

## Visualization of the Encounter Ensemble of the Transient Electron Transfer Complex of Cytochrome c and Cytochrome c Peroxidase

Qamar Bashir,<sup>†</sup> Alexander N. Volkov,<sup>†,‡</sup> G. Matthias Ullmann,<sup>§</sup> and Marcellus Ubbink\*,<sup>†</sup>

Leiden Institute of Chemistry, Leiden University, Gorlaeus Laboratories, P.O. Box 9502, 2300 RA Leiden, The Netherlands, and Structural Biology/Bioinformatics, University of Bayreuth, Universitätsstrasse 30, BGI, 95447 Bayreuth, Germany

Received July 31, 2009; E-mail: m.ubbink@chem.leidenuniv.nl

**Abstract:** Recent studies have provided experimental evidence for the existence of an encounter complex, a transient intermediate in the formation of protein complexes. We use paramagnetic relaxation enhancement NMR spectroscopy in combination with Monte Carlo simulations to characterize and visualize the ensemble of encounter orientations in the short-lived electron transfer complex of yeast cytochrome c (*Cc*) and cytochrome c peroxidase (CcP). The complete conformational space sampled by the protein molecules during the dynamic part of the interaction was mapped experimentally. The encounter complex was described by an electrostatic ensemble of orientations based on Monte Carlo calculations, considering the protein structures in atomic detail. We demonstrate that this visualization of the encounter complex, in combination with the specific complex, is in excellent agreement with the experimental data. Our results indicate that *Cc* samples only about 15% of the surface area of CcP, surrounding the specific binding interface. The encounter complex is populated for 30% of the time, representing a mere 0.5 kcal/mol difference in the free energies between the two states. This delicate balance is interpreted to be a consequence of the conflicting requirements of fast electron transfer and high turnover of the complex.

### Introduction

Electron transfer (ET) protein complexes often require high turnover and thus fast dissociation. A high affinity confers specificity on the interaction but limits the dissociation rate. Therefore, these complexes are on the border of specific and nonspecific as a result of a compromise between tight binding and fast dissociation. Early theoretical work suggested that ET complexes are highly dynamic.<sup>1</sup> NMR and kinetic studies, recently reviewed in ref 2, provided evidence for a model of the process of protein complex formation in which proteins approach each other by diffusion and initially associate in an encounter complex, followed by the formation of the specific complex. In the encounter complex, the proteins do not assume a single orientation relative to each other, but rather they rapidly change orientation, thus sampling the surface of the partner. It is thought that the encounter complex accelerates the formation of the specific complex by reduction of the dimensionality of the conformational search. In the encounter complex intermolecular interactions are dominated by electrostatic forces. Long-range charge–charge attraction prolongs the lifetime of the encounter and allows for preorientation of protein molecules, thus limiting the conformational search to a part of the binding surface.<sup>2</sup> In the specific complex the proteins assume a single,

well-defined orientation that is stabilized not only by electrostatic forces but also by short-range interactions like hydrogen bonds and the hydrophobic effect. Some ET complexes appear to be entirely or mostly nonspecific, with the encounter complex being the dominant form,<sup>3–10</sup> whereas in others the specific complex dominates.<sup>11–13</sup>

Paramagnetic relaxation enhancement (PRE) NMR spectroscopy has proven to be a useful technique for detecting the encounter complex in protein–protein<sup>13,14</sup> and protein–DNA<sup>15,16</sup> complexes as well as macromolecular self-association<sup>17,18</sup> allostery<sup>19</sup> and state equilibria.<sup>20</sup> Generally, to observe PRE, a

- (3) Xu, X.; Reinle, W.; Hannemann, F.; Konarev, P. V.; Svergun, D. I.; Bernhardt, R.; Ubbink, M. *J. Am. Chem. Soc.* **2008**, *130*, 6395–6403.
- (4) Liang, Z.-X.; Nocek, J. M.; Huang, K.; Hayes, R. T.; Kurnikov, I. V.; Beratan, D. N.; Hoffman, B. M. *J. Am. Chem. Soc.* **2002**, *124*, 6849–6859.
- (5) Volkov, A. N.; Ferrari, D.; Worrall, J. A. R.; Bonvin, A. M. J. J.; Ubbink, M. *Protein Sci.* **2005**, *14*, 799–811.
- (6) Liang, Z.-X.; Kurnikov, I. V.; Nocek, J. M.; Mauk, A. G.; Beratan, D. N.; Hoffman, B. M. *J. Am. Chem. Soc.* **2004**, *126*, 2785–2798.
- (7) Worrall, J. A. R.; Reinle, W.; Bernhardt, R.; Ubbink, M. *Biochemistry* **2003**, *42*, 7068–7076.
- (8) Worrall, J. A. R.; Liu, Y.; Crowley, P. B.; Nocek, J. M.; Hoffman, B. M.; Ubbink, M. *Biochemistry* **2002**, *41*, 11721–11730.
- (9) Hoffman, B. M.; Celis, L. M.; Cull, D. A.; Patel, A. D.; Seifert, J. L.; Wheeler, K. E.; Wang, J.; Yao, J.; Kurnikov, I. V.; Nocek, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 3564–3569.
- (10) Ubbink, M.; Bendall, D. S. *Biochemistry* **1997**, *36*, 6326–6335.
- (11) Vlasie, M. D.; Fernández-Busnadiego, R.; Prudêncio, M.; Ubbink, M. *J. Mol. Biol.* **2008**, *375*, 1405–1415.
- (12) Ubbink, M.; Ejdeback, M.; Karlsson, B. G.; Bendall, D. S. *Structure* **1998**, *6*, 323–335.
- (13) Volkov, A. N.; Worrall, J. A. R.; Holtzmann, E.; Ubbink, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18945–18950.

\* Leiden University.

† Present address: Structural Biology Brussels, Vrije Universiteit Brussel, and Department of Molecular and Cellular Interactions, VIB, Pleinlaan 2, 1050 Brussels, Belgium.

‡ University of Bayreuth.

- (1) Northrup, S.; Boles, J.; Reynolds, J. *Science* **1988**, *241*, 67–70.
- (2) Ubbink, M. *FEBS Lett.* **2009**, *583*, 1060–1066.

paramagnetic spin label (SL) is attached to the protein surface via a cysteine residue. The unpaired electron on the SL increases the relaxation rate of the nuclei in its proximity due to the magnetic dipolar interactions. The effect depends on the sixth power of the distance between the nucleus and the paramagnetic center, averaged over all positions of the nucleus and the SL (eqs 4 and 5, Materials and Methods). The effect is very strong at short distances, and consequently, even lowly populated states in which the nucleus is close to the SL can be detected. Although PRE can be used to map the surface area sampled by a partner protein in the encounter complex, visualization of the orientations of the proteins is impossible solely on the basis of experimental data because the PRE represent an average over time and space of all orientations. Several approaches have been proposed so far to visualize the encounter complex by combining modeling and PRE data, including explicit ensemble refinement,<sup>14</sup> Brownian Dynamics simulations,<sup>21</sup> and empirical ensemble simulations.<sup>3,22</sup>

The complex between yeast mitochondrial iso-1-cytochrome *c* (*Cc*) and cytochrome *c* peroxidase (CcP) represents a well-studied example of ET complexes with a short lifetime. CcP protects the cell against high concentrations of peroxides, receiving electrons from *Cc*. Here, we report the comprehensive mapping of the *Cc*–CcP encounter complex and its characterization by PRE NMR and Monte Carlo (MC) simulations. We demonstrate that a combination of the PREs calculated for the simulated ensemble and the specific complex is in excellent agreement with the experimental data. Using this approach we determine the fraction of the complex in the encounter complex to be 30%. The functional reason for the small associated free energy difference between encounter and specific complexes is discussed.

## Materials and Methods

**Protein Preparation.** Single-cysteine CcP mutants were constructed by site-directed mutagenesis using the QuikChange polymerase chain reaction protocol and CcP (MKT, C128A) plasmid<sup>13,23</sup> as a template. The proteins were expressed and purified as described before.<sup>23–26</sup> During the purification procedure of CcP mutants on a gel filtration column, dithiothreitol (1 mM) was present in the equilibration buffer as well as in the protein sample. Concentrations of ferric *Cc* and five-coordinated high-spin ferric CcP were determined according to the optical absorbance peaks.<sup>27,28</sup> The

- (14) Tang, C.; Iwahara, J.; Clore, G. M. *Nature* **2006**, *444*, 383–386.
- (15) Iwahara, J.; Clore, G. M. *Nature* **2006**, *440*, 1227–1230.
- (16) Clore, G. M. *Mol. BioSyst.* **2008**, *4*, 1058–1069.
- (17) Tang, C.; Ghirlando, R.; Clore, G. M. *J. Am. Chem. Soc.* **2008**, *130*, 4048–4056.
- (18) Tang, C.; Louis, J. M.; Aniana, A.; Suh, J.-Y.; Clore, G. M. *Nature* **2008**, *455*, 693–696.
- (19) Tang, C.; Schwieters, C. D.; Clore, G. M. *Nature* **2007**, *449*, 1078–1082.
- (20) Henzler-Wildman, K. A.; Thai, V.; Lei, M.; Ott, M.; Wolf-Watz, M.; Fenn, T.; Pozharski, E.; Wilson, M. A.; Petsko, G. A.; Karplus, M.; Hubner, C. G.; Kern, D. *Nature* **2007**, *450*, 838–844.
- (21) Kim, Y. C.; Tang, C.; Clore, G. M.; Hummer, G. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 12855–12860.
- (22) Hulsken, R.; Baranova, M. V.; Bullerjahn, G. S.; Ubbink, M. *J. Am. Chem. Soc.* **2008**, *130*, 1985–1991.
- (23) Goodin, D. B.; Davidson, M. G.; Roe, J. A.; Mauk, A. G.; Smith, M. *Biochemistry* **1991**, *30*, 4953–4962.
- (24) Morar, A. S.; Kakouras, D.; Young, G. B.; Boyd, J.; Pielak, G. J. *J. Biol. Inorg. Chem.* **1999**, *4*, 220–222.
- (25) Pollock, W. B. R.; Rosell, F. I.; Twitchett, M. B.; Dumont, M. E.; Mauk, A. G. *Biochemistry* **1998**, *37*, 6124–6131.
- (26) Teske, J. G.; Savenkova, M. I.; Mauro, J. M.; Erman, J. E.; Satterlee, J. D. *Protein Expression Purif.* **2000**, *19*, 139–147.
- (27) Margoliash, E.; Frohwirt, N. *Biochem. J.* **1959**, *71*, 570–572.

yields were, respectively, 7–8 mg and 15–25 mg/L of culture of pure *Cc* (from minimal medium) and CcP mutants (from rich medium).

Conjugation with a paramagnetic label, MTS [(1-oxyl-2,2,5,5-tetramethyl-3-pyrroline-3-methyl)methanethiosulfonate], or a diamagnetic analogue, MTS [(1-acetyl-2,2,5,5-tetramethyl-3-pyrroline-3-methyl)methanethiosulfonate], both purchased from Toronto Research Chemicals (Toronto, Canada), was carried out as published.<sup>13</sup> The yield of labeling, determined by EPR or a dithiodipyridine assay,<sup>29</sup> was more than 90%.

**NMR Experiments.** NMR samples contained 0.3 mM of a 1:1 <sup>15</sup>N *Cc*–CcP-MTS(L) complex in 20 mM sodium phosphate, 0.1 M NaCl, 6% D<sub>2</sub>O for lock, and 0.1 mM CH<sub>3</sub>CO<sup>15</sup>NH<sub>2</sub> as internal reference. The pH of the samples was adjusted to 6.00 ± 0.05 with small aliquots of 0.1 M HCl or 0.1 M NaOH. Measurements were performed at 303 K on a Bruker DMX600 spectrometer equipped with a triple-resonance TXI-Z-GRAD probe (Bruker, Karlsruhe, Germany). 2D [<sup>15</sup>N, <sup>1</sup>H] HSQC spectra were obtained with 512 and 128 complex points in the direct and indirect dimensions, respectively, and with spectral widths of 32 (<sup>15</sup>N) and 16 ppm (<sup>1</sup>H). All data were processed with the Azara suite of programs (provided by Wayne Boucher and the Department of Biochemistry, University of Cambridge, U.K.) and analyzed in ANSIG for Windows.<sup>30,31</sup> Assignments of the <sup>15</sup>N and <sup>1</sup>H nuclei of *Cc* were taken from previous work.<sup>32</sup>

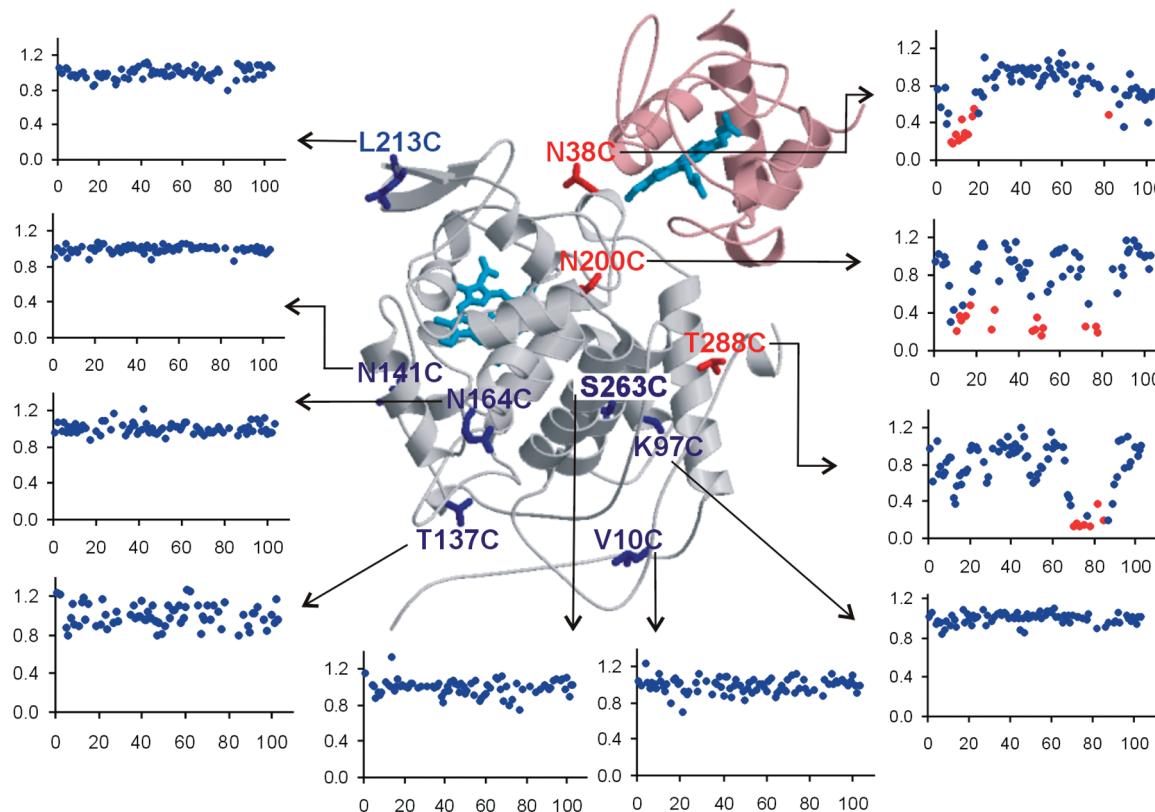
**Calculation of the PRE.** For each observed amide proton of *Cc* the intensity ratio (*I*<sub>para</sub>/*I*<sub>dia</sub>) of the resonances in the paramagnetic (CcP-MTSL) and diamagnetic (CcP-MTS) samples was determined and  $\Gamma_2$  calculated from eq 1<sup>33</sup>

$$\frac{I_{\text{para}}}{I_{\text{dia}}} = \frac{R_2 \exp(-t\Gamma_2)}{R_2 + \Gamma_2} \quad (1)$$

where *R*<sub>2</sub> is the transverse relaxation rate of *Cc* amide protons in the complex with CcP-MTS,  $\Gamma_2$  is the PRE, and *t* is the total INEPT evolution time of the HSQC. For the amides of which the resonances disappear in the paramagnetic spectrum, an upper limit for *I*<sub>para</sub> was estimated from the standard deviation of the noise level of the spectrum. For each peak, *R*<sub>2</sub> was estimated from the width at half-height ( $\Delta\nu_{1/2}$ ) of a Lorentzian fit in the proton dimension by using  $R_2 = \pi\Delta\nu_{1/2}$ . PRE values were calculated after normalization of the *I*<sub>para</sub>/*I*<sub>dia</sub> ratios. The residues with *I*<sub>para</sub>/*I*<sub>dia</sub> > 0.85 were used for this correction. The upper and lower 10% of *I*<sub>para</sub>/*I*<sub>dia</sub> values were removed, and the average of the remaining values was obtained. All *I*<sub>para</sub>/*I*<sub>dia</sub> values were divided by this average to get the normalized PREs.

**Generation of the Encounter Complex Ensemble.** A Boltzmann ensemble of encounter complex geometries was generated by a Metropolis Monte Carlo method using a previously described approach<sup>34</sup> with small modifications. *Cc* was moving in the electrostatic potential of CcP, and moves were accepted or rejected according to the Metropolis criterion. An exclusion grid was used to avoid steric overlap between the proteins.<sup>1</sup> Protein coordinates were taken from the PDB entry 2PCC.<sup>35</sup> The dielectric constants used for the protein and water are 4 and 80, respectively. The electrostatic potential was calculated by the MEAD program suite<sup>36</sup>

- (28) Vitello, L. B.; Huang, M.; Erman, J. E. *Biochemistry* **1990**, *29*, 4283–4288.
- (29) Riener, C.; Kada, G.; Gruber, H. *Anal. Bioanal. Chem.* **2002**, *373*, 266–276.
- (30) Kraulis, P. *J. Magn. Reson.* **1989**, *84*.
- (31) Helgstrand, M.; Kraulis, P.; Allard, P.; Härd, T. *J. Biomol. NMR* **2000**, *18*, 329–336.
- (32) Worrall, J. A. R.; Kolczak, U.; Canters, G. W.; Ubbink, M. *Biochemistry* **2001**, *40*, 7069–7076.
- (33) Battiste, J. L.; Wagner, G. *Biochemistry* **2000**, *39*, 5355–5365.
- (34) Ullmann, G. M.; Knapp, E.-W.; Kostic, N. M. *J. Am. Chem. Soc.* **1997**, *119*, 42–52.
- (35) Pelletier, H.; Kraut, J. *Science* **1992**, *258*, 1748–1755.



**Figure 1.** CcP spin labeling sites and intermolecular PRE effects on Cc. The ribbon representation of the specific Cc–CcP complex was drawn from PDB entry 2PCC<sup>35</sup> with PyMOL.<sup>31</sup> Cc and CcP are in pink and gray, with the heme groups shown as cyan sticks. CcP residues that were replaced by cysteines for SL attachment are labeled and shown as red (PRE observed) and blue (no PRE) sticks. For each SL position, the  $I_{\text{para}}/I_{\text{dia}}$  plots for the Cc residues are shown. The red circles in the graphs represent the residues for which the resonances disappeared in the spectrum of the paramagnetic sample. The enlarged data graphs with error bars are provided as Supporting Information, Figure S-3.

using atomic partial charges of the Charmm force field.<sup>37</sup> For histidine residues that are not coordinated to the heme iron, average Charmm charges were used, which were calculated as  $0.25 \times (\text{His\_N}^e + \text{His\_N}^{\delta}) + 0.5 \times \text{protonated His}$ . The temperature was set to 303 K and the ionic strength to 0.12 M to match the experimental conditions. When the centers of mass of the two molecules had a distance of less than 40 Å, the configuration was saved every 1000 steps. About 1600 configurations were retained for the analysis. This approach implies that configurations with a larger distance contribute negligibly to the free energy change of protein complex formation. These configurations will also be invisible in the PRE analysis.

**Ensemble Analysis.** The ensemble was analyzed by comparison with the Cc–CcP crystal structure by superposition of either of the proteins. The spherical coordinates  $\theta$  (elevation) and  $\varphi$  (azimuth) position the center of mass of a structure in the ensemble relative to the vector between the centers of mass of Cc and CcP in the crystal structure, Figure 3A. For the Cc structures around CcP, the range of  $\theta$  is small, with 95% of the angles less than 45° and 36% within 10° (Figure 3B, gray areas), whereas  $\varphi$  is evenly spread (Figure 3C). The inertial ellipsoid of CcP can be described with values 12.2, 9.3, and 9.6  $\mu\text{s}^{-1}$  for Dz, Dy, and Dx, respectively, with a rotational correlation time of 16 ns.<sup>38</sup> If CcP is approximated as a sphere, the surface area explored by Cc can be calculated. A surface element A of a sphere with radius r is given by eq 2

$$A = \int_0^\Theta \int_0^\Xi r^2 \sin(\theta) d\theta d\phi \quad (2)$$

for  $\Theta = 45^\circ$  and  $\Phi = 360^\circ$ ,  $A = 1.84r^2$ , or 14.6% of the total surface ( $12.57r^2$ ). For  $\Theta = 10^\circ$  A is 0.76% of the total surface. The fact that the lowest  $\theta$  bins are most populated implies that in the ensemble Cc is located close to the orientation in the specific complex.

For CcP structures around Cc, the bin of  $\theta = 35^\circ$  is the most frequent, indicating that the ensemble orientation is not optimal for forming the specific complex and a rotation of CcP around Cc is required. With 95% of the  $\theta$  angles between 0° and 90° (Figure 3D) and 90% of the  $\varphi$  angles between -85° and 95° (Figure 3E), it can be calculated with eq 2 that 85% of the CcP structures samples 25% of the surface area of Cc, again approximating the protein by a sphere (Dz, Dy, and Dx for Cc are 33, 26, and 28  $\mu\text{s}^{-1}$ , respectively, and the correlation time is 5.7 ns).<sup>38</sup>

The DRMS metric was defined according to eq 3 (ref 21)

$$\text{DRMS} = \frac{1}{N} \sum_{i,j} |d_{ij}^{\text{ens}} - d_{ij}^{\text{x-ray}}| \quad (3)$$

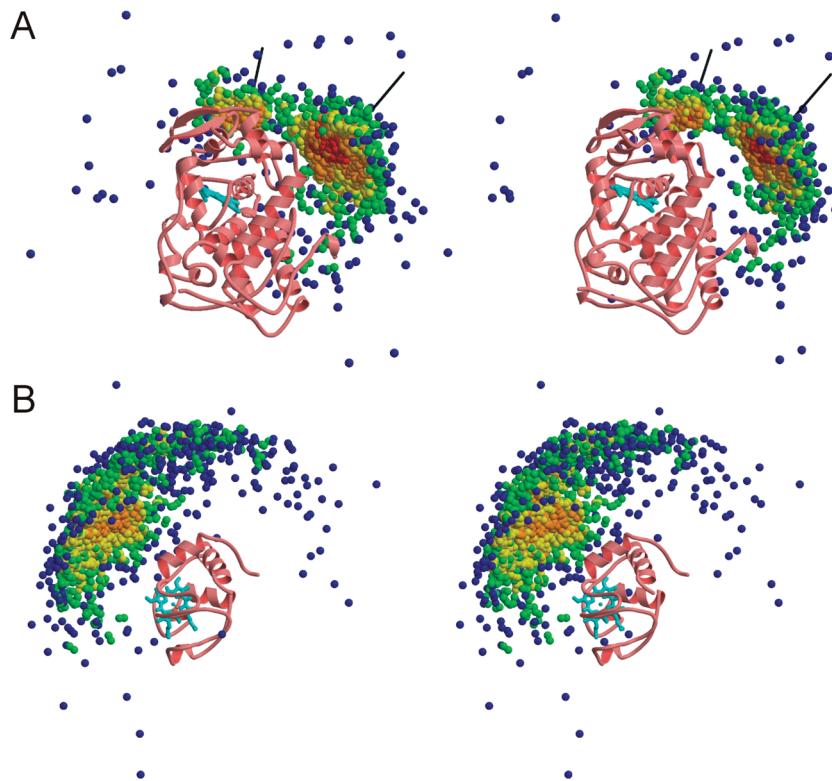
where  $d_{ij}$  is the distance between the  $C^\alpha$  atoms of two residues  $i$  and  $j$  from different proteins.  $N$  is the total number of pairs  $(i, j)$ , and  $d_{ij}^{\text{x-ray}}$  and  $d_{ij}^{\text{ens}}$  are the distance matrix elements from the specific and ensemble structures, respectively.

The distances between each Cc backbone amide hydrogen in the ensemble and the oxygen of SL on CcP were measured with XPLOR-NIH<sup>39</sup> and averaged using eq 4

$$r = \left( \frac{1}{mn} \sum_{j=1}^n \sum_{i=1}^m r_{ij}^{-6} \right)^{-1/6} \quad (4)$$

- (36) Bashford, D. In *Scientific Computing in Object-Oriented Parallel Environments*; Springer: Berlin, 1997; pp 233–240.  
 (37) MacKerell, A. D.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.  
 (38) García de la Torre, J.; Huertas, M. L.; Carrasco, B. *J. Magn. Reson.* **2000**, *147*, 138–146.

- (39) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Marius Clore, G. *J. Magn. Reson.* **2003**, *160*, 65–73.



**Figure 2.** Simulated encounter complex of the *Cc*–*CcP* complex. Stereorepresentations of the ensemble structures with *CcP* (A) and *Cc* (B) superimposed are shown in ribbons with the hemes in cyan. The centers of mass of *Cc* (A) and *CcP* (B) are shown as spheres, colored to indicate the density of the distributions, decreasing from red to blue. The highest densities denote the most favorable electrostatic orientations. Densities were determined by counting the number of neighbors within 2 Å. The lines in A indicate the two centers of high density.

where  $r$  is the effective distance for a given proton,  $m$  is the number of SL orientations used, and  $n$  is the number of structures in the ensemble. To account for the mobility of the SL, the distances to four conformations of the SL were averaged.<sup>40</sup> The use of an effective distance is justified only if the rate of interconversion ( $k_{\text{ex}}$ ) of the substrates is much higher than the PRE ( $k_{\text{ex}} \gg \Gamma_2$ ). From the chemical shift changes observed upon complex formation it can be derived that  $k_{\text{ex}}(\text{binding}) \gg 1000 \text{ s}^{-1}$ . The interconversion of the substrates in the encounter complex and between the encounter and specific complexes is necessarily much faster. Thus, it was assumed that the fast exchange regime can be applied in this case. If the distance between the oxygen of the SL and any *Cc* C $\alpha$  atom was less than 5 Å, then that SL orientation was not used for that *Cc* structure because it was assumed that steric collision between the protein and the SL would not allow that SL orientation. The effective distances were converted into  $\Gamma_2$  with eq 5

$$\Gamma_2 = \frac{\gamma^2 g^2 \beta^2}{20r^6} \left( 4\tau_c + \frac{3\tau_c}{1 + \omega_h^2 \tau_c^2} \right) \quad (5)$$

where  $\tau_c$  is the correlation time of the electron–nucleus vector (16 ns),<sup>13</sup>  $\omega_h$  and  $\gamma$  are proton Larmor frequency and gyromagnetic ratio, respectively,  $g$  is the electronic  $g$  factor, and  $\beta$  is the Bohr magneton. Given the binding constant  $1.9 \pm 0.3 \times 10^5 \text{ M}^{-1}$ <sup>13</sup> under the experimental conditions 88% of *Cc* is bound to *CcP*, so the calculated  $\Gamma_2$  was multiplied by 0.88 to be comparable with the measured values.

The  $Q$  factor was calculated from eq 6

$$Q = \sqrt{\sum_i (\Gamma_{2,i}^{\text{obs}} - \Gamma_{2,i}^{\text{calcd}})^2 / \sum_i (\Gamma_{2,i}^{\text{obs}} + \Gamma_{2,i}^{\text{calcd}})^2} \quad (6)$$

(40) Iwahara, J.; Schwieters, C. D.; Clore, G. M. *J. Am. Chem. Soc.* **2004**, 126, 5879–5896.

where  $\Gamma_{2,i}^{\text{obs}}$  and  $\Gamma_{2,i}^{\text{calcd}}$  are the observed and calculated PREs. The sum of  $\Gamma_{2,i}^{\text{obs}}$  and  $\Gamma_{2,i}^{\text{calcd}}$  was employed in the denominator, instead of the commonly used  $\Gamma_{2,i}^{\text{obs}}$  only, which leads to unjustified larger contributions to the  $Q$  factor by residues with small  $\Gamma_{2,i}^{\text{obs}}$  values. The uncertainties in the  $Q$  factors were obtained by standard propagation of the error in  $\Gamma_{2,i}^{\text{obs}}$ , eq 7. The latter was derived from the uncertainties in  $I_{\text{para}}$  and  $I_{\text{dia}}$ , based on the noise levels in the HSQC spectra.

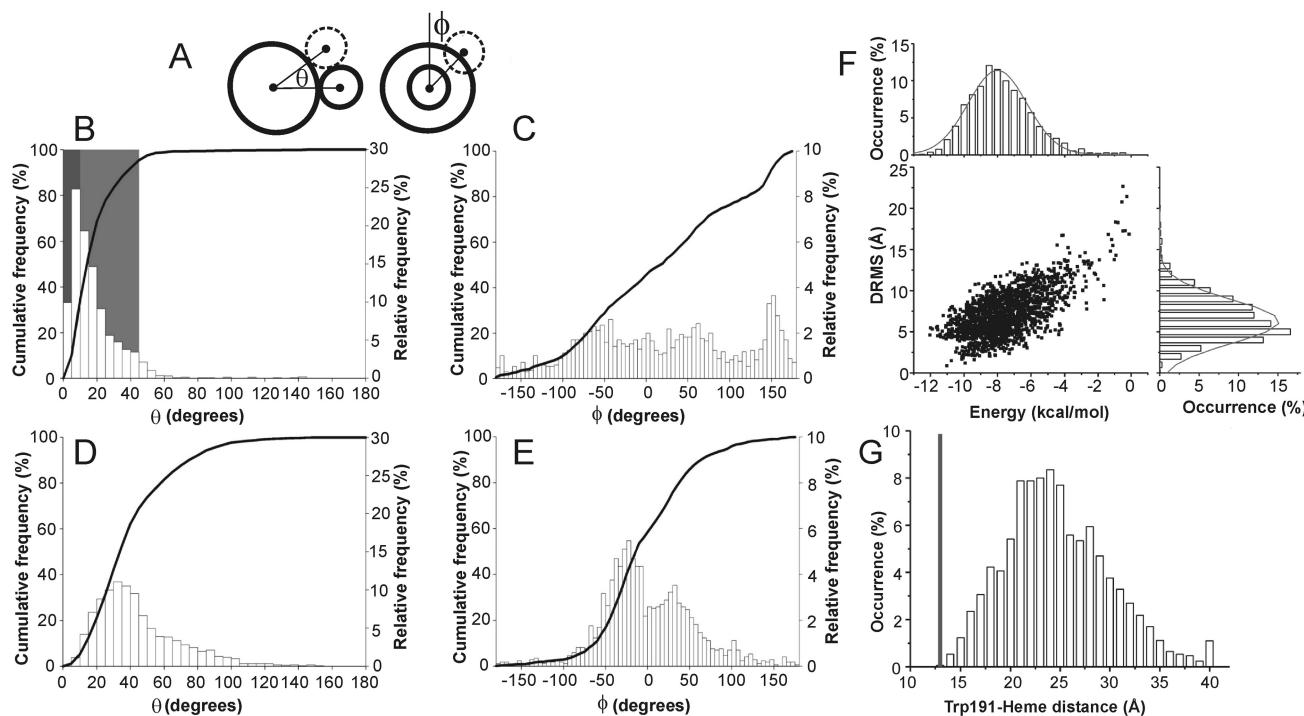
$$\delta Q = Q \sqrt{\frac{\sum_i (\Gamma_{2,i}^{\text{obs}} - \Gamma_{2,i}^{\text{calcd}})^2 \delta^2(\Gamma_{2,i}^{\text{obs}})}{\left[ \sum_i (\Gamma_{2,i}^{\text{obs}} - \Gamma_{2,i}^{\text{calcd}})^2 \right]^2} + \frac{\sum_i (\Gamma_{2,i}^{\text{obs}} + \Gamma_{2,i}^{\text{calcd}})^2 \delta^2(\Gamma_{2,i}^{\text{obs}})}{\left[ \sum_i (\Gamma_{2,i}^{\text{obs}} + \Gamma_{2,i}^{\text{calcd}})^2 \right]^2}} \quad (7)$$

The optimal fit of the experimental PRE data was found by calculating the  $Q$  factor for each value of  $p$  between 0 and 1, with increments of 0.1, where  $p$  and  $(1-p)$  are the population fractions of the encounter and specific complexes, respectively, according to eq 8

$$\Gamma_{2,i}^{\text{calcd}} = p \Gamma_{2,i}^{\text{calcd,ens}} + (1-p) \Gamma_{2,i}^{\text{calcd,specific}} \quad (8)$$

## Results

**Conformational Space of the Encounter Complex.** Nitroxide SLs were introduced, one at a time, at 10 positions, extending the initial set<sup>13</sup> to sample *Cc* interactions with the entire surface of *CcP* (Figure 1), and PREs measured for the *Cc* backbone amide protons (enlarged plots with error bars are provided as Supporting Information). Only three SLs attached to *CcP* near the crystallographic binding site (N38C, N200C and T288C, shown as red sticks in Figure 1), caused paramagnetic relaxation



**Figure 3.** Ensemble analysis. (A) Definition of the spherical coordinates. The proteins in the specific complex are in bold circles, and one of the ensemble structures is shown with a dashed line. The black dots represent the centers of mass. The right-hand picture has been turned by 90° relative to the left one, with the vector of the centers of mass of the specific complex pointing toward the reader. The angle  $\varphi$  is defined relative to an arbitrary direction. (B–E) The relative (bar plots) and cumulative frequencies (solid lines) are plotted for the angles  $\theta$  (B, D) and  $\varphi$  (C, E) in 5° bins for the centers of mass of *Cc* around *CcP* (B, C) and *CcP* around *Cc* (D, E) for a representative ensemble. The dark and light gray areas in B indicate the 0–10° and 10–45° regions (see text). (F) DRMS (distance root mean square, eq 3) is plotted against the electrostatic energy for all structures in the ensemble. The histograms show the energy and DRMS distributions with the curves representing Gaussian fits with averages (widths) of  $-8.0$  ( $3.4$ ) kcal/mol and  $6.3$  ( $5.3$ ) Å for the energies and DRMS, respectively. The interaction energy of the specific complex is  $-25.3$  kcal/mol. (G) Distribution of the edge-to-edge distance between the *Cc* Trp191 indole and the *Cc* heme in the ensemble. The solid bar marks the distance in the specific complex. The category at 40 Å contains all values of  $\geq 40$  Å.

of *Cc* amide nuclei. Attachment of a SL to any of the seven other positions (blue sticks in Figure 1) did not result in PREs. With the combined data set from 10 SL positions, the surface area of *CcP* that is sampled by *Cc* in the encounter complex can now be assessed accurately, see Figure S-1 in the Supporting Information. The results demonstrate that the conformational search of *Cc* in the encounter complex is limited to the area surrounding the interface of the specific complex.

**Monte Carlo Simulations of the Encounter Complex.** According to theoretical<sup>1,41,42</sup> and experimental<sup>43–45</sup> studies, an encounter complex exhibits very few of the short-range interactions characteristic of a specific complex and is predominantly driven by electrostatic forces. Therefore, in order to simulate the encounter complex alone and not the entire protein complex, we used rigid-body Monte Carlo simulations in which only electrostatic and steric interactions were active. It should be noted that our approach is different from that of Kim and co-workers,<sup>21</sup> who simulated the entire protein complex (i.e., combination of the specific and the encounter complexes). Also, in contrast to the coarse-grained model of Kim et al.,<sup>21</sup> our simulations employ atomic-detail descriptions of protein molecules with assigned partial charges.

Figure 2 shows the results of a typical MC run. The centers of mass of *Cc* (Figure 2A) and *CcP* (Figure 2B) are shown as spheres around the partner protein. In the Supporting Information, Figure S-2 shows both distributions superimposed on the specific complex. Also, the coordinates are provided in PDB format. The calculations yield a Boltzmann distribution of orientations with the highest densities (red) representing the lowest energies. In the MC ensemble, *Cc* samples the *CcP* surface area close to the interface of the specific complex, which is consistent with the experimental PRE mapping (see above). This restricted sampling can be attributed to the charge distribution on *CcP* that results in preorientation of the proteins during the encounter. The distribution of *Cc* around *CcP* (Figure 2A) shows two energy minima, matching those found in Brownian dynamics simulations of Northrup et al.<sup>1</sup> Similar calculations by Gabdoulline and Wade<sup>42</sup> also appear to exhibit these two minima. In our simulations, we did not see a third minimum around *CcP* residue Asp148 as described by Northrup et al.<sup>1</sup> If *CcP* is approximated as a sphere, it can be calculated that 95% of the *Cc* structures samples a mere 15% of the surface area (see Materials and Methods and Figure 3B, shaded region). One-third of all the structures is in a surface area of less than 1% (Figure 3B, dark region). Furthermore, the electrostatic preorientation of *Cc* in the encounter complex brings it very close to the position in the specific complex, as can be seen from the histogram in Figure 3B of the distribution of  $\theta$  angles ( $\theta = 0$  corresponds to the specific complex). Analysis of the distribution of *CcP* around *Cc* (Figure 2B) shows that the former interacts with the positive patch close to the exposed heme edge

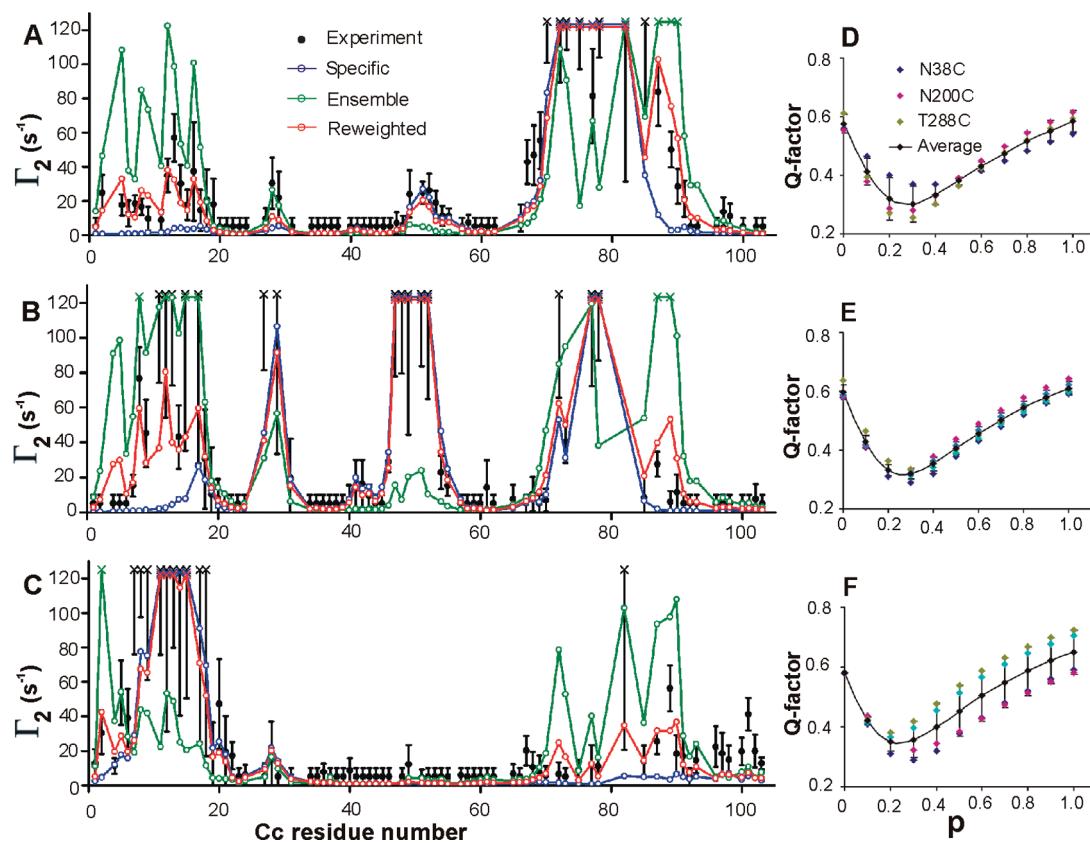
(41) Camacho, C. J.; Weng, Z.; Vajda, S.; DeLisi, C. *Biophys. J.* **1999**, *76*, 1166–1178.

(42) Gabdoulline, R. R.; Wade, R. C. *J. Mol. Biol.* **2001**, *306*, 1139–1155.

(43) Schreiber, G.; Fersht, A. R. *Nat. Struct. Mol. Biol.* **1996**, *3*, 427–431.

(44) Miyashita, O.; Onuchic, J. N.; Okamura, M. Y. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16174–16179.

(45) Suh, J.-Y.; Tang, C.; Clore, G. M. *J. Am. Chem. Soc.* **2007**, *129*, 12954–12955.



**Figure 4.** PRE analysis. (A–C) Observed (solid circles) and calculated (open symbols) PREs for *Cc* residues in the complex with T288C (A), N200C (B), and N38C (C) CcP-SL. The experimental PREs are in black, the PREs back-calculated from the specific complex in blue, the PREs of the simulated encounter complex in green, and the combination of latter two at 30% population fraction of the ensemble in red. (D) *Q* factors calculated for combinations of PREs from the specific structure and the simulated ensemble at different population fractions of the encounter complex ( $p$ , eq 8, Materials and Methods) for the SL positions N38C (blue), N200C (pink), T288C (green), and the combined data (black symbols, connected by a line). (E, F) *Q* factors of the combined data for the three SL positions with each symbol representing different SL (E) or side-chain (F) orientations. The black diamonds connected by the lines indicate the averages.

of *Cc*. CcP in the encounter complex requires an average rotation of 35° to reach the orientation found in the specific complex (histogram in Figure 3D) and explores a larger fraction of the surface of its partner, with 85% of the structures sampling 25% of the space.

The structures in the calculated ensemble were also compared with the specific complex by making use of the distance root mean square deviation metric (DRMS), according to eq 3 (Materials and Methods). Figure 3F shows the plot of DRMS against electrostatic energy of the ensemble structures. Both the energy and the DRMS distributions can be fitted with Gaussians. The average DRMS is about 6 Å.

**Experimental and Calculated PREs.** The observed PRE ( $\Gamma_2^{\text{obs}}$ ) is an ensemble average of the contributions arising from *Cc* orientations in the encounter and specific complexes, which can be back-calculated from, respectively, the simulated ensemble and the crystal structure of the complex. In Figure 4 the  $\Gamma_2^{\text{calcd}}$  values from the ensemble and the specific complex are compared with the experimental  $\Gamma_2^{\text{obs}}$  for three SL positions (N38C, N200C, and T288C). Neither the specific complex nor the encounter complex alone accounts for the experimental PRE profiles. For T288C (Figure 4A) the  $\Gamma_2^{\text{calcd}}$  values for the specific form deviate from  $\Gamma_2^{\text{obs}}$  for residues 2–8, 12–14, 67–69, and 87–90. Similarly,  $\Gamma_2^{\text{calcd}}$  for the encounter complex disagrees with  $\Gamma_2^{\text{obs}}$  for residues 2–15, 51–53, 67–69, and 87–93. However, a linear combination of the  $\Gamma_2^{\text{calcd}}$  values for the two forms yields a good fit of the experimental data. Similar differences between

observed and calculated PREs for either the simulated ensemble or the specific form are also observed for *Cc* complexes with N200C or N38C CcP-MTS (Figure 4B and 4C). Again, a combination of both PRE contributions gives an improved match with the experiment. We conclude that the MC simulations provide a good description of the encounter complex.

**Population of the Encounter Complex.** The linear combination of  $\Gamma_2^{\text{calcd}}$  for the simulated ensemble and the specific complex that fits best the experimental data was determined by varying the population fraction of the encounter complex ( $p$ ) (eq 8, Materials and Methods). To assess the goodness of the fit, we used a quality factor (*Q*) as defined by eq 6 (Materials and Methods). The plots of *Q* versus  $p$  show a clear minimum for each of the three SLs that cause PREs. The position of the minimum appears to be identical for the three mutants (T288C, N200C, and N38C) with the lowest *Q* factor found at  $p = 0.3$ . According to this analysis, *Cc* and CcP spend 70% of the lifetime of the complex in the crystallographic orientation (the specific complex) and 30% in the encounter complex (Figure 4D). To establish whether the results are influenced by the rotational freedom of the attached SLs, the *Q* factor was calculated using different sets of orientations of the SL, and the results were found to be the same (Figure 4E). The MC calculations were also repeated after randomizing the side-chain orientations of both proteins. It was found that different side-chain orientations do not affect the calculated ensemble

significantly and yield the same *p* value for the encounter complex (Figure 4F).

## Discussion

Our results show that the encounter complex represents 30% of the entire complex, implying a mere 0.5 kcal/mol difference in the free energy relative to the specific form. The delicate balance between the two states could be related to the function of the complex. On one hand, studies of non-ET complexes reported encounter population fractions of up to 10%.<sup>14,21</sup> On the other hand, small ET protein pairs have been described in which no specific complex seemed to be formed at all. For example, myoglobin–cytochrome *b*<sub>5</sub><sup>4,6,8,46</sup> and adrenodoxin–Cc<sup>3,7</sup> represent highly dynamic complexes that consist entirely of encounter complex.

For sufficiently fast biological ET ( $10^3$ – $10^4$  s<sup>-1</sup>), the maximal distance between redox centers needs to be about 16 Å.<sup>47</sup> It has been established that Trp191 acts as a hopping station for electron tunneling between the Cc and the CcP hemes.<sup>48,49</sup> In the specific complex the edge-to-edge distance between the Cc heme and the Trp191 indole group is 13.7 Å, which enables efficient intermolecular ET. In contrast, the overwhelming majority of the structures in the encounter complex ensemble are incapable of fast ET because the distance between the Cc heme and the CcP Trp 191 is more than 16 Å, with the mode being 24 Å (Figure 3G). We propose that, in contrast to the small complexes mentioned above, in big complexes the distance

between the redox centers is too large for efficient ET in most orientations of the encounter complex, which makes it necessary to form a specific complex.

Another important requirement for an ET complex is a rapid turnover. The strong electrostatic attraction and preorientation in the Cc–CcP complex ensures a high association rate<sup>50</sup> but, at the same time, stabilizes the encounter complex.<sup>45</sup> Thus, the difference in the free energy between the specific and the encounter complexes should remain small. A tight binding of the specific form would limit the turnover by a low dissociation rate. In conclusion, the nearly equally populated encounter and specific complexes may well be a consequence of biological function of the Cc–CcP complex, striking the right balance between fast ET and a high turnover rate.

**Acknowledgment.** We thank Francesco Scarpelli for EPR measurements. Q.B. was supported by a fellowship from the Higher Education Commission of Pakistan. A.V. and M.U. were supported by The Netherlands Organisation for Scientific Research, VIDI grant 700.52.425. G.M.U. was supported by the Deutsche Forschungsgemeinschaft, grant UL 174/7-1.

**Supporting Information Available:** Figures showing the SL coverage with the accompanying Pymol files, a stereorepresentation showing distributions of Cc and CcP around the partner protein in the simulated encounter complex superimposed on the specific complex with the accompanying PDB file, PREs plotted per residue for all 10 spin label positions on CcP, and complete ref 37. This material is available free of charge via the Internet at <http://pubs.acs.org>.

JA9064574

(46) Liang, Z.-X.; Jiang, M.; Ning, Q.; Hoffman, B. *J. Biol. Inorg. Chem.* **2002**, *7*, 580–588.

- (47) Moser, C. C.; Keske, J. M.; Warncke, K.; Farid, R. S.; Dutton, P. L. *Nature* **1992**, *355*, 796–802.
- (48) Miller, M. A.; Vitello, L.; Erman, J. E. *Biochemistry* **1995**, *34*, 12048–12058.
- (49) Hays Putnam, A.-M. A.; Lee, Y.-T.; Goodin, D. B. *Biochemistry* **2009**, *48*, 1–3.

(50) Schreiber, G.; Haran, G.; Zhou, H.-X. *Chem. Rev.* **2009**, *109*, 839–860.

(51) <http://www.pymol.org>.