Global Replacement of Tryptophan with Aminotryptophans Generates Non-Invasive Protein-Based Optical pH Sensors

Nediljko Budisa,* Marina Rubini, Jae H. Bae, Elisabeth Weyher, Waltraud Wenger, Ralph Golbik, Robert Huber, and Luis Moroder

Among the gene-encoded natural amino acids tryptophan occurs in proteins in only a low abundance (1.2 %),[1] but it occupies a unique position because of its ability to form numerous selective inter- and intramolecular interactions.[2] It also represents the main source of UV absorbance and of fluorescence in proteins[3] and has, therefore, recently been recognised as a useful intrinsic probe for structural and functional studies of proteins as well as an attractive target for classical protein engineering and design. More recently, the previously limiting factors of classical protein engineering procedures that rely solely on the standard set of canonical amino acids have been overcome by methods that allow for expansion of the amino acid repertoire (for example, see ref. [4]). In this context, our approach for the in vivo incorporation of noncanonical amino acids into proteins is based on the traditional use of auxotrophic *E. coli* host strains[5] that are forced to undergo selective pressure incorporation (SPI).[6] This method exploits the lack of absolute substrate specificity of aminoacyl-tRNA synthetases as the crucial enzymes in the interpretation of the genetic code,[7] a process that leads to activation and transfer of a variety of structurally and chemically similar substrate analogues onto cognate tRNAs under defined conditions (see, for example, ref. [8]). The main principle of this engineering approach is the reassignment of the coding triplets, that is, differences in the interpretation of the genetic code under the experimentally imposed selective pressure. For example, the repertoire for the tryptophan coding triplet (UGG) has been expanded with novel translationally active chromophores such as 4-fluorotryptophan, 5-fluorotryptophan, 6-fluorotryptophan,[9] 7-azatryptophan[10] and 5-hydroxytryptophan[11] or even with tryptophan-related amino acids such as β-(thienopyrrolyl)alanines[12] or β-selenolo[3,2-*b*]pyrrolyl-alanine.[13]

The main motivation for attempting global replacement of tryptophan residues in proteins with these analogues derived from earlier reports on the spectroscopic properties of substituted indoles.[14,15] In fact, it is known that substitution of indole with electron-donating groups leads to intramolecular charge transfer being extremely sensitive to pH changes.[16] For example, amino substituents on aromatic ring systems are less basic in the excited than in the ground state.[17] Thus, the protonation level in the excited states affects the resonant integration of their free electrons in conjugated ring systems which causes charge migration and is seen as blue- or red-shifted spectroscopic bands.

The amino acids **2**–**7** (Figure 1) were synthesised enzymatically from L-serine using commercially available amino- and hydroxyindoles as substrates for tryptophan synthase[18] and characterized spectroscopically[19] (the synthesis of the Trp analogues will be reported elsewhere).

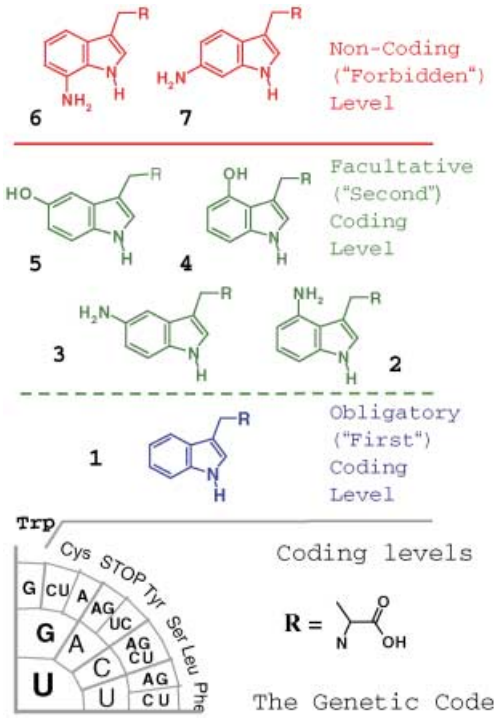

Figure 1. Genetic code engineering by expansion of the coding capacity of the tryptophan triplet UGG under the experimentally imposed selective pressure. The universal genetic code assigns UGG as the coding triplet for L-tryptophan (**1**), shown here as the obligatory or first coding level. By using the SPI method the entry of amino acids **2**–**5** into genetic code is allowed, thus creating a second or facultative coding level. Translationally inactive amino acids **6** and **7** can be regarded as prevented from entering into the genetic code (noncoding level) at this level of the development of SPI methodology.

The pseudo-wild-type variant C40A/C82A/P27A of barstar (**1**-b*) was chosen as a model protein for the expression experiments as it contains three tryptophan residues, namely, the fully or partially solvent-exposed Trp 38 and Trp 44 as well as Trp 53, which is completely buried within the protein hydrophobic core (Figure 2). The latter residue is known to be essential for the structural integrity of the protein, since it cannot be replaced by any of the other 19 canonical building blocks by routine site-directed DNA mutagenesis.[20] The three Trp residues of barstar (**1**-b*) were globally replaced with the analogues **2**–**5** by the SPI method using the Trp-auxotropic *E. coli* strain ATCC 49980 as the expression host.[21] Variant proteins were produced in yields comparable to those of the **1**-b* protein (10–30 mg L^{-1})[22] and were stable

[*] Dr. N. Budisa, M. Rubini, J. H. Bae, E. Weyher, W. Wenger, Prof. Dr. R. Huber, Prof. Dr. L. Moroder
Max-Planck-Institut für Biochemie
Am Klopferspitz 18A, 82152 Martinsried (Germany)
Fax: (+49) 89-8578-3516
E-mail: budisa@biochem.mpg.de
Dr. R. Golbik
Martin-Luther-Universität Halle-Wittenberg
06120 Halle/Saale (Germany)

Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

  0044-8249/02/4121-4066 $ 20.00+.50/0

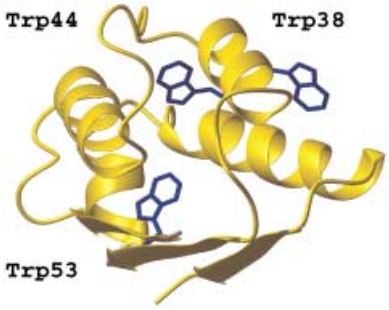

Figure 2. The structure of small ribonuclease inhibitor barstar,[23] with the tryptophan positions marked, represented as a ribbon plot (derived from the program MOLSCRIPT).[24]

in buffer solutions, with the exception of **4**-b*, which exhibits a strong tendency for aggregation. Unexpectedly, incorporation of the analogues **6** and **7** failed, most probably because of their poor recognition by the protein translation machinery.

Absorbance and fluorescence profiles of the protein variants were used as qualitative analytical criterion to monitor successful incorporation. For example, the hydroxy group at position 5 of indole leads to significant enhancement of the ^1L_b transition band[15] and thus to a pronounced spectral shoulder at 310 nm ($\varepsilon_{310} = 8180 \pm 395\,M^{-1}cm^{-1}$) in **5**-b* (see Supporting Information, Table 1). In contrast, the absorption maximum of **4**[19] is similar to that of the tyrosine whose phenolic OH group in the ground state exhibits a pK_a value of about 10.[3] Therefore, no shift in the UV spectrum of **4**-b* was observed in the examined pH range. Since the difference in molecular mass between **1** and **2**, **1** and **3**, (both 15 Da), **1** and **4**, and **1** and **5** (both 16 Da) is sufficiently large to be determined experimentally, mass spectrometry was used for quantitative analysis. While quantitative incorporation of both tryptophan analogues **2** and **3** in barstar was readily achieved (calcd: $M_r = 10298.2$; found: 10299 ± 2.0), the replacement level with the tryptophan analogues **4** and **5** never exceeded 80%. Although the parent protein could not be detected and globally substituted variants were the dominant forms (10300 ± 2.4 Da), contaminants with one 4-hydroxy- or 5-hydroxytryptophan (10269 ± 4.1 Da) and two 4-hydroxy- or 5-hydroxytryptophan residues (10285 ± 2.0 Da) were always present.

As expected, the UV properties of the tryptophan analogues **2**–**5** are fully reflected in the spectra of related barstar variants. In addition, the spectra of the **2**-b* and **3**-b* variants were found to be pH sensitive (Figure 3). The absorption profiles of **1**-b*, **2**-b*, and **3**-b* are very similar at pH 3.0 where protonation of the amino group leads to formation of a monocation. Conversely, at pH 7.0 and 9.0 the main absorption peak of **3**-b* is blue-shifted by almost 6 nm ($\lambda_{max} =$ 275 nm) while the spectral shoulder is markedly red-shifted by 15 nm ($\lambda_{max} = 310$ nm) for **3**-b* (Figure 3). A similar trend, although weaker in its intensity, is observed for **2**-b*. These changes may derive mainly from the increased basicity of the imino group in the indole moiety and, thus, from anion formation above pH 6. In fact, solution studies of 5-aminoindole revealed that a monoanion is formed by deprotonation of the indole imino group.[17]

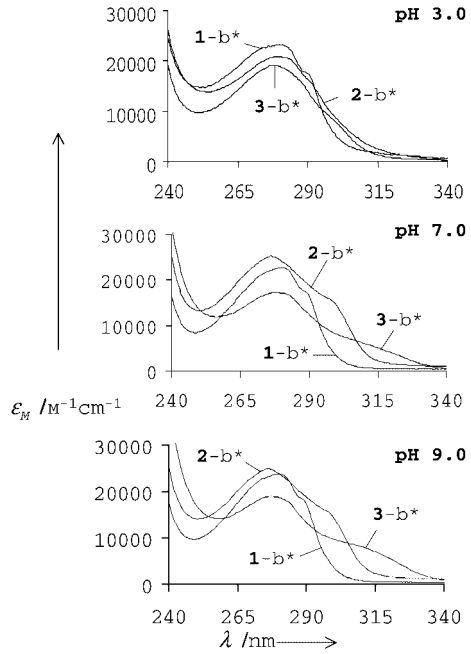

Figure 3. UV absorption spectra of **1**-b* and related variants **2**-b* and **3**-b* at different pH values. Note that under neutral and basic conditions a second prominent shoulder appears at 300 for **2**-b* and at 305 nm for **3**-b*.

Both the UV spectrum and the fluorescence of indole is composed of the two overlapping transitions ^1L_a and ^1L_b that are nearly orthogonal in polarization.[3] Similarly, the absorption profile of 5-aminoindole is a mixture of ^1L_a and ^1L_b bands, where the smaller absorption band (shoulder) at 285 nm derives from the transition of the ground to the ^1L_a state (long axis) and the main absorption, centred at 275 nm, derives from transition to the ^1L_b state (short axis). The amino group located along the longer axis (that is, at position 5) might perturb the ^1L_b state more than ^1L_a state. Therefore, the larger red shift observed for 5-aminoindole must be derived from the better delocalization of the nitrogen charge within the π cloud of the aromatic moiety than that of the oxygen atom of hydroxyindole. This is conceivable because the polarizability of the amino group is almost double that (1.44×10^{-24} mL per molecule) of the hydroxy group (0.733×10^{-24} mL per molecule),[25] and clearly indicates a higher propensity of nitrogen electrons to be spatially distorted since they are much less tightly bound than oxygen electrons. Therefore, the observation that hydroxytryptophan-b* variants **4**-b* and **5**-b* do not show significant changes in their absorbance properties in the pH range from 3.0 to 9.0 is not surprising. Correspondingly, the relative pH insensitivity of the **4**-b* and **5**-b* variants has to be attributed to the intrinsic spectroscopic properties of the hydroxytryptophan residues rather than to their protein environments—that is, the protein is only a carrier for the incorporated chromophores. Indeed, CD measurements revealed that secondary structures of the variant proteins are unchanged when compared to the **1**-b* protein variant in the pH range 6.0 to 9.0 where this protein is stable (unpublished data). Under acidic conditions (> pH 4) barstar is present in a molten globule form, while under stronger basic conditions (< pH 10) it is denatured.[20]

  0044-8249/02/4121-4067 $ 20.00+.50/0

It is also well known that the presence of electron-donating amino and hydroxy groups in different positions of the indole moiety leads to fluorescence spectra with two maxima at 350 and 520 nm.[14–16] However, it is also known that the spectral properties of indoles are strongly dependent on the solvent polarity;[3] for example, water acts as a proton donor in the S^o state, while the intramolecular charge transfer from the amino group to the indole is most effective in the S^1 excited state.[17] Therefore, the dramatic decrease in the relative fluorescence intensity of **2**-b* and **3**-b*, compared to the native protein (Figure 4), can only be explained by interactions such as the hydrogen bonding of water molecules to the aminoindole in the S^1 state. This situation leads to enhanced rates of radiationless processes and explains the absence of the second fluorescence band in the related protein variants in aqueous buffered solutions as determined by steady-state fluorescence measurements. Indeed, the second band might be restored if such proteins are dissolved in suitable organic solvents, thus providing dual-fluorescence protein-based pH sensors in nonpolar media, such as membrane environments.

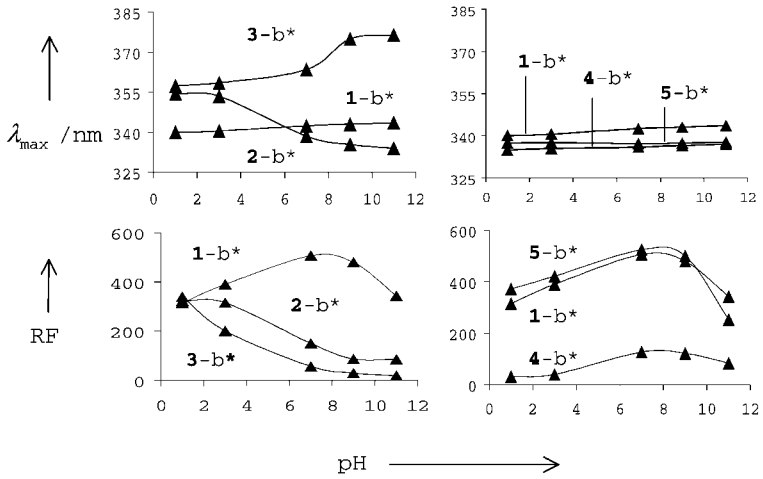

Figure 5. Plot of the variations in fluorescence emission maxima (λ_{max}) and relative intensities (RF) versus pH.

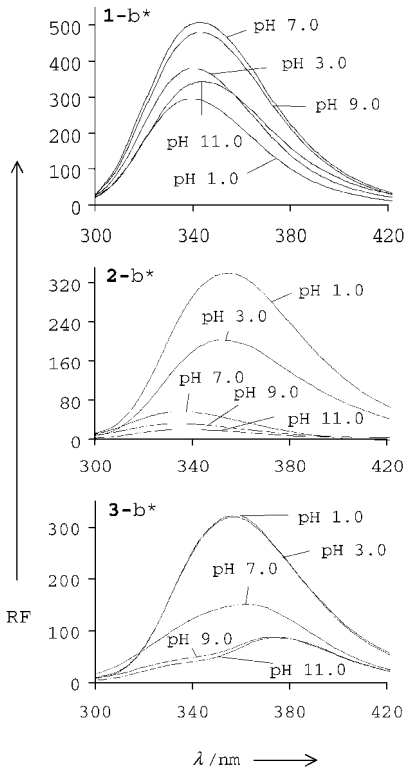

Figure 4. Effect of pH on the fluorescence emission profiles of **1**-b*, **2**-b*, and **3**-b*.

Changes in the relative fluorescence intensity of the variants **1**-b*, **4**-b*, and **5**-b* are gradual and monotonic in the pH range studied (Figure 5). Only **4**-b* was found to exhibit remarkably decreased quantum yields, probably because of its high tendency towards aggregation. The lower fluorescence intensity of these proteins under acidic conditions is the result of the transition of barstar to a molten-globule state below pH 4.0. Conversely, denaturing of barstar takes place under strongly basic conditions (pH > 10.0), and this accounts for the decreased fluorescence intensities in this pH range.[20] On the other hand, the fluorescence intensities of **2**-b* and **3**-b* in the pH range 4.0 to 9.0 results in titration curves with significantly decreasing values of fluorescence at increasing pH values which reach a plateau at pH > 9.0 (Figures 4 and 5). Observed changes in the intensity of these substituted barstars show completely different trends than the **1**-b*, **4**-b*, and **5**-b* variants upon pH titration.

The position of the emission maxima at pH 3.0 is similar for **1**-b*, **4**-b*, and **5**-b*, while for **2**-b* and **3**-b* it is red-shifted by 18 nm (Figures 4 and 5). The position of the emission maxima does not vary significantly at pH 7.0 and 9.0 for **1**-b*, **4**-b*, and **5**-b*, while two different trends were observed for **2**-b* and **3**-b*. A blue-shifted cooperative titration curve was obtained with a transition midpoint between pH 5 and 6 for the **2**-b* variant. Conversely, this cooperative transition of the emission maximum for **3**-b* is red-shifted with a transition midpoint at pH 8 (Figure 5). These large differences have to be assigned to the changes in the charge-density distributions along the indole aromatic ring which depend upon the nature, orientation, and position of the substituent in the parent indole molecule in a particular solvent. Taking this into account, the observed red shift for **3**-b* can be explained as follows: 5-aminoindole acts as a better proton donor than water in the S^1 state, that is, it is deprotonated to a greater extent in the excited state than is 4-aminoindole. Conversely, the blue shift measured for **2**-b* indicates there is a greater level of protonation in the 4-aminoindole in the excited state (that is, the solvent water molecules act as better proton donors than 4-aminoindole itself in the S^1 state). Clearly, the larger red shift is favored by the amino substituent being in position 5 of the indole (since it lays along the dipole transition moment) than in position 4, which is in full agreement with literature data on model compounds.[3,16,17] Furthermore, the observed transition arises from an equilibrium between the ionized forms in the ground and excited states.

  0044-8249/02/4121-4068 $ 20.00+.50/0

Such equilibrium is not only pH-dependent, but it also reflects the differences in the concentrations of corresponding species in the ground and excited states.

The pH sensitivity of the **2**-b* and **3**-b* variants, both in terms of fluorescence emission maximum and intensity, is the consequence of an intramolecular charge migration that originates solely from cation-to-anion transitions of the aminoindoles around pH 6. Thus, the conversion of barstar from a pH-insensitive to a pH-sensitive protein in terms of fluorescence is primarily the result of the intrinsic properties of the aminotryptophan analogues **2** and **3** integrated into the barstar structure. The multiple functions of tryptophan residues in proteins means it is indeed conceivable that proteins "tailored" with amino- and hydroxyindole or other indole-like side chains may represent useful non-invasive tools for in vivo or in vitro monitoring of protein folding as well as of protein–membrane, protein–protein, protein–ligand, and enzyme–substrate interactions. Protein variants in which tryptophan residues are globally replaced by related analogues with predefined spectral properties are readily accessible by the SPI methodology and could well offer an approach for the design of protein-based molecular sensors.

Received: March 18, 2002
Revised: June 5, 2002 [Z18926]

[1] a) M. O. Dayhoff, *Atlas of protein sequence and structure, Vol. 5*, National Biomedical Research Foundation, Washington, DC, **1972**; b) G. D. Rose, A. R. Geselowitz, G. J. Lesser, R. H. Lee, M. H. Zehfus, *Science* **1985**, *229*, 834 – 838; c) H. O. Villar, R. T. Koehler, *Biopolymers* **2000**, *53*, 226 – 232.

[2] a) D. A. Dougherty, *Science* **1996**, *271*, 163 – 168; b) U. Samanta, D. Pal, P. Chakrabarti, *Proteins* **2000**, *38*, 288 – 300; c) U. Samanta, P. Chakrabarti, *Protein Eng.* **2001**, *14*, 7 – 15.

[3] a) J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*, Kluwer Academic/Plenum Publisher, New York, **1999**; b) O. S. Andersen, D. V. Greathouse, L. L. Providence, M. D. Becker, R. E. Koeppe, *J. Am. Chem. Soc.* **1998**, *120*, 5142 – 5146; c) M. Cotton, C. L. Tian, D. D. Busath, R. B. Shirts, T. A. Cross, *Biochemistry* **1999**, *38*, 9185 – 9197.

[4] a) S. M. Hecht, B. L. Alford, Y. Kuroda, S. Kitano, *J. Biol. Chem.* **1978**, *253*, 4517 – 4520; b) J. D. Bain, C. G. Glabe, T. A. Dix, R. A. Chamberlin, R. S. Diala, *J. Am. Chem. Soc.* **1989**, *111*, 8013 – 8014; c) L. E. Steward, C. S. Collins, M. A. Gilmore, J. E. Carlson, J. B. Ross, R. A. Chamberlin, *J. Am. Chem. Soc.* **1997**, *119*, 6 – 11; d) C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, *Science* **1989**, *244*, 182 – 188; e) M. Sisido, *Prog. Polym. Sci.* **1992**, *17*, 699 – 764; f) T. Hohsaka, D. Kajihara, Y. Ashizuka, H. Murakami, M. Sisido, *J. Am. Chem. Soc.* **1999**, *121*, 34 – 40; g) N. Budisa, C. Minks, S. Alefelder, W. Wenger, F. Dong, L. Moroder, R. Huber, *FASEB J.* **1999**, *13*, 41 – 51; h) C. M. van Hest, K. L. Kiick, D. A. Tirrell, *J. Am. Chem. Soc.* **2000**, *122*, 1282 – 1288; i) D. A. Dougherty, *Curr. Opin. Chem. Biol.* **2000**, *4*, 645 – 652.

[5] a) G. Brawerman, M. Ycas, *Arch. Biochem. Biophys.* **1957**, *68*, 112 – 117; b) M. H. Richmond, *Bacteriol. Rev.* **1962**, *26*, 398 – 420; c) S. Schlessinger, *J. Biol. Chem.* **1968**, *243*, 3877 – 3883.

[6] a) C. Minks, S. Alefelder, R. Huber, L. Moroder, N. Budisa, *Tetrahedron* **2000**, *56*, 9431 – 9442; b) C. Renner, S. Alefelder, J. H. Bae, N. Budisa, R. Huber, L. Moroder, *Angew. Chem.* **2001**, *113*, 949 – 951; *Angew. Chem. Int. Ed.* **2001**, *40*, 923 – 925.

[7] N. Budisa, L. Moroder, R. Huber, *Cell. Mol. Life Sci.* **1999**, *55*, 1626 – 1635.

[8] a) J. B. A. Ross, A. G. Szabo, C. W. V. Hogue, *Methods Enzymol.* **1997**, *278*, 151 – 190; b) N. Budisa, B. Steipe, P. Demange, C. Eckerskorn, J. Kellermann, R. Huber, *Eur. J. Biochem.* **1995**, *230*, 788 – 796; c) W. Karnbrock, E. Weyher, N. Budisa, R. Huber, L. Moroder, *J. Am. Chem. Soc.* **1996**, *118*, 913 – 914; d) D. Besse, N. Budisa, W. Karnbrock, C. Minks, H. J. Musiol, S. Pegoraro, F. Siedler, E. Weyher, L. Moroder, *Biol. Chem.* **1997**, *378*, 211 – 218; e) N. Budisa, G. Pifat, *Croat. Chem. Acta* **1988**, *71*, 179 – 187; f) N. Budisa, R. Huber, R. Golbik, C. Minks, E. Weyher, L. Moroder, *Eur. J. Biochem.* **1998**, *53*, 1 – 9; g) N. Budisa, W. Karnbrock, S. Steinbacher, A. Humm, L. Prade, L. Moroder, R. Huber, *J. Mol. Biol.* **1997**, *270*, 616 – 623; h) C. Minks, R. Huber, L. Moroder, N. Budisa, *Anal. Biochem.* **2000**, *284*, 29 – 34.

[9] a) E. Pratt, C. Ho, *Biochemistry* **1975**, *14*, 3035 – 3040; b) P. M. Bronskill, J. T. Wong, *Biochem. J.* **1988**, *249*, 305 – 308; c) C. Y. Wong, M. R. Eftink, *Biochemistry* **1998**, *37*, 8938 – 8946; d) C. Minks, R. Huber, L. Moroder, N. Budisa, *Biochemistry* **1999**, *38*, 10649 – 10659.

[10] P. Soumillion, L. Jespers, J. Vervoort, J. Fastrez, *Protein Eng.* **1995**, *8*, 451 – 456.

[11] C. W. V. Hogue, I. Rasquinha, A. G. Szabo, J. P. Macmanus, *FEBS Lett.* **1992**, *310*, 269 – 272.

[12] N. Budisa, S. Alefelder, J. H. Bae, R. Golbik, C. Minks, R. Huber, L. Moroder, *Protein Sci.* **2001**, *10*, 1281 – 1292.

[13] J. H. Bae, S. Alefelder, J. T. Kaiser, R. Friedrich, L. Moroder, R. Huber, N. Budisa, *J. Mol. Biol.* **2001**, *309*, 925 – 936.

[14] S. Undenfriend, D. F. Bogdanski, H. Weisbach, *Science* **1955**, *122*, 972 – 973.

[15] T. Kishi, M. Tanaka, J. Tanaka, *Bull. Chem. Soc. Jpn.* **1977**, *50*, 1267 – 1271, and references therein.

[16] H. K. Sinha, S. K. Dogra, M. Krishnamurthy, *Bull. Chem. Soc. Jpn.* **1987**, *60*, 4401 – 4407, and references therein.

[17] R. Golbik, H. Neef, G. Hübner, S. König, B. Seliger, L. Meshalkina, G. A. Kochetov, A. Schellenberger, *Bioorg. Chem.* **1991**, *19*, 10 – 17, and references therein.

[18] E. Wilson Miles, H. Kawasaki, A. S. Ahmed, H. Morita, S. Nagata, *J. Biol. Chem.* **1989**, *264*, 6280 – 6287.

[19] Absorption maxima and molar extinction coefficients of the tryptophan analogues were determined at 25 °C. **2** at pH 3.0: $\lambda_{max} = 274$ nm, $\varepsilon_{274} = 6500\,M^{-1}cm^{-1}$; at pH 7.0: $\lambda_{max} = 272$ nm, $\varepsilon_{274} = 7800\,M^{-1}cm^{-1}$; at pH 9.0 the values are identical to those at pH 7.0. **3** at pH 3.0: $\lambda_{max} = 278$ nm, $\varepsilon_{278} = 5580\,M^{-1}cm^{-1}$; at pH 7.0: $\lambda_{max1} = 271$ nm, $\varepsilon_{271} = 4750\,M^{-1}cm^{-1}$, $\lambda_{max2} = 305$ nm, $\varepsilon_{305} = 2450\,M^{-1}cm^{-1}$; at pH 9.0: $\lambda_{max1} = 270$ nm, $\varepsilon_{270} = 5200\,M^{-1}cm^{-1}$, $\lambda_{max2} = 305$ nm, $\varepsilon_{305} = 3050\,M^{-1}cm^{-1}$. Absorbance properties as well as maximum of **1**, **4**, and **5** do not vary significantly with the change of pH in the range from 3 to 9. Average values are: **1** $\lambda_{max} = 280$ nm, $\varepsilon_{280} = 6500\,M^{-1}cm^{-1}$; **4** $\lambda_{max} = 262$ nm, $\varepsilon_{262} = 6130\,M^{-1}cm^{-1}$; **5** $\lambda_{max1} = 275$ nm, $\varepsilon_{275} = 5300\,M^{-1}cm^{-1}$, $\lambda_{max2} = 295$ nm, $\varepsilon_{295} = 4350\,M^{-1}cm^{-1}$.

[20] a) R. Swaminathan, U. Nath, J. B. Udgaonkar, N. Periasamy, C. Krishnamoorthy, *Biochemistry* **1996**, *35*, 9150 – 9157; b) R. Khurana, A. T. Hate, U. Nath, J. B. Udgaonkar, *Protein Sci.* **1995**, *4*, 1133 – 1144; c) B. R. Rami, J. B. Udgaonkar, *Biochemistry* **2002**, *41*, 1710 – 1716.

[21] The fermentation culture after overnight protein expression with **5** retained the same color as the barstar culture, while the cultures with **4** were colored like the substance itself (faint dark green). Similar effects were observed with **2** (faint orange) and **3** (dark orange).

[22] R. Golbik, G. Fischer, A. R. Fersht, *Protein Sci.* **1999**, *8*, 1505 – 1514.

[23] C. Martin, R. W. Hartley, Y. Mauguen, *FEBS Lett.* **1999**, *452*, 128 – 132.

[24] P. J. Kraulis, *J. Appl. Crystallogr.* **1991**, *24*, 946 – 950.

[25] S. K. Burley, G. A. Petsko, *Adv. Protein Chem.* **1988**, *39*, 125 – 186.

  0044-8249/02/4121-4069 $ 20.00+.50/0