Calculation of transition dipole moment in fluorescent proteins – Towards efficient energy transfer

6 AUTHORS, INCLUDING:

Hemant Kumar Srivastava
Korea Institute for Advanced Study
25 PUBLICATIONS **454** CITATIONS

SEE PROFILE

Roi Baer
Hebrew University of Jerusalem
111 PUBLICATIONS **4,179** CITATIONS

SEE PROFILE

Maarten Merkx
Technische Universiteit Eindhoven
115 PUBLICATIONS **2,797** CITATIONS

SEE PROFILE

A. Shurki
Hebrew University of Jerusalem
49 PUBLICATIONS **1,310** CITATIONS

SEE PROFILE

PCCP

PAPER

Calculation of transition dipole moment in fluorescent proteins—towards efficient energy transfer†

Tamar Ansbacher,[a] Hemant Kumar Srivastava,[a] Tamar Stein,[b] Roi Baer,[b] Maarten Merkx[c] and Avital Shurki‡*[a]

Förster Resonance Energy Transfer (FRET) between fluorescent proteins (FPs) is widely used to construct fluorescent sensor proteins, to study intracellular protein–protein interactions and to monitor conformational changes in multidomain proteins. Although FRET depends strongly on the orientation of the transition dipole moments (TDMs) of the donor and acceptor fluorophores, this orientation dependence is currently not taken into account in FRET sensor design. Similarly, studies that use FRET to derive structural constrains typically assume a κ^2 of 2/3 or use the TDM of green fluorescent protein, as this is the only FP for which the TDM has been determined experimentally. Here we used time-dependent density functional theory (TD-DFT) methods to calculate the TDM for a comprehensive list of commonly used fluorescent proteins. The method was validated against higher levels of calculation. Validation with model compounds and the experimentally determined TDM of GFP shows that the TDM is mostly determined by the structure of the π-conjugated fluorophore and is insensitive to non-conjugated side chains or the protein surrounding. Our calculations not only provide TDM for most of the currently used FPs, but also suggest an empirical rule that can be used to obtain the TDMs for newly developed fluorescent proteins in the future.

Introduction

Green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* and related fluorescent proteins have revolutionized the field of live cell imaging.[1–3] Because their fluorophores are formed autocatalytically following protein translation,[4] FPs provide excellent genetically encoded tags to monitor the fate of specific proteins within living cells. Protein engineering efforts have yielded an impressive range of GFP variants with improved pH sensitivities, folding efficiencies, maturation rates, and improved photochemical properties, most notably a range of excitation and emission spectra. These different color variants of GFP, together with more recently developed

red variants derived from several *Anthozoa*, are not only ideal for monitoring multiple proteins simultaneously,[5] but also provide excellent donor and acceptor fluorophores for application of Förster Resonance Energy Transfer (FRET). FRET is a photophysical effect in which energy that is absorbed by a donor is transferred non-radiatively to an acceptor fluorophore. The strong distance and orientation dependence of FRET makes it particularly useful to detect conformational changes on the scale of individual proteins. Fusion of donor and acceptor FPs to a conformationally responsive ligand binding domain has been widely employed to construct genetically encoded fluorescent sensor proteins for a range of intracellular messenger molecules and metal ions, but also to image enzyme activities.[3,6] Fusion of donor and acceptor FPs to two interacting proteins allows one to monitor the dynamics of this interaction in real time at the subcellular level. Finally, FRET can be used to obtain distance constrains and thus provide detailed structural information on protein complexes in different conformation states.[5,7]

The efficiency of energy transfer strongly depends on the interfluorophore distance (r) and the Förster distance (R_0) according to the Förster equation (eqn (1)). The Förster distance in turn depends on the quantum yield of the donor (Q_D), the spectral overlap between donor emission and acceptor absorption ($J(\lambda)$), the refractive index of the medium (n) and an orientational factor κ^2, which is related to the relative

[a] Department of Medicinal Chemistry, Institute for Drug Research, The Lise-Meitner Minerva Center for Computational Quantum Chemistry, The Hebrew University of Jerusalem, Jerusalem 91120, Israel. E-mail: avitalsh@ekmd.huji.ac.il; Fax: +972-2-675-7076; Tel: +972-2-675-8696
[b] Fritz Haber Center for Molecular Dynamics, Institute of Chemistry, The Hebrew University of Jerusalem, Jerusalem 91904, Israel
[c] Laboratory of Chemical Biology, Department of Biomedical Engineering, Eindhoven University of Technology, P. O. Box 513, 5600 MB Eindhoven, The Netherlands
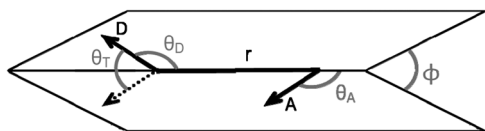
Scheme 1 Definition of angles θ_T, θ_A, θ_D and ϕ that determine κ^2. The two arrows designated by D and A represent the TDM of the donor and acceptor, respectively.

orientation of the donor emission and acceptor absorption dipole moments (eqn (2)).[8]

$$E = \frac{R_0^6}{R_0^6 + r^6} \qquad (1)$$

$$R_0 = 0.211[\kappa^2 Q_D n^{-4} J(\lambda)]^{1/6} \qquad (2)$$

Eqn (3) describes how κ^2 depends on the angles (θ_D, θ_A, θ_T) between the dipole moments of the donor and acceptor, as depicted in Scheme 1. If the orientation of donor and acceptor domains undergoes complete randomization while the donor is in the excited state, κ^2 averages out to a value of 2/3. However, if the orientation is fixed, κ^2 can vary between 0 (perpendicular), 1 (parallel) and 4 (collinear). Therefore changes in relative orientation can have significant effects on the energy transfer efficiency.

$$\kappa^2 = (\cos\theta_T - 3\cos\theta_D\cos\theta_A)^2$$
$$= (\sin\theta_D\sin\theta_A\cos\phi - 2\cos\theta_D\cos\theta_A)^2 \qquad (3)$$

The orientation dependence of FRET has been recognized as an important parameter for FRET sensor performance, *e.g.* to explain the sometimes dramatic effects of circular permutation on FRET sensor performance.[6,9–12] However, the orientation dependence is currently not taken into account in FRET sensor design. Similarly, studies that use FRET to derive structural constrains typically assume a κ^2 of 2/3 or use the transition dipole moment (TDM) of green fluorescent protein, as this is the only FP for which the TDM has been determined experimentally.

Here we used time-dependent density functional theory (TD-DFT) methods to calculate the TDM for a comprehensive list of fluorescent proteins. Validation with model compounds and the experimentally determined TDM of GFP shows that the TDM is almost exclusively determined by the structure of the π-conjugated fluorophore and is insensitive to non-conjugated side chains or protein surrounding. The results of our calculations not only provide TDM for most of the currently used FPs, but also suggest an empirical rule of thumb that can be used to estimate the TDMs for new fluorophores that are not represented by the list reported here.

Computational details

All quantum calculations of the fluorophores unless otherwise specified were carried out using the Gaussian 03 package.[13] Density functional theory (DFT) using the Becke's three-parameter hybrid functional (B3YLP)[14] was employed to optimize all structures with the 6-31G* basis set. Since inclusion of solvent was shown to have negligible effects on the geometry in previous studies of these systems,[15] structural optimization

was carried out in the gas phase. For the evaluation of excitation energies, transition dipole moments and charge distribution, the time dependent DFT (TDDFT) was utilized at the B3LYP/6-31G* level of calculation. Solvent effects, when considered, utilized the polarizable continuum model (IEF-PCM).[16] Charges were calculated using the Merz–Singh–Kollman scheme.[17,18] Since the impact of deviation from planarity on TDMs was found to be very small (see Table S1 in ESI† for more details), the reported TDMs are given for the simplified planar systems resulting from structural optimization in the gas phase. DFT (TDDFT) calculations which employ a correction to the long-range behavior as implemented in the Baer–Neuhauser–Livshits (BNL) range-separated functional were used.[19,20] These latter calculations used the BNL functional which is implemented in Q-Chem3.2.[24] The range parameter of the functional was "tuned" so that two exact conditions in DFT are enforced as closely as possible, namely, that for both the molecule and its anion $\varepsilon_H = -I$, where ε_H is the highest occupied orbital energy and I is the ionization energy. Note that although this tuning procedure does not involve any information on excited states, it has been shown that this procedure leads to successful description of excited states with significant charge transfer character.[21–23] The tuned parameter (γ) was found to be: 0.227 and 0.223 $1/a_0$ for the neutral and anionic models, respectively.

Models for the fluorophores were obtained by removing parts of the protein at appropriate positions and saturating the remaining dangling bonds by a methyl group similar to the models suggested by Tozzini *et al.*[15] Unless mentioned otherwise, side chains that are not π-conjugated were also replaced by a methyl group. All fluorophores, due to the way they are generated (composition of three sequential protein residues), share a similar basic skeleton which contains an imidazolidinone ring conjugated to additional ring/rings.

For simplicity, the direction of the TDM for all the fluorophores is given by ω, the angle with respect to the imidazolidinone carbonyl bond and is drawn as originating from the carbon (see Scheme 2). Knowledge of the absolute direction of the TDM is not important as both directions (X and ($X − 180$)) actually represent the same TDM vector and result with the same κ^2 values. Therefore for the clarity of the presentation and ease of comparison, we chose to present the ω values as obtained by a clockwise rotation from the imidazolidinone carbonyl to the TDM.

The effect of the protein surrounding was evaluated by employing the electrostatic embedding scheme. Here a one-electron term that represents the electrostatic interaction between the protein atoms' partial charges and the reacting system was introduced into the QM Hamiltonian.[25,26] Structural coordinates
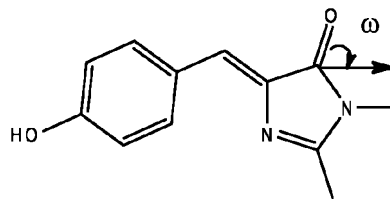
Scheme 2 Structure of the basic skeleton found in GFP, *p*-hydroxybenzylideneimidazolidinone (HBDI), showing the direction of the TDM with respect to the imidazoline carbonyl group.

of the protein atoms were obtained from the Brookhaven Protein Data Bank (PDB). For details see Table S2 (ESI†). For wt-GFP we used the 1EMB structure.[27] The partial charges for the protein were obtained using the ENZYMIX force field.[28,29]

Results and discussion

Choice of the calculation level

TD-DFT methods are relatively efficient in terms of computation effort and have been used previously to calculate various properties of GFP and its derivatives, including excitation energies.[15,30–33] Because the description of excited states with charge transfer (CT) character or with extended π systems using various DFT methods is challenging,[15,34–37] we first compared the performance of these DFT methods in calculating the TDM and excitation energy for *p*-hydroxybenzylideneimidazolidinone (HBDI), a synthetic compound commonly used to model the GFP fluorophore (**1** in Scheme 3).

Table 1 compares the results obtained for both the neutral and deprotonated forms using various levels of calculation and experimental values when available. For anions, all theoretical methods, except for SAC-CI, predict lower excitation wavelengths than reported by experiment. TDDFT predictions are lower than
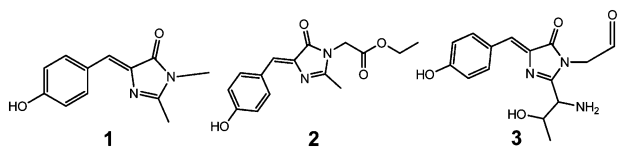
Scheme 3 Various models of the GFP fluorophore. (1) *p*-hydroxybenzylideneimidazolidinone (HBDI). (2) Ethyl 4-(4-hydroxyphenol)-methylidene-2-methyl-5-oxo-1-imidazolacetate. (3) (*Z*)-2-(2-(1-amino-2-hydroxypropyl)-4-(4-hydroxybenzylidene)-5-oxo-4,5-dihydro-1*H*-imidazol-1-yl)acetaldehyde.

Table 1 Vertical excitation wavelengths and transition dipole moment (TDM) directions of the natural and anionic forms of HBDI in the gas phase obtained by various computation levels

		Excitation/nm	$\omega^a/^\circ$
Neutral form			
a	Exp		$(68 \pm 3)^b$
b	CASPT2/CC-PVDZ^c	347	
c	CCSD/CC-PVDZ^c	310	
d	SAC-CI/DZV^d	384^e	73^f
e	TD B3LYP/CC-PVDZ	354	75
f	TD B3LYP/6-31+G*	358	74
g	TD BNL/6-31+G*	349	74
Anionic form			
h	Exp	479^g	$(75 \pm 4)^b$
i	CASPT2/CC-PVDZ^c	425	
j	CCSD/CC-PVDZ^c	408	
k	SAC_CI/DZV^d	559	71^f
l	TD B3LYP/CC-PVDZ	416	74
m	TD B3LYP/6-31+G*	403	73
n	TD BNL/6-31+G*	410	72

^{a} The angle of the TDM as defined in Scheme 2. The acute angle is presented. ^{b} Experimental results were obtained in CD$_3$OD (see ref. 38). ^{c} Data taken from ref. 37. ^{d} Data taken from ref. 39. ^{e} Data taken from ref. 40. ^{f} The angle corresponds to a derivative of the HBDI model, (*Z*)-4-(4-hydroxybenzylidene)-2-methyl-1*H*-imidazol-5(4*H*)-one. The data for the angle was obtained from the authors of ref. 39 in a private communication. ^{g} Taken from ref. 41.

those of CASPT2 but the expected overall trend of the anion having a lower excitation energy than the neutral form is kept. Furthermore, in agreement with previous studies of this particular system,[37] the performance of the TDDFT methods is comparable to those obtained using higher computational levels including CCSD and CASPT2. Importantly, the direction of the TDM obtained from the TDDFT calculations does not significantly depend on the level of calculation. The TDM direction calculated by TDDFT is virtually the same as that obtained by the extensive SAC-CI method, and both are similar to the estimated experimental value. Basis set also exhibit a minor effect on the TDM direction. Finally, the results obtained by the B3LYP functional are very similar to those obtained by the BNL functional which is an improved DFT functional that includes a correction for long range behavior (additional comparisons can be found in Table S3, ESI†). These results imply that the TDM in that system is insensitive to the computational level, showing that even when differences between calculated energies are relatively large, the calculated direction of the TDM remains unaffected.

Comparison of the charge distribution of HBDI both in the ground and the excited states using various computational methods further confirmed the validity of the TDDFT description for our system (Tables S4 and S5, ESI†). We therefore used TDDFT at the B3LYP/6-31+G* level for all further calculations.

Effect of the fluorophore model

Various derivatives of HBDI (Scheme 3) have been used as experimental models of the GFP fluorophore, differing in the non-conjugated components of the molecule. To examine the influence of these variations we calculated vertical excitation energies and TDM directions for the natural and anionic forms of compounds **1–3** and compared them to available experimental data (Table 2). To allow this comparison, excitation energies were calculated in ethanol whereas TDM directions were calculated in methanol.

The calculated excitation wavelengths are in excellent agreement with the available experimental data for the neutral forms (entries a and b). For the anionic form the agreement with the experimentally obtained excitation wavelength is less striking but still satisfactory, exhibiting a maximum deviation

Table 2 Vertical excitation wavelengths in ethanol and transition dipole moment (TDM) directions in methanol of the natural and anionic forms of various models of the GFP fluorophore^{a,b}

	Model^c	Excitation/nm	$\omega^d/^\circ$
Neutral form			
a	**1**	374 (368)^e	73 (68 $\pm$ 3)^f
b	**2**	373 (372)^g	73
c	**3**	386	77
Anionic form			
d	**1**	421 (439)^e	71 (75 $\pm$ 4)^f
e	**2**	419 (446)^g	71
f	**3**	430	76

^{a} Results are given at the TD B3LYP/6-31+G*//B3LYP/6-31G* level of calculation. Solvent effect is modelled at the IEF-PCM level. ^{b} Values in parentheses correspond to experimental data. ^{c} Model corresponds to the models presented in Scheme 3. ^{d} The angle of the TDM as defined in Scheme 2. The acute angle is presented. ^{e} Data taken from ref. 42. ^{f} Experimental results were obtained in CD$_3$OD (see ref. 38). ^{g} Data taken from ref. 43.

of 30 nm (entries d and e). The calculated TDM directions are similar for all three model compounds and in agreement with the experimentally determined TDM for model 1. Furthermore, the difference in the calculated TDM values between the charged and neutral forms of each molecule is 2° at most. These values are somewhat similar to the experimentally reported values of Tolbert and co-workers, which exhibit a difference of 7°.[38] We note in this respect that different from the experiment, where the TDM direction of the anionic form exhibits a higher angle than that of the neutral form, our calculations exhibit lower values for the anionic form. All levels of calculations exhibit this behaviour, suggesting that the difference may result from either the difficulty to properly describe this anionic species (as is perhaps suggested by the larger discrepancy in the vertical excitation energies when compared with experiment), the difficulty to properly determine the geometry, or due to particular correlation effects that none of the methods used here (including SAC-CI) accounts for.

These results show that the non-conjugated components of the molecule have virtually no effect on the direction of the TDM and imply that non-conjugated components can be disregarded in further calculations without compromising their accuracy. It also suggests that fluorophores that share the same π-conjugated skeleton, and differ from each other only by the non-conjugated components are likely to have similar TDM. Therefore, in subsequent calculations we replaced side chains that are not π-conjugated by methyl groups.

Effect of the environment

The fluorophore environment can significantly affect its photophysical behaviour, as evidenced by the differences in optical properties displayed by fluorescent proteins containing the same π-conjugated skeleton. To establish the influence of the fluorophore surrounding, we calculated the TDM direction for the HBDI model of GFP in various solvents spanning a wide range of dielectric constants as well as in the gas phase and in the protein interior (Table 3). The latter calculation involved modelling the protein atoms surrounding the fluorophore as point charges. This model provides the closest description of the actual biological system and also takes into account the possible influence of charge distribution on the fluorophore's surrounding.

Table 3 lists vertical excitation wavelengths and TDM directions of both the neutral and the anionic form of the HBDI model in different environments. Excitation energies calculated for the neutral form were found to be insensitive to the surrounding and in very good agreement with the experimental values.[42] Excitation energies calculated for the anionic form exhibited greater sensitivity to the environment but, with the exception of DMSO, also agree reasonably well with experimental values. A negligible effect of the surrounding on the TDM direction was observed, with the TDM exhibiting virtually the same angle (73° ± 1° and 71° ± 2° for the neutral and the anionic forms, respectively, both in various solvents but also in the gas phase and the protein). Most importantly, the calculated value is in good agreement with the TDM that was experimentally derived for GFP.

Table 3 Vertical excitation wavelength and transition dipole moment (TDM) directions of the natural and anionic forms of the HBDI model (1) in various environments[a,b]

		Excitation/nm	$\omega^c/°$
Neutral form			
a	Water	373 (372[d], 370[e])	73
b	DMSO	377 (370)[d]	73 (66 ± 3[f], 62 ± 4[i])
c	Methanol	373 (367)[d]	73 (68±3)[f]
d	Ethanol	374 (368)[d]	73
e	Toluene	376 (362)[d]	74
f	GFP protein[g]	361 (395)[h]	74 (67±4)[i]
g	Gas phase	358	74
Anionic form			
h	Water	417 (426[d], 426[e])	71
i	DMSO	424 (471)[d]	71
j	Methanol	417 (428)[d]	71 (75 ± 4)[f]
k	Ethanol	421 (439)[d]	71
l	Toluene	446 (445)[d]	71
m	GFP protein[g]	463 (475)[i]	70
n	Gas phase	403	73

[a] Results are given at the TD B3LYP/6-31+G*//B3LYP/6-31G* level of calculation. Solvent effect is modelled at the IEF-PCM level. [b] Values in parentheses correspond to experimental data. [c] The angle of the TDM as defined in Scheme 2. The acute angle is presented. [d] Data taken from ref. 42. [e] Data taken from ref. 41. [f] Results obtained in the deuterated form of the solvent (namely, CD₃OD or DMSO-d₆), are taken from ref. 38. [g] GFP protein coordinates are taken from PDB entry code 1EMB.[27] [h] Data taken from ref. 44. [i] Data taken from ref. 45 and supported also by ref. 46.

Transition dipole moment of representative fluorescent proteins

Since the non-conjugated parts of the fluorophore and protein surrounding had virtually no effect on the direction of the TDM vector, we grouped the various FPs according to the structure of the π-conjugated part of their fluorophore, and calculated the TDM for a representative example of each of them (Table 4). As mentioned in the computational section, side chains that are not π-conjugated were replaced by a methyl group and calculations were all carried out in the gas-phase. The fluorophores are represented both by the three letter code defining the original three amino acids comprising the fluorophore (2nd column), and by a drawing of the compound that was actually calculated as obtained after trimming of non-conjugated chains (3rd column). The entries in the table are usually ordered by ascending excitation energies and the respective emission colour is indicated in the 5th column of Table 4. Some of the fluorophores are known to have a major excitation peak representing their protonated form e.g. GFP (entry c) while others major excitation represents their deprotonated form e.g., YFP (entry g). Here we have considered the form that accounts for the major peak and the respective overall charge of the fluorophore is indicated in the 6th column. The direction of the TDM is both depicted pictorially by dashed arrows on the compounds' drawings (3rd column) and is given as the value of ω (7th column). Finally the oscillator strength (OS) of the transition is also given (8th column). A single entry can correspond to various FPs, partially listed in the 4th column with the first representative highlighted in bold. That is, for example the TDM direction calculated for SHG, Ser-His-Gly (entry b), corresponds to both BFP and EBFP as they both share identical fluorophores and differ by only two mutations in the protein environment. Moreover, due to removal of non-conjugated

Table 4 Models of various representative fluorophores, with some of their relevant FP, the emission color, the charge (in a.u.) and the direction of the TDM relative to the C → O bond vector defined by ω (in degrees)[a]
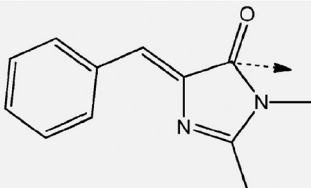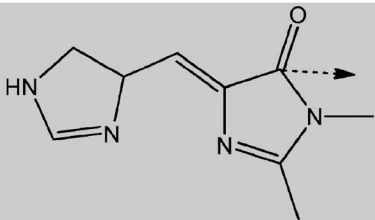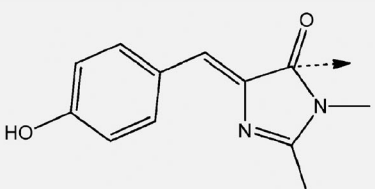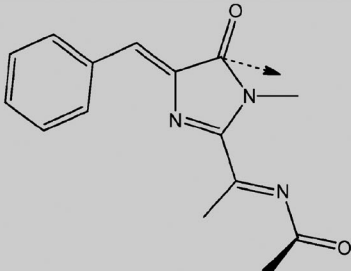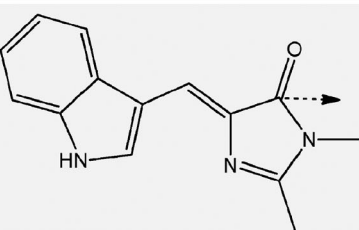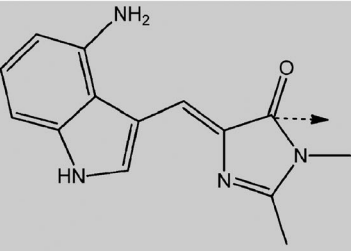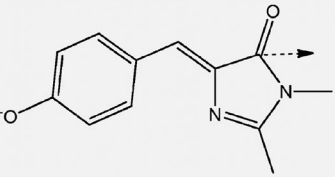
	Code		FP	Color range	C	ω[b]	OS
a	SFG		**BFPF**	Blue	0	75	0.56
b	SHG		**BFP** EBFP EBFP2	Blue	0	80	0.47
c	SYG		**GFP**	Green	0	74	0.70
d	MFG		**mBlueberry**	Blue	0	88 (87)[c]	0.54
e	TWG		**CFP** ECFP CyPet Cerulean	Cyan	0	76	0.60
f	TW'G[d]		**GoldFP**	Yellow	0	77	0.33
g	GYG		**YFP** eGFP mVenus mCitrine YPet	Yellow	-1	73	0.99

Table 4 (*continued*)
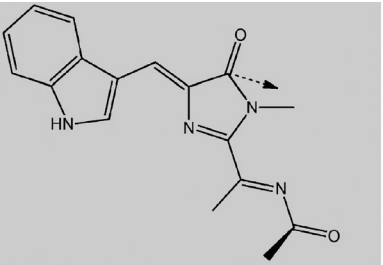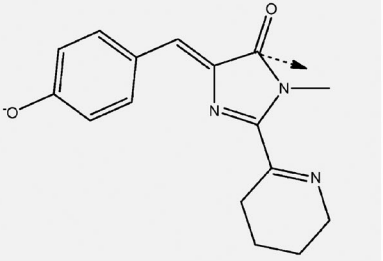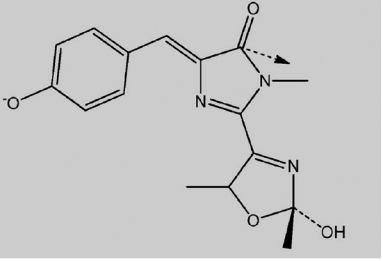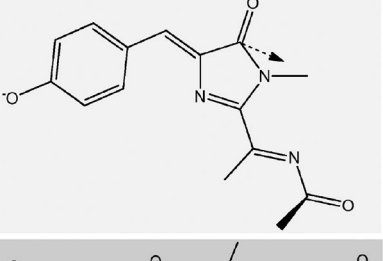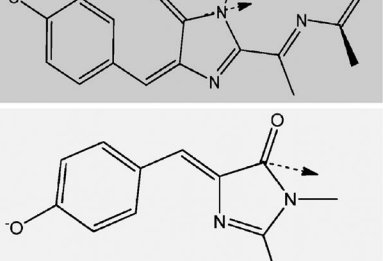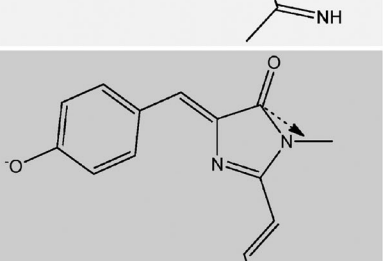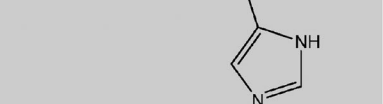
h	MWG		**mHoneydew**	Yellow	0	92 (92)[c]	0.59
i	KYG		**ZFP528**	Yellow	-1	92	0.86
j	TYG		**mOrange**	Orange	-1	94	0.89
k	XYG X=(Q,M, T,C,E)		**DsRed** mCherry mPlum	Red Cyan Violet	-1	98 (99)[c]	0.96
l	MYG		**eqFP**	Far red	-1	123 (123)[c]	1.1[f]
m	MYG		Kindling FP (KFP) family e.g., **asFP595** Dronpa KFP1	Red	-1	88 (90)[e]	0.83
n	HYG		**Kaede** KikGR Dendra2 Eos	Red	-1	111	0.79

Table 4 (*continued*)

[a] Results are given at the TD B3LYP/6-31+G*//B3LYP/6-31G* level of calculation. [b] ω the angle between the TDM vector and the vector along the imidazolidinone carbonyl bond originating at the carbon atom. The absolute direction of the TDM is not important and both directions (X and ($X-180$)) actually provide the same κ^2 values. Therefore for simplicity and consistency the calculated ω values are given here as the clockwise rotation angle from the imidazolidinone carbonyl towards the TDM vector. [c] The values correspond to a non-planar compound since the methoxymethylene formamide group is not planar.[50,51] The values when C_s symmetry was imposed on the fluorophore are given in parentheses. [d] W′ represents a synthetic non-natural Trp-like amino acid: (4-NH$_2$)-Trp. [e] Values correspond to the fluorophore as depicted in the drawing with N-unsubstituted ketimine while in parentheses values correspond to the same fluorophore where the ketimine group is replaced by a carbonyl group. [f] OS values greater than 1 suggesting that there is degeneracy of the electronic states.

elements of the molecule, most of the calculated compounds represent various FPs. For example, both DsRed and mPlum are represented by the same compound in our model (entry k, XYG), even though the first amino acid of the fluorophore is different (Gln and Met, respectively). Finally, depending on the protein environment, the three letter code can correspond to different fluorophore models due to difference in their respective formation process. Thus TYG, Thr-Tyr-Gly, corresponds to the compounds depicted in entries j and k. These various fluorophores are all a result of an autonomous multistep reaction of the three mentioned residues into an imidazolidinone heterocycle, similar to GFP. Yet, compounds in entry j (*e.g.*, mOrange) involve additional cyclization of Thr66 with the preceding carbonyl carbon to yield a partially conjugated oxazole ring whereas compounds in entry k (*e.g.*, DsRed) involve additional acylimine due to oxidation of the Cα–N bond of residue 66 which modifies the amide linkage between residues 65–66 and extends the fluorophore π-system.

The stereochemistry of the exocyclic double bond of the imidazolidinone moiety of the fluorophores can be relevant to their photo-physical behaviour.[15] For example, photo-induced isomerization between the *cis* and *trans* isomers (the *Z* and *E* isomers, respectively) is believed to account for the mechanism by which the nonfluorescent protein converts into a fluorescent protein (*e.g.*, the asFP595 fluorophore, also known as "kindling", entry m).[47–49] We typically only calculated the *cis* (*Z*) isomer, as this is the form that fluoresces, whereas the *trans* (*E*) isomer normally presents a dark state. The only exception involves DsRed and eqFP, which are two different FPs that contain the *cis* (*Z*) and *trans* (*E*) isomers of the same π-conjugated fluorophore skeleton, respectively.

Heuristic rule for guessing the TDM direction in fluorescent proteins

From inspection of the results depicted in Table 4 one finds as expected that the TDM vector always lies in the fluorophore plane. The direction of the TDM gives the polarization of the transition between two states, the ground and excited states. It is thus given by an integration over the product of the two states Ψ_{gs} and Ψ_{ex} (ground and excited states, respectively) and the dipole moment, $\hat{\mu}$, as follows:

$$\text{TDM} = \langle\Psi_{gs}|\hat{\mu}|\Psi_{ex}\rangle = e\langle\Psi_{gs}|\hat{r}|\Psi_{ex}\rangle \quad (4)$$

The excited state should be such that it involves an allowed transition, thus, all the calculated TDMs correspond to the first excitation energy with nonnegligible oscillator strength. These excitations were found to be predominantly HOMO–LUMO in all our reported systems. The HOMO and LUMO orbitals in

these systems are π orbitals delocalized over the entire molecule, and the transition is thus the symmetrically allowed π → π*. Being a singlet, the symmetry of the ground state corresponds to the fully symmetric subgroup (A′). The symmetry of the excited state is a product of the symmetries of the two singly occupied orbitals, namely, the HOMO and LUMO. Both are π-orbitals and thus belong to the same symmetry subgroup (A″). The resulting excited state, therefore, also corresponds to the fully symmetric subgroup (A′). Therefore, only transitions along the molecular plane (X and Y axes) would be symmetrically allowed and lead to a non-zero TDM. Thus, one can conclude that in general the TDM of such fluorophores would always lie in the plane of the fluorophore.

Additional analysis of the results reveals that the TDM typically follows the long axis of the π-conjugated component of the molecule with a maximal deviation of ~13° (Table S6, ESI†). Here, the long axis is defined as the eigenvector that corresponds to the largest eigenvalue of the nuclear quadrupole of the π-conjugated part of the fluorophore (see ESI† for a detailed derivation). Furthermore, looking at the various fluorophore skeletons we find that this direction can often simply be estimated by the imaginary line that connects the two farthest atoms on the π-conjugated element of the molecule. This analysis thus provides a useful rule of thumb to predict the direction of the TDM for any new fluorescent protein with a fluorophore structure that is not represented in Table 4. In cases where the largest distance between two atoms in the fluorophore is clearly defined, the imaginary line that connects them can serve as a good estimate for the TDM direction. For those cases where this largest distance is not clearly defined, one could calculate the direction of the long axis of the π-system using the method described in the ESI.†

Understanding the importance of the angular dependency of FRET

In the absence of experimental values, the TDM of other FPs are often assumed to be the same as that of GFP.[52,53] The calculated TDM obtained here should allow a more accurate interpretation of FRET effects for these FPs, both for quantitative interpretation of natural protein conformational changes and for understanding FRET sensor properties.[52,54] The design of FRET sensors is typically based on maximizing the change in interfluorophore distance, whereas orientational effects are either neglected or the fluorescent domains are assumed to be randomly distributed in both sensor states.[9] However, recent work by Nagai and co-workers suggests that intramolecular interactions between fluorescent domains may play an important role in many of these highly-optimized FRET sensors and explain
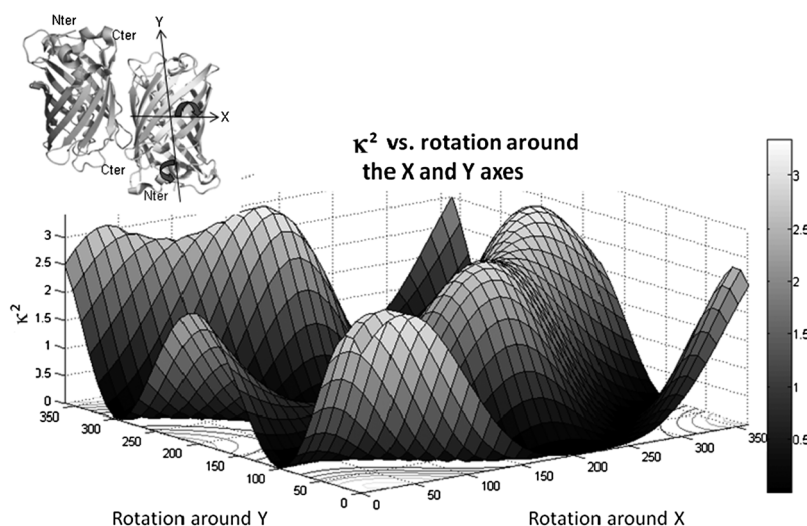
Fig. 1 Variation of κ^2 along different orientations of YFP relative to CFP. Structure of the starting complex showing the definition of X and Y axes is shown at the top left corner of the figure. CFP and YFP are represented by the darker (left) and brighter (right) proteins, respectively, in the complex (figure created with MATLAB[61] and Pymol,[62] respectively).

the sometimes dramatic effects observed by introducing circularly permuted FPs.[10–12,55] Recently, we and others have introduced the use of self-associating fluorescent domains as a design strategy to increase the dynamic range of FRET sensors.[56–59] Also in these sensors, the donor and fluorescent domains form an intra-molecular complex in either the on or the off state in which κ^2 is likely to deviate substantially from 2/3.

To illustrate the relationship between κ^2 and the relative orientation of donor and acceptor fluorescent domains, we calculated κ^2 for a broad range of relative orientations of CFP and YFP, the pair of FPs most commonly used in FRET sensors. We assumed that the calculated excitation TDMs are also good estimates for the emission TDM, which is reasonable since the absorbance and emission TDMs of GFP are known to be nearly identical.[60] As a starting point we took the structure of the GFP dimer in which two GFP domains are oriented in an anti-parallel orientation. The crystallographically determined structures of YFP and CFP were aligned with chains A and B of this GFP dimer, respectively. Following alignment of the calculated fluorophore structures with the crystallographically defined fluorophores, κ^2 was calculated using eqn (3) based on the relative orientation of the calculated TDMs (θ_T, see Scheme 1) and the angles (θ_A and θ_B) between each of the TDM and the line connecting the centres of mass of the two fluorophores. κ^2 was calculated by keeping the position of CFP fixed, while rotating YFP over all possible angles along the X- and Y-axes (Fig. 1 and ESI†). A contour plot shows a complex landscape with multiple minima and maxima. For example, in this starting anti-parallel orientation κ^2 is relatively favourable ($\kappa^2 = 2.5$), but κ^2 decreases rapidly to just 0.1 after a 90° rotation along the Y-axis. This analysis confirms that even subtle changes in relative fluorophore orientation can significantly affect κ^2 and thus energy transfer efficiency. A related plot representing the dependence of the energy transfer efficiency (E in eqn (1)) on the rotation is given in ESI†. These results could thus help to rationalize the significant effects of circular permutation on the ratiometric change observed in the optimization of FRET sensor performance.

Concluding remarks

The TDM direction of GFP was calculated at various levels and validated against experiment. It was shown that calculation of the TDM direction of the π-conjugated part of the respective fluorophore in the gas phase gives sufficiently accurate results. Thus, the TDM direction of various commonly used FPs was then calculated. Based on the results a rule of thumb was derived for approximating the TDM direction in new FPs. The expected variance in κ^2 was demonstrated for a broad range of relative orientations resulting from rotation of YFP with respect to CFP, highlighting the importance of understanding the angular dependence for optimal design of FRET based sensors.

Acknowledgements

References

1 R. Y. Tsien, *Annu. Rev. Biochem.*, 1998, **67**, 509–544.
2 A. B. Cubitt, R. Heim, S. R. Adams, A. E. Boyd, L. A. Gross and R. Y. Tsien, *Trends Biochem. Sci.*, 1995, **20**, 448–455.
3 R. H. Newman, M. D. Fosbrink and J. Zhang, *Chem. Rev.*, 2011, **111**, 3614–3666.
4 M. Ormo, A. B. Cubitt, K. Kallio, L. A. Gross, R. Y. Tsien and S. J. Remington, *Science*, 1996, **273**, 1392–1395.
5 B. N. G. Giepmans, S. R. Adams, M. H. Ellisman and R. Y. Tsien, *Science*, 2006, **312**, 217–224.
6 S. B. Van Engelenburg and A. E. Palmer, *Curr. Opin. Chem. Biol.*, 2008, **12**, 60–65.
7 L. Stryer, *Annu. Rev. Biochem.*, 1978, **47**, 819–846.
8 J. R. Lakowicz, *Principles of fluorescence spectroscopy*, Kluwer Academic/Plenum Publishers, New York, 2nd edn, 1999.
9 R. E. Campbell, *Anal. Chem.*, 2009, **81**, 5972–5979.
10 M. Mank, D. F. Reiff, N. Heim, M. W. Friedrich, A. Borst and O. Griesbeck, *Biophys. J.*, 2006, **90**, 1790–1796.

11 T. Nagai, S. Yamada, T. Tominaga, M. Ichikawa and A. Miyawaki, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 10554–10559.

12 A. E. Palmer, M. Giacomello, T. Kortemme, S. A. Hires, V. Lev-Ram, D. Baker and R. Y. Tsien, *Chem. Biol.*, 2006, **13**, 521–530.

13 G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, M. J. F. Gaussian 03 Revision C.02, Gaussian, Inc, Wallingford, CT, 2004.

14 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.

15 R. Nifosi, P. Amat and V. Tozzini, *J. Comput. Chem.*, 2007, **28**, 2366–2377.

16 S. Miertus, E. Scrocco and J. Tomasi, *Chem. Phys.*, 1981, **55**, 117–129.

17 U. C. Singh and P. A. Kollman, *J. Comput. Chem.*, 1984, **5**, 129–145.

18 B. H. Besler, K. M. Merz Jr. and P. A. Kollman, *J. Comput. Chem.*, 1990, **11**, 431–439.

19 R. Baer and D. Neuhauser, *Phys. Rev. Lett.*, 2005, **94**, 043002.

20 E. Livshits and R. Baer, *Phys. Chem. Chem. Phys.*, 2007, **9**, 2932–2941.

21 R. Baer, E. Livshits and U. Salzner, *Annu. Rev. Phys. Chem.*, 2010, **61**, 85–109.

22 T. Stein, L. Kronik and R. Baer, *J. Am. Chem. Soc.*, 2009, **131**, 2818–2820.

23 T. Stein, L. Kronik and R. Baer, *J. Chem. Phys.*, 2009, **131**, 244119–244124.

24 L. F.-M. Y. Shao, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, R. A. Distasio Jr, R. C. Lochan, T. Wang, G. J. O. Beran, N. A. Besley, J. M. Herbert, C. Y. Lin, T. Van Voorhis, S. H. Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C.-P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. M. Rhee, J. Ritchie, E. Rosta, C. D. Sherrill, A. C. Simmonett, J. E. Subotnik, H. L. Woodcock III, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. Schaefer III, J. Kong, A. I. Krylov, P. M. W. Gill and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2006, **8**, 3172–3191.

25 G. G. Hall and C. M. Smith, *Int. J. Quantum Chem.*, 1984, **25**, 881–890.

26 C. M. Smith and G. G. Hall, *Theor. Chim. Acta*, 1986, **69**, 63–69.

27 F. Yang, L. G. Moss and G. N. Phillips, *Nat. Biotechnol.*, 1996, **14**, 1246–1251.

28 F. S. Lee, Z. T. Chu and A. Warshel, *J. Comput. Chem.*, 1993, **14**, 161–185.

29 Z. Chu, J. Villa, M. Trajbl, C. Schutz, A. Shurki and A. Warshel, *University of Southern California, Los-Angeles*, 2004.

30 I. Demachy, J. Ridard, H. Laguitton-Pasquier, E. Durnerin, G. Vallverdu, P. Archirel and B. Levy, *J. Phys. Chem. B*, 2005, **109**, 24121–24133.

31 Q. K. Timerghazin, H. J. Carlson, C. Liang, R. E. Campbell and A. Brown, *J. Phys. Chem. B*, 2008, **112**, 2533–2541.

32 S. S. Patnaik, S. Trohalaki, R. R. Naik, M. O. Stone and R. Pachter, *Biopolymers*, 2007, **85**, 253–263.

33 M. A. L. Marques, X. Lopez, D. Varsano, A. Castro and A. Rubio, *Phys. Rev. Lett.*, 2003, **90**, 258101.

34 A. Dreuw, G. R. Fleming and M. Head-Gordon, *J. Phys. Chem. B*, 2003, **107**, 6500–6503.

35 A. Dreuw and M. Head-Gordon, *J. Am. Chem. Soc.*, 2004, **126**, 4007–4016.

36 S. Grimme and M. Parac, *ChemPhysChem*, 2003, **4**, 292–295.

37 C. Filippi, M. Ziccheddu and F. Buda, *J. Chem. Theory Comput.*, 2009, **5**, 2074–2087.

38 A. Usman, O. F. Mohammed, E. T. J. Nibbering, J. Dong, K. M. Solntsev and L. M. Tolbert, *J. Am. Chem. Soc.*, 2005, **127**, 11214–11215.

39 A. K. Das, J. Y. Hasegawa, T. Miyahara, M. Ehara and H. Nakatsuji, *J. Comput. Chem.*, 2003, **24**, 1421–1431.

40 J. Y. Hasegawa, K. Fujimoto, B. Swerts, T. Miyahara and H. Nakatsuji, *J. Comput. Chem.*, 2007, **28**, 2443–2452.

41 S. B. Nielsen, A. Lapierre, J. U. Andersen, U. V. Pedersen, S. Tomita and L. H. Andersen, *Phys. Rev. Lett.*, 2001, **87**, 228102.

42 J. Dong, K. M. Solntsev and L. M. Tolbert, *J. Am. Chem. Soc.*, 2006, **128**, 12038–12039.

43 A. A. Voityuk, A. D. Kummer, M. E. Michel-Beyerle and N. Rosch, *Chem. Phys.*, 2001, **269**, 83–91.

44 R. Heim, D. C. Prasher and R. Y. Tsien, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 12501–12504.

45 D. Stoner-Ma, E. H. Melief, J. Nappa, K. L. Ronayne, P. J. Tonge and S. R. Meech, *J. Phys. Chem. B*, 2006, **110**, 22009–22018.

46 X. H. Shi, J. Basran, H. E. Seward, W. Childs, C. R. Bagshaw and S. G. Boxer, *Biochemistry*, 2007, **46**, 14403–14417.

47 D. M. Chudakov, V. V. Belousov, A. G. Zaraisky, V. V. Novoselov, D. B. Staroverov, D. B. Zorov, S. Lukyanov and K. A. Lukyanov, *Nat. Biotechnol.*, 2003, **21**, 191–194.

48 M. L. Quillin, D. A. Anstrom, X. K. Shu, S. O'Leary, K. Kallio, D. A. Chudakov and S. J. Remington, *Biochemistry*, 2005, **44**, 5774–5787.

49 M. Andresen, M. C. Wahl, A. C. Stiel, F. Grater, L. V. Schafer, S. Trowitzsch, G. Weber, C. Eggeling, H. Grubmuller, S. W. Hell and S. Jakobs, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 13070–13074.

50 A. Brahimi, Y. Belmiloud and D. Kheffache, *J. Mol. Struct. (THEOCHEM)*, 2006, **759**, 1–10.

51 K. B. Wiberg, P. R. Rablen and M. Marquez, *J. Am. Chem. Soc.*, 1992, **114**, 8654–8668.

52 J. W. Borst, S. P. Laptenok, A. H. Westphal, R. Kuhnemuth, H. Hornen, N. V. Visser, S. Kalinin, J. Aker, A. van Hoek, C. A. M. Seidel and A. J. W. G. Visser, *Biophys. J.*, 2008, **95**, 5399–5411.

53 D. L. Winters, J. M. Autry, B. Svensson and D. D. Thomas, *Biochemistry*, 2008, **47**, 4246–4256.

54 T. H. Evers, M. A. M. Appelhof, P. T. H. M. De Graaf-Heuvelmans, E. W. Meijer and M. Merkx, *J. Mol. Biol.*, 2007, **374**, 411–425.

55 I. Kotera, T. Iwasaki, H. Imamura, H. Noji and T. Nagai, *ACS Chem. Biol.*, 2010, **5**, 215–222.

56 M. V. Golynskiy, W. F. Rurup and M. Merkx, *ChemBioChem*, 2010, **11**, 2264–2267.

57 J. Huang and S. Koide, *ACS Chem. Biol.*, 2010, **5**, 273–277.

58 J. L. Vinkenborg, T. H. Evers, S. W. A. Reulen, E. W. Meijer and M. Merkx, *ChemBioChem*, 2007, **8**, 1119–1121.

59 J. L. Vinkenborg, T. J. Nicolson, E. A. Bellomo, M. S. Koay, G. A. Rutter and M. Merkx, *Nat. Methods*, 2009, **6**, 737–740.

60 F. I. Rosell and S. G. Boxer, *Biochemistry*, 2003, **42**, 177–183.

61 *MATLAB The MathWorks*, Natick, MA.

62 *The PYMOL Molecular Graphics System*, Version 1.1r1 Schrodinger, LTC.