posttranslationally formed chromophore in the context of a properly folded protein and a surrounding protein matrix. Therefore, it is not simple to transfer the preferable features of the GFP chromophores to any desired protein and alternatives for intrinsic fluorescent probes are needed. In addition, the chromophore of ECFP does not accept every Trp analog within its structure: While the incorporation of 4-aminotryptophan yielded a spectrally enhanced and stable protein, 5-aminotryptophan could never be introduced into the ECFP chromophore, although translationally active.

3.9.5 Azatryptophans – endocyclic nitrogen derivatives of tryptophan

Although GFP derived proteins are routinely used as fluorescent tags for bioimaging purposes, they are not generally applicable for all desired studies and alternatives are needed. An ideal chromophore should fulfill a number of prerequisites: It has to be biocompatible, incorporated into the target protein sequences by the endogenous translational apparatus, and should not require posttranslational modifications or extensive host-engineering.

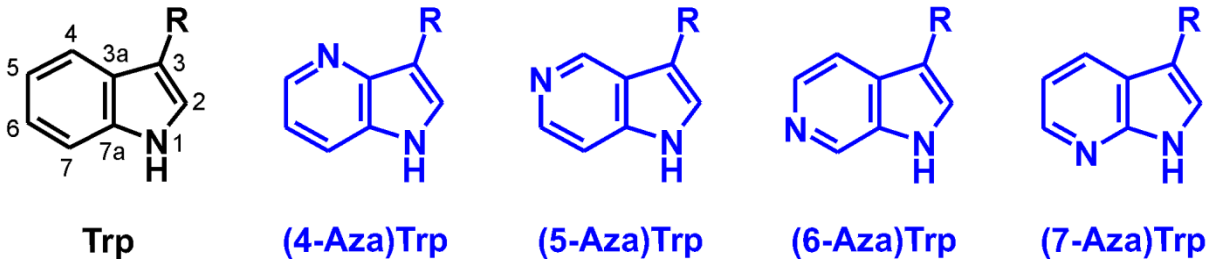


Fig. 7 Chemical structures of the canonical amino acid tryptophan and its aza-analogs. Trp: Tryptophan (with atomic numbering); (4-Aza)Trp: 4-azatryptophan; (5-Aza)Trp: 5-azatryptophan; (6-Aza)Trp: 6-azatryptophan; (7-Aza)Trp: 7-azatryptophan. R denotes $H_2CCH(NH_2)COOH$.

In addition, this chromophore should be non-invasive, introduces only minimal structural and functional perturbations into the target structures.

Azatryptophans [(Aza)Trps] meet the above described criteria (Fig. 7). In these Trp isosteres one of the endocyclic –CH– groups of Ind is substituted with nitrogen. This comprises not only the smallest possible structural alteration of all known Trp analogs but also leads to dramatic changes in the photophysics of the aromatic system. Azaindoles [(Aza)Inds] resemble nucleic acid purine bases since they contain a pyridyl nitrogen with a lone electron pair capable of hydrogen bonding and excited-state tautomerization.[81,82] Consequently, as protein building blocks, (Aza)Trps can furnish proteins with new functions, nature has exclusively reserved for nucleic acids, such as charge conductivity, a crucial feature for the design of future protein-based optoelectronic data storage devices or molecular wires for information transfer.[83,29]

To date, Trp in bacterial proteins has been replaced by 7-azatryptophan [(7-Aza)Trp] and (2-Aza)Trp, either for testing the antimetabolic properties of the Trp isosteres or for collecting information on the local environment, i.e. solvent accessibility, of Trp residues in proteins with unknown three-dimensional structures.[30,84] However, early reports on (2-Aza)Trp incorporation into proteins could not be confirmed in recent studies.[66] Steady-state absorbance and fluorescence of (7-Aza)Trp sufficiently differ from Trp, which facilitates its selective excitation and fluorescence detection (λ_{max} ~ 360 nm in a hydrophobic environment).

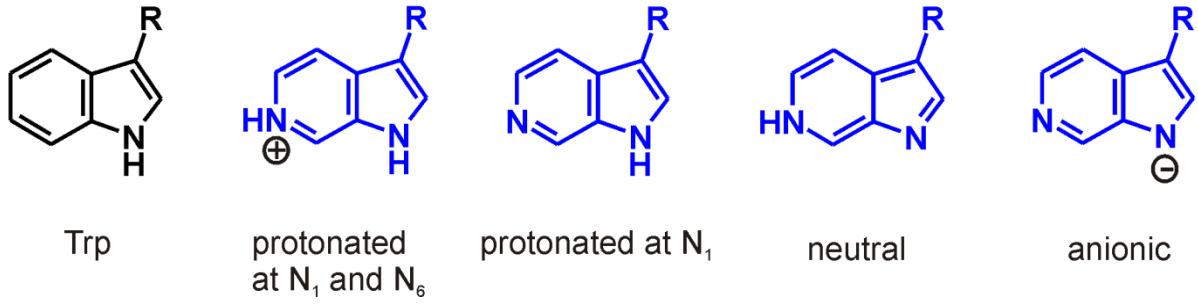


Fig. 8 **The canonical amino acid Trp and its analog 6-azatryptophan.** According to Twine *et al.*, 6-azatryptophan is capable of charge transfer and related mesomeric structures are depicted.[85] R denotes $H_2CCH(NH_2)COOH$.

Twine and Szabo reported an altered spectral behavior in respect to pH-dependent protonation/deprotonation processes for both 6-azatryptophan [(6-Aza)Trp] and (7-Aza)Trp (Fig. 8).[86] In aqueous solvents, (7-Aza)Trp´s fluorescence is strongly quenched, which limits its general use as a fluorophore.[87] The identification of other (Aza)Trps with reduced quenching sensitivity and pronounced red-shifts in physiological solutions would overcome this problem.

Recent photophysical studies on free 4-azaindole [(4-Aza)Ind] and 5-azaindole [(5-Aza)Ind] revealed a strong red-shifted fluorescence (above 400 nm, i.e. blue fluorescence) when compared with Ind in aqueous solutions.[88] Additionally, they exhibited a diminished quenching sensitivity compared to 7-azaindole [(7-Aza)Ind]. Taken together, (4-Aza)Ind- and (5-Aza)Ind exhibit all properties of suitable fluorescent probes and their incorporation into model proteins are a central topic of this thesis.

3.10 Bioorthogonal chemical reactions in proteins – click chemistry

The [3+2] cycloaddition (or 1,3-diploar cycloaddition) of an unsaturated carbohydrate with a 1,3-dipole, although already observed in 1893, was first described in detail by Huisgen.[89-91] He reported the formation of a triazole out of azide and alkyne and introduced the possible use of this reaction in organic syntheses. Recently, the studies of Sharpless, Meldal, and their coworkers expanded the scope of the Huisgen cycloaddition by using copper(I) as reaction catalyst.[92,93] This modification enabled mild reaction conditions in aqueous milieus with the predominant formation of the 1,4-regioisomer. In the literature,

this modified reaction is called copper(I)-catalyzed Huisgen cycloaddition (CCHC), also known as "click chemistry" –a term introduced by Sharpless.[94] The principle of the reaction is depicted in Fig. 9B.

Bertozzi, Tirrell, Schultz, and our group were among the first to combine the expansion of the genetic code with click chemistry, which opened up novel possibilities for rationally controlled posttranslational protein conjugations.[95-98] For example, a blue fluorescent protein, generated by the introduction of (4-Aza)Ind, would be highly useful when equipped with another chromophore for the measurement of fluorescence energy resonance transfer (also Förster energy resonance transfer, FRET) for intramolecular (single molecule spectroscopy), or intermolecular interactions (receptor-ligand binding).

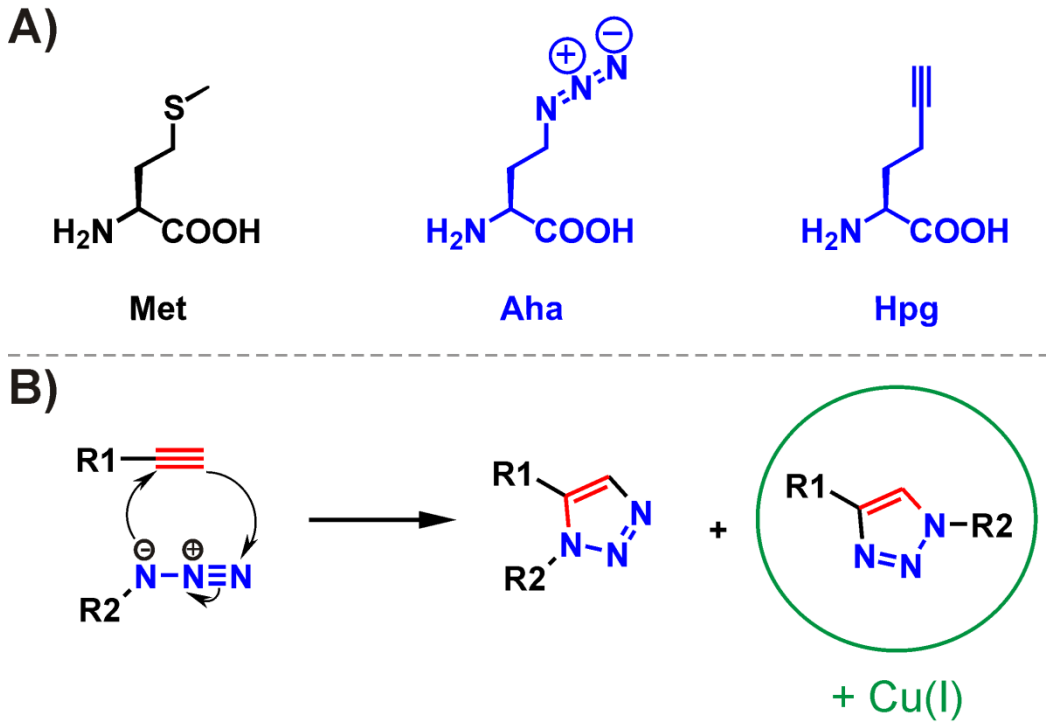


Fig. 9 **Methionine analogs for bioorthogonal chemical reactions. A)** Chemical structures of methionine (Met), azidohomoalanine (Aha), and homopropargylglycine (Hpg). Aha and Met comprise terminal azide and alkyne groups, respectively. They are regarded as almost orthogonal, nearly absent from nature. **B)** [1,2,3]-triazole formation by [3+2] cycloaddition of an alkyne and an azide by Huisgen.[90,91] After addition of copper ions, the reaction takes place at mild, e.g. physiological, conditions and only the 1,4-stereoisomer is formed. This reaction represents the copper(I)-catalyzed Huisgen cycloaddition (click chemistry).[92,93]

To combine these bioorthogonal reactions with protein engineering, the incorporation of either an azide or an alkyne function into protein sequences is necessary. In the scope of an expanded genetic code, Met analogs offering alkyne and azide groups are of particular interest. Met, like Trp, is a very rare amino acid in protein, which enables also pseudo site-directed incorporation of its analogs in proteins. In the last years, they were used as reaction partners in CCHC reactions at proteins. Two particular Met-analogs are especially suited for these chemical conjugation reactions and were already incorporated into proteins: Homopropargylglycine (Hpg) and azidohomoalanine (Aha) (Fig. 9A).[99,100,98] Translationally active Hpg comprises an alkyne and Aha an azide group.

Once incorporated via SPI into proteins, these chemical handles enable the individual and diverse functionalization of any protein. They will certainly increase the existing repertoire of posttranslational modifications and would allow the specific copying of the modification pattern of eukaryotic proteins to proteins recombinantly expressed in bacterial cells. Furthermore, the pool of modifications could be enhanced by synthetic chemical compounds, the attachment of tags for protein purification or visualization, and many more.

4 Material

4.1 Equipment

- Autoclave (Varioklav Dampfsterilisator Typ 500 E; H+P Labortechnik GmbH, Oberschleißheim, Germany)
- Balances (TE1502S; BP211D; Sartorius, Göttingen, Germany; GB2002; PC4400 Delta Range; Mettler-Toledo GmbH, Giessen, Germany)
- CD spectropolarimeter (Jasco J-715 Spectro polarimeter; Temperature control by Peltier FDCD attachment PFD-350S/350L; JASCO International Co., Ltd., Tokyo, Japan)
- Centrifuges (Avanti J-25 Centrifuge; Avanti J-20 XP Centrifuge; Beckmann, Munich, Germany; Centrifuge 5415 C/D; Zentrifuge 3200; Eppendorf, Hamburg, Germany; Universal 32R; Hettich Zentrifugen, Tuttlingen, Germany)
- Centrifuge rotors (JA 25.50; JLA 8.1000; JLA 10.500; Beckmann, Munich, Germany)
- Cuvettes (Hellma 104.002-QS; Hellma 104.002F-QS; Hellma 110-QS; Hellma, Müllheim, Germany)
- Electroporator (Electroporator 1000; Stratagene, La Jolla, CA, USA)
- Fluorescence spectrometer (Luminescence Spectrometer LS50B; PerkinElmer Life Sciences, Boston, MA, USA; F-4500 Fluorescence Spectrophotometer; Hitachi, Feldkirchen, Germany)
- FPLC instruments (Äktaexplorer; Äktabasic (GE Healthcare, Munich, Germany; columns: Ni-NTA column; HiTrap Q Sepharose; GE Healthcare, Munich, Germany)
- HPLC system (C18-RP-HPLC, Waters Alliance 2695 with photodiode array detector 996 and fluorescence detector 2475; column: Waters Xterra RP C18 3.5 µm [2.1x100 mm]; Waters, Eschborn, Germany)

- Incubator (Thermomixer comfort; Thermomixer compact; Eppendorf, Hamburg, Germany; Incubator 3033; GFL, Burgwedel, Germany; Infors-HT AG, Bottmingen, Switzerland)
- Magnetic stirrer (MR 3001; Heidolph, Kehlheim, Germany; Ikamag REO; Ika-Combimag REO; IKA® Werke GmbH & Co. KG, Staufen, Germany)
- Mass spectrometer (Pe SCIEX API 165; PerkinElmer Life Sciences, Boston, MA, USA; MicroTOF LC; Bruker Daltonik GmbH, Bremen, Germany; Q-TOF Ultima; Waters, Milford, MA, USA)
- Microscope (BX 51TRF; Olympus, Hamburg, Germany)
- PH meter (MP 220; Mettler-Toledo GmbH, Giessen, Germany)
- Sonifier (Sonifier 450 Macrotip; Branson, St. Louis, MO, USA)
- Sterile bench (Lamin Air HA244GS; Heraeus, Hanau, Germany)
- UV/Vis spectrophotometer (Ultrospec 6300 pro; Amersham Biosciences, Buckinghamshire, UK; UV/VIS spectrometer lambda 19; PerkinElmer Life Sciences, Boston, MA, USA)
- UV lamp (Olympus MT 20 monochromator; Ina-shi, Nagano, Japan)
- Vortex (Reax 1R; Heidolph, Kehlheim, Germany; Vortex Genie 2; Bender & Hobein AG, Zurich, Switzerland)
- *N*-terminal sequencer (492 protein sequencer; Applied Biosystems, Darmstadt, Germany)

4.2 Chemicals

Unless stated otherwise, all chemicals were purchased by Merck (Darmstadt, Germany), Sigma, or Fluka (Taufkirchen, Germany) and were of highest purity grade. Azaindoles [(Aza)Inds] were purchased from Molekula (Nienburg/Weser, Germany) and Biosynth (Staad, Switzerland). Homopropargylglycine (Hpg) and azidohomoalanine (Aha) were synthesized and purified by Dr. Shouliang Dong and Lars Merkel, respectively.

All aqueous buffers and solutions were prepared using bidestilled H_2O (Millipore, Billerica, MA, USA) and autoclaved or sterile filtered if required.

4.3 Buffer and solutions

- Phosphate buffered saline (15 mM $Na_2HPO_4 \cdot 2H_2O$; 1.8 mM KH_2PO_4; 140 mM NaCl; 2.7 mM KCl; pH 7.4)
- 6x SDS-PAGE sample buffer (450 mM Tris·HCl, pH 6.8; 3.6% SDS, 0.2% bromophenol blue; 30% glycerol; 45% β-mercaptoethanol)
- PAGE Running buffer (0.19 M glycine; 25 mM Tris; 3.5 mM SDS)
- Coomassie staining solution (0.1% Coomassie; 25% ethanol; 8% acetic acid)
- Coomassie destaining solution (25% ethanol; 8% acetic acid)

4.4 Media

For bacterial growth, fermentation, and protein expression two different media were used: LB medium and New Minimal Medium (NMM).[101]

The components of LB medium, Bacto™ Tryptone and Bacto™ Yeast Extract were purchased from BD Biosciences (San José, CA, USA). 10 g Bacto™ Tryptone, 5 g Bacto™ Yeast Extract, and 10 g NaCl were dissolved in 1 L H_2O and autoclaved before usage.

For the incorporation of non-canonical amino acids into protein via SPI method, NMM was used as fermentation medium. For its preparation, first an amino acid mix at pH 7.0 was prepared which contained all amino acids except Tyr and the amino acid/s which were to be exchanged, i.e. Trp or Trp and Met). For this purpose, 0.5 g of each amino acid (as lyophilized powder) were dissolved in 22 mL 1 M KH_2PO_4, 48 mL 1 M K_2HPO_4 and H_2O *add* 1 L. 0.5 g Tyr were dissolved in 50 mL H_2O separately using 32% HCl and added to the amino acid mix. The solution was sterile filtered before usage. Sterile NMM components were mixed as follows: 7.5 mM $(NH_4)_2SO_4$; 8.5 mM NaCl; 22.5 mM KH_2PO_4; 50 mM K_2HPO_4; 1 mM $MgSO_4$; 20 mM glucose; 100 mL/(L NMM) amino acid mix; 1 µg/(mL NMM) $CaCl_2$; 1 µg/(mL NMM) $FeSO_4$; trace elements 0,01 ng each/(mL NMM)[101]; 10 µg/(mL NMM) thiamine; 10 µg/mL NMM biotin; and 6.7 g Difco™ yeast nitrogen base without amino acids (BD Biosciences, San José, CA, USA).

4.5 Enzymes

- Lysozyme (application: protein purification)
- DNAse I (application: protein purification)
- RNAse A (application: protein purification)
- Tryptophan-Synthase from *Salmonella typhimurium* (TrpS; EC 4.2.1.20; application: pre-incubation of Ind analogs prior to the addition to the culture medium; expressed and purified according to the protocol of Yang *et al.*[102])
- Trypsin (application: mass spectrometry; Promega, Madison, WI, USA)

4.6 Protein molecular weight marker

- PageRuler™ Prestained Protein Ladder; Fermentas, St. Leon-Rot, Germany
- BenchMark™ Protein Ladder; Invitrogen, Karlsruhe, Germany

4.7 Plasmids

Plasmids used in this study were kindly provided by Tatjana Krywcun and Petra Birle.

4.7.1 pQE-80L Annexin A5

The pQE-80L vector (Qiagen, Hilden, Germany; modified by Tatjana Krywcun and Petra Birle) carries the annexin A5 (anxA5) gene with *N*-terminal His-Tag, the gene providing ampicillin resistance, the origin of replication Col E1 and the isopropyl-β-D-thiogalactopyranoside (IPTG; Applichem, Darmstadt, Germany) inducible promoter T5.

4.7.2 pQE-80L Barstar

The plasmid pQE-80L (Qiagen, Hilden, Germany, modified by Tatjana Krywcun and Petra Birle) Barstar carries the gene of cysteine-free pseudo wild-type barstar (ψ-b*; with mutations Pro28Ala/Cys41Ala/Cys83Ala)[1] from *Bacillus amyloliquefaciens* under the control of the T5-promotor. The gene was inserted

using the restriction enzymes *EcoR*I and *Hind*III, therefore ψ-b* is expressed without tag. Furthermore, the vector provides ampicillin resistance.

4.8 Antibiotics

- Ampicillin (working concentration 100 µg/mL)

4.9 Bacterial strains

All applied bacterial strains are derived from the *E. coli* K12 mutant.

- ATCC 49980 (auxotrophy: Trp; genotype: WP2 (trp-, uvrA-, malB-); American Type Culture Collection/LGC Promochem, Wesel, Germany)
- JE 7345 (auxotrophy: Trp, Met, Ile, Pro, His, Arg; genotype: ileS, ara, proC, galK, trp, his, argG, xyl, mtl, metA or B; National BioResource Project, Japan)

The bacterial strain ATCC 49980 was used for the expression of anxA5 and JE 7345 for the expression of ψ-b*.[1]

4.10 Software

Except standard software, Origin 6.1G (OriginLab Corporation, Northampton, MA, USA) was applied for data analysis. MAIN was used for atom replacement and energy minimization by Dr. M. Debela (chapter 6.3.3).[103]

5 Methods

5.1 Spectroscopy and spectrometry

5.1.1 UV/VIS-Spectroscopy

Molar extinction coefficients (ε_M) of (Aza)Inds were acquired by UV_{280} measurement with the UV/VIS spectrometer lambda 19 (PerkinElmer Life Sciences, Boston, MA, USA) using sample concentrations of 4 µM in phosphate buffered saline (15 mM $Na_2HPO_4 \cdot 2H_2O$; 1.8 mM KH_2PO_4; 140 mM NaCl; 2.7 mM KCl; pH 7.4).

Protein concentrations were determined by UV_{280} measurement in their standard buffer also using the UV/VIS spectrometer lambda 19 (PerkinElmer Life Sciences, Boston, MA, USA) and calculated according to the Lambert-Beer-Equation $A = \frac{logI_0}{logI} = c \cdot d \cdot \varepsilon$, A = absorbance; ε = molar extinction coefficient; c = concentration; d = path length) subsequently.

5.1.2 Estimation of the protein concentration by Bradford protein assay

For calculation of the molar extinction coefficients (ε_M) of wt-anxA5 and variants, the protein concentrations were determined using the Bradford Protein Assay (Bio-Rad, Hercules, CA). According to the supplier's instructions, the protein concentration was calculated by comparison of the OD_{595} values to those of a bovine serum albumin calibration curve. As a quality control, the molar extinction coefficient of Trp-anxA5 obtained by Bradford assay favorably compares to its coefficient calculated by the software ProtParam provided by the Expasy Proteomics Server (www.expasy.ch).[104]

5.1.3 Fluorescence spectroscopy

Fluorescence spectra of proteins were measured using the luminescence spectrometer LS50B (PerkinElmer Life Sciences, Boston, MA, USA) at 20 °C in 100 mM Tris·HCl, pH 7.5. The protein concentration was 0.5 µM. Spectra were recorded from 300 nm – 500 nm with an excitation wavelength of 280 nm (emission/excitation slits 10/10 nm).

Fluorescence spectra of Ind and (Aza)Ind were recorded at 20 °C upon excitation at the following maximum excitation wavelengths: Ind 269 nm; (4-Aza)Ind 288 nm; (5-Aza)Ind 266 nm; (7-Aza)Ind 288 nm (excitation/emission slits 6/6 nm). Samples were dissolved in 100 mM Tris·HCl, pH 7.5, or isopropanol.

5.1.4 Determination of quantum yields (QY)

QY of (Aza)Inds were recorded in 100 mM Tris·HCl buffer, pH 7.5. The absorbance maximum was determined using the UV/VIS spectrometer lambda 19 (PerkinElmer Life Sciences, Boston, MA, USA). Fluorescence spectra were recorded with the luminescence spectrometer LS50B (PerkinElmer Life Sciences, Boston, MA, USA) at 20 °C at the determined maximum absorbance wavelengths as follows and emission/excitation slits of 5/5: Ind 269 nm; (4-Aza)Ind 288 nm; (5-Aza)Ind 266 nm; (7-Aza)Ind 288 nm. Excitation spectra were corrected automatically by the instrument for artifacts originating from the xenon lamp (energy output is wavelength and time-dependent) and from the excitation monochromator (efficiency is wavelength-dependent). Emission spectra were corrected with respect to artifacts originating from the emission monochromator (efficiency is wavelength-dependent) and the sample multiplier (sensitivity is wavelength dependent). Correction files for emission spectra are yearly updated and provided by the instrument´s manufacturer (PerkinElmer Life Sciences, Boston, MA, USA). All measurements were made with constant detector current. Fluorescence spectrum integration was performed with Origin 6.1G (OriginLab Corporation, Northampton, MA, USA). QYs were calculated according to Lakowitz (1999) with Trp as standard (QY = 0.14 in H_2O).[60]

5.1.5 Circular dichroism (CD)

CD spectra and melting curves of Trp-anxA5 and (Aza)Trp variants were measured in phosphate buffered saline at protein concentrations of 0.2 mg/mL using the CD Spectropolarimeter JASCO J-715 (JASCO International Co., Ltd., Tokyo, Japan). Measurements were performed in 110-QS Hellma quartz cells (optical path-length: 0.1 cm) using temperature control by a Peltier type FDCD attachment (model PFD-350S/350L; JASCO International Co., Ltd., Tokyo, Japan). Ellipticity changes were recorded between 200 nm and 260 nm at 4° C

for CD spectra. The secondary structural proportion was calculated with CDNN CD spectra deconvolution (Version 2.1, University of Halle-Wittenberg, Germany) between 190 nm – 260 nm. Melting curves were measured at 222 nm as a function of temperature from 4° C – 95° C, gradient: 30 °C /h. The fraction of unfolded protein was calculated from the corresponding ellipticity data. T_m value and van't Hoff enthalpy (ΔH_m) were determined as described elsewhere.[28]

5.1.6 Mass spectrometry

(Aza)Trp-anxA5s were digested over night at 37 °C by modified trypsin (Promega, Madison, WI, USA) according to the supplier's instructions. The resulting peptide mixture was separated via C_{18}-RP-HPLC (Reverse phase-high performance liquid chromatography, Waters Alliance 2695 with photodiode array detector 996 and fluorescence detector 2475; column: Waters Xterra RP C_{18} 3.5 µm [2.1x100 mm]; Waters, Eschborn, Germany) and the fluorescence of the eluate was recorded on-line after excitation at 279 nm. Eluted fractions showing the characteristic (Aza)Ind fluorescence (λ_{Ex} = 400 - 420 nm) were collected and further characterized by off-line-nano-MS using a Q-TOF Ultima (Waters, Milford, MA) spectrometer in MS and MS/MS mode.

Masses of ψ-b* and its variants (0.5 mg/mL in 100 mM Tris·HCl, pH 7.5) were determined by high resolution mass analysis using liquid chromatography (LC) coupled with the electrospray ionization-time of flight (ESI-TOF) mass spectrometer MicroTOF-LC from Bruker Daltonics (Bremen, Germany). Samples were separated by Symmetry C_4 column (Waters, Eschborn, Germany) with a flow rate of 250 µL/min and 15 min linear gradient from 20% - 80% acetonitrile/0.05% trifluoroacetic acid.

5.1.7 Amino acid hydrolysis

Proteins were hydrolyzed for 24 h with 6 M HCl containing 2.5% (v/v) thioglycolic acid. Hydrolysates were derivatized using the AccQ·Tag Ultra amino acid analysis kit (Waters, Eschborn, Germany) according to the manufacturer's instructions. Derivatized hydrolysates were analyzed by high resolution mass analysis using HPLC (1100, Agilent Technologies, Waldbronn, Germany) coupled with the ESI-TOF mass spectrometer MicroTOF-LC from Bruker Daltonics (Bremen, Germany). Samples were separated by Halo C_8

column (3.0 x 150 mm, 2.7 µm; Advanced Materials Technology, Wilmington, DE, USA) with a flow rate of 0.35 mL/min and the following gradient of buffers A and B (AccQ·Tag Ultra standard buffers):

Time [min]	A [%]
0.0 → 2.0	98.0
2.0 → 28.0	91.0
28.0 → 35.0	84.0
25.0 → 41.0	79.0
41.0 → 45.5	40.0
45.5 → 48.0	40.0
48.0 → 48.5	100.0
48.5 → 52.0	100.0

Table 1 Buffer gradient of amino acid hydrolysis.

(Aza)Trps were also analyzed by the above described procedure to determine both retention times and recoveries. (4-Aza)Trp elutes after around 13 min, (5-Aza)Trp together with threonine after around 14 min, and (7-Aza)Trp after 23.4 min. The Trp signal appears after around 44 min. Since the signal of (5-Aza)Trp was not separated from the threonine peak, the incorporation level of (5-Aza)Trp was calculated from the reduction of the Trp peak and the increase of the threonine peak in comparison to the parent anxA5 protein.

5.1.8 N-*terminal sequencing*

After the gel run, the proteins were blotted on a PVDF membrane (Westran S; Whatman, Sanford, ME, USA), which was activated in methanol according to the manufacturer's protocol. Blotting was performed for 2,5 h at 200 mA and the membrane stained with Coomassie as outlined in the manual, subsequently (Transfer buffer: 2.9 g glycine, 5.8 g Tris, 0.37 g SDS, 200 mL methanol, adjust to 1 L H_2O).

The Coomassie stained bands were cut out and subjected to *N*-terminal sequence analysis using a 492 protein sequencer (Applied Biosystems Darmstadt, Germany) according to the instructions of the manufacturer.

5.2 Polyacrylamide gel electrophoresis (PAGE)

For separation of proteins according to their size, first an appropriate amount of 6x sodium dodecyl sulfate (SDS)-PAGE sample buffer (chapter 4.3) was added to the protein solution and the mix was heated to 95 °C for 5 min.

Second 12%, 17%, or 20% polyacrylamide gels with 5% stacking gels were prepared and the samples separated by their molecular weight at 120 – 200 V (PAGE Running buffer: chapter 4.3).

Composition of resolving gels (50 mL):

- 12% (16.5 mL H_2O; 20 mL 30% acryl-bisacrylamide mix; 12.5 mL 1.5 M Tris·HCl, pH 8.8; 0.5 mL 10% SDS; 0.5 mL 10% APS; 0.02 mL TEMED)
- 17% (7.98 mL H_2O; 28.5 mL 30% acryl-bisacrylamide mix; 12.5 mL 1.5 M Tris·HCl, pH 8.8; 0.5 mL 10% SDS; 0.5 mL 10% APS; 0.02 mL TEMED)
- 20% (3.5 mL H_2O; 33 mL 30% acryl-bisacrylamide mix; 12.5 mL 1.5 M Tris·HCl, pH 8.8; 0.5 mL 10% SDS; 0.5 mL 10% APS; 0.02 mL TEMED)

Composition of stacking gels (50 mL):

- 34 mL H_2O; 8.5 mL 30% acryl-bisacrylamide mix; 6.25 mL 1.5 M Tris·HCl, pH 6.8; 0.5 mL 10% SDS; 0.5 mL 10% APS; 0.05 mL TEMED

Third the gels were exposed to UV light to check for successful (4-Aza)Trp incorporation and stained by Coomassie staining solution, and finally destained by Coomassie destaining solution (chapter 4.3).

5.3 Microbiological methods

5.3.1 Production of electrocompetent cells

First cells were incubated over night at 37 °C in 5 mL LB medium without antibiotics.

On the following day, 1 L LB medium without antibiotics was inoculated using the 5 mL over night culture and incubated at 37 °C and 220 rpm until the OD_{600} of the cell culture was between 0.6 and 0.8. The cells were harvested and washed twice with ice-cold 10% glycerol, subsequently. Finally the cell pellet was resuspended in 5 mL 10% glycerol, aliquots à 100 µL were frozen in liquid nitrogen, and kept at -80 °C until further use.

5.3.2 Transformation

Bacterial cells were transformed by electroporation. For this purpose, a 100 µL competent cell aliquot of the appropriate cell strain was mixed with ~1 µg of the required plasmid in an electroporation cuvette. The cell suspension was shocked by 1650 V, mixed with 1 mL chilled LB medium, and transferred to a sterile eppendorf cup, subsequently. In the following, the cells were incubated at 37 °C and 800 rpm for 1 h. Finally the cells were distributed on agarose gel plates with the appropriate antibiotics.

On the next day, single colonies were transferred to 5 mL of LB medium with the appropriate antibiotics and incubated over night at 37 °C and 220 rpm.

5.3.3 Limitation test

For the incorporation of non-canonical amino acids into proteins during protein expression it is absolutely necessary that the applied cell strain is auxotrophic for the respective canonical amino acid. Therefore, prior to the incorporation experiment, it is essential to prove the cell´s auxotrophy and determine the optimal amino acid concentration for biomass production by a limitation test. For this purpose, 5 mL NMM with appropriate antibiotics and different concentrations of Ind, L-Trp, and L-Met were inoculated with 50 µL of pre-culture. These suspensions were incubated over night at 37 °C and 220 rpm and on the next day, cell growth was checked by measurement of OD_{600}. The amino acid

concentration which permitted cell growth up to an OD_{600} between 0.6 and 0.8 was utilized as limiting concentration for biomass production for incorporation experiments.

5.3.4 Expression test

To select the best expression clone, 1 mL of each over-night cell culture was transferred to a sterile eppendorf cup and the protein expression was induced by adding 1 mM IPTG. The cell cultures were incubated at 800 rpm and 30 °C in the case of anxA5 (6 h) or at 27 °C in the case of ψ-b* (4 h). In parallel, a non-induced sample (without addition of IPTG) was prepared.

After incubation, the cells were first harvested, second resuspended in 50 µL H_2O, and finally heated to 95 °C for 5 min after addition of 10 µL 6x SDS sample buffer. Afterwards, protein expression was checked by SDS PAGE and the best expressing clone was selected for subsequent protein expression.

5.3.5 Protein expression and purification of annexin A5

The selected bacterial clone was grown in NMM with 100 µg/mL ampicillin containing 5.5 µM Ind as natural substrate until depletion of Ind. This time-point is indicated by stable OD_{600} between 0.6 – 0.8. (Aza)Inds were added to the medium at 1 mM final concentration and 30 min later target protein expression was induced by addition of 1 mM IPTG. Protein expression was performed over night at 30 °C and 220 rpm. To support (Aza)Trp conversion, (Aza)Inds were partially converted into (Aza)Trps using recombinantly expressed and purified TrpS prior to addition to the expression batch (see 5.4.1).[102]

After cell harvest, cells were lysed by osmotic shock and the lysate cleared by high speed centrifugation (40.000 *x g*, 4 °C, 40 min).[105] Target proteins were purified from the supernatant using a Ni-NTA column (GE Healthcare, Munich, Germany) equilibrated with 50 mM sodium dihydrogen phosphate buffer (pH 8.0), 20 mM imidazole, and 300 mM NaCl. Proteins were eluted via imidazole gradient (20 – 250 mM), concentrated, and finally the purity was analyzed by 12% SDS-PAGE and Coomassie staining.

5.3.6 Protein expression and purification of barstar (ψ-b)*

For the expression of ψ-b*, the best expressing clone after the expression test was grown in NMM (with 100 µg/mL ampicillin) containing 4 µM L-Trp and 30 µM L-Met as canonical amino acid substrates at 37 °C and 220 rpm until depletion of both amino acids. 100 mg/L NMM D,L-Hpg and 100 mg/L NMM pre-converted (4-Aza)Ind were added to the culture and 30 min later target protein expression was induced by adding 1 mM IPTG. Protein expression was performed for 6 h at 27 °C and 220 rpm and success of expression was checked by loading 0.3 OD_{600} of cell material on a 20% SDS gel, subsequently.

Ψ-b* and its variants are expressed in inclusion bodies and were routinely refolded prior to purification as described elsewhere.[1]

5.4 Biochemical methods

5.4.1 Pre-conversion of azaindoles

In every bacterial cell, Ind and serine (Ser) are stoichiometrically converted to Trp by the intracellular enzyme TrpS. This conversion also takes place in case of (Aza)Inds which necessarily have to be transformed into (Aza)Trps to participate in the protein´s translation process. For accelerating this conversion and therefore reducing the pressure to the expressing cells, (Aza)Inds were pre-incubated with L-Ser using recombinantly expressed and purified TrpS.[102] The reaction mix routinely contained the following components (H_2O *add* 50 mL): 100 mg (Aza)Ind; 100 mg L-Ser; 5 mL 1 M Tris·HCl, pH 7.4; 2 mL 5 m NaCl; 0.1 mL 20 mM Pyridoxal-5´-phosphate (PLP); and 400 µL standard TrpS

All reaction batches were incubated in the dark and over night at room temperature. Afterwards, they were added to the cell culture 30 min prior to the induction of protein expression.

5.4.2 Copper(I)-catalyzed Huisgen cycloaddition (CCHC; click reaction)

For copper(I)-catalyzed Huisgen cycloaddition reactions (CCHC also click chemistry; chapter 3.10), the Met analogs Hpg or Aha were incorporated into ψ-b* using SPI (chapter 3.4).[92,93] Hpg and Aha provide reactive handles for individual bioorthogonal chemical transformations: Hpg exhibits an alkyne and

Aha an azide function. Both chemical functionalities have to be present to form the cyclic linker between the participating molecules. Only the Hpg protein variants were used for CCHC reactions and the azide function was either provided by a small reaction partner, an azido-sugar derivative or azido-polyethylene glycol (PEG, 570 Da), or by ψ-b* variant with incorporated Met analog Aha (Aha-Trp-ψ-b*).

1 mg/mL Hpg-(4-Aza)Trp-ψ-b* (in 50 mM sodium phosphate buffer, pH 7.4) was incubated for 18 h at room temperature with 3.75 mM copper sulfate, 3.75 mM ascorbic acid, 200 mM sodium phosphate buffer, pH 7.4, and required amounts of the respective azide component (1-azido-1-deoxy-β-D-glucopyranosid: 2.5 mg/mL; azido-PEG570: 5 mg/mL; Aha-Trp-ψ-b*: 1 mg/mL; reaction batch volume 200 µL). CCHC performance was checked by mass spectrometry.

6 Results and Discussion

6.1 The model proteins

6.1.1 Human annexin A5

Human annexin A5 (anxA5) belongs to the protein family of annexins. Annexins can bind to membranes in a calcium dependent manner and are in this way involved in the transport of calcium ions through membranes and in membrane fusion processes.[106,107] AnxA5 consists of 316 amino acids and comprises a molecular weight of ~35 kDa. In 1990, its three-dimensional structure was resolved by Huber and his coworkers, who reported a protein structure characterized by four homologous domains, each containing five α-helixes.[108,109] The structure displays a concave and a convex side, which offers five calcium ion binding sites. The *N*-terminal end of the peptide chain is located on the concave side of the protein.

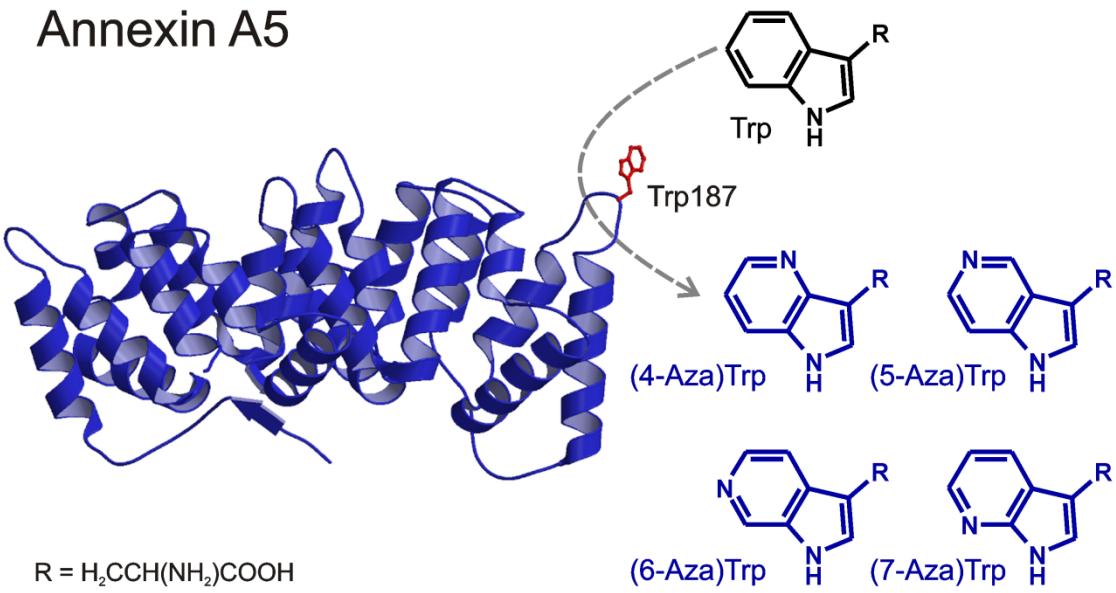


Fig. 10 Incorporation of (Aza)Trps into Annexin A5 (anxA5). Ribbon plot of human anxA5 (side view) with tryptophan at position 187 (Trp187) buried in the hydrophobic pocket at the convex side of the molecule.

AnxA5 was used as model protein for the incorporation of (4-Aza)Trp (Fig. 10) since it contains only one but essential Trp residue. Trp187 exists in two different conformations. It is either buried in a hydrophobic cavity at the convex side of the protein or solvent exposed after calcium ion addition.

6.1.2 Pseudo wild-type Barstar from Bacillus amyloliquefaciens

Barstar is the intracellular, single-domain inhibitor of the extracellular ribonuclease barnase from *Bacillus amyloliquefaciens.*[110-112] The bacterial defense protein barnase is expressed by the cells and secreted into the environment for the degradation of RNA of potential enemies. In order to prevent barnase damaging the cell´s own RNA during its expression and secretion, its inhibitor barstar is simultaneously expressed. Barstar binds tightly to barnase in a one-to-one complex. Already since the 1980s, a solid expression system for barstar has been available in *Escherichia coli*. Shortly afterwards, the protein structure was elucidated by crystallography and NMR.[110-112] The structure of barstar is characterized by a β-α-β motif, which is typical for nucleic acid binding proteins. Barstar consists of only 90 amino acid residues and comprises a molecular weight of ~10 kDa. It is commonly used as a model protein for folding and stability studies and contains three Trp residues (chapter 3.8). Several groups reported that barstar unfolds reversible upon denaturation by different procedures, e.g. temperature, chemical detergents, or extreme pH, and folding and unfolding kinetics have been determined.[113-117]

In this work, I used the cysteine-free, pseudo wild-type barstar mutant ψ-b* with the following mutations Pro28Ala/Cys41Ala/Cys83Ala[1] (Fig. 11) as model protein for the tandem incorporation of two different non-canonical amino acids.

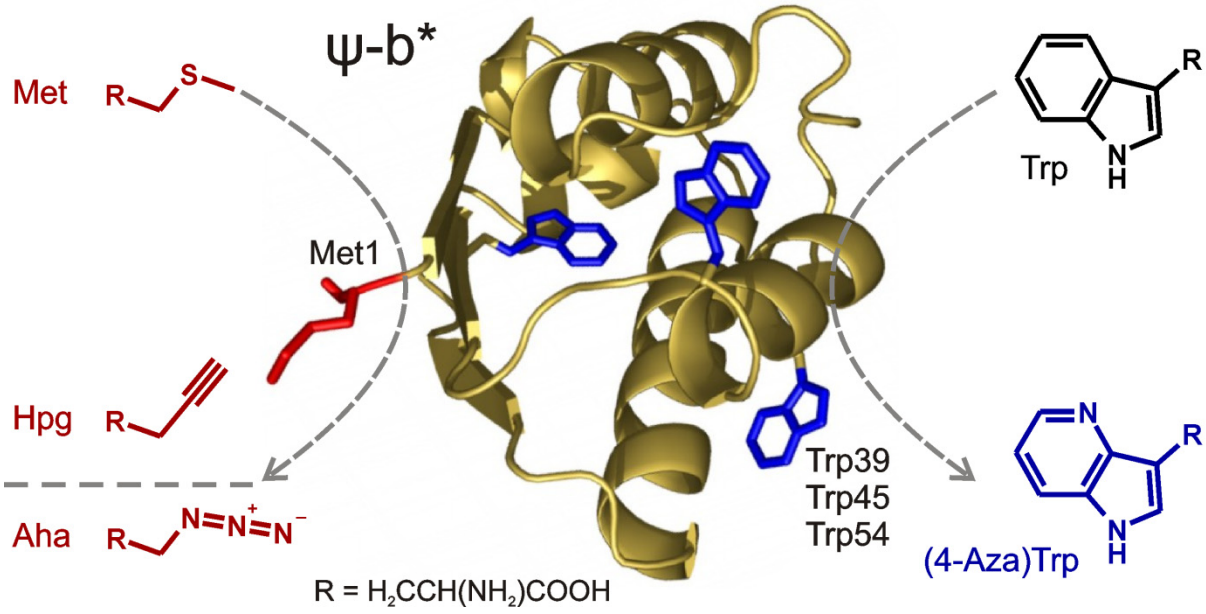


Fig. 11 Tandem labeling of pseudo wild-type barstar (ψ-b*)[1] with 4-azatryptophan [(4-Aza)Trp] and homopropargylglycine (Hpg)/azidohomoalanine (Aha). X-ray crystallography-based model of ψ-b* with the intramolecular distribution of the single *N*-terminal Met1 (in red) and three Trp side chains (in blue; Trp39, Trp45, and Trp54). The concept of tandem labeling (i.e. canonical amino acid → non-canonical analog replacement) is indicated. In a single expression/incorporation experiment, Met and three Trp side chains in ψ-b* were simultaneously substituted by their analogs Hpg/Aha and (4-Aza)Trp (chemical structures are given in red and blue, respectively).

As depicted in Fig. 11, the aim of these experiments was the substitution of the canonical amino acid Met for its analogs Hpg and Aha and the simultaneous substitution of Trp for (4-Aza)Trp. Hpg and Aha provide reactive handles for chemical functionalizations and (4-Aza)Trp serves as fluorescent probe detecting respective cells and proteins.

6.2 Incorporation of (Aza)Trps into anxA5

6.2.1 Cell growth in the presence of (Aza)Inds - blue fluorescent bacteria

Before an (Aza)Trp can be incorporated into target proteins, the precursor (Aza)Ind has to be efficiently imported into the cells. The first observation was that the Trp-auxotrophic *E. coli* strain ATCC 49980 did not grow in synthetic medium with any (Aza)Ind as the sole Ind source. Nevertheless, the same cells nearly reached stationary phase (OD_{600} ~2.0 - 2.5), if they were incubated in the presence of limiting amounts of Ind (1-5 µM) in media supplemented with 1.3 mM (4-Aza)Ind or (5-Aza)Ind, subsequently. This indicates that Ind, or more likely Trp, is essential during early growth phases but can be supplemented with (Aza)Inds in the later growth phases. Thus, both substances are sufficiently biocompatible for SPI experiments, which are routinely performed with cell cultures in mid-exponential growth phase.

In their pioneering studies in the 1950s, Pardee and Prestidge were able to demonstrate that (7-Aza)Trp incorporation into cellular proteins resulted in inactive enzymes as well as in the inhibition of phage infection of bacteria.[118] The most conceivable explanation for the lack of initial growth on Ind analogs is that in the cellular proteome a pool of proteins exists, whose members obligatorily require Ind moieties for their functional and/or structural integrity. Thus, cell growth in synthetic medium is only initiated if minute amounts of the canonical amino acid or its precursor are present.

Upon excitation with an UV lamp ($\lambda_{Em,max}$ = 360 nm, Olympus MT 20 monochromator; Ina-shi, Nagano, Japan), blue fluorescent bacteria were observed when Trp-auxotrophic *E. coli* cells were grown under the above described conditions with (4-Aza)Ind (Fig. 12). In contrast, no fluorescence was detected in the presence of Ind.

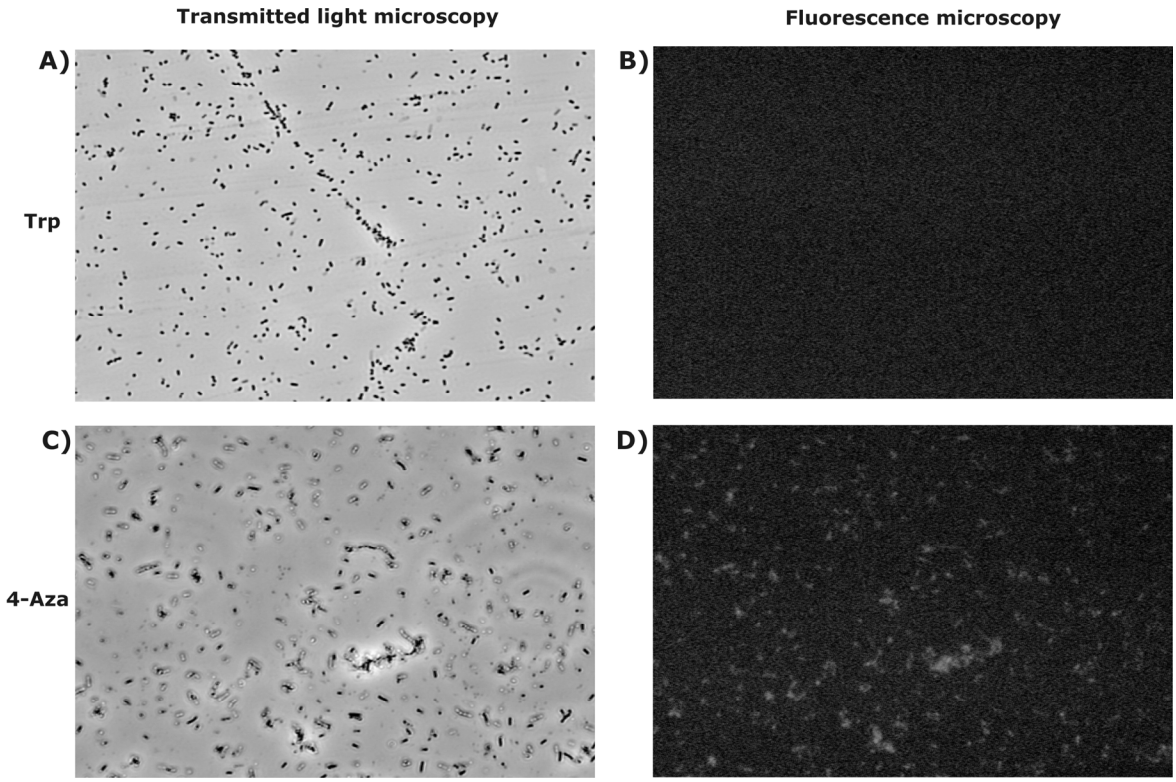


Fig. 12 Transmitted light and fluorescence microscopy images of *E. coli* ATCC 49980 cells expressing Trp-anxA5 (Trp) and (4-Aza)Trp-anxA5 (4-Aza). *E. coli* cells expressing Trp-anxA5 in transmitted light **(A)** and fluorescence **(B)** after UV excitation. *E. coli* ATCC 49980 cells expressing (4-Aza)Trp-anxA5 in transmitted light **(C)** and fluorescence **(D)** after UV excitation (Microscope: Olympus BX 51TRF, Lamp: Olympus MT 20 monochromator ($\lambda_{Em,max}$ = 360 nm); Ina-shi, Nagano, Japan).

These observations indicate (4-Aza)- and (5-Aza)Ind were uptaken by the cells, presumably via the Trp/Ind-uptake pathway.[28]

Taking into account the high level expression of the (Aza)Trp variants of anxA5, the blue fluorescence observed in bacteria most probably originates from the substituted proteins. However, it cannot be ruled out that intracellularly accumulated, free (Aza)Ind adds to the observed fluorescence.

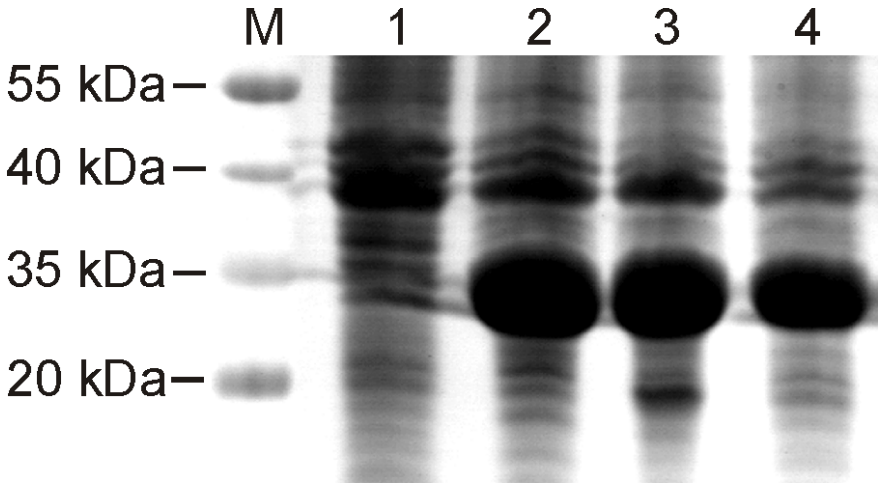


Fig. 13 Expression of anxA5 and aza-variants. Expression profiles of anxA5 in cell lysates of *E. coli* ATCC 49980 grown in NMM with Ind, (4-Aza)Ind, and (5-Aza)Ind. Note that comparably high expression levels of anxA5 were achieved with all Ind isosteres. **M)** PageRuler^{TM} Prestained Protein Ladder; **1)** non induced cell lysate; **2)** Trp-anxA5; **3)** (4-Aza)Trp-anxA5; and **4)** (5-Aza)Trp-anxA5.

6.2.2 Expression and analytics of (Aza)Trp-anxA5

For the co-translational incorporation of (Aza)Trps into any target protein, the uptaken (Aza)Inds must be efficiently converted into the related Trp analogs by TrpS. In fact, the metabolic transformation of various substituted Inds into the related amino acid analogs is described for other bacterial hosts as well.[119,120] Thus, we expected the *in situ* production of (Aza)Trps in the Trp-auxotrophic *E. coli* cells. If the fluorescent Trp analogs accumulate intracellularly up to levels high enough for sufficient activation and charging onto tryptophanyl-tRNA by the TrpRS, they will participate in ribosomal translation by re-coding the UGG (Trp) coding triplets.[28]

For efficient substitution of the single Trp187 in anxA5 with (Aza)Trps, the cell fermentation parameters were optimized. Biomass was accumulated using a defined concentration of Ind (5.5 µM) in NMM, which allowed robust cell growth to the mid-exponential phase (OD_{600} ~ 0.6 - 0.8).[38]

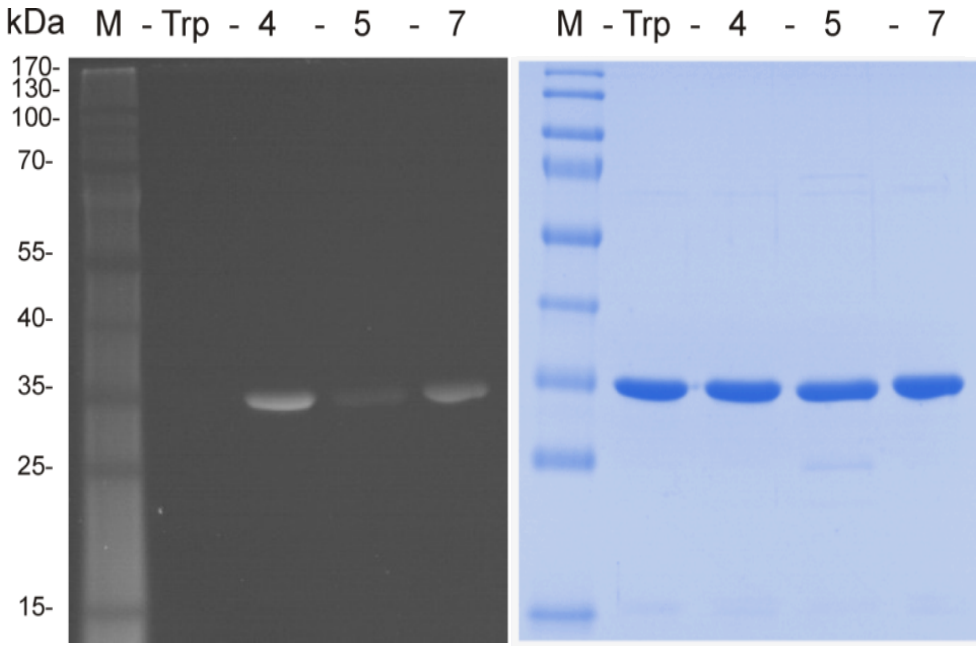


Fig. 14 SDS-PAGE profiles of purified Trp-anxA5 and corresponding aza-variants. Left: Exposed to UV light. **Right:** After Coomassie staining. **M)** PageRuler™ Prestained Protein Ladder; **Trp)** Trp-anxA5; **4)** (4-Aza)Trp-anxA5; **5)** (5-Aza)Trp-anxA5; and **7)** (7-Aza)Trp-anxA5. Note that no UV band in the lane of Trp is visible but all (Aza)Trp-substituted proteins are clearly distinguishable under UV light.

After Ind depletion, the culture was supplemented with the desired (Aza)Ind (1 mM) for 30 min, followed by the induction of the histidine-tagged anxA5 expression (see 5.3.5). The model protein, anxA5, was highly expressed by this procedure, regardless of whether Ind, (4-Aza)-, or (5-Aza)Ind was provided in the medium (Fig. 13). Furthermore, expression profiles on SDS gels of cells incubated with (4-Aza)- or (5-Aza)Ind exhibited prominent fluorescent anxA5 bands at ~35 kDa upon exposure to UV light (λ = 312 nm).

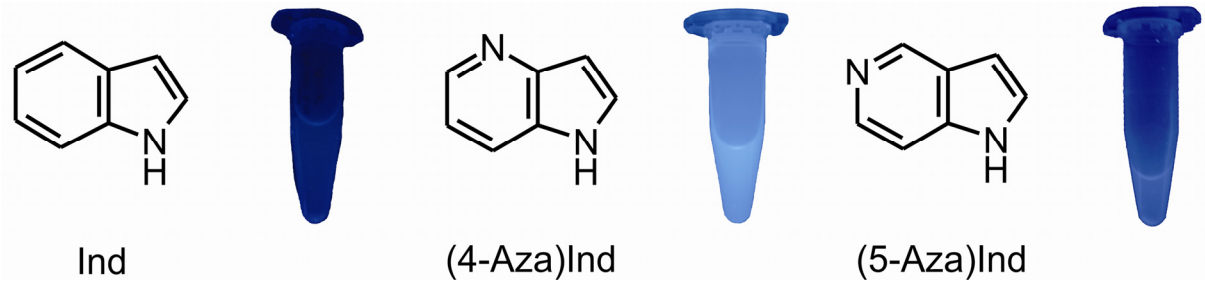


Fig. 15 Chemical structures of Ind, (4-Aza)Ind, and (5-Aza)Ind with the related purified anxA5 variants under UV light.

Indoles	**$\lambda_{Ex,max}$ [nm]**	**ε_M [M^{-1} cm^{-1}]**	**$\lambda_{Em,max}$ [nm]**	**RF[a]**	**QY**
Ind	269	5264	348	1.00	0.220
(4-Aza)Ind	288	7262	418	0.30	0.077
(5-Aza)Ind	266	2663	402	0.16	0.099
Proteins	**$\lambda_{Ex,max}$ [nm]**	**ε_M [M^{-1} cm^{-1}]**	**$\lambda_{Em,max}$ [nm]**	**RF**	
Trp-anxA5	278	21,105	318	1.00	
(4-Aza)Trp-	278	21,020	423	0.46	
(5-Aza)Trp-	278	19,730	414	0.24	

Table 2 Absorbance and emission maxima (λ_{max}) with molar extinction coefficients (ε_M) of free indole/azaindoles in buffered solution, Trp-anxA5, and related aza-variants. Experimental conditions are described in Materials and Methods and related spectra presented in Fig. 16, Fig. 19 and Fig. 24. All data were recorded in 100 mM Tris·HCl, pH 7.5.

[a]RF relative fluorescence was normalized to indole intensity at the emission maximum.

No fluorescence was observed for the parent protein. These results were the first indication for a successful conversion of (4-Aza)- or (5-Aza)Ind into (Aza)Trps and their subsequent highly efficient incorporation into anxA5.

After purification by metal affinity chromatography, the yield of anxA5 purified from cells supplemented with (4-Aza)Ind was higher (~17 mg/L) than that from the Trp-anxA5 culture (~10-15 mg/L), while anxA5 in the presence of (5-Aza)Ind was equally well expressed (~12 mg/L) (Fig. 14). Purified proteins were quantified by Bradford Protein assay (see 5.1.2) and their molar extinction coefficients (ε_M) were determined in the wavelength range between 250 nm - 320 nm as listed in Table 2.

Mass spectrometric analyses (MS and MS/MS Q-TOF mass spectrometry) provided further evidence of successful incorporation of (Aza)Trps into anxA5. Although the molecular mass difference between Trp and both (Aza)Trps represents only 1 Da, the proof for (Aza)Trp incorporation was achieved by a combined fluorescence/mass determination approach. After tryptic digestion of the protein variants, the resulting peptide fragments were separated by reversed-phase HPLC.

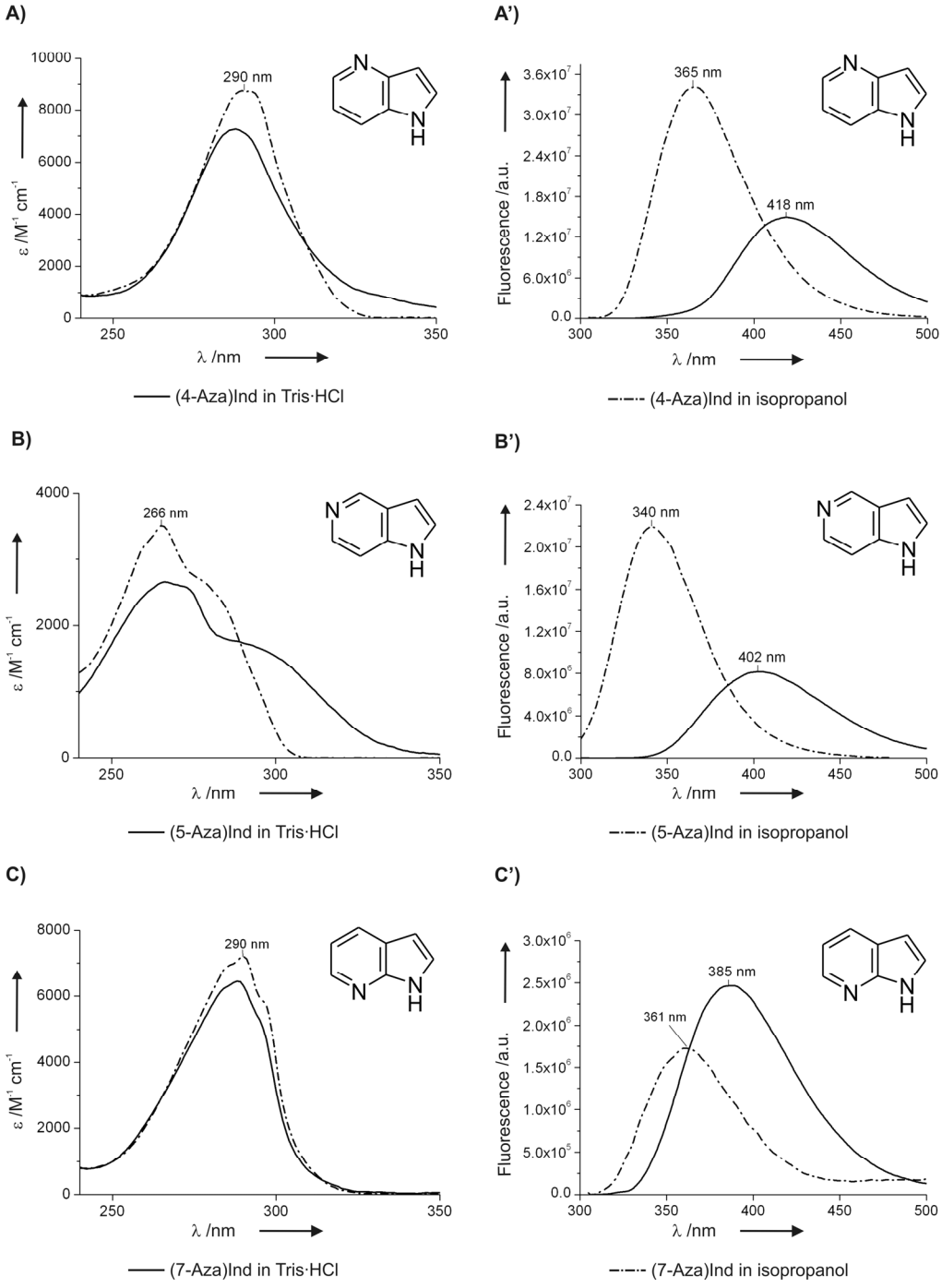


Fig. 16 UV and fluorescence spectra of (Aza)Inds in H_2O and isopropanol. Left: UV profiles; **right:** Fluorescence emission profiles. **A)** and **A')** (4-Aza)Ind; **B)** and **B')** (5-Aza)Ind; **C)** and **C')** (7-Aza)Ind. Fluorescence was excited at the respective absorbance maximum wavelength as specified in the absorbance spectra and in Table 2. Note that the fluorescence intensity of (7-Aza)Ind in both solvents is about an order of magnitude lower than of the other (Aza)Inds.

The single (Aza)Trp-containing peptide [(Aza)WGTDEEK] was identified and collected according to its characteristic fluorescence emission in the range 400 nm - 420 nm (λ_{Ex} = 279 nm). Finally it was analyzed by Q-TOF MS. As expected, the fluorescent peptide revealed a mass of m = 864.36 Da (m/z = 433.19), which agreed with the predicted mass (m = 864.36 Da). In addition, in MS/MS mode the correct amino acid sequence of this peptide was confirmed (Fig. 36, Fig. 37, Fig. 38, and Fig. 39 in Appendix).

For the quantitative estimation of the (Aza)Trp incorporation level, amino acid hydrolysis was performed using thioglycolic acid. The derivatized amino acids were analyzed by HPLC and subsequent mass spectrometry (Appendix chapter 11.2 for data presentation). Strikingly, the (4-Aza)Trp-anxA5 variant revealed a substitution of Trp of over 90%, and confirmed the high biocompatibility of this small fluorescent probe. The signal of (5-Aza)Trp from the (5-Aza)Trp-anxA5 variant could not be separated from the threonine peak. Thus, the incorporation level of (5-Aza)Trp of about 70% was calculated via the reduction of the Trp signal (-2.26%, total 3.2% in the parent protein) and the increase of the threonine peak (+2.26%) in the profile of (5-Aza)Trp-anxA5 in comparison to Trp-anxA5. (7-Aza)Trp-anxA5 was also expressed for comparative spectroscopic studies, and the level of incorporation was about 80%. These data were fully confirmed by state-of-the-art Orbitrap mass spectrometric analysis (unpublished material).

6.2.3 Dramatically red-shifted fluorescence

In order to assess the influence of the nitrogen atom in the Ind ring on the spectral properties of the free Inds or after incorporation into the model protein, fluorescence emission profiles of all Inds and proteins were compared upon excitation at their absorbance maxima, as listed in Table 2 and displayed in Fig. 17 and Fig. 18.

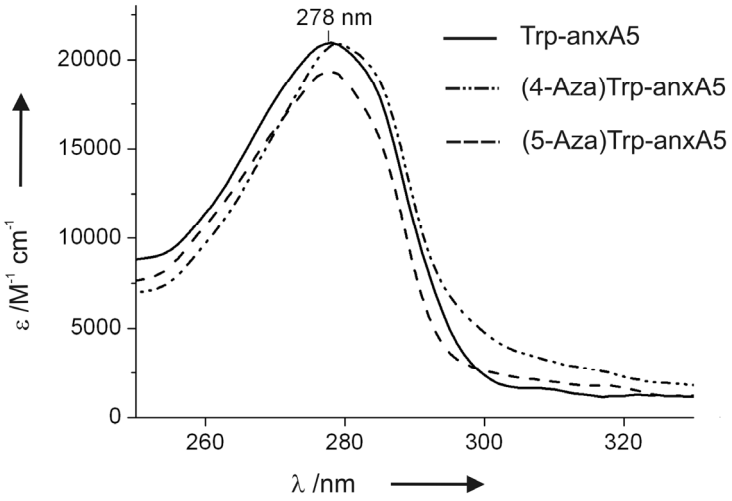


Fig. 17 UV-profiles with extinction coefficients of Trp-anxA5 and variants. Continuous line: Trp-anxA5, **dashed-dotted line:** (4-Aza)Trp-anxA5, and **dashed line:** (5-Aza)Trp-anxA5. Note that overall spectral profile and absorbance maxima (λ_{max} = 278 nm) of Trp-anxA5 and its aza-variants are not significantly affected by aza-substitutions due to the Tyr dominance in anxA5 UV profile (12 Tyr and only one Trp).

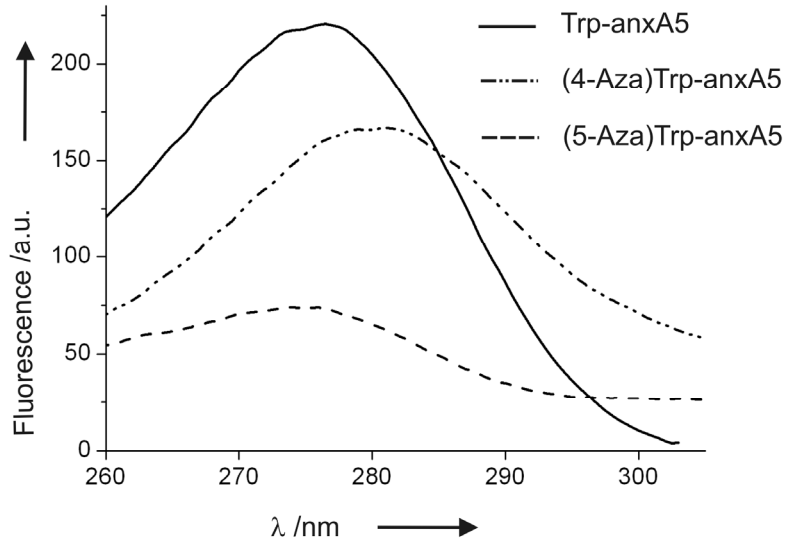


Fig. 18 Excitation spectra of Trp-anxA5 and variants. Continuous line: Trp-anxA5, **dashed-dotted line:** (4-Aza)Trp-anxA5, and **dashed line:** (5-Aza)Trp-anxA5. Spectra were recorded with 2 µM protein samples in 100 mM Tris·HCl, pH 7.5, at the respective fluorescence maximum of the (Aza)Inds (see Table 2) with excitation/emission slits of 10/10.

When compared to Ind, the fluorescence intensity maximum $\lambda_{Em,max}$ of free (4-Aza)Ind in aqueous solution was red-shifted by 70 nm, and that of (5-Aza)Ind was red-shifted by 54 nm (Fig. 19A and C, Table 2). The fluorescence characteristics of dissolved (4-Aza)- and (5-Aza)Inds were fully transmitted into the related anxA5 variants, while the red-shifts of the (Aza)Trp variants were even more pronounced.

In the case of (4-Aza)Trp-anxA5, the red-shift was about 105 nm, and the (5-Aza)Trp-anxA5 variant exhibited a 96 nm red-shifted emission maximum relative to the parent protein Trp-anxA5 (Fig. 19B and D, Table 2). In Trp-anxA5, the fluorescence emission maximum of the Trp187 Ind moiety was blue-shifted relative to free Ind in aqueous solution (Fig. 19A and B, Table 2), a consequence of the placement of Trp187 within the hydrophobic cavity (Fig. 10) of the protein, combined with the spectral contribution of the remaining 12 Tyr residues.[38]

The transmission of (Aza)Ind´s spectral properties into substituted anxA5 was not only successful but even more pronounced than among free Inds in solution. It is also reasonable that in the variant proteins, the (Aza)Ind moieties of (4-Aza)- and (5-Aza)Trp are not completely buried in the hydrophobic cavity but face the solvent environment. This conjecture was supported by the fluorescence properties of the (Aza)Trp-anxA5 variants, which (upon denaturation or Ca^{2+}-titration) resembled those of Trp-anxA5 (Fig. 20 and Fig. 21).

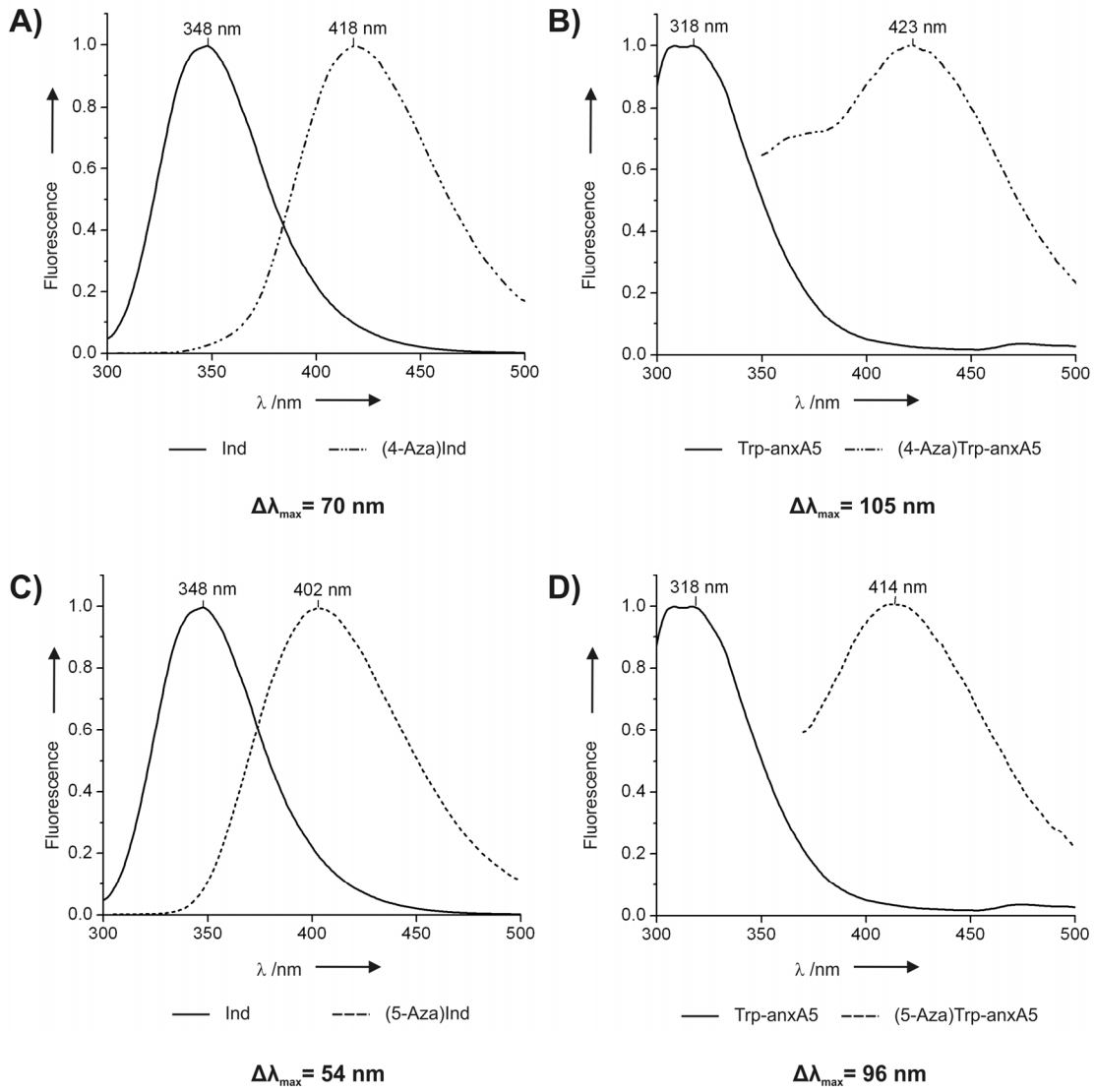


Fig. 19 Normalized fluorescence emission spectra of free indoles in buffered aqueous solution and of the (Aza)Trp protein variants upon excitation at 280 nm. Comparison of the fluorescence spectra of free indole (Ind) with free (4-Aza)Ind **(A)** and with (5-Aza)Ind **(C)**. Note the dramatic red-shift in the (4-Aza)Ind emission spectrum (70 nm). The emission maximum of free (5-Aza)Ind is less red-shifted (54 nm). The fluorescence emission maximum of (4-Aza)Trp-anxA5 **(B)** is 107 nm further red-shifted whereas that of (5-Aza)Trp-anxA5 **(D)** is 96 nm red-shifted in comparison to Trp-anxA5. Note also the spectral shoulder at 365 nm of (4-Aza)Trp-anxA5 (see text for more details).

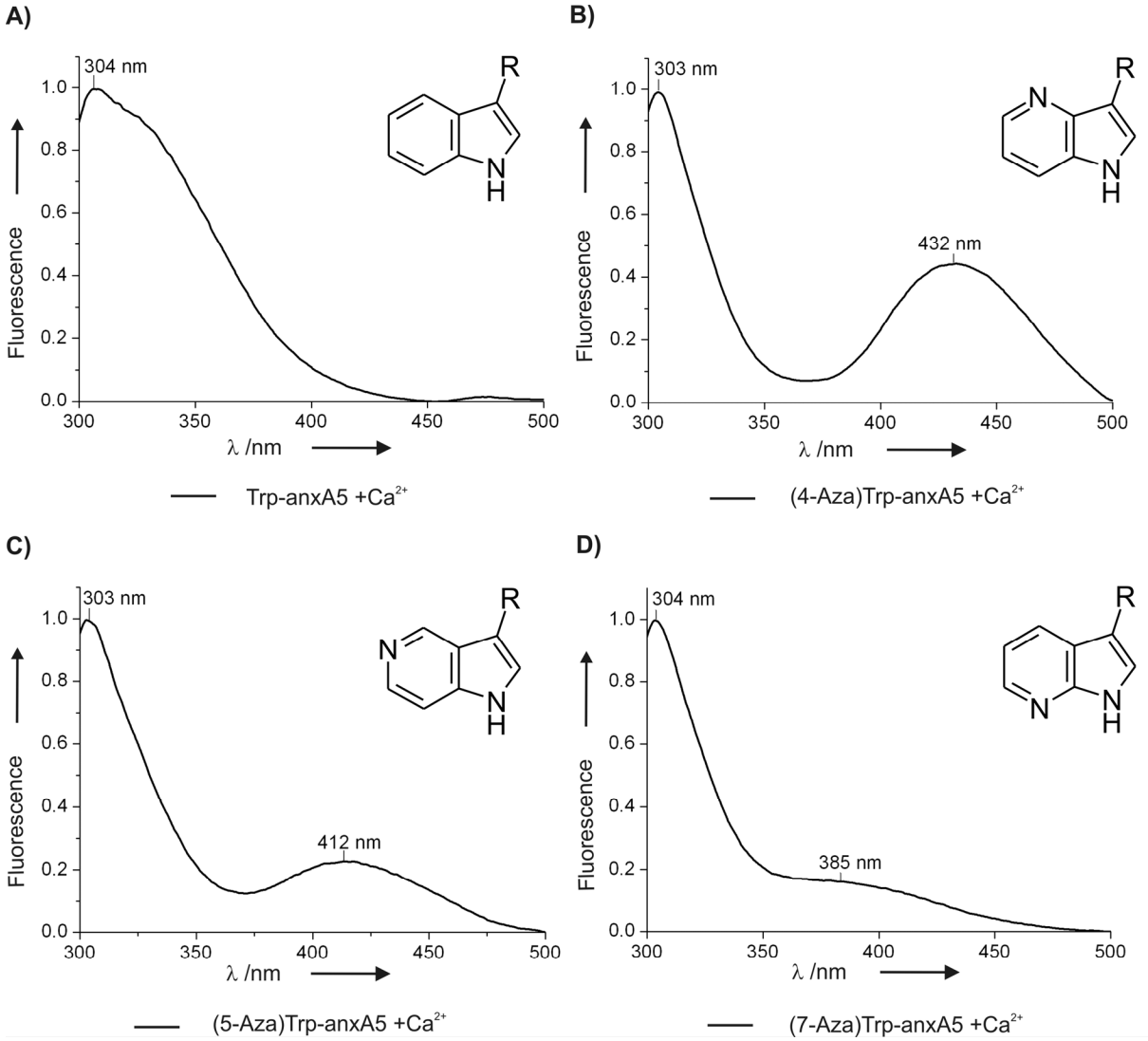


Fig. 20 Normalized fluorescence spectra of anxA5 and variants in the presence of 10 mM Ca^{2+}. A) Trp-anxA5; **B)** (4-Aza)Trp-anxA5; **C)** (5-Aza)Trp-anxA5; **D)** (7-Aza)Trp-anxA5. Spectra were recorded with 2 µM protein samples in 100 mM Tris·HCl, pH 7.5, at 278 nm with excitation/emission slits of 10/10. Data were normalized to the protein's Tyr fluorescence peak. R denotes anxA5.

Both processes are accompanied by the exposure of Trp187 to the surrounding solvent.[109] The spectral shoulder at $\lambda_{Em,max}$ = 365 nm in the fluorescence emission profile of (4-Aza)Trp-anxA5 (Fig. 19B) correlated well with the emission maximum of (4-Aza)Ind in apolar solvents. This might indicate that at least a portion of (4-Aza)Trp187 was buried in the hydrophobic cavity of the protein molecule. Upon urea denaturation or Ca^{2+}-titration of (4-Aza)Trp-anxA5, this spectral shoulder disappeared (Fig. 20 and Fig. 21).

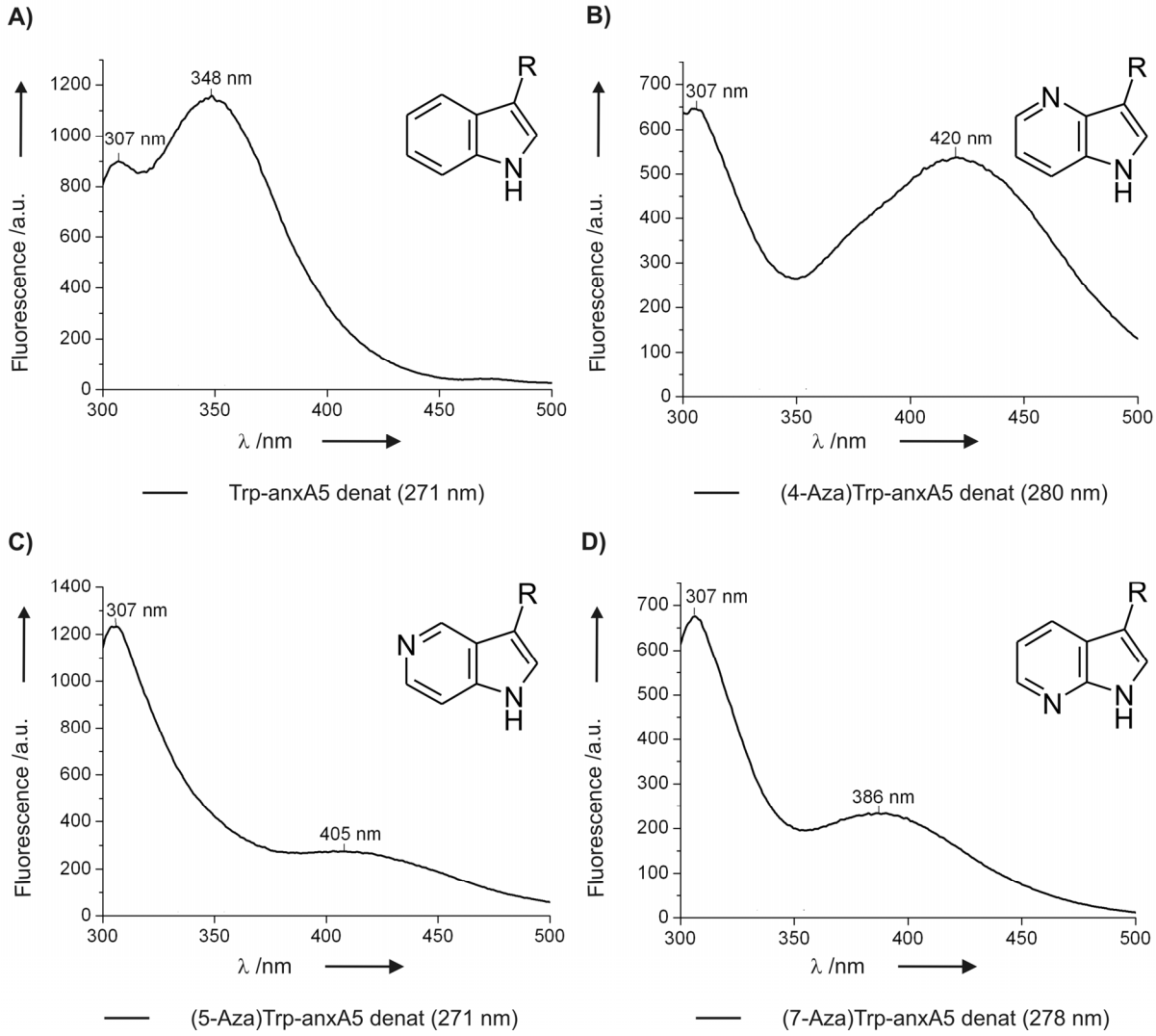


Fig. 21 Fluorescence emission profiles of chemically denatured anxA5 and variants. A) Trp-anxA5; **B)** (4-Aza)Trp-anxA5; **C)** (5-Aza)Trp-anxA5; **D)** (7-Aza)Trp-anxA5. Spectra were recorded with 8 µM protein samples in 8 M urea at the respective absorbance maximum of the (Aza)Inds (Table 2) with excitation/emission slits of 10/10. R denotes anxA5.

These observations and considerations strongly indicated the coexistence or an equilibrium between two protein conformations with either a buried or a solvent-exposed (4-Aza)Trp side chain.

6.2.4 Secondary structure and folding cooperativity

Far-UV CD spectroscopy of Trp-anxA5 and its aza-variants revealed the typical CD profiles of α-helical proteins, with two characteristic minima of similar intensity at 222 nm and 208 nm (Fig. 22). The incorporation of (Aza)Trps had only minor effects on the CD spectra indicating that the structure of anxA5 remained unchanged. The small differences in the intensity of the CD bands around 220 nm were most likely due to the contributions of the Trp residues, which were changed by replacing an Ind moiety by an (Aza)Ind moiety.

From the CD profiles, the α-helical structural content was calculated of about 60% in the secondary structure of anxA5 and its variants. Furthermore, thermally induced denaturation profiles of all three proteins had similar T_m values (± 2 °C) and enthalpy of unfolding, indicating cooperative folding (Fig. 23 and Table 3). In summary, both (4-Aza)- and (5-Aza)Trp, exhibited no significant influence on the structural integrity of anxA5.

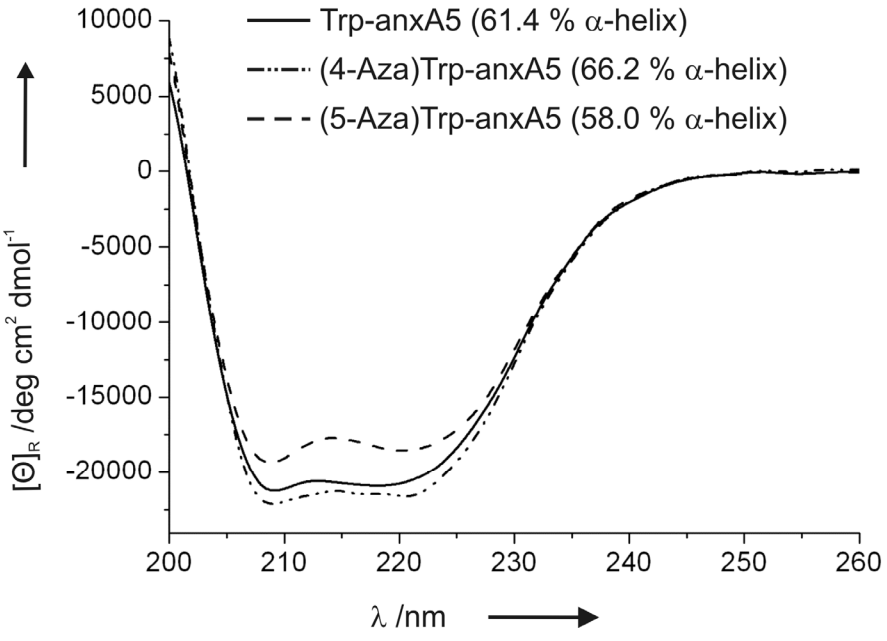


Fig. 22 Secondary structure of (Aza)Trp-anxA5 variants and the parent Trp-protein. Far UV-CD spectra from 200 - 260 nm were recorded at 4 °C.

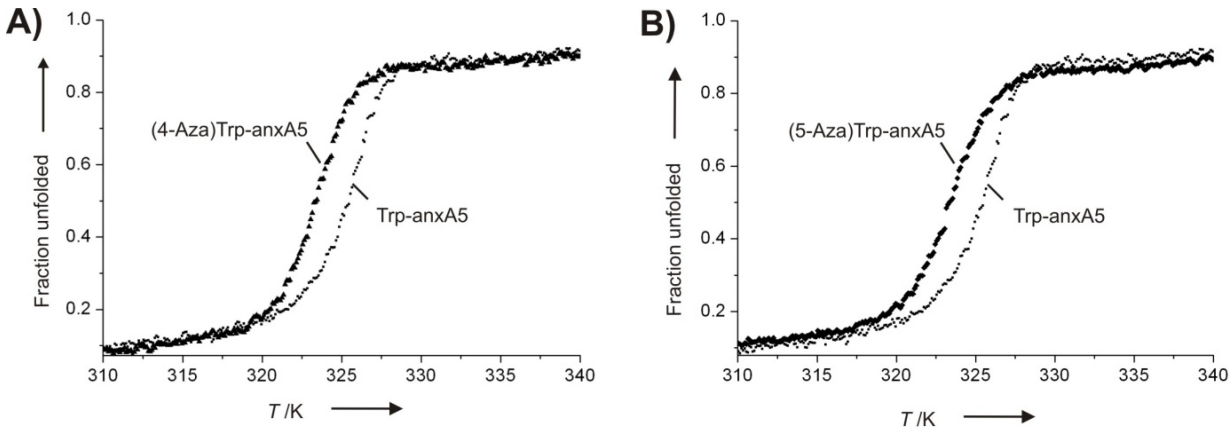


Fig. 23 Thermally induced unfolding profiles of (4-Aza)Trp-anxA5 (A) and (5-Aza)Trp-anxA5 (B) in comparison to Trp-anxA5. The fraction of unfolded protein was calculated from CD data monitored at 222 nm as described in Materials and Methods.

Protein	T_m [K]	ΔT_m [K]	ΔH_m [kJ mol^{-1}]	$\Delta\Delta H_m$ [kJ mol^{-1}]
Trp-anxA5	325.35 ± 0.07		-650.34 ± 30.79	
(4-Aza)Trp-anxA5	323.25 ± 0.04	- 2.10	-698.19 ± 20.34	+ 47.85
(5-Aza)Trp-anxA5	323.39 ± 0.01	- 1.96	-640.26 ± 06.04	- 10.08

Table 3 Thermodynamical parameters of Trp-anxA5 and its aza-variants. Experimental conditions and calculation methods for all unfolding profiles are described in Materials and Methods while related curves are presented in Fig. 23. ΔT_m and $\Delta\Delta H_m$ are the differences between T_m and ΔH_m of the parent and substituted proteins.

6.2.5 (4-Aza)Ind versus traditionally used (Aza)Inds

Recently, Twine and Szabo reported 6-azaindole [(6-Aza)Ind] as novel Ind-based chromophore with intriguing spectral properties.[86] While (6-Aza)Ind comprised a strong emission signal and large red-shift in hydrophilic environments, its fluorescence was markedly quenched in aprotic solvents (and most probably in the protein core as well). Unfortunately, the attempts generating fully substituted (6-Aza)Trp-anxA5 using the same conditions as for (4-Aza)- and (5-Aza)Ind were unsuccessful; most probably because the biocompatibility of (6-Aza)Ind is not as good as in the case of the other tested (Aza)Inds like (4-Aza)- and (5-Aza)Ind. When compared with (5-Aza)Ind, (4-Aza)Ind proved to be a better optical probe, exhibiting higher biocompatibility

and a larger red-shifted fluorescence maximum. Additionally, it exhibits higher fluorescence intensity, molar extinction coefficient, and a comparable QY (Table 2). In addition, (4-Aza)Ind in anxA5 induced similar structuring of the protein interior when compared to the parent protein and its (5-Aza)Ind-containing counterpart (Fig. 22).

As mentioned earlier, the main drawback of the traditionally used (7-Aza)Ind for protein design is that its exposure to aqueous solvents leads to dramatic quenching.[87] The fluorescence measurements of free (7-Aza)Ind in aqueous physiological buffers fully confirmed these observations (Fig. 24). Furthermore, a ~5 fold difference in QY was measured between (7-Aza)Ind and newly examined (5-Aza)- and (4-Aza)Ind (Table 2).

(7-Aza)Ind possesses a markedly red-shifted fluorescence around 360 nm when shielded by the protein matrix. However, its exposure to aqueous solvents leads to dramatic quenching (Fig. 24). For that reason it is often described in the literature as "water-quenched fluorescent probe".[88,121-123,82,87,64] The fluorescence intensity of (4-Aza)Ind in a physiological milieu is 5-fold higher and exhibits the reddest fluorescence emission among all (Aza)Inds (λ_{max} = 418 nm) (Fig. 16 and Fig. 24). In a non-aqueous medium like isopropanol, (4-Aza)Ind exhibits also a red-shifted fluorescence emission maximum, in comparison to Ind, close to the one of (7-Aza)Ind, but with markedly higher fluorescence intensity (about one order of magnitude; Fig. 16).

There are different assumptions of the issue of (7-Aza)Ind´s emission quenching in aqueous solutions. Water as solvent can interact with free (7-Aza)Ind molecules yielding tautomeric and "blocked" species. Most probably, various interactions of these species cause enhanced rates of radiationless processes upon excitation. Conversely, such radiationless processes are less pronounced in (4-Aza)Ind. One can speculate that the molecular structures of (4-Aza)Ind and (5-Aza)Ind favor different and/or "less blocking" interactions with the solvent molecules whereby radiationless processes are more suppressed.

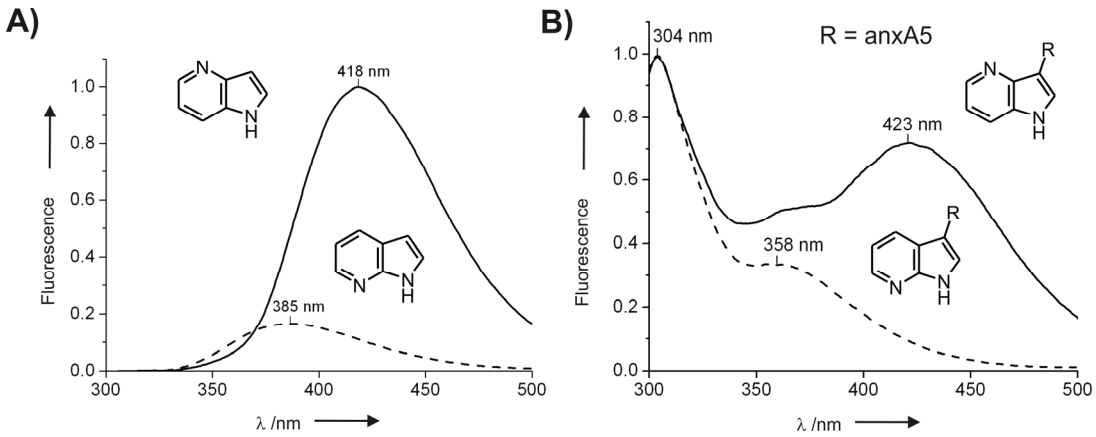


Fig. 24 Fluorescence profiles of (4-Aza)Ind and (7-Aza)Ind and related protein variants in aqueous buffered solution. A) Fluorescence spectra of (4-Aza)Ind (—) and (7-Aza)Ind (- - -) upon excitation at the respective absorbance maximum as listed in Table 2. Note the large red-shift (33 nm) in the profile of (4-Aza)Ind ($\lambda_{max,em}$ = 418 nm) when compared with (7-Aza)Ind ($\lambda_{max,em}$ = 385 nm) along with markedly higher fluorescence intensity of (4-Aza)Ind. **B)** Emission profiles (normalized to protein fluorescence) of (4-Aza)Trp-anxA5 (—) and (7-Aza)Trp-anxA5 (- - -) upon excitation at 280 nm. Note that the emission maximum $\lambda_{max,em}$ of (4-Aza)Trp-anxA5 ($\lambda_{max,em}$ = 423 nm) is 65 nm more red-shifted than of (7-Aza)Ind ($\lambda_{max,em}$ = 358 nm). The spectral shoulder of (4-Aza)Trp-anxA5 in the range between 350 nm and 370 nm is discussed in the text.

However, in the absence of spectral information from model substances any further attempt to explain the observed effects would be too speculative.

To compare (7-Aza)Ind versus (4-Aza)Ind in protein structure, related anxA5 variants were also expressed and purified. As expected, (7-Aza)Trp-anxA5 exhibited less fluorescence intensity and a diminished red-shift of the fluorescence maximum (Fig. 24B). Moreover, upon denaturation of (7-Aza)Trp-anxA5, the fluorescence intensity of the unfolded protein was almost fully quenched, reflecting full solvent accessibility of (7-Aza)Trp187. In contrast, the fluorescence intensity of the denatured or calcium-bound (4-Aza)Trp-, and also (5-Aza)Trp-anxA5 variants remained largely unchanged (Fig. 20 and Fig. 21). Finally, it has often been claimed that the main advantage of (7-Aza)Ind in protein structures is its selective excitability.[64] However, this feature also exists in the absorbance profile of (4-Aza)Ind which has a broad absorbance band around 310 nm (Fig. 19A).

6.3 Tandem incorporation of (4-Aza)Trp and Hpg/Aha into ψ-b*

*6.3.1 Expression of Hpg/Aha-(4-Aza)Trp-ψ-b**

The incorporation experiments with (4-Aza)Trp and Hpg/Aha into ψ-b* were performed as described in chapter 5.3.6. Fig. 25 clearly demonstrates the expression and effective incorporation of (4-Aza)Trp into ψ-b*, which is visible by exposing the unstained gel with the cell lysates to UV light. As expected, the expression profiles of Trp-ψ-b* are hardly visible under UV light. This simple and straightforward procedure is the first indication of high quality protein labeling with (4-Aza)Trp. After Coomassie staining of the same gel, protein expression bands of ψ-b* with (4-Aza)Trp perfectly correspond with the fluorescent signals.

Ψ-b* and its variants are expressed in inclusion bodies and were routinely refolded prior to purification as described in chapter 5.3.6. Proteins were purified in yields up to 7.5 mg/mL and kept at +4 °C in 50 mM sodium phosphate buffer, pH 7.4.

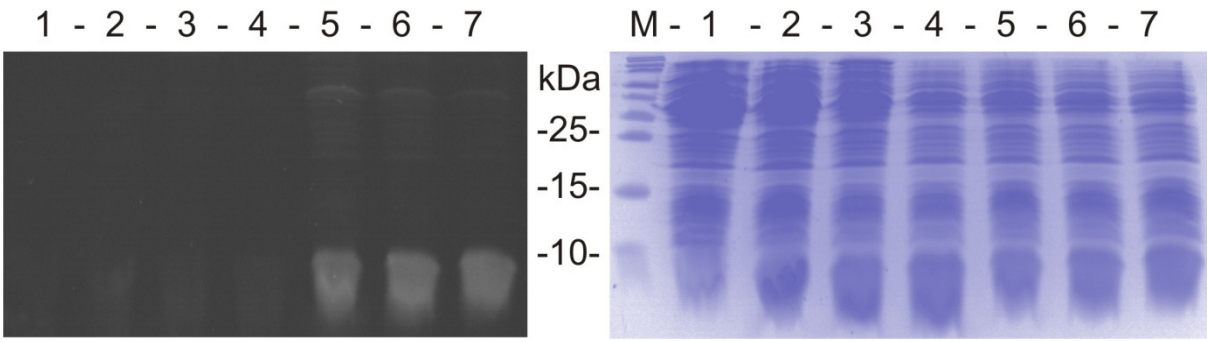


Fig. 25 Tandem labeling with (4-Aza)Trp and Hpg/Aha of ψ-b*. Ψ-b* (~10 kDa) expression profiles in *E. coli* strain JE 7345 cell lysates (0.3 OD_{600}) are visualized upon UV light exposure **(left)** and after Coomassie staining **(right)**. Molecular weight protein standard BenchMark™ Protein Ladder **(M)**, non-induced expression cell lysates **(1)**, expression in medium supplemented with Met and Trp **(2)**, Trp and Aha **(3)**, Trp and Hpg **(4)**, (4-Aza)Trp and Met **(5)**, (4-Aza)Trp and Aha **(6)**, and (4-Aza)Trp and Hpg **(7)**. In contrast to Trp-containing ψ-b*, blue fluorescence of (4-Aza)Trp-ψ-b* is easily detectable upon UV light exposure. In the background, weak fluorescence signals of other intracellular proteins, due to at least partial labeling during expression, are visible.

While the successful incorporation of (4-Aza)Trp into ψ-b* could already be estimated via simple exposure to UV light, Hpg´s incorporation was confirmed by mass analysis (chapter 6.3.2) and indicated by *N*-terminal sequencing (Appendix, chapter 11.3). Since ψ-b* contains only one Met residue the amount of incorporated Hpg can be estimated from the resulting peaks. In the case of Met-Trp-ψ-b*, one typical signal after 14 min appeared, which is the standard retention time of Met. *N*-terminal sequencing of Hpg-(4-Aza)Trp-ψ-b* revealed a significant decrease of the signal at ~14 min, while a new, very pronounced peak emerged at ~13 min, which is supposed to represent Hpg incorporated at the *N*-terminal position of ψ-b*. From the ratio of these two peaks, a substitution for Hpg of about 80% can be estimated. The results of the *N*-terminal sequencing of Aha also suggest an incorporation into ψ-b* since as well a new peak emerged after 13 min (see Appendix, chapter 11.3). However, the ratio of the two peaks (Aha and Met) presents only an incorporation of Aha of ~1%, which was unfortunately a too low amount of labeled protein for a successful detection by ESI-MS analysis.

*6.3.2 Mass spectrometry of double labeled ψ-b**

Mass spectrometry of the parent protein Met-Trp-ψ-b* revealed the molecular weights of 10252.3 Da and 10251.3 Da (calculated mass 10252.6 Da, Fig. 26). The calculated mass of fully labeled ψ-b* is 10233.6 Da, which includes a mass decrease by 22 Da by the substitution of one Met for one Hpg and a mass increase by 1 Da for each incorporated (4-Aza)Trp instead of one of the three Trp side chains.

For Hpg-(4-Aza)Trp-ψ-b*, a protein mass of 10232.6 Da was identified, correlating with the molecular weight of ψ-b* with one incorporated Hpg and two incorporated (4-Aza)Trp. The smaller peak ≥10253.6 Da (Fig. 26) indicates the presence of the parent protein and single labeled (4-Aza)Trp-ψ-b*. It is noteworthy that even an incomplete substitution of Trp residues for (4-Aza)Trps led to clearly distinguishable fluorescence signals on the SDS gel (Fig. 25). This result again demonstrates the extraordinary spectral features of (4-Aza)Trp to turn the otherwise colorless ψ-b* into a blue fluorescent protein.

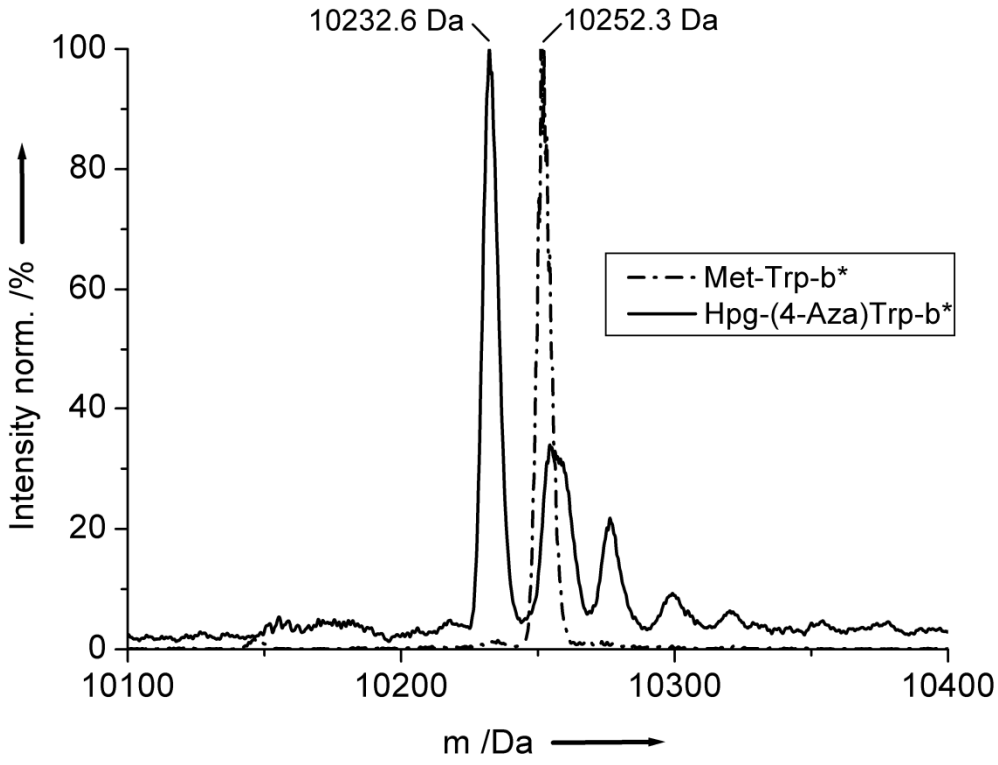


Fig. 26 Mass spectrometry of Met-Trp- and Hpg-(4-Aza)Trp-ψ-b*. Mass spectrometric analysis of Met-Trp-ψ-b* (-·-·-) and ψ-b* with incorporated Hpg and (4-Aza)Trp (—).

Although ESI-MS analysis of Aha-(4-Aza)Trp-ψ-b* confirmed successful incorporation of (4-Aza)Trp, it was not possible to determine the Aha presence in the protein variant (data not shown). One the other hand, minute amounts of Aha (presumably) were found by *N*-terminal sequencing (Appendix chapter 11.3).

In general, the double labeling of one single protein and especially the functionality of the inserted reactive handle of Hpg has to be confirmed. Thus, further experiments, particularly functionalizations by copper(I)-catalyzed Huisgen cycloaddition reactions (CCHC)[92,93], were performed with the Hpg-(4-Aza)Trp-ψ-b* variant. In addition, currently running amino acid hydrolyses will not only prove successful incorporation but will give detailed information on the incorporation yield of both non-canonical amino acids.

6.3.3 Secondary structure and folding cooperativity

Far-UV CD spectroscopy of parent ψ-b* and its double labeled variant revealed the CD profiles of proteins containing both α-helical and β-sheet structured regions (Fig. 27).[28] The CD spectrum of ψ-b* exhibited two characteristic minima at 208 nm and 222 nm.

The parallel incorporation of Hpg and (4-Aza)Trp into ψ-b* led to a shift of the typical minimum at ~208 nm to nearly 200 nm and an overall decrease of the negative CD signal, which suggest a loss in ordered structure.[124] Next to this, a dramatic reduction of the α-helical content of around 20% was measured, whereas the β-sheet content increases by nearly the same percentage (Table 4). The same increase of the β-sheet content was already measured for 4-aminotryptophan-ψ-b*, however a total reduction of structured regions in the ψ-b* variant is much more likely.[28]

In contrast, no significant influence on the secondary structure of anxA5 was detected for the incorporation of (4-Aza)Trp into anxA5 (chapter 6.2.4). In this protein, the single Trp187 is situated inside a hydrophobic cavity at the surface of the protein and is not crucial for the structural integrity of the protein.

Secondary structure element	Met-Trp- ψ-b* [%]	Hpg-(4-Aza)Trp- ψ-b* [%]	Δ [%]
α-Helix	35.9	16.8	- 19.1
Antiparallel β-sheet	4.0	25.7	+ 21.7
Parallel β-sheet	8.6	4.7	- 3.9
β-Turn	18.3	25.8	+ 7.5
Random coil	31.0	32.0	+ 1.0
Σ	97.8	105.0	+ 7.2

Table 4 Secondary structure elements of Met-Trp-ψ-b* and Hpg-(4-Aza)Trp-ψ-b*. Experimental conditions and the calculation method are described in Materials and Methods, while related curves are presented in Fig. 27.

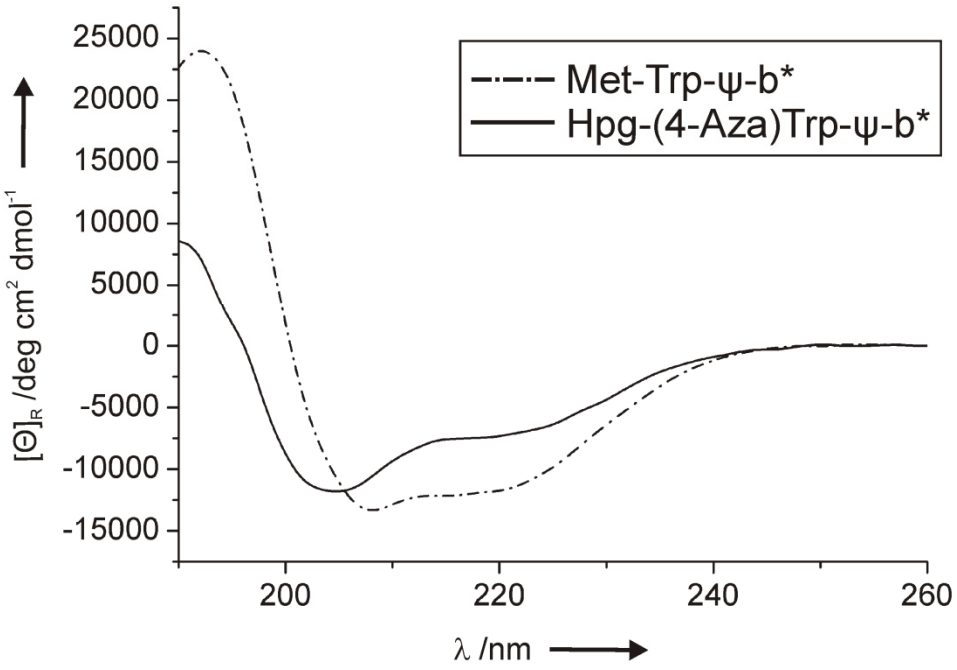


Fig. 27 CD spectra of Met-Trp-ψ-b* (-·-) and Hpg-(4-Aza)Trp-ψ-b* (—). Far-UV CD profiles from 200 nm - 260 nm were recorded at 293.15 K.

Since the CD profile of the double labeled protein differs dramatically from the curve of the parent protein, the thermally induced unfolding curve was measured as described in 5.1.5. In comparison to the parent protein, the thermally induced denaturation profile of Hpg-(4-Aza)Trp-ψ-b* revealed a T_m value decreased by about 17.61 K and a highly decreased enthalpy of unfolding (-103.24 kJ/mol), seen also as a markedly reduced steepness of the measured unfolding (Fig. 28 and Table 5).

Expectedly, the introduction of two different non-canonical amino acids, and especially the substitution of three hydrophobic Trp residues with the more hydrophilic aza-analogs, has a high thermodynamic price, which should not be underestimated. In particular, it is noteworthy that one feature characterizes the melting curve of Hpg-(4-Aza)Trp-ψ-b*: A maximum in stability around 300 K. A similar phenomenon was already observed for 4- and 5-aminotryptophan substituted ψ-b* (chapter 3.8).[28] This reduction in protein stability induced by decreasing temperatures is known as cold denaturation.[57,58]

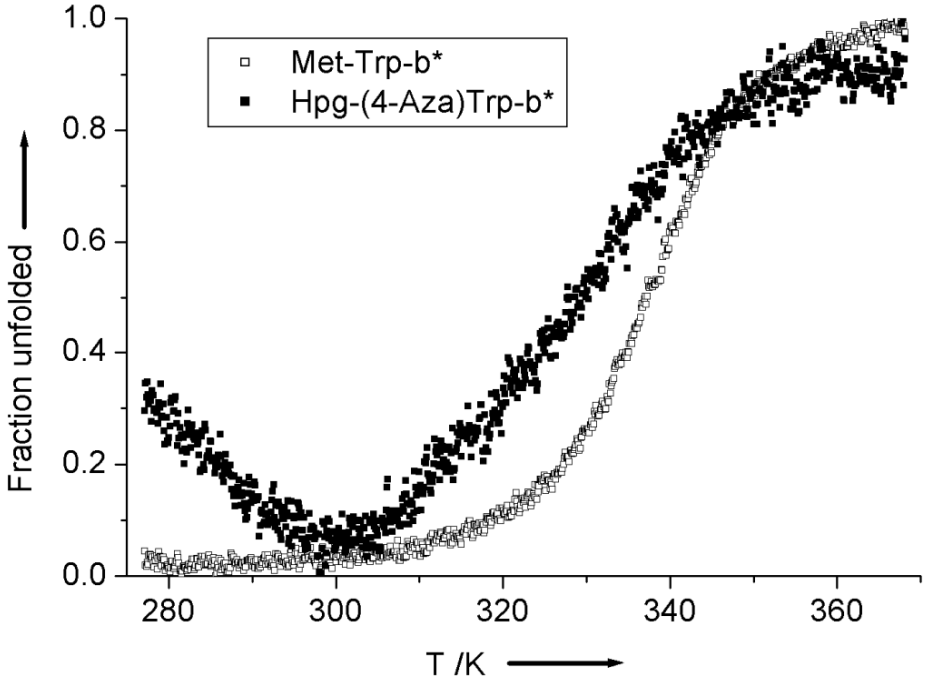


Fig. 28 Thermally induced unfolding profiles of Met-Trp-ψ-b* (▫▫▫) and Hpg-(4-Aza)Trp-ψ-b* (▪▪▪). The fraction of unfolded protein was calculated from CD data monitored at 222 nm as described in Materials and Methods.

It is difficult to speculate about the reasons for the cold denaturation behavior of Hpg-(4-Aza)Trp-ψ-b*. However, it is reasonable that the same principle underlies as in the case of aminotryptophan-substituted ψ-b* variants. Also (Aza)Trp variants of ψ-b* might represent novel tools for the study of the principles behind this behavior.

Protein	T_m [K]	ΔT_m [K]	ΔH_m [kJ mol^{-1}]	$\Delta\Delta H_m$ [kJ mol^{-1}]
Met-Trp-ψ-b*	337.87 ± 0.03		- 153.99 ± 1.07	
Hpg-(4-Aza)Trp-ψ-b*	320.26 ± 0.10	- 17.61	- 50.75 ± 1.3	+ 103.24

Table 5 Thermodynamic parameters of Met-Trp-ψ-b* and Hpg-(4-Aza)Trp-ψ-b*. Experimental conditions and calculation methods for both unfolding profiles are described in Materials and Methods, while related curves are presented in Fig. 28. ΔT_m and $\Delta\Delta H_m$ are the differences between T_m and ΔH_m of the parent and the substituted protein.

The protein structure and the melting curve of anxA5 were not significantly influenced by the incorporation of (4-Aza)Trp (see discussion above). However, the data presented here indicate that the introduction of (4-Aza)Trp into the ψ-b* structure is the major reason for the dramatic loss in the structural integrity of the double labeled protein variant. Single labeled Hpg-Trp-ψ-b* showed a similar melting behavior than the parent protein (S. Dong, personal communication). Thus, it is unlikely that the non-canonical amino acid, which is located at the *N*-terminus of the protein and solvent exposed, would have a considerable impact on the folding of the globular protein. On the other hand, all three Trp residues, and especially Trp53, are of utmost importance for the structural integrity of the small ψ-b* protein, which was already outlined in chapter 3.8. Nonetheless, it was possible to exchange this Trp to 4- and 5-aminotryptophan.[28] As in the case of Hpg-(4-Aza)-ψ-b*, the incorporation of 4-aminotryptophan into ψ-b* led to a diminished tight packing of the protein (Fig. 3). Taken together, the CD analyses revealed decreased protein integrity and reduced folding capacity of the tandemly labeled variant, which was comparable to the amino variant. Finally both ψ-b* variants exhibited similar thermally induced unfolding behavior with cold denaturation-like phenomenon as the most remarkable characteristic.

Obviously, both the exocyclic and the endocyclic nitrogen-containing Trp analogs cause a dramatic perturbation inside the hydrophobic protein interior. In this context, aza- or amino-incorporations represent the introduction of a hydrophilic group into the stabilizing hydrophobic core of the protein (chapter 3.8 for detailed discussion). In the case of the incorporation of 4-aminotryptophan at the position 53 into ψ-b*, this introduction of hydrophilicity led to a significantly reduced number of hydrophobic interactions.

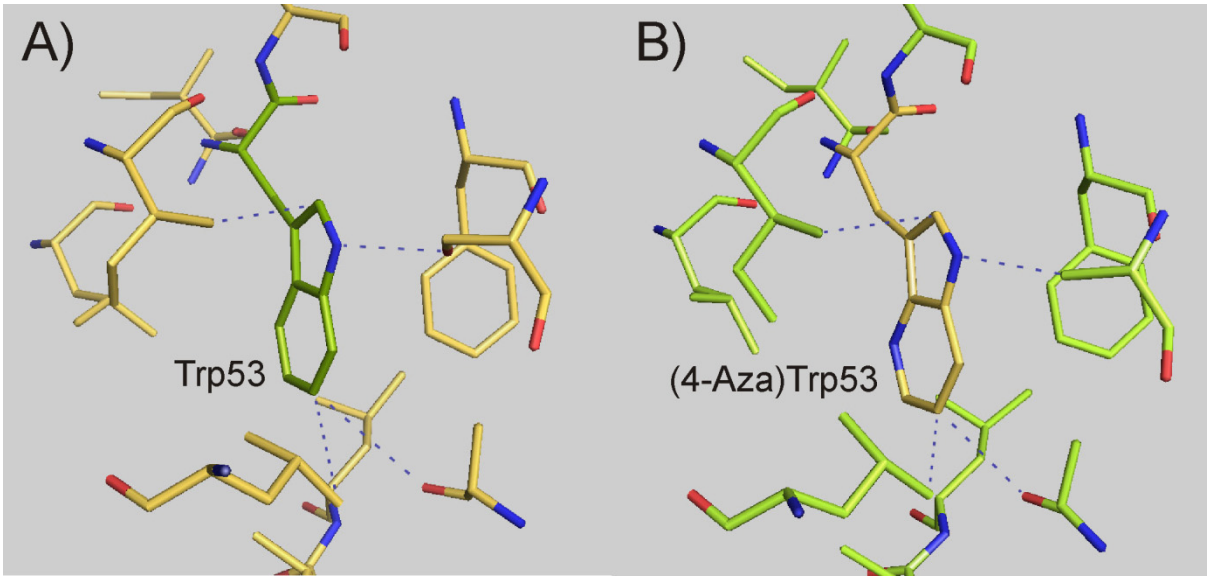


Fig. 29 Local environment of Trp53 (A) and (4-Aza)Trp53 (B) of ψ-b* and its respective variant. Atom selection and energy minimization was performed with MAIN (Turk, 1992) by Dr. M. Debela.[103] Hydrophobic interactions and hydrogen bonds are indicated by dashed lines.

As presented in chapter 3.8 and Fig. 3, the effect of 4-aminotryptophan was identified by protein crystallization and X-ray structure determination.[28] Nevertheless, the amount of Hpg-(4-Aza)Trp-ψ-b* was too low for crystallization purposes. To determine whether the same principles might explain these similar thermodynamical properties, modeling of possible interactions based on energy minimization between (4-Aza)Trp53 and the surrounding residues in ψ-b* was performed with MAIN by Dr. M. Debela.[103] For investigations of the 4-aminotryptophan53 interactions, the program LIGPLOTv4.4.2 was applied, which is presenting interactions between side chains in a two-dimensional mode, whereas MAIN allows a three-dimensional study of the structure.[28,54] As depicted in Fig. 29, the close examination of the two structures revealed no differences. Trp53 in the parent protein shows less hydrophobic interactions to the neighboring side chains than previously estimated with LIGPLOTv4.4.2 but a similar set of hydrophobic interactions and hydrogen bonds as (4-Aza)Trp53. However, although no differences in the hydrophobic interactions can be identified from the modeling data, amino- and azatryptophan brought ψ-b* to the edge of its structural integrity.

6.3.4 Copper(I)-catalyzed Huisgen cycloaddition (CCHC) with ψ-b variants*

Successful CCHC reactions with Hpg-(4-Aza)Trp-ψ-b* were performed with three different azide reaction partners and revealed not only high-yield incorporation but also the functionality of Hpg as the reactive handle after its incorporation into ψ-b*. Following azide-containing ligands were used for the click reaction: (1) the sugar derivative 1-azido-1-deoxy-β-D-glucopyranosid, (2) polyethylene glycol (PEG570), and (3) ψ-b* with incorporated Aha (Aha-Trp-ψ-b* via SPI method). The reaction products were checked by ESI-MS analyses.

Fig. 30 summarizes the results of the mass spectrometric analyses and confirms the desired outcome of the cycloaddition reactions with all azide-containing ligands. Efficient conjugation of the sugar derivative led to an increased mass of 10437.77 Da (measured 10436.4 Da).

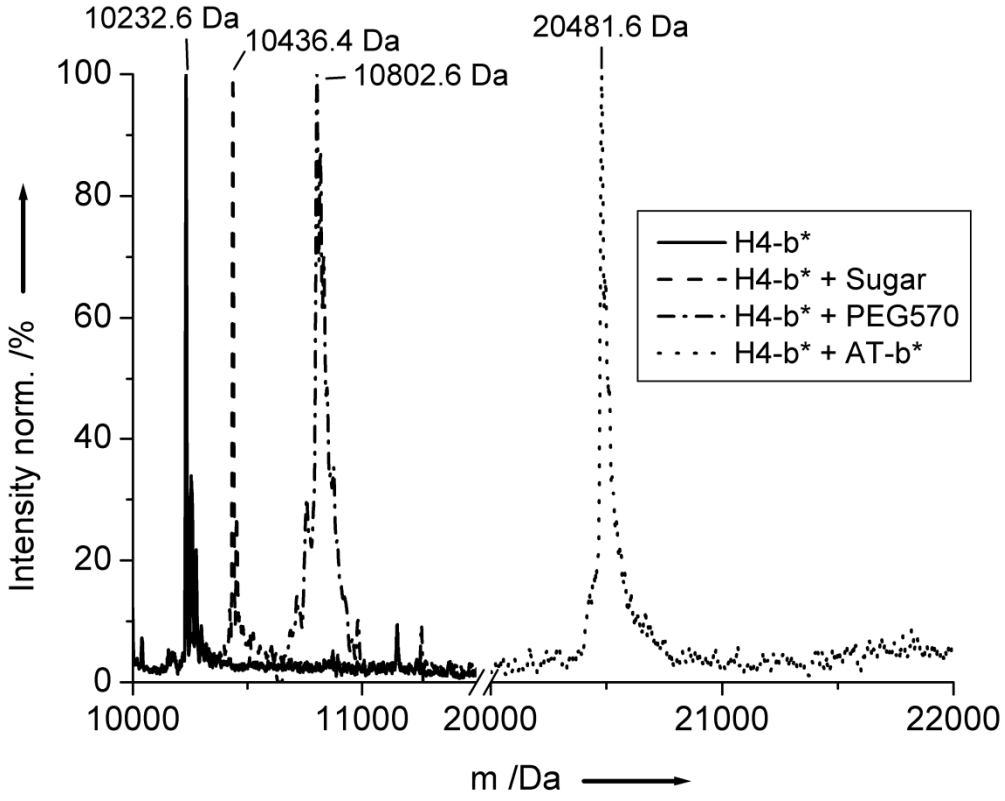


Fig. 30 Mass spectrometry of CCHC reactions with Hpg-(4-Aza)Trp-ψ-b*. Hpg-(4-Aza)Trp-ψ-b* (— H4-b*), was functionalized by CCHC reactions with different reaction partners: The sugar derivative 1-azido-1-deoxy-β-D-glucopyranosid (--- H4-b* + Sugar); with azido-PEG570 (-·-·- H4-b* + PEG570); and with Aha-Trp-ψ-b* (··· H4-b* + AT-b*).

The attachment of PEG570 to ψ-b* increases the mass to 10802.6 Da (determined mass 10802.6 Da), and finally the dimerization with Aha-Trp-ψ-b* duplicates the mass to 20481.2 Da (detected 20481.6 Da).

Glycosylations and PEGylations are among the most complex forms of protein modifications. Both modifications are in the focus of current research and are especially addressed in the context of click chemistry.[125,126,97,127,98] In comparison to small molecule drugs, proteins have a number of disadvantages as therapeutic agents since they are characterized by low oral bioavailability and very often short serum half-lives.[97,128,129] Covalently bound oligosaccharides and PEGs are known to increase a protein´s half-live by protecting the protein from proteases and reducing clearance rate by increasing the protein´s apparent size and its solubility. While these parameters are significantly influenced by the number, size and quality of the attached sugar and PEG derivatives, click chemistry, combined with classical site-directed mutagenesis, enables a highly specific and homogenous functionalization of recombinantly expressed proteins. Especially achieving a homogenous modified protein additionally requires an extensive purification procedure depending on the number of sites, which can be modified and are actually modified.[129] The mass spectrometric data demonstrated not only the successful incorporation of Hpg and (4-Aza)Trp into ψ-b* but also unambiguously proved the capacity of the reactive handle of Hpg to participate in CCHC reactions.

The click reaction batch of Hpg-(4-Aza)Trp-ψ-b* with Aha-Trp-ψ-b* was analyzed via SDS gel electrophoresis and the gel was exposed to UV light. As expected, a fluorescent band around 20 kDa was clearly visible in lane 1 (mass of two ψ-b* molecules dimerized via click reaction; see Fig. 31, left). Fluorescent bands around 10 kDa represent unconjugated (4-Aza)Trp-labeled protein and control samples. Subsequent Coomassie staining confirmed the protein band around 20 kDa (Fig. 31, right). All click reactions yielded the desired products, although further optimization is needed.

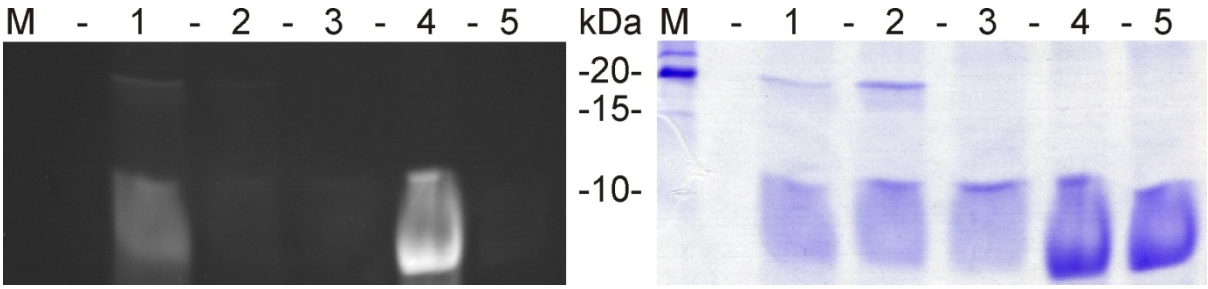


Fig. 31 Gel electrophoresis of ψ-b* and its tandemly labeled variant after CCHC reaction with Aha-Trp-ψ-b*. Under UV light **(left)** and after Coomassie staining **(right)**. **M:** Molecular weight protein standard; **1)** Hpg-(4-Aza)Trp-ψ-b* after CCHC reaction with Aha-Trp-ψ-b*; **2)** Hpg-Trp-ψ-b* after CCHC reaction with Aha-Trp-ψ-b*; **3)** Aha-Trp-ψ-b*; **4)** Hpg-(4-Aza)Trp-ψ-b*; **5)** Hpg-Trp-ψ-b*. Successful incorporation of (4-Aza)Trp into ψ-b* is clearly visible under UV light in lanes 1 and 4. The protein band ~20 kDa visible under UV light in lane 1 and after Coomassie staining in lanes 1 and 2 demonstrates successful CCHC reaction of the two ψ-b* variants (each ψ-b* ~10 kDa). Tandem incorporation of (4-Aza)Trp and Hpg into ψ-b* is demonstrated by the protein band ~20 kDa visible under UV light.

The click reaction with Aha-Trp-ψ-b* demonstrates an alternative for protein dimerization studies. Dimer- or oligomerization of proteins is extremely important in many intra- and extracellular processes.[130-132] For example, receptor dimerization is an important process in cell signaling and is therefore relevant for cancer research. On the other hand, click chemistry between proteins may represent an alternative tool to examine FRET between two fluorescent proteins and other biophysically measureable features, like conformational changes upon ligand binding. The CCHC reaction between Hpg-(4-Aza)Trp-ψ-b* and Aha-Trp-ψ-b* stands for a proof-of-principle for a new tool to dimerize proteins and the variety of possible ligands for click chemistry.

Click reactions were also performed with fluorescent dyes as azide-ligands to measure FRET (data not shown). However, these experiments proved to be difficult due to extremely low reaction yields and fluorescent dyes with suboptimal spectral properties. Further optimization of this reaction with different dyes will certainly lead to novel and useful FRET applications.

6.3.5 Secondary structure and folding cooperativity after PEGylation

As already introduced in chapter 6.3.4, PEGylation is expected to stabilize proteins.[133] In 1981, Gekko (and references cited therein) discussed that polyols, e.g. PEG, exhibited a protecting effect on proteins and biological structures from damage by heating or freezing and seemed to strengthening the hydrophobic interactions of the protein molecules.[134] The hydrophilic groups of PEG act as antagonists towards the nonpolar side chains, pushing nonpolar regions away and attracting polar regions. Recently, Naseem and Khan published a study on the effect of PEG on the acid-unfolded state of trypsinogen.[133] In the near-UV, they described a stabilizing effect on the protein´s tertiary structure and in the far-UV a transition of the unfolded towards an intermediate state, exhibiting structured elements, although different from the native one.

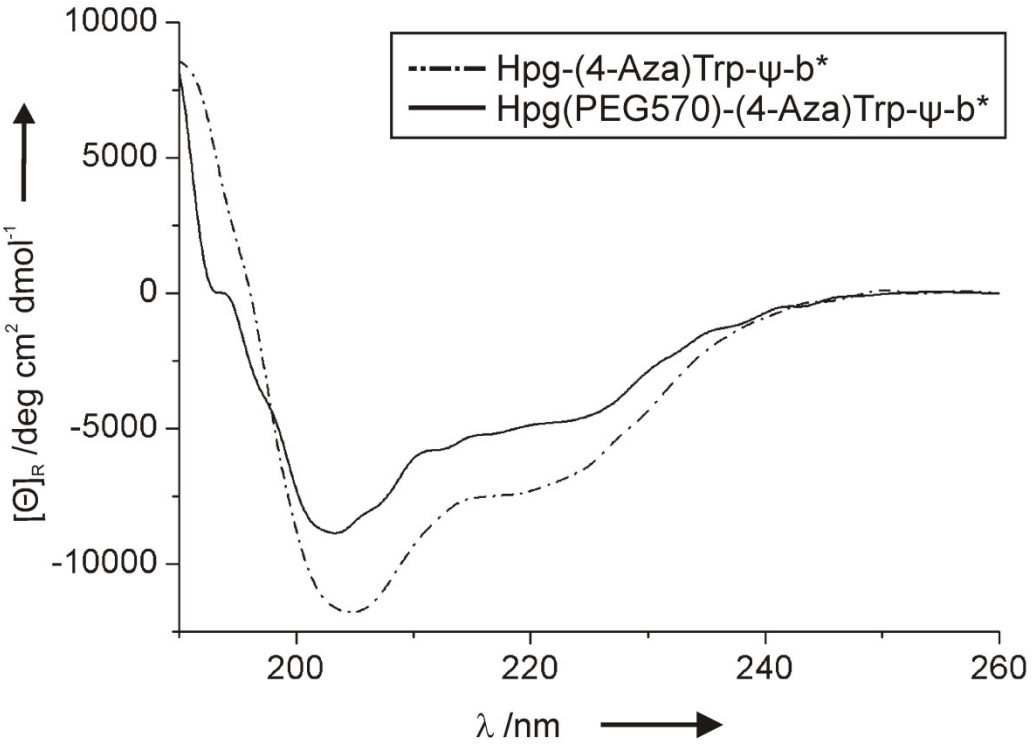


Fig. 32 CD spectra of Hpg-(4-Aza)Trp-ψ-b* (---) and Hpg(PEG570)-(4-Aza)Trp-ψ-b* (—). Far-UV CD profiles from 200 nm - 260 nm were recorded at 293.15 K.

To examine whether this stabilizing effect also accounts for PEG molecules attached by click chemistry, CD profile and thermally induced unfolding curve at 222 nm were monitored of Hpg-(4-Aza)Trp-ψ-b* after CCHC reaction with PEG570 (Hpg(PEG570)-(4-Aza)Trp-ψ-b*). As control, the CCHC reaction was also carried out with the parent protein and CD data were measured (data not shown).

As shown in Fig. 32, the CD profile of Hpg-(4-Aza)Trp-ψ-b* was not changed by the attachment of PEG570 (Hpg(PEG570)-(4-Aza)Trp-ψ-b*). However, there is a pronounced loss in signal intensity indicating protein aggregation due to the presence of copper(I). The destabilizing effect of copper(I) especially in combination with ascorbic acid is also described in the literature.[135-137] The shape of the CD profiles of Hpg-(4-Aza)Trp and its PEGylated counterpart are nearly identical indicating that the click reaction conditions did not induce any further change to the secondary structure (Fig. 32). This is in full agreement with the findings of other groups, who measured CD spectra of traditionally PEGylated interferon-α_{2b} and insulin.[138,128] They also detected no significant influence of PEG on the protein structures.

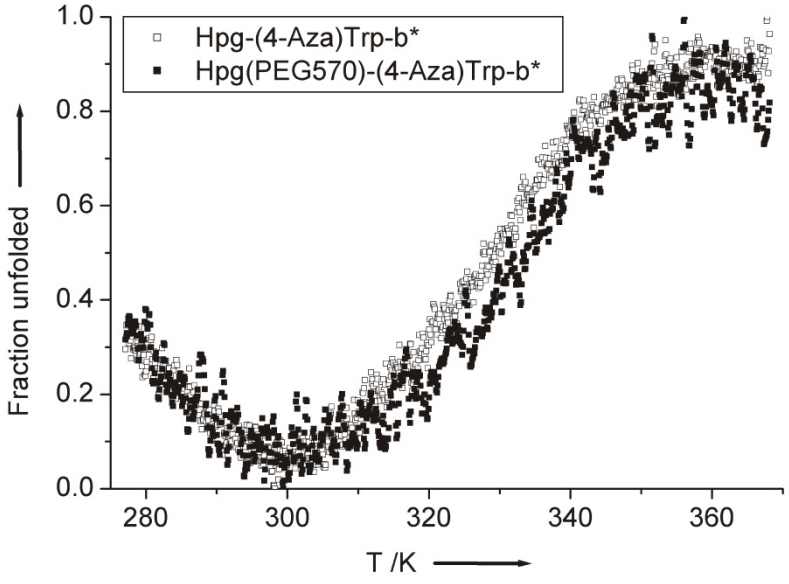


Fig. 33 Thermally induced unfolding profiles of Hpg-(4-Aza)Trp-ψ-b* (▫▫▫) and Hpg(PEG570)-(4-Aza)Trp-ψ-b* (▪▪▪). The fraction of unfolded protein was calculated from CD data monitored at 222 nm as described in Materials and Methods.

Conversely, thermally induced denaturation profiles of Hpg(PEG570)-(4-Aza)Trp-ψ-b* revealed a positive effect of the conjugated PEG570 on the stability of Hpg-(4-Aza)Trp-ψ-b* (Fig. 33). The T_m value of Hpg(PEG570)-(4-Aza)Trp-ψ-b* was significantly increased by about 22 K (T_m = 342.3 K) when compared to Hpg-(4-Aza)Trp-ψ-b* (Table 5). However, due to a quite high degree of signal noise, it should be kept in mind that the calculation of this T_m value increase might be highly overestimated. Furthermore, the melting curve of the PEGylated protein is still characterized by the cold denaturation-like phenomenon and reduced folding cooperativity.

To exclude a stabilizing effect of free PEG in the solution, also a mixture of Hpg-(4-Aza)Trp-ψ-b* and PEG570 without ascorbic acid and copper ions was measured, so that no CCHC reaction, i.e. PEGylation, could take place (data not shown). This melting curve revealed no difference when compared to the curve of Hpg-(4-Aza)Trp-ψ-b* without PEG570 and thus it can be ruled out that free PEG causes the potential increase of the T_m value.

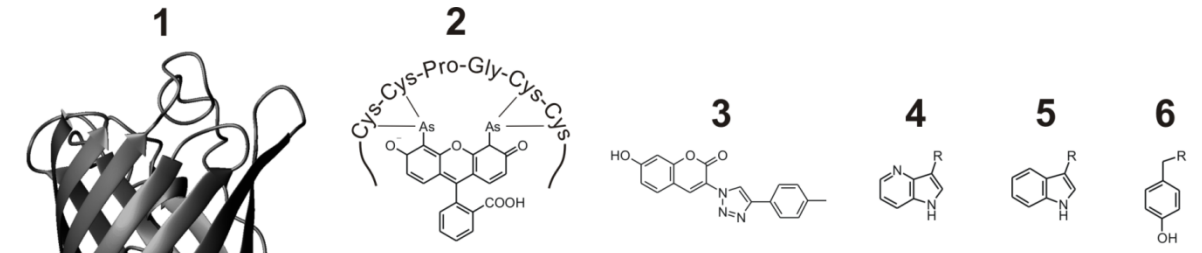
7 Conclusions and outlook

7.1 Towards new intrinsically colored proteins and cells

An ideal repertoire of fluorophores for bioimaging should be highly excitable at a single wavelength in a simple experimental set-up. In addition, their fluorescence emission should be characterized by a high quantum yield and be clearly distinguishable from each other. This Stokes-shift engineering was achieved with semiconductor nanocrystals in biological staining and diagnostics, however, the use of these materials in living biological systems is rather limited.[139,140]

Trp analogs are not only cell permeable and biocompatible but also easily co-translationally incorporated in proteins expressed in prokaryotic cells (chapter 3). The power of this protein engineering approach is best illustrated by the design of GdFP as discussed in chapter 3.9.4.[74] Nevertheless, GdFP`s red-shift cannot be transferred to any other protein and all GFP derived fluorescent proteins are relatively large in size, which can cause unpredictable perturbations within the studied biological system. To date, alternatives for intrinsic fluorescent probes are needed, but only the translationally active Trp analog (7-Aza)Ind was to some degree appropriate for this purpose. However, its fluorescence emission is strongly quenched in aqueous solutions, which limited its capacity for a broad usage in labeling and imaging applications.

This study introduces (4-Aza)Ind with its extraordinary spectral features as novel intrinsically fluorescent probe (chapter 6.2). Its incorporation into proteins enables simple and straightforward differentiation between the fluorescence emissions of different simultaneously excited protein variants. Certainly, (4-Aza)Trp protein variants open up new perspectives for multiple labeling applications as well as the design of genetically encoded physiological indicators. Fig. 34 compares commonly used fluorescent tags and imaging probes with respect to their size, applicability, and spectral properties.



	1	2	3	4	5	6
	EGFP	**FlAsH**	**3-Triazol-coumarin**	**(4-Aza)Trp**	**Trp**	**Tyr**
M_W [Da]	27700	700	319	205	204	181
λ_{max} (A) [nm]	489	508	441	289	280	274
λ_{max} (E) [nm]	509	528	495	425	350	304
Stokes-shift [nm]	20	20	53	**136**	70	30

Fig. 34 Comparison of fluorescent probes and spectral properties. Commonly used imaging tags and fluorescent probes, which are relevant for protein studies. The illustration should approximate the real size differences of the chemical components. From the biggest to the smallest compound: **1)** Member of the GFP family, EGFP, **2)** FlAsH, **3)** representative of the coumarin family, 3-triazol-coumarin, **4)** 4-azatryptophan, **5)** tryptophan, and **6)** tyrosine. All GFP derived fluorescent proteins as well as FlAsH require an engineering of the target amino acid sequence, moreover GFPs are relatively large in size, which can cause unpredictable perturbations within the studied system. The FlAsH method is based on the subsequent addition of arsenic compounds, which are toxic to many cells.[141] In contrast, 4-azatryptophan is structurally very similar to Trp, i.e. represents only a CH → N substitution. Furthermore, it comprises the largest Stokes-shift of all depicted imaging probes. The details of its spectroscopic properties are discussed in the chapter 6.2.3. R denotes $H_2CCH(NH_2)COOH$.

Apart from bulky fluorescent protein fusion tags (GFP derived proteins), even small fluorophores like FlAsH or coumarins, although they are cell-permeable, are useful only within limits: They require either the tailoring of the target gene sequences prior to labeling or bioorthogonal conjugation reactions, respectively.[141,142] In contrast, (4-Aza)Trp is intrinsically fluorescent and differs from Trp only by an "atomic exchange" (CH → N).

Currently, our group is looking for novel intrinsic fluorophores with excitation maxima above the range of 300 – 350 nm for wider applications in cell biology. Primarily, intrinsically green fluorescent optical probes, which can be excited with blue light, would represent a major advance. This "green area" is currently dominated by GFPs. However, these bulky fluorescence tags might become

redundant if synthetic chemistry provides the next generation of translationally active intrinsic optical probes, which exhibit high quantum yields and far red-shifted fluorescence covering a spectral window between 'green' and 'red'.

7.2 Between nucleic acids and proteins – Future research on nitrogen-containing fluorophores

In the scope of an expanded genetic code, in particular nitrogen-containing Trp analogs, either endocyclic or exocyclic, have led to significant changes in the spectral properties of the substituted protein.[28,143] Intriguingly, they have similarity in their structure to the purine nucleobases of DNA, adenine, guanine, and their synthetic analog 6-aminopurine, and share the capacity of the bases of pH-sensitive intramolecular charge transfer (Fig. 6, Fig. 8, and Fig. 35).[29]

To date, only Trp analogs with either exocyclic or endocyclic nitrogen atoms at the benzene ring of the Ind moiety were introduced into proteins.[29] However, DNA bases exhibit a combination of both exocyclic and endocyclic nitrogen atoms, charge transfer, and basicity. In particular, their exocyclic amino groups are key recognition sites in intra- and intermolecular interactions.[83]

The next step, after the incorporation of Trp analogs with either an aza- or an amino group, will be the generation of a purine like Trp analog, which possesses both aza- and/or amino-function/s. It is reasonable to expect that these nitrogen fitted Trp analogs may lead to a further enhancement of the spectral features with respect to the requirements of an ideal optical probe, as outlined in chapter 7.1. Endowing proteins with these "purine Trp analogs" will establish a new class of DNA/RNA-protein complexes which are able to mimic DNA or RNA, respectively ("DNA mimicry") and enable complex new interactions not only between proteins and proteins but also between proteins and DNA/RNA.[144] Furthermore, these substances may be targets for genetic code engineering by incorporating non-natural base pairs into DNA or RNA.[145]

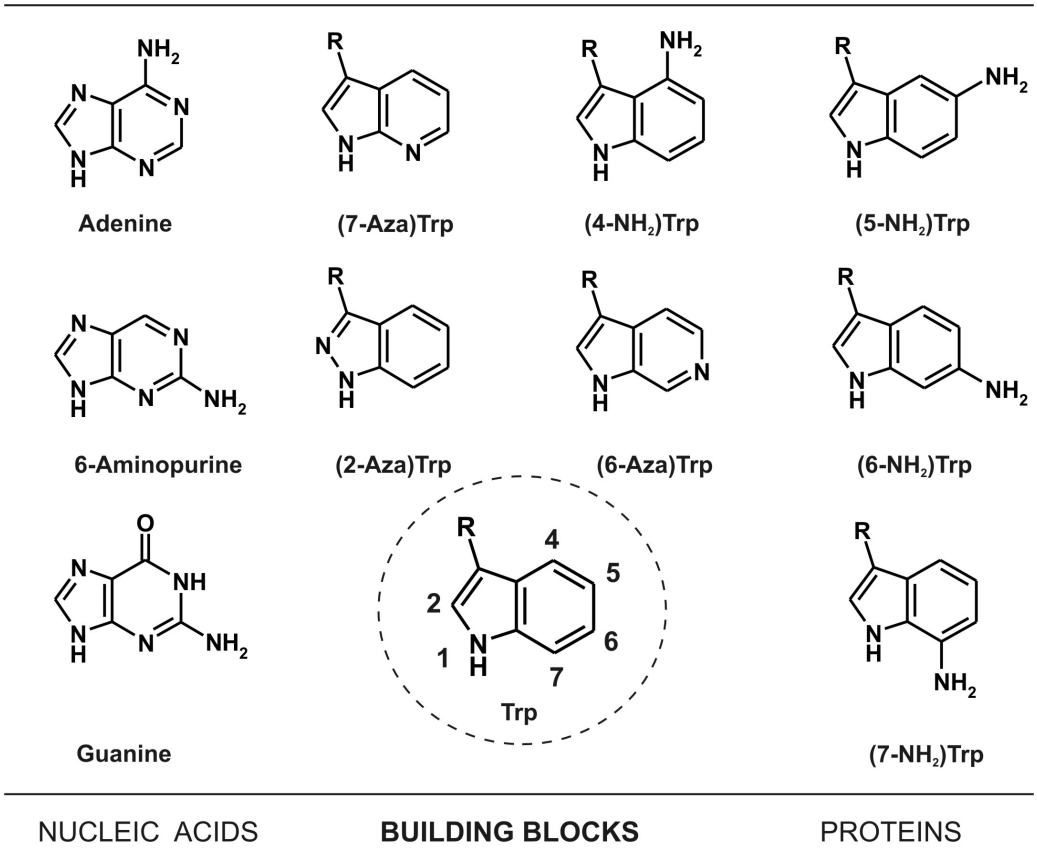


Fig. 35 Side chains of purine bases and of the canonical amino acid Trp and its exocyclic and endocyclic nitrogen-containing analogs. The purine nucleobases adenine and guanine as well as the synthetic nucleobase 6-aminopurine demonstrate a remarkable biophysical similarity to Trp and its amino- and aza-analogs.[29] The Trp analogs 4-aminotryptophan [(4-NH_2)Trp], 5-aminotryptophan [(5-NH_2)Trp], and (7-Aza)Trp are substrates of the protein synthesis *in vivo*. The translational activity of (4-Aza)Trp and (5-Aza)Trp (already depicted in Fig. 7) was presented in this study. Conversely, is has not been unambiguously demonstrated that also the analogs 6-aminotryptophan [(6-NH_2)Trp], 7-aminotryptophan [(7-NH_2)Trp], (2-Aza)Trp, and (6-Aza)Trp can be co-translationally introduced into proteins. R denotes $H_2CCH(NH_2)COOH$.

Nucleic acids and amino acids are strictly separated building blocks in the synthesis of biopolymers such as DNA/RNA on the one hand and polypeptides on the other hand. Purine-mimicking Trp analogs may build a bridge between both –until now– separated building block repertoires and may pave the way for the creation of DNA/RNA-protein "melting pots". The effect of the joined forces of nucleic bases, amino acids, and their analogs can hardly be estimated.

7.3 First tandem incorporation of two non-canonical amino acids

To date, the use of non-canonical amino acids in the frame of an expanded genetic code is a widely accepted approach for protein engineering and proved to be a powerful tool for the production of tailor-made proteins with enhanced features. The experimental methods are well-established but still in advance. Extraordinary novel features have been translated into various proteins ranging from distinct spectral properties to enhanced folding capacity and structural integrity.[27,29] Even the possibility to individually modify and equip any desired protein with chemical handles, application-aimed tags, fluorescent probes, and signal molecules has been reported.[100,98]

To date, SPI is limited to one-dimensional improvements due to the incorporation of only one target non-canonical amino acid into a protein at the same time. Consequently, only one protein feature can be modified directly and specifically: Either the biophysical properties of a certain protein were changed (fluorescence or stability) or a reactive handle for further targeted modifications was introduced. However, especially two topics are of particular interest for every protein scientist or cell biologist: The reliable and easy visualization of any desired protein in different environments, hydrophilic or hydrophobic, and the potential of specific and stoichiometric modifications for various purposes.

Therefore, the second major aim of this study was the development of an efficient method for the tandem incorporation of two distinct non-canonical amino acids simultaneously into one single protein. Stop codon suppression-based methodologies were also used for double labeling of proteins, however, low product yields and extensive engineering of the appropriate orthogonal aminoacyl-tRNA-synthetase/tRNA pair and codon optimization limits the applicability of these approaches.[146]

This thesis reports the first efficient *in vivo* tandem incorporation of two non-canonical amino acids into a single recombinant protein by using polyauxotrophic *E. coli* cells. Furthermore, it is conceivable that this approach can be easily applied for any other combination of non-canonical amino acids and for the incorporation of more than two amino acid analogs at the same time. Tandem incorporation enables the parallel tailoring of proteins with two distinct non-

canonical amino acids with different features for individual requirements. In this way, the expansion of the amino acid repertoire is potentially unlimited and of utmost importance for modern protein engineering and design.

In the future, one can expect that proteins are *in silico* designed for specific requirements and *de novo* built from a comprehensive and all-embracing library of canonical and non-canonical amino acids without limitations from cellular components, e.g. cellular uptake, metabolic interferences, ribosome architecture, or AARS proofreading mechanisms, either by using sophisticated *in vitro* expression platforms or specifically designed synthetic cells.

8 References

1. Nolting B*, et al.* (1995) Submillisecond Events in Protein-Folding. *Proc Natl Acad Sci USA* 92:10668-10672.
2. Miescher F (1871) Über die chemische Zusammensetzung der Eiterzellen. *Hoppe-Seyler´s medicinisch-chemische Untersuchungen* 4:441-460.
3. Avery OT*, et al.* (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a desoxyribonucleic acid fraction isolated from Pneumococcus type III. *Jour Exp Med* 79:137-158.
4. Franklin RE and Gosling RG (1953) Evidence for 2-Chain Helix in Crystalline Structure of Sodium Deoxyribonucleate. *Nature* 172:156-157.
5. Watson JD and Crick FHC (1953) The Structure of DNA. *Cold Spring Harbor Symposia on Quantitative Biology* 18:123-131.
6. Watson JD and Crick FHC (1953) Molecular Structure of Nucleic Acids - a Structure for Deoxyribose Nucleic Acid. *Nature* 171:737-738.
7. Watson JD and Crick FHC (1953) Genetical Implications of the Structure of Deoxyribonucleic Acid. *Nature* 171:964-967.
8. Chargaff E and Vischer E (1948) Nucleoproteins, Nucleic Acids, and Related Substances. *Annu Rev Biochem* 17:201-226.
9. Crick FHC (1957) On protein synthesis. *Symposium of the Society for Experimental Biology* 12:138-163.
10. Crick FHC (1966) Codon-Anticodon Pairing - Wobble Hypothesis. *J Mol Biol* 19:548-&.
11. Knippers R (2001) Molekulare Genetik, Vol. 8. (Georg Thieme Verlag, Stuttgart).
12. Bock A*, et al.* (1991) Selenocysteine - the 21st Amino-Acid. *Mol Microbiol* 5:515-520.

13. Srinivasan G*, et al.* (2002) Pyrrolysine encoded by UAG in Archaea: Charging of a UAG-decoding specialized tRNA. *Science* 296:1459-1462.

14. Budisa N (2005) Engineering the genetic code. (Wiley, Weinheim).

15. Chapeville F*, et al.* (1962) On Role of Soluble Ribonucleic Acid in Coding for Amino Acids. *Proc Natl Acad Sci USA* 48:1086-&.

16. Kennell D and Riezman H (1977) Transcription and Translation Initiation Frequencies of Escherichia-Coli Lac Operon. *J Mol Biol* 114:1-21.

17. Walsh CT*, et al.* (2005) Protein posttranslational modifications: The chemistry of proteome diversifications. *Angew Chem Int Ed Engl* 44:7342-7372.

18. Reinders J and Sickmann A (2007) Modificomics: Posttranslational modifications beyond protein phosphorylation and glycosylation. *Biomolecular Engineering* 24:169-177.

19. Cowie DB and Cohen GN (1957) Biosynthesis by Escherichia-Coli of Active Altered Proteins Containing Selenium Instead of Sulfur. *Biochim Biophys Acta* 26:252-261.

20. Beatty KE*, et al.* (2006) Fluorescence visualization of newly synthesized proteins in mammalian cells. *Angew Chem Int Ed Engl* 45:7364-7367.

21. Ealick SE (2000) Advances in multiple wavelength anomalous diffraction crystallography. *Curr Opin Chem Biol* 4:495-499.

22. Gerig JT (1994) Fluorine NMR of proteins. *Progress in Nuclear Magnetic Resonance Spectroscopy* 26:293-370.

23. Hohsaka T and Sisido M (2002) Incorporation of non-natural amino acids into proteins. *Curr Opin Chem Biol* 6:809-815.

24. Zhang ZW*, et al.* (2004) Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells. *Proc Natl Acad Sci USA* 101:8882-8887.

25. Kwon I and Tirrell DA (2007) Site-specific incorporation of tryptophan analogues into recombinant proteins in bacterial cells. *J Am Chem Soc* 129:10431-10437.

26. Cropp TA*, et al.* (2007) Reprogramming the amino-acid substrate specificity of orthogonal aminoacyl-tRNA synthetases to expand the genetic code of eukaryotic cells. *Nat Protoc* 2:2590-2600.

27. Budisa N (2004) Prolegomena to future experimental efforts on genetic code engineering by expanding its amino acid repertoire. *Angew Chem Int Ed Engl* 43:6426-6463.

28. Rubini M*, et al.* (2006) Aminotryptophan-containing barstar: Structure-function tradeoff in protein design and engineering with an expanded genetic code. *Biochim Biophys Acta -Proteins and Proteomics* 1764:1147-1158.

29. Lepthien S*, et al.* (2006) In vivo engineering of proteins with nitrogen-containing tryptophan analogs. *Appl Microbiol Biotechnol* 73:740-754.

30. Richmond MH (1962) Effect of Amino Acid Analogues on Growth and Protein Synthesis in Microorganisms. *Bacteriol Rev* 26:398-&.

31. Hortin G and Boime I (1983) Applications of Amino-Acid-Analogs for Studying Co-Translational and Posttranslational Modifications of Proteins. *Methods Enzymol* 96:777-784.

32. Wilson MJ and Hatfield DL (1984) Incorporation of Modified Amino-Acids into Proteins Invivo. *Biochim Biophys Acta* 781:205-215.

33. Kirk KL (1991) Biochemistry of the Elements, Vol. 9b. Biochemistry of Halogenated Organic Compounds. *Plenum Press: New York, USA; London, England, UK. Illus* XVII+362P.

34. Vanwinkle LJ (1993) Endogenous Amino-Acid-Transport Systems and Expression of Mammalian Amino-Acid-Transport Proteins in Xenopus-Oocytes. *Biochim Biophys Acta* 1154:157-172.

35. Reig N*, et al.* (2007) Functional and structural characterization of the first prokaryotic member of the L-amino acid transporter (LAT) family - A model for APC transporters. *J Biol Chem* 282:13270-13281.

36. Kramer R (1994) Systems and Mechanisms of Amino-Acid-Uptake and Excretion in Prokaryotes. *Archives of Microbiology* 162:1-13.

37. Giese C*, et al.* (2008) Inhibitory activity and intracellular uptake of aromatic fluorinated amino acids in human breast cancer cells. *ChemMedChem* 3:1449-1456.

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

39. Fowden L*, et al.* (1967) Toxic Amino Acids - Their Action as Antimetabolites. *Advances in Enzymology and Related Areas of Molecular Biology* 29:89-&.

40. Evans CS and Bell EA (1980) Neuroactive Plant Amino-Acids and Amines. *Trends in Neurosciences* 3:70-72.

41. O'Brien PJ and Herschlag D (1999) Catalytic promiscuity and the evolution of new enzymatic activities. *Chem Biol* 6:R91-R105.

42. Lepthien S (2004) Herstellung maßgeschneiderter Proteine mittels metabolischen Engineerings und erweiterten genetischen Codes. Diploma thesis, Leibniz University Hanover, Germany.

43. Link AJ*, et al.* (2003) Non-canonical amino acids in protein engineering. *Curr Opin Biotechnol* 14:603-609.

44. Kast P and Hennecke H (1991) Amino-Acid Substrate-Specificity of Escherichia-Coli Phenylalanyl-Transfer Rna-Synthetase Altered by Distinct Mutations. *J Mol Biol* 222:99-124.

45. Xie JM and Schultz PG (2006) Innovation: A chemical toolkit for proteins - an expanded genetic code. *Nat Rev Mol Cell Biol* 7:775-782.

46. Rodnina MV and Wintermeyer W (2001) Ribosome fidelity: tRNA discrimination, proofreading and induced fit. *Trends Biochem Sci* 26:124-130.

47. Hohsaka T*, et al.* (1993) Adaptability of Nonnatural Aromatic-Amino-Acids to the Active-Center of the Escherichia-Coli Ribosomal a-Site. *FEBS Lett* 335:47-50.

48. Dedkova LM*, et al.* (2003) Enhanced D-amino acid incorporation into protein by modified ribosomes. *J Am Chem Soc* 125:6616-6617.

49. Rose GD and Wolfenden R (1993) Hydrogen-Bonding, Hydrophobicity, Packing, and Protein-Folding. *Annu Rev Biophys Biomol Struct* 22:381-415.

50. Wolfenden R and Radzicka A (1986) How Hydrophilic Is Tryptophan. *Trends Biochem Sci* 11:69-70.

51. Trinquier G and Sanejouand YH (1998) Which effective property of amino acids is best preserved by the genetic code? *Protein Eng* 11:153-169.

52. Kauzmann W (1959) Some Factors in the Interpretation of Protein Denaturation. *Advances in Protein Chemistry* 14:1-63.

53. Baldwin RL (2002) Protein folding - Making a network of hydrophobic clusters. *Science* 295:1657-1658.

54. Wallace AC*, et al.* (1995) Ligplot - a Program to Generate Schematic Diagrams of Protein Ligand Interactions. *Protein Eng* 8:127-134.

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

56. Sridevi K*, et al.* (2000) The slow folding reaction of barstar: The core tryptophan region attains tight packing before substantial secondary and tertiary structure formation and final compaction of the polypeptide chain. *J Mol Biol* 302:479-495.

57. Wong KB*, et al.* (1996) Cold denaturation of barstar: H-1, N-15 and C-13 NMR assignment and characterisation of residual structure. *J Mol Biol* 259:805-818.

58. Babu CR*, et al.* (2004) Direct access to the cooperative substructure of proteins and the protein ensemble via cold denaturation. *Nature Structural & Molecular Biology* 11:352-357.

59. Rosenthal G (1982) Plant nonprotein amino and imino acids. Biological, biochemical, and toxicological properties. (Academic, New York).

60. Lakowitz JR (1999) Protein fluorescence, 2. Edition (Kluwer Academic/Plenum Publisher, New York).

61. Platt JR (1951) Isoconjugate Spectra and Variconjugate Sequences. *Journal of Chemical Physics* 19:101-118.

62. Sinha HK*, et al.* (1987) Excited-State and Ground-State Proton-Transfer Reactions in 5-Aminoindole. *Bulletin of the Chemical Society of Japan* 60:4401-4407.

63. Kishi T*, et al.* (1977) Electronic Absorption and Fluorescence-Spectra of 5-Hydroxytryptamine (Serotonin) - Protonation in Excited-State. *Bulletin of the Chemical Society of Japan* 50:1267-1271.

64. Ross JBA*, et al.* in Fluorescence Spectroscopy, Vol. 278 151-1901997).

65. Budisa N*, et al.* (2001) Proteins with beta-(thienopyrrolyl)alanines as alternative chromophores and pharmaceutically active amino acids. *Protein Sci* 10:1281-1292.

66. Budisa N and Pal PP (2004) Designing novel spectral classes of proteins with a tryptophan-expanded genetic code. *Biol Chem* 385:893-904.

67. Brawerman G and Ycas M (1957) Incorporation of the Amino Acid Analog Tryptazan into the Protein of Escherichia-Coli. *Arch Biochem Biophys* 68:112-117.

68. Pardee AB*, et al.* (1956) Incorporation of Azatryptophan into Proteins of Bacteria and Bacteriophage. *Biochim Biophys Acta* 21:406-407.

69. Schlesinger S and Schlesinger MJ (1967) Effect of Amino Acid Analogues on Alkaline Phosphatase Formation I Escherichia Coli K-12. 1. Substitution of Triazolealanine for Histidine. *J Biol Chem* 242:3369-&.

70. Sykes BD*, et al.* (1974) Fluorotyrosine Alkaline-Phosphatase from Escherichia-Coli - Preparation, Properties, and Fluorine-19 Nuclear Magnetic-Resonance Spectrum. *Proc Natl Acad Sci USA* 71:469-473.

71. Ross JBA*, et al.* in Trends in fluorescence spectroscopy, Vol. 6. (Editor Lakowitz JR) (Plenum, New York; 2000).

72. Bae JH*, et al.* (2001) Incorporation of beta-selenolo[3,2-b]pyrrolyl-alanine into proteins for phase determination in protein X-ray crystallography. *J Mol Biol* 309:925-936.

73. Budisa N*, et al.* (2002) Global replacement of tryptophan with aminotryptophans generates non-invasive protein-based optical pH sensors. *Angew Chem Int Ed Engl* 41:4066-4069.

74. Bae JH*, et al.* (2003) Expansion of the genetic code enables design of a novel "gold'' class of green fluorescent proteins. *J Mol Biol* 328:1071-1081.

75. Patterson GH*, et al.* (1997) Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. *Biophys J* 73:2782-2790.

76. Tsien RY (1998) The green fluorescent protein. *Annu Rev Biochem* 67:509-544.

77. Campbell RE*, et al.* (2002) A monomeric red fluorescent protein. *Proc Natl Acad Sci USA* 99:7877-7882.

78. Baird GS*, et al.* (2000) Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *Proc Natl Acad Sci USA* 97:11984-11989.

79. Gross LA*, et al.* (2000) The structure of the chromophore within DsRed, a red fluorescent protein from coral. *Proc Natl Acad Sci USA* 97:11990-11995.

80. Martynov VI*, et al.* (2001) Alternative cyclization in GFP-like proteins family - The formation and structure of the chromophore of a purple chromoprotein from Anemonia sulcata. *J Biol Chem* 276:21012-21016.

81. Wang K*, et al.* (2002) Synthesis and fluorescence study of 7-azaindole in DNA oligonucleotides replacing a purine base. *Spectrochim Acta A Mol Biomol Spectrosc* 58:2595-2603.

82. Smirnov AV*, et al.* (1997) Photophysics and biological applications of 7-azaindole and its analogs. *J Phys Chem B* 101:2758-2769.

83. Shafirovich V and Geacintov NE in Charge Transfer in DNA, pp175-193, Editor Wagenknecht HA (Wiley-VCH, Weinheim; 2005).

84. Soumillion P*, et al.* (1995) Biosynthetic Incorporation of 7-Azatryptophan into the Phage-Lambda Lysozyme - Estimation of Tryptophan Accessibility, Effect on Enzymatic-Activity and Protein Stability. *Protein Eng* 8:451-456.

85. Twine SM*, et al.* (2003) The photophysical properties of 6-azaindole. *J Phys Chem B* 107:637-645.

86. Twine SM and Szabo AG (2003) Fluorescent Amino Acid Analogs. *Methods Enzymol* 360:104-127.

87. Rich RL*, et al.* (1993) Steady-State and Time-Resolved Fluorescence Anisotropy of 7-Azaindole and Its Derivatives. *J Phys Chem B* 97:1781-1788.

88. Cash MT*, et al.* (2005) Excited state tautomerization of azaindole. *Org Biomol Chem* 3:3701-3706.

89. Michael A (1893) Über die Einwirkung von Diazobenzolimid auf Acetylendicarbonsäuremethylester. *J. Prakt. Chem.* 48:94-95.

90. Huisgen R (1963) 1.3-Dipolare Cycloadditionen - Rückschau und Ausblick. *Angew Chem Int Ed Engl* 75:604-+.

91. Huisgen R (1963) Kinetik Und Mechanismus 1.3-Dipolarer Cycloadditionen. *Angew Chem Int Ed Engl* 75:742-&.

92. Rostovtsev VV*, et al.* (2002) A stepwise Huisgen cycloaddition process: Copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew Chem Int Ed Engl* 41:2596-+.

93. Tornoe CW*, et al.* (2002) Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J Org Chem* 67:3057-3064.

94. Kolb HC*, et al.* (2001) Click chemistry: Diverse chemical function from a few good reactions. *Angew Chem Int Ed Engl* 40:2004-+.

95. Agard NJ*, et al.* (2004) A strain-promoted [3+2] azide-alkyne cycloaddition for covalent modification of blomolecules in living systems. *J Am Chem Soc* 126:15046-15047.

96. Link AJ*, et al.* (2006) Discovery of aminoacyl-tRNA synthetase activity through cell-surface display of noncanonical amino acids. *Proc Natl Acad Sci USA* 103:10180-10185.

97. Deiters A*, et al.* (2004) Site-specific PEGylation of proteins containing unnatural amino acids. *Bioorg Med Chem Lett* 14:5743-5745.

98. Merkel L*, et al.* (2008) Efficient N-terminal glycoconjugation of proteins by the N-end rule. *Chembiochem* 9:1220-1224.

99. van Hest JCM*, et al.* (2000) Efficient incorporation of unsaturated methionine analogues into proteins in vivo. *J Am Chem Soc* 122:1282-1288.

100. Link AJ*, et al.* (2004) Presentation and detection of azide functionality in bacterial cell surface proteins. *J Am Chem Soc* 126:10598-10602.

101. Budisa N*, et al.* (1995) High-Level Biosynthetic Substitution of Methionine in Proteins by Its Analogs 2-Aminohexanoic Acid, Selenomethionine, Telluromethionine and Ethionine in Escherichia-Coli. *Eur J Biochem* 230:788-796.

102. Yang LH*, et al.* (1996) PCR mutagenesis and overexpression of tryptophan synthase from Salmonella typhimurium: On the roles of beta(2) subunit Lys-382. *Protein Expr Purif* 8:126-136.

103. Turk D (1992) Weiterentwicklung eines Programms für Molekülgraphik und Elektronendichte-Manipulation und seine Anwendung auf verschiedene Protein-Strukturaufklärungen, PhD Thesis, Technical University, Munich, Germany

104. Gasteiger E*, et al.* in The Proteomics Protocols Handbook, pp571-607, Editor Walker JM (Humana Press, Heidelberg; 2005).

105. Burger A*, et al.* (1993) A Rapid and Efficient Purification Method for Recombinant Annexin-V for Biophysical Studies. *FEBS Lett* 329:25-28.

106. Blackwood RA and Ernst JD (1990) Characterization of Ca-2+-Dependent Phospholipid Binding, Vesicle Aggregation and Membrane-Fusion by Annexins. *Biochem J* 266:195-200.

107. Ernst JD*, et al.* (1994) Annexins Possess Functionally Distinguishable Ca2+ and Phospholipid-Binding Domains. *Biochem Biophys Res Commun* 200:867-876.

108. Huber R*, et al.* (1990) The Calcium-Binding Sites in Human Annexin-V by Crystal-Structure Analysis at 2.0 a Resolution - Implications for Membrane-Binding and Calcium-Channel Activity. *FEBS Lett* 275:15-21.

109. Huber R*, et al.* (1990) The Crystal and Molecular-Structure of Human Annexin-V, an Anticoagulant Protein That Binds to Calcium and Membranes. *EMBO J* 9:3867-3874.

110. Hartley RW (1988) Barnase and Barstar - Expression of Its Cloned Inhibitor Permits Expression of a Cloned Ribonuclease. *J Mol Biol* 202:913-915.

111. Guillet V*, et al.* (1993) Recognition between a Bacterial Ribonuclease, Barnase, and Its Natural Inhibitor, Barstar. *Structure* 1:165-176.

112. Lubienski MJ*, et al.* (1994) 3-Dimensional Solution Structure and C-13 Assignments of Barstar Using Nuclear-Magnetic-Resonance Spectroscopy. *Biochemistry* 33:8866-8877.

113. Khurana R and Udgaonkar JB (1994) Equilibrium Unfolding Studies of Barstar - Evidence for an Alternative Conformation Which Resembles a Molten Globule. *Biochemistry* 33:106-115.

114. Swaminathan R*, et al.* (1994) Molten Globule-Like Conformation of Barstar - a Study by Fluorescence Dynamics. *Journal of Physical Chemistry* 98:9270-9278.

115. Schreiber G and Fersht AR (1993) The Refolding of Cis-Peptidylprolyl and Trans-Peptidylprolyl Isomers of Barstar. *Biochemistry* 32:11195-11203.

116. Nolting B*, et al.* (1997) The folding pathway of a protein at high resolution from microseconds to seconds. *Proc Natl Acad Sci USA* 94:826-830.

117. Nath U*, et al.* (1996) Initial loss of secondary structure in the unfolding of barstar. *Nature Structural Biology* 3:920-923.

118. Pardee AB*, et al.* (1956) Incorporation of Azatryptophan into Proteins of Bacteria and Bacteriophage. *Biochim Biophys Acta* 21:406-407.

119. Broos J*, et al.* (2003) Efficient biosynthetic incorporation of tryptophan and indole analogs in an integral membrane protein. *Protein Sci* 12:1991-2000.

120. El Khattabi M*, et al.* (2008) Lactococcus lactis as expression host for the biosynthetic incorporation of tryptophan analogues into recombinant proteins. *Biochem J* 409:193-198.

121. Chen Y*, et al.* (1993) Fluorescent Species of 7-Azaindole and 7-Azatryptophan in Water. *J Phys Chem B* 97:1770-1780.

122. Gai F*, et al.* (1992) Nonradiative Pathways of 7-Azaindole in Water. *J Am Chem Soc* 114:8343-8345.

123. Negrerie M*, et al.* (1991) Photophysics of a Novel Optical Probe - 7-Azaindole. *J Phys Chem B* 95:8663-8670.

124. Fasman GDE (1996) Circular Dichroism and the Conformational Analysis of Biomolecules. (Plenum Press, New York).

125. Davis BG (2002) Synthesis of glycoproteins. *Chem Rev* 102:579-601.

126. Veronese FM and Harris JM (2002) Preface - Introduction and overview of peptide and protein pegylation. *Adv Drug Deliv Rev* 54:453-456.

127. Le Drournaguet B and Velonia K (2008) Click chemistry: A powerful tool to create polymer-based macromolecular chimeras. *Macromolecular Rapid Communications* 29:1073-1089.

128. Hinds KD and Kim SW (2002) Effects of PEG conjugation on insulin properties. *Adv Drug Deliv Rev* 54:505-530.

129. Roberts MJ*, et al.* (2002) Chemistry for peptide and protein PEGylation. *Adv Drug Deliv Rev* 54:459-476.

130. Wang JH and Norcross M (2008) Dimerization of chemokine receptors in living cells: key to receptor function and novel targets for therapy. *Drug Discov Today* 13:625-632.

131. Shimozono S and Miyawaki A (2008) Engineering FRET constructs using CFP and YFP. *Fluorescent Proteins, Second Edition* 85:381-+.

132. Ferguson KM (2008) Structure-based view of epidermal growth factor receptor regulation. *Annu Rev Biophys* 37:353-373.

133. Naseem F and Khan RH (2003) Effect of ethylene glycol and polyethylene glycol on the acid-unfolded state of trypsinogen. *J Protein Chem* 22:677-682.

134. Gekko K (1981) Mechanism of Polyol-Induced Protein Stabilization - Solubility of Amino-Acids and Diglycine in Aqueous Polyol Solutions. *J Biochem* 90:1633-1641.

135. Chiou SH (1983) DNA - Scission and Proteolytic Activities of Ascorbate in the Presence of Copper-Ion and a Copper-Peptide Complex. *Federation Proceedings* 42:2179-2179.

136. Stadtman ER (1990) Metal Ion-Catalyzed Oxidation of Proteins - Biochemical-Mechanism and Biological Consequences. *Free Radic Biol Med* 9:315-325.

137. Stadtman ER*, et al.* (2003) Oxidation of methionine residues of proteins: Biological consequences. *Antioxidants & Redox Signaling* 5:577-582.

138. Grace M*, et al.* (2001) Structural and biologic characterization of pegylated recombinant IFN-alpha 2b. *J Interferon Cytokine Res* 21:1103-1115.

139. Bruchez M*, et al.* (1998) Semiconductor nanocrystals as fluorescent biological labels. *Science* 281:2013-2016.

140. Chan WCW and Nie SM (1998) Quantum dot bioconjugates for ultrasensitive nonisotopic detection. *Science* 281:2016-2018.

141. Griffin BA*, et al.* (1998) Specific covalent labeling of recombinant protein molecules inside live cells. *Science* 281:269-272.

142. Zhao YR*, et al.* (2004) New caged coumarin fluorophores with extraordinary uncaging cross sections suitable for biological imaging applications. *J Am Chem Soc* 126:4653-4663.

143. Lepthien S*, et al.* (2008) Azatryptophans endow proteins with intrinsic blue fluorescence. *Proc Natl Acad Sci USA* in press.

144. Dryden DTF and Tock MR (2006) DNA mimicry by proteins. *Biochemical Society Transactions* 34:317-319.

145. Kimoto M*, et al.* (2004) Site-specific incorporation of a photo-crosslinking component into RNA by T7 transcription mediated by unnatural base pairs. *Chem Biol* 11:47-55.

146. Hohsaka T*, et al.* (1999) Incorporation of two different nonnatural amino acids independently into a single protein through extension of the genetic code. *J Am Chem Soc* 121:12194-12195.

9 Figure List

10 Table List

11 Appendix

11.1 Detailed presentations of mass spectrometric analyses of anxA5 variants

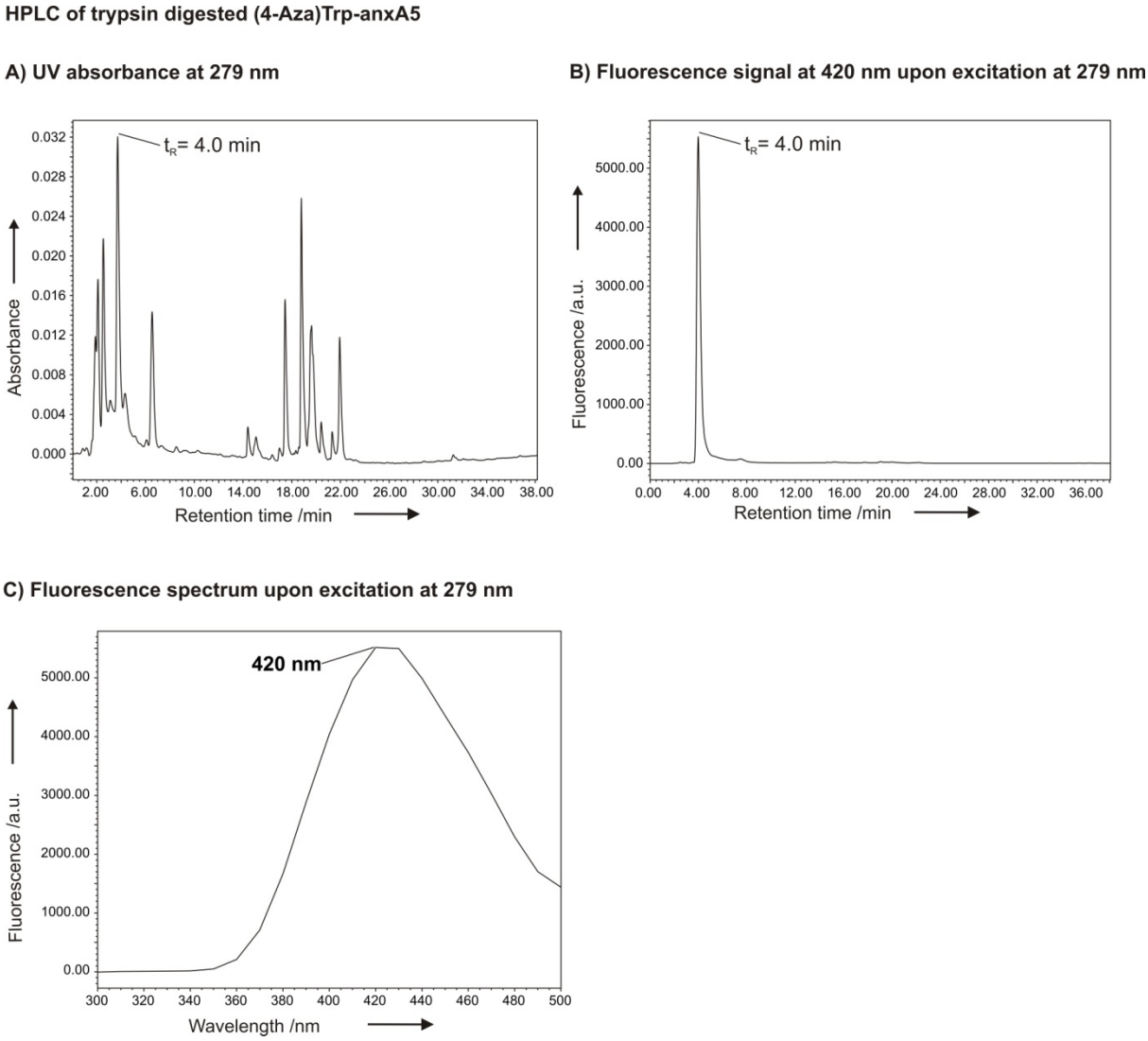


Fig. 36 HPLC analysis of trypsin digested (4-Aza)Trp-anxA5.

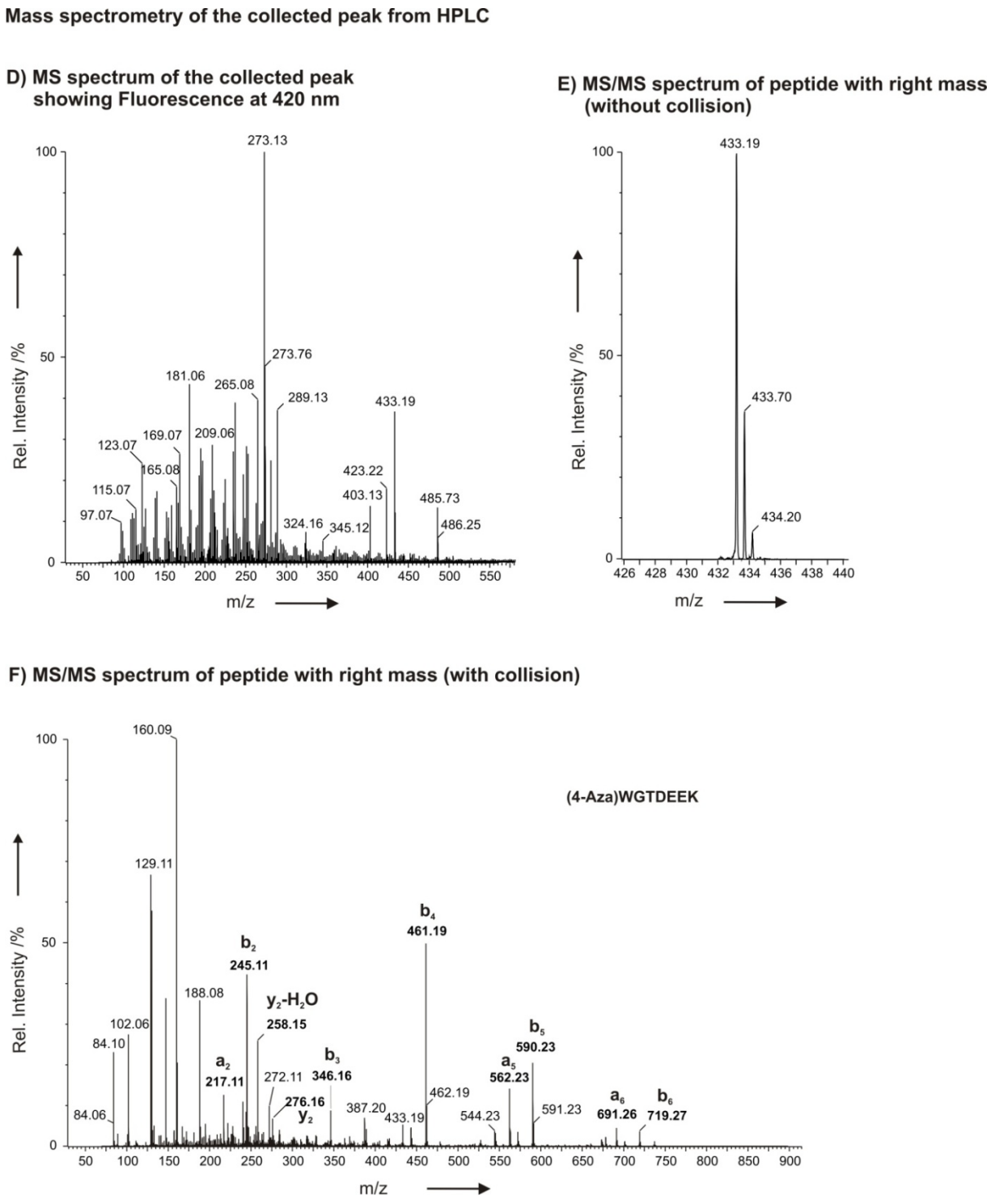


Fig. 37 Mass spectrometric analysis of (4-Aza)Trp-anxA5.

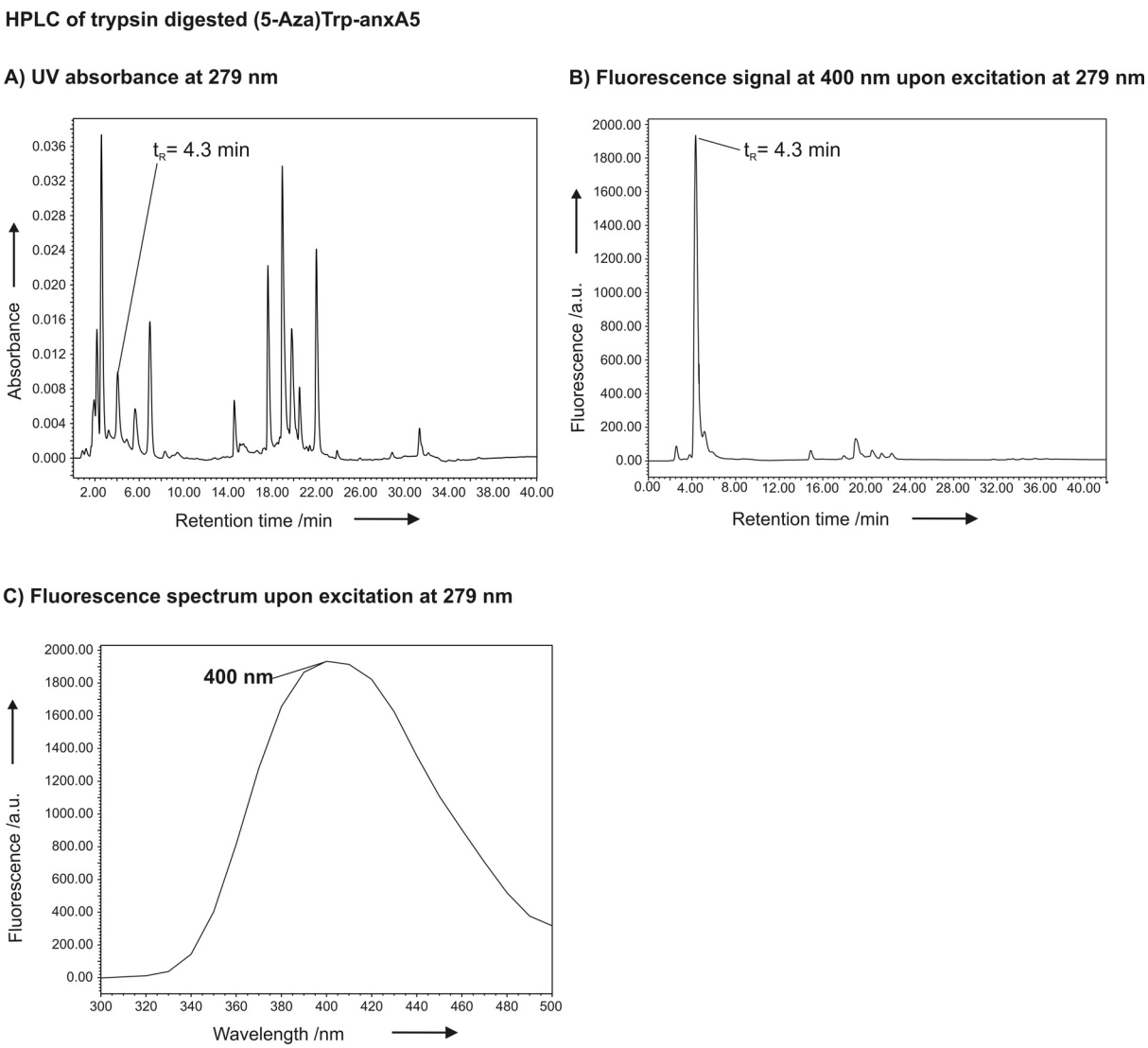
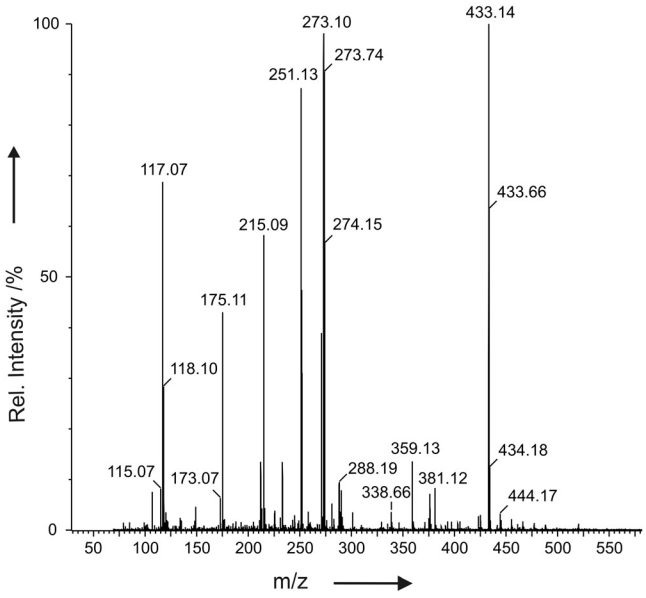


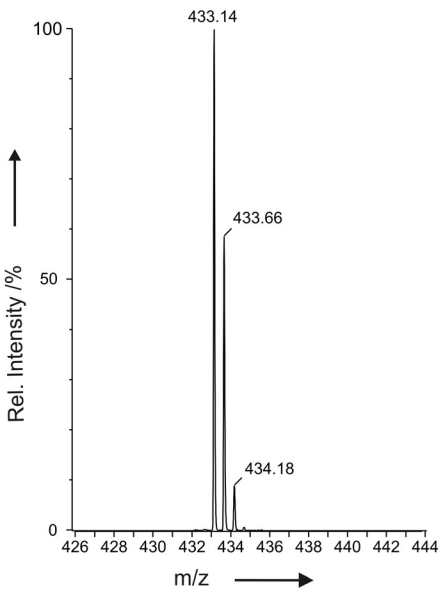
Fig. 38 HPLC analysis of trypsin digested (5-Aza)Trp-anxA5.

Mass spectrometry of the collected peak from HPLC

D) MS spectrum of the collected peak showing Fluorescence at 400 nm



E) MS/MS spectrum of peptide with right mass (without collision)



F) MS/MS spectrum of peptide with right mass (with collision)

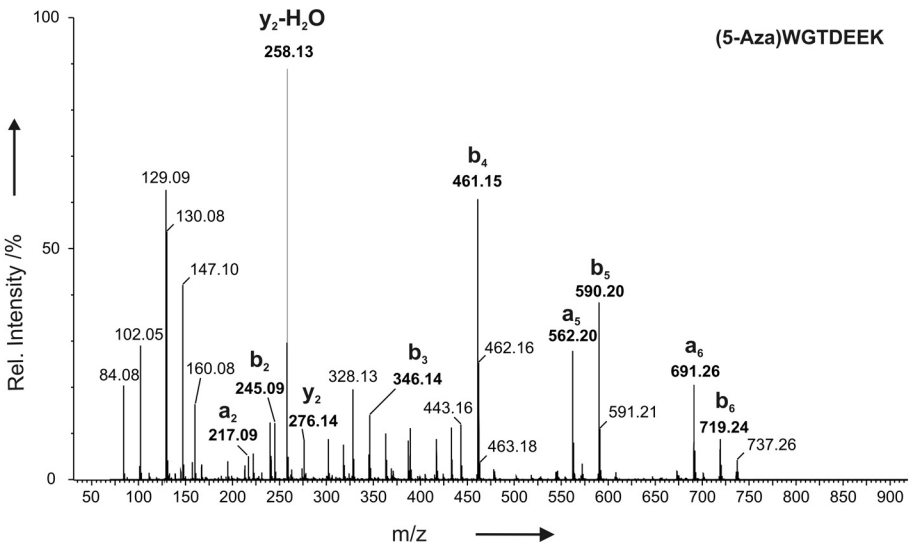


Fig. 39 Mass spectrometric analysis of (5-Aza)Trp-anxA5.

11.2 Amino acid hydrolysis – Data presentation

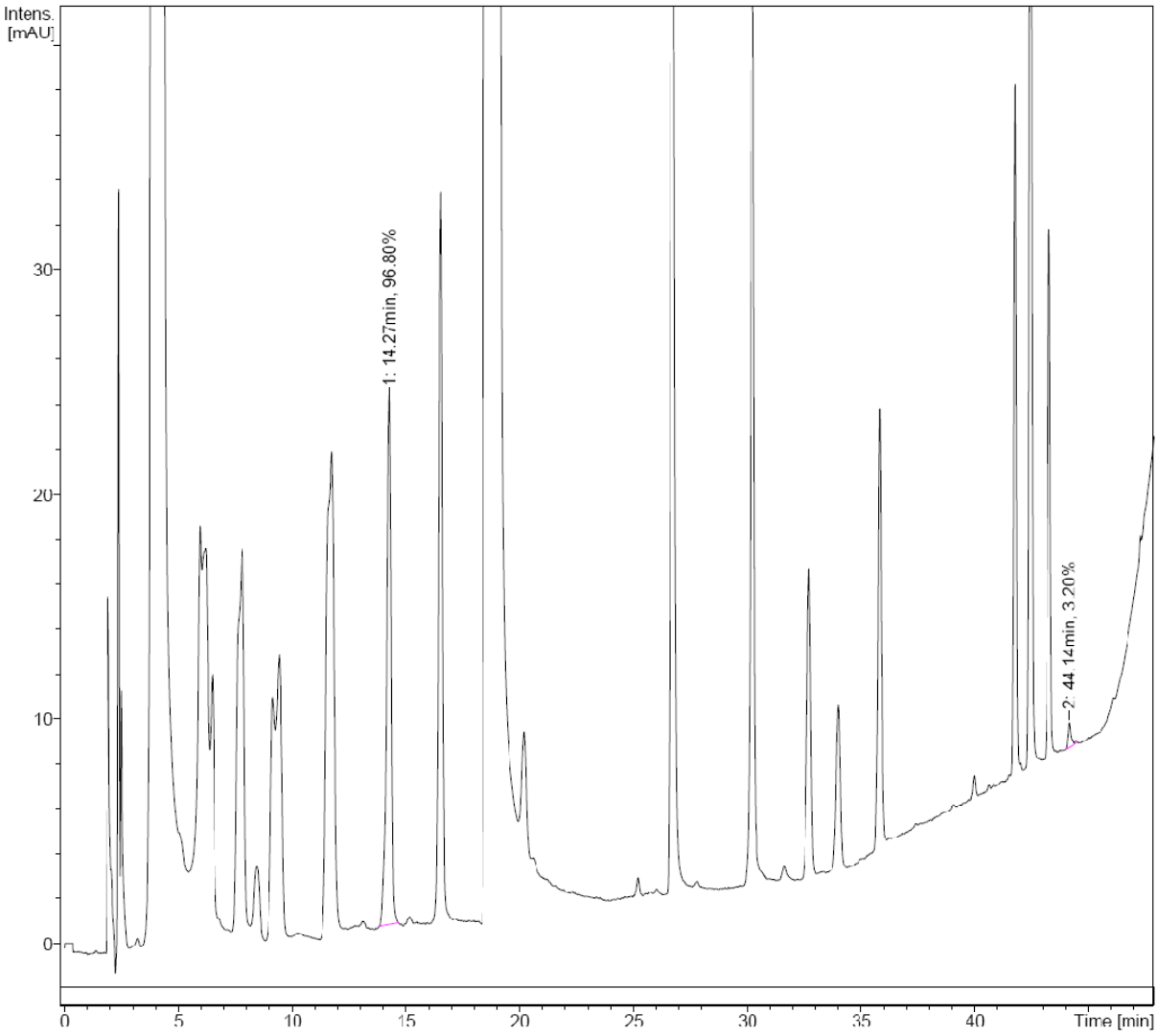


Fig. 40 Chromatogram of the amino acid hydrolysis of the parent protein Trp-anxA5 (Peak 1: Threonine; peak 2: Trp). The threonine and the Trp peaks were used for the calculation of the (5-Aza)Trp incorporation level as described in Material and Methods, chapter 5.1.7.

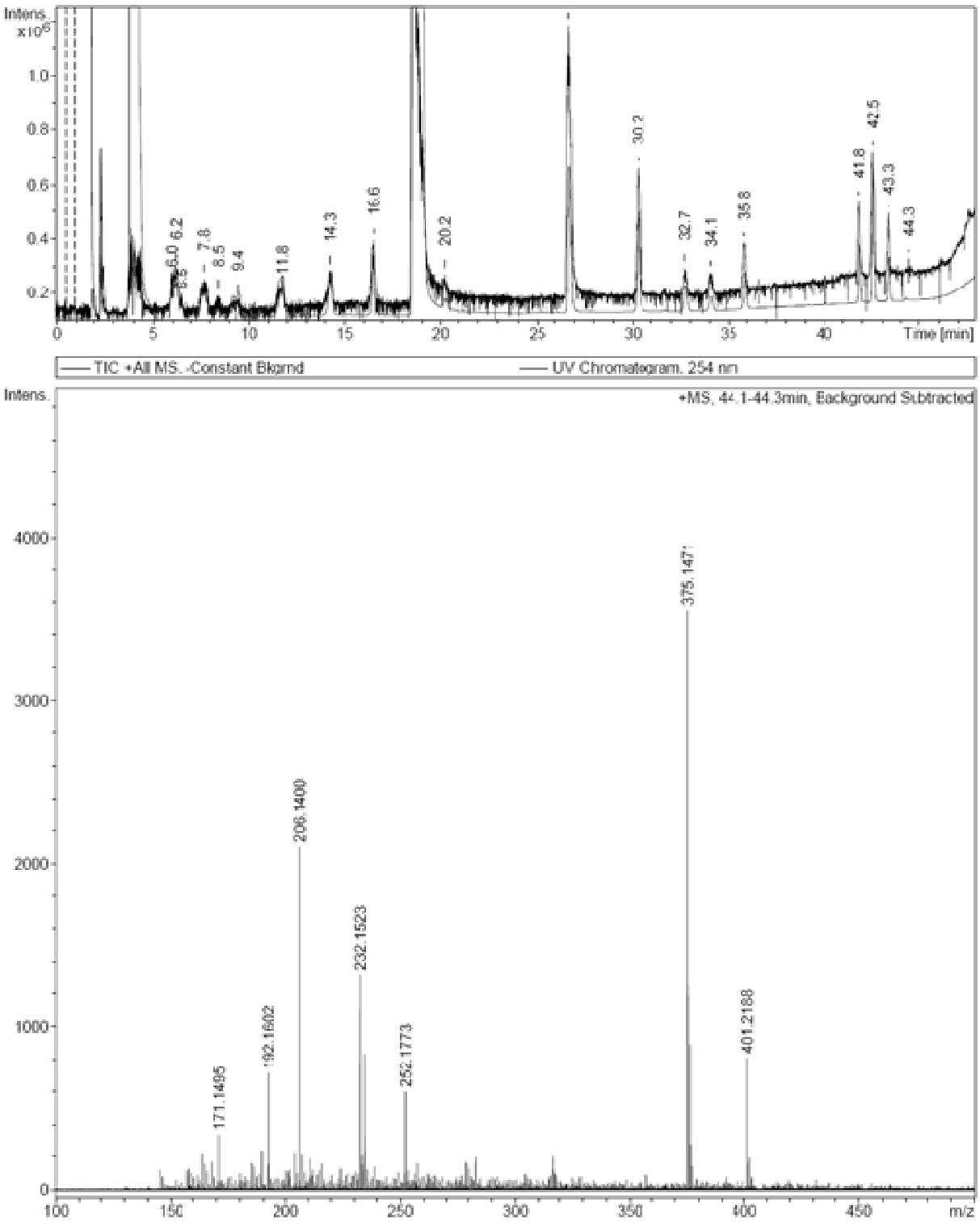


Fig. 41 Mass spectrometric analysis of the hydrolyzed parent protein Trp-anxA5 with the mass of derivatized Trp (375.15 Da).

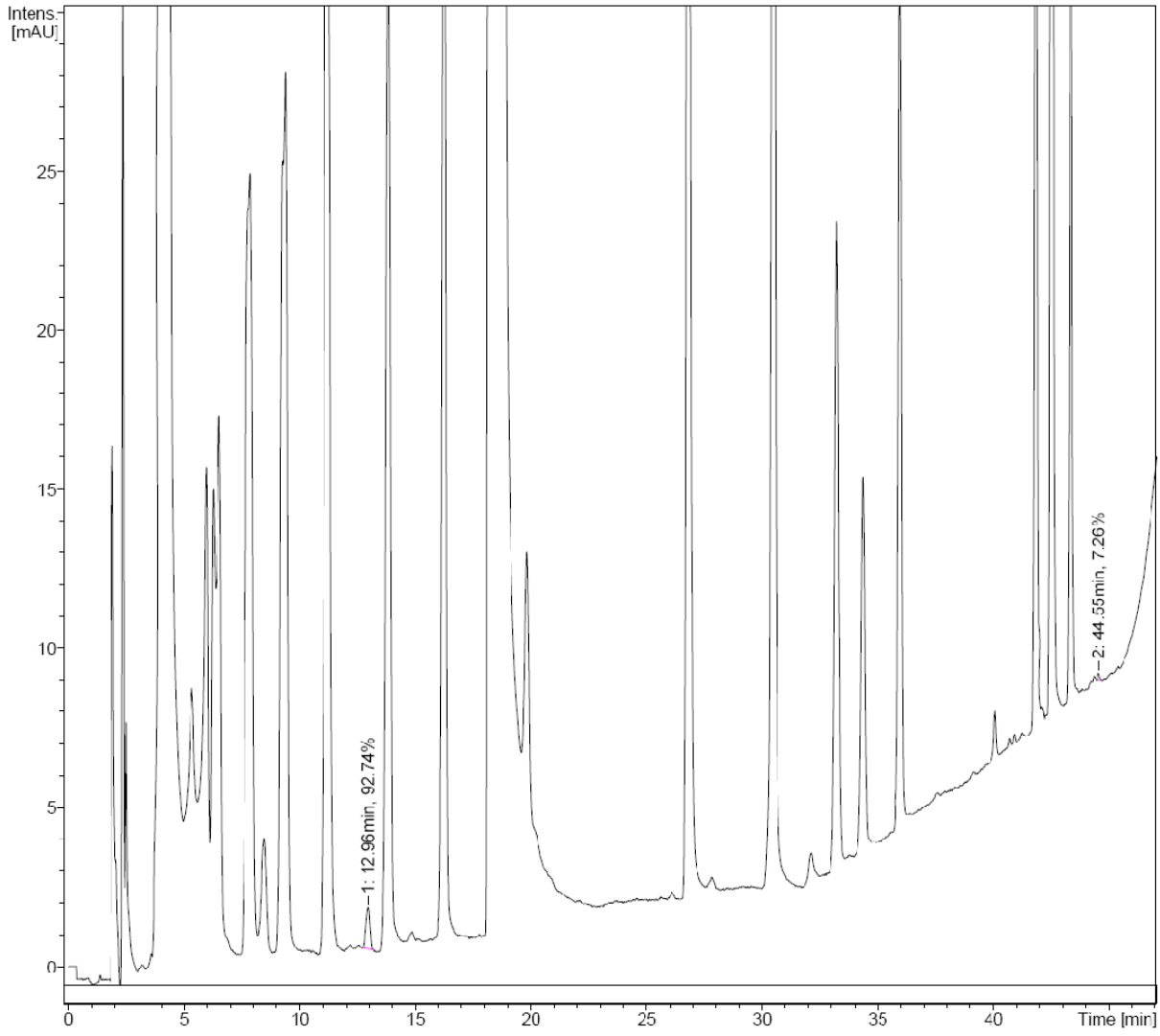


Fig. 42 Chromatogram of the amino acid hydrolysis of (4-Aza)Trp-anxA5 (Peak 1: (4-Aza)Trp; peak 2: Trp). The analysis revealed a substitution of the single Trp residue of anxA5 by (4-Aza)Trp of over 90%.

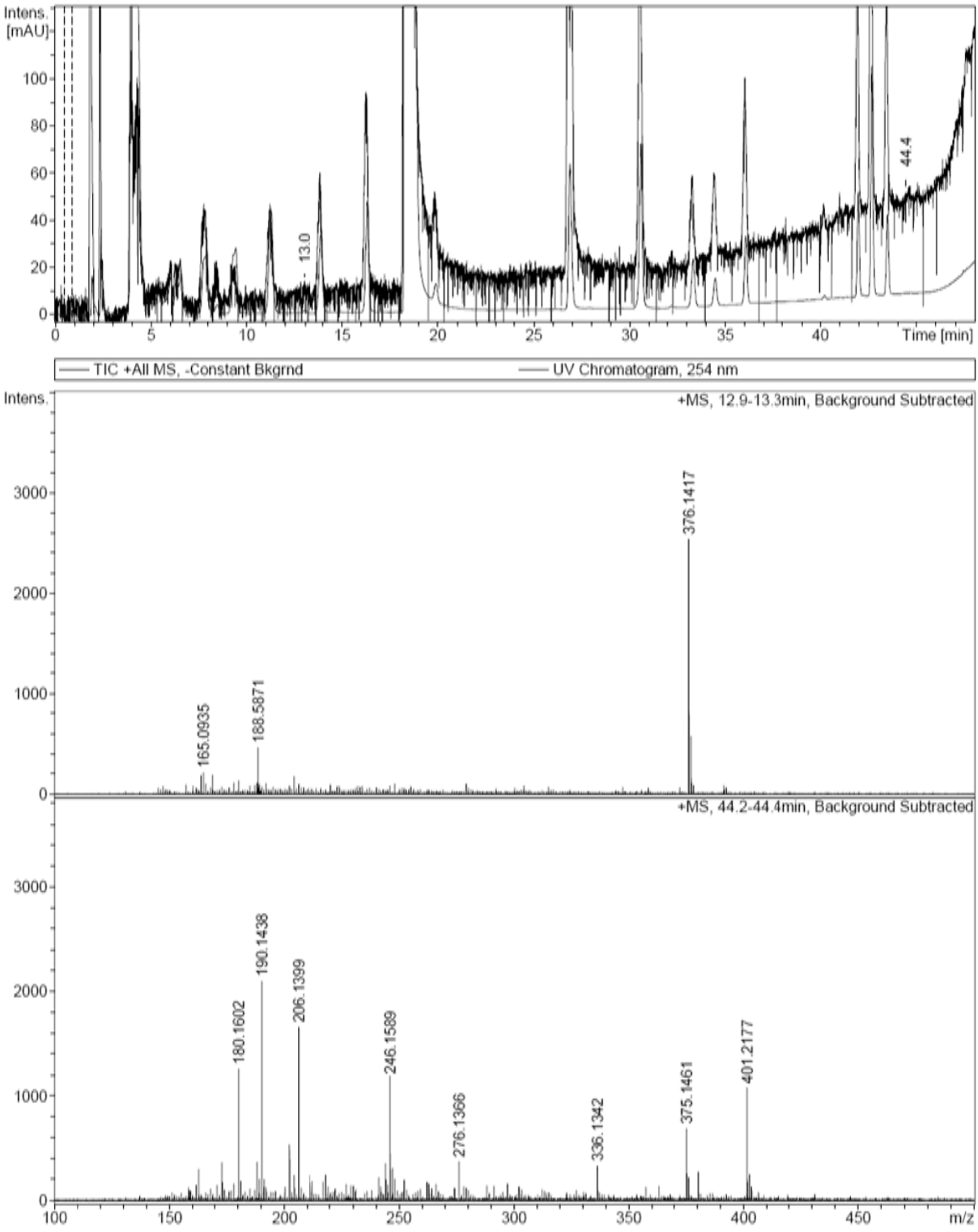


Fig. 43 Mass spectrometric analysis of hydrolyzed (4-Aza)Trp-anxA5 with the masses of derivatized (4-Aza)Trp (376.14 Da) and Trp (375.15 Da).

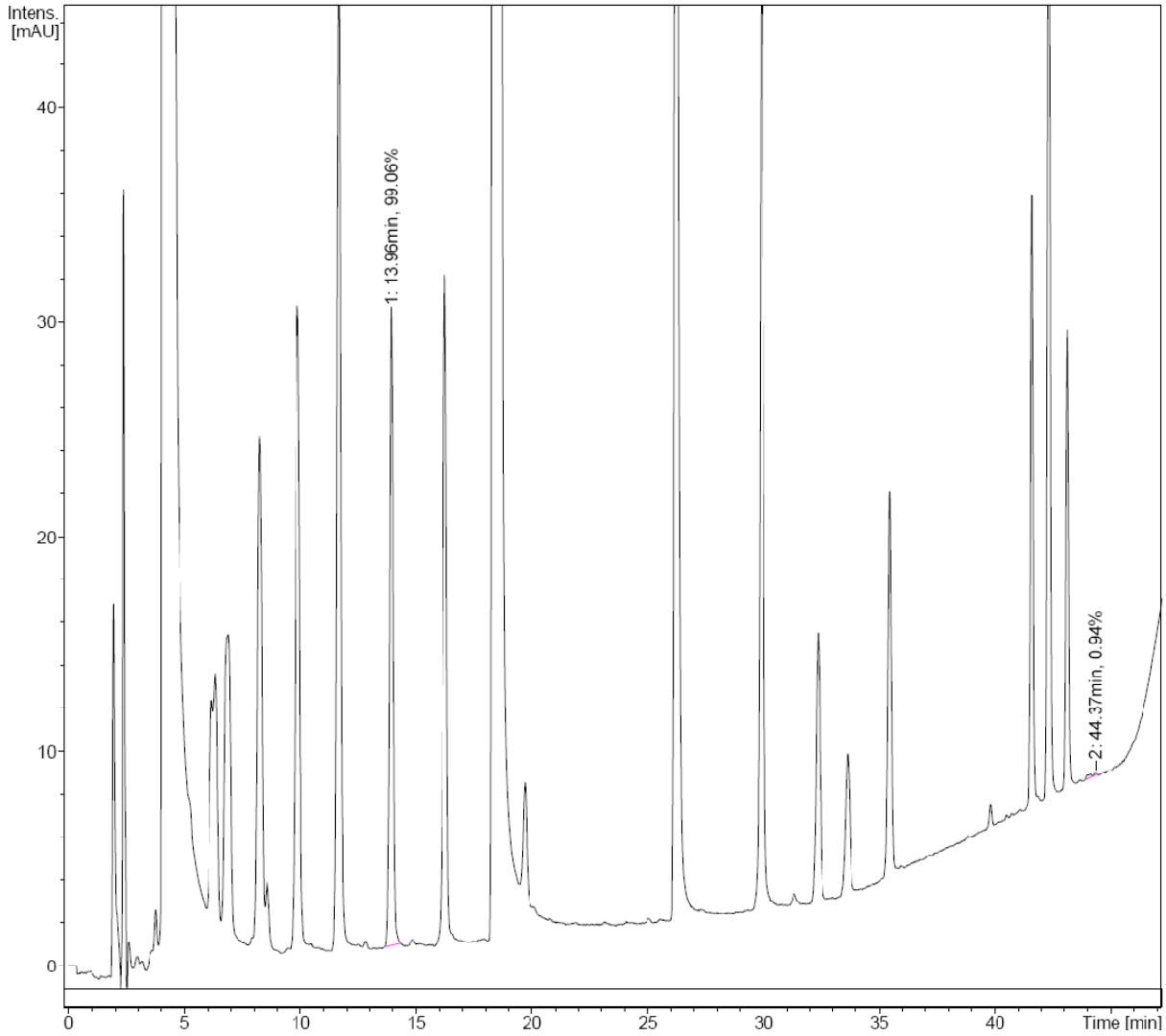


Fig. 44 Chromatogram of the amino acid hydrolysis of (5-Aza)Trp-anxA5 (Peak 1: Threonine and (5-Aza)Trp; peak 2: Trp). The analysis revealed a substitution of the single Trp residue of anxA5 by (5-Aza)Trp of about 60%, by comparison of the Trp and threonine peaks of (5-Aza)Trp-anx5 and of the parent protein (see Fig. 40).

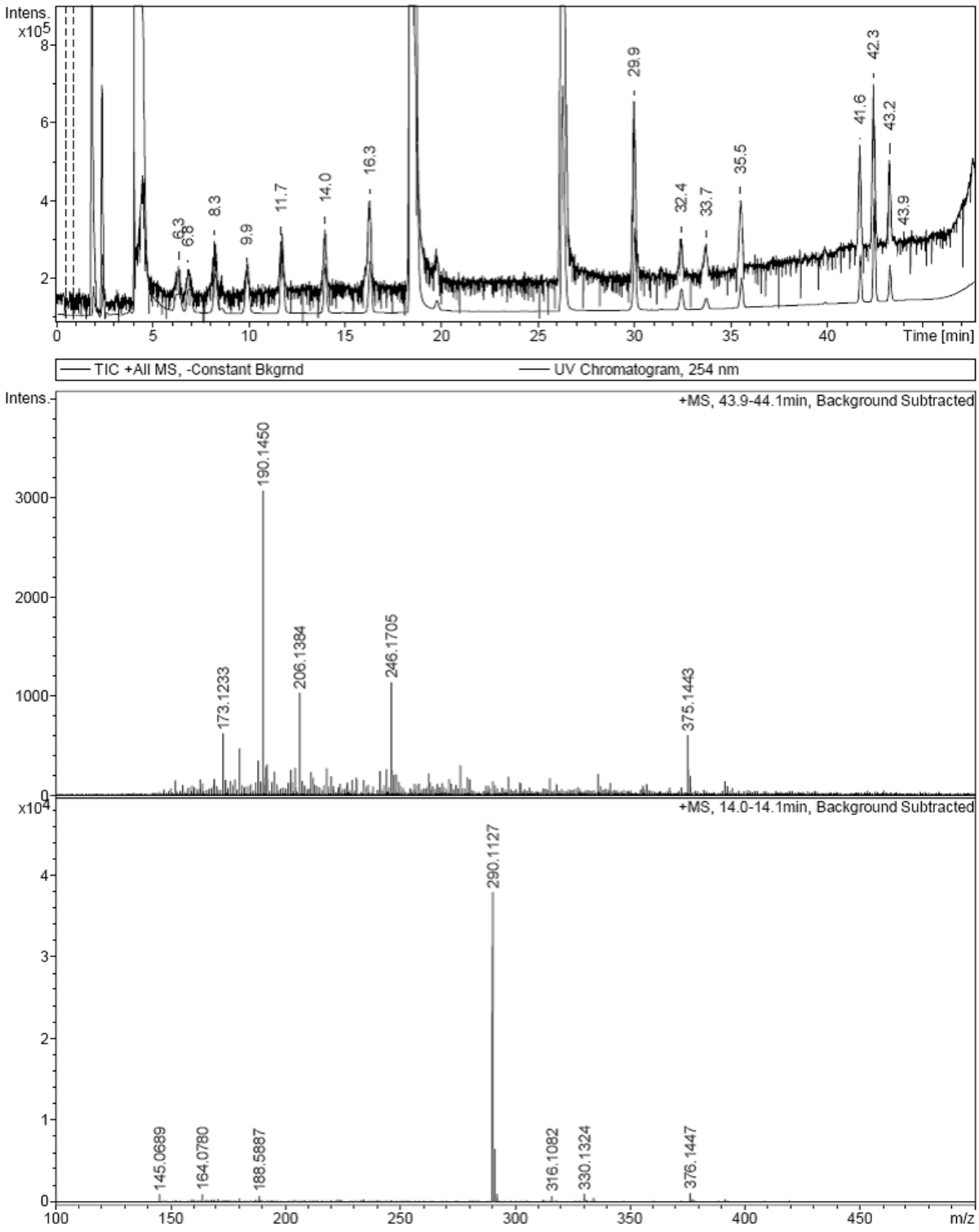


Fig. 45 Mass spectrometric analysis of hydrolyzed (5-Aza)Trp-anxA5 with the masses of derivatized (5-Aza)Trp (376.14 Da), threonine (290.11 Da), and Trp (375.14 Da).

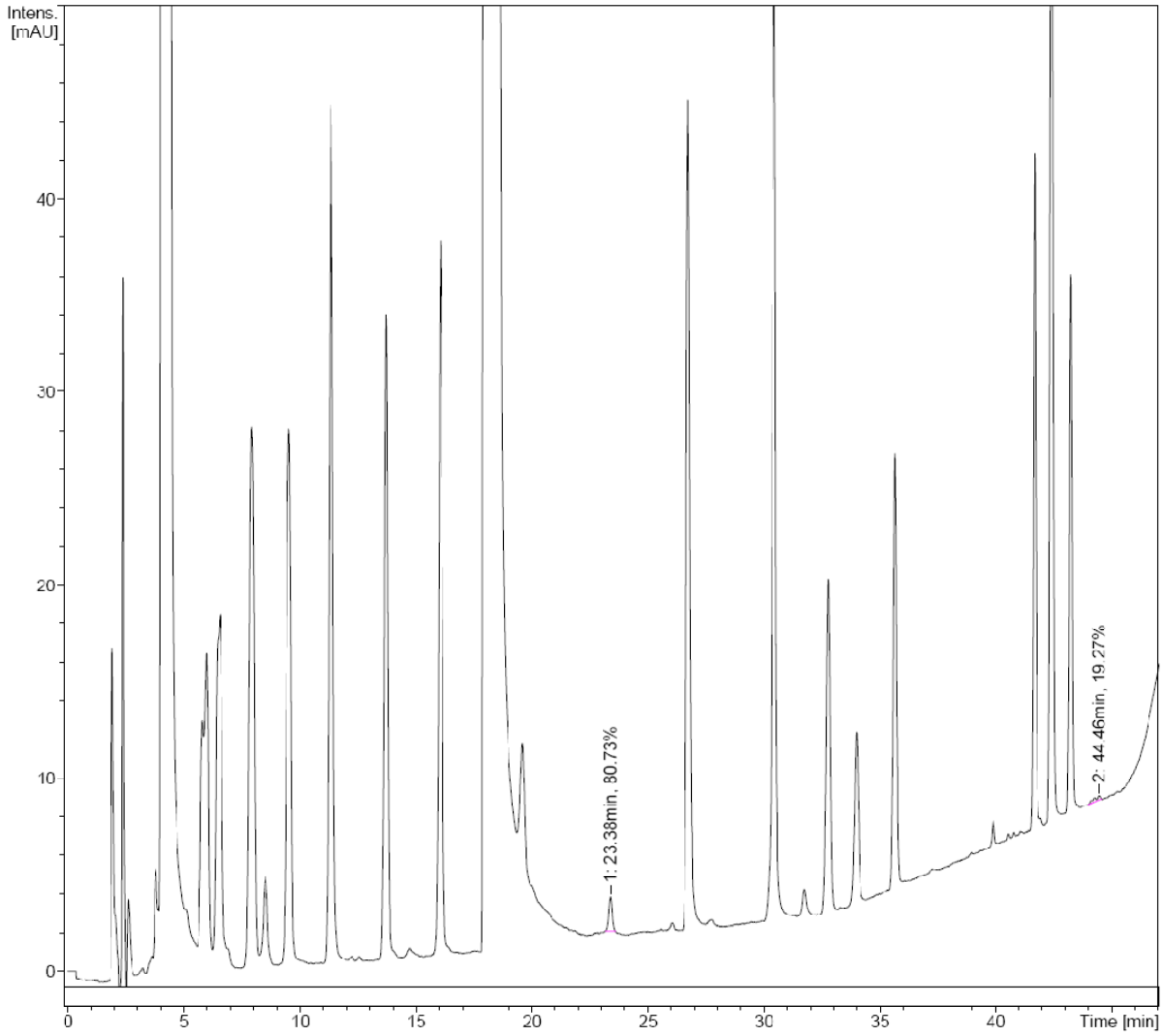


Fig. 46 Chromatogram of the amino acid hydrolysis of (7-Aza)Trp-anxA5 (Peak 1: (7-Aza)Trp; peak 2: Trp). The analysis revealed a substitution of the single Trp residue of anxA5 for (7-Aza)Trp of about 80%.

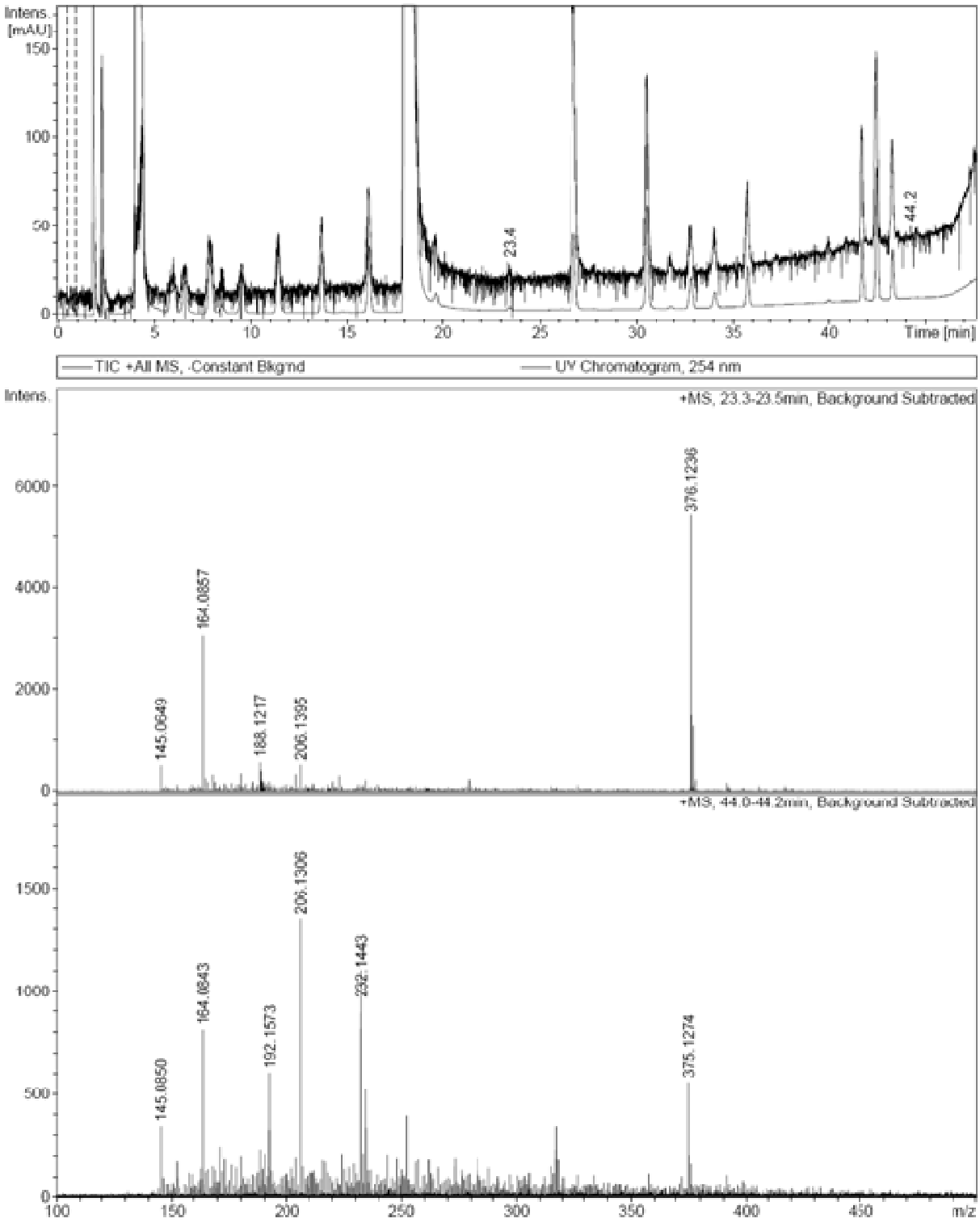


Fig. 47 Mass spectrometric analysis of hydrolyzed (7-Aza)Trp-anxA5 with the masses of derivatized (7-Aza)Trp (376.12 Da) and Trp (375.13 Da).

11.3 Results of *N*-terminal sequencing

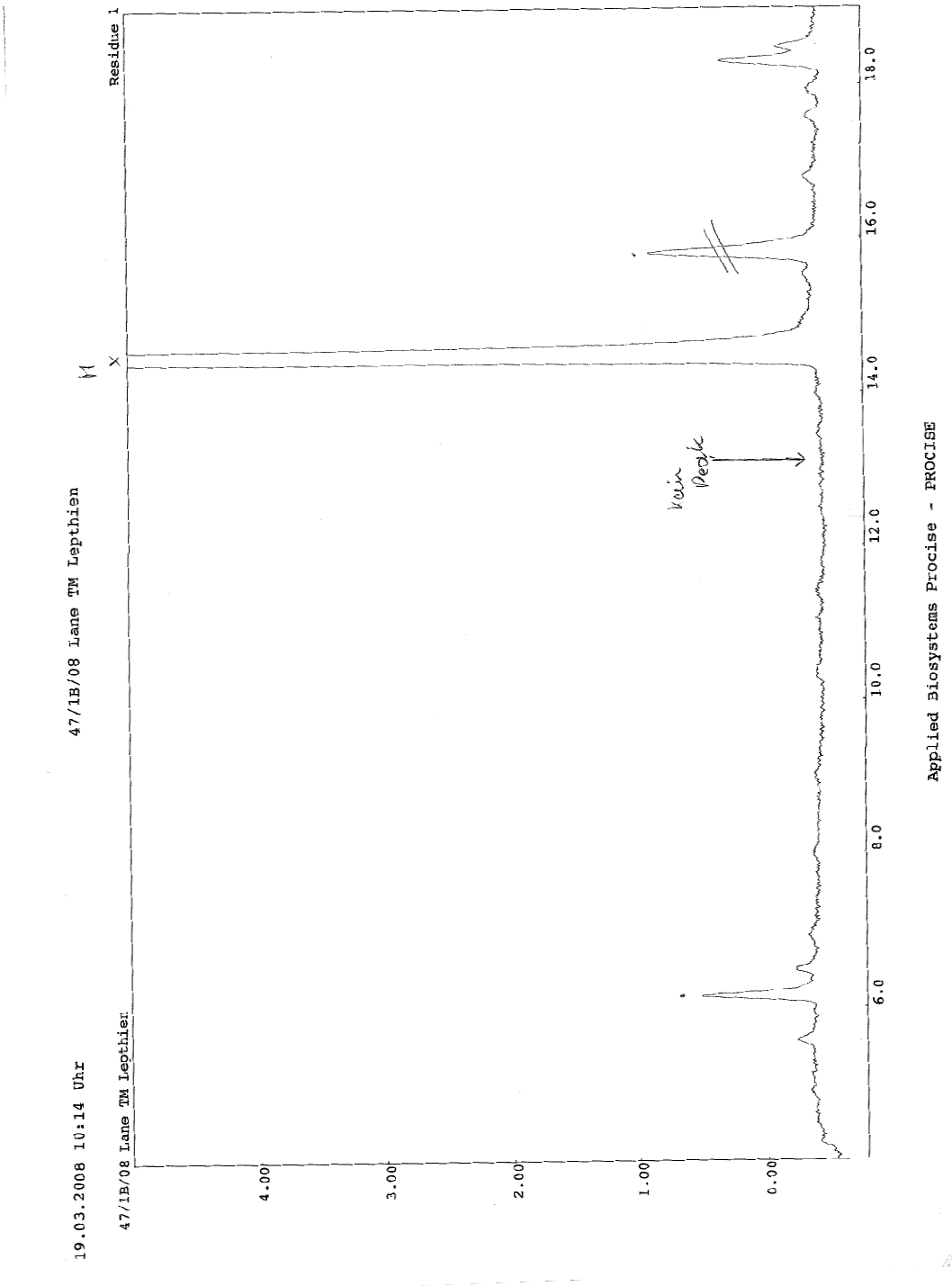


Fig. 48 *Nt*-sequencing of Met-Trp-ψ-b*. The *N*-terminal amino acid Met gives a signal after ~14 min.

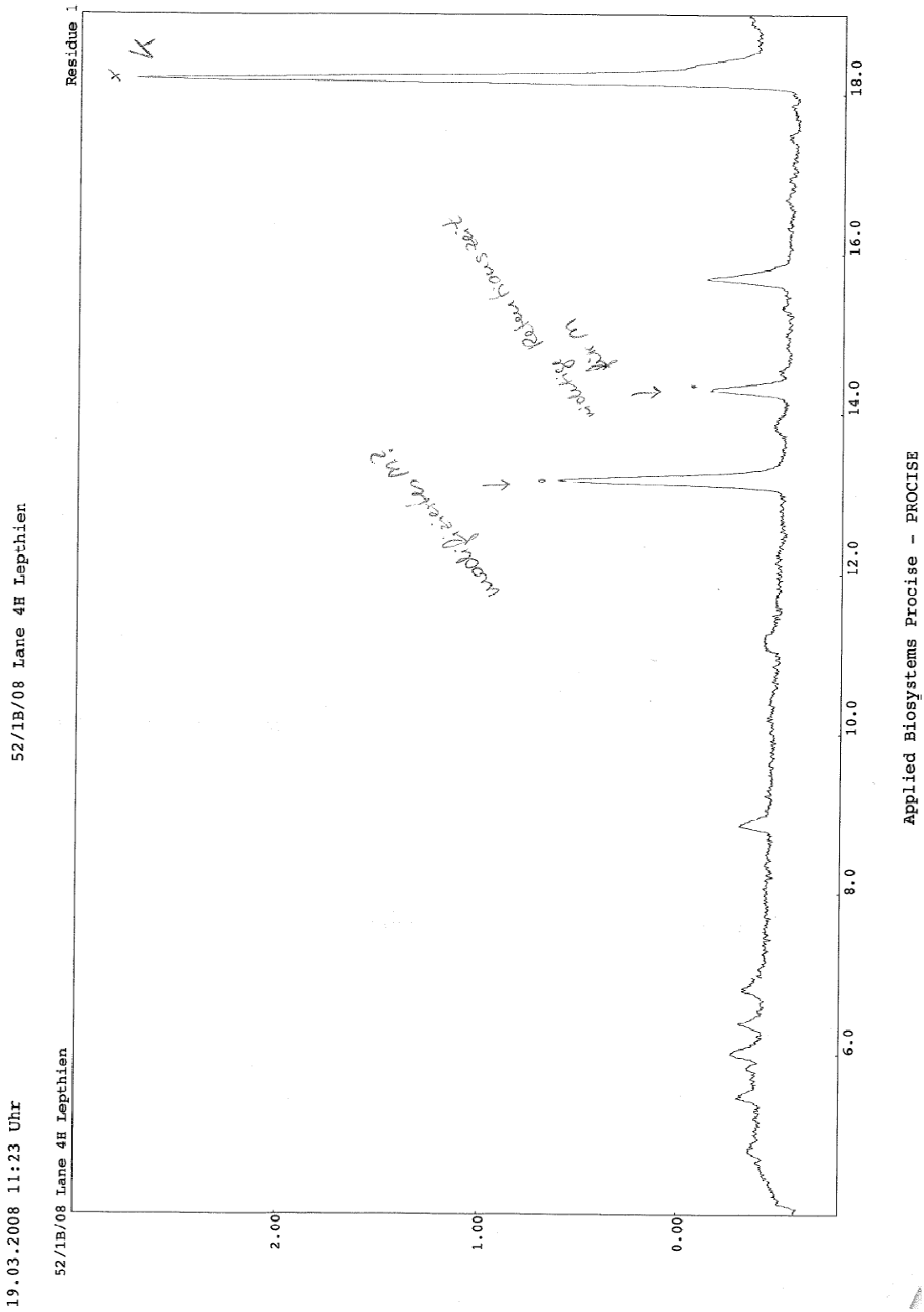


Fig. 49 ***Nt*-sequencing of Hpg-(4-Aza)Trp-ψ-b*.** The *N*-terminal amino acid after incorporation should be Hpg, which was identified to give a signal at ~13 min. In proteins, which were not substituted by Hpg, Met is still the ultimate amino acid and gives a signal after ~14 min (Fig. 48).

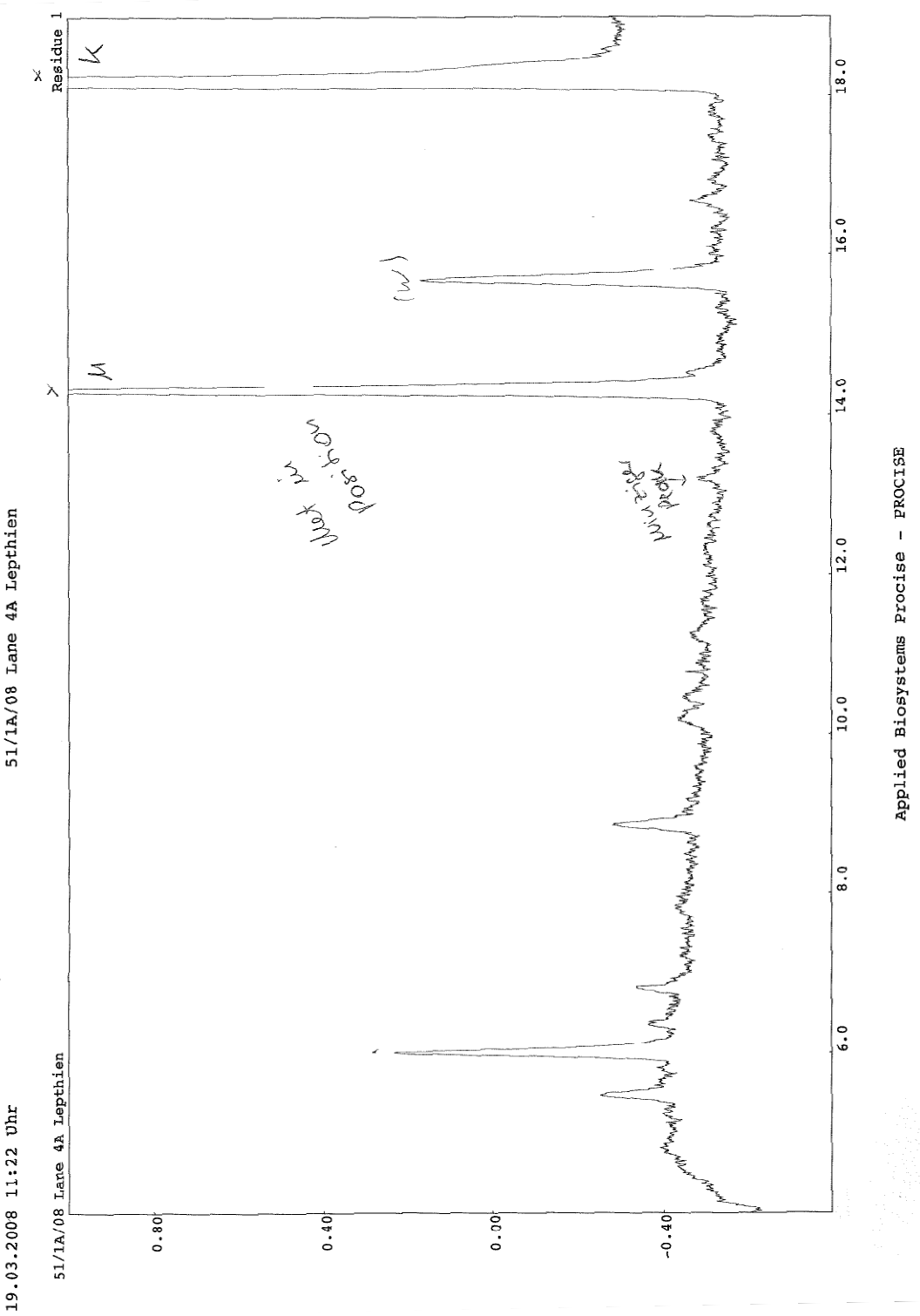


Fig. 50 *Nt*-sequencing of Aha-(4-Aza)Trp-ψ-b*. The *N*-terminal amino acid after incorporation should be Aha, which was identified to give a slight signal at ~13 min. In proteins, which were not substituted by Aha, Met is still the ultimate amino acid and gives a signal after ~14 min (Fig. 48).