

# Supporting Material

## Intrinsic Coordination for Revealing Local Structural Changes in Protein Folding-Unfolding

Ying-Jen Shiu,<sup>a,b</sup> Michitoshi Hayashi,<sup>c,\*</sup> Orion Shih,<sup>a</sup> Charlene Su,<sup>b</sup> Min-Yeh Tsai,<sup>d</sup> Yi-Qi Yeh,<sup>a</sup> Chun-Jen Su,<sup>a</sup> Yu-Shan Huang,<sup>a</sup> Sheng-Hsien Lin<sup>b,d</sup> and U-Ser Jeng,<sup>a,e,\*</sup>

<sup>a</sup> National Synchrotron Radiation Research Center, Hsinchu 30076, Taiwan.

<sup>b</sup> Institute of Atomic and Molecular Sciences, Academia Sinica, P. O. Box 23-166, Taipei 106, Taiwan.

<sup>c</sup> Center for Condensed Matter Sciences, National Taiwan University, Taipei 106, Taiwan.

<sup>d</sup> Department of Applied Chemistry, National Chiao Tung University, Hsinchu 300, Taiwan.

<sup>e</sup> Department of Chemical Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan.

### S1 Transformation of the matrixes.

The Euler angle rotation matrix R is given by

$$R = \begin{pmatrix} \cos\theta & \cos\varphi\sin\theta & \sin\varphi\sin\theta \\ -\cos\chi\sin\theta & \cos\chi\cos\varphi\cos\theta - \sin\chi\sin\varphi & \cos\chi\cos\theta\sin\varphi + \cos\varphi\sin\chi \\ \sin\chi\sin\theta & -\cos\varphi\cos\theta\sin\chi - \cos\chi\sin\varphi & -\cos\theta\sin\chi\sin\varphi + \cos\chi\cos\varphi \end{pmatrix}. \quad (\text{S-1})$$

This matrix operation consists of the sequence of the three successive rotations around the axes X-n(Z)-x by the angles  $\varphi$ - $\theta$ - $\chi$ , as shown in Figure 1.

The T-matrix is composed of the elements of transition dipole moment; these elements are given by projecting the transition dipole moment and the normal vector onto the xyz frame in protein coordinate system. Thus T-matrix is given by

$$T = \begin{pmatrix} \mu_{ex} & \mu'_{ax} & N_x \\ \mu_{ey} & \mu'_{ay} & N_y \\ \mu_{ez} & \mu'_{az} & N_z \end{pmatrix}. \quad (\text{S-2})$$

To obtain  $\mu'_a$ , the Gram-Schmidt Method was used:  $\mu'_a = \mu_a - (\mu_a \cdot \mu_e) \cdot \hat{e}_e$  where  $\hat{e}_e$  is the unit vector of  $\mu_e$ .

### S2 TRFA analysis of a symmetric rotor.

The decay profile of TRFA of the asymmetric rotor is given by

$$r(t) = \frac{6}{5} \left\{ \sum_{i=1}^3 c_i \exp(-t/\tau_i) + \frac{F+G}{4} \exp[-(6D-2\Delta)t] + \frac{F-G}{4} \exp[-(6D+2\Delta)t] \right\}. \quad (\text{S-3})$$

where  $D = \frac{D_1 + D_2 + D_3}{3}$  and  $\Delta = \sqrt{D_1^2 + D_2^2 + D_3^2 - D_1D_2 - D_1D_3 - D_2D_3}$ , and

$c_i = \alpha_j \alpha_k \varepsilon_j \varepsilon_k$  with  $(i,j,k) = (1,2,3), (2,3,1)$  or  $(3,1,2)$ .

Here,  $\alpha_1, \alpha_2$ , and  $\alpha_3$  correspond to the cosine projections of the absorption transition dipole moment with respect to the principle axis of rotation;  $\varepsilon_1, \varepsilon_2, \varepsilon_3$  are the cosine projections of the emission transition dipole moment in the rotor.

The decay constants  $\tau_i$  are given by

$$\tau_i = \frac{1}{3D + 3D_i}, \quad (\text{S-4})$$

F and G are given by

$$F = \sum_{i=1}^3 \alpha_i^2 \varepsilon_i^2 - \frac{1}{3}, \quad (\text{S-5})$$

and

$$G \cdot \Delta = \sum_{i=1}^3 D_i (\alpha_i^2 \varepsilon_i^2 + \alpha_j^2 \varepsilon_k^2 + \alpha_k^2 \varepsilon_j^2) - D; \quad i \neq j \neq k \neq i, \quad (\text{S-6})$$

respectively.

For a symmetric rotor prolate, the rotational diffusion constant should be  $D_1 = D_{//}$  and  $D_2 = D_3 = D_{\perp}$ . The emission transition dipole moment in the xyz frame projected directions at the principle axis of the rotor is given by

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix} = R^{-1} \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos \theta \\ \sin \theta \\ 0 \end{pmatrix}; \quad (\text{S-7})$$

while the direction cosines of the absorption transition dipole moment projections is given by

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = R^{-1} \cdot R_{\gamma} \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos \theta \cos \gamma + \sin \theta \sin \chi \sin \gamma \\ \sin \theta \cos \gamma - \cos \theta \sin \chi \sin \gamma \\ \cos \chi \sin \gamma \end{pmatrix}, \quad (\text{S-8})$$

where  $R_\gamma = \begin{pmatrix} \sin \gamma & \cos \gamma & 0 \\ -\cos \gamma & \sin \gamma & 0 \\ 0 & 0 & 1 \end{pmatrix}$  by the rotation angle  $90 - \gamma$  degree.

The anisotropic correlation function  $r(t)$  in Eq.(2) is finally reduced to

$$r(t) = \frac{6}{5} \{ (A \cdot B) \exp(-(D_{//} + 5D_{\perp})t) + \left( \frac{2A^2 + B^2 + C^2 - \frac{2}{3}}{4} \right) \exp(-6D_{\perp}t) + \left( \frac{B^2 - C^2}{4} \right) \exp(-(4D_{//} + 2D_{\perp})t) \} \quad (\text{S-9})$$

### S3. TRFA analysis of a spherical rotor.

For  $\vec{\mu}_a = \vec{\mu}_e$  or  $\gamma = 0$ , the anisotropic correlation function of the symmetric rotor becomes

$$r(t) = 2/5 \cdot \{ 3 \cos^2 \theta \sin^2 \theta \exp(-(D_{//} + 5D_{\perp})t) + 1/4 \cdot (3 \cos^2 \theta - 1)^2 \exp(-6D_{\perp}t) + 3/4 \cdot \sin^4 \theta \exp(-(4D_{//} + 2D_{\perp})t) \}, \quad (\text{S-10})$$

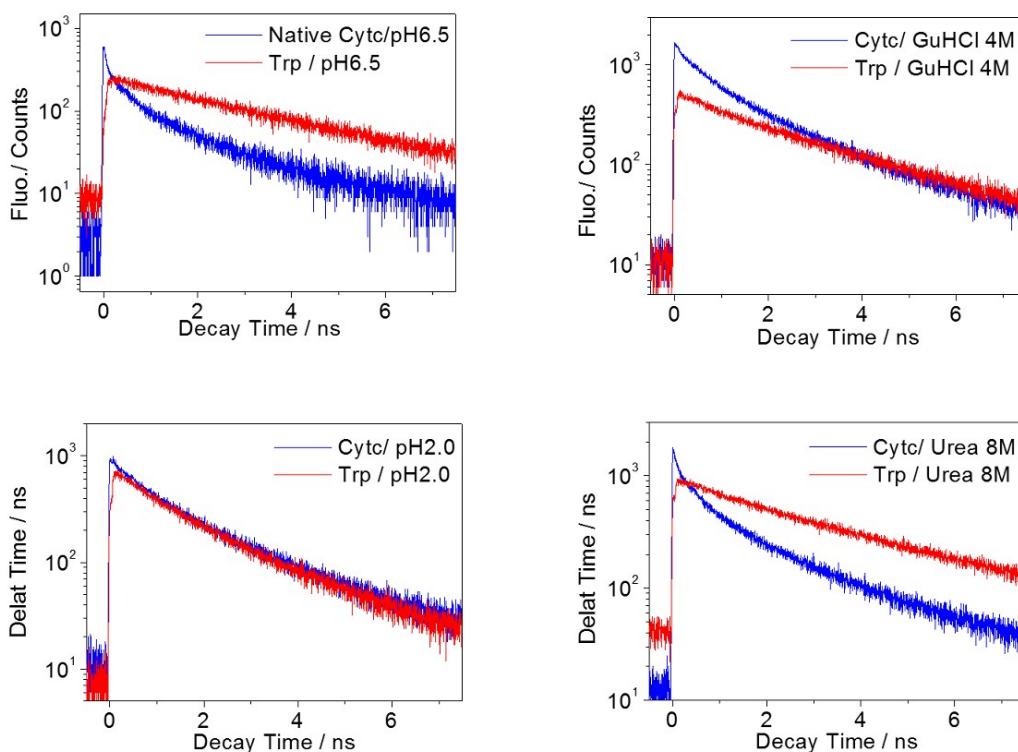
For a sphere rotor  $D_{//} = D_{\perp} = D$  and  $\chi = 0$ , Eq.(S-10) can be further simplified to

$$r(t) = \frac{1}{5} (3 \cos^2 \gamma - 1) \exp(-6Dt). \quad (\text{S-11})$$

The above two equations agree with the known results (1). It is noted that the anisotropic correlation function at zero time ( $t=0$ ) is determined only by the included angle between two transition dipole moments  $\vec{\mu}_a$  and  $\vec{\mu}_e$ . The deduced equation also agrees with the case of  $r(0) = 0.4$  at zero time for the parallel vectors of  $\vec{\mu}_a$  and  $\vec{\mu}_e$ , and  $r(0) = 0$  when the angle is  $\gamma = 54.7^\circ$ .

### S4.

We have measured and compared the time-resolved fluorescence decay profiles for free tryptophans and the Trp59 of the protein in same environments (pH6.5, 4M GuHCl and 8M urea), as that shown in the following figures.



**Fig. S1:** Time-resolved fluorescence decay profiles for free tryptophans and the Trp59 of cyt c in the environments indicated.

The fluorescence decay behaviors of the free tryptophan deviate largely from that of the Trp59 of the native and unfolded protein in the corresponding solutions. The much faster decays of the latter indicate that the Trp59 of the native and the unfolded protein may not be exposed to solvent, urea, or GuHCl in the solution as the free tryptophan does. The results suggest that unfolded cyt c conformation still could hide/enclose the Trp59 reasonably well in the urea- and GuHCl-induced protein unfolding cases, to hinder largely the free rotating of the indole ring for Trp59 of the protein in the unfolding environment. Nevertheless, the decay profile for the Trp59 of the protein in pH2 is not much faster than that for free tryptophans, suggesting that Trp59 might be partially surrounded by acid protons. However our SAXS results show that the cyt c conformation in pH2 is still relatively compact with a relatively smaller aspect ratio 3.3, compared with that (above 6.0) for the urea and GuHCl cases. Therefore, the Trp59 of the unfolded protein in low-pH environment might still be buried inside the relatively compact structure of the denatured cyt c.

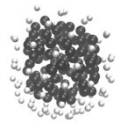
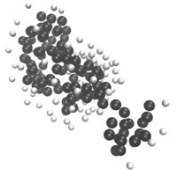
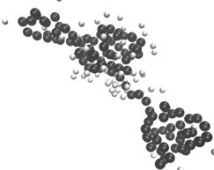
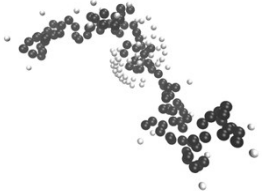
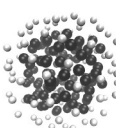
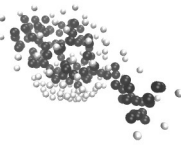
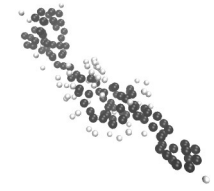

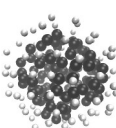
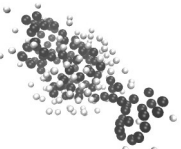


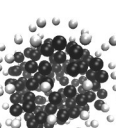
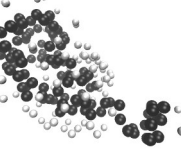
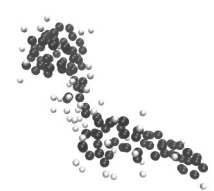

### S5. The ensemble effect to the unfolded protein structures

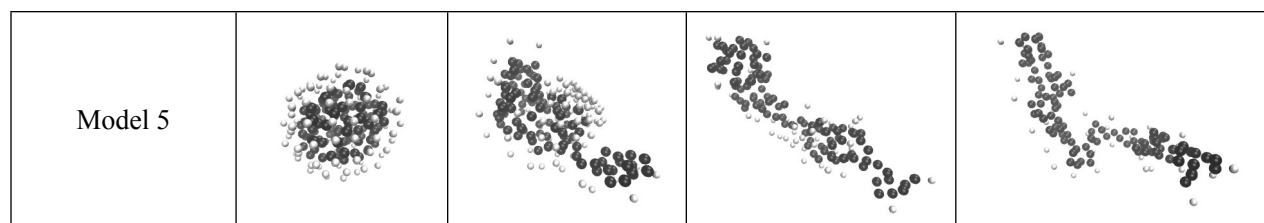
We have fitted the SAXS data using GASBOR model (2), with a chain-like ensemble of dummy residues for ab initio reconstruction of the unfolded protein structure. The ensemble conformations hence fitted suggest a relatively converged shape of an approximated ellipsoid shape. From the ensembles, we have obtained the distribution of the aspect ratios for the unfolded protein, as summarized in the Table S1 below. The fitted ensemble conformations, in

general, can be approximated by an ellipsoidal shape for an aspect ratio needed in the TRFA analysis. Nevertheless, we agree that the ellipsoid shape is not an ideal shape but may be acceptable for the GuHCl case; the associated larger uncertainty in the aspect ratio  $\rho = 6.3 \pm 1.8$  of the case (cf. Table S1), however, does not change significantly the fitting result of TRFA associated with the ellipsoid model fitting result (with  $\rho = 6.6$ ), as that show in Table S2.

Overall, the aspect ratios extracted from the average GABSOR conformations are consistent with that obtained using the analytical ellipsoid model (Table S1,S2) in all the cases. The differences in the mean aspect ratio  $\rho$  are in the range of 2 to 13%, which affect not much (a few to 10 percent) of the fitted parameters of TRFA. The small effect may also be realized from the similar TRFA fitting results for the urea and GuHCl unfolding cases, having a ca. 10% difference (Table S2) in the aspect ratio of the protein (Table S1). With these checks, the proposed SAXS-TRFA approach bears more trusts now to provide useful hints and information for the different unfold protein conformations observed in solution. The selected sensitive TRFA parameters fitted for cyt c, using the values of the aspect ratio  $\rho$  extracted from the GASOR model in Table S3.

**Table S1** The five equally well fitted models (with same probability) for the native and unfolded protein conformations obtained with the SAXS data of the protein at the conditional indicated, using the GASBOR model.

Cyt c	Native	pH 1.9	8M Urea	4M GuHCl
Representative Mode 1				
Model 2				
Model 3				
Model 4				



**Table S2** Aspect ratios of the native and unfolded cyt c obtained from fitting the SAXS data with the chain model of GASBOR (compared with that obtained using the ellipsoidal model).

Aspect ratio	Native	Acid (pH 1.9)	Urea (8 M)	GuHCl (4 M)
$\rho$ (GASBOR)	$1.3 \pm 0.1$	$3.8 \pm 0.6$	$6.1 \pm 1.3$	$6.3 \pm 1.8$
$\rho$ (Ellipsoid)	$1.27 \pm 0.08$	$3.30 \pm 0.04$	$6.0 \pm 0.1$	$6.6 \pm 0.1$

**Table S3.** The selected sensitive TRFA parameters fitted for the native and unfolded cyt c obtained, using the values of the aspect ratio  $\rho$  extracted from the GASOR model fitting with the SAXS data

	Native	Acid (pH 1.9)	Urea (8 M)	GuHCl(4 M)
$\rho$	$1.3 \pm 0.1$	$3.8 \pm 0.6$	$6.1 \pm 1.3$	$6.3 \pm 1.8$
$D_{\perp} / D_{\parallel}$	0.83	0.23	0.112	0.107
$D_{\perp}$ (ns <sup>-1</sup> )	$0.26 \pm 0.01$	$0.49 \pm 0.04$	$0.22 \pm 0.2$	$0.24 \pm 0.02$
$\theta$	$76.8 \pm 42$	$79.8 \pm 5$	$51.3 \pm 1.3$	$54.4 \pm 0.4$

**S6. Table S4.** Anisotropy fitting results with sphere and ellipsoidal models for a free tryptophan in acid pH=1.9.

	Sphere	Ellipsoid
$\gamma$ (degree)	$44.2 \pm 0.08$	$44.6 \pm 0.1$
$D_{\perp} / D_{\parallel}$	1	$0.47 \pm 0.05$
$D_{\perp}$ (ns <sup>-1</sup> )	1.67	0.66
$D_{\parallel}$ (ns <sup>-1</sup> )	$1.67 \pm 0.02$	$1.39 \pm 0.07$

### S7. Quantum chemistry model calculations for transition states of Trp59.

The 0-0 transition energies and the oscillation strengths from the ground state up to the 5-th excited states of the Trp59 in the native cyt c are listed in Table S5. A realistic calculation of the optical properties of the molecule in the protein environment can be carried out by using time-dependent functional theory method combined with QM/MM method where the optically active

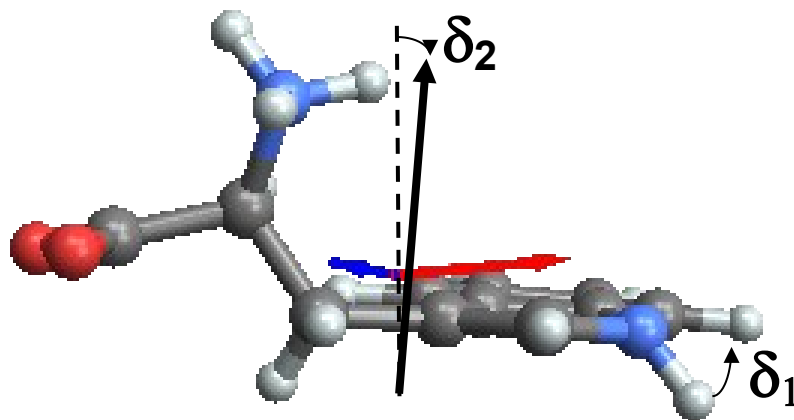
molecule is treated by TD-DFT (quantum mechanics) and the protein environment can be modeled by molecular mechanics (force field). However, such calculations may require an intensive and careful study of the effects of the size dependent of the protein environment on the optical properties (3). To grasp the essence of the transition moments and excitation energies, we adopted, as the first step, a semi-empirical quantum chemistry method, ZINDO/S, using Gaussian03 package (4).

**Table S5.**

Ground to excited state	Oscillator strength	Excitation(nm)
1st	0.0000	320.74
2nd	0.0001	300.50
3rd	0.0136	297.08
4th	0.2695	289.14
5th	0.2548	240.03

**S8. N-H tilt angle to the indole ring of Trp59 of cyt c.**

For distinguishing the two N-H tilt angles, the corresponding transition dipole moments and the elongated orientations in the protein were calculated. The angle  $\delta_1$  in Fig. S2 denotes the tilt angle of N-H bending to the indole ring, and  $\delta_2$  is the deviation angel between the normal vectors from the  $\vec{\mu}_e$  and  $\vec{\mu}_a$  plane (the black arrow) to the indole ring (the dashed line). We examined the aspect ratios along the corresponding elongated orientation in cyt c (code 1HRC). As a result, aspect ratios 1.7 and 1.28 are obtained for N-H tilts  $\delta_1$  within  $+35^\circ \sim +40^\circ$  and  $-44^\circ \sim -47^\circ$ , respectively. Since the latter is consistent with that (1.27) for the native cyt c obtained by SAXS, we therefore adopted the N-H tilt angle  $-44^\circ \sim -47^\circ$  of Trp59 in the related model calculations. The tilt angles of  $\delta_1$  for cyt c in the native and unfolded states are shown in the Table S6, at  $\delta_2 = 5^\circ$



**Figure S2.** N-H tilt angle to the indole ring of Trp59 of cyt c.

**Table S6.**  $\delta_1$  values for cyt c in the native and unfolded states as indicated, at  $\delta_2 = 5^\circ$

Cyt c	Native	pH 1.9	8M Urea	4M GuHCl
$\delta_1$	-45.5°	-47.5°	-44.0°	-47.0°

### S9. Total dipole moment $\overset{\vee}{\nu}_D$ of native cyt c.

The total dipole moment of  $n$  atoms in the native cyt c (code 1HRC) is calculated with  $\overset{\vee}{\nu}_D = \sum_{i=1}^n (q_i - q_0) \overset{\vee}{r}_i$ , where  $q_i$  and  $\overset{\vee}{r}_i$  are the partial charge and the coordinates of the  $i$ -th atom, respectively, and  $q_0 = \sum_{i=1}^n q_i / n$ . Here, the partial charges are given by CHARMM 27 force field (5), and that of the Heme are developed by Luthey-Schulten's group (6). This force field has been successfully examined for the calculations of the folding-unfolding conformations of cyt c (7).

### S10. Table S7. List of the deduced unit vectors based on the cyt c crystalline structure (code 1HRC).

$\overset{\vee}{X}'_I$	$x$	$y$	$z$	
$I=N$ Native	-0.584	0.0267	0.811	
$I=U$	pH 1.9	-0.597	0.0176	0.802
	8M Urea	-0.828	0.154	0.538
	4M GuHCl	-0.770	0.327	0.548
$\overset{\vee}{\mu}_e$	$x$	$y$	$z$	
Native	-0.824	0.460	-0.331	
pH 1.9	-0.827	0.425	-0.368	
8M Urea	-0.820	0.486	-0.302	
4M GuHCl	-0.827	0.434	-0.359	
$\overset{\vee}{\mu}_a$	$x$	$y$	$z$	
Native	-0.563	-0.466	-0.819	
pH 1.9	-0.567	-0.460	-0.817	
8M Urea	-0.560	-0.471	-0.820	
4M GuHCl	-0.566	-0.461	-0.818	
Unit vectors	$x$	$y$	$z$	
$\overset{\vee}{\alpha}_N$	0.489	-0.846	-0.214	
$\overset{\vee}{\alpha}_C$	-0.703	-0.101	-0.703	
$\overset{\vee}{\alpha}_{60}$	0.519	-0.725	0.452	
$\overset{\vee}{\nu}_{Fe-S(Met80)}$	-0.933	-0.332	-0.136	
$\overset{\vee}{\nu}_{Fe-Trp59}$	0.0302	-0.917	0.398	
$\overset{\vee}{\nu}_D$	-0.546	0.676	-0.494	

### S11. Table S8. Relative angles between the emission transition dipole moment $\overset{\vee}{\mu}_e$ and the unit vectors in the native cyt c (code 1HRC) and the unfolded states indicated.

Unit vectors	Native	pH 1.9	8M Urea	4M GuHCl
--------------	--------	--------	---------	----------



$\alpha_N$	136°	133°	138°	134°
$\alpha_C$	40°	37°	42°	38°
$\alpha_{60}$	156°	155°	156°	155°
$\nu_{Fe-S(Met80)}$	49°	47°	50°	47°
$\nu_{Fe-Trp59}$	125°	124°	126°	124°
$\nu_D$	22°	23°	22°	23°

## S12. Molecular dynamics (MD) simulations.

MD simulations with isothermal–isobaric ensemble were performed using Amber 14 package (8) with the ff14SB force field. A charge set and CHARMM parameters for the Heme group based on density functional theory (6) were modified for Amber-14. The initial structure of cyt c was taken from Protein Data Bank entries 1HRC. Periodic boundary conditions along with build-in 8M urea solvent model were used. The system contained 1694 urea molecules, 8253 water and 6 Cl<sup>-</sup> ions. SHAKE constraints were applied to maintain the bond lengths and angles of urea and water. A non-bonded cutoff of 8 Å was used. The integration time step was 2 fs and the coordinates were saved every 2 ps for analysis. After initial energy minimization, the system was heated to 300 K for 50 ps and equilibrated for 150 ps. The simulation was then followed by a 250ns production MD simulation with 1 atm constant pressure using GPU-accelerated Amber code (9).

## Supporting References:

1. Casalini, S., G. Battistuzzi, M. Borsari, C. A. Bortolotti, G. Di Rocco, A. Ranieri, and M. Sola, 2010. Electron transfer properties and hydrogen peroxide electrocatalysis of cytochrome c variants at positions 67 and 80. *J. Phys. Chem. B.* 114: 1698-1706.
2. Svergun, D.I., Petoukhov, M.V. and Koch, M.H.J. (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.*, **80**, 2946-2953.
3. Li, J. H., J. D. Chai, G. Y. Guo, M. Hayashi. 2012. Significant role of the DNA backbone in mediating the transition origin of electronic excitations of B-DNA - implication from long range corrected TDDFT and quantified NTO analysis. *Chem. Phys. Phys. Chem.* 14: 9092-9103.
4. Frisch, M. J., G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Rob, 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. Ayala, P. Y. Ochterski, 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. 2003. Gaussian 03. Gaussian, Inc. Wallingford. CT.
5. MacKerell, A. D. Jr., D. Bashford, M. Bellott, R. L. Jr. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. 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. Wiörkiewicz-Kuczera, D. Yin, and M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102: 3586-3616.
  6. Autenrieth, F., E. Tajkhorshid, J. Baudry, and Z. Luthey-Schulten, 2004. Classical force field parameters for the heme prosthetic group of cytochrome c. *J. Comput. Chem.* 25: 1613-1622.
  7. Tsai, M. Y., A. N. Morozov, K. Y. Chu, and S. H. Lin. 2009. Molecular dynamics insight into the role of tertiary (foldon) interactions on unfolding in cytochrome c. *Chem. Phys. Lett.* 475: 111–115.
  8. Case, D. A., T. E. III Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Jr. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods. 2005. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* 26: 1668-1688.
  9. Salomon-Ferrer, R., A. W. Götz, D. Poole, S. Le Grand, and R. C. Walker. 2013. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. explicit solvent particle mesh ewald. *J. Chem. Theory Comput.* 9: 3878-3888.