Parametric Models to Compute Tryptophan Fluorescence Wavelengths from Classical Protein Simulations

Alvaro J. Lopez and Leandro Martínez *

Fluorescence spectroscopy is an important method to study protein conformational dynamics and solvation structures. Tryptophan (Trp) residues are the most important and practical intrinsic probes for protein fluorescence due to the variability of their fluorescence wavelengths: Trp residues emit in wavelengths ranging from 308 to 360 nm depending on the local molecular environment. Fluorescence involves electronic transitions, thus its computational modeling is a challenging task. We show that it is possible to predict the wavelength of emission of a Trp residue from classical molecular dynamics simulations by computing the solvent-accessible surface area or the electrostatic interaction between the indole group and the rest of the system. Linear parametric models are obtained to predict the maximum emission wavelengths with standard errors of the order 5 nm. In a set of 19 proteins with emission wavelengths ranging from 308 to 352 nm, the best model predicts the maximum wavelength of emission with a standard error of 4.89 nm and a quadratic Pearson correlation coefficient of 0.81. These models can be used for the interpretation of fluorescence spectra of proteins with multiple Trp residues, or for which local Trp environmental variability exists and can be probed by classical molecular dynamics simulations. © 2018 Wiley Periodicals, Inc.

DOI: 10.1002/jcc.25188

Introduction

Fluorescence spectroscopy is an important method for studying structure, dynamics and interactions of proteins.[1,2] Tryptophan, tyrosine and phenylalanine are the three naturally occurring fluorescent amino acids. Tryptophan is the most important intrinsic probe to monitor fluorescence in proteins due to i) the environmental sensitivity of its fluorescence emission and quantum yield; and ii) its low abundance in proteins, which allows the study of specific molecular environments.[3,4] The dependency of the fluorescence emission wavelength (λ_{em}) and quantum yield of the Trp residues with the molecular environment is, however, its most distinctive feature. Indeed, the λ_{em} of Trp residues can vary between 308 and 360 nm, depending on their local molecular interactions.[5,6]

The indole group of the Trp is the chromophore responsible for the fluorescence. The photophysics of indole has been elucidated and depends on two main electronic transitions, involving excited states known as L_a and L_b states.[7,8] The L_a state is sensitive to the local environment and is responsible for the protein fluorescence.[6,9,10] Therefore, in principle, the prediction of the emission wavelength of a Trp residue depends on the computation of the energy of the ground and L_a states in the appropriate molecular surroundings.

The computation of the fluorescence spectra from first principles depends on the computation of the electronic structure of ground and excited states of the chromophore in condensed media. Therefore, hybrid Quantum-Mechanics/Molecular-Mechanics (QM/MM) methods have been developed to predict the λ_{em} of Trp in proteins.[6,11] For example, Vivian and Callis[6] sampled protein conformations using classical molecular dynamics (MD) simulations in which the charges of the 3-methyl-indole (3MI) group of the Trp residue were set either to those of the Ground State (GS) or excited L_a states. The molecular structures obtained from these simulations were used to compute the energy of vertical transitions using a QM methodology, where the water and the remaining protein atoms were treated as points charges. Excitation and emission wavelengths could be estimated from the GS and L_a simulations, respectively, with good quantitative agreements with experimental data: the predicted fluorescence wavelengths agreed with experimental measurements with standard deviation of 6.87 nm.

QM/MM methodologies for the calculation of fluorescence spectra depend, however, on significant computational resources and, more importantly, a detailed understanding of the properties of the electronic transitions involved. For instance, the analysis of the transition energy for the identification of the correct states to compute the energy of vertical transitions requires specific expertise. On the other side, previous studies reported that the emission wavelengths of Trp residues can be correlated in specific cases to simple structural parameters, as the solvent exposure of specific atoms,[12,13] or the presence of local hydrogen bonds.[14]

Motivated by the large availability of experimental data on protein Trp fluorescence, we studied the possibility of

A. J. Lopez, L. Martínez
Institute of Chemistry and Center for Computational Engineering & Sciences, University of Campinas, Campinas, SP, Brazil
E-mail: leandro@iqm.unicamp.br

Contract grant sponsor: Fundação de Amparo à Pesquisa do Estado de São Paulo; Contract grant numbers: 2010/16947-9; 2013/08293-7; 2013/05475-7; Contract grant sponsor: Conselho Nacional de Desenvolvimento Científico e Tecnológico; Contract grant number: 161789/2014-5

Table 1. Proteins used, experimental maximum emission wavelengths, and structural and energetic parameters computed from the simulations. $SASA_{ind}$, $SASA_{ben}$ and $SASA_{pyr}$ are the SASAs of the complete indole, or benzene and pyrrole groups. Q_T, Q_p and Q_w are the electrostatic interaction energies of the indole group with the whole environment (protein + water), with the protein and with water, respectively. The $\langle\rangle$ indicate simulation averages, and the standard deviation of the averages is shown. Wavelengths are reported in nm, SASA in $Å^2$ and energies in kJ × mol^{-1}.

Protein	PDB id.	λ_{em}^{exp}	$\langle SASA_{ind}\rangle$	$\langle SASA_{ben}\rangle$	$\langle SASA_{pyr}\rangle$	$\langle Q_T\rangle$	$\langle Q_p\rangle$	$\langle Q_w\rangle$
Azurin	4AZU[37]	308[38]	0.45 ± 0.78	0.21 ± 0.48	0.24 ± 0.50	−8.99±5.05	−5.94±3.06	−3.05±4.31
Parvalbumin	1B8R[39]	316[1]	6.39 ± 3.03	3.22 ± 2.44	3.17 ± 2.59	−3.29±8.68	3.60 ±4.13	−6.89±8.28
Myoglobin	1MYT[40]	321[1]	24.23 ± 13.89	5.63 ± 5.48	18.60 ± 12.46	−25.93±13.81	−22.31±20.75	−3.62±16.30
Ribonuclease T1	9RNT[41]	328[26]	1.79 ± 1.20	1.19 ± 1.07	0.60 ± 0.58	−17.93±8.05	−9.56±4.75	−8.37±7.11
FKBP12	1D6O[42]	330[43]	14.87 ± 5.56	10.74 ± 5.07	4.13 ± 2.66	−12.64±7.65	−0.43±3.95	−12.21±7.43
T4-lysozyme W126Y W138Y	1LYD[44]	330[45]	43.04 ± 9.74	5.68 ± 4.46	37.36 ± 6.05	−28.77±15.96	−15.29±14.12	−13.48±10.72
T4-lysozyme W126Y W158Y	1LYD[44]	330[45]	5.01 ± 2.42	0.32 ± 0.64	4.69 ± 2.28	−24.22±6.75	−29.10±6.08	4.88 ±4.69
T4-lysozyme W138Y W158Y	1LYD[44]	330[45]	42.88 ± 6.52	4.60 ± 3.19	38.28 ± 4.75	19.30 ±11.35	0.74 ± 6.46	−20.04±11.59
Subtilisin Carlsberg	1SBC[46]	331[47]	61.98 ± 8.86	18.59 ± 5.88	43.39 ± 4.45	−24.80±10.79	−5.89±4.07	−18.91±10.64
Staphylococcal Nuclease	1STN[48]	334[25]	48.46 ± 7.51	48.34 ± 7.45	0.12 ± 0.27	−35.87±8.03	−17.98±6.42	−17.89±8.96
Che-Y	1CHN[49]	335[1]	10.20 ± 4.52	4.05 ± 3.22	6.15 ± 4.17	−20.08±11.47	−7.24±5.58	−12.84±11.27
Cobra toxin	1CTX[50]	340[1]	81.75 ± 7.85	46.13 ± 6.18	35.62 ± 5.25	−32.97±12.17	−11.86±6.17	−21.11±13.12
HSA	1BM0[51]	340[35]	49.83 ± 7.93	23.21 ± 6.64	26.62 ± 5.76	−31.63±11.61	−6.40±5.37	−25.23±11.46
Monellin	1MOL[52]	342[26]	60.51 ± 15.26	19.44 ± 10.72	41.07 ± 8.67	−30.80±13.18	−7.26±12.17	−23.54±15.70
Cholera toxin B (monomer)	1CHP[53]	345[1]	58.07 ± 14.91	42.09 ± 13.50	15.98 ± 4.09	−28.84±11.69	−9.42±5.84	−19.42±11.76
Thioredoxin W31A	2TRX[54]	345[55]	73.99 ± 14.08	70.78 ± 12.40	3.21 ± 2.84	−44.26±14.05	−27.91±12.16	−16.35±11.70
Melittin	2MLT[56]	346[1]	135.49 ± 14.03	77.51 ± 13.66	57.98 ± 11.70	−36.25±13.82	1.78 ± 6.88	−38.03±15.67
Phospholipase A2	2BPP[57]	348[58]	149.38 ± 30.20	87.75 ± 31.09	61.63 ± 9.49	−37.21±13.92	−6.21±5.07	−31.00±14.42
Glucagon	1GCN[59]	352[26]	154.47 ± 13.28	84.27 ± 11.82	70.20 ± 7.59	−37.30±13.76	11.09 ± 4.13	−48.39±14.12

obtaining estimates of the emission wavelengths of Trp in proteins from classical simulations, using standard force-fields. We found linear correlations between the solvent-accessible surface area (SASA) and the electrostatic interactions of indole with the Trp emission wavelength. Based on this correlation, we developed parametrized models for the prediction of λ_{em} of Trp in proteins from classical MD simulations. These models can be used to interpret fluorescence spectra of proteins with multiple Trp residues or for which conformational variability leads to the exposure of the chromophores to different molecular environments.

Methodology

Molecular dynamics simulation were performed for 19 single-tryptophan proteins, which are listed in Table 1. Crystallographic water molecules were maintained, and the proteins were solvated with Packmol.[15,16] To render the systems neutral, Sodium and Chloride ions were used. The protein, water and ions were simulated with CHARMM27 force-fields,[17,18] and the TIP3P model of water was used.[19] Periodic boundary conditions were used. The systems were equilibrated in three steps: i) 1000 steps of Conjugate-Gradient minimization followed by 200 ps of molecular dynamics simulations (MD) with the all protein atoms fixed, ii) 500 steps of Conjugate-Gradient minimization followed by 200 ps MD with the Cα atoms of the proteins fixed, iii) 2.0 ns MD without any constraint. The final structures were used as starting configurations for production runs of 20 ns. All simulations were run at the NPT ensemble at 1 atm (controlled by a Langevin barostat with 200 fs period and 100 ps decay) and 298.15 K (controlled by a Langevin bath with 10 ps damping coefficient). All simulations were performed with the NAMD software.[20] The computation of the SASAs, protein side-chain mutations and pictures of molecular structures were performed with VMD.[21] The computation of electrostatic interactions was performed with the MDAnalysis suite.[22]

Results and Discussion

Models based on tryptophan SASA

The indole group of the Trp residues is aromatic and quite hydrophobic, thus it tends to be partially or completely buried in hydrophobic pockets of proteins. This property is extensively exploited in protein folding studies, as the denaturation of the proteins exposes the Trp residues. This exposure results in the increased Trp fluorescence emission wavelength.[2,4,23,24] Therefore, it is expected that some correlation exists between the solvent exposure of the indole group of Trp residues and λ_{em}.

Figure 1 shows the temporal evolution of the SASA of the indole moiety of the Trp residues of 3 proteins: Parvalbumin,[1] Staphylococcal Nuclease[25] and Glucagon.[26] These proteins were chosen due to their steady-state emission wavelength maxima (λ_{em}): Parvalbumin, Staphylococcal Nuclease and Glucagon display the lowest, median and the highest values of λ_{em}, respectively, within the set of proteins used in this study

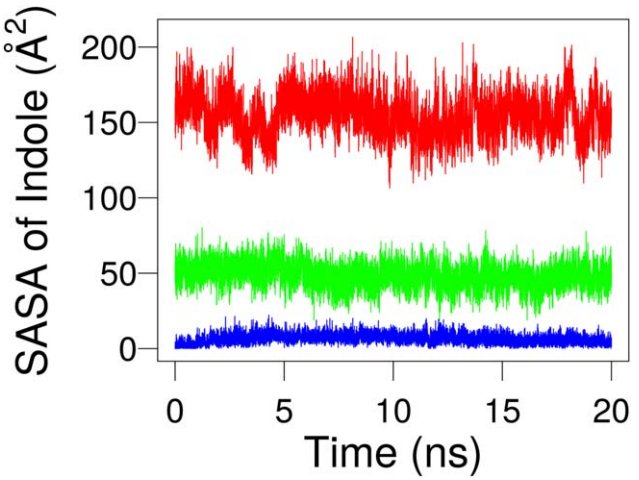

Figure 1. Temporal evolution of the SASA of indole of Parvalbumin (blue), Staphylococcal Nuclease (green) and Glucagon (red). Increased solvent exposure is associated with increased fluorescence emission wavelength. [Color figure can be viewed at wileyonlinelibrary.com]

(Table 1). Clearly, the greater solvent exposure of the indole group of the Trp residue is correlated with the fluorescence emission wavelength.

Similar calculations were performed for a set 19 single-tryptophan proteins, for which experimental λ_{em} is available. These proteins are representative of the diversity of λ_{em} found in nature and cover a range from 308 to 352 nm (Table 1). The SASA of the indole was computed for every snapshot of the simulations. The averages and standard deviations of the SASA of the indole groups and their decompositions on the exposure of the benzene and pyrrole groups are shown in Table 1.

Figure 2 shows the correlation between the experimental emission wavelength and the SASA of the indole. It is clear that the solvent exposure of the indole groups is correlated with the maximum emission wavelength of the proteins. Thus, we can propose that a greater (or lower) fluorescence emission wavelength of the Trp is due to the greater (or lower) exposition of the Trp indole group to water. Two different models based on the accessibility to the solvent of indole were studied.

The first model (Model 1) we propose consists of a simple linear dependence of the fluorescence emission wavelength with the average SASA of the indole. The SASA data used and the model equation and parameters obtained are reported in Tables 1 and 2, respectively.

Model 1 predicts the experimental emission wavelengths with a standard deviation of 6.80 nm, and with a correlation coefficient of $R^2=0.64$. Figure 3A displays the fit of the predicted versus experimental maximum wavelengths. In this model the lower limit of the emission wavelength of the indole when it is totally inaccessible to the solvent is 324 nm. This limiting wavelength is overestimated, as Trp residues emitting at wavelengths smaller than 324 nm exist. Indeed, three of the proteins emitting below 324 nm, Azurin, Parvalbumin and Myoglobin, were used for parametrization. For these three proteins, the predicted λ_{em} is overestimated by 16 nm, 9 nm and 7 nm, respectively (Table 3). For the proteins with λ_{em} above 324 nm, the error in the estimates is consistently below 10 nm, and the standard error is reduced to 5.27 nm.

The indole molecule consists of a benzene ring fused to a pyrrole ring. To know what moiety of indole more strongly affects the fluorescence emission wavelength, we divided the SASA of the indole as the sum of SASA of benzene and SASA of pyrrole. Based on this, we build a second model (Model 2), with three parameters:

$$\lambda_{em}=B_0+B_1\times\langle SASA_{ben}\rangle+B_2\times\langle SASA_{pyr}\rangle, \quad (1)$$

where B_0 is the limiting wavelength at low solvent exposures, and B_1 and B_2 the weight factors associating the SASA of the indole and pyrrole groups to λ_{em}, respectively. The resulting fit is shown in Table 2 and Figure 3B. In this case, the two atoms which are shared by both groups were considered to belong to the pyrrole group (other alternatives are discussed in Appendix A).

Model 2 provides a lower limit for the emission wavelength of $B_0=324$ nm, thus the increased complexity does not solve the limitation of Model 1 in predicting the λ_{em} of Azurin, Parvalbumin and Myoglobin. The correlation to the experimental data increases slightly, to $R^2=0.66$ and the standard deviation to experimental values decreases to 6.58 nm, as shown in Table 2.

The addition of one parameter, therefore, does not significantly improve the quality of the fit. Interestingly, these results from the fact that the contribution of the benzene ring to the predicted wavelength is much greater than the contribution of the pyrrole ring. The ratio B_1/B_2 is 2.39 and contrasts with the ratio of the solvent exposures of the two groups, $\overline{\langle SASA_{ben}\rangle/\langle SASA_{pyr}\rangle}=1.18$. Therefore, the benzene ring is determinant for λ_{em} even though the exposure of the two groups to water is somewhat similar. The importance of the exposure of benzene part of indole for λ_{em} is also reported by Vivian and Callis:[6] The water protons achieve an effective stabilization of the benzene group by an approximation to the

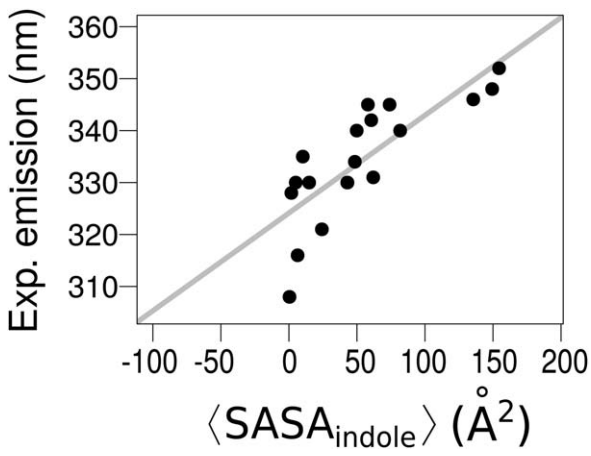

Figure 2. Correlation of experimental emission wavelength and average SASA of indole.

Table 2. Models and model parameters for the prediction of fluorescence emission wavelengths from classical MD simulations. R^2 is the squared Pearson correlation coefficient of the model predictions to experimental data, and SD is the standard deviation of the predictions from experimental data. Q_T in Model 3 is the total electrostatic interaction between the indole and the environment. Q_w and Q_p in Model 4 are the electrostatic interactions of the indole with water and protein atoms, respectively. Wavelengths are obtained in nm for SASA in Å^2 or energies in kJ × mol^{-1}.

Model	Description	Equation	R^2	SD
1	SASA indole	$\lambda_{em}=324+0.188\langle SASA_{ind}\rangle$	0.64	6.80
2	SASA benzene(i)/pyrrole(c)	$\lambda_{em}=324+0.246\langle SASA_{ben}\rangle+0.103\langle SASA_{pyr}\rangle$	0.66	6.58
3	Total electrostatics	$\lambda_{em}=311-0.884\langle Q_T\rangle$	0.69	6.36
4	Electrostatics water/protein	$\lambda_{em}=311-0.998\langle Q_w\rangle-0.591\langle Q_p\rangle$	0.81	4.89

carbon atoms CE3 and CZ2. However, the pyrrole part of indole was found to be unable to interact with waters strongly enough to cause an emission red-shift.[6]

The predicted λ_{em} using Model 2 are reported in Table 3. The lowest and highest deviations correspond to T4-lysozyme W126Y W138Y (∼−0.20 nm) and Azurin (∼17 nm), respectively. The computed versus experimental λ_{em} are plotted in Figure 3B. These models based on the SASA of the indole show that there is a reasonable correlation between SASA and emission wavelength for Trp which are at least partially exposed to solvent. Burstein et al. proposed that the tryptophan fluorescence of proteins can be classified into five discrete classes:[14,27] Class A ($\lambda_{em}=308$ nm, completely buried TRP), Class S ($\lambda_{em}=316$ nm, buried TRP), Class I ($\lambda_{em}=330$–332 nm, partial buried TRP), Class II ($\lambda_{em}=340$–342 nm, exposed TRP) and class III($\lambda_{em}=350$–353 nm, completely exposed TRP). The proteins used in this study can be classified according to the discrete states proposed by Burstein et al. as:

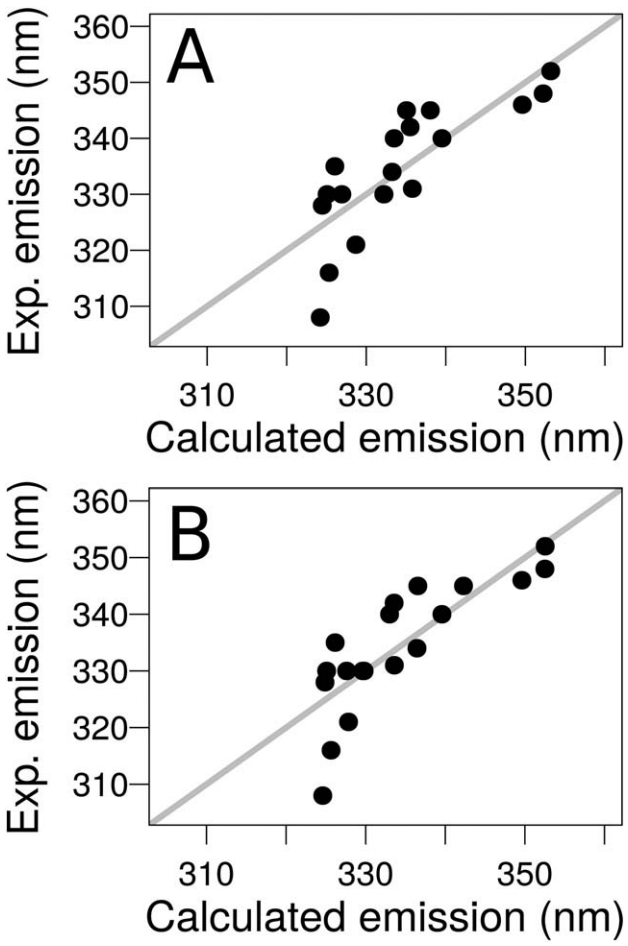

Figure 3. Experimental versus predicted maximum emission wavelength of the Trp residues using models based on the SASA. A) Model 1: Two-parameter model based on the ⟨SASA⟩ of the indole group. B) Model 2: Three-parameter model with different weights for the ⟨SASA⟩ of pyrrole and benzene groups of the Trp side-chain.

- Class A: Azurin
- Class S proteins: Parvalbumin, Myoglobin and Ribonuclease T1
- Class I proteins: FKBP12, T4 Lysozyme W126Y W138Y, T4 Lysozyme W126Y W158Y, T4 Lysozyme W138Y W158Y, S. Carlsberg, S. Nuclease and Che-Y
- Class II proteins: Cobra toxin, HSA, Monelin, Cholera toxin and Thioredoxin W31A
- Class III proteins: Melittin, Phospholipase A2 and Glucagon

The average values of SASA of the proteins of classes A, S, I, II and III are 0.45, 10.80, 32.35, 64.83 and 146.45 Å^2, respectively, showing a good correspondence with the discrete classes of Burstein. The notable exception, however, is Myoglobin, for which the solvent exposure observed is much greater than what is expected from the λ_{em} alone. From the SASA of Myoglobin one would expect an emission in ∼330 nm (Table 3), and its classification in Class I.

Models based on tryptophan electrostatic interactions

The simulations permit the computation of electrostatic interactions parameters of the indole with the environment.[6] These interactions are fundamental for the stabilization of the excited states, which display increased polarity than the ground state. Therefore, a correlation between the nature of electrostatic interactions experienced by the indole group and the emission wavelength is expected.[12] The interactions considered here are those with the ground state charges obtained directly from classical MD force-fields, for simplicity. We show that it is possible to obtain good parametric models for the emission wavelength from these ground-state, classical models.

The electrostatic interaction between indole group and environment were calculated using the classical equation, for each time step t of the simulations:

Table 3. Experimental and calculated fluorescence emission wavelengths of Trp: λ_{em}^{exp} is the experimentally observed maximum emission wavelength. $\lambda_{em}^{(1-4)}$ and $\Delta\lambda_{em}^{(1-4)}$ are the predicted emission wavelength and its deviation relative to the experimental value using Models 1 to 4. Models 1 and 2 are based on SASAs. Models 3 and 4 are based on electrostatic interactions. Wavelengths and deviation units are nm.

Protein	λ_{em}^{exp}	$\lambda_{em}^{(1)}$	$\Delta\lambda_{em}^{(1)}$	$\lambda_{em}^{(2)}$	$\Delta\lambda_{em}^{(2)}$	$\lambda_{em}^{(3)}$	$\Delta\lambda_{em}^{(3)}$	$\lambda_{em}^{(4)}$	$\Delta\lambda_{em}^{(4)}$
Azurin	308[38]	324.22	16.22	324.61	16.61	318.90	10.90	318.04	10.04
Parvalbumin	316[1]	325.34	9.34	325.65	9.65	313.86	−2.14	316.24	0.24
Myoglobin	321[1]	328.69	7.69	327.84	6.84	333.87	12.87	328.28	7.28
Ribonuclease T1	328[26]	324.47	−3.53	324.89	−3.11	326.80	−1.20	325.49	−2.51
FKBP12	330[43]	326.93	−3.07	327.60	−2.40	322.12	−7.88	323.93	−6.07
T4-lysozyme W126Y W138Y	330[45]	332.23	2.23	329.80	−0.20	336.38	6.38	333.97	3.97
T4-lysozyme W126Y W158Y	330[45]	325.08	−4.92	325.10	−4.90	332.36	2.36	323.81	−6.19
T4-lysozyme W138Y W158Y	330[45]	332.20	2.20	329.63	−0.37	328.02	−1.98	331.06	1.06
Subtilisin Carlsberg	331[47]	335.80	4.80	333.60	2.60	332.88	1.88	333.85	2.85
Staphylococcal Nuclease	334[25]	333.25	−0.75	336.44	2.44	342.66	8.66	339.97	5.97
Che-Y	335[1]	326.05	−8.95	326.16	−8.84	328.70	−6.30	328.58	−6.42
Cobra toxin	340[1]	339.52	−0.48	339.58	−0.42	340.09	0.09	339.57	−0.43
HSA	340[35]	333.51	−6.49	333.00	−7.00	338.91	−1.09	340.46	0.46
Monellin	342[26]	335.52	−6.48	333.57	−8.43	338.17	−3.83	339.27	−2.73
Cholera toxin B (monomer)	345[1]	335.06	−9.94	336.55	−8.45	336.44	−8.56	336.44	−8.56
Thioredoxin W31A	345[55]	338.06	−6.94	342.29	−2.71	350.07	5.07	344.30	−0.70
Melittin	346[1]	349.63	3.63	349.62	3.62	343.00	−3.00	348.40	2.40
Phospholipase A2	348[58]	352.24	4.24	352.52	4.52	343.84	−4.16	346.10	−1.90
Glucagon	352[26]	353.20	1.20	352.55	0.55	343.92	−8.08	353.24	1.24

$$E(t) = \frac{1}{4\pi\varepsilon_0}\sum_i\sum_j \frac{q_i q_j}{d_{ij}} \quad (2)$$

where ε_0 is the vacuum permittivity, d_{ij} is the distance between atom i and atom j, Q is the point charges, i is the atoms of the Trp group considered (indole atoms, for example) and j is the other atoms of the system considering appropriate bonded exclusions determined by the CHARMM force-fields. All atomic charges were obtained from the CHARMM27 force-fields, with which the simulations were performed.[17,18] $E(t)$ can be trivially decomposed in the contributions of interactions of any groups, provided an appropriate decomposition of the sum above.

Two models based on electrostatic interactions were built. Model 3 associates the total electrostatic interaction between the indole group and environment with the emission wavelength. Model 4 discriminates the contributions of the electrostatic interactions with the protein and with water molecules. The electrostatic parameters computed from the simulations are shown in Table 1, and the details of the models are shown in Table 2.

Model 3, which is based on the total electrostatic interactions of the indole group provides an overall similar prediction of λ_{em} than the models based on the SASA. The correlation coefficient is $R^2 = 0.69$ and the standard deviation to experimental data is 6.36 nm. However, the lower limit wavelength of emission is 311 nm, thus the use of the electrostatics corrects the error of the SASA models in this range of λ_{em}. Indeed, as shown in Figure 4A and Table 3, the emission wavelength of Parvalbumin, which as incorrectly predicted by the SASA models, is now obtained with a deviation of only −2.14 nm to experimental data. The error on the emission wavelength of Azurin decreases to about ∼10 nm. However,

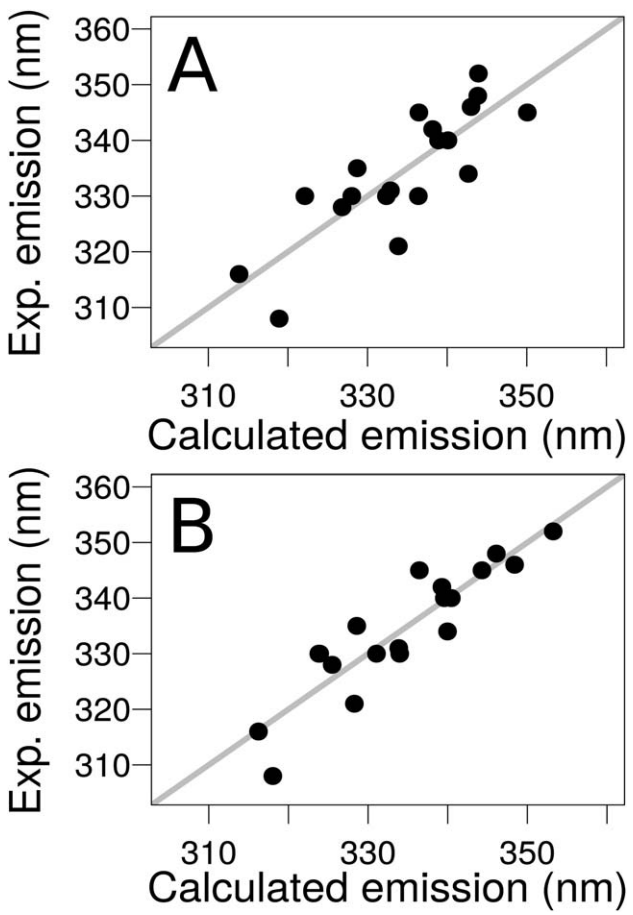

Figure 4. Experimental versus predicted maximum emission wavelength of the Trp residues using models based on electrostatic interactions. A) Model 3: Two-parameter full electrostatic model. B) Model 4: Three-parameter model with different weights for interactions with water and the protein.

this improvement comes with the side effect of reducing the quality of the predictions of greater wavelengths.

The decomposition of the electrostatic interactions of the indole into water and protein components, on the other side, improves significantly the predictive capacity of the models. Model 4, with three parameters, was constructed by linearly fitting the experimental wavelengths to the electrostatic interactions with water and the protein with different weights. The electrostatic interactions obtained from the simulations are reported in Table 1, and the parameters of the model obtained are shown in Table 2.

Model 4 predictions have a significantly higher correlation coefficient with experimental data than other models, as shown in Figure 4B. The correlation is $R^2=0.81$, and the standard deviation of the predicted values to experimental values is reduced to 4.89 nm. More importantly, however, as shown in Figure 4B, the predicted values do not display particularly large errors at the minimum or maximum limiting emission wavelengths, as observed for the other models. For Model 4, all predicted wavelengths are within ~10 nm of the experimental values. The greatest deviation is observed for Azurin, which appears to display greater solvent exposure and electrostatic interactions than what is expected from the experimentally observed spectrum.

The contribution of electrostatic interactions with water molecules is more important for the prediction of λ_{em} than the interactions with the protein. As shown in Table 2, the slope associating λ_{em} with electrostatic interactions with water, $\langle Q_w \rangle$, is -0.998 (kJ $\times$ mol^{-1})$^{-1}$ nm, while the slope associated with electrostatic interactions with the proteins, $\langle Q_p \rangle$, is -0.591 (kJ $\times$ mol^{-1})$^{-1}$ nm. This is probably due to the greater possible directionality of water–indole interactions.[12]

Physical basis for empirical correlations

The high correlations that can be obtained between SASA and electrostatic interactions of ground-state indole groups of Trp with the emission wavelength are, perhaps, surprising. The emission wavelength is dependent on the stabilization of the excited state of the indole, which displays a different charge distribution than the ground state. In particular, it is known that the electron density shifts from pyrrole to the benzene ring of the indole upon excitation, increasing the polarity of the group.[28]

Because of the increased polarity of the excited state, however, the emission wavelength can be associated with the polarity of the environment. Local environments composed of polar groups can rearrange to interact effectively with the increased polarity of the excited state. Of course, this depends on the mobility or flexibility of the polar groups participating in the interactions with the indole group.

The present parametric models suggest that the SASA and the electrostatic interactions of the ground-state indole are reasonable predictors of the local polarity around the Trp. The correlations obtained with the SASA models show that the local polarity is mostly dependent on the exposure of the indole group to water. Correspondingly, these models fail to predict emission wavelengths when the indole is completely buried into to the protein.

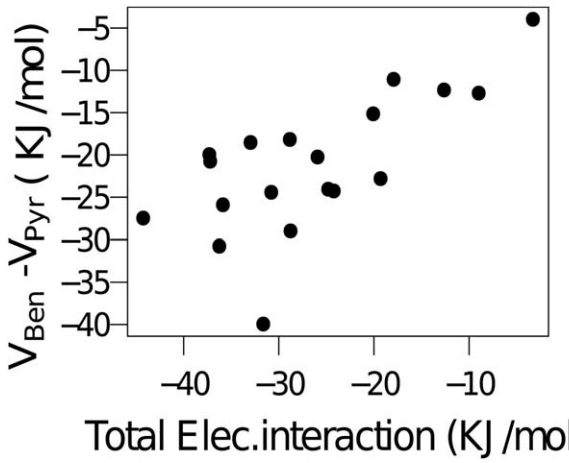

Figure 5. Correlation between the difference of electrostatic interactions of the benzene and pyrrole rings and the total electrostatic interaction of the indole group.

The use of the electrostatic interactions to obtain the models improves the predictions because they take into consideration interactions with the protein, which might be also polar. Of course, the indole group with ground-state charges can only be a reasonable probe of the local polarity of the medium because of the variable partial charges of its atoms. At the same time, the excited state of the indole displays a polarity resulting from the uneven distribution of charges between the pyrrole and benzene rings. Therefore, for the total electrostatic interactions of the indole be correlated with the stabilization of its excited state, there should be a correlation between total indole interactions and the difference in interactions between benzene and pyrrole groups. In Figure 5, we show that this correlation indeed can be verified in the set of proteins studied here. This correlation stems from the fact that electrostatic interactions are stronger if the system is polar. Therefore, if there is at least some polarization between benzene and pyrrole groups in the ground state of the indole, this will be reflected in the total interactions. Thus, these interactions can be used to probe the polarity of the molecular environment. Of course, if the interactions of the indole and pyrrole groups are considered as independent variables for model parametrization, the quality of the correlations can be increased. Several models with these considerations in mind are discussed in Appendix A.

From these arguments, it follows that all models suggested here would fail if there were polar interactions that disfavor the excited state and are too rigid to be rearranged following excitation. The good correlations obtained here indicate that this is not a common scenario. At the same time, it is likely that the improper account of the different flexibilities of the different groups interacting with the indole group is a source of partial deviations between predicted and experimental emission wavelengths.

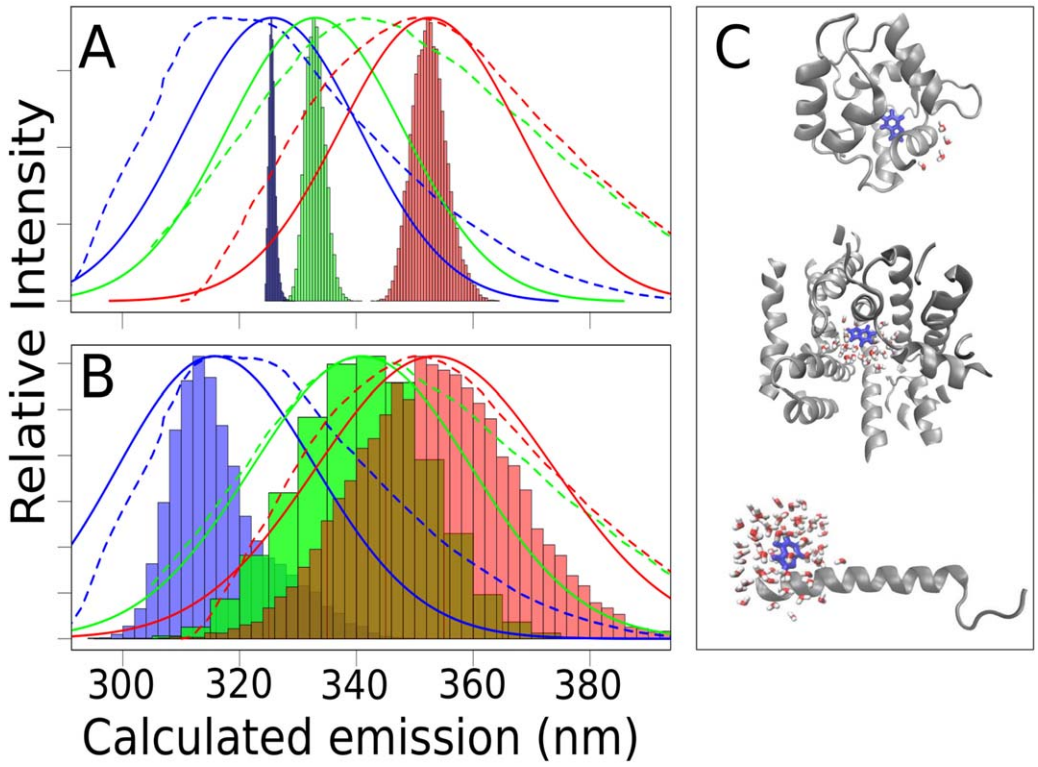

Figure 6. Calculated fluorescence emission spectra of proteins Parvalbumin, Human Serum Albumin and Glucagon using Model A) 2 and B) Model 4. Gaussian fits were applied with a band broadening of 15 nm (continuous lines). Experimental emission spectra for Parvalbumin,[34] Human Serum Albumin[35] and Glucagon[36] are also shown (dashed lines). C) Representative molecular structures of Parvalbumin, Human Serum Albumin and Glucagon (from top to bottom). Indole (blue), nearest water (red and white) and protein structure (gray). [Color figure can be viewed at wileyonlinelibrary.com]

Variability of emission wavelength within the simulations

The variability of the molecular environment and its effects on the Trp conformations has been used to study the time-dependence of fluorescence wavelength and anisotropy decay.[29–31] Here, different protein and solvent conformations obtained in the simulations are associated with different solvent accessibilities and electrostatic interactions of the indole group. Therefore, it is possible to construct the distribution of wavelengths predicted from the simulations. These distributions correspond to the emission spectra associated with the single electronic transition that is responsible for the maximum wavelength of emission. The experimental spectra are expected to display additional band broadening due to the differences in geometry between ground and excited states.[7,32]

Using our linear models, we can predict the time-dependence of λ_{em}, that is, we compute the instantaneous fluorescence emission wavelength at each snapshot sampled by the MD simulations. The histogram constructed with every "instantaneous" λ_{em} is an approximation of the emission spectra of the Trp conformation probed within each simulation. Figures 6A and 6B show the computed emission spectra of Parvalbumin, Human Serum Albumin, and Glucagon, computed using Model 2 and Model 4, respectively. As expected, the variability of the emission wavelengths predicted by the simulations is smaller than that observed experimentally.[13] However, the experimental broadening can be reasonably fit considering a broadening of 15 nm. In this case, a qualitative agreement between the experimental and predicted spectra can be obtained.

Representative molecular structures of these three proteins are shown in Figure 6C. The indole groups and the nearest accessible water molecules are shown. For Parvalbumin, the indole is located in the interior of protein and only two water molecules are found within 6Å. For the Human Serum Albumin, the indole is partially buried, while for Glucagon the indole is completely exposed to water. These results illustrate the different solvent exposures of the indole group which determines the maximum emission wavelengths.

Model robustness

Our models were parametrized with the complete set of 19 single Trp proteins (Table 1). Here, we probe how robust are the predictions relative to the training set used for parametrization.

We define new training sets for model building and parameter estimation and evaluate the prediction of the emission wavelength of the Trp of each protein using these new training sets.[33] For each protein P, the models were parametrized and the maximum emission wavelength of P was predicted using four different types of training sets:

1. Using all proteins of Table 1.
2. Using all but protein P.

Table 4. Sensitiveness of Model 2 relative to the training set used. λ_{em}^{calc} is the emission wavelength computed using all proteins in the training set. The deviation of the wavelength predicted when removing each protein from the training set is reported as $\Delta\lambda_{em}^{calc}(1)$. The average and maximum (in parenthesis) deviation associated with removing 2 proteins are reported in the $\Delta\lambda_{em}^{calc}(2)$ column. The average and maximum deviations (in parenthesis) associated with removing 3 proteins is shown in column $\Delta\lambda_{em}^{calc}(2)$. The average deviations are always smaller than 3 nm, indicating that the model is robust relative to the training set used. Wavelengths and deviation units are nm.

Protein	λ_{em}^{calc}	$\Delta\lambda_{em}^{calc}(1)$	$\Delta\lambda_{em}^{calc}(2)$	$\Delta\lambda_{em}^{calc}(3)$
Azurin	324.61	2.33	2.33 (4.11)	2.33 (5.29)
Parvalbumin	325.65	1.17	1.17 (3.84)	1.17 (4.98)
Myoglobin	327.84	0.65	0.66 (2.48)	0.66 (3.97)
Ribonuclease T1	324.89	−0.43	−0.43 (2.18)	−0.44 (4.23)
FKBP12	327.60	−0.26	−0.26 (1.98)	−0.27 (3.74)
T4-lysozyme W126Y W138Y	329.80	−0.03	−0.02 (−1.69)	−0.008 (−2.61)
T4-lysozyme W126Y W158Y	325.10	−0.60	−0.61 (−1.83)	−0.61 (3.66)
T4-lysozyme W138Y W158Y	329.63	−0.08	−0.06 (−1.83)	−0.05 (−2.79)
Subtilisin Carlsberg	333.60	0.44	0.46 (1.18)	0.48 (1.99)
Staphylococcal Nuclease	336.44	0.78	0.76 (2.89)	0.73 (−4.88)
Che-Y	326.16	−0.96	−0.97 (−1.73)	−0.97 (2.88)
Cobra toxin	339.58	−0.03	−0.02 (0.75)	−0.009 (1.94)
HSA	333.00	−0.43	−0.43 (−1.20)	−0.42 (−1.88)
Monellin	333.57	−1.20	−1.19 (−1.85)	−1.18 (−2.43)
Cholera toxin B (monomer)	336.55	−0.88	−0.89 (−2.31)	−0.91 (−3.10)
Thioredoxin W31A	342.29	−1.77	−1.74 (−4.83)	−1.71 (−6.22)
Melittin	349.62	0.97	1.02 (3.55)	1.07 (8.78)
Phospholipase A2	352.52	1.68	1.73 (4.03)	1.78 (9.92)
Glucagon	352.55	0.24	0.31 (3.29)	0.39 (10.39)

3. Removing protein P plus one protein: all possible combinations.
4. Removing protein P plus two proteins: all possible combinations.

Training set 1 consists of the model parametrized with all available data. We evaluate how the models are sensitive to the choice of the training set, by comparing the predictions obtained with each training set relative to those obtained using Training set 1. For sets 3 and 4, we tested all possible combinations of removing other proteins besides the reference one, and report the average and maximum deviations. The sensitiveness of Models 2 and 4 are reported in Tables 4 and 5, respectively.

The models are observed to be very robust relative to the training set used. The average deviations of the predictions obtained by reducing the training set are always below 3 nm. The greatest deviation in the prediction is observed for Glucagon. The maximum deviation of the predictions of most other proteins often falls below 6 nm. Therefore, the

Table 5. Sensitiveness of Model 4 relative to the training set used. λ_{em}^{calc} is the emission wavelength computed using all proteins in the training set. The deviation of the wavelength predicted when removing each protein from the training set is reported as $\Delta\lambda_{em}^{calc}(1)$. The average and maximum (in parenthesis) deviation associated with removing 2 proteins are reported in the $\Delta\lambda_{em}^{calc}(2)$ column. The average and maximum deviations (in parenthesis) associated with removing 3 proteins is shown in column $\Delta\lambda_{em}^{calc}(2)$. The average deviations are always smaller than 3 nm, indicating that the model is robust relative to the training set used. Wavelengths and deviation units are nm.

Protein	λ_{em}^{calc}	$\Delta\lambda_{em}^{calc}(1)$	$\Delta\lambda_{em}^{calc}(2)$	$\Delta\lambda_{em}^{calc}(3)$
Azurin	318.04	2.56	2.58 (4.25)	2.61 (5.36)
Parvalbumin	316.24	0.12	0.16 (5.32)	0.21 (7.21)
Myoglobin	328.28	1.36	1.35 (2.53)	1.33 (3.44)
Ribonuclease T1	325.49	−0.26	−0.26 (1.50)	−0.26 (2.62)
FKBP12	323.93	−1.12	−1.11 (−1.99)	−1.10 (−3.20)
T4-lysozyme W126Y W138Y	333.97	0.32	0.32 (1.29)	0.31 (2.21)
T4-lysozyme W126Y W158Y	323.81	−2.67	−2.70 (−4.69)	−2.75 (−7.88)
T4-lysozyme W138Y W158Y	331.06	0.13	0.15 (1.64)	0.17 (3.12)
Subtilisin Carlsberg	333.85	0.17	0.18 (1.04)	0.19 (1.69)
Staphylococcal Nuclease	339.97	0.82	0.82 (1.93)	0.82 (2.84)
Che-Y	328.58	−0.50	−0.50 (−1.32)	−0.50 (−2.01)
Cobra toxin	339.57	−0.03	−0.04 (−0.63)	−0.04 (1.37)
HAS	340.46	0.04	0.05 (−0.54)	0.06 (1.25)
Monellin	339.27	−0.19	−0.19 (−0.77)	−0.18 (−1.10)
Cholera toxin B (monomer)	336.44	−0.50	−0.51 (−1.01)	−0.51 (−1.52)
Thioredoxin W31A	344.30	−0.33	−0.37 (−3.27)	−0.41 (−4.71)
Melittin	348.40	0.60	0.64 (2.01)	0.69 (6.56)
Phospholipase A2	346.10	−0.28	−0.27 (−0.97)	−0.26 (−1.61)
Glucagon	353.24	0.84	0.93 (4.98)	1.01 (10.33)

Table 6. Models and model parameters using alternative group decompositions. Wavelengths are obtained in nm for SASA in Å^2 or energies in kJ × mol^{-1}. The subscripts ben and pyr indicate the contributions of benzene and pyrrole groups. Marks (c) and (i) indicate if the groups are complete or incomplete (missing shared atoms). When both groups are complete, the contributions of the shared atoms were considered repeatedly.

Model	Description	Equation	R^2	SD
A.1	SASA benzene(c)/pyrrole(i)	$\lambda_{em}=324+0.241\langle SASA_{ben}\rangle+0.094\langle SASA_{pyr}\rangle$	0.67	6.56
A.2	SASA benzene(c)/pyrrole(c)	$\lambda_{em}=324+0.284\langle SASA_{ben}\rangle-0.005\langle SASA_{pyr}\rangle$	0.59	7.23
A.3	Electrostatics benzene(i)/pyrrole(c)	$\lambda_{em}=312-1.808\langle Q_{ben}\rangle-0.506\langle Q_{pyr}\rangle$	0.74	5.77
A.4	Electrostatics benzene(i)/pyrrole(c)	$\lambda_{em}=313-1.930\langle Q_{ben-w}\rangle-1.263\langle Q_{ben-p}\rangle-0.602\langle Q_{pyr-w}\rangle-0.388\langle Q_{pyr-p}\rangle$	0.85	4.45
A.5	Electrostatics benzene(c)/pyrrole(i)	$\lambda_{em}=311-1.032\langle Q_{ben}\rangle-0.854\langle Q_{pyr}\rangle$	0.69	6.31
A.6	Electrostatics benzene(c)/pyrrole(i)	$\lambda_{em}=313-1.18\langle Q_{ben-w}\rangle-0.762\langle Q_{ben-p}\rangle-0.958\langle Q_{pyr-w}\rangle-0.549\langle Q_{pyrr-p}\rangle$	0.82	4.81
A.7	Electrostatics benzene(c)/pyrrole(c)	$\lambda_{em}=315-0.523\langle Q_{ben}\rangle-0.998\langle Q_{pyr}\rangle$	0.59	7.27
A.8	Electrostatics benzene(c)/pyrrole(c)	$\lambda_{em}=315-0.719\langle Q_{ben-w}\rangle-0.608\langle Q_{ben-p}\rangle-1.113\langle Q_{pyr-w}\rangle-0.595\langle Q_{pyr-p}\rangle$	0.75	5.68

deviations relative to experimental data are not importantly associated with the data set used, but to intrinsic limitations of the models.

Conclusions

In this article, classical parametrized models are proposed to predict fluorescence emission wavelengths (λ_{em}) of Trp in proteins. Linear correlations of the maximum emission wavelength to SASAs of the indole group and to their electrostatic interaction energies are observed. The best correlation was obtained providing different weights for indole–water and indole–protein electrostatic interactions. This model predicts λ_{em} with a standard deviation from experimental data of 4.89 nm, and predicted and experimental values display a correlation coefficient of $R^2=0.81$. All models tested are robust relative to the training set used. These models can be used for the study of the fluorescence emission wavelengths of proteins using classical molecular dynamics simulations.

Only single Trp proteins were used in this study, but many proteins exist with more than one tryptophan, exhibiting complex fluorescence emission spectra due to the contribution of every fluorophore. The present models can aid the interpretation of such spectra by decomposition into the contribution of each Trp residue.

Appendix A

Study of Pyrrole and Benzene Contributions

To compute the contributions of the benzene and pyrrole in the construction of the models, a choice must be made concerning the two carbon atoms which are shared by the groups. We explored three alternatives: 1) The SASA and electrostatic interactions were computed for both groups as if they were independent, that is, having 6 (benzene) and 5 (pyrrole) carbon atoms; 2) The two atoms were considered as being part of the benzene group; 3) The two atoms were considered as being part of the pyrrole group. In (1) the two groups are consistent with their chemical definitions, but the shared atoms are considered in duplicate in the calculations of SASA and interactions, and this is shown to reduce the quality of the fittings. In (2) and (3), one the groups loses its precise chemical definition and becomes a subset of the indole atoms only. However, in these cases, the sum of the contributions of the SASA or electrostatic interactions corresponds to that of the complete indole group. Thus, the decomposition increases the number of parameters of the fit and, therefore, necessarily improves the correlations. The goal of this decomposition was to evaluate the roles of benzene and pyrrole groups in the correlations. The resulting models are shown in Table 6.

In Model 2 of Table 2 and Models A.1 and, A.2 of Table 6, we see that the exposure of the benzene group to water is mostly determinant for the emission wavelength. Shifting the shared atoms from the pyrrole to the benzene group does not affect significantly the obtained model. The fit is the worst if the atoms are considered in both groups.

The decomposition of the contributions of the benzene and pyrrole groups in models based on electrostatic interactions results in Models A.3 to A.8 from Table 6. In most cases, the participation of the benzene group is more important than that of the pyrrole group, in agreement with what is observed for the SASA models. These results are probably derived from the fact that the pyrrole, being more constrained relative to the protein, suffers smaller variations both in SASA or electrostatic interactions. Thus, the emission wavelengths are mostly determined by the variable environment of the benzene group. However, the role of pyrrole is relatively increased if compared with the SASA models, and this appears to be an important factor for obtaining a good correlation, particularly for indole groups protected from solvent. The decomposition of the contributions of the interactions with water and the protein improved the correlation even further (Models A.3 and A.6), as expected from the fact that additional parameters are fit. The small improvement relative to Model 4 (Table 2) does not seem to justify the adoption of this model on a physical basis.

The greatest contribution of benzene atoms is verified in all models but those that consider both groups as complete, thus doubling the contribution of the shared atoms. These models display a poor correlation with the experimental wavelengths compared to models with the same number of parameters.

Keywords: protein fluorescence · indole · solvent-accessible surface area · molecular dynamics simulations · tryptophan

How to cite this article: A. J. Lopez, L. Martínez. *J. Comput. Chem*. **2018**, *39*, 1249–1258. DOI: 10.1002/jcc.25188

[1] M. R. Eftink, *Methods Biochem. Anal.* **1991**, *35*, 127.
[2] M. R. Eftink, *Biophys. J.* **1994**, *66*, 482.
[3] J. R. Lakowicz, In *Principles of Fluorescence Spectroscopy*, Vol. *3*, Chapter 16; J. R. Lakowicz, Ed.; Springer: New York, **2006**.
[4] C. A. Royer, *Chem. Rev.* **2006**, *106*, 1769.
[5] E. A. Burstein, In *Model Studies: Science and Technology Results, Biophys*, Vol. *6*; VINITI: Moscow, **1976**.
[6] J. T. Vivian, P. R. Callis, *Biophys. J.* **2001**, *80*, 2093.
[7] P. R. Callis, *Methods Enzymol.* **1997**, *278*, 113.
[8] B. Valeur, G. Weber, *Photochem. Photobiol.* **1977**, *25*, 441.
[9] M. R. Eftink, L. A. Selvidge, P. R. Callis, A. A. Rehms, *J. Phys. Chem.* **1990**, *94*, 3469.
[10] K. W. Short, P. R. Callis, *J. Chem. Phys.* **2000**, *113*, 5235.
[11] D. Brisker-Kaiman, A. Dreuw, *ChemPhysChem* **2015**, *16*, 1695.
[12] C. Shen, R. Menon, D. Das, N. Bansal, N. Nahar, N. Guduru, S. Jaegle, J. Peckham, Y. K. Reshetnyak, *Proteins* **2008**, *71*, 1744.
[13] M. S. Najor, K. W. Olsen, D. J. Graham, D. M. Freitas, *Protein Sci.* **2014**, *23*, 1392.
[14] Y. K. Reshetnyak, E. A. Burstein, *Biophys. J.* **2001**, *81*, 1710.
[15] J. M. Martínez, L. Martínez, *J. Comput. Chem.* **2003**, *24*, 819.
[16] L. Martínez, R. Andrade, E. G. Birgin, J. M. Martínez, *J. Comput. Chem.* **2009**, *30*, 2157.
[17] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, *102*, 3586.
[18] A. D. MacKerell, Jr., M. Feig, C. L. Brooks, *J. Comput. Chem.* **2004**, *25*, 1400.
[19] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *102*, 3586.
[20] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten, *J. Comput. Chem.* **2005**, *26*, 1781.
[21] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **1996**, *14*, 33.
[22] L. Martínez, MDAnalysis. Version 17.224. http://leandro.iqm.unicamp.br/mdanalysis, **2017**.
[23] A. Vallée-Bélisle, S. W. Michnick, *Nat. Struct. Mol. Biol.* **2012**, *19*, 731.
[24] C. A. Royer, C. J. Mann, C. R. Matthews, *Protein Sci.* **1993**, *2*, 1844.
[25] G. Wright, R. B. Freedman, *Protein Eng.* **1989**, *2*, 583.
[26] R. Swaminathan, G. Krishnamoorthy, N. Periasamy, *Biophys. J.* **1994**, *67*, 2013.
[27] E. A. Burstein, N. S. Vedenkina, M. N. Ivkova, *Photochem. Photobiol.* **1973**, *18*, 263.
[28] P. R. Callis, T. Liu, *J. Phys. Chem. B* **2004**, *108*, 4248.
[29] R. F. Chen, J. R. Knutson, H. Ziffer, D. Porter, *Biochemistry* **1991**, *30*, 5184.
[30] A. G. Szabo, D. M. Rayner, *J. Am. Chem. Soc.* **1980**, *102*, 554.
[31] M. R. Batista, L. Martínez, *Biophys. J.* **2013**, *105*, 1670.
[32] P. R. Callis, *J. Mol. Struct.* **2014**, *1077*, 14.
[33] M. Kuhn, K. Johnson, *Applied Predictive Modeling*; Springer: New York, **2013**.
[34] K. Sudhakar, C. M. Phillips, S. A. Williams, J. M. Vanderkooi, *Biophys. J.* **1993**, *64*, 1503.
[35] M. Dockal, M. Chang, D. C. Carter, F. Ruker, *Protein Sci.* **2000**, *9*, 1455.
[36] S. A. Cockle, A. G. Szabo, *Photochem. Photobiol.* **1981**, *34*, 23.
[37] H. Nar, A. Messerschmidt, R. Huber, M. van de Kamp, G. W. Canters, *J. Mol. Biol.* **1991**, *221*, 765.
[38] J. Broos, K. Tveen-Jensen, E. de Waal, B. H. Hesp, J. B. Jackson, G. W. Canters, P. R. Callis, *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 5137.
[39] M. S. Cates, M. B. Berry, E. L. Ho, Q. Li, J. D. Potter, G. N. Phillips, Jr., *Structure* **1999**, *7*, 1269.
[40] G. I. Birnbaum, S. V. Evans, M. Przybylska, D. R. Rose, *Acta Crystallogr. Sect. D* **1994**, *50*, 283.
[41] J. Martinez-Oyanedel, H. W. Choe, U. Heinemann, W. Saenger, *J. Mol. Biol.* **1991**, *222*, 335.
[42] P. Burkhard, P. Taylor, M. D. Walkinshaw, *J. Mol. Biol.* **2000**, *295*, 953.
[43] D. A. Egan, T. M. Logan, H. Liang, E. Matayoshi, S. W. Fesik, T. F. Holzman, *Biochemistry* **1993**, *32*, 1920.
[44] D. R. Rose, J. Phipps, J. Michniewicz, G. I. Birnbaum, F. R. Ahmed, A. Muir, W. F. Anderson, S. Narang, *Protein Eng.* **1988**, *2*, 277.
[45] D. L. Harris, B. S. Hudson, *Chem. Phys.* **1991**, *158*, 353.
[46] D. J. Neidhart, G. A. Petsko, *Protein Eng.* **1988**, *2*, 271.
[47] P. M. Bayley, J. M. Janot, S. R. Martin, *FEB* **1989**, *250*, 389.
[48] T. R. Hynes, R. O. Fox, *Proteins* **1991**, *10*, 92.
[49] L. Bellsolell, J. Prieto, L. Serrano, M. Coll, *J. Mol. Biol.* **1994**, *238*, 489.
[50] M. D. Walkinshaw, W. Saenger, A. Maelicke, *Proc. Natl. Acad. Sci. U S A* **1980**, *77*, 2400.
[51] S. Sugio, A. Kashima, S. Mochizuki, M. Noda, K. Kobayashi, *Protein Eng.* **1999**, *12*, 439.
[52] J. R. Somoza, F. Jiang, L. Tong, C. H. Kang, J. M. Cho, S. H. Kim, *J. Mol. Biol.* **1993**, *234*, 390.
[53] E. A. Merritt, S. Sarfaty, T. T. Chang, L. M. Palmer, M. G. Jobling, R. K. Holmes, W. G. Hol, *Structure* **1995**, *3*, 561.
[54] S. K. Katti, D. M. LeMaster, H. Eklund, *J. Mol. Biol.* **1990**, *212*, 167.
[55] P. R. Callis, W. R. Kirk, F. G. Prendergast, *Biophys. J. (Abstr.)* **2000**, *78*, 127a.
[56] T. C. Terwilliger, D. Eisenberg, *J. Biol. Chem.* **1982**, *257*, 6010.
[57] J. P. Noel, C. A. Bingman, T. L. Deng, C. M. Dupureur, K. J. Hamilton, R. T. Jiang, J. G. Kwak, C. Sekharudu, M. Sundaralingam, M. D. Tsai, *Biochemistry* **1991**, *30*, 11801.
[58] L. Zhao, S. K. Pal, T. Xia, A. H. Zewail, *Angew. Chem. Int. Ed. Engl.* **2003**, *43*, 60.
[59] K. Sasaki, S. Dockerill, D. A. Adamiak, I. J. Tickle, T. Blundell, *Nature* **1995**, *257*, 751.

Received: 10 October 2017
Revised: 22 December 2017
Accepted: 28 January 2018
Published online on 26 February 2018