

Supporting Information

Spatial and electronic correlations in the PE545 light-harvesting complex

Lucas Viani[†], Carles Curutchet[‡], and Benedetta Mennucci[†]

[†]Dipartimento di Chimica e Chimica Industriale, Università di Pisa, via Risorgimento 35, 56126 Pisa, Italy.

[‡]Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Av. Joan XXIII s/n, 08028 Barcelona, Spain.

Correlation Coefficient

The *Pearson coefficients* are commonly used to estimate the correlation between two one-dimensional sets of data but it can be extended to a multidimensional space and in the particular case of three-dimensions. In such a case the correlation coefficient is defined as the covariance of the positional fluctuations (Δr_i) normalized by their standard deviations.

$$I_{ij}^P = \frac{\text{cov}(\Delta r_i, \Delta r_j)}{[\text{var}(\Delta r_i) \cdot \text{var}(\Delta r_j)]^{1/2}} = \frac{\langle (\mathbf{r}_i - \bar{\mathbf{r}}_i) \cdot (\mathbf{r}_j - \bar{\mathbf{r}}_j) \rangle}{[\langle (\mathbf{r}_i - \bar{\mathbf{r}}_i)^2 \rangle \cdot \langle (\mathbf{r}_j - \bar{\mathbf{r}}_j)^2 \rangle]^{1/2}} \quad (1)$$

where $\mathbf{r}_i$ is a three-dimensional vector and $\bar{\mathbf{r}}_i$ is its mean value.

This methodology is based on the assumption of collinear motion assuming a Gaussian distribution in the space of configurations, and is dependent on the relative orientation of the positional fluctuations of the random variables.¹ One way to avoid these limitations and compute any correlation motion is the use of an approach based on the mutual information (MI) between the positional fluctuations.

The MI approach is based on the calculation of the entropy of the random variable (H_{ij}) and its marginal entropies (H_i). In the case of correlations between pairs of atoms it is then defined as,^{2,3}

$$I_{ij}^{MI} = H_i + H_j - H_{ij} \quad (2)$$

and it relies on the well-known inequality, $H_{ij} \leq H_i + H_j$, which becomes an equality if the variables are independent. In the general MI framework, indicated here as GMI, the entropies are extracted from a MD trajectory using the nearest-neighbors estimator developed by Kraskov et al.,⁴ with the k parameter equals to 6 (number of nearest-neighbors), leading to the generalized correlation coefficient (I_{ij}^G). The impact of the nonlinear contributions in the correlations can be singled out by confronting the GMI approach with the so-called linear mutual information (LMI) approach.¹ In this approximation, only the collinear contributions are captured since it is assumed a Gaussian probability distribution of the random variables. The LMI approach is similar to the Pearson method although it is not restricted to the relative orientation of the atomic spatial fluctuations. In this framework, assuming the normal distribution,^{2,3}

$$g_{ij} = \frac{1}{[(2\pi)^n \det \mathbf{C}_{(ij)}]^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \mathbf{v} \mathbf{C}_{(ij)}^{-1} \mathbf{v}^T\right) \quad (3)$$

with $\mathbf{v}$ the positional fluctuation vector, $\det \mathbf{C}_{(ij)}$ the determinant of their covariance matrix, and n the number of components, an analytical formulation of the entropy is obtained,^{2,3}

$$H_{ij} = \frac{1}{2} [2n(1 + \ln 2\pi) - \ln \det \mathbf{C}_{(ij)}] \quad (4)$$

Thus, from Eq. (2), the LMI correlation coefficient is expressed as,

$$I_{ij}^L = \frac{1}{2} [\ln \det \mathbf{C}_{(i)} + \ln \det \mathbf{C}_{(j)} - \ln \det \mathbf{C}_{(ij)}] \quad (5)$$

The MI approaches provide correlation coefficients varying between zero and infinity, with no obvious intuitive interpretation as in the Pearson coefficients. Therefore MI values were rescaled to the [0.0, 1.0] by applying a nonlinear relation,^{1,5} which may relatively larger coefficients, especially for the lower ones.

Computational details

The MD simulations of PE545 in water and in vacuo were performed using the parm99SB⁶ and the GAFF⁷ Amber force fields to describe the protein and the pigments, respectively; in addition,

for the simulation in water, the solvent was described using the TIP3P water model.⁸ All the details can be found in ref. [9]

The molecular representation as well as the MD trajectory analysis was done using the MView¹⁰ software, plots were created using the Qtiplo¹¹ software.

In the QM/MMPol calculations, the protein residues, solvent molecules, and bilin chromophores not included in the QM region, were described using atom centered charges and isotropic polarizabilities. The corresponding polarizable FF was derived from DFT calculations using the scheme outlined in Ref. [12]. Atomic isotropic polarizabilities were calculated at the DFT(B3LYP)/aug-cc-pVDZ level using the LoProp¹³ approach as implemented in the Molcas code,¹⁴ whereas atomic point charges were obtained from DFT(B3LYP)/cc-pVTZ ESP calculations followed by RESP constrained fittings as implemented in the Gaussian09¹⁵ and Amber9¹⁶ programs, respectively. Charges and polarizabilities were derived corresponding to the initial crystal structure and then used in all QM/MM calculations. The parameters for water, derived using the same strategy, were taken from Ref.[17].

All QM/MMPol calculations were performed using a locally modified version of Gaussian09.¹⁵

References

- (1) Lange, O. F.; Grubmüller, H. *Proteins* **2006**, *62*, 1053–61.
- (2) Hyvärinen, A.; Karhunen, J.; Oja, E. *Independent Component Analysis*; Wiley-Interscience, Ed.; Firrst.; 2001.
- (3) Cover, T. M.; Thomas, J. A. *Elements of Information Theory*; Second.; Wiley-Interscience, 2006.
- (4) Kraskov, A.; Stögbauer, H.; Grassberger, P. *Physical Review E* **2004**, *69*, 1–16.
- (5) Khan, S.; Bandyopadhyay, S.; Ganguly, A. R.; Saigal, S.; Erickson, D. J.; Protopopescu, V.; Ostrouchov, G. *Physical review. E, Statistical, nonlinear, and soft matter physics* **2007**, *76*, 026209.
- (6) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.

- (7) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. a; Case, D. a *J. Comput. Chem.* **2004**, *25*, 1157–74.
- (8) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
- (9) Curutchet, C.; Novoderezhkin, V. I.; Kongsted, J.; Muñoz-Losa, A.; Van Grondelle, R.; Scholes, G. D.; Mennucci, B. *J. Phys. Chem. B* **2012**.
- (10) Viani, L. MView: Designed as a tool for visualization and analysis of molecular properties. www.mview-tools.com.
- (11) QtiPlot Data Analysis and Scientific Plotting; <http://soft.proindependent.com/qtiplot.html>.
- (12) Söderhjelm, P.; Ryde, U. *J. Phys. Chem. A* **2009**, *113*, 617–27.
- (13) Gagliardi, L.; Lindh, R.; Karlström, G. *J. Chem. Phys.* **2004**, *121*, 4494–500.
- (14) Karlström, G.; Lindh, R.; Malmqvist, P.-Å.; Roos, B. O.; Ryde, U.; Veryazov, V.; Widmark, P.-O.; Cossi, M.; Schimmelpfennig, B.; Neogady, P.; Seijo, L. *Comput. Materials Science* **2003**, *28*, 222–239.
- (15) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Jr., J. A. M.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomas, J.; Coss, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Vot, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; A. D. Daniels, Ö. F.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09* **2009**.
- (16) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M. *Amber 9* **2006**.
- (17) Curutchet, C.; Muñoz-Losa, A.; Monti, S.; Kongsted, J.; Scholes, G. D.; Mennucci, B. *J. Chem. Theory Comput.* **2009**, *5*, 1838–1848.

Additional Tables and Figures

	DBV _A	DBV _B	PEB _{158C}	PEB _{158D}	PEB _{50/61C}	PEB _{50/61D}	PEB _{82C}
DBV _B	45.27						
PEB _{158C}	20.86	48.35					
PEB _{158D}	47.01	20.00	45.67				
PEB _{50/61C}	31.75	24.07	26.21	38.40			
PEB _{50/61D}	23.19	31.16	24.24	37.07	16.89		
PEB _{82C}	34.52	24.72	37.16	22.76	23.17	22.72	
PEB _{82D}	24.08	34.57	39.67	25.29	34.14	33.06	36.71

Table S1: Mean intermolecular distances averaged over the 60.000 frames of the simulation. All the distances are given in angstroms.

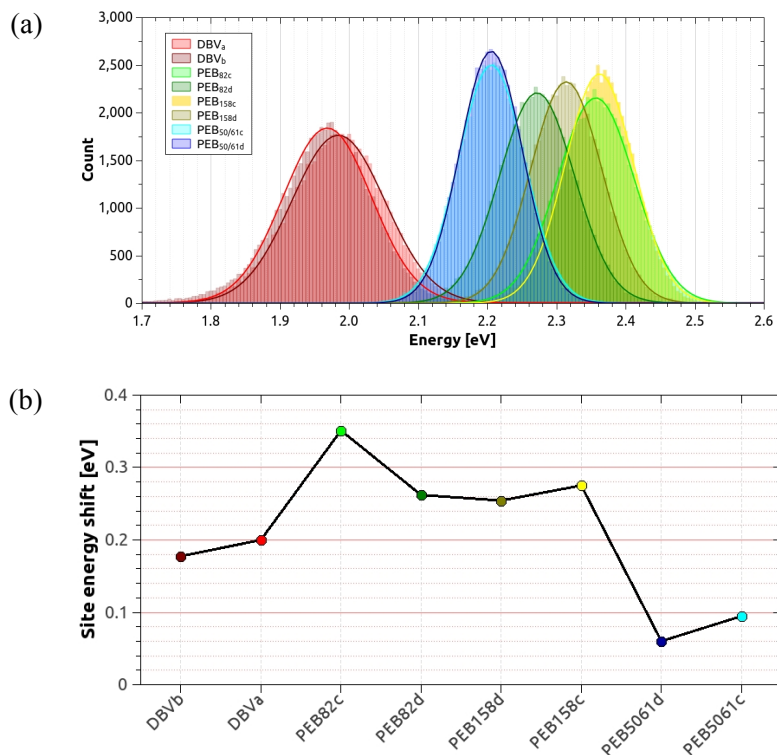


Figure S1: (a) Distribution of the site energies computed using the QM/MM method. (b) Variation of the mean site energies computed using a QM/MMPol and QM/MM methods.

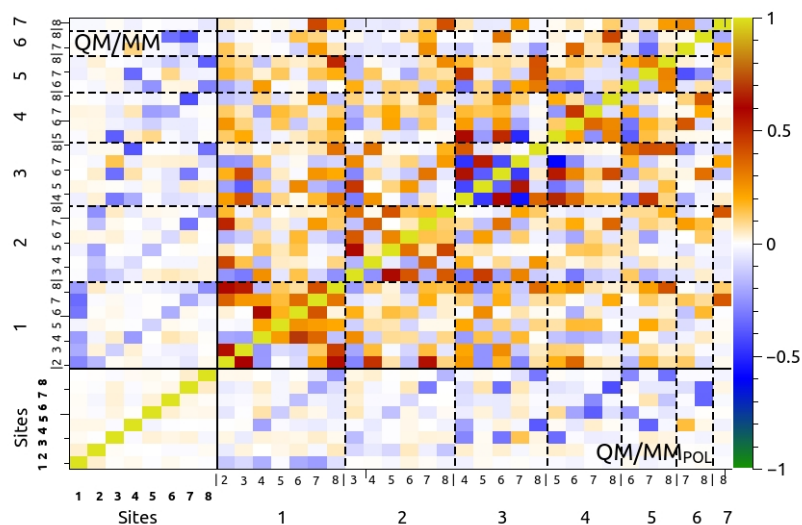


Figure S2: Correlation coefficients between the fluctuation of the site energies and excitonic couplings computed using the quantum chemistry, QM approach, based on results of QM/MM (upper triangle) and QM/MMPol (lower triangle) methods. The chromophores are labeled from 1-8 in the following order: DBV_A, DBV_B, PEB_{50/61C}, PEB_{158C}, PEB_{82C}, PEB_{50/61D}, PEB_{158D}, and PEB_{82D}.