Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics

Calculation of Transition Dipole Moment in Fluorescent Proteins–Towards Efficient Energy Transfer

Tamar Ansbacher,[1] Hemant Kumar Srivastava,[1] Tamar Stein[2], Roi Baer[2], Maarten Merkx[3] and Avital Shurki[1†]*

[1]*Department of Medicinal Chemistry, Institute of Drug Research, The Lise-Meitner Minerva Center for Computational Quantum Chemistry, The Hebrew University of Jerusalem, Jerusalem 91120, Israel*

[2] Fritz Haber Center for Molecular Dynamics, Institute of Chemistry, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

[3]*Laboratory of Chemical Biology, Department of Biomedical Engineering, Eindhoven University of Technology, P. O. Box 513, 5600 MB Eindhoven, The Netherlands*

Supporting Information

Effect of Deviation from Fluorophore Planarity on the TDM Direction

Calculations of the TDM were performed on optimized structures of the fluorophores. As a result the calculated structure ended up being planar and in some cases partially planar (e.g. mBlueberrry, mHoneydew, DsRed and eqFP have their methoxymethylene formamide group out of plane). The TDM is lying within the plane in all cases. We note that symmetry was not imposed on the structures during the optimization, but was obtained as its result. The experimentally known structures are usually not entirely planar and sometimes deviation from planarity is significant (e.g., the fluorophore of asFP525 as given in entry code 1A52 of the Protein Data Bank). We decided therefore, to examine the impact of the geometry and in particular the symmetry on the resulting TDM. Table S1 compares the TDM values obtained for both calculated and experimental structures of each one of the fluorophores where experimental data is available. The experimental structure relates to the first FP representative of each fluorophore (Table 4 in the text, 4th column in bold). Coordinates in that case are taken from the crystallographic data as published in the Protein Data Bank (see Table S2) with no further optimization. Similar to the optimized structures, the TDM was calculated using only the π-conjugated portion of the molecule in each case and side chains which are not π-conjugated were replaced by a methyl group. Overall, the TDM directions calculated based on the experimental and the gas-phase optimized structures are very similar. The highest deviation that is found is 6º (CFP). It is therefore reasonable for further discussion to relate to a simplified planar system in all cases, as the impact on the TDM direction seems to be small and in the same order as the error of the calculation.

Table S1. Comparison of calculated TDM direction, ω (in degrees), of experimental vs. computationally optimized structures of various FP fluorophores.[a]

FP[b]	Charge	ω for Exp. Structure[c]	ω for Opt. Structure[d]
BFP (SHG)	0	76	80
GFP (SYG)	0	80	74
CFP (TWG)	0	82	76
GoldFP (TW'G)	0	81	77
ZFP528 (KYG)	-1	95	92
mOrange (TYG)	-1	93	94
DsRed (QYG)	-1	94	98 (99)[e]
eqFP (MYG)	-1	125	123
asFP595 (MYG)	-1	88	88
Kaede (HYG)	-1	111	111

[a] TDM was calculated at the TD B3LYP/6-31+G* level of calculation for all structures. ω is the angle between the TDM vector and the vector along the imidazolidinone carbonyl bond originating at the carbon atom.
[b] The three letter code representing the fluorophore is depicted in parenthesis. The actual molecule for which TDM was calculated can be found in the 3rd column of the relevant entry in Table 4 in the text.
[c] A full list of the exact protein source and protein data bank (PDB) entry codes of the relevant proteins is given in Table S2 in Supporting Information. When the PDB involved a multimer rather than a monomer, the corresponding fluorophore structure was taken from the first chain.
[d] Structures were optimized at the B3LYP/6-31G* level of calculation.
[e] The value corresponds to a non-planar compound since the methoxymethylene formamide group is not planar.[1, 2] The values when Cs symmetry was imposed on the fluorophore are given in parenthesis.

Table S2 Source and protein data bank (PDB) entry code of the various FP fluorophores presented in this work.

FP[a]	source	PDB code
BFP (SHG)	Engineered	1BFP[3]
GFP (SYG)	Auquorea Victoria	1EMB[4]
CFP (TWG)	Engineered	1OXD[5]
Gold FP (TW'G)	Engineered	1OXF[5]
Zfp (KYG)	Zoanthus yellow	1XAE[6]
mOrangeTYG)	Engineered	2H5O[7]
Dsred (QYG)	Discosoma red	1GGX[8]
eqFP (MYG)	Entacmaea Quadricolor	1UIS[9]
asFP595 (MYG)	Anemonia sulcata	2A52[10]
Kaede (HYG)	Trachyphyllia geoffroyi	2GW4[11]

[a] The three letter code representing the fluorophore is depicted in parenthesis. The actual molecule for which TDM was calculated can be found in the 3rd column of the relevant entry in Table 4 in the text.

Comparison of TD-B3LYP and TD-BNL

Further support for the validity of TD-B3LYP for this study is given by comparison of its results to the results obtained by TD-BNL. The BNL functional involves implementation of a recent improvement of DFT which includes a correction to long-range behavior.[12] The predictive power of this range separated hybrid functional (BNL) was proved to be reliable for charge transfer excitations of molecular complexes.[13, 14] Thus, the performance of TD-B3LYP in calculating the TDM direction is compared to the performance of TD-BNL. Three different fluorophores have been examined, GFP, GoldFP and Kaede. The results obtained by the two functionals are very similar indicating that the B3LYP functional is sufficient for the purposes of this study.

Table S3. Comparison of calculated TDM direction, ω (in degrees), of BNL functional vs. B3LYP functional

FP[b]	Charge	ω for B3LYP[c]	ω for BNL[c]
GFP (SYG)	0	74	74
GoldFP (TW'G)	0	77	76
Kaede (HYG)	-1	111	107

[a]ω is the angle between the TDM vector and the vector along the imidazolidinone carbonyl bond originating at the carbon atom.
[b] The three letter code representing the fluorophore is depicted in parenthesis. The actual molecule for which TDM was calculated can be found in the 3rd column of the relevant entry in Table 4 in the text.
[c] Structures are optimized at the B3LYP/6-31G* level of calculation. TDM calculations were done using Q-Chem [15]

Comparison of charge distribution

DFT vs. Ab-initio

To further confirm the validity of the TDDFT description for our system we analyzed the charge distribution of the HBDI in the ground and the excited states. We then compared the results to results obtained by *ab-initio* approach. Thus, we compared our results, obtained with B3LYP/6-31+G* (for ground state, GS) and TD B3LYP/6-31+G* (for excited state, Ex) to those obtained by Helms et al.[16] at the RHF/6-31G* and CIS/6-31G* levels of calculation. For simplicity the fluorophore was divided into three groups: the phenol ring, the bridge atoms and the imidazolidinone ring.

Table S4 summarizes the overall partial charges on these groups in the ground and excited states (GS and Ex, respectively) as well as the respective difference (Δ_{Ex-GS}). The charges of the excited state and hence the difference, correspond to a vertical excitation. The results for both the neutral and the anionic forms of the fluorophores are presented. The difference in the charge distribution between the two states represents both the amount and the direction of CT upon excitation. Despite the fact that the partial charges obtained by the two approaches are not the same, the

difference between the two states is usually very similar. It is thus concluded that for this system TDDFT gives a sufficiently reliable description, which is comparable with *ab-initio* based methods. We therefore used TDDFT at the B3LYP/6-31+G* level for all further calculations.

Table S4. Gas phase charge distribution of HBDI in the ground and first excited states along with the respective difference calculated by *ab initio* vs. DFT approaches.[a]

	Computational approach	Phenyl ring	Bridging atoms	Imidazolidinone ring
Neutral form				
GS[b]	*ab initio*	0.23	-0.20	-0.03
	DFT	0.16	-0.06	-0.10
Ex[b]	*ab initio*	0.30	-0.34	0.04
	DFT	0.25	-0.27	0.02
Δ_{Ex-GS}[b]	*ab initio*	0.07	-0.14	0.07
	DFT	0.09	-0.21	0.10
Anionic form				
GS[b]	*ab initio*	-0.49	-0.04	-0.47
	DFT	-0.54	0.03	-0.49
Ex[b]	*ab initio*	-0.26	-0.50	-0.24
	DFT	-0.33	-0.27	-0.31
Δ_{Ex-GS}[b]	*ab initio*	0.23	-0.46	0.23
	DFT	0.21	-0.30	0.08

[a] Charges calculated at the RHF/6-31G* vs. B3LYP/6-31G* levels for the ground state and at the CIS/6-31G* vs. TD B3LYP/6-31G* levels for the excited state. Charges are obtained using the Mertz-Kollman procedure.[17, 18]

[b] GS and Ex stand for ground and excited states respectively, whereas Δ_{Ex-GS} stands for their respective difference.

TD-B3LYP vs. TD-BNL

Additional comparison to results obtained by TD-BNL which was proved to properly describe the distribution in cases of charge separation,[13, 14] is given in Table S5. The three different fluorophores, GFP, GoldFP and Kaede, have been examined. The results obtained by the two functionals are again usually similar indicating that the B3LYP functional is probably sufficient for the purposes of this study.

Table S5. Mulliken charge distribution of GFP, GoldFP and Kaede in the ground and first excited states along with the respective difference calculated by TD-B3LYP vs. TD-BNL.[a]

	Computational Level	Phenyl ring[d]	Bridging atoms	Imidazolidinone ring	Additional chain[e]
GFP(SYG)[b]					
GS[c]	TD-BNL	-0.10	-0.03	0.12	
	TD-B3LYP	0.03	-0.06	0.03	
Ex[c]	TD-BNL	-0.01	-0.25	0.26	
	TD-B3LYP	0.12	-0.26	0.14	
Δ_{Ex-GS}[c]	TD-BNL	-0.08	0.22	-0.14	
	TD-B3LYP	-0.08	0.20	-0.12	
GoldFP (TW'G)					
GS[c]	TD-BNL	-0.69	0.10	0.59	
	TD-B3LYP	-0.33	-0.08	0.41	
Ex[c]	TD-BNL	-0.28	-0.20	0.48	
	TD-B3LYP	0.28	-0.38	0.10	
Δ_{Ex-GS}[c]	TD-BNL	-0.40	0.29	0.11	
	TD-B3LYP	-0.61	0.30	0.31	
Kaede (HYG)					
GS[c]	TD-BNL	-0.61	-0.19	0.34	-0.54
	TD-B3LYP	-0.50	-0.18	0.34	-0.66
Ex[c]	TD-BNL	-0.40	-0.34	0.62	-0.88
	TD-B3LYP	-0.19	-0.27	0.62	-1.16
Δ_{Ex-GS}[c]	TD-BNL	-0.20	0.15	-0.29	0.34
	TD-B3LYP	-0.31	0.09	-0.28	0.50

[a] Calculations were done using Q-Chem[15], using the Mulliken procedure.
[b] The three letter code representing the fluorophore is depicted in parenthesis. The actual molecule for which TDM was calculated can be found in the 3rd column of the relevant entry in Table 4 in the text
[c] GS and Ex stand for ground and excited states respectively, whereas Δ_{Ex-GS} stands for their respective difference.
[d] 3H-inden-4-amine for GoldFP (see Table 4 entry f in text)
[e] (E)-5-(prop-1-enyl)-1H-imidazole for kaede only (see Table 4 entry n in text)

Deviation of the TDM from the Long axis of the molecule

Table S6 Deviation (in degrees) of the TDM from the long axis defined as the eigenvector that corresponds to the largest eigenvalue of the nuclear quadrupole.

FP[a]	Deviation
SFG (BFP)	6
SHG (BFPF)	3
SYG (GFP)	3
GYG (YFP)	4
MFG (mBlueberry)	14 (8)[b]
TWG (CFP)	12
TW'G (GoldFP)	13
MWG (mHoneydew)	11 (5)[b]
KYG (zfp528)	8
TYG (mOrange)	10
XYG (Dsred)	0 (1)[b]
MYG(eqFP)	2 (0)[b]
MYG (asFP595)	3 (4)[c]
HYG (Kaede)	3

[a] The three letter code representing the fluorophore and a representative FP is depicted in parenthesis. The actual molecule for which TDM was calculated can be found in the 3rd column of the relevant entry in Table 4 in the text.
[b] The value corresponds to a non-planar compound since the methoxymethylene formamide group is not planar.[1, 2] The values when Cs symmetry was imposed on the fluorophore are given in parenthesis.
[c] Values correspond to the fluorophore as depicted in the drawing with N-unsubstituted ketimine while in paranthesis values correspond to the same fluorophore where the ketimine group is replaced by a carbonyl group.

All calculations unless otherwise mentioned were done using the Gaussian 03 package.[19]

Deriving the eigenvectors of the nuclear quadrupole moment:

a) Calculate center of charge of the molecule by:

$$X = \frac{\sum_i q_i x_i}{\sum_i q_i}; Y = \frac{\sum_i q_i y_i}{\sum_i q_i}; Z = \frac{\sum_i q_i z_i}{\sum_i q_i}$$

where q_i is the charge of i*th* nuclei, and x_i, y_i and z_i are the corresponding coordinates describing its position.

b) Define the center of charge (the above X, Y, Z) as {0.0, 0.0, 0.0} and translate the coordinates of the entire molecule accordingly.

c) Calculate the traceless matrix elements of the nuclear quadrupole moment, Q_{kj} by:

$$Q_{kj} = -\sum_i q_i ((x_k)_i (x_j)_i - r_i^2 \delta_{kj})$$

where x_k and x_j run over the Cartesian coordinates x y and z of each i*th* nuclei and q is the charge of the i*th* nuclei. δ is the Dirac delta function, which is zero for all values except where k=j (meaning the xx, yy , zz matrix elements), and $r_i^2=x_i^2+y_i^2+z_i^2$. For example:

$$Q_{xx} = \sum_i q_i (y_i y_i + z_i z_i) \quad ; \quad Q_{xy} = -\sum_i q_i x_i y_i \quad ; \quad etc$$

d) Diagonalize the 3x3 matrix Q and find the highest eigenvalue.

e) The eigenvector that corresponds to this eigenvalue is the long axis of the molecule.

Calculation details of κ^2 and energy transfer efficiency values for different orientations of YFP relative to CFP

The structure of the GFP homo-dimer (PDB entry code 1GFL[4]) was used as the starting structure. The structures of both YFP and CFP (PDB entry codes 1YFP[20] and 1OXD[5], respectively) were then aligned with chains A and B of the GFP dimer in the following manner:

1. The coordinates of chain A of YFP were first transformed (translated and rotated) so that the center of nuclear charge defines YFP's origin (0.0 0.0 0.0) and the eigenvectors of the nuclear quadrupole define the X, Y, Z axes. This orientation will be referred to as the standard orientation of YFP.
2. We then aligned chain A of the GFP dimer with chain A of YFP keeping YFP in its standard orientation (RMSD=0.223 for the alignment). This orientation of GFP will be referred to as the new orientation.
3. Finally we aligned CFP with chain B of the GFP dimer now keeping GFP in its new orientation (RMSD=0.359 for the alignment). This orientation of CFP will be referred to as the new orientation.

The resulting YFP-CFP complex (involving YFP in its standard orientation and CFP in its new orientation) served as the so called "starting complex structure" (depicted in Figure 1b in the text).

The calculated fluorophore structures of both YFP and CFP were aligned with the crystallographically defined fluorophores in the "starting complex structure" (RMSD=0.032 and RMSD=0.109 for YFP and CFP respectively), keeping the complex structure intact) and the respective TDM vectors were transformed and expressed by this new coordinates system.

Having the relative orientations of the two TDM as well as the angle between each TDM and a virtual line connecting the center of mass of the two fluorophores we could calculate θ_T, θ_D and θ_A and therefore also κ^2 (equation 3 in the text).

Finally, we rotated YFP over all possible angles along the X- and Y-axes every 10º while keeping the coordinates of CFP fixed and recalculated κ^2 (Figure 1 in the text).

In order to obtain the respective energy transfer efficiency values, the Förster distance (R_0) was first calculated based on the following Eq:

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

using κ^2 values that were previously calculated. n, the refractive index of the medium and $Q_{D,}$ the quantum yield of the donor were given a value of 1.4 and 0.4 respectively. All values were taken from the work of Evers et al.[21] A representative value for the overlap integral, $J(\lambda)$, of 2.29 $10^{-15}M^{-1}cm^{-1}nm^4$, was used.

Finally, using the Förster equation shown below, the efficiency was calculated using a constant typical value of 40Å for r.

$$E = \frac{R_0^6}{R_0^6 + r^6}$$

The corresponding plot of the relative efficiency values versus the rotation angles is presented in figure 1S.

All alignments were done by PYMOL[22].

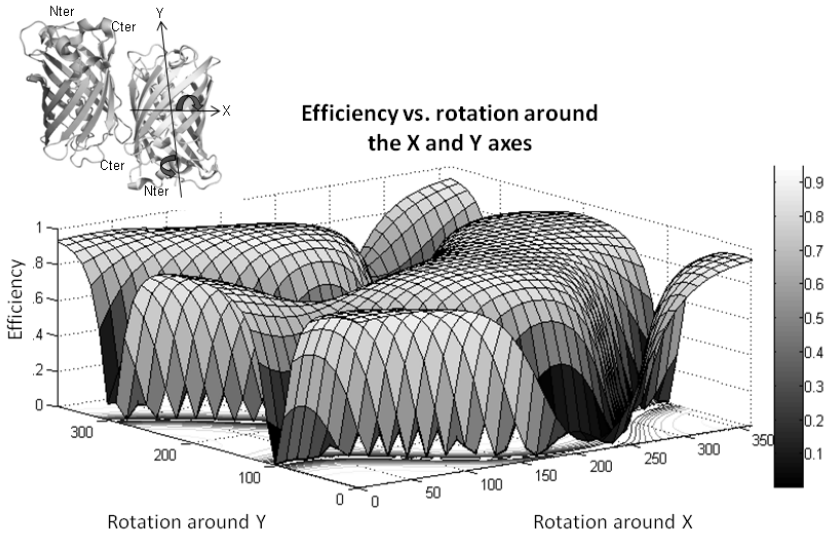

Figure 1S Variation of the FRET efficiency along different orientations of YFP relative to CFP. Structure of the starting complex showing the definition of the X and Y axes is shown at the top left corner of the Figure. CFP and YFP are represented by the darker (left) and brighter (right) proteins, respectively, in the complex (Figure created with MATLAB[23] and Pymol[22], respectively).

Supporting References

1. A. Brahimi, Y. Belmiloud and D. Kheffache, *J. Mol. Struc-Theochem*, 2006, **759**, 1-10.
2. K. B. Wiberg, P. R. Rablen and M. Marquez, *J. Am. Chem. Soc.* , 1992, **114**, 8654-8668.
3. R. M. Wachter, B. A. King, R. Heim, K. Kallio, R. Y. Tsien, S. G. Boxer and S. J. Remington, *Biochemistry*, 1997, **36**, 9759-9765.
4. F. Yang, L. G. Moss and G. N. Phillips, *Nat. Biotechnol.*, 1996, **14**, 1246-1251.

5. J. H. Bae, M. Rubini, G. Jung, G. Wiegand, M. H. J. Seifert, M. K. Azim, J. S. Kim, A. Zumbusch, T. A. Holak, L. Moroder, R. Huber and N. Budisa, *J. Mol. Biol.*, 2003, **328**, 1071-1081.
6. S. J. Remington, R. M. Wachter, D. K. Yarbrough, B. Branchaud, D. C. Anderson, K. Kallio and K. A. Lukyanov, *Biochemistry*, 2005, **44**, 202-212.
7. X. K. Shu, N. C. Shaner, C. A. Yarbrough, R. Y. Tsien and S. J. Remington, *Biochemistry*, 2006, **45**, 9639-9647.
8. M. A. Wall, M. Socolich and R. Ranganathan, *Nat. Struct. Biol.*, 2000, **7**, 1133-1138.
9. J. Petersen, P. G. Wilmann, T. Beddoe, A. J. Oakley, R. J. Devenish, M. Prescott and J. Rossjohn, *J. Biol. Chem.*, 2003, **278**, 44626-44631.
10. 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.
11. I. Hayashi, H. Mizuno, K. I. Tong, T. Furuta, F. Tanaka, M. Yoshimura, A. Miyawaki and M. Ikura, *J. Mol. Biol.*, 2007, **372**, 918-926.
12. R. Baer and D. Neuhauser, *Phys. Rev. Lett.*, 2005, **94**.
13. T. Stein, L. Kronik and R. Baer, *J. Am. Chem. Soc.* , 2009, **131**, 2818-2820.
14. T. Stein, L. Kronik and R. Baer, *J. Chem. Phys.*, 2009, **131**, 244119-244124.
15. 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, M. Head-Gordon,, *Phys. Chem. Chem. Phys.*, 2006, **8**, 3172-3191.
16. V. Helms, C. Winstead and P. W. Langhoff, *J. Mol. Struc-Theochem*, 2000, **506**, 179-189.
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. M. J. F. Gaussian 03 Revision C.02, 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, and J. A. Pople, Gaussian, Inc., Wallingford CT, 2004.
20. R. M. Wachter, M. A. Elsliger, K. Kallio, G. T. Hanson and S. J. Remington, *Struct. Fold. Des.*, 1998, **6**, 1267-1277.
21. T. H. Evers, E. M. W. M. van Dongen, A. C. Faesen, E. W. Meijer and M. Merkx, *Biochemistry*, 2006, **45**, 13183-13192.
22. *The PYMOL Molecular Graphics System, Version 1.1r1 Schrodinger, LTC.*
23. *MATLAB The MathWorks, Natick, MA*.