Decomposition of Protein Tryptophan Fluorescence Spectra into Log-Normal Components. III. Correlation between Fluorescence and Microenvironment Parameters of Individual Tryptophan Residues

Yana K. Reshetnyak,* Yuly Koshevnik,† and Edward A. Burstein*
*Institute of Theoretical and Experimental Biophysics, Russia Academy of Sciences, Pushchino, Moscow Region, Russia 142290; and †MCI WorldCom, Inc., Richardson, Texas 75081 USA

ABSTRACT In our previous paper (Reshetnyak, Ya. K., and E. A. Burstein. 2001. *Biophys. J.* 81:1710–1734) we confirmed the existence of five statistically discrete classes of emitting tryptophan fluorophores in proteins. The differences in fluorescence properties of tryptophan residues of these five classes reflect differences in interactions of excited states of tryptophan fluorophores with their microenvironment in proteins. Here we present a system of describing physical and structural parameters of microenvironments of tryptophan residues based on analysis of atomic crystal structures of proteins. The application of multidimensional statistical methods of cluster and discriminant analyses for the set of microenvironment parameters of 137 tryptophan residues of 48 proteins with known three-dimensional structures allowed us to 1) demonstrate the discrete nature of ensembles of structural parameters of tryptophan residues in proteins; 2) assign spectral components obtained after decomposition of tryptophan fluorescence spectra to individual tryptophan residues; 3) find a correlation between spectroscopic and physico-structural features of the microenvironment; and 4) reveal differences in structural and physical parameters of the microenvironment of tryptophan residues belonging to various spectral classes.

INTRODUCTION

In the two previous articles of this series we have presented algorithms of decomposition composite fluorescence spectra of tryptophan residues in proteins into log-normal components (Burstein et al., 2001) and results of such decomposition of spectra of ~100 various proteins and their conformers induced by denaturants, complexing with ions or organic ligands, varying pH, etc. (Reshetnyak and Burstein, 2001). An obtained database of spectral maximum positions and external quencher accessibilities for >300 components supposedly belonging to individual tryptophan residues allowed us to hope that statistical analysis of the database may improve our understanding of the nature of wide variation of fluorescence properties of tryptophan residues in proteins.

Studies on intrinsic protein fluorescence began in the late 1950s (Shore and Pardee, 1956; Duggan and Udenfriend, 1956; Vladimirov and Konev, 1957; Konev, 1957, 1959; Steel and Szent Györgyi, 1958; Teale and Weber, 1958; Vladimirov, 1959; Teale, 1960; Vladimirov and Burstein, 1960; Weber, 1960; Burstein, 1961). Over the past 40 years significant progress in the development of spectroscopic technique has been achieved and a great amount of experimental and theoretical work on model compounds and proteins has been carried out. However, there remains a gap between the knowledge of physical principles of spectroscopy of molecular interactions and relaxation in solutions and the level at which they are applied in the interpretation of protein emission spectroscopy data.

The wide variation in the fluorescence properties of tryptophan residues in proteins reflects differences in processes in their excited state, which have to depend on features of fluorophore microenvironments in macromolecules. That is why one can observe the rousing interest in analyzing tryptophan location and its surrounding in proteins revealed by x-ray crystallography. More than 20 years ago the first attempts were done to reveal a correlation between emission spectral parameters and structural characteristics of tryptophan fluorophore environment based on atomic structures of several proteins (Pelley and Horowitz, 1976; Brown et al., 1977; Rousslang et al., 1979). The rapid progress in x-ray crystallography and NMR methods led to an increase in the amount of work on analyzing the measured fluorescence properties of individual proteins in relation to the structural features (for example, Turoverov et al., 1984, 1985; Desie et al., 1986; Dolashka et al., 1992; Bhaskaran et al., 1996; Mely et al., 1997; Reshetnyak and Burstein, 1997a, b; Kuznetsova and Turoverov, 1998; Kuznetsova et al., 1999; Orlov et al., 1999; Turoverov, 1999). The number of databases and algorithms were proposed for calculating different structural and physical characteristics of individual residues in proteins (Gray et al., 1996; Hogue et al., 1996; Islamov et al., 1997; Laskowski et al., 1997). However, still there is a very scarce basis for wide generalizations relating to this problem.

To reveal a correlation between spectroscopic and structural features it was necessary to correctly associate fluorescence characteristics to individual fluorophores in a protein. Therefore, first searches of such correlation have been carried out for a rather limited number of single-tryptophan-

Received for publication 12 January 2001 and in final form 7 June 2001.
Address reprint requests to (present address) Dr. Yana K. Reshetnyak, Department of Molecular Biology and Immunology, Institute for Cancer Research, University of North Texas Health Science Center, 3500 Camp Bowie Blvd., Fort Worth, TX 76107. Tel.: 817-735-5417; Fax: 817-735-2133; E-mail: yreshetn@hsc.unt.edu.

0006-3495/01/09/1735/24 $2.00

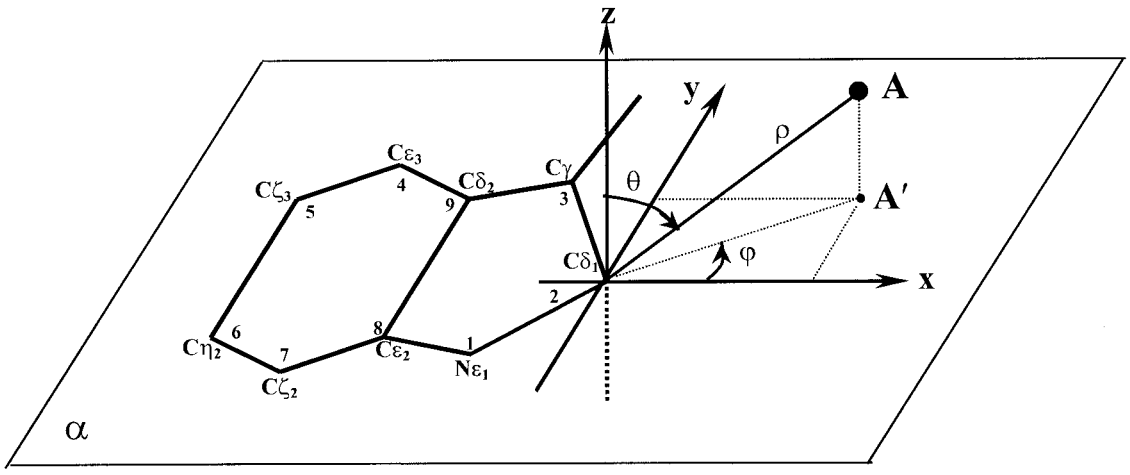

FIGURE 1 Spherical systems of coordinates (ρ, distance; φ, azimuth; and θ, elevation) centered at each of nine atoms of indole moiety by turns.

containing proteins. To extend the possibility of such analysis on multi-tryptophan-containing proteins we have developed algorithms of stable decomposition of composite steady-state tryptophan fluorescence spectra of proteins into log-normal components (Abornev and Burstein, 1992; Burstein and Emelyanenko, 1996; Burstein et al., 2001). Then, we decomposed tryptophan fluorescence spectra of ~100 proteins, some in various structural states, and created the database of obtained steady-state log-normal components of protein spectra of tryptophan fluorescence (Reshetnyak and Burstein, 2001). The statistical analysis of the distributions' spectral maximum positions of the components confirmed the statistical discreteness of at least five main classes of emitting tryptophan fluorophores in proteins, first proposed earlier (Burstein et al., 1973; Burstein, 1977a, 1983). It was natural to assume that such a situation might be a result of the existence of a limited number of combinations of specific and/or universal interactions of individual fluorophores with their environment, mainly during the lifetime of the fluorescent excited state. It means that tryptophan residues from different spectroscopic classes have to have different features of fluorophore microenvironments in proteins. However, the large increase in the number of proteins with known highly resolved x-ray structures gave us an opportunity to assign the log-normal components to individual fluorophores and, then, carry out a detailed analysis of the environments of tryptophan residues. In the present work we succeeded in performing such an assignment for 137 tryptophan residues of a representative set of ~50 proteins.

The goals of the present investigation were to 1) assign spectral fluorescence components to individual tryptophan residues of proteins with known atomic structure; 2) analyze physical and structural characteristics of microenvironments of individual indole fluorophores in proteins and search for the correlation between fluorescent and environment parameters using multidimensional statistical methods of cluster and discriminant analyses; and 3) reveal differences in structural and physical parameters of microenvironments of tryptophan residues belonging to various spectral classes.

STRUCTURAL AND PHYSICAL PARAMETERS OF MICROENVIRONMENTS OF TRYPTOPHAN FLUOROPHORES IN PROTEINS

The set of structural and physical parameters of microenvironments of individual tryptophan residues was estimated for 48 well-resolved structures of proteins with known characteristics of log-normal components of their fluorescence spectra (Table 2, "Mean Best Fit" in Reshetnyak and Burstein, 2001). The parameters were calculated from atomic coordinates of proteins from the Protein Data Bank (PDB; Bernstein et al., 1977) using the original algorithm realized in Visual Basic. In cases when there are several structural PDB entries for a protein or the PDB entry for a protein contains atomic coordinates for several polypeptide chains, the estimated microenvironment parameters were averaged for every equivalent tryptophan residue.

Spherical coordinate systems for indole ring atoms

Because each atom of indole fluorophore may contribute differently in specific interactions with the environment, we estimated environmental characteristics for every indolic atom. The parameters of nine indole atoms of tryptophan residues were then averaged or summed to characterize the environment of the given fluorophore as a whole. To describe the surrounding of each of nine atoms of an indole ring independently, we introduced spherical systems of coordinates (see Fig. 1, where ρ is the distance, φ is the azimuth, and θ is the elevation) centered at each atom. In the

microenvironment of fluorophores we included all neighboring atoms of protein and structure-defined solvent located in near and far layers from 0 to 5.5 Å and from 5.5 to 7.5 Å from indole atoms, respectively. The program output the complete lists of distances (ρ) and orientations (φ and θ) of "polar" (nitrogen, oxygen, and sulfur) and carbon atoms located in each layer. Then, the numbers of "polar" and carbon atoms in two layers were estimated separately for a fluorophore as a whole.

Eventual hydrogen-bonding partners

The most probable mechanism of the exciplex formation between excited fluorophore and surrounding atoms is hydrogen-bonding (Lumry and Hershberger, 1978; Hershberger et al., 1981). We revealed eventual hydrogen bond donors and acceptors from neighboring polar groups around indole atoms. It is now generally accepted that hydrogen bonds are strictly directional in nature (Legon and Millen, 1987). According to the geometric criteria of hydrogen bond, $\cos(\theta)$ must be near 1 for possible donors (main-chain nitrogen atoms; Sγ of Cys; Nϵ2 and Nδ1 of His; Nζ of Lys; Nδ2 of Asn; Nϵ2 of Gln; Nϵ, Nη1, and Nη2 of Arg; Oγ of Ser; Oγ1 of Thr; Oη of Tyr, Sγ of Cys); and $\cos(\theta)$ and $\cos(\varphi)$ must be near 0 and 1, respectively, for possible acceptors (main-chain carbonyl oxygen; Oδ1 and Oδ2 of Asp and Asn; Sγ of Cys; Oϵ1 and Oϵ2 of Glu and Gln; Nδ1 of His; Sδ of Met; Oγ of Ser; Oγ of Thr; Oη of Tyr) (McDonald and Thronton, 1994). We considered all atoms listed above as potential partners for hydrogen bonding if they were located within a cone with angles differing by $<\sim 20°$ from the ideal geometry of H-bonds, i.e., $\cos(\theta) > 0.9$ for possible donors in near (*Ndon1*) and far (*Ndon2*) layers, and $\cos(\theta) < 0.35$ and $\cos(\varphi) > 0.9$ for possible acceptors in near (*Nac1*) and far (*Nac2*) layers.

Solvent accessibility and environment packing density

It is well known that the dipole reorientational relaxation in polar environment during excitation lifetime (the universal interactions) may play an essential role in the shift of fluorescence spectra to longer wavelengths (Mataga et al., 1955, 1956; Lippert, 1957; Bilot and Kawski, 1962; Liptay, 1965; Bakhshiev, 1972; Vincent et al., 1995, 1997; Toptygin and Brand, 2000). Such a relaxation-induced spectral shift may be effective for tryptophan fluorescence due to a large increase in indole dipole moment in excited ^1L_a compared with the ground state (by up to >10 D) (Konev, 1967; Burstein, 1976; Lakowicz, 1983; Muiño and Callis, 1994; Pierce and Boxer, 1995; Callis, 1997). Recent quantum-mechanical studies showed that much electron density is lost from Nϵ1 and Cγ atoms and deposited at Cϵ3, Cζ2, Cδ2 atoms of indole ring during excitation in main fluorescent ^1L_a state (Callis, 1997). Therefore, the emission band of tryptophan fluorophores that are accessible to bulk water, possessing the high density of large and fast-relaxing dipoles, is expected to be shifted maximally to longer wavelengths (Burstein, 1976). We estimated relative accessibility of each atom of indole moieties (especially Nϵ1 (*Acc1*) and Cζ2 (*Acc7*) atoms) and accessibility of the indole ring as a whole (*Acc*) using an algorithm based on the Lee and Richards approach (1971). The relative accessibility was expressed as the percent ratio of the accessible area of an atom in a protein for the 1.4 Å spherical probe to that in free tryptophan. The solvent environment surrounding a protein is generally divided into two types: bulk (free) water that is fluid and not bound to the protein, and water that is either partially or strongly bound (Edsall and McKenzie, 1983; Otting et al., 1991; Levitt and Park, 1993). Our program, therefore, presented the accessibility of the indole ring to the solvent in absence (*Acc*, *Acc1*, *Acc7*) and presence (*Accw*, *Accw1*, *Accw7*) of bound water molecules, which are included in atomic coordinates in PDB files.

The packing density

The packing density, i.e., the number of neighbor atoms within the layers up to 5.5 Å (*Den1*) or up to 7.5 Å (*Den2*) around indole ring, reflects the degree of burying of fluorophore into protein matrix and/or the presence of hollow crevices in the structure. In compact structure the packing density may serve as an inverse measure of the accessibility. An example of a hollow cavity that surrounds the completely buried Trp-48 in azurin was proposed by Turoverov et al. (1984, 1985). The packing density of Trp-48 is rather low compared with other buried tryptophan residues.

The measure of relative polarity of the microenvironment

The relative polarity of the fluorophore microenvironment in two surrounding layers we expressed as:

$$A1 = [S1 \cdot (100 + Acc)]/100 = S1 + (Acc \cdot S1)/100$$

$$A2 = [S2 \cdot (100 + Acc)]/100 = S2 + (Acc \cdot S2)/100$$

where *S1* and *S2* are percent portions of atoms of polar groups among all atoms in near and far layers, respectively. For buried tryptophan residues, when accessibility is about zero, the parameters *A1* and *A2* become equal to *S1* and *S2*, respectively; e.g., the relative polarity of the environment is determined by only the "polar" atoms. However, in cases of partially or completely accessible fluorophores, the values of parameters *A1* and *A2* increase and reflect the quantity of highly polar water nearby them.

Temperature factor and "dynamic accessibility" measure

The occurrence of the above-mentioned fluorescence spectral shift due to dipole relaxation in the surrounding dielectric (Mataga et al., 1955, 1956; Lippert, 1957; Bilot and Kawski, 1962; Liptay, 1965; Bakhshiev, 1972) critically depends on the ratio of the medium relaxation time to the fluorophore fluorescence lifetime (Mazurenko, 1973; Mazurenko and Udaltsov, 1978). The relaxation of the large dipoles of bulk water is very fast (a few hundreds of femtoseconds), while the relaxation time of dipoles of protein groups and bound water can be much longer and may reach many nanoseconds (Callis, 1997). Therefore, it was important to estimate parameters that could reflect a relative mobility of polar groups around fluorophores. Protein crystallography provides such a parameter as a crystallographic temperature factor for individual atoms (Debye-Waller factor or B-factor). Although temperature B-factors in principle measure static or dynamic disorders of atomic position, they are often indistinguishable from a measure of genuine vibration of the atom around its mean position (Frauenfelder et al., 1979; Giacovazzo, 1992; Glusker et al., 1994; Carugo and Argos, 1997). Usually, temperature factors are considered in relative or normalized form (Carugo and Argos, 1997, 1998). We used the temperature B-factors of "polar" atoms as normalized to the mean B-factor value of all Cα atoms in crystal structure within both near (*B1*) and far (*B2*) layers around indole atoms.

Parameters R1 and R2

To account for a common effect of both mobility of neighbor polar atoms of protein and free water around indole rings we introduced parameters *R1* and *R2*, which may be considered as a measure of "dynamic accessibility" for the bulk water:

$$R1 = Acc \cdot B1$$

$$R2 = Acc \cdot B2$$

All parameters calculated from atomic structures are static ones, and only B-factors (*B1*, *B2*) and "dynamic accessibility" (*R1*, *R2*) are kinds of dynamic characteristics of microenvironment.

EVENTUAL INTRAMOLECULAR FLUORESCENCE QUENCHING AND ENERGY HOMO-TRANSFER EFFICIENCY

Fluorescence of individual tryptophan residues might be partially or totally quenched by some protein groups (Cowgill, 1970; Bushueva et al., 1974, 1975; Burstein, 1977a, 1983; Willaert and Engelborghs, 1991; Chen and Barkley, 1998; Yuan et al., 1998) or due to resonance energy homo-transfer to other indole fluorophore(s) (Konev, 1967; Burstein, 1976). A detailed analysis of the distance and orientation of potentially quenching groups (cystein SH and S–S groups, histidine imidazole or imidazolium, arginine guanidinium, hydrated amides, etc.) near indole moiety allowed us to predict eventual quenchers of tryptophan fluorescence.

The probability of excitation energy homo-transfer was estimated using the Förster equation with parameters taken from Dale and Eisinger (1974) and orientation factors calculated from mutual orientation of transition moments of donors in the ^1L_a state and of acceptors in either ^1L_a or ^1L_b states from atomic structure in cases when the centers of their indolic rings in proteins were separated by <12 Å.

ASSIGNMENT OF SPECTRAL COMPONENTS TO INDIVIDUAL TRYPTOPHAN RESIDUES

Among the ~100 proteins and their 60 conformers with measured fluorescence spectra decomposed into log-normal components (Reshetnyak and Burstein, 2001), we have found 48 proteins with well-resolved atomic structures included in the PDB. They contain in sum 137 tryptophan residues. These proteins are listed in Table 1. The preliminary assignments of log-normal spectral components to individual tryptophan residues in a protein were carried out according to the findings that tryptophan residues that are more accessible to solvent emit at longer wavelengths (Burstein et al., 1973; Burstein, 1976, 1977a; Reshetnyak and Burstein, 1997a). The calculated inter-tryptophan energy transfer efficiencies were taken into account for such an assignment to fluorophores in multitryptophan proteins. High homo-transfer efficiency revealed the existence of tryptophan clusters of two or more fluorophores with highly effective internal energy migration in many cases (Reshetnyak and Burstein, 1997b). In such cases we assumed that the most accessible fluorophore in a cluster serves as an emitting acceptor. The final assignment was carried out later, based on the iterative refinements in the procedure of discriminant analysis of microenvironment parameters of all tryptophan fluorophores.

THE CHOICE OF MICROENVIRONMENT PARAMETERS FOR STATISTICAL ANALYSES

Table 2 contains 18 physical and structural parameters of microenvironments of 137 tryptophan residues. We tested each of 18 parameters and their various combinations by cluster and discriminant analyses and revealed that all these characteristics have a tendency to correlate with emission features of fluorophores. However, we later excluded from calculation their accessibility in the presence of water molecules (*Accw1*, *Accw7*, *Accw*) because these parameters were not uniformly determined for various proteins due to

TABLE 1 Assignment of protein tryptophan residues to the classes, identified by cluster and discriminant analyses based on the fluorescence spectra and structural and physical characteristics of indole moiety microenvironment in protein structures

No.	Protein	Protein Code	PDB Entries (resolution, Å)	Trp Residue Position	λ_m (nm)	Class (P, %)*	Comments
			1-Tryptophan-Containing Proteins				
1	Albumin (human serum), N-form	ASH.N	1AO6 (2.5); 1BJ5 (2.5)	W214	344.8	II (78) I (22)	Two log-normal components (328.7 and 344.8). There are difficulties in complete purification from covalently bound ligands (e.g., acetylsalicylate).
2	L-Asparaginase (*E. coli* B)	ASP.N	3ECA (2.4)	W66	323.9	S (96.5)	
3	Azurin (*Pseudomonas aeruginosa*)	AZU	1JOI (2.05); 4AZU (1.9)	W48	307.9	A (34) S (65)	
4	Glucagon	GLG	1GCN (3.0)	W25	350.8	III (100)	
5	Protein G, streptococcal	GPS	1IGD (1.1)	W48	342.5	II (91)	
6	Telokin KRP, myosin light-chain kinase (chicken gizzard)	KRP	1TLK (2.8)	W75	332.4	I (89)	
7	Monellin (*Dioscoreophyllum cummensii*)	MON	1MON (1.7)	W3	340.3	II (95)	
8	Nuclease (*Staphylococcus aureus*)	NST	1STN (1.7)	W140	340.7	II (63) I (37)	Two log-normal components (330.9 and 340.7). In solution it could be two Pro isomers that lead to two conformers of Trp-140 environment (Ikura et al., 1997; Veeraraghavan et al., 1997; Maki et al., 1999).
9	Parvalbumin II (cod *Gadus morrhua*) Ca-form	PAC.CA	Model by M. Laberge	W102	326.5	S (88)	
10	Parvalbumin (whiting *Gadus merlangus*)	PAM	1A75 (1.9)	W102	317.9	S (75) I (20)	Two log-normal components (317.9 and 341.5). Ca^{2+}-binding site might be incompletely occupied.
11	Phospholipase A_2 (bovine pancreas)	PLB	1UNE (1.5); 2BPP (1.8)	W3	352.3	III (100)	Two log-normal components (331.7 and 352.3), the discrepancy is unclear.
12	Phospholipase A_2 (porcine pancreas)	PLS	1P2P (2.6); 4P2P (2.4)	W3	349.4	III (100)	Two log-normal components (323.7 and 349.4), the discrepancy is unclear.
13	Ricin chain A (*Ricinus communis*)	RCA	1RTC (2.3)	W211	324.9	S (65.5) I (34)	Two log-normal components (342.9 and 345.5), the discrepancy is unclear.
14	RNAse T_1 (*Aspergillus oryzae*)	RNT	9RNT (1.5)	W59	325.1	S (97)	
15	Somatotropin, growth hormone (human)	STP	1HGU (2.5)	W86	(335.0)	S (93)	13 residues near W59 were changed by Ala during refinement that might result in erroneous environment parameters.
16	Vipoxin, protein A (*Vipera ammodytes ammodytes*); high ionic strength (dimers)	VTA.HIS	3D model by B. Atanasov	W30	323.6	S (81)	Two log-normal components (323.6 and 346.2). It may be incompletely dimerized protein sample.
17	Vipoxin, protein A (*Vipera ammodytes ammodytes*); low ionic strength (monomers)	VTA.LIS	1VPI (1.76)	W31	349.1	III (100)	
			2-Tryptophan-Containing Proteins				
18	α1-Antitrypsin (human)	A1AT	2PSI (2.9); 7API (3.0); 8API (3.1); 9API (3.0)	W194	324.4	S (73) I (26)	
				W238	340.1	II (98)	
19	Annexin VI (human)	AX6	1AVC (2.9)	W192	346.5	III (100)	
				W343	325.5	S (99)	
20	Melittin, tetramer (*Apis mellifera* venom)	MLT.TETR	2MLT (2.0)	W19	343.9	II (89)	
				W219	329.5	I (40) II (60)	

No.	Protein	Protein Code	PDB Entries (resolution, Å)	Trp Residue Position	λ_m (nm)	Class (P, %)*	Comments
21	Neurotoxin I (cobra *Naja naja oxiana* venom)	NO1	1NTN (1.9)	W26	346.6	II (100)	
				W30	346.6	II (100)	
22	Neurotoxin II (cobra *Naja naja oxiana* venom)	NO2	1NOR (NMR)	W27	344.4	III (100)	
				W28	344.4	III (100)	
23	HIV-1 protease	PRH.LIS	1HHP (2.7)	W6	345.4	III (100)	
				W42	345.4	III (96)	
24	HIV-1 protease complex with pepstatin	PRH.PST	5HVP (2.0)	W6	344.9	III (100)	
				W42	344.9	III (94)	
25	Protease K (*Tritirachium album* Limber)	PRK	2PRK (1.5)	W8	326.9	I (80)	
				W212	344.2	II (72)	
						I (28)	
			3-Tryptophan-Containing Proteins				
26	Agglutinin (wheat germ)	AWG	7WGA (2.0); 9WGA (1.8)	W41 (Q)†	(349.9)	II (92.5)	W41 (Q): 6 sulfur atoms at 4.0–7.3 Å (5 Cys; 1 Met)
				W107	349.1	III (100)	
				W150	349.1	III (97.5)	
27	Subtilisin BPN′	BPN	1ST1 (1.8)	W106	346.2	II (100)	
				W113	346.2	II (97)	
				W241	346.2	II (94)	
28	Elastase (human neutrophils)	ENH	1HNE (1.84); 1PPF (1.8)	W27	336.6	II (88)	
				W141	327.7	S (98)	
				W237	336.6	II (60)	
						I (40)	
29	α-Lactalbumin (human milk)	LAH	1HLM (1.7)	W60 (Q)	(322.4)	I (69)	Two log-normal components (322.4 and 341.9 mn):
						S27	Ca^{2+}-binding site might be incompletely occupied.
				W104	322.4	S (94)	W60 (Q): 3 sulfur atoms of Cys at 4.2–6.7 Å;
				W118 (Q)	(341.9)	I (85)	W118 (Q): 2 sulfur atoms of Cys (S–S bond) at 3.4–5.4 Å.
30	Ovalbumin (hen egg white)	OVH.LIS	1OVA (1.95)	W160	336.9	II (76)	
						I (24)	
				W194	325.8	S (98)	
				W275	336.9	II (89)	
31	Phospholipase A_2 (cobra *Naja naja oxiana* venom)	PAO.HIS	1PSH (2.3)	W18 (ET)‡	(331.4)	II (95)	Two log-normal components (331.4 and 351.6): shortened emission lifetime (underrelaxation). W18: (ET) W18 → W19, 75% and/or Ca^{2+}-binding site is incompletely occupied.
				W19	351.6	III (100)	
				W61	351.6	III (100)	
32	Phosphatase alkalaine (*Escherichia coli*)	PHA	1ALK (2.0)	W109	322.8	S (97)	
				W220	345.8	II (100)	
				W268	345.8	II (66)	
						I (34)	
33	Pyruvate kinase (rabbit skeletal muscle)	PYK	1PKN (2.9)	W157 (Q)	(344.5)	II (91)	Two log-normal components (328.2 and 344.5 nm): shortened lifetime (underrelaxation). W157 (Q): H-bond: (2.9 Å to Asn-145 Oδ1); W481 (ET):W481 → W514, 71%; Arg-488 (3.4 Å); W514 (Q): Arg-515 (3.4 Å); Asn-522 (4.9 Å)
				W481ET, Q	(344.5)	II (100)	
				W514 (Q)	344.5	III (100)	
36	Vipoxin, complex protein A, and protein B (*Vipera ammodytes ammodytes*)	VTC	1AOK (2.0)	W31	334.2	I (56)	
						S (42.5)	

No.	Protein	Protein Code	PDB Entries (resolution, Å)	Trp Residue Position	λ_m (nm)	Class (P, %)*	Comments
				W220 (ET)	(334.2)	II (58)	W220 (ET): W220 $\overset{26\%}{\longleftrightarrow}$ W31 $\overset{61\%}{\longleftrightarrow}$ W231
						I (42)	
				W231	334.2	I (84)	
			4-Tryptophan-Containing Proteins				
37	G-Actin (rabbit skeletal muscule)	ACR.G	1ATN (2.8)	W79 (Q; ET)	(332.8)	II (92)	W79 (Q):_S Met119 (4.9 Å); (ET): W79 ↔ W86, 75%
				W86	320.9	S (73)	
						I (19)	
				W340	320.9	S (98)	
				W356	332.8	I (77)	
						S (20)	
38	Calcium-binding protein (*Astacus leptodactylus*, sarcoplasmic reticulum)	CBC	2SAS (2.4)	W62	337.2	I (79)	
						II (19)	
				W79	347.8	II (80)	
				W86	315.5	S (89)	
				W102	315.5	S (86)	
39	α-Lactalbumin (cow milk)	LAB	1HFZ (2.3)	W26	322.4	S (97)	
				W60	338.3	I (84)	
				W104	322.4	S (89)	
				W118	3338.3	I (78)	
						II (13)	
40	Trypsinogen (bovine)	TRG	1TGB (1.8); 1TGN (1.65)	W51	327.0	I (70)	
						S (28)	
				W141	327.0	S (91)	
				W215	337.7	II (97)	
				W237	327.0	I (61)	
						II (31)	
41	Trypsin (bovine)	TRY	1TLD (1.5); 1TPO (1.7);	W51	328.8	I (85)	
			2PTN (1.55); 3PTN (1.7)	W141	328.8	S (97)	
				W215	(346.8)	II (55)	
					or 328.8	I (44)	
				W237	328.8	I (66)	
						II (32)	
			5-Tryptophan-Containing Proteins				
42	Myosin subfragment 1 (rabbit skeletal muscle)	MS1	2MYS (2.8)	W113	343.5	II (100)	
				W131	343.5	III (100)	
				W440	333.1	I (66)	
						II (32.5)	
				W510	343.5	III (98)	
				W595	318.6	S (40)	
						I (59)	
43	Papain (*Carica papaya*)	PAP.N	1PPN (1.6); 9PAP (1.65)	W7 (Q)	(341.0)	I (82)	W7 (Q): A net of H-bonds between water molecules, ε-amines of Lys-10 and Lys-39, Asn-127 amide at 3–6 Å from the fluorophore, able to form an electron trap.
				W26	317.7	S (72)	
						I (28)	
				W69	349.6	III (99)	
				W177	341.0	II (86)	

No.	Protein	Protein Code	PDB Entries (resolution, Å)	Trp Residue Position	λ_m (nm)	Class (P, %)*	Comments
				W181 (ET)	ET	I (73)	W181 (ET): W181 → W177, 78%
						S (26)	
44	Pepsin (porcine)	PES	4PEP (1.8); 5PEP (2.34)	W39	325.5	S (72)	
						I (26)	
				W141	345.1	II (100)	
				W181	345.1	II (99)	
				W190	325.5	S (99)	
				W299	325.5	I (81)	
			6-Tryptophan-Containing Proteins				
45	Lysozyme (hen egg white); pH 2.2	LZC.A	193L (1.33)	W28	317.4	S (99)	
				W62	345.7	III (100)	
				W63	345.7	II (82.5)	
				W108	317.4	S (89)	
				W111	333.6	I (53)	
						S (46.5)	
				W123	345.7	II (73)	
						I (27)	
46	Lysozyme (hen egg white); pH 7.7	LZC.N	194L (1.4); 1BWH (1.8); 1HEL (1.7); ILZA (1.6); 1LZT (1.97); 1RFP (1.75)	W28	313.9	S (99)	
				W62	343.0	III (100)	
				W63	343.0	II (90)	
				W108	313.9	S (87)	
				W111	334.8	I (71)	
					S (27)		
				W123	343.0	II (69)	
					I (31)		
			7-Tryptophan-Containing Proteins				
47	Carbonic anhydrase II (human)	CAH	1CA2 (2.0); 2CBA (1.54); 2CBB (1.67); 2CBC (1.88); 2CBD (1.67); 2CBE (1.82)	W5	340.3	II (93)	
				W16	317.1	S (96)	
				W97	317.1	S (99)	
				W123	331.1	I (65)	
					S (34)		
				W192	317.1	S (51)	
						I (48)	
				W209	317.1	S (99)	
				W245	331.1	I (78)	
						II (14)	
			8-Tryptophan-Containing Proteins				
48	Chymotrypsinogen A (bovine)	CHG	2CGA (1.8)	W27	327.5	S (97)	
				W29	327.5	S (99)	
				W51	327.5	I (81)	
				W141	327.5	S (72)	
						I (28)	
				W172	339.4	II (60)	
						I (40)	
				W207	339.4	II (66)	
						I (34)	

No.	Protein	Protein Code	PDB Entries (resolution, Å)	Trp Residue Position	λ_m (nm)	Class (P, %)*	Comments
				W215	339.4	II (60)	
						I (39.5)	
				W237	339.4	II (83)	
49	α-Chymotrypsin (bovine)	CHT	4CHA (1.68); 5CHA	W27	328.2	S (96)	
			(1.67)	W29	328.2	S (99.5)	
				W51	328.2	I (79)	
						II (19)	
				W141	328.2	S (75)	
						I (25)	
				W172	341.0	II (40)	
						I (59)	
				W207	328.2	I (72)	
						II (26)	
				W215	341.0	II (28)	
						I (70)	
				W237	341.0	II (75)	
						I (25)	
50	Carboxypeptidase A (bovine)	CPA	2CTB (1.5); 5CPA (1.54)	W63	326.2	S (96)	
				W73	326.2	I (81)	
				W81	326.2 or	I (65)	
					(342.6)	II (33)	
				W126	326.2	I (73)	
						S (24)	
				W147	326.2	S (97)	
				W257	326.2	I (82)	
				W294	326.2 or	I (77.5)	
					(342.6)	II (22)	

*Classification probabilities that a tryptophan belongs to a particular class.
†Emission of the tryptophan residue may be partially quenched by neighbor groups.
‡Emission of the tryptophan residue may be partially quenched due to energy transfer processes.

TABLE 2 18 physical and structural parameters of microenvironments of 137 tryptophan residues of 48 proteins

Prot. Trp	*1 Accw1*	*2 Acc1*	*3 Accw7*	*4 Acc7*	*5 Accw*	***6 Acc***	*7 Den1*	***8 Den2***	*9 B1*	*10 B2*	*11 R1*	*12 R2*	*13 A1*	*14 A2*	*15 Nac1*	*16 Nac2*	*17 Ndon1*	*18 Ndon2*	***Acc1-7***	B_{av}	R_{av}	A_{av}
											Class	**A**										
AZU W48	0	0	0	0	0.02	**0.02**	58	**138**	0.54	0.68	0.08	0.1	19.6	27.3	1	2.3	2	6.4	**0**	**0.61**	**0.09**	**23.5**
											Class	**S**										
LZC.N W28	0	0	0	0	0.07	**0.07**	68	**159**	0.87	0.97	0.06	0.07	25.5	31	2.8	7.8	0.3	2.5	**0**	**0.92**	**0.07**	**28.3**
LZC.N W108	0	8.33	0	0	0	**4.18**	69	**146**	0.99	1.04	4.13	4.34	34.5	40.3	4.7	7.7	0	1.5	**4.16**	**1.02**	**4.24**	**37.4**
CBC W86	0	0	0	0	0	**0**	64	**144**	1	1	0	0	25.1	30.3	2	3	0	3	**0**	**1**	**0**	**27.7**
CBC W102	0	0	0	0	0	**0**	63	**151**	1.17	1.12	0	0	30.8	34.9	2	5	2	2	**0**	**1.15**	**0**	**32.8**
CAH W16	0	0	0	0	0	**0.33**	76	**159**	1.03	1.22	0.34	0.42	35.4	42.1	5.8	5.7	1	1	**0**	**1.13**	**0.38**	**38.8**
CAH W97	0	0	0	0	0	**0**	71	**157**	0.64	0.79	0	0	35.9	33.6	4.3	9.4	2	2.5	**0**	**0.72**	**0**	**34.8**
CAH W192	0	8.75	12.1	12.5	1.85	**3.7**	54	**136**	0.9	1.1	3.34	4.12	23.9	39.4	1.9	6.5	0	8.1	**10.63**	**1**	**3.73**	**31.7**
CAH W209	0	0	0	7.1	0	**1.1**	76	**160**	0.57	0.57	0.63	0.62	39.3	40.4	7	8.5	3.8	3	**3.55**	**0.57**	**0.63**	**39.8**
LZC.AC W28	0	0	0	0	0	**0**	69	**162**	0.87	1	0	0	25.9	30.9	2	10.5	2	3	**0**	**0.94**	**0**	**28.4**
LZC.AC W108	0	8.33	0	0	0	**4.13**	69	**149**	1.01	1.09	4.16	4.49	35	41.3	4	8	0	2	**4.16**	**1.05**	**4.32**	**38.2**
PAP.N W26	0	0	0	0	0	**0**	77	**155**	1.16	1.24	0	0	44.8	43.9	5	7	4	3	**0**	**1.2**	**0**	**44.3**
PAM W102	0	0	0	0	0.22	**0.22**	56	**136**	0.86	1.1	0.18	0.26	19.9	24.1	0.5	3	0.5	3	**0**	**0.98**	**0.22**	**22**
MS1 W595	0	0	0	0	0	**0**	62	**141**	1.19	1.24	0	0	29.1	28.9	3	2	1	3	**0**	**1.22**	**0**	**29**
ACR W86	0	0	0	0	0	**0**	56	**130**	0.68	0.74	0	0	31.2	29.1	1	4	0	1	**0**	**0.71**	**0**	**30.2**
ACR W340	0	0	0	0	0	**0**	78	**149**	0.66	0.69	0	0	38.1	40.1	6	4	2	4	**0**	**0.67**	**0**	**39.1**
LAH W104	0	0	0	0	0.38	**1.36**	64	**141**	0.69	0.82	0.94	1.11	28.3	35.1	4	7	0	2	**0**	**0.76**	**1.03**	**31.7**
LAB W26	0	0	0	0	0	**0**	64	**153**	0.84	0.89	0	0	19.9	29.1	4.8	5.8	0.8	0.8	**0**	**0.87**	**0**	**24.5**
LAB W104	0	0	0	0	0.57	**0.98**	64	**138**	0.79	0.81	0.78	0.79	25.5	31.2	5.8	7.5	0	1	**0**	**0.8**	**0.78**	**28.4**
PHA W109	2.09	2.09	0	0	0.23	**0.23**	59	**155**	0.5	0.59	0.1	0.12	34	35.8	4	3.5	2	2.5	**1.05**	**0.55**	**0.11**	**34.9**
VTA.HIS W30	6.25	6.25	0	0	1.66	**1.66**	60	**137**	0.69	0.96	1.16	1.58	28.3	38.5	2.5	8.8	3	4	**3.13**	**0.83**	**1.37**	**33.4**
ASP.N W66	0	0	5.36	14.3	0.6	**1.78**	67	**147**	0.77	0.73	1.39	1.31	36.9	39.6	4	5	2	4	**7.15**	**0.75**	**1.35**	**38.3**
A1AT W194	0	0	0	0	0	**0**	71	**139**	0.82	1.03	0	0	35.6	32.4	5.5	4.3	2.3	1.8	**0**	**0.93**	**0**	**34**
RCA W211	0	0	0	0	1.19	**4.14**	59	**131**	0.78	0.7	3.25	2.88	39.8	42.6	5	3	4	2	**0**	**0.74**	**3.07**	**41.2**
RNT W59	0	0	0	0	0	**0**	73	**154**	0.9	1.12	0	0	32.5	35.3	2	7	2	4	**0**	**1.01**	**0**	**33.9**
PEB.4-5 W39	0	0	0	1.79	0	**0.2**	63	**134**	0.8	0.77	0.12	0.13	32.3	37.4	1.5	3.5	2	3.5	**0.9**	**0.78**	**0.13**	**34.8**
PEB.4-5 W190	0	0	0	0	0	**0**	71	**157**	0.84	0.92	0	0	26.6	35	5.5	5.5	1.5	3	**0**	**0.88**	**0**	**30.8**
AX6 W343	0	0	0	0	0	**0**	67	**154**	0.85	0.91	0	0	29.9	39.1	6	9	3	2	**0**	**0.88**	**0**	**34.5**
OVH.LIS W194	0	3.13	0	0.89	0	**4.38**	72	**144**	0.61	0.74	2.61	3.19	38.5	37.5	7.8	5	0	2.5	**2.01**	**0.67**	**2.9**	**38**
CPA W63	0	0	0	0	0	**0**	66	**148**	0.98	0.79	0	0	26.6	29.1	2.5	2.5	2.5	3.5	**0**	**0.89**	**0**	**27.9**
CPA W147	0	0	0	1.8	0	**0.2**	73	**154**	0.92	1.01	0.21	0.24	33.3	42.3	2	3	2.5	7	**0.9**	**0.97**	**0.23**	**37.8**
PAC.CA W102	0	0	0	0	0	**0**	63	**141**	0.85	0.9	0	0	19.7	24.1	0	1	0	3	**0**	**0.88**	**0**	**21.9**
TRG W141	0	0	0	0	0	**0**	74	**154**	1.03	1.07	0	0	36	45.4	11	6.5	2.5	3	**0**	**1.05**	**0**	**40.7**
CHG W27	0	0	0	0	0	**0**	59	**147**	0.58	0.81	0	0	29.5	35.6	2	3	2	3.5	**0**	**0.7**	**0**	**32.6**
CHG W29	0	2.09	0	0	0	**0.23**	71	**154**	0.69	0.78	0.16	0.18	32.4	41.5	5	8.5	0	4.5	**1.05**	**0.74**	**0.17**	**37**
CHG W141	0	0	0	0	0	**0**	77	**156**	1.21	1.17	0	0	43.9	48.1	9	8	2.5	3	**0**	**1.19**	**0**	**46**
ENH W141	0	0	0	0	0	**0**	72	**155**	0.8	0.98	0	0	34.5	47.5	9	7.5	2	2.5	**0**	**0.89**	**0**	**41**
CHT W27	0	0	4.47	6.25	0.59	**1**	59	**148**	0.85	0.88	0.85	0.88	31.5	38.7	2.7	3	2.3	3.7	**3.13**	**0.87**	**0.87**	**35.1**
CHT W29	0	1.04	0	0	0	**0.12**	74	**161**	0.7	0.76	0.09	0.1	32.2	40.7	4.7	8.3	0	5.3	**0.52**	**0.73**	**0.1**	**36.5**
CHT W141	0	0	0	0	0	**0**	73	**149**	1.13	0.99	0	0	41.1	43.5	9	7.7	3	2.7	**0**	**1.06**	**0**	**42.3**
TRY W141	0	0	0	0	0	**0**	74	**152**	0.88	0.9	0	0	36.4	47.8	7.8	6.3	3	2.5	**0**	**0.89**	**0**	**42.1**
STP W86	0	0	0	0	2.05	**2.05**	58	**144**	0.89	0.93	1.8	1.9	29.4	34.5	4	6	0	3	**0**	**0.91**	**1.85**	**32**
											Class	**I**										
PEB.4-5 W299	2.09	25	1.79	1.79	0.62	**3.17**	54	**134**	1.37	1.44	4.38	4.6	34.3	34.6	1	4	3	1.5	**13.4**	**1.41**	**4.49**	**34.5**
CPA W73	0	0	0	0	2.88	**8.86**	64	**132**	1.15	1.02	10.2	9.07	51.1	44.4	2.5	8	2.5	2.5	**0**	**1.09**	**9.64**	**47.8**
CPA W81	14.6	50	8.93	26.7	2.61	**13.5**	61	**138**	1.48	1.34	19.7	18	41	44	1.5	5.5	1.5	6	**38.35**	**1.41**	**18.85**	**42.5**

Prot. Trp	*1* *Accw1*	*2* *Acc1*	*3* *Accw7*	*4* *Acc7*	*5* *Accw*	***6*** ***Acc***	*7* *Den1*	***8*** ***Den2***	*9* *B1*	*10* *B2*	*11* *R1*	*12* *R2*	*13* *A1*	*14* *A2*	*15* *Nac1*	*16* *Nac2*	*17* *Ndon1*	*18* *Ndon2*	***Acc1-7***	B_{av}	R_{av}	A_{av}
CPA W126	0	0	0	0	3.1	**10.1**	65	**138**	1.4	1	14.2	10.1	47.4	48	4.5	8	3	1	**0**	**1.2**	**12.15**	**47.7**
CPA W257	20.8	20.8	0	0	2.4	**2.48**	64	**132**	1.32	1.19	3.28	3.08	34.8	41.9	4	4	2	4	**10.4**	**1.26**	**3.18**	**38.3**
CPA W294	2.1	12.5	0	0	0.23	**2.54**	59	**127**	1.7	1.1	4.3	2.9	23.9	32.1	1.5	2	1	4	**6.25**	**1.4**	**3.6**	**28**
PRK W8	0	0	0	0	0	**4.22**	72	**136**	1.51	1.3	6.36	5.49	46.7	53.2	4	5	2	3	**0**	**1.41**	**5.93**	**50**
TRG W51	6.25	14.6	0	0	1.19	**2.28**	64	**134**	0.95	0.94	2.16	2.17	32.7	45	4	5.5	2	4	**7.3**	**0.95**	**2.17**	**38.8**
CHG W51	0	31.3	0	0	0	**4.29**	67	**132**	1.22	1.22	5.39	5.21	34	45.3	2.5	5	2	3.5	**15.65**	**1.22**	**5.3**	**39.7**
CHT W51	24	33.3	0	0	2.83	**4.03**	66	**126**	1.11	1.03	4.5	4.2	31.3	42.3	2.7	4.7	1.7	3	**16.65**	**1.07**	**4.35**	**36.8**
CHT W207	6.25	12.5	22.3	33	7.09	**9.55**	53	**119**	0.86	0.85	8.28	8.12	30.7	47.8	4	8	2	3	**22.75**	**0.86**	**8.2**	**39.3**
TRY W51	6.25	19.8	0	0	1.19	**2.94**	64	**130**	0.96	1.1	2.8	3.3	31.8	44.9	3.3	5.3	2	3.8	**9.9**	**1.03**	**3.05**	**38.3**
TRY W215	0	0	1.79	3.57	10.4	**10.6**	55	**113**	1.03	0.95	11	10.2	35.2	47.1	4.8	7.8	0.3	4	**1.79**	**0.99**	**10.6**	**41.2**
TRY W237	21.9	21.9	0.89	0.89	7.68	**7.92**	56	**110**	0.82	0.85	6.52	6.74	34.2	35.4	1.8	2	2	3.3	**11.4**	**0.84**	**6.63**	**34.8**
MLT.TETR W219	33.3	33.3	3.57	3.57	6.55	**6.71**	58	**119**	1.25	1.25	8.4	8.4	32.3	39.4	4	7	0	3.5	**18.44**	**1.25**	**8.4**	**35.8**
KRP W75	0	0	0	0	0.33	**0.33**	64	**136**	1.4	1.42	0.46	0.47	33.8	37.8	3	5	3	3	**0**	**1.41**	**0.47**	**35.8**
ACR W356	20.8	20.8	3.57	3.57	3.69	**3.69**	65	**133**	1.03	1.05	3.8	3.9	30.5	38	6	3	0	0	**12.19**	**1.04**	**3.85**	**34.3**
CAH W123	0	0	0.79	12.9	0.39	**2.2**	72	**147**	1.11	1.33	2.41	2.89	41.1	44.2	3.9	6.3	2.6	3.4	**6.45**	**1.22**	**2.65**	**42.7**
CAH W245	0	0	17.2	27.5	8.27	**14.3**	62	**124**	0.85	0.9	12.1	13	55.6	48.4	6.3	3	2	1	**13.75**	**0.88**	**12.55**	**52**
MS1 W440	41.7	41.7	7.14	7.14	9.02	**9.02**	51	**122**	0.99	1.28	8.9	11.6	27.3	38.8	2	5	1	4	**24.42**	**1.14**	**10.25**	**33.1**
LZC.AC W111	0	0	3.57	17.86	2.31	**3.9**	62	**140**	0.82	1.15	3.19	4.5	37.5	47.5	4	7	4	2	**8.93**	**0.99**	**3.85**	**42.5**
VTC W31	4.17	25	0	10.7	0.46	**5.39**	60	**136**	1.09	0.8	5.86	4.33	31.2	41.1	6	9	0	3	**17.85**	**0.95**	**5.1**	**36.2**
VTC W231	0	20.8	0	0	0.38	**8.85**	61	**133**	1.34	1	11.8	8.81	38.9	48.7	5	4	7	3	**10.4**	**1.17**	**10.31**	**43.8**
LZC/N W111	0	0	6.6	19.1	3.2	**4.5**	62	**134**	0.84	1.06	3.86	4.74	37.4	46.8	4.2	6.5	3.5	2.5	**9.55**	**0.95**	**4.3**	**42.1**
CBC W62	29.2	29.2	0	0	6.18	**6.18**	60	**125**	1.12	1.1	6.93	6.81	32	41.1	2	5	3	4	**14.6**	**1.11**	**6.87**	**36.6**
TRG W237	25	25	0	0	6.5	**6.7**	59	**112**	0.96	0.87	6.47	6.08	33.4	36.2	2	2	2	3	**12.5**	**0.92**	**6.28**	**34.8**
LAB W60	6.25	6.25	11.6	11.6	2.66	**2.66**	64	**122**	0.98	1.06	2.57	2.76	26.3	36.4	2.3	3.8	2	2	**8.93**	**1.02**	**2.67**	**31.4**
LAB W118	13.5	16.7	22.3	28.6	9.73	**11.5**	59	**124**	0.86	0.81	9.21	8.87	42.5	43.1	1.8	1.3	2.5	4	**22.65**	**0.84**	**9.04**	**42.8**
LAH W60	0	12.5	7.14	10.7	1.56	**3.35**	64	**124**	0.64	0.92	2.14	3.1	24.8	39	3	3	2	2	**11.6**	**0.78**	**2.62**	**31.9**
LAH W118	0	4.17	10.7	14.3	3.87	**6.27**	61	**132**	1.18	1.09	7.41	6.86	45.1	42.6	3	1	5	4	**9.23**	**1.14**	**7.14**	**43.8**
PAP.N W7	0	0	0	3.57	3.64	**8.26**	60	**129**	1.24	1.46	10.2	12	38	42.5	8	6	1	2	**1.79**	**1.35**	**11.1**	**40.3**
PAP.N W181	0	0	0	0	1.15	**1.8**	68	**148**	1.3	1.42	2.29	2.66	38.4	40.7	4.5	6.5	3	3	**0**	**1.36**	**2.48**	**39.6**
											Class	**II**										
ENH W27	0	0	0	26.8	5.4	**17.8**	44	**109**	1.46	1.09	25.9	19.2	47.6	48.5	3	2.5	1.5	1.5	**13.4**	**1.28**	**22.55**	**48.1**
ENH W237	2.09	41.7	0	0	3.5	**9.21**	57	**116**	0.96	1.11	8.84	10.2	41.6	40.5	3.5	6.5	1	2	**20.85**	**1.04**	**9.52**	**41.1**
OVH.LIS W160	28.1	28.1	8.92	23.2	8.04	**10.8**	49	**115**	0.93	1.06	10	11.4	49.5	47.3	4.8	5.3	1.5	2	**25.65**	**1**	**10.7**	**48.4**
OVH.LIS W275	5.2	5.2	0	0	1.07	**1.07**	56	**113**	1.7	1.43	1.73	1.48	29	33.3	2.3	3.5	1.3	1.5	**2.6**	**1.57**	**1.61**	**31.2**
TRG W215	0	0	1.79	5.36	12.2	**12.6**	55	**105**	1.03	0.93	12.9	11.7	32.8	46.7	2.5	6	1	3.5	**2.68**	**0.98**	**12.3**	**39.8**
CHG W172	0	0	0	0	3.83	**12.8**	57	**111**	1.21	1.26	15.5	16.3	38.5	44.6	5.5	3.5	0	1	**0**	**1.24**	**15.9**	**41.6**
CHG W207	0	6.25	10.7	16.1	1.19	**6.07**	57	**113**	1.22	0.96	7.41	5.78	35.4	44.6	4	7.5	0.5	2.5	**11.18**	**1.09**	**6.6**	**40**
CHG W215	0	0	1.79	5.36	1.68	**9.35**	58	**112**	1.15	1.11	10.7	10.4	37.5	50.6	3.5	8.5	0	3	**2.68**	**1.13**	**10.55**	**44.1**
CHG W237	20.8	43.8	1.79	5.36	4.31	**10.9**	57	**116**	1.25	1.11	13.4	12.1	49.4	43.6	3	2	3.5	4.5	**24.58**	**1.18**	**12.75**	**46.5**
A1AT W238	60.4	69.8	60.7	60.7	19.4	**20.5**	39	**103**	0.87	0.91	17.8	18.7	30.3	49.5	1	2.8	1.3	3.5	**65.25**	**0.89**	**18.25**	**39.9**
CAH W5	44.6	61.5	14.6	18.9	7.28	**11.1**	61	**122**	1.49	1.42	17	15.6	38.7	46	4.1	6.9	0.9	1.1	**40.2**	**1.46**	**16.3**	**42.3**
MON W3	14.6	52.1	46.4	53.6	24	**31.4**	43	**83**	1.39	1.33	43.9	41.9	55.6	69.2	3.5	5	1.5	2	**52.85**	**1.36**	**42.9**	**62.4**
NST W140	0	0	0	0	12	**12.4**	54	**112**	1.44	1.11	17.8	13.8	37.1	49.7	3	1	2	2	**0**	**1.28**	**15.8**	**43.4**
CHT W172	0	0	0	0	9.9	**11.4**	55	**112**	1.1	1.3	12.6	14.4	30.6	40.6	4.3	5.3	0.7	1	**0**	**1.2**	**13.5**	**35.6**
CHT W215	0	0	3.57	6.25	16.1	**17**	56	**114**	0.96	1.07	16.2	18.1	39	52.7	4	8.3	0.3	3	**3.13**	**1.02**	**17.15**	**45.8**
CHT W237	37.5	49	2.68	3.57	10.4	**12.5**	53	**111**	0.98	0.86	12.3	10.8	45.2	43.4	2.3	2.3	3.3	0.4	**26.29**	**0.92**	**11.55**	**44.3**
PAP.N W177	4.2	16.7	16.1	37.5	9.2	**18.2**	62	**126**	1.7	1.6	29.5	30	51.2	50	7.5	9.5	2.5	1	**27.1**	**1.65**	**29.75**	**50.6**
GPS W48	0	33.3	0	3.57	0	**6.39**	60	**143**	2.29	2.34	14.6	14.9	40.4	48.7	3	10	2	2	**18.44**	**2.32**	**14.75**	**44.6**

Prot. Trp	*1 Accw1*	*2 Acc1*	*3 Accw7*	*4 Acc7*	*5 Accw*	***6 Acc***	*7 Den1*	***8 Den2***	*9 B1*	*10 B2*	*11 R1*	*12 R2*	*13 A1*	*14 A2*	*15 Nac1*	*16 Nac2*	*17 Ndon1*	*18 Ndon2*	***Acc1-7***	***B*$_{av}$**	***R*$_{av}$**	***A*$_{av}$**
LZC.N W63	13.2	43.1	22	33.3	5	**10.3**	58	**122**	1.34	1.35	13.8	13.9	38.1	51.2	4.3	4.2	2	1.7	**38.2**	**1.35**	**13.85**	**44.7**
LZC.N W123	0.7	30.6	6	40.5	7.2	**22.7**	55	**123**	1.4	1.2	31.3	28.1	46.1	52.8	3.7	5	1.7	4	**35.55**	**1.3**	**29.7**	**49.5**
MS1 W113	0	0	35.7	35.7	21.8	**21.8**	34	**87**	1.96	1.89	42.8	41.2	51.4	39.5	0	3	2	2	**17.85**	**1.93**	**42**	**45.5**
MLT.TETR W19	75	75	17.9	17.9	11.6	**13.9**	52	**113**	1.13	1.12	15.7	15.5	30.9	40.8	2	8	0	5	**46.45**	**1.13**	**15.6**	**35.8**
PRK W212	0	25	14.3	25	4.2	**9.3**	57	**116**	1	1	9.4	9.6	46.3	47.3	2	4	3	4	**25**	**1**	**9.5**	**46.8**
PYK W157	0	0	28.6	28.6	8.54	**8.54**	58	**103**	1.36	0.54	11.6	4.64	49.5	46.5	4	2	2	2	**14.3**	**0.95**	**8.12**	**48**
PYK W481	0	0	14.3	14.3	16.2	**16.2**	38	**85**	1.76	1.1	28.5	17.8	41.8	49.9	1	4	2	1	**7.15**	**1.43**	**23.15**	**45.8**
ASH.N W214	31.3	31.3	52.7	52.7	31.1	**31.1**	37	**98**	0.67	0.7	21.8	21.7	34	42.7	1.3	3	0	3.3	**42**	**0.68**	**21.75**	**38.3**
PEB.4-5 W141	60.4	77.1	57.1	80.4	19.2	**33.7**	40	**100**	1	1.03	33.7	34.9	71.7	61.8	3	3	1	3.5	**78.75**	**1.02**	**34.3**	**66.8**
PEB.4-5 W181	39.6	54.2	37.5	39.3	25.9	**27.9**	47	**99**	1.13	1.28	30.6	36.6	43.4	53.9	2.5	3	2	1	**46.75**	**1.21**	**33.6**	**48.7**
LZC.AC W63	0	37.5	10.7	25	2.34	**8.8**	60	**127**	1.45	1.45	12.8	12.8	39	51.5	5	5	2	4	**31.25**	**1.45**	**12.8**	**45.3**
LZC.AC W123	0	33.3	7.1	39.3	6	**23.8**	57	**125**	1.5	1.3	34.5	30.4	48.4	54.7	5	4	2	5	**36.3**	**1.4**	**32.45**	**51.6**
PHA W220	35.4	35.4	41.1	41.1	22.4	**22.4**	39	**78**	1.72	1.77	40.3	41	44.3	56.6	0	1.1	1.1	1.1	**38.25**	**1.75**	**40.65**	**50.5**
PHA W268	27.1	27.1	61.7	61.7	17.2	**17.8**	57	**117**	0.91	0.8	15.8	14	35.4	50.2	2	6	2.5	5.5	**44.4**	**0.86**	**14.9**	**42.8**
BPN W106	16.7	16.7	28.6	28.6	9.51	**13.2**	53	**102**	1.7	1.63	22.5	21.5	41.9	51.9	3	3	2	4	**22.65**	**1.67**	**22**	**46.9**
BPN W113	33.3	50	7.14	28.6	7.76	**14**	45	**104**	1.12	0.81	15.7	11.3	46.5	48.1	1	4	2	5	**39.3**	**0.97**	**13.5**	**47.3**
BPN W241	0	8.3	3.6	10.7	1.4	**3.2**	59	**113**	1.64	1.09	5.2	3.4	44.6	49.2	3	3	2	1	**9.5**	**1.37**	**4.3**	**46.9**
NO1 W26	20.8	41.7	57.1	71.4	11.4	**19.3**	56	**101**	1.21	1.34	23.5	26	51	51.4	2	2	3	3	**56.55**	**1.28**	**24.75**	**51.2**
CBC W79	62.5	62.5	21.4	21.4	13.9	**13.9**	55	**109**	0.99	0.91	13.7	12.7	23.8	42.1	3	3	1	3	**41.95**	**0.95**	**13.2**	**33**
AWG W41	0.17	6.25	7.15	16.1	1.68	**10.6**	60	**114**	1.29	1.09	13.8	11.5	59.4	64.8	3.5	4	5.5	5	**11.18**	**1.19**	**12.65**	**62.1**
PAO.HIS W18	13.9	20.8	41.7	42.9	10	**13.6**	50	**100**	1.08	1.03	14.5	13.9	27.9	41.1	2	1.8	1.6	3.7	**31.85**	**1.06**	**14.2**	**34.5**
ACR W79	58.3	58.3	3.57	3.57	12.4	**12.4**	51	**101**	0.89	0.89	11.1	10.2	33.5	45.9	1	1	4	2	**30.94**	**0.89**	**10.65**	**39.7**
VTC W220	16.7	25	0	0	2.83	**5.07**	57	**117**	1.33	1.09	6.75	5.53	29.1	43	3	4	4	1	**12.5**	**1.21**	**6.14**	**36.1**
											Class	**III**										
LZC.N W62	47.2	61.1	73.2	74.4	46.6	**49.7**	33	**73**	1.76	1.77	87.9	88.8	54.6	63.9	0.2	3.2	0.8	1	**67.75**	**1.77**	**88.35**	**59.3**
MS1 W131	91.7	91.7	60.7	60.7	62.1	**62.1**	23	**50**	2.1	2.91	130.5	180.5	64.9	50.3	0	0	2	2	**76.2**	**2.51**	**155.5**	**57.6**
MS1 W510	20.8	20.8	46.4	46.4	28.6	**28.9**	35	**75**	2.1	2.33	60	66.8	43.7	33.6	1	3	2	1	**33.6**	**2.22**	**63.4**	**38.7**
NO2 W27	58.3	58.3	82.1	82.1	42	**42**	46	**91**	2.6	2.7	109.3	170	58.5	57.1	3	5	3	1	**70.2**	**2.65**	**139.7**	**57.8**
NO2 W28	58.3	58.3	64.3	64.3	39.5	**39.5**	33	**85**	1.51	2.7	59.6	170	41.3	58.2	2	2	2	3	**61.3**	**2.11**	**114.8**	**49.8**
PYK W514	54.2	54.2	100	100	66.3	**66.3**	20	**48**	1.19	1.47	78.6	97.5	62.2	85.1	0	4	0	0	**77.1**	**1.33**	**88.05**	**73.6**
PRH.PST W6	35.4	75	62.5	75	60.7	**60.3**	31	**57**	1.07	1	64.8	60.3	73.6	96.3	1.5	0.5	2	1.5	**75**	**1.04**	**62.55**	**85**
PRH.PST W42	75	83.3	71.4	80.4	26.7	**34.9**	44	**79**	1.25	2.4	43	48.8	56.1	64	2.5	2	0.5	2.5	**81.85**	**1.83**	**45.9**	**60.1**
PRH.LIS W6	70.8	70.8	78.6	78.6	61.5	**61.5**	21	**78**	0.81	1.09	49.5	67	62.8	62.8	1	1	0	1	**74.7**	**0.95**	**58.25**	**62.8**
PRH.LIS W42	87.5	87.5	85.7	85.7	39.7	**39.7**	35	**69**	1.26	1.37	49.9	54.3	43.2	56.9	1	1	1	1	**86.6**	**1.32**	**52.1**	**50.1**
LZC.AC W62	41.7	72.9	75	82.1	49.6	**56**	32	**76**	1.78	1.85	99.9	103.4	50.2	72.8	0	2	1	1	**77.5**	**1.82**	**101.7**	**61.5**
AX6 W192	95.8	95.8	100	100	86.2	**86.2**	10	**22**	1.27	1.12	109.5	96.6	48.4	95.5	1	1	0	0	**97.9**	**1.2**	**103.1**	**72**
NO1 W30	95.8	95.8	96.4	96.4	61.5	**61.5**	19	**37**	2.18	2.06	134.2	126.5	48.6	71.5	2	2	1	2	**96.1**	**2.12**	**130.4**	**60.1**
VTA.LIS W31	100	100	100	100	81.9	**86.5**	13	**37**	0.82	0.93	71.3	80.7	68.6	116.4	1	0	0	0	**100**	**0.88**	**76**	**92.5**
AWG W107	60.4	77.1	55.4	60.7	54.8	**63**	23	**51**	1.27	1.35	79.7	84.4	69.6	102.3	1	1	2	0	**68.9**	**1.31**	**82.05**	**86**
AWG W150	25	52.1	1.79	23.2	9.72	**37.7**	50	**84**	1.81	1.21	68.3	45.5	66.9	69.5	2	4	2	1	**37.65**	**1.51**	**56.9**	**68.2**
PLS W3	31.3	41.7	44.6	58.9	50.9	**53.7**	25	**55**	0.62	0.71	33.3	38.4	57	77.9	0.5	0	2	2	**50.3**	**0.66**	**35.85**	**67.5**
PAP.N W69	14.6	43.8	41.1	42.9	29.9	**33.9**	46	**89**	2.02	1.9	68.4	64.3	40.6	60.3	2.5	3	1	3.5	**43.35**	**1.96**	**66.35**	**50.5**
GLG W25	100	100	96.4	96.4	74.8	**74.8**	14	**43**	0.87	1.15	64.9	86	52.1	62.9	0	0	0	0	**98.2**	**1.01**	**75.45**	**57.5**
PAO.HIS W19	55.6	77.8	15.5	73.8	46.3	**62**	23	**56**	1.34	1.22	83.3	82.3	75.4	84.7	1.7	2.4	1	0.3	**75.8**	**1.28**	**82.8**	**80.1**
PAO.HIS W61	16.7	63.9	27.4	73.8	41.2	**59.5**	25	**62**	1.03	0.97	62.5	58	93.5	80.8	1.8	3.8	0.6	0.3	**68.85**	**1**	**60.25**	**87.1**
PLB W3	33.3	39.6	50	50	52.7	**56.9**	27	**63**	1.47	1.22	84.4	70.2	60.3	68.4	1	0	2	0.5	**44.8**	**1.35**	**77.3**	**64.3**

Six parameters used in discriminant and cluster analyses are shown in bold.

different numbers of the bound water molecules contained in their structures. The number of water molecules included in crystallographic models depended mainly on the resolution at which the structure has been solved. On average, at 2.0 Å resolution one water molecule per residue is included, while at 1.0 Å resolution this number reaches values of ~1.6–1.7 (Carugo and Bordo, 1999). Aside from that, to avoid an extra redundancy in the set of excluded several parameters of similar physical meaning. As a result, in the final step of statistical analyses, we used only six parameters of tryptophan microenvironment: *Acc1–7*, *Acc*, *Den2*, B_{av}, R_{av}, A_{av}. The last three parameters are averaged values of normalized temperature factor, dynamic accessibility, and relative polarity of environment in the layer 0 to 7.5 Å from a fluorophore. *Acc1–7* is a mean accessibility of Nϵ1 (*Acc1*) and Cζ2 (*Acc7*) atoms.

CLUSTER ANALYSIS AND DISCRETE NATURE OF STRUCTURAL PARAMETERS

In our previous paper (Reshetnyak and Burstein, 2001) we confirmed existence of several states of emitting tryptophan fluorophores in proteins by analyzing the histograms of distribution of spectral maximum position of >300 log-normal components of protein fluorescence spectra. To check whether there is any analogous discreteness of fluorophore microenvironment, we applied cluster analysis (STATISTICA for Windows 5.0, StatSoft, Inc., 1984–1995) to the above-described set of the physical and structural environment parameters. We chose this method because it does not require any a priori assumption about data distribution and allows us to reveal naturally existing classes and quantitatively estimate degrees of their distinctions. The purpose of the joining (or tree-clustering) algorithm applied here is to join objects (i.e., tryptophan residues) into clusters by using an appropriate measure of distance between objects in the multi-dimensional space of analyzed parameters, and then applying an amalgamation (linkage) rule. As a measure of the distance, *D*, between objects we selected the power distance form, i.e.:

$$D(x, y) = (\Sigma_i |x_i - y_i|P)^{1/r},$$

with $p = 1$ and $r = 4$. The power distance can be viewed as a generalized Euclidean distance with weights. In the case when $p < 2$ and $r > 2$, the distance between nearest neighbors tends to increase compared to the Euclidean distance, while the highly differed objects tend to be less distanced with respect to a modified distance function. As an amalgamation rule we implemented Ward's method based on an analysis of variance approach to evaluate the distances between clusters (Ward, 1963). The results of cluster analysis are most obviously presentable in the form of hierarchical trees (dendrograms).

Hierarchical tree for the set of tryptophan residues of single-, two-, and three-tryptophan-containing proteins

Because the assignment of spectral components to individual fluorophores in proteins with a few tryptophan residues is much more evident, we first applied the cluster analysis to the set of fluorophores of proteins containing not more than three tryptophans (36 proteins; 60 tryptophan residues). Six microenvironment parameters (marked bold in Table 2) were normalized (minimal and maximal values of each parameter being 0 and 1, respectively) to get equal weights for each parameter. The result of cluster analysis as a dendrogram is presented in Fig. 2. Here, the ordinate represents the relative (percent) linkage distance of the maximal distance, and the abscissa presents all objects (tryptophan residues) under analysis. The algorithm starts with selection of each tryptophan residue into a cluster by itself. Step by step, more and more tryptophans become linked together and aggregated (amalgamated) into progressively larger clusters of increasingly dissimilar objects according to rise in distances between lesser clusters. Finally, all residues become linked. The presented dendrogram possesses a clear "structure" in terms of clusters (branches) of tryptophan residues that are more similar to each other. The dendrogram reveals three large clusters marked "Classes A and S," "Classes I and II," and "Class III." The most distant is "Class III," the distance between the first two clusters being ~45% of the maximal distance (100%), separating them from "Class III."

The preliminary assignment of spectral components to fluorophores of proteins with one, two, or three tryptophan residues (Table 1, columns 5 and 6) was more obvious. However, for several single-tryptophan-containing proteins the number of spectral components exceeded the number of tryptophan residues. Such cases are reflected and sometimes interpreted in the last column of Table 1. For example, the log-normal decomposition of tryptophan fluorescence spectra of single-tryptophan-containing staphylococcal nuclease stably gave two components, which might be a result of the equilibrium between *cis* and *trans* isomers of Pro-117 and/or Pro-47 in the protein solution, unlike crystal (Fox et al., 1986; Evans et al., 1987; Ikura et al., 1997; Veeraraghavan et al., 1997; Maki et al., 1999).

The assignment of spectral components to tryptophan residues allowed us to estimate mean values of emission maximum positions λ_m for fluorophores in each large cluster (see Fig. 2). The λ_m values for classes A, S, and III almost coincide with those seen in the histogram of occurrence of maximum positions of log-normal components in emission spectra of >160 proteins (~326 and 350 nm, respectively; Reshetnyak and Burstein, 2001). This means that the ensemble of microenvironment parameters rather strongly correlate with discretely distributed fluorescence properties of tryptophan residues in native proteins. However, the fluorophores of classes I and II, best separated in

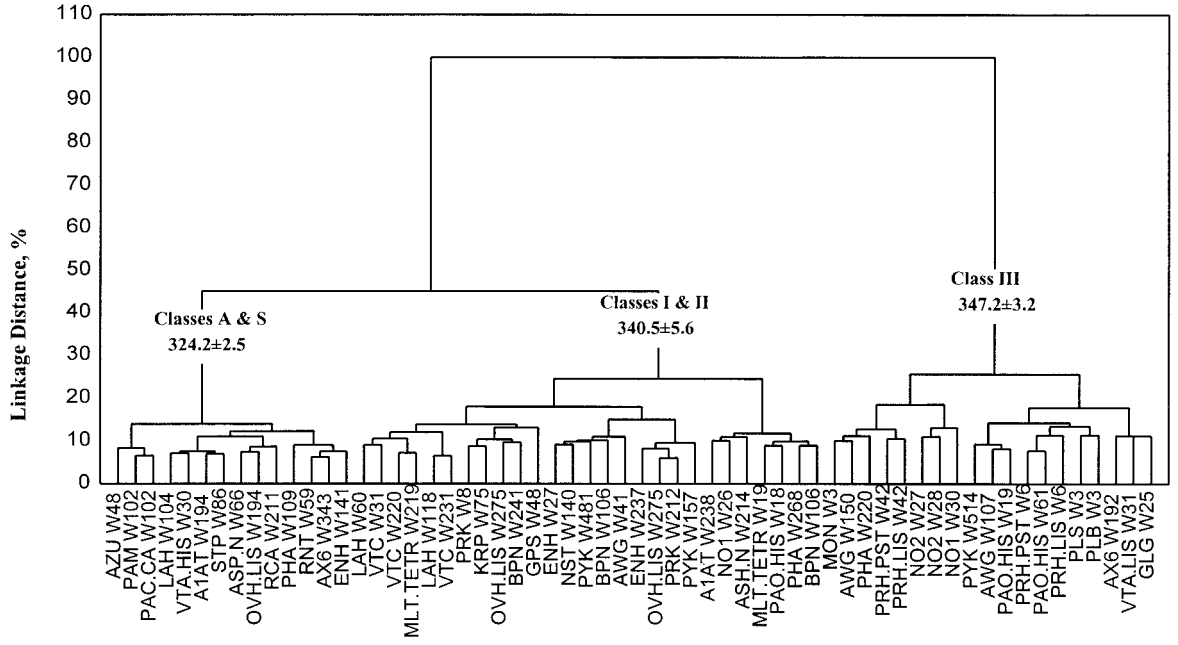

FIGURE 2 A dendrogram constructed based on six normalized structural parameters of 60 tryptophan residues of 36 1-2-3-tryptophan-containing proteins.

the histograms (Reshetnyak and Burstein, 2001), here are amalgamated as one mixed cluster.

Hierarchical tree for the set of tryptophan residues of all proteins under study

The result of cluster analysis of microenvironment parameters of the whole set of 137 fluorophores in 48 proteins is presented in Fig. 3. The distance between "Classes A and S" and "Classes I and II" increased up to 62% in this dendrogram, compared with ~45% in Fig. 2. However, tryptophan residues of classes I and II also remain inseparable and form a single mixed cluster. This fact may mean that the striking discrimination of classes I and II in the fluorescence-maximum-position histograms is scarcely reflected by static features of fluorophore microenvironment, and the cause of their emission distinctions may lie rather in qualitative differences of kinetics of excited fluorophore interactions with dielectric medium. However, we made an attempt to discriminate all five discrete classes of tryptophan residues by their microenvironment parameters using a more refined multidimensional statistical method of discriminant analysis (STATISTICA for Windows 5.0, StatSoft, Inc., 1984–1995), although this analysis requires constructing a preliminary "teaching" classification.

DISCRIMINANT FUNCTION ANALYSIS AND CANONICAL VARIATES

Main concepts of discriminant analysis

The main idea of discriminant analysis is similar to the analysis of variance (Gnanadesikan, 1977). Discriminant analysis is rather widely used to detect the variables that provide the most efficient discrimination between estimated numbers of classes (in our case it would be five classes), and to classify objects by assigning them to different classes based on "training" data sets of parameters (Burgess, 1995; Zhang, 1997; Andrade et al., 1998; Chou and Elrod, 1998). The procedure of discriminant analysis develops similarly to multivariate analysis of variance, when a set of variables are included in the study to see which one(s) contribute to the discrimination between classes. A total variance-covariance matrix and a pooled-within-group variance-covariance matrix is computed. Those two matrices are compared via multivariate F-tests to determine whether there is any significant difference between classes. Discriminant analysis searches for the best discriminant functions (also called canonical variates, canonical coordinates, or roots), which are independent (mutually non-correlated) optimal combinations of investigated parameters that provide the highest overall resolution between classes. Discrimination is carried out in the space of canonical coordinates using, for example, Euclidean or Mahalanobis distances between class means for an allocation rule (Gnanadesikan, 1977).

We performed a stepwise discriminant analysis (the forward mode). At each step the program reviewed all six microenvironment parameters and evaluated which one would contribute most to the discrimination among five classes. The selected parameter then would be included in the model, and the program would proceed. The stepwise procedure was "guided" by the respective F-test-to-enter value. Such a value for a parameter indicates its statistical

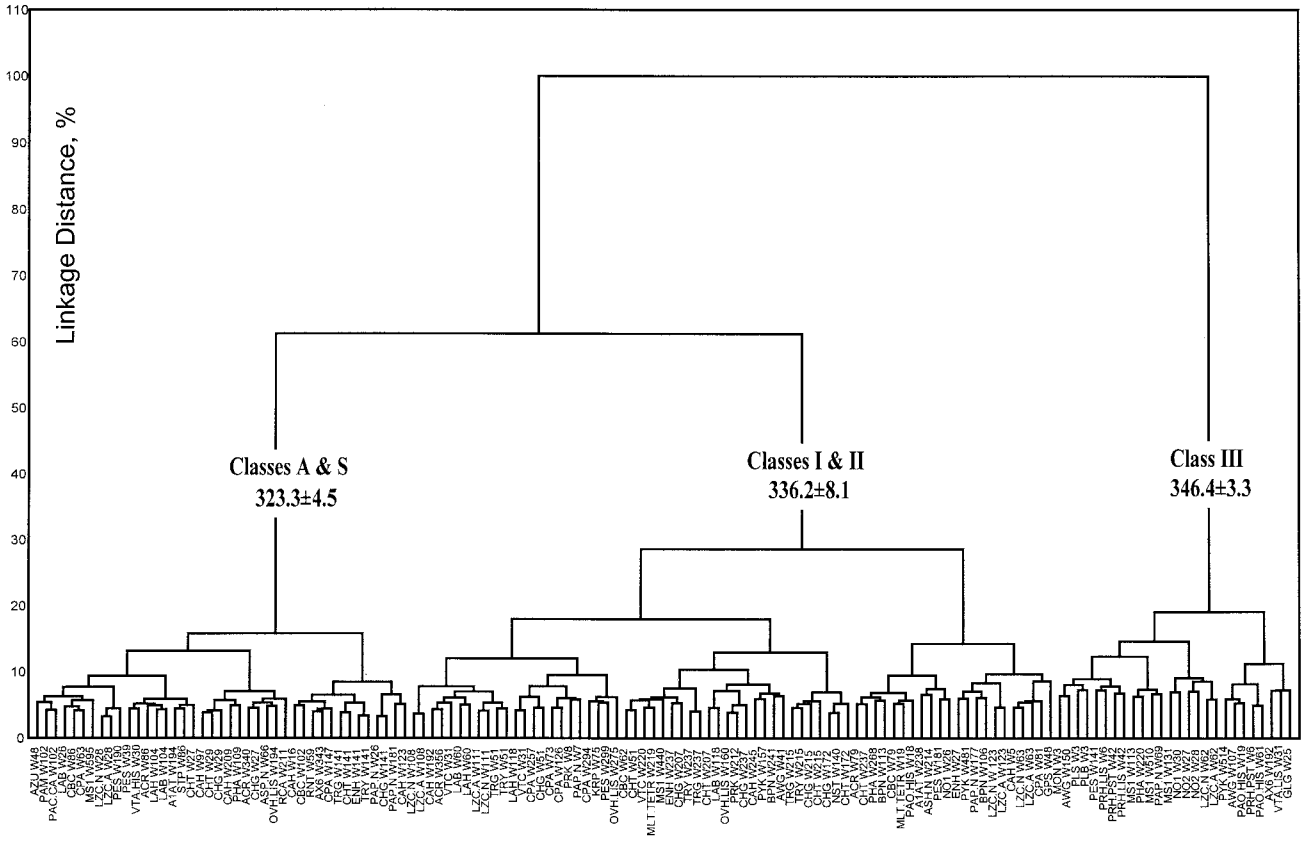

FIGURE 3 A dendrogram constructed based on six normalized structural parameters of 137 tryptophan residues of 48 proteins.

significance for the discrimination; therefore it is an appropriate measure of the impact brought by a parameter into classification. The program continues to choose parameters to be included in the model, as long as the respective F values for those parameters are >1.

Classification of tryptophan residues using discriminant analysis

To obtain the "training" dataset we used the preliminary assignment of tryptophan residues of proteins containing fewer than three tryptophan residues to five classes according to the model of discrete states (Burstein et al., 1973; Burstein, 1977a, 1983) using 1) the histograms of occurrence of protein fluorophores with various fluorescence maximum positions (Reshetnyak and Burstein, 2001), and 2) the classification reached in cluster analysis of microenvironment parameters of tryptophan fluorophores (Fig. 2). Class A contained only Trp-48 of azurin. Class S consisted of all residues of clusters A and S from the dendrogram in Fig. 2, excluding the Trp-48 of azurin. Class I contained the tryptophan residues of clusters I and II with $\lambda_m < 337$ nm, i.e., less than the deep minimum position in the histograms (Reshetnyak and Burstein, 2001), and class II contained those with $\lambda_m > 337$ nm. Class III consisted of all residues of cluster class III in the dendrogram. For the "training" data set we chose tryptophan residues with values of six environment parameters maximally similar to their mean values for a class, differing from the mean by less than two root-mean-square deviations. Then, in a cycle of several iterations of discriminant analysis, the "training" dataset was refined by changes in assignment and classification of some tryptophan residues. Later, the dataset was extended by including microenvironment parameters of fluorophores of proteins containing up to eight tryptophan residues. According to results of proceeding iteration, we obtained the final classification with a very low number of residues, which have λ_m values incompatible to the expected belonging to corresponding class.

The final results of discriminant analysis are presented in Tables 3 and 4. All six chosen structural parameters were included in the model. The column "F to enter" in Table 3 presents the values of F statistics for each parameter. The higher these values are, the stronger the discriminating power of the parameter. Wilks' lambda values associated with the unique contribution of the respective parameter to the discriminatory power of the model, and it may vary in the range from 0 (perfect discrimination) to 1 (no discrimination) (Rao, 1951).

The results of discriminant analysis allow us to conclude that the best discriminating parameter is *Den2* (F to enter is

TABLE 3 Summary of stepwise discriminant analysis

Parameters	F to Enter	Wilks' Lambda	*P*-Level
Den2	194.98	0.146	0.0000
Acc	26.15	0.081	0.0000
B_{av}	14.81	0.056	0.0000
R_{av}	8.63	0.044	0.0000
Acc1-7	2.86	0.041	0.0000
A_{av}	3.07	0.037	0.0000

maximal). The other five parameters also possess discriminatory power because the values of Wilks' lambda successively decreased from 0.146, when only parameter *Den2* was included into the model, to 0.037, when all six parameters were included. Table 4 brings up the squared Mahalanobis distances between the classes' centroids. The Mahalanobis distance extends the common Euclidean distance taking into account the correlations between parameters. If parameters are not correlated, the Mahalanobis distance coincides with the simple Euclidian measure. However, when parameters are correlated, the axes would become non-orthogonal. Thus, the Mahalanobis distance would adequately take correlations into account. The larger the distances, the farther the respective classes are apart from each other, and the more discriminatory power the current model possesses between the respective two classes. The most distant from the others is class III (distance of 40.39); the distances between other neighboring classes are smaller and rather similar (see Table 4).

The most important result of discriminant analysis is the derived classification of tryptophan residues. First the class centroids were found, and then the program calculated the Mahalanobis paired distances from each tryptophan residue to every other class centroid. A fluorophore was assigned to that class to which centroid it was closest in terms of Mahalanobis distance. Then, the program derived probabilities of classification of tryptophan residues, which are proportional to the distances. The classification and corresponding probabilities (column "Class (*P*, %)") for all individual tryptophan residues obtained at the last iteration are listed in Table 1. The mean values and the standard deviations of the microenvironment parameters for every class based on final classification of tryptophan residues are presented in Table 5. Boldface type indicates six parameters used in discriminant analysis.

TABLE 4 Squared Mahalanobis distances among classes

	Class A	Class S	Class I	Class II	Class III
Class A	—	5.00	14.30	31.27	111.61
Class S	5.00	—	6.29	21.58	97.89
Class I	14.30	6.29	—	4.88	65.36
Class II	31.27	21.58	4.88	—	40.39
Class III	111.61	97.89	65.36	40.39	—

Canonical variates

Discriminant analysis searches for the best discriminant functions (canonical variates or roots) that possess the maximal discriminatory power. The discriminant analysis computed four canonical variates for five classes, which are the independent linear combinations of investigated parameters. Table 6 presents the results of the step-down test (a sequential testing procedure) for the significance of the canonical roots (Mendoza et al., 1978). The first row contains the test parameters of significance for all four roots combined; the second row contains the significance of the remaining roots, after removing the first one (Root 1), and so on. The data collected in Table 6 allowed us to conclude that only Root 1 is a statistically significant canonical variate, e.g., the best discriminant function, and all further discussion would be directed to analyze the structure of Root 1. Table 7 presents the factor loading matrix, whose elements reflect the correlations of each parameter with the best canonical variate (generated by Root 1). They could be thought of as linear regression coefficients of the parameters on the canonical variate, which are interpreted in the same manner as factor coefficients in the principal component analysis. The factor structure coefficients have produced the conclusion that *Den2*, *Acc* and, to a lesser degree, R_{av} and *Acc1–7* defined Root 1 and, as a result, were mostly included in discrimination among the five classes. Fig. 4 *A* is an illustration of the discrimination by the best canonical variate (Root 1) of 137 tryptophan residues of 48 proteins among five classes in canonical coordinates (Root 3 vs. Root 1). This figure demonstrates rather well separated clouds of the tryptophan residues belonging to different classes. Despite the fact that clouds are partly overlapped, the centers of each class are well distant (see Fig. 4 *B* and Table 8). Analogous discrimination of five discrete classes of tryptophan residues was obtained in discriminant analysis applied to all 18 microenvironment parameters listed in Table 5.

Pairwise discriminant analyses for "neighbor" fluorophore classes

To refine the classification obtained on the full set of fluorophores and to find out which of the structural parameters determine the discrimination between neighbor classes, we applied discriminant analysis for the tryptophan residues belonging to pairs of neighboring classes by turn. Discriminant analyses performed for the pairs of neighbor classes confirmed the classification obtained on the full set of tryptophan residues (data not shown). The most significant factor-loading coefficients are presented in Table 9 for each model separately. The tryptophan residues of class A are separated from other classes mainly by polarity, A_{av}, and flexibility, B_{av}, of microenvironment. The discrimination between classes S and I occurs mainly due to the surrounding packing density,

TABLE 5 Mean values and standard deviations of the structural parameters

Parameters	Class A	Class S	Class I	Class II	Class III
Number of Trp in Classes	1	41	32	41	22
1 Accw1	0	0.2 ± 1.0	8.7 ± 11.9	17.62 ± 22.01	57.70 ± 28.58
2 Acc1	0	1.0 ± 2.4	14.9 ± 14.2	28.5 ± 23.9	69.2 ± 21.8
3 Accw7	0	0.5 ± 2.1	4.1 ± 6.4	18.2 ± 19.9	64.9 ± 27.9
4 Acc7	0	1.1 ± 3.2	7.4 ± 10.1	25.0 ± 21.4	73.0 ± 20.6
5 Accw	1.9	0.2 ± 0.5	3.5 ± 3.1	10.2 ± 7.6	50.6 ± 18.5
6 Acc	**1.9**	**0.8 ± 1.4**	**6.0 ± 3.6**	**14.8 ± 7.5**	**55.3 ± 15.9**
7 Den1	58.3	67.1 ± 6.7	61.7 ± 4.8	52.3 ± 7.6	28.5 ± 11.0
8 Den2	**138.3**	**148.3 ± 8.5**	**129.3 ± 9.1**	**109.3 ± 12.6**	**62.7 ± 18.8**
9 B1	0.54	0.85 ± 0.18	1.21 ± 0.24	1.29 ± 0.34	1.46 ± 0.52
10 B2	0.68	0.92 ± 0.17	1.10 ± 0.20	1.18 ± 0.33	1.61 ± 0.65
11 R1	0.1	0.7 ± 1.2	7.2 ± 4.2	19.0 ± 10.3	80.7 ± 26.6
12 R2	0.1	0.7 ± 1.3	6.6 ± 3.8	17.5 ± 10.5	89.0 ± 40.5
13 A1	19.6	31.9 ± 6.1	36.1 ± 7.4	41.6 ± 9.5	58.7 ± 12.9
14 A2	27.3	37.0 ± 6.1	42.5 ± 4.8	48.5 ± 6.8	72.3 ± 18.9
15 Nac1	1	4.4 ± 2.6	3.5 ± 1.6	3.0 ± 1.5	1.2 ± 0.9
16 Nac2	2.3	5.8 ± 2.4	5.0 ± 2.1	4.3 ± 2.3	1.9 ± 1.5
17 Ndon1	2	1.6 ± 1.3	2.2 ± 1.4	1.8 ± 1.2	1.2 ± 0.9
18 Ndon2	6.4	3.0 ± 1.4	3.0 ± 1.2	2.6 ± 1.4	1.1 ± 1.0
Acc1-7	**0.0**	**1.0 ± 2.2**	**11.2 ± 8.5**	**26.7 ± 19.1**	**71.1 ± 19.5**
B_{av}	**0.61**	**0.89 ± 0.17**	**1.11 ± 0.20**	**1.23 ± 0.32**	**1.54 ± 0.55**
R_{av}	**0.9**	**0.7 ± 1.2**	**6.7 ± 4.0**	**18.2 ± 10.3**	**85.2 ± 30.9**
A_{av}	**23.5**	**34.5 ± 5.8**	**39.3 ± 5.5**	**45.1 ± 7.4**	**65.5 ± 13.9**

Six parameters used in discriminant and cluster analyses are shown in bold.

Den2, and, to a lesser degree, the solvent accessibility of hypothetical hydrogen-bonding atoms, *Acc1–7*. In the case of classes I and II, all six parameters were included in the model; however, packing density, *Den2*, total solvent accessibility, *Acc*, and "dynamic accessibility" parameter R_{av} define the discriminant function to the highest degree. The total *Acc* and its related *Den2* evidently separate the most distant class III.

Cluster analysis applied to the best canonical variate

The application of cluster analysis to six normalized parameters of tryptophan residues (Figs. 2 and 3) proved the discrete nature of microenvironment parameters, although classes I and II were mixed. Discriminant analysis yielded the classification of tryptophan residues and provided an optimal combination of microenvironment parameters via the best canonical variate (Root 1). After that, it was important to apply the cluster analysis for this canonical variate. Such cluster analysis revealed four separate clusters (Fig. 5). Class A consists of one object, and therefore could not be identified as an individual cluster. The most distant is class III (100% of linkage distance). Class S is distanced from classes I and II by ~85%. Classes I and II, which were mixed before, now form separate clusters with a relative distance between them of ~47%. The optimal combination of microenvironment parameters (discriminant function Root 1) obtained in discriminant analysis allowed us to separate all five classes.

TABLE 6 χ^2 test for the canonical variates

	Eigenvalue	Canonical R	Wilks' Lambda	χ^2	*P*-Level
1	**10.92**	**0.957**	**0.037**	**433.3**	**0.0000**
2	1.10	0.724	0.442	107.4	0.0000
3	0.07	0.254	0.929	9.7	0.2891
4	0.01	0.082	0.993	0.9	0.8282

RELATIONSHIP BETWEEN FLUORESCENCE AND MICROENVIRONMENT PARAMETERS

As described above, statistical methods allowed us to reveal the existence of discreteness in the set of physical and structural parameters of the microenvironment of tryptophan residues in proteins in parallel with discreteness of their fluorescence spectral properties. Moreover, these methods provided the classification of the tryptophan resi-

TABLE 7 Factor structure coefficients of Root 1

Parameters	Factor Structure Coefficients
Den2	**0.73**
Acc	**−0.72**
B	−0.22
R	**−0.61**
Acc1-7	**−0.53**
A	−0.40

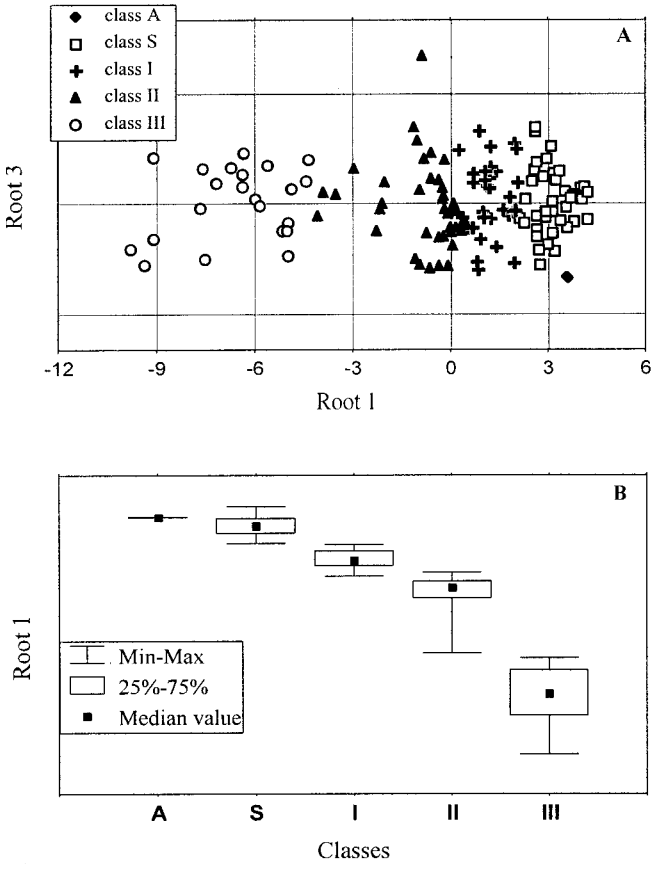

FIGURE 4 (*A*) Discrimination of 137 tryptophan residues of 48 proteins presented in canonical coordinates (Root 3 vs. Root 1). (*B*) Box plot-style picture of the central tendency (median) and range (quartiles) of Root 1 in five classes.

dues based on their microenvironment properties, which practically coincided with the classification made using their emission features. The goal of our investigation was to see how the structural parameters correlate with spectrofluorescent ones. Fig. 6 *A* illustrates the correlation between values of Root 1, which is a combination of six microenvironment parameters, and emission spectral maximum positions taken from the results of assignments of log-normal components to individual protein fluorophores (Table 1, column "λ_m (nm)"; values in parentheses in the body of the table were excluded from consideration because they belong to tryptophan residues partially quenched by neighbor groups or due to energy homo-transfer). The plot

TABLE 8 Mean values for Root 1

Classes	Means of Root 1
Class A	3.62
Class S	3.17
Class I	1.22
Class II	−0.78
Class III	−6.55

TABLE 9 Factor structure coefficients of the parameters included in the models based on discriminant analyses performed for neighbor classes

	Classes A & S	Classes S & I	Classes I & II	Classes II & III
Acc1-7	—	**−0.46**	−0.41	—
Acc	—	—	**−0.58**	**0.82**
Den2	—	**0.58**	**0.72**	**−0.71**
B_{av}	**−0.69**	−0.33	−0.18	−0.17
R_{av}	—	—	**−0.57**	—
A_{av}	**−0.78**	−0.23	−0.35	—

shows that fluorophores belonging to various structural classes revealed by statistical methods have different values of spectral maximum positions. Although the clouds of various classes are partly overlapped, the centers of each class are well distant, as is seen in Fig. 6 *B*, where the averaged maximum positions are shown versus classes revealed using parameters of microenvironment. Calculated mean values of spectral maximum positions, λ_m, and relative accessibility to quenchers, K_{rel}, for tryptophan residues of different classes are collected in Table 10.

Then, we constructed five histograms reflecting distributions of maximum positions of spectral components assigned to tryptophan residues from each structure-based class (Fig. 7). The "Total" histogram is a sum of those five. Each histogram was constructed as a sum of individual Gauss distributions (for the details of the constructing of histograms see Reshetnyak and Burstein, 2001) with mode of each individual Gauss curve coinciding with the maximum position of spectral components (see Table 1, λ_m (nm)). Dispersions and maximal amplitudes were taken as 1.0 and 1.0 nm, respectively. The "Total" histogram is rather similar to that constructed in our previous paper for >300 spectral components (Figs. 3–5 in Reshetnyak and Burstein, 2001). The global minimum is at 332–336 nm and five maxima at ~308, 317, 327, 340, and 346 and two shoulders at 338–339 and 350–353 nm are seen here. The curves representing the spectral maximum position distributions in each class are partly overlapped (especially, those for classes S and I, and classes II and III). Unexpectedly, the distributions for all classes (except class A) have polymodal character, which has to be investigated in further work.

We have to note also that tryptophan residues of five highly homologous proteins (trypsin, trypsinogen, chymotrypsin, chymotrypsinogen, and elastase) have maximum positions of one of spectral components at 327–329 nm. In chymotrypsin and chymotrypsinogen the fluorophores with such spectral maximum positions are parts of large "clusters" of fluorophores with highly efficient energy migration inside them (see Fucaloro and Forster, 1985; Desie et al., 1986; Reshetnyak and Burstein, 1997b). Therefore, the emission of these tryptophan residues might be quenched; however, we did not exclude them because it is not evident

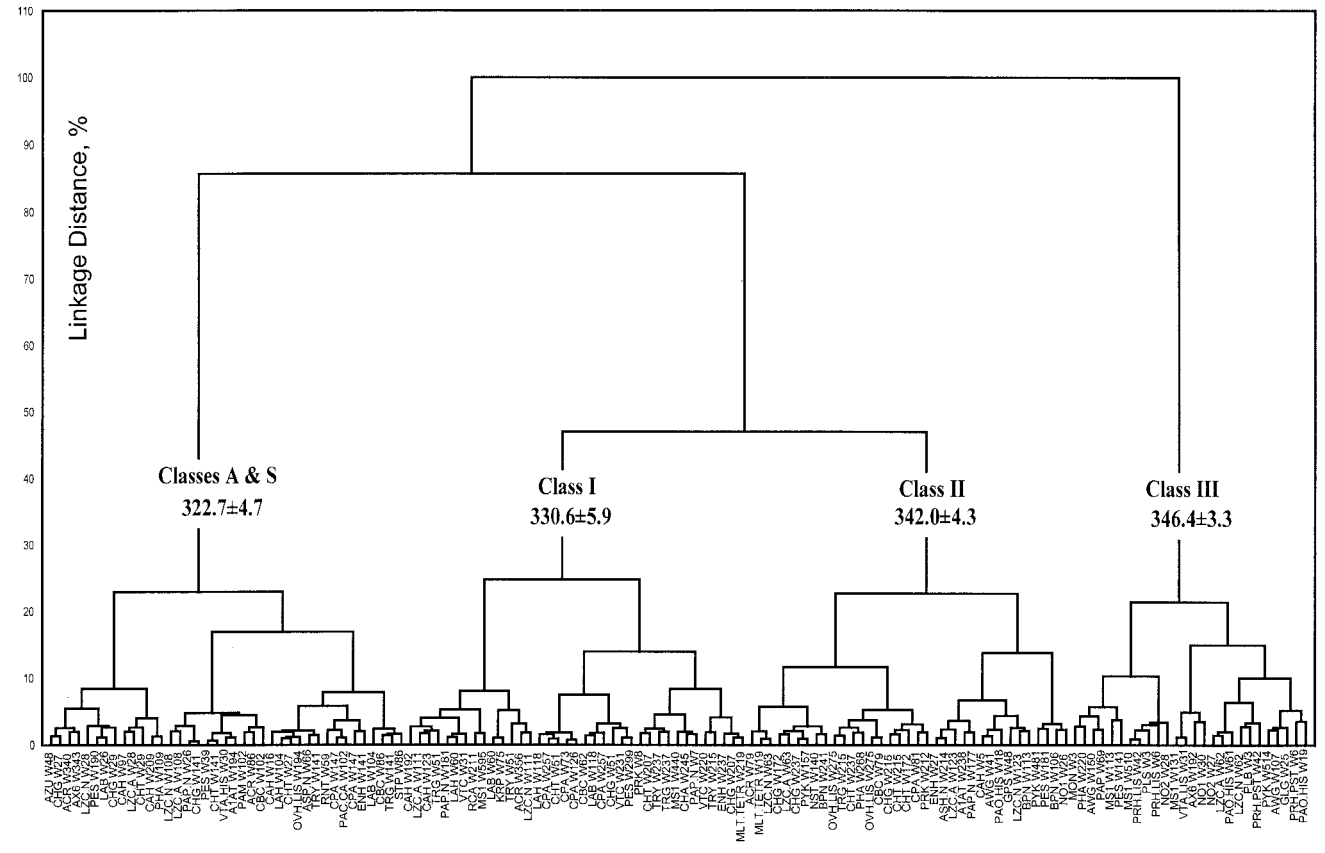

FIGURE 5 A dendrogram constructed based on canonical variate (Root 1).

which of the fluorophores could be emitting energy transfer acceptors in the cluster.

DISCUSSION

In our previous paper (Reshetnyak and Burstein, 2001), we succeeded in confirming the assumed (1973–1977; Burstein et al., 1973; Burstein, 1977a) discreteness of spectral parameters by analyzing the distribution of maximum positions of log-normal components obtained by decomposition of tryptophan fluorescence spectra of ~100 proteins and some of their conformers. Because the variety of fluorescence properties of tryptophan residues in proteins has to be a result of combinations of various excited-state interactions of individual fluorophores with their environment, we investigated some physical and structural characteristics of microenvironment of 137 individual fluorophores in 48 proteins with known atomic structure and compared them with spectral parameters of log-normal components assigned to individual fluorophores. Now we can consider features of photophysical interactions inherent in fluorophores of various classes, based on newly identified microenvironment parameters, which define the discrimination between classes.

In the model of discrete states (Burstein et al., 1973; Burstein, 1977a, 1983) it was assumed that the well-structured spectra with main peak at 307–308 nm of class A fluorophores (Trp-48 of azurin (Burstein et al., 1977; Burstein, 1977b)) and some tryptophan residue(s) of bacteriorhodopsin (Permyakov and Shnyrov, 1983) are emitted from the excited state, which has no hydrogen bond with neighbors. Analysis of vibrational structure of azurin emission spectrum revealed that emission occurs from the ^1L_b singlet excited state (Callis, 1997). Our analysis shows that the single Trp-48 of azurin is located in a rather hydrophobic environment. Several atoms surrounding its indole ring, which might eventually form hydrogen bonds (see Table 2), are involved into stabilization of the β-barrel secondary structure of azurin and, indeed, cannot interact with excited fluorophore without breaking β-sheets. The fluorescence study performed on two azurin mutants (I7S and F110S) in holoform showed the red shift up to 312–313 nm of spectral maximum position (Gilardi et al., 1994). Analysis of x-ray structures of these mutants revealed that the Oγ atom of Ser-7 located at 4.3 Å from the Nϵ1 atom of Trp-48 provides a possibility of hydrogen bond formation. However, mutation F110S creates an additional cavity near Trp-48 with several water molecules included inside (Hammann et

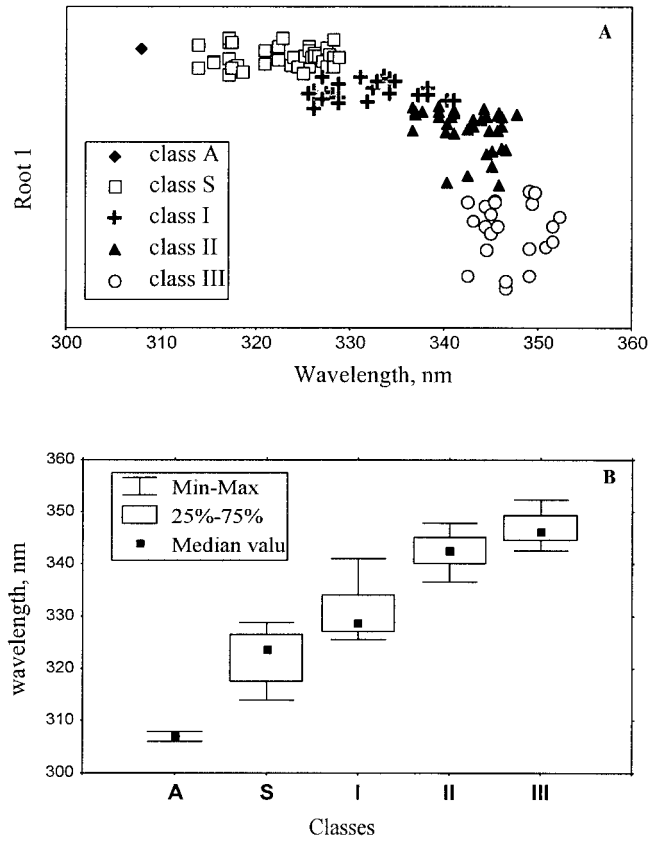

FIGURE 6 (*A*) The dependence of canonical variate (Root 1), which was calculated based on structural parameters of the microenvironment of tryptophan residues versus spectral maximum positions of log-normal components assigned to individual protein fluorophores. (*B*) Box plot-style picture of the central tendency (median) and range (quartiles) of spectral maximum positions in five structure-based classes.

al., 1996). Such changes around Trp-48 in two mutated proteins might increase in probability of hydrogen-bonded exciplex formation and, as a result, induce the lead to shift in tryptophan emission to longer wavelengths.

Based on the data of spectrofluorometric titration of 3-methyl-indole in apolar solvent with alcohols and other polar co-solvents (Walker et al., 1967; Lumry and Hershberger, 1978; Hershberger et al., 1981), fluorophores of class S, having structured spectra with main peaks at ~317 nm, has been assumed to form hydrogen-bonded exciplexes in 1:1 stoichiometry (Burstein, 1977a, b, 1983). Emission spectra of classes A and S do not undergo any spectral shift by freezing the protein solutions down to −196°C, i.e., the dipole moments of these fluorophores are scarcely changed at excitation and no dipole relaxation of dielectric medium can be observed (Burstein, 1977a, 1983). In the present work, the mean value of the maximum positions of spectral components of fluorophores of structure-based class S are shown to be at 322.5 ± 4.6 which is, however, longer than 317 nm. Approximation of structured spectra with central peaks at 316–317 nm using smooth log-normal function usually revealed effective maximum positions of spectra at 320–326 nm. However, it seems that tryptophan residues assigned to the structure-based class S may emit at 320–325 nm as well. Class S is best discriminated from class A by higher relative polarity (A_{av}) and flexibility (B_{av}) of microenvironment (see Tables 5 and 9). Also, there are evident free partners for hydrogen bond formation near the fluorophores of class S.

The model of discrete states assumed that tryptophan residues possessing emission spectra with maxima at wavelengths ~330 nm and longer (classes I–III) form H-bonded exciplexes with stoichiometry not less than 2:1 (Burstein, 1977a, b, 1983). This assumption was also based on the experiments with 3-methyl-indole in cycloheptane at varying concentrations of alcohol (Walker et al., 1967; Lumry and Hershberger, 1978; Hershberger et al., 1981). The maximum position of the exciplex 2:1 spectrum was red-shifted upon increasing alcohol concentration, which reflected the rise of a Stokes shift induced by the universal solvent dipole relaxation in response to a large change of fluorophore dipole moment in its excited state, according to the Bakhshiev's theory of universal interactions (Bakhshiev, 1972; Mazurenko, 1973; Mazurenko and Udaltsov, 1978).

In this study we found that fluorophores of class I have averaged maximum position of fluorescence at 331.0 ± 4.8 nm, which excellently coincides with that in the model of discrete states. The parameters, mostly discriminating class S and I, are packing density (*Den2*) and averaged accessibility (*Acc1–7*) of Nε1 and Cζ2 atoms (Tables 5 and 9). It is known that in most cases the indole ring of tryptophan residues in proteins are oriented toward polar environment (water molecules) by their Nε1, Cη2, Cζ2, and Cε2 atoms. Our investigation of the accessibility of individual atoms of indole rings confirmed this statement (data not shown). Thus, it seems that interaction of Nε1 and Cζ2 atoms of excited indole rings of tryptophan residues with their surrounding (especially with water molecules) is very important, and possibly just these two atoms may be the best candidates for hydrogen bond formation in the excited state. Moreover, the recent quantum mechanical studies indicated

TABLE 10 Mean values and standard deviations of the spectral parameters in each class based on classification of tryptophan residues obtained in discriminant analysis

Parameters	Class A	Class S	Class I	Class II	Class III
λ_m (nm)	307.9	322.5 ± 4.6	331.0 ± 4.8	342.3 ± 3.3	347.0 ± 3.1
K_{rel}	0.0	0.10 ± 0.11	0.11 ± 0.16	0.44 ± 0.44	0.76 ± 0.71

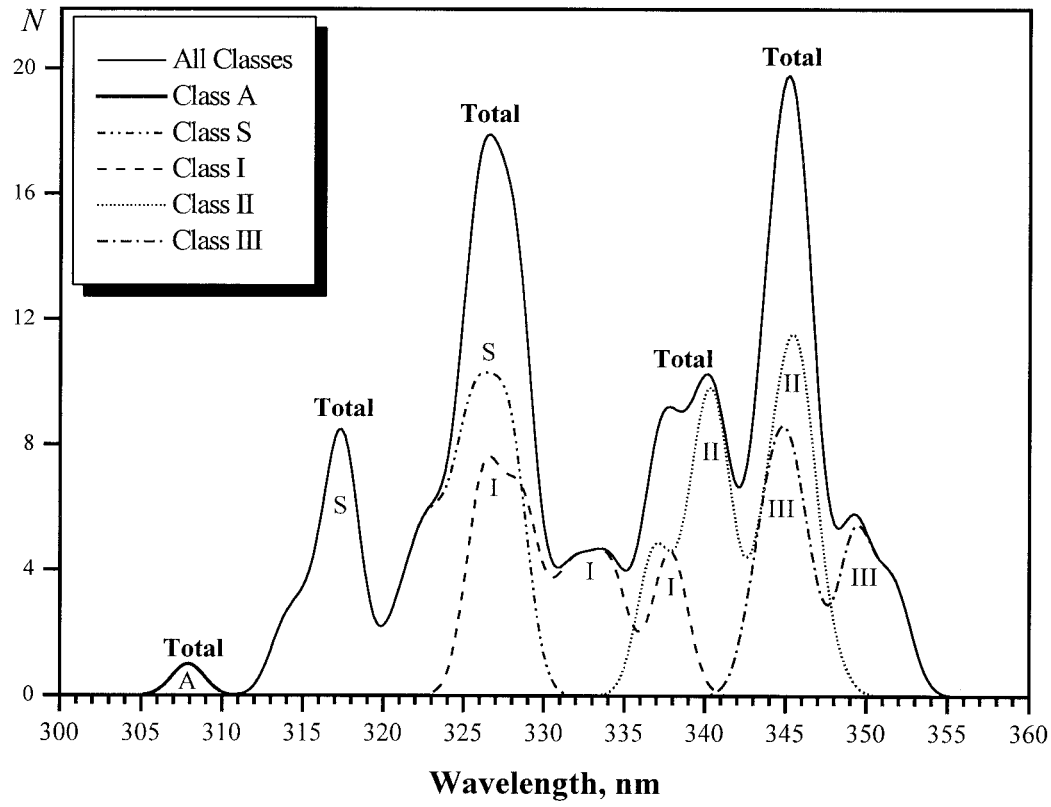

FIGURE 7 The distributions of spectral components assigned to tryptophan residues from different structure-based classes.

that electron density is mostly lost from Nϵ1 and Cγ atoms and mostly rises at the Cϵ3, Cζ2, and Cδ2 atoms of the indole ring during excitation in the ^{1}La state (Callis, 1997). The averaged number of atoms in the 7.5 Å layer (packing density) around tryptophan residues of classes S and I are ~148 and 129, respectively. Because the total accessibility of indole rings of class I is very small (~6%, see Table 5), the decrease in *Den2* in class I could not be associated with the presence of many free water molecules near the fluorophores. Such a lowering of the packing density of the environments of tryptophan residues of class I compared with those of class S may result in an increase in frequency and/or amplitude of structural mobility of the environment favoring both hydrogen-bonded exciplex formation and dipole relaxation during the lifetime of fluorophore excited state.

It was suggested that tryptophan residues of class II are partially exposed to solvent (Burstein, 1977a, b, 1983). Here we estimated the mean value of maximum position of fluorescence of the structure-based class II fluorophores at 342.3 ± 3.3 nm. However, unexpectedly, the statistical analysis revealed that microenvironment parameters of tryptophan residues of classes I and II are overlapped, while all histograms of spectral maximum positions (Reshetnyak and Burstein, 2001) gave the deep global minimum just between these classes. The parameters best discriminating classes I and II are the packing density, *Den2*, total solvent accessibility, *Acc*, and "dynamic accessibility," R_{av}. The increase in accessibility and decrease in packing density around tryptophan residues belonging to class II evidently confirmed their partial exposure to solvent. However, ~31% of total solvent accessibility of class II fluorophores is due to their contact with bound water molecules (compare *Accw* and *Acc* values in Table 5). High-resolution NMR measurements of protein hydration in aqueous solutions showed that there exist two qualitatively different types of hydration: 1) the water molecules in the protein interior, identically located in both crystals and solution with residence times in the range 10^{-8} up to 10^{-2} s, and 2) the water molecules forming outer hydration shells at the protein surface and having residence times in the range from >10 ns down to subnanoseconds (Otting et al., 1991). Water appears to act as a plasticizer for the protein polymer, giving the globule surface almost fluid-like properties (Vanderkooi, 1998). The mobility of water molecules included in x-ray structures was shown to correlate with values of their B-factors and occupancies in crystals (Sanschagrin and Kuhn, 1998; Craig et al, 1998). The increase in dynamic accessibility for class II fluorophores suggests that their environment, which consists of protein atoms and bound water molecules, is essentially more flexible than the environment of class I indole rings. These structural characteristics are in good

agreement with the high accessibility of class II fluorophores to external small fluorescence quenchers, in contrast to those of class I (Burstein et al., 1973; Burstein, 1977a, b).

The microenvironment characteristics of fluorophores of classes I and II allow us to assume that emission of class II fluorophores takes place after completion of relaxation process of dipoles surrounding excited fluorophores, in contrast to fluorophores of class I, which emit from the unrelaxed state. Such an interpretation can explain the existence of a deep gap in occurrence histograms between these two classes by very low statistical probability (statistical weights) of a situation where dipole relaxation time of environment, τ_r, is comparable with the fluorophore excited-state lifetime, τ_{em}, in contrast to situations with $\tau_r > \tau_{em}$ (class I) or $\tau_r < \tau_{em}$ (class II). This hypothesis can be checked in experiments with time dependencies of position of time-resolved instant emission spectra during excitation lifetimes measured with proteins representing steady-state fluorescence spectra of classes I or II. If the hypothesis is valid, such time-resolved emission spectra of class II fluorophores, in contrast to those of class I, are expected to experience the effective relaxation-induced shift to longer wavelengths during excited-state lifetimes.

The model of discrete states assumed that tryptophan residues of class III, possessing spectra almost coinciding in position and shape with those of free aqueous tryptophan or completely unfolded proteins, are highly exposed to solvent (Burstein et al., 1973; Burstein, 1977a, 1983). Indeed, the mean emission maximum position of fluorophores of class III in the structure-based classification is 347.0 $\pm$ 3.1 nm (Table 1), which is close to the maximum position of free aqueous tryptophan solution (352 nm). In both cluster and discriminant analyses, class III is the best-discriminated from other classes. The primary separating parameter for this class from class II is the total accessibility, *Acc*; i.e., the class III fluorophores are well exposed to the bulk water, possessing a high density of large and rapidly relaxing dipoles able to provide very fast (hundreds of femtoseconds, Muiño and Callis, 1994; Callis, 1997) shift of emission spectra to a completely relaxed position.

The difference between the two accessible classes II and III may be interpreted as a result of essential distinction in dipole densities and relaxation rates of fluorophore dielectric environment. Although the ratios τ_r/τ_{em} are less than unity for both class II and III, the bulk water surrounding class III fluorophores contains a great density of large dipoles relaxing in the subpicosecond time range, but the dipoles surrounding the fluorophores of class II belong mostly to protein groups and bound water that are much less abundant and, therefore, cannot induce long-wavelength relaxation shifts, even after relaxation completion. Moreover, the relaxation-rate-limiting mobility of protein and bound-water dipoles is much lower than of those in the bulk water.

Statistical methods allowed us to reveal the structure-based classification of tryptophan residues using six physical and structural microenvironment parameters. However, we assume that these six chosen parameters may not be exhaustive enough for the overwhelming description of all processes of excited fluorophores of different spectral classes. It seems to be reasonable to more carefully analyze all 18 parameters in the future, and maybe include several additional characteristics that would reflect the magnitude and direction of protein electric field, since Callis (1997) suggested that the fluorescence could be strongly blue-shifted if the local electric field is oriented against the tryptophan dipole change, and vice versa. Moreover, the quantitative accounting of eventual hydrogen-bonding partners near fluorophore might become very informative characteristics, discriminating members of classes A, S, and I.

The authors thank Dr. Boris P. Atanasov (Institute of Organic Chemistry, Bulgarian Academy of Sciences, Sofia, Bulgaria) for 3D models of viperotoxin subunits; Drs. Monique Laberge and Jane Vanderkooi (University of Pennsylvania, Philadelphia, PA) for the 3D model of cod parvalbumin; Dr. Julian Borejdo (University of North Texas Health Science Center) for support; Dr. Robert V. Polozov (Institute of Theoretical and Experimental Biophysics, Pushchino, Russia); Dr. K. K. Turoverov (Institute of Cytology, St. Petersburg, Russia), and Dr. Oleg A. Andreev (University of North Texas Health Science Center) for stimulating consultations and discussion.

This work was supported in part by Grants 95-04-12935, 97-04-49449, and 00-04-48127 from Russia Foundation of Basic Research.

REFERENCES

Abornev, S. M., and E. A. Burstein. 1992. Resolution of protein tryptophan fluorescence spectra into elementary components. *Molecular Biology (Moscow).* 26:890–987 [In Russian; English translation].

Andrade, M. A., S. I. O'Donoghue, and B. Rost. 1998. Adaptation of protein surfaces to subcellular location. *J. Mol. Biol.* 276:517–525.

Bakhshiev, N. G. 1972. Spectroscopy of Intermolecular Interactions. Nauka, Leningrad [in Russian].

Bernstein, F. C., T. F. Koetzle, G. J. Williams, E. E., Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. 1977. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112:535–542.

Bhaskaran, R., M. Prabhakaran, G. Jayaraman, C. Yu, and P. K. Ponnuswamy. 1996. Internal packing conditions and fluctuations of amino acid residues in globular proteins. *J. Biomol. Struct. Dyn.* 13:627–639.

Bilot, L., and A. Kawski. 1962. Zur Theorie des Einflüsses von Lösungsmitteln auf die Elektronenspektren der Moleküle. *Z. Naturforsch.* 17a:621–627.

Brown, M. F., S. Omar, R. A. Raubach, and T. Schleich. 1977. Quenching of the tyrosyl and tryptophyl fluorescence of subtilisins Carlsberg and Novo by iodide. *Biochemistry.* 16:987–992.

Burgess, S. A. 1995. Rigor and relaxed outer dynein arms in replicas of cryofixed motile flagella. *J. Mol. Biol.* 250:52–63.

Burstein, E. A. 1961. Luminescence of aromatic amino acids and proteins in solutions at excitation in the far ultraviolet region. *Biophysics (Moscow).* 6:753–763 [In Russian; English translation].

Burstein, E. A. 1976. Luminescence of protein chromophores (model studies). *In* Advances in Science and Technology (Itogi Nauki i Tekhniki), ser. Biophysics, vol. 6. VINITI, Moscow. [In Russian].

Burstein, E. A. 1977a. Intrinsic protein luminescence (the nature and application). *In* Advances in Science and Technology (Itogi Nauki i Tekhniki), ser. Biophysics, vol. 7. VINITI, Moscow. [In Russian].

Burstein, E. A. 1977b. The study of fast dynamics of protein structure using the intrinsic fluorescence methods. *In* Equilibrium Dynamics of Native Protein Structure. E. A. Burstein, editor. Center of Biological Research, RAS, Pushchino. 60–83. [In Russian].

Burstein, E. A. 1983. The intrinsic luminescence of proteins is a method for studies of the fast structural dynamics. *Molecular Biology (Moscow).* 17:455–467 [In Russian; English translation].

Burstein, E. A., S. M. Abornev, and Ya. K. Reshetnyak. 2001. Decomposition of protein tryptophan fluorescence spectra into log-normal components. I. Decomposition algorithms. *Biophys. J.* 81:1699–1709.

Burstein, E. A., and V. I. Emelyanenko. 1996. Log-normal description of fluorescence spectra of organic fluorophores. *Photochem. Photobiol.* 64:316–320.

Burstein, E. A., E. A. Permyakov, V. A. Yashin, S. A. Burkhanov, and A. Finazzi Agrò. 1977. The fine structure of luminescence spectra of azurin. *Biochim. Biophys. Acta.* 491:155–159.

Burstein, E. A., N. S. Vedenkina, and M. N. Ivkova. 1973. Fluorescence and the location of tryptophan residues in protein molecules. *Photochem. Photobiol.* 18:263–279.

Bushueva, T. L., E. P. Busel, and E. A. Burstein. 1975. The interaction of protein functional groups with indole chromophore. III. Amine, amide, and thiol groups. *Studia Biophys.* 52:41–52.

Bushueva, T. L., E. P. Busel, V. N. Bushueva, and E. A. Burstein. 1974. The interaction of protein functional groups with indole chromophore. I. Imidazole group. *Studia Biophys.* 44:129–132.

Callis, P. R. 1997. ^1L_a and ^1L_b transitions of tryptophan: applications of theory and experimental observations to fluorescence of proteins. *Methods Enzymol.* 278:113–151.

Carugo, O., and P. Argos. 1997. Correlation between side chain mobility and conformation in protein structures. *Protein Eng.* 10:777–787.

Carugo, O., and P. Argos. 1998. Accessibility to internal cavities and ligand binding sites monitored by protein crystallographic temperature factors. *Protein.* 31:210–213.

Carugo, O., and D. Bordo. 1999. How many water molecules can be detected by protein crystallography? *Acta Crystallogr. D.* 55:479–483.

Chen, Yu, and M. D. Barkley. 1998. Toward understanding tryptophan fluorescence in proteins. *Biochemistry.* 37:9976–9982.

Chou, K. C., and D. W. Elrod. 1998. Using discriminant function for prediction of subcellular location of prokaryotic proteins. *Biochem. Biophys. Res. Commun.* 252:63–68.

Cowgill, R. W. 1970. Fluorescence and the structure of proteins. XVIII. Spatial requirements for quenching by disulfide groups. *Biochim. Biophys. Acta.* 207:556–559.

Craig, L., P. C. Sanschagrin, A. Rozek, S. Lackie, L. A. Kuhn, and J. K. Scott. 1998. The role of structure in antibody cross-reactivity between peptides and folded proteins. *J. Mol. Biol.* 281:183–201.

Dale, R. E., and J. Eisinger. 1974. Intramolecular distances determined by energy transfer. Dependence on orientational freedom of donor and acceptor. *Biopolymers.* 13:1573–1605.

Desie, G., V. Boens, and C. D. Schryver. 1986. Study of the time-resolved tryptophan fluorescence of crystalline alpha-chymotrypsin. *Biochemistry.* 25:8301–8308.

Dolashka, P., I. Dimov, N. Genov, I. Svendsen, K. S. Wilson, and C. Betzel. 1992. Fluorescence properties of native and photooxidised proteinase K: the x-ray model in the region of the two tryptophans. *Biochim. Biophys. Acta.* 1118:303–312.

Duggan, D. E., and S. Udenfriend. 1956. The spectrofluorometric determination of tryptophan in plasma and of tryptophan and tyrosine in protein hydrolysates. *J. Biol. Chem.* 223:313.

Edsall, J. T., and H. A. McKenzie. 1983. Water and proteins. II. The location and dynamics of water in protein systems and its relation to their stability and properties *Adv. Biophys.* 16:53–183.

Evans, P. A., C. M. Dobson, R. A. Kautz, G. Hatfull, and R. O. Fox. 1987. Proline isomerism in staphylococcal nuclease characterized by NMR and site-directed mutagenesis. *Nature.* 329:266–268.

Fox, R. O., P. A. Evans, and C. M. Dobson. 1986. Multiple conformations of a protein demonstrated by magnetization transfer NMR spectroscopy. *Nature.* 320:191–194.

Frauenfelder, H., G. A. Petsko, and D. Tsernoglou. 1979. Temperature-dependent x-ray diffraction as a probe of protein structural dynamics. *Nature.* 280:558–563.

Fucaloro, A. F., and L. S. Forster. 1985. Conformational fluctuations in alpha-chymotrypsinogen A powders. *Photochem. Photobiol.* 41:91–93.

Giacovazzo, C. 1992. Fundamentals of Crystallography. Oxford Univ. Press, Oxford.

Gilardi, G., G. Mei, N. Rosato, G. W. Canters, and A. Finazzi Agrò. 1994. Unique environment of Trp-48 in *Pseudomonas aeruginosa* azurin as probed by site-directed mutagenesis and dynamic fluorescence spectroscopy. *Biochemistry.* 33:1425–1432.

Glusker, J. P., M. Lewis, and M. Rossi. 1994. Crystal Structure Analysis for Chemists and Biologists. VCH, New York.

Gnanadesikan, R. 1977. Methods for Statistical Data Analysis of Multivariate Observations. John Wiley, New York.

Gray, P. M. D., G. J. L. Kemp, C. J. Rawlings, N. P. Brown, C. Sander, J. M. Thornton, C. M. Orengo, S. J. Wodak, and J. Richelle. 1996. Macromolecular structure information and database. *Trends Biochem. Sci.* 21:251–256.

Hammann, Ch., A. Messerschmidt, R. Huber, H. Nar, G. Gilardi, and G. W. Canters. 1996. X-ray crystal structure of the two site-specific mutants Ile7Ser and Phe110Ser of azurin from *Pseudomonas aeruginosa. J. Mol. Biol.* 255:362–366.

Hershberger, M. V., R. Lumry, and R. Verrall. 1981. The 3-methylindole/*n*-butanol exciplexes evidence for two exciplex sites in indole compounds. *Photochem. Photobiol.* 33:609–617.

Hogue, C. W. V., H. Ohkawa, and S. H. Bryant. 1996. A dynamic look at structures: WWW-entrez and the molecular modeling database. *Trends Biochem. Sci.* 21:226–229.

Ikura, T., G. P. Tsurupa, and K. Kuwajima. 1997. Kinetic folding and *cis/trans* prolyl isomerization of staphylococcal nuclease. A study by stopped-flow absorption, stopped-flow circular dichroism, and molecular dynamics simulations. *Biochemistry.* 36:6529–6538.

Islamov, A. S., A. B. Rumjantsev, V. S. Skvortsov, and A. I. Archakov. 1997. ONIX: an interactive PC program for the examination of protein 3D structures from PDB. *Comput. Appl. Biosci.* 13:111–113.

Konev, S. V. 1957. Fluorescence spectra and fluorescence action spectra of several proteins. *Proc. Acad. Sci. USSR.* 116:594–597. [In Russian].

Konev, S. V. 1959. Fluorescence action spectra of proteins. *Trans. Acad. Sci. USSR.* 23:90–93. [In Russian].

Konev, S. V. 1967. Fluorescence and Phosphorescence of Proteins and Nucleic Acids. Plenum Press, New York.

Kuznetsova, I. M., and K. K. Turoverov. 1998. What determines the characteristics of the intrinsic UV-fluorescence of proteins? Analysis of the properties of the microenvironment and features of the localization of their tryptophan residues. *Tsitologiia.* 40:747–762. [In Russian].

Kuznetsova, I. M., T. A. Yakusheva, and K. K. Turoverov. 1999. Contribution of separate tryptophan residues to intrinsic fluorescence of actin. Analysis of 3D structure. *FEBS Lett.* 452:205–210.

Lakowicz, J. R. 1983. Principles of Fluorescence Spectroscopy. Plenum Press, New York.

Laskowski, R. A., E. Hutchinson, A. D. Michie, A. C. Wallace, M. L. Jones, and J. M. Thornton. 1997. PDBsum: a web-based database of summaries and analyses of all PDB structures. *Trends Biochem. Sci.* 22:488–490.

Lee, B., and F. M. Richards. 1971. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* 55:379–400.

Legon, A. C., and D. J. Millen. 1987. Directional character, strength, and nature of the hydrogen bond in gas-phase dimers. *Acc. Chem. Res.* 20:39–45.

Levitt, M., and B. H. Park. 1993. Water: Now you see it, now you don't. *Structure.* 1:223–226.

Lippert, E. 1957. Spektroskopische Bestimmungen des Dipolmomentes aromatischer Verbindungen im ersten angeregten Singulettzustand. *Z. Elektrochem.* 61:962.

Liptay, W. 1965. Die Lösungsmittelabhägigkeit der Wellenzahl von Elektronenbanden und chemisch-physikalischen Grundlagen. *Z. Naturforsch.* 20a:1441–1471.

Lumry, R., and M. Hershberger. 1978. Status of indole photochemistry with special reference to biological application. *Photochem. Photobiol.* 27:819–840.

Maki, K., T. Ikura, T. Hayano, N. Takahashi, and K. Kuwajima. 1999. Effects of proline mutations on the folding of staphylococcal nuclease. *Biochemistry.* 38:2213–2223.

Mataga, N., Y. Kaifu, and M. Koizumi. 1955. The solvent effect on fluorescence spectrum. Change of solute-solvent interaction during the lifetime of excited solute molecule. *Bull. Chem. Soc. Japan.* 28: 690–691.

Mataga, N., Y. Kaifu, and M. Koizumi. 1956. Solvent effects upon fluorescence spectra and the dipole-moments of excited molecules. *Bull. Chem. Soc. Japan.* 29:465–470.

Mazurenko, Yu. T. 1973. On the temperature dependence of luminescence spectra of viscous solution. *Optics and Spectrosc. (USSR).* 34:917–923. [In Russian].

Mazurenko, Yu. T., and V. S. Udaltsov. 1978. Spectral relaxations of fluorescence. I. Kinetics of spectra related to the orientational solvent relaxation. *Optics and Spectrosc. (USSR).* 44:714–719. [In Russian].

McDonald, I., and J. Thronton. 1994. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* 238:777–793.

Mely, Y., M. Cadene, I. Sylte, and J. G. Bieth. 1997. Mapping the suramin-binding sites of human neutrophil elastase: investigation by fluorescence resonance energy transfer and molecular modeling. *Biochemistry.* 36:15624–15631.

Mendoza, J. L., V. H. Markos, and R. Gonter. 1978. A new perspective on sequential testing procedures in canonical analysis: a Monte Carlo evaluation. *Multivariate Behavioral Res.* 13:371–382.

Muiño, P. L., and P. R. Callis. 1994. Hybrid simulations of solvation effects on electronic spectra: indoles in water. *J. Chem. Phys.* 100: 4093–4109.

Orlov, N. Ya, T. G. Orlova, Ya. K. Reshetnyak, E. A. Burstein, and N. Kimura. 1999. Comparative study of recombinant rat nucleoside diphosphate kinases alpha and beta by intrinsic protein fluorescence. *J. Biomol. Struct. and Dyn.* 16:955–968.

Otting, G., E. Liepinsh, and K. Wuthrich. 1991. Protein hydration in aqueous solution. *Science.* 254:974–980.

Pelley, R., and P. Horowitz. 1976. Fluorimetric studies of tryptophyl exposure in concanavalin A. *Biochim. Biophys. Acta.* 427:359–363.

Permyakov, E. A., and V. L. Shnyrov. 1983. A spectrofluorometric study of the environment of tryptophans in bacteriorhodopsin. *Biophys. Chem.* 18:145–152.

Pierce, D. W., and S. G. Boxer. 1995. Stark effect spectroscopy of tryptophan. *Biophys. J.* 68:1583–1591.

Rao, C. R. 1951. An asymptotic expansion of the distribution of Wilks' criterion. *Bull. Int. Statist. Inst.* 33:177–181.

Reshetnyak, Ya. K., and E. A. Burstein. 1997a. Assignment of log-normal components of protein fluorescence spectra to individual tryptophan residues using their microenvironment properties in three-dimensional structure. *Biophysics (Moscow).* 42:293–300 [In Russian; English translation].

Reshetnyak, Ya. K., and E. A. Burstein. 1997b. Assignment of components of fluorescence spectra of serine proteases to clusters of tryptophan residues. *Biophysics (Moscow).* 42:785–795 [In Russian; English translation].

Reshetnyak, Ya. K., and E. A. Burstein. 2001. Decomposition of protein tryptophan fluorescence spectra into log-normal components. II. The statistical proof of discreteness of tryptophan classes in proteins. *Biophys. J.* 81:1710–1734.

Rousslang, K. W., J. M. Thomasson, J. B. Rose, and A. L. Kwiram. 1979. Triplet state of tryptophan in proteins. 2. Differentiation between tryptophan residues 62 and 108 in lysozyme. *Biochemistry.* 18:2296–2300.

Sanschagrin, P. C., and L. A. Kuhn. 1998. Cluster analysis of consensus water sites in thrombin and trypsin shows conservation between serine proteinases and contributions to ligand specificity. *Protein Sci.* 7:2054–2064.

Shore, V. G., and A. B. Pardee. 1956. Fluorescence of some proteins, nucleic acids and related compounds. *Arch. Biochem. Biophys.* 60: 100–107.

Steele, R. H., and A. Szent Györgyi. 1958. Studies on the excited states of proteins. *Proc. Natl. Acad. Sci. U.S.A.* 44:540–545.

Teale, F. W. J. 1960. The ultraviolet fluorescence of proteins in neutral solutions. *Biochem. J.* 76:381–388.

Teale, F. W. J., and G. Weber. 1958. Ultraviolet fluorescence of proteins. *Biochem. J.* 72:15.

Toptygin, D., and L. Brand. 2000. Spectrally- and time-resolved fluorescence emission of indole during solvent relaxation: a quantitative model. *Chem. Phys. Lett.* 322:496–502.

Turoverov, K. K. 1999. Intrinsic protein fluorescence: structure and dynamics of macromolecules. Thesis, Doctor Phys. and Math. Sci., St. Petersburg State Technical University, St. Petersburg, Russia.

Turoverov, K. K., I. M. Kuznetsova, and V. M. Zaitsev. 1984. Azurin UV-fluorescence interpretation on the basis of x-ray data. *Bioorg. Khim. (Moscow).* 10:792–806. [In Russian].

Turoverov, K. K., I. M. Kuznetsova, and V. N. Zaitsev. 1985. The environment of the tryptophan residue in *Pseudomonas aeruginosa* azurin and its fluorescence properties. *Biophys. Chem.* 23:79–89.

Vanderkooi, J. M. 1998. The protein state of matter. *Biochim. Biophys. Acta.* 1386:242–253.

Veeraraghavan, S., B. T. Nall, and A. L. Fink. 1997. Effect of prolyl isomerase on the folding reactions of staphylococcal nuclease. *Biochemistry.* 36:15134–15139.

Vincent, M., J. Gallay, and A. P. Demchenko. 1995. Solvent relaxation around the excited state of indole: analysis of fluorescence lifetime distributions and time-dependence spectral shifts. *J. Phys. Chem.* 99: 14931–14941.

Vincent, M., J. Gallay, and A. P. Demchenko. 1997. Dipolar relaxation around indole as evidenced by fluorescence lifetime distributions and time-dependence spectral shifts. *J. Fluoresc.* 7:107S–110S.

Vladimirov, Yu. A. 1959. Fluorescence of aromatic amino acids in solutions, crystals and proteins. *Trans. Acad. Sci. USSR.* 23:86–89. [In Russian].

Vladimirov, Yu. A., and E. A. Burstein. 1960. Luminescence spectra of aromatic amino acids and proteins. *Biophysics (Moscow).* 5:385–392 [In Russian; English translation].

Vladimirov, Yu. A., and S. V. Konev. 1957. On the possibility of energy migration in protein molecules. *Biophysics (Moscow).* 2:3–19 [In Russian; English translation].

Walker, M. S., T. W. Bednar, and R. Lumry. 1967. Exciplex studies. II. Indole and indole derivatives. *J. Chem. Phys.* 47:1020–1028.

Ward, J. H. 1963. Hierarchical grouping to optimize an objective function. *J. Am. Statist. Assoc.* 58:236.

Weber, G. 1960. Fluorescence polarization spectrum and electronic energy transfer in tyrosine, tryptophan and related compounds. *Biochem. J.* 207:522–531.

Willaert, K., and Y. Engelborghs. 1991. The quenching of tryptophan fluorescence by protonated and unprotonated imidazole. *Eur. Biophys. J.* 20:177–182.

Yuan, T., A. M. Weljie, and H. J. Vogel. 1998. Tryptophan fluorescence quenching by methionine and selenomethionine residues of calmodulin: orientation of peptide and protein binding. *Biochemistry.* 37:3187–3195.

Zhang, M. Q. 1997. Identification of protein coding regions in the human genome by quadratic discriminant analysis. *Proc. Natl. Acad. Sci. U.S.A.* 94:565–568.